Encyclopedic Dictionary of Mathematics

Second Edition

Encyclopedic Dictionary of Mathematics

Second Edition

by the
Mathematical Society
of Japan

edited by
Kiyosi Itô

Volume I
A–N

The MIT Press
Cambridge, Massachusetts,
and London, England

First MIT Press paperback edition, 1993

Originally published in Japanese in 1954 by Iwanami Shoten, Publishers, Tokyo, under the title *Iwanami Sūgaku Ziten*.

This book was set in Monolaser Times Roman by Asco Trade Typesetting Ltd., Hong Kong, and printed and bound by Arcata Graphics, Kingsport in the United States of America.

Library of Congress Cataloging-in-Publication Data

Iwanami sūgaku jiten (ziten). English.
Encyclopedic dictionary of mathematics.
Translation of: Iwanami sūgaku jiten.
Includes bibliographies and indexes.
1. Mathematics—Dictionaries. I. Itô, Kiyosi, 1915- . II. Nihon Sūgakkai. III. Title.
QA5.I8313 1986 510'.3'21 86-21092
ISBN 0-262-09026-0 (HB), 0-262-59020-4 (PB)

Contents

Foreword to the Second English Edition

The Mathematical Society of Japan takes pleasure in presenting this second edition of our *Encyclopedic Dictionary of Mathematics* to every researcher and user of mathematics. It is intended to be a compact, up-to-date source of information comprising, as completely as possible, all significant results in all fields of our Science, pure and applied, from the elementary to the advanced level. The success of the first edition owed much to the kind assistance given by the American Mathematical Society. As described in the preface, the members of our Society have taken responsibility for compiling this new edition. We hope that it will be as useful to the mathematicians of today as the first edition was to the mathematicians of yesterday. We also hope that this edition will be followed in years to come by subsequent ones incorporating the future development of our Science.

Hikosaburo Komatsu
President 1985–1987
Mathematical Society of Japan

Foreword to the First English Edition

The American Mathematical Society welcomes the publication of the *Encyclopedic Dictionary of Mathematics*. For many years we have been fascinated by the publication in Japanese, *Iwanami Sūgaku Ziten*, because we saw that this was an encyclopedia that contained effective and penetrating information about all the fields of advanced mathematical research. We were also frustrated because we could not read Japanese and so we could not really reach out to this expert and effective source of information. We now welcome the fact that the second Japanese edition has been translated into English and we look forward to the fascination which we can now have in getting at this rich mine of information.

Saunders MacLane
President 1973–1974
American Mathematical Society

Preface to the Second English Edition

This second edition of the *Encyclopedic Dictionary of Mathematics* is in substance an English version of the third edition of *Iwanami Sūgaku Ziten* (in Japanese). We shall explain how these two versions are related to each other and how the present edition differs from the first English edition. For the sake of simplicity we abbreviate *Encyclopedic Dictionary of Mathematics* and *Iwanami Sūgaku Ziten* to EDM and ISZ, respectively, and indicate the numbers of editions beyond the first by Arabic numerals.

The prefaces of the previous editions will clarify how ISZ, its augumented edition, ISZ2, ISZ3, and EDM came into existence in 1954, 1960, 1968, 1985, and 1977, respectively. EDM, published nine years later than ISZ2, consisted of its English translation and some new materials added to update its contents. In retrospect, it was felt that this procedure was not adequate to cope with the recent rapid progress of mathematics, and a suggestion was raised in the Mathematical Society of Japan that ISZ3 and EDM2 be produced simultaneously. The favorable reaction of the mathematical public to EDM encouraged us greatly.

In 1978 an agreement was made among the Society, Iwanami Shoten, Publishers, and The MIT Press for the publication of ISZ3 and EDM2. An editorial committee was established in the Society with the members named in the preface to ISZ3. Manuscripts were prepared simultaneously in Japanese and English by each contributor, with Yuji Ito acting as linguistic consultant.

I should mention that we benefited greatly from the kind comments on EDM by the following mathematicians: J. F. Adams, M. Atiyah, A. Borel, H. Cartan, K. Chandrasekharan, S. S. Chern, J. Dieudonné, E. Hewitt, F. Hirzebruch, O. Lehto, J. L. Lions, L. Markus, and J. P. Serre. In particular, we are deeply grateful to J. Dieudonné for his many detailed comments.

Compared with ISZ3 and EDM, EDM2 has the following characteristics:

(1) EDM2 contains many more new mathematical results than EDM. For the details we refer the reader to the main points of revision mentioned in the preface to ISZ3, because EDM2 has the same mathematical content as ISZ3 and the additions made to ISZ2 and subsequently in EDM are of relatively minor importance.

(2) The Japanese textbooks listed in the article references of ISZ3 are replaced in EDM2 by references to standard textbooks written in English. Such replacement was not done in compiling EDM.

(3) Years of birth and death are included in the Name Index of EDM2 as far as possible. This information was given in ISZ2 but was removed in compiling EDM.

(4) The Subject Index of EDM2 is so designed that every concept consisting of two or more words can be traced from each of these component words. (This principle was adopted in ISZ2 and EDM but was not rigidly followed in ISZ3 due to the lack of space.)

(5) While all editions of ISZ were in one volume and EDM was in two volumes, EDM2 is in four volumes: I. Forewords, prefaces, introduction, text A–E; II. Text F–N; III. Text O–Z; IV. Appendices, indexes and other backmatter. The Systematic List of Articles appears in volume IV, page 1857.

While EDM2 is more voluminous than ISZ3, we hope that, being written in English and printed with generous margins, it can easily provide readers with information on every significant result of today's mathematics and so will be useful to a large number of mathematicians. I am responsible for any shortcomings that may exist despite all our efforts, and I would appreciate any remarks from the readers. I hope to have occasion to remedy any such shortcomings in the future.

Finally I would like to repeat my thanks to all the collaborators named in my preface to ISZ3. Also I express my gratitude to Yuji Ito for his tremendous effort to polish the English of all manuscripts, to Shigeru Iitaka for his laborious assistance in the final stage of the work, to Mrs. M. Nawata for her excellent secretarial work, and to the staff of the Department of Mathematics at Gakushuin University for providing me with an office for compilation for eight years.

Kiyosi Itô
December 1986

Preface to the Third Japanese Edition

The first edition of *Iwanami Sūgaku Ziten* was published in 1954, and a revised and augmented edition appeared in 1960. Extensive revision of the work was carried out subsequently, and the result was published in 1968 as the second edition, which has retained its valuable and useful life for the past 17 years. The English translation of this second edition published by The MIT Press has achieved international recognition. In the meantime, remarkable progress has been made in mathematical sciences: recognition of the interrelationships amongst various branches within mathematics has increased significantly, and the formulation of mathematics as a synthetic entity is in the making. Furthermore, advanced mathematical theories have been utilized frequently in the physical, biological, and social sciences, and expectations for mathematics to be the basis of all the sciences have been increasing. In order to cope with these developments, it was decided that the second edition of *Sūgaku Ziten* should be further revised and an updated version be published as the third edition. The English translation of this third edition will also be published by The MIT Press.

The main points of the revision are as follows:

(1) On the Size and Scope of the Encyclopedia: The total number of articles has increased only slightly from that of the second edition, from 436 to 450. However, in view of the rapid and extensive development in mathematics in recent years, many old articles have been reorganized and unified, and much new material has been added. Consequently, the mathematical content of the encyclopedia has increased a great deal over that of the second edition, and this, together with the authors' efforts to render the articles more readable than before, necessitated a 50% increase in the number of pages devoted to the main text. The account of computer science has been taken out of the subject area of Numerical Analysis, and a new subject area has been created for it; thus mathematics is now classified into 21 subject areas altogether instead of 20.

(2) On the Arrangement of the Articles: In the second edition, the title of each article was spelled out in Romanized Japanese, and articles were arranged alphabetically. In the third edition, however, the titles of the articles are arranged in the order of the Japanese syllabary (*katakana*) and each article is numbered accordingly. To each article a parenthetical number is also attached, giving the subject area to which the article belongs and its place in that list.

(3) On the Text: Sections within each article are headed by letters A, B, C, ...; the initial section of each article gives an outline and is devoted to an introductory and general account of the topic with which the article is concerned. In the actual exposition of the articles, every effort has been made to pay close attention to the interrelationships amongst various fields of mathematics. Even when there was no need to revise the original article, changes were made in the bibliography to list items that would be more accessible to the readers.

(4) On the Appendices: The appendices have been revised and augmented along with the text. Numerical tables were deleted whenever it was felt that scientific calculators or microcomputers could easily reproduce their contents.

(5) On the Indexes: The listing of mathematical terms in Japanese has been changed, as in the case of the arrangement of articles, from the alphabetical order to the order of the Japanese syllabary. However, terms in Japanese that start with a European word have been separated out, and, in order to facilitate finding, have been listed independently in an Index for Mathematical Terms in Japanese with European Headings. The location of a term listed in the indexes is indicated by the article number and the letter heading of the section in the article where the term appears.

We here outline briefly how the process of compilation of this edition was organized. In the summer of 1978 the Mathematical Society of Japan decided, following a proposal by Professors Shôkichi Iyanaga and Yukiyosi Kawada, the chief editors of the previous editions of *Sūgaku Ziten*, to undertake the compilation of the third edition. Following a resolution by the members of the Society, Kunihiko Kodaira, Sigeru Mizohata, Itiro Tamura, Nagayoshi Iwahori, Tosihusa Kimura, and myself have joined Iyanaga and Kawada to form an editorial committee and have formulated basic guidelines for the compilation of the third edition. The work of editing and compiling articles in each subject area was delegated to the editors designated for the areas, and I was asked to undertake the responsibility of putting the entire volume in order.

The subject areas and their editors are as follows:

I Logic and Foundations: Shôji Maehara
II Sets, General Topology, and Categories: Shôji Maehara, Hikosaburo Komatsu, Masayoshi Nagata
III Algebra: Masayoshi Nagata
IV Group Theory: Nagayoshi Iwahori
V Number Theory: Takayoshi Mitsui
VI Euclidean and Projective Geometry: Itiro Tamura
VII Differential Geometry: Morio Obata, Shigeru Iitaka, Itiro Tamura
VIII Algebraic Geometry: Shigeru Iitaka
IX Topology: Itiro Tamura
X Analysis: Sigeru Mizohata, Kiyosi Itô
XI Complex Analysis: Yukio Kusunoki, Shigeru Iitaka
XII Functional Analysis: Hikosaburo Komatsu
XIII Differential, Integral, and Functional Equations: Tosihusa Kimura, Sigeru Mizohata
XIV Special Functions: Sin Hitotumatu
XV Numerical Analysis: Masaya Yamaguti
XVI Computer Science and Combinatorics: Sin Hitotumatu
XVII Probability Theory: Kiyosi Itô
XVIII Statistics: Kei Takeuchi
XIX Mathematical Programming and Operations Research: Shigeru Furuya
XX Mechanics and Theoretical Physics: Huzihiro Araki
XXI History of Mathematics: Shôkichi Iyanaga
Appendices: Sin Hitotumatu

Before the actual process of compilation started, the editorial committee had met several times, and two meetings of all the subject-area editors had been called. Furthermore, editors for each area consulted with other experts in their fields as they selected the titles of articles to be included. In areas where remarkable development had taken place, such as differential geometry, functional analysis, probability theory, and applied mathematics, extensive revision and insertion of many new articles were proposed by the respective editors. For this reason, the original articles were thoroughly reorganized and systematized. We resolved to keep the whole work in one volume, even though a number of new articles have had to be added.

The selection of articles and of their respective authors was finally completed in the spring of 1980, and 197 colleagues were asked to write the articles. The names of these contributors and those of the previous editions are listed elsewhere in this volume. To all of the authors we express our sincere gratitude.

Editing of the entire manuscript has been carried out since the summer of 1982, and we are deeply indebted to the following colleagues for their painstaking efforts in checking cross references and in proofreading: Nobuyosi Motohashi (Foundations, Set Theory), Takeo Yokonuma (Algebra, Group Theory), Takayoshi Mitsui (Number Theory), Tetsuro Kawasaki (Geometry, Topology), Shigeru Iitaka, Isao Wakabayashi (Algebraic Geometry), Morio Obata, Koichi Ogiue, Osamu Kobayashi (Differential Geometry), Seizô Ito, Hisashi Okamoto (Analysis), Hiroaki Aikawa, Makoto Ohtsuka (Computer Analysis), Hikosaburo Komatsu, Akihiko Miyachi (Functional Analysis), Kazuo Okamoto, Daisuke Fujiwara (Differential Equations), Sin Hitotumatu (Special Functions), Teruo Ushijima (Numerical Analysis), Hideo Wada (Computer Science), Yasunori Okabe (Probability Theory), Mituaki Huzii, Yoshihiro Yajima (Statistics), Shigeru Furuya (Mathematical Programming), Koichi Nakamura (Theoretical Physics), Shûichi Okamoto (History of Mathematics), Kosaku Okutsu (History of Mathematics, Number Theory).

On those occasions when it became necessary to rapidly revise manuscripts in order to unify the presentation, we were forced to go through the revision without consulting the authors of the manuscripts. I am responsible for all such revisions and hereby express my apologies to the authors concerned.

As for the indexes, we received assistance from Takeo Yokonuma, Koichi Yano, Hiroaki Aikawa, and Hisashi Okamoto in all aspects of the work, which lasted for two years, through the final proofreading. Preliminary work on the Name Index was carried out by staff members of the University of Tokyo, and Seizô Ito supervised its final compilation. We also asked many other colleagues, and in particular Nobuyuki Ikeda and Huzihiro Araki, to participate in the preparation of the Name Index, using resources available at different universities. To all of these people goes our sincere gratitude for their assistance.

Last August, when the completion of the work was drawing near, it became necessary for me to leave Japan for some time, and this made it imperative that I entrust Shigeru Iitaka with responsibility for supervising the work at the final stage of editing. I am most grateful to Iitaka for having agreed to take over this responsibility and for having brought the work to completion under the pressure of time.

From the beginning of this project, we have received an unlimited amount of assistance from the members of the Dictionary Department of Iwanami Shoten, Publishers. Messrs. Ikutaro Sasaki, Hirotomo Ushida, Kazuhiko Uetake, and Nagao Sato in particular, have made supreme efforts and have come up fre-

quently with innovative ideas to make this dictionary polished and perfect. To them and also to those at Dai-Nippon Printing Co., who typeset with the use of computers the entire text, including the complicated mathematical formulas, and who have cooperated with us fully in the process of editing the indexes, goes our deep gratitude.

Kiyosi Itô
October 1985

Preface to the First English Edition

The first and second editions of *Iwanami Sūgaku Ziten* (in Japanese) were published, respectively, in April 1954 and June 1968 by Iwanami Shoten, Publishers, Tokyo. Beginning in the late 1950s, a number of unsuccessful attempts were made to arrange for translating the *Sūgaku Ziten* into European languages. Finally an agreement for an English translation was made between The MIT Press and the Mathematical Society of Japan in July 1968. The discussions were carried on first by Professor Kôsaku Yosida, then president of the Mathematical Society of Japan, and later by Yukiyosi Kawada, who succeeded him in April 1968. Throughout these initial negotiations, which lasted from 1966 to 1968, we received the kindest assistance from Dr. Gordon Walker, Executive Director of the American Mathematical Society, and from Professors W. T. Martin and Shizuo Kakutani.

The agreement for the project was shortly followed by the establishment of a committee for the English edition of *Sūgaku Ziten* within the Mathematical Society of Japan, with the following membership: Professors Yasuo Akizuki, Shigeru Furuya, Sin Hitotumatu, Masuo Hukuhara, Isao Imai, Shôkichi Iyanaga, Yukiyosi Kawada, Kunihiko Kodaira, Atuo Komatu, Hirokichi Kudō, Shôji Maehara, Yukio Mimura, Kiyoshi Noshiro, Shigeo Sasaki, Shoji Ura, Nobuo Yoneda, and Kôsaku Yosida. This committee requested the original authors of the articles and other members of the Society to translate the work. A list of translators will be found at the end of this work.

In November 1968, an advisory committee for the project was formed with the following membership: Professor Edwin Hewitt (chairman), Dr. Sydney H. Gould, Professor Shizuo Kakutani, Professor Kenneth O. May, and Professor Isaac Namioka.

As the translating began, we were immediately faced with problems concerning unification of terminology and style, some of which were inherent in the differences between the structures of our two languages—for example, the fact that the Japanese language makes no distinction between singular and plural forms of nouns.

In August 1969, Professor Hewitt kindly arranged a meeting at the University of Washington, Seattle, that included the members of the Japanese and American committees and a representative of The MIT Press. It was agreed during this meeting that the transla-

tion should be faithful, with only a minimum number of changes, such as the correction of mathematical errors; whereas the references to each article might be augmented considerably for the convenience of Western readers. Professor Kenneth O. May volunteered to review the entire translated manuscript, and Professors Isaac Namioka and Shizuo Kakutani, who are proficient in both Japanese and English, proposed to read through some of the manuscript of the translated articles. It was also agreed that the Systematic List of Articles should appear in French, German, and Russian, as well as in English.

We owe very much to the American committee: Professor Hewitt organized the whole work, and Professor May revised the entire manuscript and gave us important advice concerning the appendices, according to which we deleted some of the numerical tables which may be easily found in readily accessible Western books. Professor Namioka reviewed a great part of the manuscript, transmitting his views to Professor May, who forwarded them to us with his comments. All of this assistance helped us greatly in making our final decisions. Professor Kakutani gave us very detailed and important advice on the choice of reference works.

We were also assisted concerning English terminology and reference books by the following Japanese mathematicians working in American universities: Tadatoshi Akiba, Professors Kiyosi Itô, Tatsuji Kambayashi, Tosio Kato, Teruhisa Matsusaka, Katsumi Nomizu, Ichiro Satake, Michio Suzuki, and Gaisi Takeuti.

The Mathematical Society of Japan established the following double reviewing system: group A with its twenty subgroups, each headed by the members listed below, reviewed their respective subjects; while group B reviewed the whole manuscript, mainly from the linguistic standpoint.

Group A

1. Foundations of mathematics: Shôji Maehara
2. Set theory: Atuo Komatu and Shôji Maehara
3. Algebra: Akira Hattori, Masayoshi Nagata, and Hideyuki Matsumura
4. Group theory: Shingo Murakami, Mitsuo Sugiura, and Reiji Takahashi
5. Number theory: Yukiyosi Kawada and Tomio Kubota
6. Geometry, 7. Differential geometry: Shôkichi Iyanaga, Shigeo Sasaki, and Kentaro Yano
8. Algebraic geometry: Yasuo Akizuki and Kunihiko Kodaira
9. Topology: Atuo Komatu
10. Real analysis: Sin Hitotumatu, Shunji Kametani, and Shigeki Yano
11. Complex analysis: Kiyoshi Noshiro
12. Functional analysis: Yukio Mimura and Kôsaku Yosida
13. Differential equations: Masuo Hukuhara and Sigeru Mizohata
14. Special functions: Sin Hitotumatu
15. Numerical analysis, 16. Probability theory: Kiyosi Itô
17. Statistics: Hirokichi Kudō
18. Information theory: Tosio Kitagawa and Hirofumi Uzawa
19. Theoretical physics: Isao Imai and Kazuhiko Nishijima
20. History of mathematics: Tamotsu Murata

Group B

Kenichi Iyanaga and Mitsuyo Iyanaga

Professor Sin Hitotumatu also assisted us in translating the titles of Japanese books given in the references and the explanations attached to the lists of formulas and numerical tables in the appendices. We are also grateful for the generous cooperation offered to us by our colleagues in the Department of Mathematics, Faculty of Science, University of Tokyo: Hiroshi Fujita, Shigeru Furuya, Akio Hattori, Seizô Itô, Nagayoshi Iwahori, Tosihusa Kimura, Kunihiko Kodaira, Hikosaburo Komatsu, Akihiro Nozaki, and Itiro Tamura. We are indebted as well to Professors Walter L. Bailey of the University of Chicago and Yuji Ito of Brown University for many valuable consultations concerning both mathematical and linguistic questions.

In translating the Systematic List of Articles into French, German, and Russian, we were assisted by Professor Hideya Matsumoto in Paris, Professor Emanuel Sperner in Hamburg, Professor Katsuhiro Chiba in Tokyo, and Professor Arkadiĭ Maltsev in Moscow.

We began to send the manuscript to The MIT Press in March 1970 and finished sending it in July 1972. The manuscript was edited there, then sent to Professors May, Namioka, and Kakutani, and finally was sent back to us with their comments and questions. All of the references were carefully checked by Laura Platt.

Iwanami Shoten, Publishers, have always been cooperative with us. In the office of the Mathematical Society of Japan, Yōko Endo, Reiko Nagase, and Chieko Sagawa helped us with their efficient secretarial work.

The fruition of this project was made possible only by the gracious assistance offered to us by many people, including those already mentioned. We should like to express our most sincere gratitude to all those who have helped us so kindly.

Shôkichi Iyanaga, Yukiyosi Kawada
August 1973

Addition made in January 1976

After the procedures described above, the whole manuscript of this Encyclopedia was sent to The MIT Press in August 1973. It was, however, toward the end of November 1975 that the final decision was made by The MIT Press to send the manuscript to composition in early 1976 in order that the work be published in 1977.

At the same time, we were asked to review and update the manuscript up to the end of February 1976. We are now making our best effort to this effect with the kind help, especially from the linguistic viewpoint, of Dr. E. J. Brody.

In so doing, we have noticed that perhaps too much emphasis has been given to results obtained by Japanese mathematicians and that there are still many things in this book which should be improved.

We hope that ongoing revisions will be carried out in subsequent editions.

Preface to the Second Japanese Edition

Seven and a half years have passed since the revised and augmented edition of *Iwanami Sūgaku Ziten* was published. The nature and purpose of this book remain the same as described in the preface to the first edition: It is an encyclopedic dictionary with articles of medium length aimed at presenting the whole of mathematics in a lucid system, giving exact definitions of important terms in both pure and applied mathematics, and describing the present state of research in each field, together with historical background and some perspectives for the future. However, mathematical science is in rapid motion, and the "present state of research" changes constantly. The present updated second edition has been published to remedy this situation as far as possible.

The main points of revision are as follows:

(1) On the Articles and the Size of the Encyclopedia: From the articles of the last edition, we have removed those whose importance has diminished recently (e.g., Geometry of Triangles), while we have added new articles in domains of growing importance (e.g., Categories and Functors; K-Theory). Many articles concerning applied mathematics in the first edition were short; in this edition, they have been combined into articles of medium length to save space and to systematize the presentation. The number of articles, 593 in the first edition, has thus diminished to 436. We have made every effort to keep the size of the encyclopedia as it was, but the substantial augmentation of content has necessarily brought about an enlargement of about 30%.

(2) On the Text: When the title of an article has remained the same as in the first edition, we have reviewed the whole text and revised whenever necessary. Especially for the fundamental ideas, we have endeavored to give thorough explanations. In the first edition we gave English, French, and German translations of article titles; in the present edition, we have also added Russian. The bibliographies at the ends of articles have been updated.

(3) On Terminology: In previous editions we endeavored to unify the terminology of the whole encyclopedia so that the reader would have no difficulties with cross-references. Here we have done this once again with the hope of attaining results more perfect than before.

(4) On the Appendices: The appendices were designed to supplement the text efficiently. Some overlapping of the appendices with the text found in previous editions has been removed. Also deleted in this edition are elemen-

tary formulas in analytic geometry and tables of statistical distributions, which can be easily found in other books. However, we have added some formulas in topology, the theory of probability, and statistics, as well as tables of characters of finite groups, etc.

(5) On the Indexes: Important terms are listed multiply in the indexes to facilitate finding, e.g., the term *transcendental singularity* appears under both *transcendental* and *singularity*. Both names in the text and those in the references are included in the Name Index of this edition. The numbers of items in the Index of Mathematical Terms in Japanese and in European Languages and in the Name Index in this edition are 17740, 10124, and 2438, compared with 8254, 8070, and 1279, respectively, in the previous edition.

The compilation of this edition was organized as follows. In the spring of 1964 we began to select the titles of articles with the aid of the following colleagues: in set theory and foundations of mathematics, Shôji Maehara; in algebra and number theory, Yasuo Akizuki, Yukiyosi Kawada; in differential geometry, theory of Lie groups, and topology, Yozo Matsushima, Atuo Komatu; in analysis, Masuo Hukuhara, Kôsaku Yosida, Shunji Kametani, Sin Hitotumatu; in probability theory, statistics, and mathematics for programming, Kiyosi Itô, Hirokichi Kudō, and Shigeru Furuya; in theoretical physics, Isao Imai; and for the appendices, Sin Hitotumatu. I have participated in compiling the articles on geometry and the history of mathematics. Kawada and Hitotumatu undertook the responsibility of putting the volume in order.

The work of selecting titles was completed in the summer of 1964. We then asked 173 colleagues to contribute articles. The names of these contributors and those of the previous editions are listed elsewhere. To all of them goes our most sincere gratitude.

In editing the manuscript, we were assisted by the following colleagues in addition to those mentioned already: in set theory and foundations of mathematics, Setsuya Seki and Tsurane Iwamura; in algebra and number theory, Masayosi Nagata, Akira Hattori, Hideyuki Matsumura, Ichiro Satake, Tikao Tatuzawa; in geometry, theory of Lie groups, and topology, Singo Murakami, Hideki Ozeki, Noboru Tanaka, Kiiti Morita, Hirosi Toda, Minoru Nakaoka, Masahiro Sugawara, Shôrô Araki; in analysis, Kiyoshi Noshiro, Yûsaku Komatu, Seizo Ito, Hiroshi Fujita, Shige-Toshi Kuroda, Sigeru Mizohata, Masaya Yamaguti, Tosiya Saito, Tosihusa Kimura, Masahiro Iwano; in probability theory, statistics, and mathematics of programming, Nobuyuki Ikeda, Tadashi Ueno, Masashi Okamoto, Haruki Morimoto, Kei Takeuchi, Goro Ishii, Tokitake Kusama, Hukukane Nikaido, Toshio Kitagawa; and in theoretical physics, Ryogo Kubo, Hironari Miyazawa, Yoshihide Kozai.

After the summer of 1965, we entered into the period of finer technical editing, in which we were assisted by the following colleagues: in algebra, Keijiro Yamazaki, Shin-ichiro Ihara, Takeshi Kondo; in geometry, Tadashi Nagano, Mitsuo Sugiura, Ichiro Tamura, Kiyoshi Katase; in analysis, Nobuyuki Suita, Kotaro Oikawa, Kenkiti Kasahara, Tosinobu Muramatsu, Hikosaburo Komatsu, Setuzô Yosida, Hiroshi Tanaka; and in history, Tamotsu Murata.

We were also assisted by Katsuhiro Chiba in the Russian-language translation of the article titles, and by Osamu Kôta and Kiyoshi Katase in the indexing.

Proof sheets began appearing in the spring of 1966. In proofreading, Kaoru Sekino, Osamu Kôta, Kiyoshi Katase, and Teruo Ushijima helped us, as well as Mrs. Rieko Fujisaki. Miss Yoko Endo worked with all of us throughout the entire revision at the office of the Mathematical Society of Japan. She helped us especially in looking for and checking references and preparing the Name Index.

Yukiyosi Kawada supervised the whole work, succeeding me in the role I had played in the compilation of the first and augmented editions. Sin Hitotumatu collaborated with him throughout, especially on the appendices. The second and the third proof sheets of the text were read by Kawada; the fourth proof by Hitotumatu; the proof sheets of appendices by both Kawada and Hitotumatu.

The Editorial Committee of the Mathematical Society of Japan asked me to write this preface. Having edited the previous editions and realizing full well the difficulty of the task, I would like to express my particular gratitude to Kawada. In the Dictionary Department of Iwanami Shoten, Publishers, Messrs. Hiroshi Horie, Tetsuo Misaka, Shigeki Kobayashi, and Toshio Kouda were very cooperative in their collaboration with us. To them and also to those who typeset and printed this book at Dai-Nippon Printing Co. and Shaken Co. goes our gratitude.

S. Iyanaga
March 1968

Preface to the Revised and Augmented Japanese Edition

Six years have passed since the publication of the first edition of this encyclopedia. This revised and augmented edition incorporates the achievement of these years. It contains, together with the correction of errors found in the first edition, some new articles such as Abelian Varieties, Automata, Sheaves, Homological Algebra, Information Theory, and also supplements to articles in the first edition such as Complex Multiplication, Computers, and Manifolds. These additional items render the new edition 93 pages longer than the previous one. Each article has been thoroughly revised, and the indexes have been completely rewritten.

We were assisted by the following colleagues in selecting articles, writing, and proofreading: In set theory and foundations of mathematics, Sigekatu Kuroda, Setsuya Seki; in algebra and number theory, Tadao Tannaka, Tsuneo Tamagawa: in real analysis, Shunji Kametani, Kôsaku Yosida; in function theory, Kiyoshi Noshiro, Sin Hitotumatu; in theory of differential and functional equations, Masuo Hukuhara, Masahiro Iwano, Ken Yamanaka; in functional analysis, Kôsaku Yosida, Seizô Itô; in geometry, Shigeo Sasaki, Nagayoshi Iwahori; in topology, Atuo Komatu, Itiro Tamura, Nobuo Yoneda; in theory of probability, Kiyosi Itô, Seizô Itô; in statistics, Toshio Kitagawa, Sigeiti Moriguti, Tatsuo Kawata; in applied mathematics, Sigeiti Moriguti; and in mechanics and theoretical physics, Takahiko Yamanouchi, Isao Imai. The revision and augmentation of the articles concerning the history of mathematics was done by myself. Mlles. Yôko Tao, Eiko Miyagawa, and Mutsuko Nogami worked in the office of the Society.

The names of authors who contributed to the completion of this edition have been added to the original list of contributors.

The project of editing this edition started in the summer of 1958. We acknowledge our deep gratitude to all those who have collaborated with us since that time.

S. Iyanaga
August 1960

Preface to the First Japanese Edition

This encyclopedia, *Iwanami Sūgaku Ziten*, was compiled by the Mathematical Society of Japan at the request of the Iwanami Shoten, Publishers, who have hitherto published a series of scientific dictionaries such as *Iwanami Rikagaku-Ziten* (*Iwanami Encyclopedia of Physics and Chemistry*) and *Iwanami Tetugaku-Ziten* (*Iwanami Encyclopedia of Philosophy*). As mentioned in the prefaces to these volumes, the importance of such encyclopedias in clarifying the present state of each science is obvious if we observe the rapid pace of contemporary research. Mathematics is also in rapid motion. As a fundamental part of exact science, it serves as a basis of all science and technology. It also retains its close contact with philosophy. Therefore, the significance of having an encyclopedia of mathematics cannot be overemphasized.

Mathematics have made remarkable progress in the 20th century. As for the situation toward the end of the 19th century, we quote the following passage from the article Mathematics in the 19th Century of this encyclopedia: "Toward the end of [the 19th] century, the subjects of mathematical research became highly differentiated. Branches split into more specialized areas of study, while unexpected relations were found between previously unconnected fields. The situation became so complicated that it was difficult to view mathematics as a whole. It was in these circumstances that in 1898, at the suggestion of Franz Meyer and under the sponsorship of the Academies of Göttingen, Berlin, and Vienna, a project was initiated to compile an encyclopedia of the mathematical sciences. Entitled the *Enzyklopädie der mathematischen Wissenschaften*, it was completed in 20 years...."

One of the characteristics of 20th-century mathematics is the conscious utilization of the axiomatic method and of general concepts such as sets and mappings, which serve as foundations of different theories. Indeed, mathematics is being reorganized on the basis of topology and algebra. One such example of reorganization is found in Bourbaki's *Eléments de mathématique*; some fifteen volumes of this series have been published since 1939, and more are coming. This encyclopedia, with its limited size, cannot contain proofs for theorems. However, we intend to present a lucid view of the totality of mathematics, including its historical background and future possibilities.

Each article of this encyclopedia is of medium length—sufficiently short to permit

the reader to find exact definitions of notions, and sufficiently long to contain explanations clarifying how important concepts in the same field are related to each other. The problem of choosing adequate titles required some deliberation. The Systematic List of Articles, classified according to specific fields, shows those we have chosen. The Index of Terms contains detailed references for each notion. The appendices, including formulas and tables, supplement the text, and will be particularly useful for applied mathematicians.

The project of compiling this encyclopedia was proposed in the spring of 1947 by the Steering Committee of the Mathematical Society of Japan. It was promptly adopted, and the selection of articles in specific fields was started by the sectional committees of the Society. After seven years, our encyclopedia is finally appearing. We shall not give a detailed description of how our work proceeded through all these years. We list simply the names of those who assisted us greatly and to whom we should like to express our deep gratitude.

The president of the Society at the start of this project was the late Professor Tadahiko Kubota; but our work has been supported also by Professors Teiji Takagi, Zyoiti Suetuna, and Masatsugu Tsuji as well as by other leading members of the Society.

At the stage of selecting articles, we were assisted by the following colleagues: in history and the foundations of mathematics, Sigekatu Kuroda, Motokiti Kondo; in algebra and number theory, Kenjiro Shoda, Tadasi Nakayama, Masao Sugawara, Yukiyosi Kawada, Kenkichi Iwasawa; in geometry, Kentaro Yano, Asajiro Ichida; in function theory, Kiyoshi Noshiro, Yûsaku Komatu; in the theory of differential and functional equations, Masuo Hukuhara, Shigeru Furuya; in topology, Atuo Komatu, Ryoji Shizuma; in functional analysis, Yukio Mimura, Shizuo Kakutani, Kôsaku Yosida; in the theory of probability and statistics, Tatsuo Kawata, Toshio Kitagawa, Junjiro Ogawa; and in applied mathematics, Ayao Amemiya, Isao Imai, Kunihiko Kodaira, Shigeiti Moriguti.

We asked 190 colleagues to contribute articles, which were collected in 1949. Since then we have spent an unexpectedly long time editing them. Terminology had to be unified throughout the encyclopedia so that the reader would have no trouble with cross-references. Repetitions had to be eliminated and gaps filled. Part of the manuscript thus had to be rewritten a number of times. Wc have made out utmost effort in this editing work, but we are not completely without apprehension that our result has still left something to be desired. For any shortcomings in the work, I take complete responsibility, as I have acted as the editor-in-chief. Also, since we have rewritten the manuscript, as already mentioned, we have refrained from printing the name of the original author of each article; for this, I must request the understanding of the contributors.

In the stage of editing and proofreading, we were assisted by the following colleagues: Yukio Mimura, Yukiyosi Kawada, Kazuo Matsuzaka, Sin Hitotumatu, Setsuo Fukutomi, Setsuya Seki, Shoji Irie, Shigeo Sasaki, Tatsuo Kawata, Sigekatu Kuroda, Yûsaku Komatu, Ayao Amemiya, Isao Iami, Tosio Kato, Tsurane Iwamura, Morikuni Goto, Kôsaku Yosida, Jirô Tamura, Yasuo Akizuki, Kiyoshi Noshiro, Motosaburo Masuyama, Sigeiti Moriguti, Osamu Kôta, Nobuo Yoneda, Tsuneo Tamagawa, Jun-ichi Hano; and more particularly, in the foundations of mathematics, Sigekatu Kuroda, Tsurane Iwamura; in algebra and number theory, Kazuo Matsuzaka, Yukiyosi Kawada; in algebraic geometry, Yakuo Akizuki; in real analysis, Tatsuo Kawata; in complex analysis, Yûsaku Komatu, Sin Hitotumatu, Jirô Tamura; in functional analysis, Kôsaku Yosida; in topology, Setsuo Fukutomi, Nobuo Yoneda; in the theory of probability and statistics, Motosaburo Masuyama, Sigeiti Moriguti; and in applied mathematics, Ayao Amemiya, Isao Imai, Tosio Kato, Sigeiti Moriguti.

The portraits of Abel and Riemann were kindly loaned to us by Torataro Shimomura.

The formulas in the appendices were compiled by Isao Imai, Sin Hitotumatu, and Sigeiti Moriguti; the Subject Index (in Japanese and European languages) by Osamu Kôta and Mrs. Hiroko Ide; the Name Index and Comments on Journals and Serials by Setsuo Fukutomi. Setsuo Fukutomi has taken an active part in our work ever since 1948 and given much effort to collecting and rewriting the manuscript and to unifying terminology. The editorial staff of Iwanami Shoten, Publishers, has always been cooperative. Without their generous support, this encyclopedia could never have been published.

I should like to express my sincere gratitude to all those who have collaborated with us directly or indirectly.

S. Iyanaga
March 1954

Introduction

The text of this Encyclopedic Dictionary consists of 450 *articles* arranged alphabetically, beginning with 1 Abel and ending with 450 Zeta Functions. Most of these articles are divided into *sections*, indicated by A, B, C, ..., AA, BB, ... Cross-references to articles, e.g., to the second article, are of the form: (→ 2 Abelian Groups) or (→ 2 Abelian Groups A), according as the whole article or a specific section is being referred to. Citations in the indexes are also given in terms of article and section numbers.

Key terms accompanied by their definitions in the text are printed in boldface. All of these terms are found in the Subject Index at the end of the volume.

The sign † means that the term preceded by it can be found in the index. A list of special notations used throughout the work (with explanations of their meanings) appears after the appendices.

A Systematic List of Articles, showing the general structure of the work, will be found on p. 1857 (vol. IV). The number in parentheses after each article title refers to this systematic classification into subject areas; e.g., "Abelian Varieties (VIII.5)" means that the article on Abelian varieties is the fifth article in Section VIII of the systematic list.

Books and articles in journals are cited in the text by numbers in brackets: [1], [2], ... At the end of each article there is a section of references in which, for books, the name of the author or authors, title, name of the publisher, year of publication, and the number of the edition are given; for journals, the name of the author, title of the article, name of the journal, and the volume numbers and inclusive page numbers are given in this order. (The names of journals and publishers are abbreviated as indicated in the lists at the end of the work.)

The Cyrillic alphabet is transliterated as follows:

Cyrillic Alphabet			**Transliteration**
А	а	(a)	a
Б	б	(be)	b
В	в	(ve)	v
Г	г	(ge)	g
Д	д	(de)	d
Е	е	(ye)	e
Ё	ё	(yo)	ë
Ж	ж	(zhe)	zh
З	з	(ze)	z
И	и	(i)	i
Й	й	(i kratkoe)	ĭ
К	к	(ka)	k
Л	л	(el')	l
М	м	(em)	m
Н	н	(en)	n
О	о	(o)	o
П	п	(pe)	p
Р	р	(er)	r
С	с	(es)	s
Т	т	(te)	t
У	у	(u)	u
Ф	ф	(ef)	f
Х	х	(kha)	kh
Ц	ц	(tse)	ts̀
Ч	ч	(che)	ch
Ш	ш	(sha)	sh
Щ	щ	(shcha)	shch
Ъ	ъ	(tvërdyĭ znak)	"
Ы	ы	(yery)	y
(Ь)	ь	(myagkiĭ znak)	'
Э	э	(e)	è
Ю	ю	(yu)	yu
Я	я	(ya)	ya

A

1 (XXI.12)
Abel, Niels Henrik

Niels Henrik Abel (August 5, 1802–April 6, 1829) was born a son of a poor pastor in the hamlet of Findö in Norway. In 1822, he entered the University of Christiania; however, he studied mathematics almost entirely on his own. He was recognized as a promising student by his senior, Holmboe, and after graduation he studied abroad in Berlin and Paris. In Berlin he met and was aided by A. Crelle, the founder of the *Journal für die Reine und Angewandte Mathematik*, and participated in the founding of this journal. Although he did brilliant work in Paris, he did not gain the fame he deserved. He returned to Norway in May 1827, but, unable to find a job, he was obliged to fight poverty while continuing his research. He died at twenty-six of tuberculosis.

His best-known works are the result that algebraic equations of order five or above cannot generally be solved algebraically, the result that †Abelian equations can be solved algebraically, the theory of †binomial series and of series in general, the theory of †elliptic functions and more generally of †algebraic functions, the introduction of †Abelian integrals, and the establishment of †Abel's theorem. His work in both algebra and analysis, written in a style conducive to easy comprehension, reached the highest level of attainment of his time.

References

[1] N. H. Abel, Oeuvres complètes I, II, L. Sylow and S. Lie (eds.), Grondahl & Son, new edition, 1881 (Johnson, n.d.).
[2] C. A. Bjerknes, Niels Henrik Abel, Gauthier-Villars, 1885.
[3] F. Klein, Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert I, Springer, 1926 (Chelsea, 1956).

2 (IV.2)
Abelian Groups

A. General Remarks

A †group G is called an **Abelian group** (or **commutative group**) if G satisfies the commutative law $ab = ba$ for all $a, b \in G$. In this article, G always denotes an Abelian group. Every †subgroup of G is a †normal subgroup, and all the elements of finite order in G form a subgroup T, for which the †factor group G/T has no elements of finite order except the identity e. T is called the **(maximal) torsion subgroup** of G. If $G = T$, then the Abelian group G is called a **torsion group** (or **periodic group**). On the other hand, if $T = \{e\}$, then G is called **torsion-free**; if $T \neq G$, $T \neq \{e\}$, then G is called **mixed**. If the order of every element of a torsion group G is a power of a fixed prime number p, then G is called an **Abelian p-group** (or **primary Abelian group**). An Abelian torsion group is the †direct sum of primary Abelian groups. Thus the study of torsion groups is reduced to that of primary Abelian groups.

B. Finite Abelian Groups

The following fundamental theorem on finite Abelian groups was established by L. Kronecker, G. Frobenius, and L. Stickelberger in the 1870s. An Abelian group G of order p^n, where p is a prime number, is a direct product of †cyclic subgroups $Z_1, \ldots, Z_r$: $G = Z_1 \times \ldots \times Z_r$. If Z_i is of order p^{n_i}, then $n = n_1 + \ldots + n_r$, and we can assume that $n_i \geqslant n_{i+1}$. A direct product decomposition of G, as above, is not unique, but $n_1, \ldots, n_r$ are determined uniquely by G. The system $\{p^{n_1}, \ldots, p^{n_r}\}$ or $\{n_1, \ldots, n_r\}$ is called the system of **invariants** (or **type**) of G, and a system of †generators $\{z_1, \ldots, z_r\}$ of $Z_1, \ldots, Z_r$ is called a **basis** of G. An Abelian group of type $(p, p, \ldots, p)$ is called an **elementary Abelian group**. The decomposition of a finite Abelian group into a †direct sum of subsets (not necessarily of subgroups) was considered by G. Hajós (1942) and applied successfully to a problem of number theory (→ 151 Finite Groups).

C. Finitely Generated Abelian Groups

The theory of †finitely generated Abelian groups, i.e., Abelian groups generated by a finite number of elements, is as old as that of finite Abelian groups. The direct product of †infinite cyclic groups is called a **free Abelian group**. A finitely generated Abelian group G is the direct product of a finite Abelian group and a free Abelian group. The finite factor is the **torsion subgroup** of G. The free factor of the group G is not necessarily unique; however, the number of infinite cyclic factors of the free factor is uniquely determined and is called the **rank** of G. Two finitely generated Abelian groups are isomorphic if they have isomorphic maximal torsion subgroups and the same rank. This theory can be extended to the theory of †modules over a †principal ideal domain (→ 67 Commutative Rings K).

D. Torsion Groups

The structure of Abelian p-groups is relatively well known, compared with other infinitely generated Abelian torsion groups. In the 1920s, H. Prüfer made the first important contribution to the study of Abelian p-groups, and H. Ulm and L. Zippin completed the theory for countable groups in the 1930s. The uncountable case was first treated by L. Kulikov in the 1940s, but the study of this case is still in progress.

An Abelian p-group $G \neq \{e\}$ is called **divisible** (or **complete**) if for any $a \in G$ there is an element $x \in G$ satisfying $x^p = a$. A divisible group is a †direct sum of **Abelian groups of type** p^∞ (Prüfer). Here a group of type p^∞ is isomorphic to the †multiplicative group of all the p^nth roots of unity ($n = 1, 2, \ldots$) in the complex number field. Let G be any Abelian p-group. The maximal divisible subgroup V of G is a direct product factor of $G: G = V \times R$, where R has no divisible subgroups. An Abelian p-group without a divisible subgroup is called a **reduced Abelian group**.

An element x of an Abelian p-group G is said to have **infinite height** if for any n there is an element $y_n \in G$ satisfying $x = y_n^{p^n}$. The elements of infinite height form a subgroup G^1 of G. If $G^1 = \{e\}$ and G is countable, then G is decomposed uniquely into the direct sum of cyclic groups. This assertion fails if the hypothesis of countability is dropped. By †transfinite induction we can define G^β as follows. If β is an †isolated ordinal number, then $G^\beta = (G^{\beta-1})^1$; if β is a †limit ordinal number, then $G^\beta = \bigcap_{\alpha<\beta} G^\alpha$. For the least ordinal number τ such that $G^\tau = G^{\tau+1}$, G^τ is the maximal divisible subgroup of G. If G is reduced, then $G^\tau = \{e\}$. We call τ the **type** of an Abelian p-group G. For $\alpha < \tau$, $\bar{G}^\alpha = G^\alpha / G^{\alpha+1}$ is called the **Ulm factor** of G, and the sequence $\bar{G}^0, \ldots, \bar{G}^\alpha, \ldots$ $(\alpha < \tau)$ is called the **sequence of Ulm factors** of G. Each Ulm factor $\bar{G}^\alpha$ has no element of infinite height, and if $\alpha < \tau - 1$, then $\bar{G}^\alpha$ has an element of arbitrarily large order. Let τ be a countable ordinal number; assume that for any ordinal number $\alpha < \tau$ there is given a countable Abelian p-group A_α such that A_α has no element of infinite height, and that for $\alpha \neq \tau - 1$ A_α has an element of arbitrarily large order. Then there is a reduced countable Abelian p-group which is of type τ with a sequence of Ulm factors isomorphic to $A_0, A_1, \ldots, A_\alpha, \ldots$ $(\alpha < \tau)$ (Zippin). Two reduced countable Abelian p-groups A and B are isomorphic if they have the same type τ; and for any $\alpha < \tau$ the Ulm factors $\bar{A}^\alpha$ and $\bar{B}^\alpha$ are isomorphic. The assertion fails if the hypothesis of countability is dropped.

E. Torsion-Free Groups

In Abelian groups, the group operation is often written $a+b$, using the additive notation; an additively written group, called an **additive group**, is generally assumed to be Abelian. In the rest of this article we consider exclusively additive Abelian groups, of which the additive group $\mathbf{Z}$ of rational integers is the most primitive example. In such a group the identity element is called the **zero element** and is denoted by 0; the inverse of a is denoted by $-a$, and we write $a+(-b) = a-b$. The direct sum of additive groups A_λ $(\lambda \in \Lambda)$ is called a **free additive group** if each A_λ is isomorphic to $\mathbf{Z}$. An additive group G is regarded as a †module over the †ring $\mathbf{Z}$, to which the notion of linear independence is applicable (→ 277 Modules). Elements $a_1, \ldots, a_r$ of G are **linearly dependent** if there are integers $n_1, \ldots, n_r$ not all of which are zero such that $n_1 a_1 + \ldots + n_r a_r = 0$. Those that are not linearly dependent are termed **linearly independent**. An infinite set of elements of G is called linearly independent if the elements of any finite subset are linearly independent. If there are N elements of G that are linearly independent, but if any $N+1$ elements of G are linearly dependent, then N is called the **rank** of G. Such a system of N linearly independent elements is called a **maximal independent system**. A torsion-free additive group G is not necessarily free if G is not finitely generated.

The first important work on torsion-free additive groups was done by F. W. Levi (1917). A. G. Kurosh (1937) completed the theory in the case of finite rank. In the general case little is known, and I. Kaplansky, J. Rotman, and others are continuing the investigation.

The additive group $\mathbf{Q}$ of rational numbers is of rank 1, and conversely any additive group of rank 1 is isomorphic to some subgroup of $\mathbf{Q}$. An additive group G is called **divisible** (or **complete**) if for any $a \in G$ and for any integer n there is an element $x_n \in G$ such that $nx_n = a$. A divisible torsion-free additive group is isomorphic to a direct sum of some copies of $\mathbf{Q}$. For any torsion-free additive group G there is a divisible torsion-free additive group containing G. A minimal additive group F among these groups is uniquely determined up to isomorphism and has the rational number field $\mathbf{Q}$ as an †operator domain. Let $\mathbf{Q}^{(p)} = \{a/b \mid (a,b) = 1, p \nmid b\}$ be the ring of p-integers in $\mathbf{Q}$, and let G_p be the smallest $\mathbf{Q}^{(p)}$-subgroup of F containing G. Let $\mathbf{Q}_p$ be the †p-adic number field and $\mathbf{Z}_p$ the ring of †p-adic integers. Extending the operator domain $\mathbf{Q}$ to $\mathbf{Q}_p$ we obtain naturally a $\mathbf{Q}_p$-module F_p from F. Let $\bar{G}_p$ be the natural closure of G_p in F_p. Then $\bar{G}_p$ has $\mathbf{Z}_p$ as an

operator domain and thus becomes a $\mathbf{Z}_p$-module. A $\mathbf{Z}_p$-module of rank N is isomorphic to the direct sum of κ_p copies of $\mathbf{Q}_p$ and $N-\kappa_p$ copies of $\mathbf{Z}_p$: $\bar{G}_p \cong \sum_n \mathbf{Q}_p v_n + \sum_m \mathbf{Z}_p w_m$ ($n=1, \ldots, \kappa_p; m=1, \ldots, N-\kappa_p$). Here κ_p is called the **p-rank** of G. As the invariants of G, Kurosh gives the rank, the p-ranks for all primes p, and a certain equivalence class of the sequence of the matrices $\mathfrak{M}_p$. Here p ranges over all primes, and $\mathfrak{M}_p$ is the matrix of coefficients when the elements of a maximal independent system of F are written as linear combinations of (v_n, w_m).

F. General Abelian Groups

An Abelian group is, in general, an †extension of a torsion group by a torsion-free group. A torsion group T is called **bounded** if there is an integer n such that $t^n=1$ for all $t \in T$. Suppose there is a torsion group T. Then T is a direct summand of an Abelian group G which contains T as its maximal torsion subgroup if and only if T is the direct product of a divisible group and a bounded group (R. Baer and S. Fomin).

G. Characters

A **character** χ of an Abelian group G is a function which assigns to each $a \in G$ a complex number $\chi(a)$ of absolute value 1 and satisfies $\chi(ab)=\chi(a)\chi(b)$ for all $a, b \in G$. The product $\chi=\chi_1\chi_2$ of two characters χ_1 and χ_2 is defined by $\chi(a)=\chi_1(a)\chi_2(a)$, and χ is also a character of G. Thus all the characters of G form an Abelian group $C(G)$, which is called the **character group** of G. The identity element of the character group is the **identity character** (or **principal character**) χ such that $\chi(a)=1$ for all $a \in G$. If G is finite, then $G \cong C(G)$. This implies the duality $G=C[C(G)]$. This fact was extended by L. S. Pontryagin to †locally compact †topological Abelian groups (→ 422 Topological Abelian Groups B–D). For additive Abelian groups with operator domains → 277 Modules.

References

[1] L. Fuchs, Abelian groups, Pergamon, 1960.
[2] I. Kaplansky, Infinite Abelian groups, Univ. of Michigan, 1954.
[3] A. G. Kurosh, Theory of groups I, II, Chelsea, 1960. (Original in Russian, 1953.)

3 (VIII.5) Abelian Varieties

A. History

Except for C. F. Gauss, whose work on this subject saw the light many years after his death, N. H. Abel was the first to consider †algebraic functions as functions of complex variables and to discover double periods of †elliptic functions such as $x=x(u)$, which is the inverse function of an elliptic integral

$$u=\int^x \frac{dx}{\sqrt{f_4(x)}},$$

where $f_4(x)$ is a polynomial in x of degree 4. C. G. Jacobi expressed elliptic functions explicitly as ratios of theta series. As a natural generalization of elliptic functions, which are †Abelian functions of genus 1, Jacobi and his successors studied the inverse function of the †hyperelliptic integral, or more generally, the inverse function of the Abelian integral. By investigating hyperelliptic integrals of the first kind of genus 2,

$$\int^{x_1} \frac{dx_1}{\sqrt{f_6(x_1)}}, \qquad \int^{x_2} \frac{x_2 dx_2}{\sqrt{f_6(x_2)}},$$

where $f_6(x)$ is a polynomial of degree 6 in x, Jacobi obtained multiple-valued functions with quadruple periods. He then discovered the remarkable fact that if we consider sums of two integrals

$$u_1=\int^{x_1} \frac{dx_1}{\sqrt{f_6(x_1)}}+\int^{x_2} \frac{dx_2}{\sqrt{f_6(x_2)}},$$

$$u_2=\int^{x_1} \frac{x_1 dx_1}{\sqrt{f_6(x_1)}}+\int^{x_2} \frac{x_2 dx_2}{\sqrt{f_6(x_2)}},$$

the elementary symmetric functions $s_1=x_1+x_2$ and $s_2=x_1x_2$ of x_1 and x_2 are single-valued functions of u_1 and u_2 with quadruple periods. He also conjectured that these functions s_1 and s_2 might be expressed explicitly in terms of theta series of u_1 and u_2; this conjecture was later confirmed by J. G. Rosenhain and A. Göpel.

In the latter half of the 19th century, the general theory of Abelian functions was established. The central subject was Jacobi's inverse problem (→ Section L), which is a natural generalization of the above results. B. Riemann solved this problem by expressing elementary Abelian functions (→ Section M) as rational functions of theta functions.

The theory of functions with multiple periods was developed further by H. Poincaré, G. Frobenius, and E. Picard. In the 20th cen-

tury, the importance of the theories of Abelian functions and Abelian varieties has become more obvious with the development of the theory of functions of several complex variables and algebraic geometry. In particular, problems intimately related to number theory have given rise to the development of a purely algebraic theory of Abelian varieties. Valuable contributions have been made by S. Lefschetz, C. L. Siegel (†Siegel modular forms), A. Weil (algebraic theory of Abelian varieties, systematization of the theory of theta functions), and by D. Mumford (†moduli theory, algebraic theory of theta functions).

B. Algebraic Theory

When a †group variety is †complete as a variety, the group law is commutative; such a group variety is called an **Abelian variety** (→ 12 Algebraic Varieties H.). Let B be a subvariety of an Abelian variety A, and assume that B is a subgroup of A as an abstract group. Then B has the structure of an Abelian variety whose law of composition is induced by that of A, and B is called an **Abelian subvariety** of A. More generally, when an algebraic subset $\mathfrak{B}$ is a subgroup of A, then the component B of $\mathfrak{B}$ containing the identity element is an Abelian subvariety, and $\mathfrak{B}$ is a union of B and a finite number of cosets of B. When A is defined over a field k, then any Abelian subvariety of A is defined over a finite †separable extension of k (W. L. Chow's theorem). An Abelian variety A is called **simple** if A has no Abelian subvarieties other than A itself and 0.

Every †rational mapping of an algebraic variety V into an Abelian variety is defined at each simple point of V. This implies that an Abelian variety is †absolutely minimal.

C. Homomorphisms

A rational mapping of an Abelian variety A into an Abelian variety B is called a **rational homomorphism** (or simply **homomorphism**) if f is a group homomorphism. Let F be a rational mapping of A into B; then F can be uniquely expressed as follows: $F(x)=F_0(x)+F(0)$ $(x\in A)$, where F_0 is a homomorphism and $F(0)$ is the image of the unit element 0 of A. Hence the structure of an Abelian variety (as a group variety) is uniquely determined by the underlying algebraic-variety structure.

When a rational homomorphism f is birational, f is called a **birational isomorphism** (or simply **isomorphism**). It is clear that a rational isomorphism is an abstract isomorphism, but the converse is not necessarily true. Let A, B be two Abelian varieties. We say that A is **isogenous** to B if the dimension of A is equal to that of B and there exists a surjective homomorphism of A onto B, or equivalently, if there exists a surjective homomorphism of A onto B whose kernel is finite. The relation of isogeny is an equivalence relation. For an Abelian variety A and an Abelian subvariety X of A there is an Abelian subvariety Y of A such that the natural homomorphism $X\times Y\rightarrow A:(x,y)\rightarrow x+y$ is an isogeny (**Poincaré's complete reducibility theorem**). In particular, every Abelian variety is isogenous to a product of simple Abelian varieties that are determined uniquely up to isogeny and order.

Let A, B be two Abelian varieties; we denote by $\mathrm{Hom}(A,B)$ the additive group of rational homomorphisms of A into B. When a rational homomorphism λ is surjective, then the **degree** $\nu(\lambda)$ of λ is defined by $\lambda(A)=\nu(\lambda)B$ as †algebraic cycles. If λ is an isogeny, then $\nu(\lambda)\neq 0$, and the order of the kernel $\{t \mid t\in A, \lambda t=0\}$ is at most $\nu(\lambda)$ and equal to $\nu(\lambda)$ if and only if λ is †separable. The additive group $\mathrm{Hom}(A,B)$ is shown to be free of rank $\leqslant 4\dim A\cdot\dim B$ by the l-adic representation (→ Section E). If $A=B$, then $\mathrm{Hom}(A,A)$ has a ring structure; it is called the **ring of endomorphisms** (or **endomorphism ring**) of A and is denoted by $\mathfrak{A}(A)$. The tensor product $\mathfrak{A}_0(A)=\mathfrak{A}(A)\otimes\mathbf{Q}$, where $\mathbf{Q}$ is the field of rational numbers, is an †associative algebra over $\mathbf{Q}$. If A is simple, then $\mathfrak{A}_0(A)$ is a †division algebra. More generally, $\mathfrak{A}_0(A)$ is isomorphic to a direct product of some †total matrix algebras over division algebras; thus $\mathfrak{A}_0(A)$ is †semisimple. In particular, if A is 1-dimensional (in other words, if A is an †elliptic curve), the types of $\mathfrak{A}_0(A)$ are well known; when the characteristic $p=0$, then $\mathfrak{A}_0(A)$ is either the field of rational numbers or an †imaginary quadratic field. When $p>0$, aside from these two fields, we have a †quaternion algebra over $\mathbf{Q}$ as a possible type of $\mathfrak{A}_0(A)$.

Let k be a finite field with q elements. An algebraic integer is called a **Weil number** for q if every conjugate of it has absolute value $\sqrt{q}$. If A is an Abelian variety defined and simple over k, the qth power endomorphism of A: $x\rightarrow x^q$ determines a conjugacy class of Weil numbers for q, as Weil showed (→ 450 Zeta Functions). Moreover, we have the following classification theorem (J. Tate, T. Honda): There is a one-to-one correspondence between the set of all k-isogeny classes of k-simple Abelian varieties over k and the set of all conjugacy classes of Weil numbers for q. Tate also determined the structure of the division algebra $\mathfrak{A}_0(A)$ over $\mathbf{Q}$, which is described in terms of the decomposition of the qth power endomorphism into prime ideals.

D. Divisors

Let $\mathfrak{G}$ be the additive group of †divisors on an Abelian variety A and $\mathfrak{G}_a$ be the subgroup of divisors that are †algebraically equivalent to 0. Then the factor group $\mathfrak{G}/\mathfrak{G}_a$ has no †torsion part; this implies that for an Abelian variety †numerical equivalence coincides with †algebraic equivalence. We denote this relation by $\equiv$. Given an element a of A, the translation $T_a: A \ni x \rightarrow x+a \in A$ gives a †birational transformation, which is everywhere †biregular, on the underlying variety of A; we denote by X_a the image of a divisor X on A. Then $X \equiv 0$ if and only if X_a is †linearly equivalent to X for each point a of A. The †Albanese variety of an Abelian variety A is A itself, and the †Picard variety $\hat{A}$ of A is isogenous to A. In particular, for the †Jacobian variety J of an algebraic curve, $\hat{J}$ is isomorphic to J itself. The Picard variety $\hat{\hat{A}}$ of $\hat{A}$ is isomorphic to A (**duality theorem**). Let X be a divisor on A; the mapping $a \rightarrow$the †linear equivalence class of the divisor $X_a - X$, $a \in A$, is a rational homomorphism of A into $\hat{A}$, and we denote it by φ_X. If $\varphi_X = 0$, then $X \equiv 0$, and vice versa; hence we have an additive map of $\mathfrak{G}/\mathfrak{G}_a$ into $\mathrm{Hom}(A, \hat{A})$. If φ_X is surjective, we say that X is **nondegenerate**. A †positive divisor X is nondegenerate if and only if X is †ample, and then nX is †very ample for $n \geqslant 3$. There always exist positive nondegenerate divisors on an Abelian variety; therefore an Abelian variety is a †projective variety. For a given divisor X on A, we can find n suitable points $u_1, \ldots, u_n$, where n is the dimension of A, so that the †intersection product $X_{u_1} \cdot \ldots \cdot X_{u_n}$ is defined. We denote by $(X^{(n)})$ the †degree of the zero cycle $X_{u_1} \cdot \ldots \cdot X_{u_n}$. If X is positive nondegenerate, then the dimension $l(X)$ of the †defining module of the †complete linear system determined by X is equal to $(X^{(n)})/n!$ (**Poincaré's theorem**). Furthermore, the degree $\nu(\varphi_X)$ of φ_X, where X is any divisor on A, is given by the formula $\nu(\varphi_X) = ((X^{(n)})/n!)^2$ (**Frobenius's theorem**).

For a nondegenerate divisor X on A there is a unique integer $i = i(X)$, $0 \leqslant i \leqslant \dim A$, called the **index** of X, such that $H^p(A, \mathcal{O}(X)) = 0$ for $p \neq i$ and $H^i(A, \mathcal{O}(X)) \neq 0$. We have $i(-X) = \dim A - i(X)$ and $i(X) = 0$ if and only if X is ample.

E. l-adic Representations

Let A be an Abelian variety of dimension n. For a given prime number l, let $\mathfrak{G}_l(A)$ denote the group of points on A whose order is a power of l. If l is different from the characteristic of the base field of A, then the group $\mathfrak{G}_l(A)$ is isomorphic to the direct product of $2n$ factor groups $\mathbf{Q}_l/\mathbf{Z}_l$, where $\mathbf{Q}_l$ is the field of l-adic numbers and $\mathbf{Z}_l$ is the group of l-adic integers (→ 439 Valuations). We call such an isomorphism the **l-adic coordinate system** of $\mathfrak{G}_l(A)$. Now let λ be a rational homomorphism of A into an Abelian variety B of dimension m. Then we can see that λ induces a homomorphism of $\mathfrak{G}_l(A)$ into $\mathfrak{G}_l(B)$. This shows that by placing l-adic coordinate systems on $\mathfrak{G}_l(A)$ and $\mathfrak{G}_l(B)$ respectively, we get a matrix representation $M_l(\lambda)$ of λ with $2m$ rows and $2n$ columns. The representation $\lambda \rightarrow M_l(\lambda)$ is faithful, and $M_l(\lambda)$ is called the **l-adic representation** of λ. When $A = B$, then $\lambda \rightarrow M_l(\lambda)$ is a faithful representation of the ring of endomorphisms $\mathfrak{A}(A)$. This representation can be naturally extended to the representation of the algebra $\mathfrak{A}_0(A)$; the characteristic polynomial of the l-adic representation $M_l(\lambda)$ (where λ is an element of $\mathfrak{A}_0(A)$) is a polynomial with coefficients in $\mathbf{Q}$. Moreover, the polynomial does not depend on the choice of the prime number l. When $\lambda \in \mathfrak{A}(A)$, then $\nu(\lambda)$ is equal to $\det M_l(\lambda)$. The trace of $M_l(\lambda)$ is usually written as $\sigma(\lambda)$.

Let λ be a rational homomorphism of A into B and Y be a divisor on B. Then by the correspondence $\mathrm{cl}(Y) \rightarrow \mathrm{cl}(\lambda^{-1}(Y))$, we obtain a rational homomorphism of $\hat{B}$ into $\hat{A}$, called the **transpose** of λ and denoted by ${}^t\lambda$, where cl means the linear equivalence class (→ 16 Algebraic Varieties M). If X is a nondegenerate divisor on A, then the composition map $\beta: \hat{A} \rightarrow A$ of ${}^t\varphi_X: \hat{A} \rightarrow \hat{\hat{A}}$ and the canonical isomorphism $\hat{\hat{A}} \rightarrow A$ satisfy the equality $\beta \circ \varphi_X = \nu(\varphi_X)\delta$ ($\delta =$ the identity mapping of A). We denote by φ_X^{-1} the element $(1/\nu(\varphi_X))\,\beta$ in $\mathrm{Hom}(\hat{A}, A) \otimes \mathbf{Q}$. The correspondence $\alpha \rightarrow \alpha'$, $\alpha' = \varphi_X^{-1} \circ {}^t\alpha \circ \varphi_X$ ($\alpha \in \mathfrak{A}_0(A)$) is an †involution of $\mathfrak{A}_0(A)$ and is of order 1 or 2. If $\alpha \neq 0$, then $\sigma(\alpha' \circ \alpha) > 0$ (**Castelnuovo's lemma**). A. Weil was the first to recognize the importance of this theorem in connection with †Riemann's hypothesis on †congruence zeta functions.

F. Differential Forms

A †differential form ω on an Abelian variety A of dimension n is called **invariant** if $T_a^* \omega = \omega$ for every point $a \in A$, where T_a is the translation by a (→ Section D). The differential form †of the first kind is invariant, and conversely, every invariant differential form is of the first kind. Let K be the †universal domain and $K(A)$ be the †function field of A. The set of the linear differential forms of the first kind on A is a linear space over K of dimension n, and its basis becomes a basis over $K(A)$ of the linear space consisting of all linear differential forms on A. An **invariant derivation** on A is a derivation D in $K(A)$ satisfying $(Df) \circ T_a = D(f \circ T_a)$

for any element f of $K(A)$ and every point a of A. For a linear differential form $\omega=\sum f_i du_i$ and a derivation D, we put $\langle\omega, D\rangle=\sum f_i Du_i$. Then $\langle\omega, D\rangle$ is a bilinear form in ω and D. A derivation D is invariant if and only if $\langle\omega, D\rangle$ is a constant function for every invariant linear differential form ω. Similarly, a linear differential form ω is invariant if and only if $\langle\omega, D\rangle$ is a constant function for every invariant derivation D. The linear space of invariant linear differential forms and that of invariant derivations are dual to each other with respect to the bilinear form $\langle\omega, D\rangle$.

Now consider the case when the characteristic p of the universal domain is positive. The automorphism $a\rightarrow a^p$, $a\in K$, of the universal domain K induces a group isomorphism of A; we denote by A^p the image of A and by x^p the image of a point x of A. The image A^p is an Abelian variety, and the group isomorphism $\pi: x\rightarrow x^p$, $x\in A$, is an isogeny of A onto A^p. Let B be another Abelian variety and let λ be an isogeny of A onto B. If there is an isogeny μ: $B\rightarrow A^p$ such that $\pi=\mu\circ\lambda$, then we say that λ is of **height** 1. The function field $K(B)$ of B can be considered as a subfield of the function field $K(A)$ of A by the mapping λ. If λ is of height 1 and $v(\lambda)=p$, there exists an invariant derivation D of $K(A)$ with the constant field $K(B)$, uniquely determined up to constants. Moreover, we can choose D so that $D^p=D$ or $D^p=0$. In the first case λ is said to be of type (i_1); in the second case it is said to be of type (i_2). An isogeny whose degree is a prime different from the characteristic p is said to be of type (s_1), and a †separable isogeny whose degree is p is said to be of type (s_2). Any isogeny can be written as a product of isogenies of these four types.

G. Polarized Abelian Varieties

Let X be a divisor on an Abelian variety A; we denote by $\mathfrak{X}$ the class of divisors X' such that $mX\equiv m'X'$ for suitably chosen positive integers m, m'. When the class $\mathfrak{X}$ contains positive nondegenerate divisors, we say that $\mathfrak{X}$ determines a **polarization** on A, and the couple $(A,\mathfrak{X})$ is called a **polarized Abelian variety**. In particular, if A is a †Jacobian variety whose polarization $\mathfrak{X}$ is determined by a theta divisor, we call $(A,\mathfrak{X})$ the **canonically polarized Jacobian variety**. If an endomorphism α of A keeps the polarization invariant, i.e., if the class determined by $\alpha^{-1}(X)$ coincides with the class $\mathfrak{X}$, then α is called an **endomorphism** of the polarized Abelian variety $(A,\mathfrak{X})$. In particular, if α is an automorphism of A, then we say that α is an **automorphism** of $(A,\mathfrak{X})$. The group of all automorphisms of a polarized Abelian variety is finite. In particular, the group of automorphisms of a canonically polarized Jacobian variety is finite. Hence follows the famous theorem concerning the finiteness of the group of automorphisms of an †algebraic curve of genus not less than 2.

On the other hand, the algebraic equivalence class of a nondegenerate divisor is called an **inhomogeneous polarization**. (The above polarization is then sometimes called a homogeneous polarization.) An inhomogeneous polarization X determines an isogeny φ_X: $A\rightarrow\hat{A}$ uniquely. If φ_X is an isomorphism, the polarization is called **principal**. An endomorphism of an inhomogeneously polarized Abelian variety can be defined similarly.

H. Analytic Theory

For the rest of this article we take the complex number field $\mathbf{C}$ as the universal domain, and in this case we can utilize analytic and topological methods.

Let $\mathbf{C}^n$ be an n-dimensional vector space over $\mathbf{C}$. In a natural way, the space $\mathbf{C}^n$ becomes a $2n$-dimensional vector space $\mathbf{R}^{2n}$ over the real field $\mathbf{R}$, and the mapping $J: z\rightarrow\sqrt{-1}\,z$, $z\in\mathbf{C}^n$, is an $\mathbf{R}$-linear automorphism of $\mathbf{R}^{2n}$ such that $J^2=-1$. Conversely, if for an even-dimensional $\mathbf{R}$-vector space $\mathbf{R}^{2n}$ such a mapping J is given, then by putting $(a+\sqrt{-1}\,b)x=ax+bJx$ $(x\in\mathbf{R}^{2n}; a,b\in\mathbf{R})$, we can introduce an n-dimensional complex linear structure into $\mathbf{R}^{2n}$. We then say that J determines a **complex structure** on $\mathbf{R}^{2n}$; we denote by $\mathbf{C}^n=(\mathbf{R}^{2n}, J)$ the space having the complex structure determined by J. Let $\omega_1,\ldots,\omega_{2n}$ be $2n$ $\mathbf{R}$ linearly independent points on $\mathbf{C}^n=(\mathbf{R}^{2n}, J)$. Then the †lattice $\mathbf{D}$ generated by these points is discrete, and the factor group $\mathbf{T}^n=\mathbf{C}^n/\mathbf{D}$ is a **complex torus** of dimension n. We fix a basis of $\mathbf{C}^n$ and introduce a complex coordinate system on $\mathbf{C}^n$. Utilizing the basis $\omega_1,\ldots,\omega_{2n}$ of $\mathbf{R}^{2n}$, we also introduce a real coordinate system on $\mathbf{R}^{2n}$. We then obtain an $n\times 2n$ matrix $\Omega=(\omega_{ij})$, where the $(\omega_{1i},\ldots,\omega_{ni})$ are the complex coordinates of ω_i; the matrix Ω is called the **period matrix** of $\mathbf{T}^n$. Let $(z_1,\ldots,z_n)$ be the complex coordinates of a point $z\in\mathbf{C}^n$ and $(x_1,\ldots,x_{2n})$ be the real coordinates. Then we have ${}^t(z_1,\ldots,z_n)=\Omega\,{}^t(x_1,\ldots,x_{2n})$. If we let the same symbol J stand for the representation matrix of the linear transformation J with respect to the basis $\omega_1,\ldots,\omega_{2n}$, then we have $\sqrt{-1}\,\Omega=\Omega J$. The underlying real Lie group of $\mathbf{T}^n$ is a $2n$-dimensional (real) †torus group $(\mathbf{R}/\mathbf{Z})^{2n}$. Hence the †Poincaré polynomial of $\mathbf{T}^n$ is $(1+x)^{2n}$. Any †Hermitian metric on $\mathbf{C}^n$ (as a vector space) induces a †flat †Kähler metric; hence the

Hodge numbers (→ Kähler Manifolds C) are $h^{p,q}=\binom{n}{p}\binom{n}{q}$.

I. Theta Functions

A holomorphic function $f(z)$ on $\mathbf{C}^n=(\mathbf{R}^n, J)$ is called a **theta function** if for every $d\in\mathbf{D}$ we have $f(z+d)=f(z)\exp(l_d(z)+c_d)$, where $l_d(z)$ is a linear form on $\mathbf{C}^n$ which, as for c_d, depends on d. The set of zeros of a theta function f, which we write as (f), determines an analytic divisor on $\mathbf{T}^n$. Conversely, for every effective analytic divisor X on $\mathbf{T}^n$, there exists a theta function f such that $(f)=X$. With respect to the real coordinate system $x_1,\ldots,x_{2n}$ determined by the basis $\omega_1,\ldots,\omega_{2n}$, we can find $2n\times 2n$ matrices A, A_0 and a $1\times 2n$ matrix b, with elements in $\mathbf{C}$, such that the **transformation formula** $f(x+a)=f(x)\exp(2\pi\sqrt{-1}({}^t aAx+\frac{1}{2}{}^t aA_0a+{}^t ba))$ (where $A\equiv A_0 \pmod{\mathbf{Z}}$, ${}^tA_0=A_0$) holds for every $1\times 2n$ matrix a whose elements are rational integers. Moreover, if we put $E=A-{}^tA$, then E is an †alternating matrix whose elements are rational integers, and $S=EJ$ is a †positive semidefinite symmetric matrix. Conversely, if there exists such an alternating matrix E we can find a theta function. (There does exist, however, a complex torus on which no theta function exists other than trivial ones, i.e., ones of the form $\exp(\varphi(z))$, where $\varphi(z)$ is a polynomial of degree at most 2.)

A theta function f is called **nondegenerate** if it cannot be a function of $n-1$ complex variables, and f is nondegenerate if and only if the matrix $S=EJ$ is †positive definite. A complex torus has the structure of an Abelian variety if and only if there exists a nondegenerate theta function, i.e., if and only if there exists an alternating matrix E whose elements are rational integers such that EJ is a positive definite symmetric matrix. The latter condition is satisfied if and only if there exists an alternating matrix E whose elements are rational integers such that $\Omega{}^tE^{-1}{}^t\Omega=0$, $\sqrt{-1}\,\Omega{}^tE^{-1}{}^t\bar{\Omega}>0$ (positive definite Hermitian matrix). In particular, a period matrix Ω satisfying these conditions is called a **Riemann matrix**, and the rational matrix ${}^tE^{-1}$ is called the **principal matrix** belonging to Ω.

Determining a polarization on an Abelian variety can be reduced to designating a class of principal matrices obtained from a principal matrix by multiplying by positive integers. Let X be a positive divisor on an Abelian variety $\mathbf{T}^n=\mathbf{C}^n/\mathbf{D}$, and let f be a theta function such that $(f)=X$. Then the divisor X is nondegenerate if and only if the theta function f is nondegenerate, and the latter statement holds if and only if the alternating matrix E obtained from f is †invertible. For a given principal matrix ${}^tE^{-1}$, we can choose suitable coordinate systems of $\mathbf{C}^n$ and $\mathbf{D}$ so that $\Omega=(I_n, F)$ and $E=\begin{pmatrix}0 & \Delta\\ -\Delta & 0\end{pmatrix}$, where I_n is a unit matrix and Δ is a diagonal matrix whose elements are †elementary divisors of E. In these situations, ΔF is symmetric and its imaginary part positive definite, i.e., ΔF is a point of the †Siegel upper half-space $\mathfrak{S}_n$. Thus to the polarized Abelian variety (where the polarization is determined by ${}^tE^{-1}$) there corresponds a point of $\mathfrak{S}_n$. This gives a one-to-one correspondence between the isomorphism classes of Abelian varieties inhomogeneously polarized by a principal matrix with given elementary divisors and the points of the factor space $\mathfrak{S}_n/\Gamma_n(\Delta)$, where $\Gamma_n(\Delta)$ is a subgroup that is †commensurable to the Siegel †modular group of degree n and operates on $\mathfrak{S}_n$ discontinuously. The case of principal polarization corresponds to the case when $\Delta=I_n$, and then $\Gamma_n(\Delta)$ is the Siegel modular group itself. $\mathfrak{S}_n/\Gamma_n(\Delta)$ is the coarse †moduli space of Abelian varieties, polarized as above; the projective embedding of $\mathfrak{S}_n/\Gamma_n(\Delta)$ is given by means of †Siegel modular forms.

J. Abelian Functions

Meromorphic periodic functions on $\mathbf{C}^n$ with periods $\omega_1,\ldots,\omega_{2n}$ (i.e., meromorphic functions on $\mathbf{T}=\mathbf{C}^n/\mathbf{D}$, $\mathbf{D}=\mathbf{Z}\omega_1\oplus\ldots\oplus\mathbf{Z}\omega_{2n}$) are called **Abelian functions** with periods $\mathbf{D}$ (or on $\mathbf{T}$). The quotient of two theta functions having the same periods and the same transformation formulas is an Abelian function, and conversely any Abelian function can be written as such. All Abelian functions on $\mathbf{T}$ form a field $\mathbf{C}(\mathbf{T})$ called the **Abelian function field**. If $\mathbf{T}$ has the structure of an Abelian variety, then $\mathbf{C}(\mathbf{T})$ coincides with the field of rational functions on $\mathbf{T}$. In general, for any complex torus $\mathbf{T}$ there are an Abelian variety $\mathbf{T}'$ (possibly of dim 0) and a surjective homomorphism $\lambda:\mathbf{T}\rightarrow\mathbf{T}'$ which induces an isomorphism $\lambda^*:\mathbf{C}(\mathbf{T}')\cong\mathbf{C}(\mathbf{T})$. Such a pair $(\mathbf{T}',\lambda:\mathbf{T}\rightarrow\mathbf{T}')$ is unique up to isomorphism, and is called the †algebraic reduction of $\mathbf{T}$.

K. Homomorphisms

Let $\mathbf{C}^{n_i}=(\mathbf{R}^{2n_i}, J_i)$ $(i=1,2)$ be complex linear spaces; an $\mathbf{R}$-linear mapping $f:\mathbf{R}^{2n_1}\rightarrow\mathbf{R}^{2n_2}$ is $\mathbf{C}$-linear if and only if the relation $f\circ J_1=J_2\circ f$ holds. Let D_i be †lattice groups of $\mathbf{C}^{n_i}$ for $i=1,2$. If a $\mathbf{C}$-linear mapping $\Lambda:\mathbf{C}^{n_1}\rightarrow\mathbf{C}^{n_2}$ satisfies $\Lambda(\mathbf{D}_1)\subset\mathbf{D}_2$, then Λ induces a complex analytic homomorphism of $\mathbf{T}_1=\mathbf{C}^{n_1}/\mathbf{D}_1$ to $\mathbf{T}_2=\mathbf{C}^{n_2}/\mathbf{D}_2$. Conversely, any complex analytic homomorphism of $\mathbf{T}_1$ to $\mathbf{T}_2$ is ob-

tained in this way. Let $\mathbf{T}_1$ and $\mathbf{T}_2$ be Abelian varieties, and let $\Omega_1=(\omega_1^{(1)},\ldots,\omega_{2n_1}^{(1)})$ and $\Omega_2=(\omega_1^{(2)},\ldots,\omega_{2n_2}^{(2)})$ be their Riemann matrices. Then for a homomorphism $\lambda:\mathbf{T}_1\to\mathbf{T}_2$, we can find a representation matrix $W(\lambda)$ with complex coefficients; and with respect to the real coordinate systems $(\omega_1^{(1)},\ldots,\omega_{2n_1}^{(1)})$ and $(\omega_1^{(2)},\ldots,\omega_{2n_2}^{(2)})$, we can find a representation matrix $M(\lambda)$ with coefficients in $\mathbf{Z}$ such that $W(\lambda)\Omega_1=\Omega_2 M(\lambda)$. Conversely, if for a complex matrix W there is a matrix M, with coefficients in $\mathbf{Z}$, satisfying the relation given above, then W gives a homomorphism of $\mathbf{T}_1$ to $\mathbf{T}_2$. The above equation is called **Hurwitz's relation.** The notion of l-adic coordinate system, which is valid for a general characteristic, corresponds to that of the lattice group, and the l-adic representation $M_l(\lambda)$ of λ is the abstraction of the integral representation $M(\lambda)$.

L. Abelian Integrals

Let $\mathfrak{R}$ be a compact †Riemann surface of genus $g\geqslant 1$ (→ 11 Algebraic Functions) and let ω be a sum of †Abelian differentials of the first kind or of the second kind. Then the †period of ω along a cycle γ depends only on the †homology class of γ. The set of all differentials of the first kind forms a complex linear space of dimension g; we denote it by $\mathfrak{D}_0$. Let P be a point and P_0 a fixed point of R; then we denote by $u(P)$ the vector integral $(\int_{P_0}^{P}\omega_1,\ldots,\int_{P_0}^{P}\omega_g)$, where $(\omega_1,\ldots,\omega_g)$ is a basis of $\mathfrak{D}_0$ and the path from P_0 to P is common to every integral. The correspondence $P\to u(P)$ is not a single-valued mapping; the totality of differences of values of $u(P)$ coincides with the group $\mathbf{D}$ consisting of periods $(\int_\gamma\omega_1,\ldots,\int_\gamma\omega_g)$, where γ varies over all cycles. Let a set of cycles $\{\gamma_1,\ldots,\gamma_{2g}\}$ be a basis of the homology group, with coefficients in $\mathbf{Z}$; then $2g$ column vectors of the $g\times 2g$ matrix $\Omega=(\omega_{ij})$, $\omega_{ij}\int_{\gamma_j}\omega_i$, are linearly independent over $\mathbf{R}$. Since the group $\mathbf{D}$ coincides with the set of linear combinations, with coefficients in $\mathbf{Z}$, of column vectors of Ω, this $\mathbf{D}$ is a †lattice of rank $2g$, and the matrix Ω is a period matrix of the complex torus $\mathbf{T}^g=\mathbf{C}^g/\mathbf{D}$ of dimension g.

For a basis of the homology group with coefficients in $\mathbf{Z}$, we take †normal sections $\alpha_1,\ldots,\alpha_g,\alpha_{g+1},\ldots,\alpha_{2g}$ of $\mathfrak{R}$, and let the same symbol Ω stand for the period matrix (ω_{ij}), $\omega_{ij}=\int_{\alpha_j}\omega_i$; then we have

$$\Omega E{}^t\Omega=0,\qquad \sqrt{-1}\,\Omega E{}^t\bar{\Omega}>0$$

(positive definite Hermitian matrix), where

$$E=\begin{pmatrix}0 & I_g\\ -I_g & 0\end{pmatrix}\quad\text{(with } I_g \text{ the unit matrix).}$$

This implies that E is a principal matrix belonging to Ω; we call the equality and inequality just given **Riemann's period relation** and **Riemann's period inequality**, respectively. Furthermore, we can choose a suitable basis of $\mathfrak{D}_0$ so that the period matrix Ω is of the form (I_g,F) with $F\in\mathfrak{S}_g$ (→ Section I). We consider the function $\vartheta(u)$, $u=(u_1,\ldots,u_g)$, defined by an infinite series

$$\vartheta(u)=\sum_m\exp(2\pi\sqrt{-1}\,(\mathbf{m}{}^tu+\tfrac{1}{2}\mathbf{m}F{}^t\mathbf{m})),$$

where the sum is taken over all row vectors $\mathbf{m}=(m_1,\ldots,m_g)$ with coefficients in $\mathbf{Z}$. If u is in a bounded region, then the series for $\vartheta(u)$ is uniformly and absolutely convergent. Hence $\vartheta(u)$ is a holomorphic function of u. This is a theta function corresponding to the principal matrix $E=\begin{pmatrix}0 & I_g\\ -I_g & 0\end{pmatrix}$ and is called the **Riemann theta function**. As $\vartheta(u)$ is nondegenerate, the complex torus $\mathbf{T}^g=\mathbf{C}^g/\mathbf{D}$ has the structure of an Abelian variety. If we regard the Riemann surface $\mathfrak{R}$ as an algebraic curve, this Abelian variety $\mathbf{T}^g$ is precisely the Jacobian variety of the curve $\mathfrak{R}$. The collection of zeros of $\vartheta(u)$, a divisor on $\mathbf{T}^g$, defines the canonical polarization on the Jacobian variety (→ Section G).

The correspondence $P\to u(P)$, $p\in\mathfrak{R}$ induces a well-defined mapping φ of $\mathfrak{R}$ into $\mathbf{T}^g=\mathbf{C}^g/\mathbf{D}$. Moreover, if we set $\varphi(A)=\sum_{j=1}^r\varphi(P_j)-\sum_{j=1}^r\varphi(Q_j)$ for any divisor $A=P_1\ldots P_r/Q_1\ldots Q_r$ of degree 0, then $\varphi(A)$ is a point on $\mathbf{T}^g=\mathbf{C}^g/\mathbf{D}$, which is represented by the vector $(\sum_{j=1}^r\int_{Q_j}^{P_j}\omega_1,\ldots,\sum_{j=1}^r\int_{Q_j}^{P_j}\omega_g)$ of $\mathbf{C}^g$, and the mapping $A\to\varphi(A)$ is a homomorphism of the group $\mathfrak{G}_0$ of divisors of degree 0 onto $\mathbf{T}^g=\mathbf{C}^g/\mathbf{D}$. The kernel of this homomorphism coincides with the group $\mathfrak{G}_l$ of †principal divisors (**Abel's theorem**). Hence a divisor $A=P_1\ldots P_r/Q_1\ldots Q_r$ of degree 0 is a divisor of some function if and only if we have $\sum_{j=1}^r\int_{Q_j}^{P_j}\omega_i\equiv 0\ (\mathrm{mod}\,\mathbf{D})$ $(i=1,\ldots,g)$, or equivalently, the left-hand side is 0 for a suitable path.

Given g fixed points $P_1,\ldots,P_g$ on $\mathfrak{R}$ and given $(u_1,\ldots,u_g)$ as any vector of $\mathbf{C}^g$, the problem of finding g points $Q_1,\ldots,Q_g$ satisfying relations $\sum_{j=1}^g\int_{P_j}^{Q_j}\omega_i\equiv u_i\ (\mathrm{mod}\,\mathbf{D})$, $i=1,\ldots,g$, is called **Jacobi's inverse problem**. To solve the problem, we take a divisor A of degree 0 such that the class $\varphi(A)\ (\mathrm{mod}\,\mathbf{D})$ is represented by $(u_1,\ldots,u_g)$; then, by virtue of the †Riemann-Roch theorem, there exists a divisor $Q_1\ldots Q_g$ satisfying

$$A\equiv Q_1\ldots Q_g/P_1\ldots P_g\ (\mathrm{mod}\,\mathfrak{G}_l).$$

Abel's theorem implies that the set of points $\{Q_1,\ldots,Q_g\}$ is a solution of Jacobi's problem. Moreover, for general $(u_1,\ldots,u_g)$, the solution is unique; i.e., there exists a subvariety $\tilde{\mathfrak{X}}$ of dimension $g-2$ on $\mathbf{C}^g$ such that the solution is

unique up to order if and only if $(u_1, \ldots, u_g)$ does not lie on $\tilde{\mathfrak{X}}$. In particular, if every point P_i coincides with the fixed point P_0 that appeared in the definition of $u(P)$, the subvariety on $\mathbf{T}^g = \mathbf{C}^g/\mathbf{D}$ determined by $\tilde{\mathfrak{X}}$ is obtained in the following way: Let $W_1 + \ldots + W_{2g-2}$ be a canonical divisor on $\mathfrak{R}$, and put $c = \varphi(W_1) + \ldots + \varphi(W_{2g-2})$. Then the locus of points $c - \varphi(R_1) - \ldots - \varphi(R_{g-2})$ (where $g-2$ points $R_1, \ldots, R_{g-2}$ are taken independently over all points of $\mathfrak{R}$) is the desired subvariety.

In terms of the theory of complex manifolds the above result can be rewritten as follows: Let $S^g(\mathfrak{R})$ be the symmetric product of $\mathfrak{R}$ and $\mathfrak{X}$ the subvariety of $\mathbf{T}^g = \mathbf{C}^g/\mathbf{D}$ induced from $\tilde{\mathfrak{X}}$. Then the holomorphic map $S^g(\varphi): S^g(\mathfrak{R}) \rightarrow \mathbf{T}^g$ induced from $\varphi: \mathfrak{R} \rightarrow \mathbf{T}^g$ is bimeromorphic and isomorphic outside that $\mathfrak{X}$ whose elements correspond to effective divisor classes of degree g contained in canonical divisors on $\mathfrak{R}$.

M. Elementary Abelian Functions

Let z be a nonconstant meromorphic function on $\mathfrak{R}$. Then for any $u = (u_1, \ldots, u_g) \in \mathbf{C}^g$ that does not lie on $\tilde{\mathfrak{X}}$, there exist g points $Q_1, \ldots, Q_g$ (uniquely determined up to order) as the solution of Jacobi's problem. Therefore the elementary symmetric functions

$$s_1(u; z) = \sum_{j=1}^{g} z(Q_j),$$

$$s_2(u; z) = \sum_{i<j} z(Q_i) z(Q_j), \ldots, s_g(u; z) = \prod_{j=1}^{g} z(Q_j)$$

are well defined if u lies outside the variety $\tilde{\mathfrak{X}}$ of dimension $g-2$. Each function $s_i(u; z)$, regarded as a function of u, can be extended uniquely to an Abelian function in the whole space $\mathbf{C}^g$; the function is denoted by the same symbol $s_i(u; z)$. The Abelian functions $s_1(u; z), \ldots, s_g(u; z)$ are called the **elementary Abelian functions** obtained from z.

Now let K be the field of Abelian functions on $\mathbf{C}^n/\mathbf{D}$ and k the field of meromorphic functions on $\mathfrak{R}$; then the dimension of K over $\mathbf{C}$ is g, and $[K : \mathbf{C}(s_1(u; z), \ldots, s_g(u; z))] = r^g$, where r is the degree of the function z that is given by $[k : \mathbf{C}(z)]$. Moreover, if we take any function w such that $k = \mathbf{C}(z, w)$, then we have

$$K = \mathbf{C}(s_1(u; z), \ldots, s_g(u; z); s_1(u; w), \ldots, s_g(u; w)),$$

where $s_1(u; w), \ldots, s_g(u; w)$ are the elementary Abelian functions obtained from w.

We can write any elementary Abelian function as a rational function of Riemann theta functions; therefore any Abelian function can be written as a rational function of Riemann theta functions. Furthermore, if u and v are variable vectors, then $s_i(u+v; z)$ can be represented as an algebraic function of $s_1(u; z), \ldots, s_g(u; z), s_1(v; z), \ldots, s_g(v; z)$; i.e., we can choose a suitable polynomial $H_i(Z; X_1, \ldots, X_g; Y_1, \ldots, Y_g)$ with coefficients in $\mathbf{C}$ so that $H_i(s_i(u+v; z); s_1(u; z), \ldots, s_g(u; z); s_1(v; z), \ldots, s_g(v; z)) = 0$. This **algebraic addition formula** with respect to the elementary Abelian functions $s_i(u; z)$, $i = 1, \ldots, g$, is a function-theoretic interpretation of the fact that the addition map $X \times X \rightarrow X : (x, y) \rightarrow x + y$ is a †morphism of algebraic varieties.

As the study of †Abelian integrals of the first kind led us to the theory of Jacobian varieties, †Abelian integrals of the second and the third kind give rise to the theory of †generalized Jacobian varieties (→ 9 Algebraic Curves).

The theory of Abelian varieties has significant applications to number theory, as shown by the following examples: the theory of †unramified Abelian extensions with respect to a function field of several variables defined over a finite field (S. Lang), the theory of heights of points on an Abelian variety (Weil, A. Néron, J. Tate), and the theory of complex multiplication (→ 73 Complex Multiplication) in the case of higher dimensions (Y. Taniyama, G. Shimura).

N. Some Recent Results

(1) Level Structure, Moduli of Abelian Varieties. Let A be an Abelian variety over k of dimension g and n a positive integer which the characteristic of k does not divide. A **level n structure** on A is defined to be a set of $2g$ points $\sigma_1, \ldots, \sigma_{2g}$ on A which form a basis for the group of points of order n on A.

Let $A(g, d, n; k)$ be the set of triples: (i) an Abelian variety A over k of dimension g, (ii) an inhomogeneous polarization X on A with $v(\varphi_X) = d^2$, and (iii) a level n structure $\sigma_1, \ldots, \sigma_{2g}$ of A, all up to isomorphisms. Similarly we can define $A(g, d, n; S)$ for Abelian schemes over a scheme S. The correspondence $S \rightarrow A(g, d, n; S)$ defines a functor $\mathscr{A}(g, d, n)$. D. Mumford has shown that there exists the †coarse moduli scheme $A(g, d, n)$ quasi-projective over $\mathrm{Spec}(\mathbf{Z}[1/n])$, and that it is even fine if $n \geqslant 3$ [6]. He used the technique of †Hilbert schemes and †stable points (→ 16 Algebraic Varieties). One of the key steps of his proof is to show that for an embedding $\phi : A \rightarrow \mathbf{P}^m$ of degree r (i.e., the degree of $\phi(A)$ in $\mathbf{P}^m$) over an algebraically closed field k and a positive integer n such that char(k) does not divide n and $n > \sqrt{(m+1)r}$, the point $(\phi(x_i))_{i=1,\ldots,n^{2g}}$ in $(\mathbf{P}^m)^{n^{2g}}$ is stable with respect to the action of $PGL(m)$, where the x_i are the points of order n on A (with an arbitrary order). Mumford

later showed another method of constructing the moduli of polarized Abelian varieties by using algebraic theta constants [7].

(2) Néron Minimal Models, Good and Stable Reduction. Let R be a discrete valuation ring with residue field k and quotient field K. For an Abelian variety A over K, there exists a smooth group scheme $\mathscr{A}$ of finite type over $S = \mathrm{Spec}(R)$, called the **Néron minimal model** of A, such that for every scheme S' smooth over S there is a canonical isomorphism

$$\mathrm{Hom}_S(S', \mathscr{A}) \cong \mathrm{Hom}_K(S'_K, A),$$

where S'_K is the pullback of S' by $\mathrm{Spec}(K) \to \mathrm{Spec}(R)$ (A. Néron, M. Raynaud). In particular, we have $\mathscr{A}_K \cong A$. Denote by A_0 the fiber of $\mathscr{A}$ over the closed point of S.

If $\mathscr{A}$ is proper over S, we say that A has a **good reduction** at R. If the connected component A_0^0 of A_0 containing 0 has no unipotent radical (or equivalently, A_0^0 is an extension of an Abelian variety by an †algebraic torus over a finite algebraic extension of k), we say that A has a **stable reduction**. If there is a finite separable extension K' of K with a prolongation R' of R to K' such that $A \times_K K'$ has good (stable) reduction, we say that A has a **potential good (stable) reduction** at R. Let K_s be a separable closure of K and $\bar{R}$ a prolongation of R to K_s. For a prime number $l \neq \mathrm{char}(k)$, we have a canonical homomorphism $\rho: \mathrm{Gal}(K_s/K) \to \mathrm{Aut}(\mathfrak{G}_l(A))$, called a monodromy. Then A has potential good reduction if and only if the image of the †inertia group $I(\bar{R})$ by ρ is a finite group (J.-P. Serre and Tate). Every Abelian variety A over K has potential stable reduction at R (**stable reduction theorem**, A. Grothendieck [16]).

(3) Graded Ring of Theta Functions. If f is a theta function with a period system $\mathbf{D}$, f^n is also a theta function with the same period system for every positive n. We denote by S_n the vector space of the theta functions with the period system $\mathbf{D}$ subject to the same transformation law as f^n. If we denote by X the effective divisor on $\mathbf{T} = \mathbf{C}^g/\mathbf{D}$ defined by f, then S_n can be naturally considered as the †defining module of the †complete linear system of nX, and the dimension of S_n $(= l(nX))$ is equal to the product of the nonzero diagonal elements of $n\Delta$ (→ Section I) (Frobenius). If $g \in S_m$ and $h \in S_n$, then $gh \in S_{m+n}$; hence $S = \bigoplus_{n \geqslant 0} S_n$ is a †graded ring, which is normal and finitely generated. For $m \geqslant 2$ and $n \geqslant 3$, the product map $S_m \times S_n \to S_{m+n}$ is surjective (D. Mumford, S. Koizumi). If the elementary divisors of E can be divided by an integer $\geqslant 4$, the kernel of the natural graded mapping $\mathrm{S}(S_1) \to S$ (where $\mathrm{S}(S_1)$ denotes the †symmetric algebra over S_1) is generated by the quadratic relations (i.e., the part of degree 2) for sufficiently large degrees (Mumford). Geometrically this means that if $X = (f)$, $f \in S_1$, is nondegenerate, then, with the projective embedding $i: \mathbf{T} \to \mathbf{P}^N$ defined by the complete linear system of X, $i(\mathbf{T})$ is an intersection of quadrics in $\mathbf{P}^N$ containing $i(\mathbf{T})$.

Mumford has developed a theory of algebraic theta functions that also works for the positive characteristic case [7] and has proved the above results in general.

References

[1] Sém. H. Cartan, vol. 2, Fonctions analytiques de plusieurs variables I–V, 1951–1952, Benjamin, 1967.
[2] F. Conforto, Abelsche Funktionen und algebraische Geometrie, Springer, 1956.
[3] A. Krazer and W. Wirtinger, Abelsche Funktionen und allgemeine Thetafunktionen, Enzykl. Math. Wiss., II, 2, B7 (1920), 604–873.
[4] J. Igusa, Theta functions, Springer, 1972.
[5] S. Lang, Abelian varieties, Interscience, 1959.
[6] D. Mumford, Geometric invariant theory, Erg. Math., Springer, 1965; second edition, 1982.
[7] D. Mumford, On the equations defining Abelian varieties I–III, Inventiones Math., 1 (1966), 287–354; 3 (1967), 75–135, 215–244.
[8] D. Mumford, Varieties defined by quadratic equations, Questioni sulle variete algebraiche, Corsi dal C. I. M. E., Cremona, 1969, III Ciclo, Cremonese, 1970, 31–94.
[9] D. Mumford, Abelian varieties, Tata Inst. studies in math., Oxford Univ. Press, 1974.
[10] A. Néron, Modèles minimaux des variétés abéliennes sur les corps locaux et globaux, Publ. Math. Inst. HES, 21 (1964), 5–128.
[11] W. F. Osgood, Lehrbuch der Funktionentheorie II_2, Teubner, 1932.
[12] M. Raynaud, Modèles de Néron, C. R. Acad. Sci. Paris, 262 (1966), 345–347.
[13] B. Riemann, Theorie der Abelschen Funktionen, J. Reine Angew. Math., 54 (1857), 115–155.
[14] J.-P. Serre, Quelques propriétés des variétés abéliennes en caractéristique p, Amer. J. Math., 80 (1958), 715–739.
[15] J.-P. Serre and J. Tate, Good reduction of Abelian varieties, Ann. Math., (2) 88 (1968), 492–517.
[16] Séminaire de géométrie algébriques (SGA) 7, Groupes de monodromie en géométrie algébrique, Lecture notes in math. 288, 340, Springer, 1972, 1973.

[17] G. Shimura, Moduli and fibre systems of Abelian varieties, Ann. Math., (2) 83 (1966), 294–338.
[18] C. L. Siegel, Analytic functions of several complex variables, Lecture notes, Institute for Advanced Study, Princeton, 1948–1949.
[19] J. Tate, Classes d'isogénie des variétés abéliennes sur un corps fini (d'après T. Honda), Sém. Bourbaki, 1968–1969, no. 352; Lecture notes in math. 179, Springer, 1971.
[20] A. Weil, Théorèmes fondamentaux de la théorie des fonctions thêta, Séminaire Bourbaki, no. 16, 1949, Benjamin, 1966.
[21] A. Weil, Variétés abéliennes et courbes algébriques, Actualités Sci. Ind., Hermann, 1948.
[22] A. Weil, Variétés kählériennes, Actualités Sci. Ind., Hermann, 1958.
[23] A. Weil, Généralisation des fonctions abéliennes, J. Math. Pures Appl., (9) 17 (1938), 47–87.
[24] D. Mumford, Tata lecture notes on theta functions I, II, Birkäuser, 1983.

4 (V.5) Additive Number Theory

A. General Remarks

Let $\mathbf{N}$ be the set of natural numbers, and let A, $B, \ldots$ be subsets of $\mathbf{N}$. The sum $C=A+B$ is defined as the set $\{c \mid c\in A, c\in B, \text{ or } c=a+b, a\in A, b\in B\}$. A finite sum of subsets of $\mathbf{N}$ is defined similarly. If $\mathbf{N}$ is the sum $A+\ldots+A$ (r times), then we say that A is a **basis of order r in $\mathbf{N}$**. Let $A(x)$ denote the number of integers in A that do not exceed x. The **density** of A is defined as $\inf A(x)/x$. If $A(n)\geqslant \alpha n$, $B(n)\geqslant \beta n$ $(0<\alpha, \beta<1)$ for all $n\in\mathbf{N}$, then we have $(A+B)(n)\geqslant(\min(1,\alpha+\beta))n$. This result was stated by E. Landau (1937) without proof; A. J. Khinchin had given a proof (1932) for the case $\alpha=\beta$, and H. B. Mann (1942) and E. Artin and P. Scherk (1943) succeeded in proving the statement for the general case. Suppose that the densities of A and B are, respectively, α and 0. If B is a basis of finite order in $\mathbf{N}$, then the density of $A+B$ is greater than that of A (Khinchin and P. Erdös, 1936). Let P be the set of all prime numbers. Though the density of P is 0, the density of $P+P$ is positive (L. G. Shnirel'man, 1930). Hence P is a basis of finite order in $\mathbf{N}$; in other words, there exists a positive integer r such that every natural number can be expressed as a sum of at most r primes. Though the density of the set Q of the kth powers of natural numbers is 0, there exists a positive integer $s(k)$ such that the sum $Q+\ldots+Q$ ($s(k)$ times) is a set of positive density. L. K. Hua (1956) gave a simple proof of this fact, based on Yu. V. Linnik's idea. It follows, therefore, that any natural number n can be expressed as a sum $n=a_1^k+\ldots+a_t^k$, where $a_i\in\mathbf{N}$, $t\geqslant s(k)$. This result had already been shown by D. Hilbert (1909).

An ancient method of finding prime numbers is †Eratosthenes' sieve. V. Brun (1920) devised a new sieve method to express an arbitrary integer n as the sum of two integers $n=a+b$, where the number of prime factors of a and b is as small as possible. This method was improved by H. A. Rademacher (1924), Landau (1931), A. Bustab (1937), and others. Among these the method found by A. Selberg (1952) is notable (→ 123 Distribution of Prime Numbers E).

B. Farey Sequences

Let τ be a positive integer. We arrange in increasing order the set of all positive irreducible fractions lying between 0 and 1 whose denominators do not exceed τ. This sequence is called the **Farey sequence** of order τ. For example, the Farey sequence of order 5 consists of

$$\frac{0}{1},\ \frac{1}{5},\ \frac{1}{4},\ \frac{1}{3},\ \frac{2}{5},\ \frac{1}{2},\ \frac{3}{5},\ \frac{2}{3},\ \frac{3}{4},\ \frac{4}{5},\ \frac{1}{1}.$$

A necessary and sufficient condition that a fraction a/b be directly followed by a fraction c/d in the Farey sequence of order n is $b+d\geqslant n+1$, $bc-ad=1$. In this case the fraction $(a+c)/(b+d)$ is called the **mediant** of a/b and c/d. Interpolating the Farey sequence of order n with such mediants $(a+c)/(b+d)$ satisfying $b+d=n+1$, we obtain the Farey sequence of order $n+1$.

Let a/q be a fraction in the Farey sequence of order τ, and a'/q', a''/q'' be adjacent members of a/q in the sequence such that $a'/q'<a/q<a''/q''$. The interval $[(a'+a)/(q'+q), (a+a'')/(q+q'')]$ is known as the **Farey arc** surrounding a/q. In particular, if $a/q=0/1$, then we set $[-1/n, 1/n]$ to be the Farey arc surrounding 0/1, where $n=[\tau]+1$ ([ ] is the †Gauss symbol). We can thus decompose the interval $[-1/n, 1-1/n]$ into a disjoint union of Farey arcs. If α is contained in the Farey arc surrounding a/q, then $|\alpha-a/q|<1/q\tau$. Therefore, for a given $\tau\geqslant 1$ and a real α, we can prove the existence of a/q such that

$$(|a|,q)=1, \quad 0<q\leqslant\tau, \quad |\alpha-a/q|<1/q\tau.$$

Let $f(\zeta)=\sum_{n=0}^{\infty}a_n\zeta^n$ be a power series which is convergent in the disk $|\zeta|\leqslant 1$. Then, by Cauchy's integral formula,

$$a_n = \frac{1}{2\pi i}\int_{|\zeta|=1} \frac{f(\zeta)}{\zeta^{n+1}} d\zeta = \int_0^1 f(e^{2\pi i\alpha}) e^{-2\pi i n\alpha} d\alpha.$$

To estimate such an integral, we sometimes utilize the decomposition of the interval [0, 1] into a disjoint union of Farey arcs as mentioned above. This method is called the **circle method**, and the subdivision of the interval is known as the **Farey dissection**.

Given a positive number c, a Farey arc around a/q is usually called a **major arc** (or **basic interval**) if q does not exceed the given bound c; otherwise, it is called a **minor arc** (or **supplementary interval**). Usually, the principal part of the previously mentioned integral is derived from the integral over the major arcs, and the residual part is provided by the integral over the minor arcs.

C. Goldbach's Problem

Goldbach's problem is found in letters (1742) he exchanged with L. Euler. In them he stated that every positive integer can be expressed as the sum of primes. More precisely, he conjectured that any even integer not smaller than 6 can be expressed as the sum of two odd primes and that any odd integer not smaller than 9 can be expressed as the sum of three odd primes.

I. M. Vinogradov (1937) proved that every sufficiently large odd integer can be expressed as the sum of three primes. Let N be a sufficiently large odd integer. If we write

$$A(q, N) = \frac{\mu(q)}{\varphi^3(q)} \sum_{\substack{1 \leqslant a \leqslant q \\ (a,q)=1}} \exp\left(-2\pi i \frac{a}{q} N\right),$$

then the series $S(N) = \sum_{q=1}^{\infty} A(q, N)$ is absolutely convergent and is equal to

$$\prod_p \left(1 + \frac{1}{(p-1)^3}\right) \prod_{p|N} \left(1 - \frac{1}{p^2 - 3p + 3}\right).$$

It is known that $S(N) > 6/\pi^2$ for all N. If we denote by $r(N)$ the number of solutions of $N = p_1 + p_2 + p_3$, then $r(N) \sim (N^2/2(\log N)^3) \cdot S(N)$. To prove this, Vinogradov used the circle method. He employed the †prime number theorem for arithmetic progressions to estimate the integrals over the major arcs and devised an ingenious method to estimate the series $\sum_{p \leqslant N} \exp(2\pi i \alpha p)$ in the computation of the integrals on the minor arcs. A finite or infinite sum of exponential functions such as this is called a **trigonometric sum**. More generally, we consider trigonometric sums of several variables. Vinogradov provided detailed remarks and calculations [6].

In the case of even integers, the problem is still unsolved, although J. G. van der Corput, T. Estermann, and N. G. Chudanov proved simultaneously (1938) that almost all even integers (i.e., except a set of density 0) can be expressed as the sum of two primes. For these problems, Linnik (1946) and Chudanov (1947) introduced function-theoretic methods. They obtained the density theorem concerning the zeros of L-series, and A. Zulauf (1952) continued along the same lines. These methods had been suggested by G. H. Hardy and J. E. Littlewood (c. 1919), although they had assumed that the †extended Riemann hypothesis held.

We denote by P_j a number having at most j prime factors counted with their multiplicity. Using his sieve method, Brun (1919) proved that every sufficiently large even integer $2n$ can be expressed in the form $2n = P_9 + P_9$. Rademacher (1924) improved this result and obtained $2n = P_7 + P_7$. Applying his new sieve method to this problem, Selberg (1950) proved that $2n = P_2 + P_3$. On the other hand, using the †large sieve, A. Rényi proved that $2n = P_1 + P_k$ for some k. Afterwards, combining Richert's sieve with a large sieve, Chen Jing-Run (1973) proved that $2n = P_1 + P_2$ for large $2n$ [11].

D. Polygonal Numbers

Let m be an integer greater than 3, and let $a_1 = 1$, $a_{n+1} - a_n = (m-2)n + 1$ $(n = 1, 2, \ldots)$. The sequence $\{a_n\}$ forms the system of **polygonal numbers of order** m. The general term of $\{a_n\}$ is given by $n + \frac{1}{2}(m-2)(n^2 - n)$ $(n = 1, 2, \ldots)$. Such a_n are said to be **triangular numbers** if $m = 3$, **square numbers** if $m = 4$, and **pentagonal numbers** if $m = 5$.

P. Fermat (1636) stated that every natural number can be expressed as the sum of m polygonal numbers of order m. This conjecture was proved by A. M. Legendre (1798) for $m = 3$, by J. L. Lagrange (1772) for $m = 4$, and by A. L. Cauchy (1813) for the general case. With regard to Lagrange's result, Legendre noticed that in order that a positive integer n be expressed as the sum of three squares, it is necessary and sufficient that n not be of the form $4^a(8m + 7)$.

Given a positive integer n, the number of integral solutions of the equation $x_1^2 + x_2^2 + \ldots + x_s^2 = n$ is denoted by $r_s(n)$. For example, $r_2(5) = 8$. The †generating function $\sum_{n=1}^{\infty} r_2(n) n^{-s}$ can be expressed as $\sum_{m,n=-\infty}^{+\infty} (m^2 + n^2)^{-s}$, where the term corresponding to $m = n = 0$ is omitted; the function is equal to $4\zeta_K(s)$, where $\zeta_K(s)$ is the †Dedekind zeta function of the Gaussian number field $K = \mathbf{Q}(\sqrt{-1})$. The equation $4\zeta_K(s) = 4\zeta(s) L(s, \chi)$ (where $\chi(n) = (-4/n)$) leads to

$$r_2(n) = 4 \sum_{m|n}{}' (-1)^{(m-1)/2},$$

where Σ' means the sum over all odd factors m of n. This result was obtained by C. G. J. Jacobi (1829). He also obtained the following formula:

$$r_4(n)=8\sum_{m|n,4\nmid m}' m,$$

where Σ' means the sum over all divisors m of n not divisible by 4 (Hardy and E. M. Wright, C. L. Siegel, 1964). Let $q=\exp(2\pi i\tau)$ $(\operatorname{Im}\tau>0)$, and

$$f(q)=1+\sum_{n=1}^{\infty} r_s(n)q^n$$
$$=(1+2q+2q^4+2q^9+\ldots)^s.$$

Hardy (1920) considered the variation of $f(q)$ for $q=a\exp(2\pi ih/k)$ (with $0<a<1$) as $a\to 1$; he obtained

$$f(q)\sim\pi^{s/2}\left(\frac{S_{h,k}}{k}\right)^s\log\left(\frac{1}{a}\right)^{-s/2},$$

where $S_{h,k}=\sum_{j=1}^k\exp(2\pi i(h/k)j^2)$. Furthermore, he constructed the **singular series**

$$\rho_s(n)=\frac{\pi^{s/2}n^{(s/2)-1}}{\Gamma(s/2)}\sum_{k=1}^{\infty}A_k,\quad A_1=1,$$

$$A_k=k^{-s}\sum_{\substack{1\leqslant h\leqslant k\\(h,k)=1}}(S_{h,k})^s\exp\left(-2\pi i\frac{h}{k}n\right),$$

and showed that $r_s(n)=\rho_s(n)$ for $s=5$, 6, 7, and 8. P. T. Bateman (1951) proved the same equation for $s=3$ and 4. If $s\geqslant 9$ then $r_s(n)=\rho_s(n)+O(n^{s/4})$ (Hardy, Littlewood, and S. Ramanujan, c. 1919). The detailed exposition of this result is in the notes of A. Z. Walfisz (1952) and Rademacher [4]. H. D. Kloostermann (1926) and Estermann (1962) studied the equation $ax_1^2+bx_2^2+cx_3^2+dx_4^2=n$, which led to a new field of study concerning the †Kloostermann sum. For instance, estimates such as

$$\left|\sum_{x=1}^{p-1}\exp\left(2\pi i\frac{cx+d/x}{p}\right)\right|\leqslant 2\sqrt{p}$$

are obtained by using the theory of zeta functions in algebraic function fields in one variable (A. Weil, 1948) (→ 450 Zeta Functions).

E. Waring's Problem

The first formulation of **Waring's problem** is found in E. Waring, *Meditationes algebraicae* (1770), in which he discusses the problem of expressing an arbitrary positive integer as the sum of at most nine cubes or as the sum of at most nineteen biquadratics. Hilbert proved (→ Section A) that there exists a positive integer $s(k)$ such that, for any integer N, the equation

$$x_1^k+x_2^k+\ldots+x_s^k=N$$

has a nonnegative integral solution if $s\geqslant s(k)$. We denote by $g(k)$ the least value of $s(k)$, and by $G(k)$ the least value of $s(k)$ for which the equation is solvable with at most finitely many exceptions of N. Research concerning $g(k)$ and $G(k)$ received its initial impetus from the circle method considered by Hardy and Littlewood, and it underwent considerable development in the works of H. Weyl and Vinogradov.

Let $r_s(N)$ be the number of solutions of the above equation. We then have

$$r_s(N)=\int_0^1\left(\sum_{x\leqslant N^{1/k}}\exp(2\pi i\alpha x^k)\right)^s\times\exp(-2\pi iN\alpha)\,d\alpha.$$

If we make the Farey dissection, translating the interval $[0,1]$ slightly, then the main term of $r_s(N)$ is provided by the integrals over major arcs, and the residual term is derived from the integrals over minor arcs. Let p be a prime, and let $M(N,p^l)$ denote the number of solutions of the congruence equation

$$x_1^k+x_2^k+\ldots+x_s^k\equiv N(\operatorname{mod}p^l).$$

Then $\lim_{l\to\infty}M(N,p^l)/p^{l(s-1)}=\chi_p(N)$ is not zero, and the infinite product $\prod_p\chi_p(N)=S(N)$ converges for $s\geqslant 4k$, where $S(N)$ is larger than a positive constant which is determined independently of the choice of N. On the other hand, let

$$S(a,q)=\sum_{x=0}^{q-1}\exp\left(2\pi i\frac{a}{q}x^k\right),$$

$$A(q,N)=q^{-s}\sum_{\substack{1\leqslant a\leqslant q\\(a,q)=1}}S(a,q)^s\exp\left(-2\pi i\frac{a}{q}N\right).$$

Then $\sum_{q=1}^{\infty}A(q,N)$ is absolutely convergent, and the sum is equal to $S(N)$. According to Hua (1959) we have

$$r_s(N)\sim S(N)\frac{\Gamma(1+1/k)^s}{\Gamma(s/k)}N^{s/k-1},$$

provided that $s\geqslant 2k^2(2\log k+\log\log k+c)$. Next, if we denote by $V(N,\delta)$ the volume of the closed region satisfying $N\leqslant x_1^k+x_2^k+\ldots+x_s^k\leqslant N+\delta$ in the s-dimensional Euclidean space, then $\lim_{\delta\to 0}V(N,\delta)/\delta=\chi_\infty(N)$ exists and is equal to $(\Gamma(1+1/k)^s/\Gamma(s/k))N^{s/k-1}$. Hence we can show that the principal part of $r_s(N)$ is equal to the infinite product $\prod_p\chi_p(N)$, where p runs over all finite and infinite prime spots in $\mathbf{Q}$. This is a generalization of the singular series studied by Hardy.

With regard to $g(k)$, there are studies by L. E. Dickson (1936) and others. It is easy to see that $g(k)\geqslant 2^k+(3/2)^k-2$ and that $G(k)\geqslant k+1$. It has been shown that $g(3)=9$, $G(3)\leqslant 7$ (G. L. Watson, Linnik, 1947) and that $g(4)\leqslant 37$, $G(4)=16$ (H. Davenport, 1939). More generally, Vinogradov (1959) proved that

$$G(k) \leqslant 2k\log k + 4k\log\log k + 2k\log\log\log k + ck.$$

To prove this, Vinogradov (1934) introduced the following integral, which is closely related to the †prime number theorem:

$$I(P) = \int_0^1 \dots \int_0^1 \left| \sum_{x \leqslant P} \exp(2\pi i(\alpha_1 x + \alpha_2 x^2 + \dots + \alpha_k x^k)) \right|^{2s} d\alpha_1 \dots d\alpha_k.$$

Hua (1949) improved Vinogradov's result and proved that if $s \geqslant \frac{1}{4}k(k+1) + lk$, then

$$I(P) \leqslant (5s)^{5sl}(\log P)^{2l} P^{2s - k(k+1)/2 + \delta},$$

where $\delta = \frac{1}{2}k(k+1)(1-1/k)^{l-1}$. Concerning $I(P)$, another notable approach was made by A. A. Karacuba and N. M. Korobov (1963). Further investigation proved that $I(P) = c_1 c_2 P^{2s-k(k+1)/2} + o(P^{2s-k(k+1)/2})$ if $s \geqslant ck^2 \log k$ (Vinogradov, Hua, 1959). The result is called the **Vinogradov mean value theorem.**

There are many variations and generalizations of this theorem. Vinogradov and Hua (1944) studied the problem of representing an arbitrary N as $N = p_1^k + p_2^k + \dots + p_s^k$ (with p_i prime). Hua (1937) and others also considered the problem of representing N as $N = f(x_1) + f(x_2) + \dots + f(x_s)$, where $f(x)$ is a given polynomial. Also, let

$$C(x_1, x_2, \dots, x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} C_{ijk} x_i x_j x_k$$

be a homogeneous polynomial of degree 3 with integral coefficients. Davenport (1963) proved that if $n \geqslant 16$, then the equation $C(x_1, x_2, \dots, x_n) = 0$ has at least one nontrivial integral solution. There are further developments of the theory of representations of integers by forms in many variables by V. A. Tartakovskiĭ, H. Davenport, B. J. Birch, and D. J. Lewis.

F. Additive Problems in an Algebraic Number Field

Siegel (1922) considered the generalization of Hardy's square sum problem to the case of an algebraic number field. He later (1945) studied the generalized Waring's problem in an algebraic number field K of finite degree: Let I be the principal order of K, and let J_k be the subring of I generated by the kth powers of integers in K. It is easily seen that the index $(I : J_k)$ is finite. Hence our concern regarding $s(k)$ must be restricted to integers contained in J_k. Another question is how to extend the Farey dissection to an algebraic number field. Siegel succeeded in solving these difficulties. His ingenuity is seen in his way of dealing with the minor arcs, which provided a stimulus to the research of T. Mitsui (1960).

A generalization of Goldbach's problem to the case of an algebraic number field was obtained by Mitsui (1960) and O. Körner (1961).

As another extension of the Vinogradov three primes theorem, Mitsui (1971) proved the following theorem: Let K be an algebraic number field of degree n. Let C be the principal ideal class generated by a totally positive number in K, and P be the set of prime ideals of degree 1 contained in C. Let N be a positive integer and $I_s(N)$ be the number of representations of N as the sum of the norms of s prime ideals belonging to P,

$$I_s(N) = \sum_{N = N\mathfrak{p}_1 + \dots + N\mathfrak{p}_s} 1, \quad \mathfrak{p}_i \in P \quad (1 \leqslant i \leqslant s).$$

If N is sufficiently large and $s \geqslant 3$, we have the asymptotic formula

$$I_s(N) = A_s S(N) \frac{N^{s-1}}{(\log N)^s} + O\left(N^{s-1} \frac{\log\log N}{(\log N)^{s+1}} \right),$$

where A_s is a positive constant depending on s and K independent of N, and $S(N)$ denotes the singular series. If $s \equiv N \pmod{2D}$, where D is the discriminant of K, then $S(N) \geqslant c > 0$, where c is a constant. Later this problem was extended by Mitsui and T. Tatuzawa (1981).

References

[1] R. G. Ayoub, An introduction to the analytic theory of numbers, Amer. Math. Soc. Math. Surveys, 1963.
[2] G. H. Hardy and E. M. Wright, An introduction to the theory of numbers, Clarendon Press, fourth edition, 1963.
[3] H. H. Ostmann, Additive Zahlentheorie I, II, Erg. Math., Springer, 1956.
[4] H. A. Rademacher, Lectures on analytic number theory, Lecture notes, Tata Inst., 1954–1955.
[5] E. G. H. Landau, Über einige neuere Fortschritte der additiven Zahlentheorie, Cambridge Univ. Press, 1937.
[6] L. K. Hua, Additive Primzahltheorie, Teubner, 1959.
[7] L. K. Hua, Die Abschätzung von Exponentialsummen und ihre Anwendung in der Zahlentheorie, Enzykl. Math., Bd. I, 2, Heft 13, Teil I, 1959.
[8] I. M. Vinogradov, Selected works (in Russian), Akad. Nauk SSSR, 1952.
[9] I. M. Vinogradov, Upper boundary of $G(n)$ (in Russian), Izv. Akad. Nauk SSSR, 23 (1959), 637–642.
[10] T. Mitsui, On the Goldbach problem in

an algebraic number field I, II, J. Math. Soc. Japan, 12 (1960), 290–324, 325–372.
[11] J.-R. Chen, On the representation of a large even integer as the sum of a prime and the product of at most two primes, Sci. Sinica, 16 (1973), 157–176.
[12] T. Mitsui and T. Tatuzawa, Sur un problème en théorie additive des nombres, Monographies Scientifiques de la Maison Franco-Japonaise, fasc. 3, 1981.

5 (XVII.9) Additive Processes

A. General Remarks

The study of sums of †independent random variables has been one of the main topics in modern probability theory (→ 250 Limit Theorems in Probability Theory). Combining the ideas of this study with the consideration of †stochastic processes with continuous time parameter, we get the notion of additive processes.

B. Definitions and Fundamental Properties

A real-valued †stochastic process $\{X_t\}_{0\leqslant t<\infty}$, denoted by $X(t)$ $(0\leqslant t<\infty)$ for the rest of this article, where for simplicity we assume that $X(0)=0$, is called an **additive process** (or **process with independent increments**) if for any $t_0<t_1<\ldots<t_n$, $X(t_i)-X(t_{i-1})$ $(i=1,2,\ldots,n)$ are †independent. An additive process is essentially the same as a †spatially homogeneous †Markov process (i.e., a Markov process on $\mathbf{R}^1$ that is invariant under translations). When, for any $h>0$ and $t>s$, $X(t+h)-X(s+h)$ and $X(t)-X(s)$ have the same law, i.e., the distribution law of $X(t)-X(s)$ depends only on $t-s$, we call the additive process $X(t)$ **temporally homogeneous**. This is essentially the same notion as a temporally and spatially homogeneous Markov process.

Let $X(t)$ be a given additive process. If $f(t)$ is a function of t only, then clearly $Y(t)=X(t)-f(t)$ is also an additive process, and we can choose $f(t)$ in such a way that for every $t>0$ and for every sequence $s_n\uparrow t$ $(s_n\downarrow t)$, $Y(s_n)$ converges with probability one. Here, $\lim Y(s_n)$ is independent of a particular choice of s_n, and we denote it by $Y(t-)$ $(Y(t+))$. We call such $Y(t)$ a **central process** and also say that $Y(t)$ is obtained from $X(t)$ by **centering**. This $f(t)$ is given, for example, by the condition $E(\arctan(X(t)-f(t)))=0$.

Let $Y(t)$ be a centered additive process. Then $Y(t-)=Y(t)=Y(t+)$ for all $t>0$, except on an at most countable t-set S, and $t\in S$ is called a **fixed point of discontinuity** of $Y(t)$. Then $Y_1(t)=\lim_{n\to\infty}U_n(t)$ exists with probability 1, where

$$U_n(t)=\sum_{\substack{0\leqslant i\leqslant n\\ s_i<t}}(Y(s_i+)-Y(s_i-))+Y(t)-Y(t-)-C_t^n,$$

C_t^n being a constant determined by $E(\arctan U_n(t))=0$, and $S=\{s_j\}$ $(j=1,2,\ldots)$. $Y_2(t)=Y(t)-Y_1(t)$ is a centered additive process without any fixed point of discontinuity. Furthermore, $Y_1(t)$ and $Y_2(t)$ are independent. Thus we have a decomposition of $Y(t)$: $Y(t)=Y_1(t)+Y_2(t)$, where $Y_1(t)$ and $Y_2(t)$ are mutually independent additive processes. The structure of $Y_1(t)$ is simple, and it is not worthwhile to study its behavior in more detail. On the other hand, since $Y_2(t)$ is a centered additive process without any fixed point of discontinuity, it is an additive process that is †continuous in probability. Let $\tilde{Y}_2(t)$ be a †separable modification of $Y_2(t)$. Then the discontinuities of almost all sample functions of $\tilde{Y}_2(t)$ are of the first kind. If we set $Y_2^*(t)=\tilde{Y}_2(t+)$, $Y_2^*(t)$ is a †modification of $Y_2(t)$, and almost all sample functions of $Y_2^*(t)$ are right continuous and have a left-hand limit at every t. In the study of the process $Y_2(t)$, it is always convenient to take such a modification. Thus we give the following general definition: An additive process is called a **Lévy process** if it is continuous in probability and almost all sample functions are right continuous and have a left-hand limit at every $t\in[0,\infty)$ [3,9].

The notions of additive processes and Lévy processes can also be considered for $\mathbf{R}^N$-valued processes.

C. Additive Processes and Infinitely Divisible Distributions

Let $X(t)$ be a Lévy process and Φ_{st} $(s<t)$ be the †distribution of $X(t)-X(s)$. Then Φ_{st} is an †infinitely divisible distribution (→ 341 Probability Measures G). Conversely, for a given infinitely divisible distribution Φ we can construct an essentially unique temporally homogeneous Lévy process $X(t)$ such that Φ coincides with the distribution of $X(1)$. If $X(t)$ is temporally homogeneous, the †characteristic function $\varphi_{st}(z)=E(e^{iz(X(t)-X(s))})$ of the distribution Φ_{st} is given in the form $\varphi_{st}=\exp((t-s)\psi(z))$; hence the law of the process $X(t)$ is completely determined by the function $\psi(z)$. By the †Lévy-Khinchin canonical form, $\psi(z)$ is

written in the form

$$\psi(z)=imz-\frac{v}{2}z^2+\int_{-\infty}^{\infty}\left(e^{izu}-1-\frac{izu}{1+u^2}\right)n(du), \tag{1}$$

where $m, v\in\mathbf{R}$, $v\geqslant 0$, and $n(du)$ is a nonnegative measure on $\mathbf{R}-\{0\}$ such that $\int_{-\infty}^{\infty}\frac{u^2}{1+u^2}n(du)<\infty$. These m, v, and $n(du)$ are uniquely determined by $\psi(z)$.

D. Basic Additive Processes

Wiener Process. When almost all sample functions of a Lévy process $X(t)$ are continuous, the distribution of $X(t)-X(s)$ is a †normal distribution. If, further, $X(t)$ is temporally homogeneous, $\psi(z)$ has the form $\psi(z)=imz-\frac{1}{2}vz^2$. In particular, if $m=0$ and $v=1$, then $X(t)$ is called a **Wiener process** or **Brownian motion**. This stochastic process was introduced by N. Wiener (1923) as a mathematical model for the random movement of colloidal particles first observed by a British botanist, R. Brown. This is one of the most fundamental and important stochastic processes in modern probability theory (→ 45 Brownian Motion).

Poisson Processes. When almost all sample functions of a Lévy process are increasing step functions with only jumps of size 1, the distribution of $X(t)-X(s)$ is a †Poisson distribution. If, further, $X(t)$ is temporally homogeneous, $\psi(z)$ in (1) has the form $\psi(z)=\lambda(e^{iz}-1)$ $(\lambda>0)$, and $X(t)$ is called a **Poisson process**. Let $X(t)$ be a Poisson process, and let T_0, T_0+T_1, $T_0+T_1+T_2, \ldots$ be successive jumping times of a sample function $X(t)$. Then $T_0, T_1, T_2, \ldots$ is a sequence of mutually independent random variables with the common exponential distribution $P(T\in dt)=\lambda e^{-\lambda t}dt$. Conversely, given such a sequence $\{T_n\}$, if we define $X(t)=\inf\{n \mid T_0+T_1+\ldots+T_n>t\}$, then $X(t)$ is a Poisson process. Thus, for example, the number of telephone calls at a switchboard is a Poisson process when the intervals between successive calls can be regarded as independent and having a common exponential distribution.

E. The Structure of the General Lévy Process [8, 9]

In this section we restrict ourselves for simplicity to temporally homogeneous Lévy processes. As we noted, the probability law of the process $X(t)$ is determined by the function $\psi(z)$. The Lévy-Khinchin formula (1) in a certain sense shows that ψ is a combination of a Wiener process and Poisson processes. This fact can be seen more clearly from the **Lévy-Itô theorem**, which states that the sample function of $X(t)$ itself can be expressed as a composite of those of a Wiener process and Poisson processes. The Lévy-Itô theorem actually implies formula (1) and, moreover, clarifies its probabilistic meaning.

The Lévy-Itô theorem can be summarized as follows: Let U be a Borel subset of $\mathbf{R}$ which has a positive distance from the origin, and let $N(t, U)$ be the number of s such that $X(s)-X(s-)\in U$, $s\leqslant t$. Then $N(t, U)$ is a Poisson process. The expectation $E(N(t, U))$ can be written in the form $tn(U)$, where $n(U)$ defines a nonnegative Borel measure on $\mathbf{R}-\{0\}$; furthermore, it satisfies

$$\int_{-\infty}^{\infty}\frac{u^2}{1+u^2}n(du)<\infty.$$

Next, we set

$$S_\varepsilon(t)=\int_{|u|>\varepsilon}uN(t,du) = \sum_{\substack{s\leqslant t\\ |X(s)-X(s-)|\geqslant\varepsilon}}(X(s)-X(s-)).$$

Generally, $S_\varepsilon(t)$ diverges as $\varepsilon\downarrow 0$. However, with probability one, a centered process

$$\bar{S}_\varepsilon(t)=S_\varepsilon(t)-t\int_{|u|>\varepsilon}\frac{u}{1+u^2}n(du)$$

converges uniformly in t on every finite interval as $\varepsilon\downarrow 0$. Furthermore, $X(t)-\lim_{\varepsilon\downarrow 0}\bar{S}_\varepsilon(t)$ is continuous with probability one. However, a Lévy process $X(t)$ of which almost all sample functions are continuous has the form $mt+\sqrt{v}\,B(t)$, where $m, v\geqslant 0$ are constants and $B(t)$ is a Wiener process. Hence we have

$$X(t)=mt+\sqrt{v}\,B(t) + \lim_{\varepsilon\downarrow 0}\int_{|u|>\varepsilon}\left(uN(t,du)-\frac{u}{1+u^2}tn(du)\right). \tag{2}$$

Furthermore, we can show that if

$U_1, U_2, \ldots, U_n$

are disjoint, then

$B(t), N(t, U_1), N(t, U_2), \ldots, N(t, U_n)$

are mutually independent Lévy processes. In particular, in (2) the terms are mutually independent. The m, v, and $n(du)$ in (2) correspond, of course, to those in (1). Conversely, given m, v, and $n(du)$, we can construct $B(t)$ and $N(t, U)$ with the properties above, and then (2) defines a Lévy process which corresponds to $\psi(z)$ given by (1). The measure $n(du)$ is called the **Lévy measure** of $X(t)$.

F. Examples of Lévy Processes

Compound Poisson Processes. A temporally homogeneous Lévy process is called a **compound Poisson process** if almost all sample functions are step functions, namely, if $\psi(z)$ in (1) is given by

$$\psi(z)=\int_{-\infty}^{\infty}(e^{izu}-1)n(du),$$

$$\lambda=\int_{-\infty}^{\infty}n(du)<\infty.$$

If we set $\Phi(du)=(1/\lambda)n(du)$, then $\Phi(du)$ is a probability distribution on $\mathbf{R}$. Φ is the distribution of the size of jumps when they occur. A compound Poisson process is constructed in the following way: Let $T_0, T_1, T_2, \ldots; U_1, U_2, \ldots$ be mutually independent random variables such that

$$P(T_n\in dt)=\lambda e^{-\lambda t}dt, \qquad t>0,$$

$$P(U_n\in du)=\Phi(du),$$

and let

$$N(t)=U_1+U_2+\ldots+U_{N(t)},$$

where

$$n=\inf\{n\mid T_0+T_1+\ldots+T_n>t\}.$$

Then $X(t)$ is a compound Poisson process. Thus the number of jumps of $X(t)$ follows a Poisson process, and the size of each jump obeys the distribution Φ.

Stable Processes. A temporally homogeneous Lévy process $X(t)$ is called a **stable process** if for every $a>0$ we can find $b>0$ and c real such that the processes $X_1(t)=X(at)$ and $X_2(t)=bX(t)+ct$ are equivalent in law. It is called a **strictly stable process** if in the above c can be chosen to be 0. $X(t)$ is a stable process (resp. a strictly stable process) if and only if the corresponding infinitely divisible distribution is a †quasistable distribution (resp. †stable distribution). The †exponent α $(0<\alpha\leqslant 2)$ of the quasistable distribution is called the **exponent** (or **index**) **of the stable process.** $\psi(z)$ in (1) corresponding to a stable process is given as follows:

$$\psi(z)=imz+C_+\int_0^{\infty}(e^{izu}-1)\frac{du}{u^{1+\alpha}}$$

$$+C_-\int_{-\infty}^{0}(e^{izu}-1)\frac{du}{|u|^{1+\alpha}}, \quad 0<\alpha<1,$$

$$\psi(z)=imz+C_+\int_0^{\infty}\left(e^{izu}-1-\frac{izu}{1+u^2}\right)\frac{du}{u^2}$$

$$+C_-\int_{-\infty}^{0}\left(e^{izu}-1-\frac{izu}{1+u^2}\right)\frac{du}{|u|^2}, \quad \alpha=1,$$

$$\psi(z)=imz+C_+\int_0^{\infty}(e^{izu}-1-izu)\frac{du}{u^{1+\alpha}}$$

$$+C_-\int_{-\infty}^{0}(e^{izu}-1-izu)\frac{du}{|u|^{1+\alpha}}, \quad 1<\alpha<2,$$

$$\psi(z)=imz-\frac{v}{2}z^2, \quad \alpha=2.$$

Here m, C_+, C_- , and v are real constants such that $C_+\geqslant 0$, $C_-\geqslant 0$ with $C_++C_->0$ and $v>0$. $\psi(z)$ corresponds to a strictly stable process if and only if $m=0$ when $\alpha\neq 1$ and $C_+=C_-$ when $\alpha=1$. The above $\psi(z)$ is also expressed as follows:

(i) if $0<\alpha\leqslant 2$, $\alpha\neq 1$, then

$$\psi(z)=imz-c_0|z|^{\alpha}\left(1-i\beta\tan\frac{\pi\alpha}{2}\frac{z}{|z|}\right),$$

where

$$c_0=\frac{v}{2} \quad \text{if} \quad \alpha=2$$

and

$$c_0=-(C_++C_-)\Gamma(-\alpha)\cos\frac{\pi\alpha}{2},$$

$$\beta=\frac{C_+-C_-}{C_++C_-} \quad \text{if} \quad \alpha\neq 2;$$

(ii) if $\alpha=1$, then

$$\psi(z)=i\gamma z-c_0|z|\left(1+i\frac{2}{\pi}\beta\frac{z}{|z|}\log|z|\right),$$

where

$$\gamma=m+a(C_++C_-)$$

$$\left(a=\int_0^{\infty}\left(\sin u-\frac{u}{1+u^2}\right)\frac{du}{u^2}\right),$$

$$c_0=\frac{\pi}{2}(C_+-C_-), \quad \beta=\frac{C_+-C_-}{C_++C_-}.$$

When $\alpha=2$, it is thus essentially a Wiener process: $X(t)=mt+\sqrt{v}\,B(t)$, where $B(t)$ is a Wiener process. When $\alpha=1$, it is called a **Cauchy process**, a **symmetric Cauchy process** if $m=0$ and $C_+=C_-$ or equivalently $\gamma=\beta=0$, or an **asymmetric Cauchy process** if $\beta\neq 0$. Generally it is called a **symmetric stable process** if $m=0$ and $C_+=C_-$ or $\alpha=2$ and $m=0$. In particular, for the symmetric Cauchy process corresponding to $\psi(z)=-|z|$, we have

$$P(X(t)<x)=\frac{1}{\pi}\int_{-\infty}^{x}\frac{t}{t^2+y^2}dy.$$

Next, when $0<\alpha<1$ and $m=C_-=0$, almost all sample functions of $X(t)$ are purely discontinuous increasing functions (i.e., sums of positive jumps). In this case, $X(t)$ is called a **one-sided stable process of the exponent** α (or **subordinator of the exponent** α). Now, if $X(t)$ is

a symmetric stable process of the exponent β $(0<\beta\leqslant 2)$ and $Y(t)$ is a subordinator of the exponent α which is independent of $X(t)$, then $Z(t)=X(Y(t))$ is a symmetric stable process of the exponent $\alpha\beta$. This operation is called a **subordination** and is closely related to the theory of †fractional powers of †infinitesimal generators of semigroups (→ 261 Markov Processes) [2].

A stable process $X(t)$ is defined in a similar way when $X(t)$ takes values in an N-dimensional space $\mathbf{R}^N$. In particular, if $X(t)$ is a symmetric stable process of the exponent α $(0<\alpha\leqslant 2)$ given by

$$E(e^{i(z,X(t))})=e^{-t|z|^\alpha},\quad z\in\mathbf{R}^N,$$

then for every bounded measurable function $f(x)$ with compact support, we have

$$E\left(\int_0^\infty f(x+X(t))\,dt\right)=\frac{\Gamma((N-\alpha)/2)}{2^\alpha\pi^{N/2}\Gamma(\alpha/2)}\int_{\mathbf{R}^N}|x-y|^{\alpha-N}f(y)\,dy.$$

The right-hand side is the †Riesz potential of order α (→ 338 Potential Theory). This fact is a generalization of a well-known relation between Brownian motion and Newtonian potential (→ 45 Brownian Motion), and through this relation we can study several properties of sample functions and also compute various quantities related to stable processes [1].

G. Sample Path Properties of Lévy Processes

Let $X(t)$ be a temporally homogeneous Lévy process. For a Borel set B in $\mathbf{R}$, the hitting time σ_B is defined by

$$\sigma_B=\inf\{t>0\,|\,X(t)\in B\}.$$

Recurrence. $X(t)$ is called **recurrent** if $\sigma_B<\infty$ a.s. for every nonempty open set B. Otherwise it is called **transient**. $X(t)$ is recurrent if and only if $\int_{-1}^{1}\mathrm{Re}\frac{1}{-\psi(z)}dz=\infty$ (Port and Stone, Ornstein). If, in particular, $E(X(1))$ exists, it is recurrent if and only if $E(X(1))=0$. When $X(t)$ is a stable process, it is recurrent if and only if $\alpha>1$, or $\alpha=1$ and $\beta=0$.

Hitting Probabilities for Single Points. If $B=\{a\}$, σ_B is denoted by σ_a. $0\in\mathbf{R}$ is said to be **regular** for $X(t)$ if $\sigma_0=0$ a.s. Set $C=\{x\in\mathbf{R}\,|\,P(\sigma_x<\infty)>0\}$. The following result is due to Kesten [10]:

(i) If $v\neq 0$, then $C=\mathbf{R}$ and 0 is regular;
(ii) If $v=0$ and $\int_{-1}^{1}|u|n(du)=\infty$, then either $C=\mathbf{R}$ and 0 is regular, or $C=\varnothing$ and 0 is not regular according to whether $\int_{\mathbf{R}}\mathrm{Re}\frac{1}{\lambda-\psi(z)}dz<\infty$ for all $\lambda>0$ or $=\infty$ for some $\lambda>0$;
(iii) If $v=0$ and $\int_{-1}^{1}|u|n(du)<\infty$, then the following four cases are possible, where $a=m-\int_{\mathbf{R}}u(1+u^2)^{-1}n(du)$ and $S(n)$ is the support of the Lévy measure n:
(a) $C=\varnothing$ when $a=0$,
(b) $C=(0,\infty)$ when $a>0$ and $S(n)\subset(0,\infty)$,
(c) $C=(-\infty,0)$ when $a<0$ and $S(n)\subset(-\infty,0)$, and
(d) $C=\mathbf{R}$ in the remaining case.

For further properties of sample functions → [4].

H. Generalization of Additive Processes

A temporally homogeneous Lévy process is, as we have seen, essentially a temporally homogeneous Markov process on $\mathbf{R}$ which is homogeneous in space (i.e., invariant under translations of the space). Thus, on a homogeneous space when homogeneity in space makes sense, we can generalize the notion of additive processes. Let M be a †homogeneous space with transformation group G. A temporally homogeneous Markov process $X(t)$ is called an **invariant Markov process**, (or **homogeneous Markov process**) if its system of †transition probabilities $\{P(t,x,E)\}$ satisfies $P(t,x,E)=P(t,gx,gE)$ for all $g\in G$. Thus an additive process is exactly an invariant Markov process on $\mathbf{R}^N$ when G is the group of translations. G. A. Hunt determined all invariant Markov processes when M is a †Lie group or a factor space of a Lie group [6].

Let G be a Lie group and $\Lambda=\Lambda(G)$ be the (left-invariant) †Lie algebra of G. Let $G_c=G\cup\{\omega\}$ be a †one-point compactification of G, and C be the set of all continuous functions on G_c. We can define $Yf(Y\in\Lambda, f\in C)$ as usual by

$$Y(f)=\lim_{t\downarrow 0}\frac{R_{\eta(t)}f-f}{t},$$

$$\eta(t)=\exp tY,\ R_\sigma f(\tau)=f(\tau\sigma),$$

when the limit exists uniformly. Let $C_2=\{f\in C\,|\,Y(Xf)$ exists for every $X,Y\in\Lambda\}$. Let $X_1,X_2,\ldots,X_d$ be a basis of $\Lambda(G)$, and let $x_1,x_2,\ldots,x_d$ be functions in C_2 such that $x_i(e)=0$ and $X_i(x_j)(e)=\delta_{ij}$ $(i,j=1,2,\ldots,d$; e is the unit element of $G)$. Take a neighborhood of e, and define a function $\varphi(g)=\sum_{i=1}^d x_i^2(g)$ for g contained in the neighborhood, and extend this function to G_c in such a way that $\varphi\in C_2$ and $\varphi\geqslant k>0$ $(k=\text{constant})$ outside of the neighborhood of e. Then $g\in G$ defines a transformation of G_c by $\tau_g\sigma=g\sigma$, $\tau_g\omega=\omega$, and in this way G_c is supplied with the structure of a †homogeneous space with the transformation

group G. Now let $X(t)$ be an invariant Markov process on G_c which is †continuous in probability. Then the †semigroup T_t (which is a †strongly continuous semigroup on C) of the process $X(t)$ is characterized as follows: The domain of the infinitesimal generator A of T_t contains C_2, and for $f \in C_2$

$$Af(\tau)=\sum_{i=1}^{d} a_i(X_i f)(\tau)+\sum_{i,j=1}^{d} a_{ij}X_i(X_j f)(\tau)$$
$$+\int_{G_c-\{e\}}\left\{f(\tau\sigma)-f(\tau)-\sum_{i=1}^{d} X_i f(\tau)\cdot x_i(\sigma)\right\}n(d\sigma),$$

where a_i, a_{ij} are real numbers $(i,j=1,2,\ldots,d)$ such that (a_{ij}) is a symmetric nonnegative definite matrix, and $n(d\sigma)$ is a nonnegative measure on $G_c-\{e\}$ such that $\int_{G_c-\{e\}}\varphi(\sigma)\cdot n(d\sigma)<\infty$. Conversely, given such a_i, a_{ij} and $n(d\sigma)$, there exists one and only one invariant Markov process on G_c whose infinitesimal generator is given as above.

A similar result is obtained when M is a factor space of a Lie group by its compact subgroup. Furthermore, for more concrete homogeneous spaces such as spheres or Lobachevskiĭ spaces (more generally, †symmetric Riemannian spaces) the canonical form of the invariant Markov processes and infinitely divisible laws is obtained by making use of harmonic analysis [5].

References

[1] R. M. Blumenthal, R. K. Getoor, and D. B. Ray, On the distribution of the first hits for the symmetric stable processes, Trans. Amer. Math. Soc., 99 (1961), 540–554.
[2] S. Bochner, Harmonic analysis and the theory of probability, Univ. of California Press, 1955.
[3] J. L. Doob, Stochastic processes, Wiley, 1953.
[4] B. Fristedt, Sample functions of stochastic processes with stationary, independent increments, Advances in Probability and Related Topics III, P. Ney and S. Port (eds.), Dekker, 1974, 241–396.
[5] R. Gangolli, Isotropic infinitely divisible measures on symmetric spaces, Acta Math., 111 (1964), 213–246.
[6] G. A. Hunt, Semi-groups of measures on Lie groups, Trans. Amer. Math. Soc., 81 (1956), 264–293.
[7] K. Itô. On stochastic processes I, Japan. J. Math., 18 (1942), 261–301.
[8] K. Itô, On stochastic processes, Lecture notes, Tata Inst., 1960.
[9] K. Itô, Stochastic processes, Lecture Notes Series, no. 16, Aarhus Univ., 1969.
[10] H. Kesten, Hitting probabilities of single points for processes with stationary independent increments, Mem. Amer. Math. Soc., 93 (1969).
[11] P. Lévy, Théorie de l'addition des variables aléatoires, Gauthier-Villars, 1937.
[12] P. Lévy, Processus stochastiques et mouvement brownien, Gauthier-Villars, 1948.

6 (III.20) Adeles and Ideles

A. Introduction

The concept of idele was first introduced by C. Chevalley (*J. Math. Pures Appl.*, (9) 15 (1936); *Ann. Math.*, (2) 41 (1940)), for †algebraic number fields. Later on, this concept and the allied concept of adele were defined for †simple algebras and also for †algebraic groups over algebraic number fields, and the two concepts became important in the arithmetical theory of these objects. We shall first explain the general concept of restricted direct product, by means of which adeles and ideles will be defined.

B. Restricted Direct Product

Let I be an index set. Suppose we are given, for each $\mathfrak{p}\in I$, a †locally compact group $G_\mathfrak{p}$, and for each $\mathfrak{p}$ except for a given finite set, say $\mathfrak{p}_1,\mathfrak{p}_2,\ldots,\mathfrak{p}_r$, a compact open subgroup $U_\mathfrak{p}$ of $G_\mathfrak{p}$. Let G be the subgroup of the direct product $\prod_{\mathfrak{p}\in I} G_\mathfrak{p}$ consisting of elements $(g_\mathfrak{p})$ whose $G_\mathfrak{p}$-components $g_\mathfrak{p}$ lie in $U_\mathfrak{p}$, except for a finite number of $\mathfrak{p}$. Put $U=\prod_{i=1}^{r} G_{\mathfrak{p}_i}\times\prod_{\mathfrak{p}\neq\mathfrak{p}_i} U_\mathfrak{p}$. Then U is a locally compact group with respect to the †product topology. The group G can be supplied naturally with a topology with respect to which G is a locally compact group and the quotient space G/U is discrete. The group G together with this topology is called the **restricted direct product** of $\{G_\mathfrak{p}\}$ with respect to $\{U_\mathfrak{p}\}$.

C. Adeles and Ideles

Let k be an †algebraic number field of finite degree and I be the totality of finite and infinite †prime divisors of k. For each $\mathfrak{p}\in I$ we denote by $k_\mathfrak{p}$ and $k_\mathfrak{p}^\times$ the †completion of k with respect to $\mathfrak{p}$ and the multiplicative group of nonzero elements of $k_\mathfrak{p}$, respectively. Furthermore, for each finite prime divisor $\mathfrak{p}$, we denote by $\mathfrak{o}_\mathfrak{p}$ and $\mathfrak{u}_\mathfrak{p}$ the ring of †$\mathfrak{p}$-adic integers

of k and the multiplicative group of †units of $\mathfrak{o}_\mathfrak{p}$, respectively.

(1) Since $\mathfrak{o}_\mathfrak{p}$ is a compact open subgroup of $k_\mathfrak{p}$ as an additive group, we can construct the restricted direct product $\mathbf{A}_k$ of $\{k_\mathfrak{p}\}$ with respect to $\{\mathfrak{o}_\mathfrak{p}\}$. Then $\mathbf{A}_k$ is a locally compact ring with respect to the componentwise ring operations. We call $\mathbf{A}_k$ the **adele ring** (or **ring of valuation vectors**) of k, and an element of $\mathbf{A}_k$ an **adele** (or **valuation vector**) of k. The element of the direct product $\prod_\mathfrak{p} k_\mathfrak{p}$ whose $\mathfrak{p}$-component is a fixed element of k for all $\mathfrak{p}$ is an adele. We call such an adele a **principal adele**. Since $\mathfrak{u}_\mathfrak{p}$ is a compact open subgroup of $k_\mathfrak{p}^\times$ for each finite prime $\mathfrak{p}$, we can construct the restricted direct product $\mathbf{J}_k$ of $\{k_\mathfrak{p}^\times\}$ with respect to $\{\mathfrak{u}_\mathfrak{p}\}$. We call $\mathbf{J}_k$ the **idele group** of k and an element of $\mathbf{J}_k$ an **idele** of k. The element of the direct product $\prod k_\mathfrak{p}^\times$ whose $\mathfrak{p}$-component is a fixed element of k for all $\mathfrak{p}$ is an idele. We call such an idele a **principal idele**. Each element $\mathfrak{b}$ of $\mathbf{J}_k$ induces an automorphism $f_\mathfrak{b}$ of the additive group $\mathbf{A}_k$ defined by $f_\mathfrak{b}(\mathfrak{a}) = \mathfrak{b}\cdot\mathfrak{a}$ ($\mathfrak{a}\in\mathbf{A}_k$). Thus $\mathbf{J}_k$ can be regarded as a subgroup of the automorphism group $\mathrm{Aut}(\mathbf{A}_k)$ of the additive group $\mathbf{A}_k$. The topology of $\mathbf{J}_k$ coincides with the relative topology of $\mathbf{J}_k$ as a subgroup of $\mathrm{Aut}(\mathbf{A}_k)$. We note, however, that the topology of $\mathbf{J}_k$ is different from the relative topology of $\mathbf{J}_k$ as a subspace of $\mathbf{A}_k$, and the former is stronger than the latter. Finally, for a †function field in one variable over a †finite field, the adele ring and the idele group can be defined similarly.

(2) Let $\mathfrak{R}$ be a †normal simple algebra over k and $\mathfrak{O}$ be a †maximal order of $\mathfrak{R}$. For each $\mathfrak{p}\in I$ put $\mathfrak{R}_\mathfrak{p} = \mathfrak{R}\otimes_k k_\mathfrak{p}$, and for each finite prime divisor $\mathfrak{p}$ put $\mathfrak{O}_\mathfrak{p} = \mathfrak{o}_\mathfrak{p}\cdot\mathfrak{O}$. Then $\mathfrak{O}_\mathfrak{p}$ is a compact open additive subgroup of $\mathfrak{R}_\mathfrak{p}$. By the adele ring $\mathbf{A}_\mathfrak{R}$ of $\mathfrak{R}$ we mean the restricted direct product of $\{\mathfrak{R}_\mathfrak{p}\}$ with respect to $\{\mathfrak{O}_\mathfrak{p}\}$. Let $\mathfrak{R}_\mathfrak{p}^\times$ and $\mathfrak{U}_\mathfrak{p}$ be the multiplicative group of nonzero divisors of $\mathfrak{R}_\mathfrak{p}$ and the multiplicative group of the units of $\mathfrak{O}_\mathfrak{p}$, respectively. ($\mathfrak{U}_\mathfrak{p}$ can be defined only if $\mathfrak{p}$ is a finite prime divisor.) By the idele group $\mathbf{J}_\mathfrak{R}$ of $\mathfrak{R}$ we mean the restricted direct product of $\{\mathfrak{R}_\mathfrak{p}^\times\}$ with respect to $\{\mathfrak{U}_\mathfrak{p}\}$. The notion of principal adele (or idele) of $\mathfrak{R}$ can be defined similarly, as in (1). The structures, as topological groups, of $\mathbf{A}_\mathfrak{R}$ and $\mathbf{J}_\mathfrak{R}$ do not depend on the choice of a maximal order $\mathfrak{O}$. The adele ring $\mathbf{A}_k$ and the idele group $\mathbf{J}_k$ described in (1) are special cases of $\mathbf{A}_\mathfrak{R}$ and $\mathbf{J}_\mathfrak{R}$, respectively.

(3) Let G be a linear †algebraic group defined over k, and let $G_\mathfrak{p}$ be the set of $k_\mathfrak{p}$-†rational points of the group for each $\mathfrak{p}\in I$. For each finite prime divisor $\mathfrak{p}$, let $U_\mathfrak{p}$ be the set of elements α of $G_\mathfrak{p}$ such that the coordinates of both α and α^{-1} are $\mathfrak{p}$-adic integers. We can then construct the restricted direct product of $\{G_\mathfrak{p}\}$ with respect to $\{U_\mathfrak{p}\}$, which is called the **idele group** (or **adele group**) of G.

In the following section we focus on describing the fundamental properties of adeles and ideles of an algebraic number field k. We shall start, however, by observing more generally those adeles and ideles of a normal simple algebra $\mathfrak{R}$ over k. (For the properties of the adele group of algebraic groups → [7]; 13 Algebraic Groups.)

D. The Structures of the Adele Ring and Idele Group

Let $\mathfrak{R}$ be a normal simple algebra over an algebraic number field of finite degree k. We identify the totality of principal adeles of $\mathfrak{R}$ (principal ideles of $\mathfrak{R}$) with $\mathfrak{R}$ ($\mathfrak{R}^\times$), and denote it by the same letter $\mathfrak{R}$ ($\mathfrak{R}^\times$). Then $\mathfrak{R}$ ($\mathfrak{R}^\times$) is a discrete subgroup of $\mathbf{A}_\mathfrak{R}$ ($\mathbf{J}_\mathfrak{R}$). The quotient group $\mathbf{A}_\mathfrak{R}/\mathfrak{R}$ is compact. Denoting by $|\alpha_\mathfrak{p}|_\mathfrak{p}$ ($\alpha_\mathfrak{p}\in k_\mathfrak{p}$) and $N_\mathfrak{p}(\alpha_\mathfrak{p})$ ($\alpha_\mathfrak{p}\in\mathfrak{R}_\mathfrak{p}$) the †normalized valuation of $k_\mathfrak{p}$ and the †reduced norm from $\mathfrak{R}_\mathfrak{p}$ to $k_\mathfrak{p}$, respectively, we define, for $\mathfrak{a}\in\mathbf{J}_\mathfrak{R}$, a positive number: $V(\mathfrak{a}) = \prod_{\mathfrak{p}\in I}|N_\mathfrak{p}(\alpha_\mathfrak{p})|_\mathfrak{p}$, where $\mathfrak{a} = (\alpha_\mathfrak{p})$. We call $V(\mathfrak{a})$ the **volume** of $\mathfrak{a}$. If $\mathfrak{a}$ is a principal idele, we have $V(\mathfrak{a}) = 1$ by the †product formula on valuations. Denote by $\mathbf{J}_\mathfrak{R}^0$ the set of ideles $\mathfrak{a}$ with $V(\mathfrak{a}) = 1$ and put $\mathbf{C}_\mathfrak{R}^0 = \mathbf{J}_\mathfrak{R}^0/\mathfrak{R}^\times$. Then $\mathbf{C}_\mathfrak{R}^0$ has finite volume with respect to the †Haar measure of $\mathbf{J}_\mathfrak{R}$. Furthermore, $\mathbf{C}_\mathfrak{R}^0$ is compact if and only if $\mathfrak{R}$ is a †division algebra. In particular, $\mathbf{C}_k^0$ is compact. Let $\mathbf{Q}$ be the field of rational numbers. For each rational prime p, we define a character λ_p of the completion $\mathbf{Q}_p$ of $\mathbf{Q}$ with respect to the p-adic topology (by a character of $\mathbf{Q}_p$, we mean a continuous homomorphism from $\mathbf{Q}_p$ to the 1-dimensional torus $\mathbf{R}/\mathbf{Z}$): If $p = p_\infty$ is the infinite prime of $\mathbf{Q}$, then we put $\lambda_{p_\infty}(x) \equiv -x \bmod \mathbf{Z}$ ($x\in\mathbf{Q}_p$). If p is finite, then we let λ_p be the composite of the following three canonical homomorphisms, namely, the one from $\mathbf{Q}_p$ to $\mathbf{Q}_p/\mathbf{Z}_p$, the one from $\mathbf{Q}_p/\mathbf{Z}_p$ to $\mathbf{Q}/\mathbf{Z}$, and the one from $\mathbf{Q}/\mathbf{Z}$ to $\mathbf{R}/\mathbf{Z}$. We define a character $\lambda_\mathfrak{p}$ of $\mathfrak{R}_\mathfrak{p}$ as follows: $\lambda_\mathfrak{p} = \lambda_p \circ \mathrm{tr}(\mathfrak{R}_\mathfrak{p}/\mathbf{Q}_p)$, where p is the rational prime divisible by $\mathfrak{p}$ and $\mathrm{tr}(\mathfrak{R}_\mathfrak{p}/\mathbf{Q}_p)$ denotes the †reduced trace from $\mathfrak{R}_\mathfrak{p}$ to $\mathbf{Q}_p$. For $x, y\in\mathfrak{R}_\mathfrak{p}$, put $(x,y)_\mathfrak{p} = \exp(2\pi i\lambda_\mathfrak{p}(xy))$. Then the additive group $\mathfrak{R}_\mathfrak{p}$ is self-dual relative to $(x,y)_\mathfrak{p}$. Furthermore, if we put $\langle\mathfrak{a},\mathfrak{b}\rangle = \prod_\mathfrak{p}(a_\mathfrak{p}, b_\mathfrak{p})_\mathfrak{p}$ for $\mathfrak{a} = (a_\mathfrak{p})$ and $\mathfrak{b} = (b_\mathfrak{p})\in\mathbf{A}_\mathfrak{R}$ then $\mathbf{A}_\mathfrak{R}$ is self-dual relative to $\langle\mathfrak{a},\mathfrak{b}\rangle$. The †annihilator of the group of principal adeles with respect to $\langle\mathfrak{a},\mathfrak{b}\rangle$ is $\mathfrak{R}$. Hence it follows from †Pontryagin's duality theorem that $\mathbf{A}_\mathfrak{R}/\mathfrak{R}$ is compact. Henceforth let $\mathfrak{R} = k$. We call the quotient group $\mathbf{C}_k = \mathbf{J}_k/k^\times$ (an element of $\mathbf{C}_k$) the **idele class group** of k (an **idele class**). If a character χ of $\mathbf{J}_k$ satisfies the condition $\chi(\alpha) = 1$ for all $\alpha\in k$ (i.e.,

if χ is a character of $\mathbf{C}_k$), we call such a character a **Grössencharakter** (or **Hecke character**). Grössencharakters were introduced by E. Hecke as characters of a certain type of the †ideal group of k (*Math. Z.*, 1 (1918), 5 (1920)), but they are essentially the same as the ones defined above [1]. Let $\mathbf{D}_k$ be the connected component of the identity element of $\mathbf{C}_k$. Then $\mathbf{C}_k/\mathbf{D}_k$ is totally disconnected and compact. Hence a Grössencharakter χ is of finite order if and only if $\chi(\mathbf{D}_k)=1$. We can prove by †class field theory that $\mathbf{C}_k/\mathbf{D}_k$ is canonically isomorphic to the Galois group over k of the maximal Abelian extension of k (→ 59 Class Field Theory). For the structure of $\mathbf{D}_k$, the following fact is known: Let r_1 and r_2 be the number of †real infinite prime divisors and †imaginary infinite prime divisors of k, respectively. Then the dual group of $\mathbf{D}_k$ is isomorphic to $\mathbf{R}\times\mathbf{Q}^{r_1+r_2-1}\times\mathbf{Z}^{r_2}$, where $\mathbf{R}$ is the additive group of real numbers with the usual topology and $\mathbf{Q}$ ($\mathbf{Z}$) is the additive group of rational numbers (rational integers) with the discrete topology. Let F be a function field in one variable over a finite field $\mathbf{F}_0$. The properties of the adele ring and idele group of F are similar to the properties of $\mathbf{A}_k$ and $\mathbf{J}_k$, while the group $\mathbf{C}_F$ has a simpler structure than $\mathbf{C}_k$. To explain the structure of $\mathbf{C}_F$, let $\tilde{F}$ be the maximal Abelian extension of F, $\tilde{G}$ be the Galois group of $\tilde{F}/F$, and G_F be the subgroup of $\tilde{G}$ consisting of the elements σ such that $\sigma(\alpha)=\alpha^{q^n}$ for all $\alpha\in\bar{\mathbf{F}}_0$ (= the †algebraic closure of $\mathbf{F}_0$), where q is the number of elements of the finite field $\mathbf{F}_0$ and n is a given rational integer. Also, let G_F^0 be the subgroup of G_F consisting of the elements inducing the identity map on $\bar{\mathbf{F}}_0$. G_F^0 is a compact group with respect to the †Krull topology. G_F can be naturally supplied with a topology such that the group G_F is a locally compact group and the quotient group G_F/G_F^0 is discrete. Then class field theory implies that G_F is isomorphic to $\mathbf{C}_F$ as a topological group.

The following characterization of the adele ring of a number field or function field in one variable over a finite field is the work of K. Iwasawa (*Ann. Math.*, (2) 57 (1953)). Let $\mathbf{A}$ be a †semisimple commutative and locally compact topological ring with unity 1. Assume that $\mathbf{A}$ is neither discrete nor compact, and moreover that $\mathbf{A}$ contains a discrete subfield $k\ni 1$ and $\mathbf{A}/k$ is compact. Then k is an algebraic number field or a function field in one variable over a finite field, and $\mathbf{A}$ is isomorphic to the adele ring of k as a topological ring.

E. Ideles and Cohomology

Let K be a Galois extension of finite degree of an algebraic number field k, and $\mathfrak{G}$ be the Galois group of the extension K/k. $\mathfrak{G}$ operates naturally on the idele group $\mathbf{J}_K$ and the idele class group $\mathbf{C}_K$ of K. The structures of the †cohomology groups of $\mathfrak{G}$ with the coefficient groups $\mathbf{J}_K$ and $\mathbf{C}_K$ were investigated by G. Hochschild, T. Nakayama, E. Artin, J. Tate, and others. In particular, we have $H^1(\mathfrak{G},\mathbf{C}_K)=\{0\}$ and $H^2(\mathfrak{G},\mathbf{C}_K)\cong\mathbf{Z}/n\mathbf{Z}$ (cyclic group of order n), where $n=[K:k]$. These facts play an important role in one of the proofs of class field theory (→ [3]; 59 Class Field Theory). Furthermore, A. Weil introduced the so-called **Weil group**, which is a †group extension of a certain type of $\mathbf{C}_K$ by $\mathfrak{G}$. He defined the most general L-functions, which include both †Artin L-functions and †Hecke L-functions with Grössencharakters (→ [2]; 450 Zeta Functions).

F. Fourier Analysis on the Adele Group

†Dedekind zeta functions and Hecke L-functions are †meromorphic on the whole complex plane and satisfy functional equations of certain types. This can be proved by methods of Fourier analysis on the adele group $\mathbf{A}_k$ (Artin, Iwasawa, Tate [1, 8, 9]). For a continuous complex-valued function $\varphi(\mathfrak{a})$ on $\mathbf{A}_k$ satisfying suitable conditions, we define the Fourier transform of $\varphi(\mathfrak{a})$ as follows:

$$\hat{\varphi}(\mathfrak{a})=\int_{A_k}\varphi(\mathfrak{a})\langle\mathfrak{a},\mathfrak{b}\rangle d\mathfrak{b},$$

where $d\mathfrak{b}$ denotes the Haar measure on $\mathbf{A}_k$. By normalizing $d\mathfrak{b}$ suitably and applying †Poisson's summation formula, we get, for each idele $\mathfrak{a}$ of k,

$$\sum_{\alpha\in k}\varphi(\mathfrak{a}\alpha)=V(\mathfrak{a})^{-1}\sum_{\alpha\in k}\hat{\varphi}(\mathfrak{a}^{-1}\alpha).$$

This is called the **Θ-formula**. Consider the following integral on $\mathbf{J}_k$:

$$\xi(s)=\int_{J_k}V(\mathfrak{a})^s\chi(\mathfrak{a})\varphi(\mathfrak{a})d^*\mathfrak{a},$$

where $d^*\mathfrak{a}$ denotes the Haar measure on $\mathbf{J}_k$, s is a complex number, and χ is a Grössencharakter of k, namely, a character of $\mathbf{C}_k$. This integral converges if $s>1$, and by using the Θ-formula one can show that $\xi(s)$ is meromorphic on the whole complex plane and satisfies a functional equation of a certain type. When the function φ is of special type, then the above integral can be explicitly expressed as the product of L-functions, Γ-functions, and exponential functions. This method of expressing L-functions by integrals on $\mathbf{J}_k$ and applying the Θ-formula can be applied to investigate †Hey zeta functions and L-functions of various types defined for a simple algebra (→

450 Zeta Functions) (G. Fujisaki [6]; T. Tamagawa, *Ann. Math.*, 77 (1963)).

References

[1] J. Tate, Fourier analysis in number fields and Hecke's zeta-functions, Thesis, Princeton Univ., 1950. (Algebraic number theory, edited by J. W. S. Cassels and A. Fröhlich, Academic Press, 1967, ch. 15.)
[2] A. Weil, Sur la théorie du corps de classes, J. Math. Soc. Japan, 3 (1951), 1–35.
[3] E. Artin and J. Tate, Class field theory, Lecture notes, Harvard Univ., 1961 (Benjamin, 1967).
[4] E. Artin, Representatives of the connected component of the idele class group, Proc. Intern. Symp. Alg. Number Theory, Tokyo and Nikko, 1955, 51–54.
[5] G. Shimura and Y. Taniyama, Modern number theory (in Japanese), Kyôritu, 1957.
[6] G. Fujisaki, On the zeta-function of the simple algebra over the field of rational numbers, J. Fac. Sci. Univ. Tokyo, 7 (1958), 567–604.
[7] A. Weil, Adeles and algebraic groups, Lecture notes, Institute for Advanced Study, Princeton, 1961.
[8] S. Lang, Algebraic numbers, Addison-Wesley, 1964.
[9] A. Weil, Basic number theory, Springer, 1967.

7 (VI.15) Affine Geometry

A. Construction of Affine Spaces

An affine space A is constructed as follows: Let V be a †vector space over a †field K, and let A be a nonempty set. For any vector $\mathbf{a} \in V$ and any element p of A, suppose that an addition $p + \mathbf{a} \in A$ is defined so as to satisfy the following three conditions: (i) $p+\mathbf{0}=p$ ($\mathbf{0}$ being a zero vector); (ii) $(p+\mathbf{a})+\mathbf{b}=p+(\mathbf{a}+\mathbf{b})$ ($\mathbf{a}, \mathbf{b} \in V$); and (iii) for any $q \in A$ there exists a unique vector $\mathbf{a} \in V$ such that $q=p+\mathbf{a}$. (Condition (i) follows from (ii) and (iii).) Then we call A an **affine space**, V the **standard vector space** of A, and K the **coefficient field** of A. Each element of A is called a **point**.

If we fix an arbitrary point $o \in A$, there is a one-to-one correspondence between A and V given by the mapping sending $p \in A$ to $\mathbf{a} \in V$ such that $p=o+\mathbf{a}$. Such an element $\mathbf{a}$ of V is called a **position vector** of p with the **initial point** o and is denoted by $\overrightarrow{op}$. We say that $r+1$ points p_α $(0 \leqslant \alpha \leqslant r)$ of A are **independent** if r vectors $\mathbf{a}_i=\overrightarrow{p_0p_i}$ $(1 \leqslant i \leqslant r)$ are linearly independent in V; otherwise, they are said to be **dependent**. This definition of dependence of points p_α is independent of the choice of the initial point among them. If V is of dimension n, we say that A is of **dimension** n, $\dim A=n$; in this case, we sometimes write A^n instead of A and V^n instead of V. The affine space A is of dimension n if and only if the maximum number of independent points in A is $n+1$.

Next, for any vector subspace V^k of V^n and an arbitrary point $p \in A^n$, we put $A^k_p=\{q \in A^n \mid q=p+\mathbf{x}, \mathbf{x} \in V^k\}$ and call it a **subspace** of A^n. It is an affine space of dimension k. Conversely, every subset of A^n that is an affine space under the affine space structure of A can be expressed in this form. A^1, A^2, and A^{n-1} in A^n are called a **line**, **plane**, and **hyperplane**, respectively. A set that consists of only one point is also considered as a subspace A^0. For subspaces A^r and A^s of A^n, we denote by $A^r \cap A^s$ the **intersection** (i.e., the set-theoretic intersection) of A^r and A^s, and by $A^r \cup A^s$ the join of A^r and A^s (i.e., the intersection of all subspaces that contain both A^r and A^s). Then $A^r \cap A^s$ is the affine space of highest dimension contained in A^r and A^s, and $A^r \cup A^s$ is the affine space of lowest dimension that contains A^r and A^s. If $r+1$ points are given in A^n, there always exists a subspace A^r that contains all these points. In particular, if the points are independent, then such an A^r is unique. Moreover, if $A^r \cap A^s \neq \varnothing$ ($\varnothing$ is the empty set), then we have $r+s=\dim(A^r \cup A^s)+\dim(A^r \cap A^s)$. This is called the **dimension theorem** (or **intersection theorem**) of affine geometry.

Next suppose that $r+1$ points p_α $(0 \leqslant \alpha \leqslant r)$ in A^n are independent, and put $p_{r+1}=p_0$. Let q_α be an arbitrary point on $p_\alpha \cup p_{\alpha+1}$ that differs from p_α and $p_{\alpha+1}$. If λ^α are elements of K such that $\lambda^\alpha \cdot \overrightarrow{p_\alpha q_\alpha}=\overrightarrow{q_\alpha p_{\alpha+1}}$, then $q_0, \ldots, q_r$ are dependent if and only if $\lambda^0\lambda^1 \ldots \lambda^r=(-1)^{r+1}$. And if $r \geqslant 2$ and $\sigma_\alpha=q_\alpha \cup p_{\alpha+2} \cup \ldots \cup p_{\alpha+r}$ $(p_{i+r}=p_{i-1}$ if $i \geqslant 1)$, then $\sigma_0, \ldots, \sigma_r$ have a point in common if and only if $\lambda^0\lambda^1 \ldots \lambda^r=1$. The former is called **Menelaus's theorem**, and the latter **Ceva's theorem**.

The set $L(A)$ of all subspaces (including $\varnothing$ considered as an affine space of dimension -1) constitutes a †lattice with respect to the inclusion relation.

B. Parallelism in Affine Spaces

Let A^r and A^s be subspaces of A^n. We say that A^r and A^s are **parallel in the wider sense** if either of the following conditions holds: (i) $A^r \supset A^s$ or $A^s \supset A^r$; or (ii) $A^r \cap A^s=\varnothing$ and $\dim(A^r \cup A^s) \leqslant r+s$. Next, let A^r and

B^r be subspaces of A^n of the same dimension. If A^r and B^r coincide, or $A^r \cap B^r = \emptyset$ and $\dim(A^r \cup B^r) = r+1$, then they are said to be **parallel in the narrower sense** (or simply **parallel**), and we denote the relation by $A^r // B^r$. If $r=s=1$, the definitions of parallelism in the narrower sense and wider sense are equivalent. If $r>1$, parallelism in the narrower sense implies parallelism in the wider sense. For two sets (a_α) and (b_α) $(0 \leqslant \alpha \leqslant r)$ of $r+1$ independent points, let V^r and W^r be vector spaces with bases $\overrightarrow{a_0 a_i}$ and $\overrightarrow{b_0 b_i}$ $(1 \leqslant i \leqslant r)$, respectively. Then $A^r = a_0 \cup \ldots \cup a_r$ and $B^r = b_0 \cup \ldots \cup b_r$ are parallel if and only if $V^r = W^r$; and for an arbitrary point p, there exists a unique r-dimensional subspace that is parallel to A^r and passes through point p. If A^r and B^s are parallel in the wider sense, then there exist subspaces A^t and B^t $(t \geqslant 1)$ of A^r and B^s that are parallel to each other. Moreover, if neither A^r nor B^s is contained in the other and if t is the largest integer with the property just given, then we have $t = r+s+1-\dim(A^r \cup B^s)$.

Parallelism between subspaces of A^n is an equivalence relation. Specifically, the equivalence class of a 1-dimensional subspace A^1 is called a **point at infinity** and is denoted by A^0_∞. Given a subspace A^r of A^n, the set of points at infinity A^0_∞ represented by lines A^1 contained in A^r is denoted by A^{r-1}_∞; we have $A^r // B^r$ if and only if $A^{r-1}_\infty = B^{r-1}_\infty$. A^{r-1}_∞ is called a **space at infinity**. In particular, the set A^{n-1}_∞ is called the **hyperplane at infinity**. The set-theoretic sum $A^n \cup A^{n-1}_\infty = \bar{A}^n$ is supplied with the structure of a †projective space; the "points" in $\bar{A}^n$ are elements of $\bar{A}^n$, and the "lines" in $\bar{A}^n$ are $A^1 \cup A^0_\infty$ and A^1_∞.

C. Coordinates of Affine Spaces

If we fix a point o in A^n and a basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ of the standard vector space V^n, then any point p in A^n is uniquely expressed as

$$p = o + \sum_{i=1}^{n} x^i \cdot \mathbf{e}_i, \qquad x^i \in K. \tag{1}$$

The system $\mathfrak{F} = (o; \mathbf{e}_1, \ldots, \mathbf{e}_n)$ is called an **affine frame** (simply the **frame**) of A^n; the point o is called its **origin**, and $\mathbf{e}_i$ is called the ith **unit vector**. The mapping sending p to $(x^1, \ldots, x^n)$ gives a †bijection of A^n to K^n; we call $(x^1, \ldots, x^n)$ **affine coordinates** of p with respect to $\mathfrak{F}$, and x^i the ith **affine coordinate**. In particular, if K is a topological field (e.g., the real number field $\mathbf{R}$ or the complex number field $\mathbf{C}$), this bijection $A^n \to K^n$ induces a topology of A^n, which can be shown to be independent of the choice of $\mathfrak{F}$. For the remainder of this article, by "coordinates" we mean affine coordinates unless otherwise stated. Putting $a_i = o + \mathbf{e}_i$ $(1 \leqslant i \leqslant n)$, we sometimes call $(o; a_1, \ldots, a_n)$ an affine frame. Further, putting $l_i = o \cup a_i$, $\pi_i = o \cup a_1 \cup \ldots \cup a_{i-1} \cup a_{i+1} \cup \ldots \cup a_n$, we call a_i, l_i, and π_i the ith **unit point**, the ith **coordinate axis**, and the ith **coordinate hyperplane**, respectively.

Assume that subspaces A^r and A^s $(r, s > 0, r+s=n)$ are not parallel in the wider sense. For a point p of A^n, denote by $A^r(p)$ the subspace that passes through p and is parallel to A^r, and put $q = A^r(p) \cap A^s$. A mapping $\varphi: A^n \to A^s$ defined by $\varphi(p) = q$ is called a **parallel projection** on A^s with respect to A^r. In particular, if $A^r = \pi_i$ and $A^s = l_i$ $(r = n-1, s = 1)$, we write $\varphi(p) = p_i$. Then the ith coordinate x^i of p is an element of K such that $\overrightarrow{op_i} = x^i \overrightarrow{oa_i}$. Hence such coordinates are also called **parallel coordinates** (or **Cartesian coordinates**).

Suppose that we are given $r+1$ points $b_0, \ldots, b_r$ of A^n and $r+1$ elements $\lambda^0, \ldots, \lambda^r$ of K such that $\sum_{\alpha=0}^{r} \lambda^\alpha = 1$. We fix a point o of A^n. If a point p in A^n satisfies $\overrightarrow{op} = \sum_{\alpha=0}^{r} \lambda^\alpha \overrightarrow{ob_\alpha}$, then $\overrightarrow{b_0 p} = \sum_{i=1}^{r} \lambda^i \overrightarrow{b_0 b_i}$; hence p is contained in the subspace $b_0 \cup \ldots \cup b_r$. Conversely, if a point p is contained in the latter subspace, then there exists a system $(\lambda^0, \ldots, \lambda^r)$ such that

$$\overrightarrow{op} = \textstyle\sum_{\alpha=0}^{r} \lambda^\alpha \overrightarrow{ob_\alpha} \quad \text{and} \quad \sum_{\alpha=0}^{r} \lambda^\alpha = 1.$$

The system $(\lambda^0, \ldots, \lambda^r)$ has a geometric meaning since we also have $\overrightarrow{o'p} = \sum_{\alpha=0}^{r} \lambda^\alpha \overrightarrow{o'b_\alpha}$ if we replace the point o by any other point o' of A^n. The elements $\lambda^0, \ldots, \lambda^r$ are called **barycentric coordinates** of p with respect to $\{b_0, \ldots, b_r\}$. In particular, if $\{b_0, \ldots, b_r\}$ are independent, then the barycentric coordinates $(\lambda^0, \ldots, \lambda^r)$ are uniquely determined by the point p on $b_0 \cup \ldots \cup b_r$. Furthermore, let $(y^1, \ldots, y^n)$ be affine coordinates of p with respect to an affine frame $\mathfrak{F}$, and let $(x^1_\alpha, \ldots, x^n_\alpha)$ be affine coordinates of b_α $(\alpha = 0, \ldots, r)$. Then p belongs to the subspace $A^r = b_0 \cup \ldots \cup b_r$ if and only if $y^i = \sum_{\alpha=0}^{r} \lambda^\alpha x^i_\alpha$ $(i = 1, \ldots, n)$. In this case we say that the system of the linear equations $y^i = \sum \lambda^\alpha x^i_\alpha$ $(\sum \lambda^\alpha = 1)$ gives a **parametric representation** of the subspace A^r (by parameters λ^α). Specifically, if $r = n-1$, the solvability of the system of equations $y^i = \sum_{\alpha=0}^{n-1} \lambda^\alpha x^i_\alpha$ $(i = 1, \ldots, n)$, $1 = \sum \lambda^\alpha$ implies the equation $\sum_{i=1}^{n} y^i \mu_i = \mu_0$ for some nontrivial constants $\mu_0, \ldots, \mu_n$. Hence the latter equation represents the hyperplane $\pi = \mathrm{A}^{n-1}$. If a point p has barycentric coordinates $\lambda^0 = \lambda^1 = \ldots = \lambda^r = (r+1)^{-1}$ with respect to $\{b_0, \ldots, b_r\}$, it is called the **barycenter** of $b_0, b_1, \ldots, b_r$. The barycenter is uniquely determined by the set $\{b_0, \ldots, b_r\}$ and is denoted by $g(b_0, \ldots, b_r)$. Specifically, the barycenter of two points b_0 and b_1 is called the **midpoint** (or **middle point**) of b_0 and b_1. If we divide $B = \{b_0, \ldots, b_r\}$ into two sets of points and if g_1 and g_2 are barycenters of these two sets of points, respectively, then $g_1 \cup g_2$ passes

through the barycenter of B. More generally, a point with barycentric coordinates $(\lambda^0, \ldots, \lambda^r)$ with respect to $\{b_0, \ldots, b_r\}$ is called a barycenter of $b_0, \ldots, b_r$ with **weights** $\lambda^0, \ldots, \lambda^r$.

D. Affine Spaces over Ordered Fields

Suppose the coefficient field K is an †ordered field (e.g., the real number field $\mathbf{R}$). Given a hyperplane π of A^n, we take an affine frame for which π is the nth coordinate hyperplane. If we denote coordinates of points with respect to this frame by $(x^1, \ldots, x^n)$, then the equation of π is given by $x^n = 0$. Let A^n_+ and A^n_- be sets of points whose nth coordinates are positive and negative, respectively. They are called **half-spaces** of A^n divided by π. The union of π and a half-space is called a **closed half-space**. A half-space of a subspace A^r of A^n (divided by some A^{r-1} on A^r) is called a half-space of dimension r. For a point p of A^n that does not lie on π, the half-space containing p is called the **side** of p with respect to π. In particular, when $n=1$, let p and q be two points on a line l. The closed side of q with respect to p is called the **(closed) half-line** (or **ray**) from p to q. The intersection of the closed half-lines emanating from p to q and from q to p is called the **segment** joining p and q and is denoted by $\overline{pq}$. Clearly $\overline{pq} = \overline{qp}$. A subset C of A^n is called a **convex set** if the segment joining two arbitrary points of C is contained in C. Each half-space of each dimension is convex. For any family C_τ of convex sets, $\bigcap_\tau C_\tau$ is also convex. Therefore, for any subset D in A^n there exists a minimal convex set that contains D. It is called the **convex closure** (or **convex hull**) of D. The convex closure $C(P)$ of a finite set of points $P = \{p_0, \ldots, p_k\}$ in A^n is called a **convex cell**, and $\dim(p_0 \cup \ldots \cup p_k)$ is called the **dimension** of the convex cell. In particular, when $p_0, \ldots, p_k$ are independent, $C(P)$ is called a k-dimensional **simplex** with **vertices** $p_0, \ldots, p_k$. The 1-dimensional simplex having two distinct points p and q as vertices is the segment $\overline{pq}$, and the vertices p and q are called **ends** of the segment. A point is regarded as a 0-dimensional simplex. Each 2-dimensional or 3-dimensional simplex is called **a triangle** or **tetrahedron**, respectively. A k-dimensional simplex S with vertices $p_0, \ldots, p_k$ is a set of points whose barycentric coordinates λ^α $(0 \leqslant \alpha \leqslant k, \sum \lambda^\alpha = 1)$ with respect to the vertices satisfy $\lambda^\alpha \geqslant 0$. On the other hand, if we put $A^k = p_0 \cup \ldots \cup p_k$ and $\pi_\alpha = p_0 \cup \ldots \cup p_{\alpha-1} \cup p_{\alpha+1} \cup \ldots \cup p_k$, and denote by A^k_α the side of p_α in A^k and by $\bar{A}^k_\alpha$ the closed side of p_α in A^k with respect to π_α, then the simplex S is given by $\bigcap_{\alpha=0}^k \bar{A}^k_\alpha$, and $\bigcap_{\alpha=0}^k A^k_\alpha$ is called an **open simplex**. A^n has the structure of a †topological space in which the set of open n-dimensional simplexes forms a †base of †open sets. In particular, if K is $\mathbf{R}$, the topology of A^n thus defined is compatible with the one that is naturally induced by the topology of $\mathbf{R}$. With respect to this topology, A^n is a †Hausdorff space. The terms *open* and *closed* used before for n-dimensional simplexes agree with the corresponding notions with respect to this topology.

A subset of A^n is said to be **bounded** if it is contained in some simplex. A bounded set obtained through a finite process of constructing intersections and unions from a finite number of closed half-spaces is called a **polyhedron**. The points of a convex polyhedron are characterized by several linear inequalities satisfied by their coordinates. A set of points whose coordinates $(x^1, \ldots, x^n)$ satisfy $h^i \leqslant x^i \leqslant k^i$ for $k^i, h^i \in K$ is called a **parallelotope**; it is a polyhedron whose †interior is called an **open parallelotope**. A simplex is a polyhedron, and polyhedra admit †simplicial decompositions. A polyhedron can also be defined as the set-theoretic union of a finite number of simplexes.

Let P be a finite set of points, and let its convex closure $C(P)$ be a convex cell of dimension m. Then $C(P)$ is contained in an m-dimensional subspace A^m. The †boundary of $C(P)$ in the topological space A^m is called the **boundary** of a convex cell $C(P)$. We can take a subset Q of P so that $\dim C(Q) = m-1$ and $C(Q)$ is the intersection of the boundary of $C(P)$ and an $(m-1)$-dimensional subspace. Such a $C(Q)$ is called a **face** of $C(P)$, and we denote this relation by $C(P) \succ C(Q)$. If $C(P) \succ C(P_1) \succ \ldots \succ C(P_s)$, then $C(P_s)$ is called an $(m-s)$-dimensional face of $C(P)$. A 0-dimensional face is called a **vertex**, and a 1-dimensional face is called an **edge**. Suppose that $C(P) \succ C(Q)$ for $P = \{p_0, \ldots, p_k\}$ and $Q = \{p_{i_0}, \ldots, p_{i_{k-1}}\}$. Then $F = p_{i_0} \cup \ldots \cup p_{i_{k-1}}$ is a hyperplane of $E = p_0 \cup \ldots \cup p_k$, and $C(P)$ is contained in a closed side of E divided by F. Therefore, if $C(P)$ has d $(m-1)$-dimensional faces, then $C(P)$ is expressed as the intersection of d m-dimensional closed half-spaces. This shows that any convex cell is a polyhedron.

E. Affine Transformations

A mapping $\varphi: A^n \to A^m$ is an **affine mapping** if there is a linear mapping $\bar{\varphi}: V^n \to V^m$ of the standard vector spaces of A^n and A^m such that $\varphi(p + \mathbf{x}) = \varphi(p) + \bar{\varphi}(\mathbf{x})$ holds for any $p \in A^n$ and any $\mathbf{x} \in V^n$. An affine mapping of A^n into itself is called an **affine transformation** (or **affinity**) of A^n. Specifically, a bijective affine transformation is called a **regular** (or **proper**) **affine transformation**. An affine transformation φ of A^n is characterized by each one of the following

properties: (i) Let $o \in A^n$ be a fixed point. Then φ is a mapping of A^n onto itself that can be expressed as

$$\varphi(o+\mathbf{x}) = o+\mathbf{a}+f(\mathbf{x}), \qquad (2)$$

where $\mathbf{a}$ is a fixed vector of V^n and f is a linear transformation of V^n. (ii) The mapping $\varphi: A^n \rightarrow A^n$ is a mapping such that $\overrightarrow{\varphi(a)\varphi(b)} = \lambda \cdot \overrightarrow{\varphi(p)\varphi(q)}$ if $\overrightarrow{ab} = \lambda \cdot \overrightarrow{pq}$ $(\lambda \in K)$. Moreover, if the †characteristic of K is not equal to 2, an affine transformation is also characterized as follows: (iii) φ is a mapping that sends lines into lines and preserves the ratio of each pair of parallel segments.

The set $\mathfrak{A}(A^n)$ of all regular affine transformations of A^n constitutes a group that we call the **group of affine transformations.** If the linear mapping f associated with a regular affine transformation φ is the identity mapping, then φ is called a **translation.** The set $\mathfrak{T}(A^n)$ of all translations is a normal subgroup of $\mathfrak{A}(A^n)$ and is called the **group of translations.** The group of translations is isomorphic to V regarded as an additive group. The vector group $\mathfrak{V}(A^n)$ (i.e., an additive group of a linear space) acts †simply transitively on A^n. We see that $\mathfrak{A}(A^n)/\mathfrak{V}(A^n) \cong GL(n, K)$, where $GL(n, K)$ denotes the †general linear group. The set of all regular affine transformations that leave a point o of A^n invariant constitutes a subgroup $\mathfrak{G}(A^n)$ of $\mathfrak{A}(A^n)$; it is called an †isotropy group at o and is isomorphic to $GL(n, K)$. Let $\mathfrak{F} = (o; \mathbf{e}_1, \ldots, \mathbf{e}_n)$ be an affine frame of A^n with origin o, and let φ be a regular affine transformation of A^n given by (2); put $\mathbf{x} = \sum x^i \mathbf{e}_i$, $\varphi(o+\mathbf{x}) = o + \sum \bar{x}^i \mathbf{e}_i$, $\mathbf{a} = \sum a^i \mathbf{e}_i$, and $f(\mathbf{e}_i) = \sum a_i^j \mathbf{e}_j$. Then φ is expressed with respect to $\mathfrak{F}$ by the following equation:

$$\bar{x}^i = a^i + \sum_{k=1}^{n} a_k^i x^k, \quad \det(a_j^i) \neq 0, \quad 1 \leqslant i \leqslant n. \qquad (3)$$

Conversely, a transformation that is given by (3) is a regular affine transformation. Elements of $\mathfrak{V}(A^n)$ and $\mathfrak{G}(A^n)$ are expressed with respect to $\mathfrak{F}$ by

$$\bar{x}^i = x^i + a^i, \qquad 1 \leqslant i \leqslant n, \qquad (4)$$

and

$$\bar{x}^i = \sum_{k=1}^{n} a_k^i x^k, \qquad \det(a_j^i) \neq 0, \qquad 1 \leqslant i \leqslant n, \qquad (5)$$

respectively. Hence $\mathfrak{A}(A^n)$ is represented as a †semidirect product group of $\mathfrak{V}(A^n)$ and $\mathfrak{G}(A^n)$. In particular, a regular affine transformation that is represented by $\bar{x}^i = ax^i$ $(1 \leqslant i \leqslant n)$ for some $a \in K$ $(a \neq 0)$ is called a **similarity** with the origin 0 as center.

According to F. Klein, the objects we deal with in affine geometry are the properties (parallelism, barycenters, etc.) that are invariant under regular affine transformations.

Subsets S_1 and S_2 of A^n are called **affinely congruent** if there exists a regular affine transformation φ sending S_1 onto S_2. For a fixed k, two k-dimensional simplexes are affinely congruent. Now we fix an affine frame $\mathfrak{F}$ in A^n and denote by x_α^i the coordinates of $n+1$ points p_α $(0 \leqslant \alpha \leqslant n)$ in A^n. Then the quantity

$$V(p_0, \ldots, p_n) = \frac{1}{n!} \begin{vmatrix} 1 & 1 & \ldots & 1 \\ x_0^1 & x_1^1 & \ldots & x_n^1 \\ \ldots & \ldots & \ldots & \ldots \\ x_0^n & x_1^n & \ldots & x_n^n \end{vmatrix} \qquad (6)$$

is called the **volume** with respect to $\mathfrak{F}$ of the n-dimensional simplex with vertices $p_0, \ldots, p_n$. If φ is a regular affine transformation given by (3), we have $\bar{V} = \det(a_j^i) V$. Hence the ratio of volumes of two n-dimensional simplexes is independent of the choice of coordinate systems, and is invariant under regular affine transformations.

A regular affine transformation given by (3) satisfying $\det(a_j^i) = 1$ is called an **equivalent affnity.** The set of all equivalent affinities constitutes a subgroup of $\mathfrak{A}(A^n)$. The geometry belonging to this group is called **affine geometry in the narrower sense.** For instance, the concept of volume is an invariant in affine geometry in the narrower sense.

F. Relation to Projective Geometry

Let $\mathbf{P}^n$ be a projective space over a coefficient field K (→ 343 Projective Geometry). If we fix a hyperplane π_∞ in $\mathbf{P}^n$, then the set of projective transformations that leave π_∞ invariant constitutes a subgroup of the group of †projective transformations of $\mathbf{P}^n$; this subgroup is isomorphic to a group of regular affine transformations. Actually, if we use a †projective frame $[a_0, a_1, \ldots, a_n, u]$ such that $a_1, \ldots, a_n$ are points on π_∞, then each projective transformation leaving π_∞ invariant is expressed by equations of the same form as (3) with respect to the †inhomogeneous projective coordinates. The point set A^n complementary to π_∞ in $\mathbf{P}^n$ is an affine space, and π_∞ coincides with the hyperplane at infinity. Moreover, two distinct lines in $\mathbf{P}^n$ are parallel in A^n if they meet on the hyperplane at infinity. Hence, denoting by $(0, l^1, \ldots, l^n)$ the †homogeneous projective coordinates of the intersection of a line l in A^n and π_∞, we call $(l^1, \ldots, l^n)$ the **direction ratio** of l. A projective transformation leaving each point of π_∞ invariant induces a translation. The †principle of duality that holds in projective geometry does not hold in affine geometry. The †pole of the hyperplane at infinity, with respect to a quadric hypersurface, is called the **center** of the quadric hypersurface. A regular quadric hypersurface is called **central**

or **noncentral** according as its center belongs to A^n or is a point at infinity. Quadric hypersurfaces in an affine space are classified in several ways, by taking account of their relations with the hyperplane at infinity (→ 78 Conic Sections, 350 Quadric Surfaces).

References

[1] O. Schreier and E. Sperner, Einführung in die analytische Geometrie und Algebra I, II, Teubner, 1931, 1935; English translation, Introduction to modern algebra and matrix theory, Chelsea, second edition, 1961.
[2] E. Sperner, Einführung in die analytische Geometrie und Algebra I, II, Vandenhoeck & Ruprecht, 1948, 1951.
[3] E. Artin, Geometric algebra, Interscience, 1957.
[4] H. Weyl, Raum, Zeit, Materie, Springer, fifth edition, 1923; English translation, Space, time, matter, Dover, 1952.
[5] S. Iyanaga and K. Matsuzaka, Affine geometry and projective geometry, J. Fac. Sci. Univ. Tokyo, 14 (1967), 171-196.

8 (III.1)
Algebra

Though the word "algebra" usually refers to a field of mathematics as will be explained below, the word may also denote specific mathematical structures such as †associative algebras, †Jordan algebras, †Clifford algebras, etc. The first concepts concerning "unknowns" in algebra originated in India, whence came also our decimal positional system of numeration. These ideas were introduced to Europe through Arabia in the Renaissance period. F. †Viète systematized them into a symbolic method, called **algebra,** representing numbers by letters. The first problem of algebra was solving equations. Before Viète, G. Cardano and L. Ferrari had solved algebraic equations of degrees 3 and 4; the solution of equations of lower degree had been known from antiquity. The effort to solve equations of higher degree remained unresolved until the middle of the 19th Century, when N. H. †Abel and E. †Galois proved the nonexistence of algebraic solutions of such equations. They considered not only individual roots of these equations but also any rational transforms of their roots at the same time, and thus were led to the concept of †fields. They also noticed that the problem of algebraic solution could be characterized by properties of permutation groups of the roots. After the discovery of the Galois group, group theory and group-theoretical considerations maintained the central position in algebra for some time (→ 172 Galois Theory). These developed into the "abstract algebra" of this Century in the general atmosphere of arithmetization and of axiomatization of mathematics. At the turn of the century the monumental textbook in three volumes by H. Weber [1] was considered the standard work on algebra. Then there appeared in 1910 an epoch-making paper [2] by E. Steinitz on the abstract theory of fields.

The main objects of algebra today are †algebraic systems of various kinds, such as †groups, †rings, †fields, and †modules. Another fundamental concept of algebra is that of †isomorphism or of †homomorphism. The collection of algebraic systems of a given kind, together with the homomorphisms among them, gives rise to the notion of †category; a functor is a sort of homomorphism between categories (→ 52 Categories and Functors). These notions were first used in †homological algebra, created in the 1940s by methods transferred from topology to algebra; now they are of basic significance to the whole of mathematics.

An important branch of algebra with wide applications is the theory of †vector spaces, or more generally that of †modules over a ring. This branch is called **linear algebra.** Homomorphisms between finitely generated modules can be represented by †matrices. Another branch of algebra, called †representation theory, is concerned with representations of groups or rings by matrices. The methods of modern algebra provide useful and powerful tools for the whole of mathematics, in particular for the theory of numbers and algebraic geometry.

The present development of algebra owes much to the activity of the German school in the late 1920s represented by E. Noether, E. Artin, W. Krull, and B. L. van der Waerden. The book by van der Waerden [3] has had a great impact on mathematics. N. Bourbaki [4] has been influenced by van der Waerden but gives accounts of more recent developments, particularly in linear algebra. In Japan, M. Sono, who worked at about the same period as E. Noether, was a forerunner in this field; after him, algebraists of the Kyoto School, Y. Akizuki, M. Nagata, and their followers, did notable research, especially in algebraic geometry. On the other hand, K. Shoda studied with E. Noether toward 1930 in Germany; his school includes such algebraists as T. Nakayama, K. Asano, and G. Azumaya. Finally K. Morita and his disciples have made significant contributions to homological algebra.

References

[1] H. M. Weber, Lehrbuch der Algebra, Vieweg, I, 1894; II, 1896; III, 1891.
[2] E. Steinitz, Algebraische Theorie der Körper, J. Reine Angew. Math., 137 (1910), 167–309.
[3] B. L. van der Waerden, Moderne Algebra, Springer, first edition, I, 1930; II, 1931; English translation, Algebra I, II, Ungar, 1970.
[4] N. Bourbaki, Eléments de mathématique II. Algèbre, ch. 1-9, Actualités Sci. Ind., 1144b, 1236b, 1044, 1102b, 1179a, 1261a, 1272a, Hermann, 1958-1964.
[5] G. Birkhoff and S. MacLane, A survey of modern algebra, Macmillan, third edition, 1965.

9 (VIII.2) Algebraic Curves

A. General Remarks

An †algebraic variety of dimension 1 is called an **algebraic curve** (for analytic theory → 11 Algebraic Functions). The theory of algebraic curves has two aspects, the geometry of 1-dimensional complex manifolds in projective spaces and the theory of function fields of transcendence degree 1 (→ 3 Abelian Varieties, 16 Algebraic Varieties). The number-theoretic study of algebraic function fields concerns the latter theory (- 73 Complex Multiplication, 450 Zeta Functions). In this article, the geometric aspect of the theory is emphasized.
We denote the †universal domain by **K.**

B. Classical Results on Plane Algebraic Curves

Let $f(X, Y)$ be a polynomial of degree m in two variables X and Y. A point set in an affine two-space defined by $f(X, Y) = 0$ is called the plane algebraic curve of degree m. If we set $F(X_0, X_1, X_2) = X_0^m f(X_1/X_0, X_2/X_0)$, the homogeneous polynomial F defines an algebraic curve of degree m in a projective plane P^2. The curve is called **irreducible** if the polynomial $f(X, Y)$ is irreducible. A curve of degree 1 is said to be a line. Some results of this section are valid only in the case where the characteristic of **K** is zero.

Let C be a plane curve defined by the equation $f(X, Y) = 0$. A point $P = (a, b)$ on C is called an **r-ple point** if $f(X + a, Y + b)$ has no term of degree $< r$ in X and Y. At an r-ple point there are r tangent straight lines (counting multiplicity). An r-ple point with $r > 1$ is called a **multiple point** (or **singular point).** If these tangents are distinct, P is called ordinary. An ordinary double point is called a node; e.g., the origin for $X^3 + Y^3$ -3X Y = 0. An algebraic curve can be transformed into a plane curve that has only ordinary multiple points by a finite number of plane †Cremona transformations (†quadratic transformations of the projective plane into itself).

Let C be an irreducible plane curve of degree $m > 1$ in a projective plane. The set of all tangent lines at nonsingular points of C determines a set of points in the dual projective plane, and its closure is an algebraic curve $\hat{C}$, that is called the **dual curve** of C. The dual curve of $\hat{C}$ becomes the original C. The degree m' of $\hat{C}$ is said to be the **class** of C, which is equal to the number of tangent lines to C drawn from a general point. A nonsingular point P is called a **point of inflection** (or a flex) if the tangent line at P has contact of order > 2. If C is defined by an irreducible homogeneous polynomial $F(X_0, X_1,$ X,), the curve defined by $\det(\partial^2 F/\partial X_i \partial X_j) = 0$ is said to be the **Hessian** of C. A nonsingular point P is a point of inflection if and only if P is contained in the Hessian. A singular point P is said to be a **cusp** if C is defined by an equation $Y^2 = X^3$ + higher terms, in terms of suitable affine coordinates X and Y where $P = (0, 0)$. Whenever the singular points of C are only v nodes and y cusps, the effective genus g is given by the formula g = $(m - 1)(m - 2)/2 -$ v -y. In addition, suppose that the dual curve has only v' nodes and y' cusps as singularities. Then C has y' points of inflection, and m' and y' are given by $m' = m(m-1) - 2\nu - 3\gamma$ and $\gamma' = 3m(m-2) - 6\nu - 8\gamma$. Moreover, $m = m'(m'-1) - 2\nu' - 3\gamma'$ and y = $3m'(m'-2)$ $6\nu'$ $8\gamma'$ hold. These formulas are called **Plücker formulas.**

For example, a nonsingular plane curve of degree 3 is an elliptic curve, i.e., g = 1, of class 6, which has 9 points of inflection. A nonsingular plane curve of degree 4 has, in general, 24 points of inflection and 28 bitangents. This results from the Plücker formulas.

C. Fundamental Notions

In what follows, by a curve we mean an algebraic variety of dimension 1. Let Γ be a nonsingular complete irreducible curve. An element of the free Abelian group generated by points of Γ is called a **divisor.** A divisor is written in the form a = $\sum n_i P_i$, with $n_i \in$ Z. The integer $n = \sum n_i$ is called the **degree** of a and is denoted by deg a. The expression for a divisor a is said to be reduced if $P_i \neq P_j$ for $i \neq j$. A divisor whose reduced expression has only positive coefficients is called a **positive divisor** (or **effec-**

tive divisor, or **integral divisor),** and this is denoted by $\mathfrak{a}\succ 0$. The group of divisors on Γ is denoted by $G(\Gamma)$, and the subgroup consisting of divisors of degree 0 is denoted by $G_0(\Gamma)$. Let P be a point of Γ. The subset of the function field $\mathbf{K}(\Gamma)$ of Γ consisting of functions regular at P forms a valuation ring R_P for a †discrete valuation of K(T). A prime element t of R_P is called a **local parameter** at P. Let v_P be the †normalized valuation of $\mathbf{K}(\Gamma)$ defined by R_P; the integer $v_P(f)$ is called the **order** off at P. The point P is a **zero** off if $v_P(f)>0$; it is a **pole** off if $v_P(f)<0$. There are only a finite number of poles and zeros of a given function f. The divisor $\sum v_P(f)P$ is called the **divisor of the function** f and is denoted by (f). The set of divisors of functions forms a subgroup G_l of G,. Any divisor a in G_l is called a **principal divisor** (we also say that a is †linearly equivalent to zero and write a $\sim$ 0).

Let a be an arbitrary divisor. The set of all positive divisors that are linearly equivalent to a forms a **complete linear system** $|\mathfrak{a}|$ determined by a. We set L(a) = $\{f\in$ **K(r)** **(f)** + $\mathfrak{a}\succ 0\}\cup\{0\}$. Then L(a) is a finite-dimensional vector space over **K,** and 1-dimensional subspaces of L(a) correspond bijectively to the elements of $\mathfrak{a}|$. We set J(a) = dim, L(a) and $\dim|\mathfrak{a}|=l(\mathfrak{a})-1$. Then $\dim|\mathfrak{a}|$ is called the **dimension** of $\mathfrak{a}|$. For any divisor a, the integer deg a $-$ dim| a is nonnegative and bounded. The supremum g of such integers is called the **genus** of Γ. The nonnegative integer i(a) = g $-$ deg a + dim a is called the **speciality index** of a.

Let ω be a †differential form on Γ, P be a point of Γ, and t be a local parameter at P. Then ω can be written in the form $\omega=fdt$. We now set $v_P(\omega)=v_P(f)$ and (w) = $\sum v_P(\omega)P$. Then (w) is a well-defined divisor, and the class of (w) in G/G_l is independent of the choice of ω. This divisor class is called the **canonical class**; any divisor in this class is called a **canonical divisor** (or **differential divisor)** and is denoted by $\mathfrak{k}$. We have J(f) = y, deg $\mathfrak{k}=2g-2$. Given a divisor a, the index i(a) is equal to the number of linearly independent differentials ω such that $(\omega)\succ\mathfrak{a}$, $i(\mathfrak{a})=l(\mathfrak{k}-\mathfrak{a})$. The equality J(a)= deg a $-g+1+$ i(a) is called the **Riemann-Roch theorem.**

For any irreducible algebraic curve Γ, there exists a birationally equivalent nonsingular curve $\tilde{\Gamma}$ that is unique up to isomorphism. The genus of $\tilde{\Gamma}$ is called the **effective genus** of Γ. A curve whose effective genus is zero is called a **rational (or unicursal)** curve. An **elliptic curve** is a curve whose effective genus is 1. An irreducible curve Γ with genus $\geqslant$ **2** is called a **hyperelliptic curve** if $\mathbf{K}(\Gamma)$ is a quadratic extension of a field **K(t)** for some t.

A positive divisor a on a nonsingular complete irreducible curve Γ is called a **special divisor** if i(a) > 0. In this case, 2 (J(a) $-$ 1) $\leqslant$ deg a, where equality holds if and only if a = 0 or a $\sim$ $\mathfrak{k}$, or Γ is a hyperelliptic curve (Clifford's theorem). Let Γ be a nonsingular complete irreducible curve and k be a subfield of **K** such that Γ is defined over k. Denoting by k the algebraic closure of k, we call a divisor p = $\sum n_iP_i$ **on** Γ a **prime rational divisor over** k if p satisfies the following three conditions: (i) p is invariant under any automorphism σ of $\bar{k}/k$; (ii) for any j, there exists an automorphism σ_j of $\bar{k}/k$ such that $P_j=P_1^{\sigma_j}$; (iii) $n_1=\ldots=n_t=[k(P_1):k]_i$. An element in the subgroup of $G(\Gamma)$ generated by prime rational k-divisors is called a **k-rational divisor.** Let $k(\Gamma)$ be the subset of $\mathbf{K}(\Gamma)$ consisting of functions f defined over k. Then $k(\Gamma)$ is a subfield of $\mathbf{K}(\Gamma)$, and the quotient field of $k(\Gamma)\otimes_k$**K = K(T).** $k(\Gamma)$ is called the **function field of** Γ **over** k. Let p be a prime rational k-divisor, and let P be a point of $\mathfrak{p}$. Then $R_P\cap k(\Gamma)$ is a valuation ring of $k(\Gamma)$ uniquely determined by p and independent of the choice of the point P in p. We call this valuation ring the valuation ring determined by $\mathfrak{p}$.

D. Algehraic Function Fields

Let k be a field, and let K be a finite separable extension of a purely transcendental extension $k(x)$ of k such that k is maximally algebraic in K. Then K is called an **algehraic function field over** k **of dimension** 1 (or **of transcendence degree** 1). The equivalence class of †exponential valuations of K that are trivial over k is called a **prime divisor** of K/k. An element of the free Abelian group generated by prime divisors is called a **divisor** of K/k. The group operation in the divisor group of K/k is usually denoted multiplicatively. Let R_P be the valuation ring of the prime divisor P, and let M_P be the maximal ideal of R,. The **degree** deg P of the prime divisor P is defined by $[(R_P/M_P):k]$. If we replace the terms: curve Γ by function field K/k; $\mathbf{K}(\Gamma)$ by K; **K** by k; and points on Γ by prime divisors of K/k, we can develop the theory of the function field K/k, which is similar to the theory of nonsingular curves Γ (- Sections B, C). Thus we define the **genus of the function field** K/k.

Suppose we are given an algebraic function field K/k of dimension 1. An algebraic curve Γ defined over k is called a **model** of K/k if $k(\Gamma)$ and K are k-isomorphic. For any function field of dimension 1, there always exist two elements x and y in K such that $K=k(x,y)$. Let $f(X, Y)$ be an irreducible polynomial such that $f(x,y)=0$. Then the plane curve defined by the equation $f(X,Y)=0$ is a model of K/k. Among

the models of K/k there exists a †normal model Γ_0 over k that is unique up to isomorphism (and the uniqueness of the normal model of the function field within the birational equivalence class of varieties holds only for curves). In particular, if k is the complex number field, the normal model Γ_0 is the †Riemann surface of the function field K/k. If Γ_0 has no singular point, the theory of the curve Γ_0 and the theory of the function field K/k are essentially identical. (This occurs, for example, when k is †perfect.) In that case the genus of Γ_0 is equal to that of K/k. In general, the genus of the function field is not less than the genus of the normal model Γ_0, and it is greater than the latter if Γ_0 has a singular point. If the genus of K/k is zero, we can take a plane quadratic curve as a model of K/k. Moreover, K/k has a prime divisor of degree 1 if and only if K is a purely transcendental extension of k. A function field K/k of genus 1 is called an **elliptic function field.** If K has a prime divisor of degree 1, an elliptic function field K has a model of a plane cubic curve. Moreover, if the characteristic of the field k is different from 2, we can take as the model Γ_0 the curve defined by an equation of the form $Y^2=4X^3-g_2X-g_3$. This is called **Weierstrass's canonical form.** The number $j=(g_2{}^3-27g_3{}^2)^{-1}g_2{}^3$ (#O)is a birational invariant of Γ_0.

E. Jacobian Varieties

Let Γ be a nonsingular curve. A †group variety J is called the **Jacobian variety** of Γ if it has the following four properties (we fix an algebraically closed †field k of definition for Γ and J): (i) There exists an isomorphism Φ (of abstract groups) of $G_0(\Gamma)/G_l(\Gamma)$ into J. (ii) Φ is continuous in the following sense: Let ii, $\bar{b}$ be elements of $G_0(\Gamma)/G_l(\Gamma)$ represented by a, b. If b is a specialization of a over a field $K(\supset k)$, then $\Phi(\bar{b})$ is also a specialization of $\Phi(\bar{a})$ over K. (iii) If there exists a K-rational divisor in the class $\bar{a}$, then the point $\Phi(\bar{a})$ is also K-rational. (iv) For any $\xi\in J$, there exists a $k(\xi)$-rational divisor a in G_0 such that $\Phi(\mathfrak{a} \bmod G_l)$ = ξ. A group variety J satisfying these conditions is necessarily a complete variety, hence an †Abelian variety, and is determined uniquely up to isomorphism. The construction of Jacobian varieties over a field of arbitrary characteristic is due to A. Weil [27] (for analytic construction → 11 Algebraic Functions C).

Let P be a †generic point of Γ over k, and let P_0 be a fixed k-rational point. Then $\varphi(P)=\Phi(P-P_0)$ defines a rational mapping of Γ into J, and φ, which is an isomorphism of Γ and its image $\varphi(\Gamma)$, is determined uniquely by Φ up to translation on J. This mapping φ is called the **canonical function on** Γ. The dimension of J is equal to the genus g of Γ. If $P_1,\ldots,P_g$ are independent generic points of Γ over k, then $k(P_1,\ldots,P_g)_s$ is the function field of J over k, where $k(P_1,\ldots,P_g)_s$ is the subfield invariant under the group of g! automorphisms $(P_1,\ldots,P_g)\rightarrow(P_{\alpha_1},\ldots,P_{\alpha_g})$. The Jacobian variety of Γ is also the †Picard variety of Γ, and it is equal to the †Albanese variety of Γ (→ 16 Algebraic Varieties P). Hence for any function f on Γ with values in an Abelian variety A, there exists a unique homomorphism λ of J into A such that $f=\lambda\circ\varphi+\text{const}$. This λ is called the **linear extension** of f.

Let Θ be the set of points on J that can be written as $\varphi(P_1)+\ldots+\varphi(P_{g-1})$. Then Θ is an irreducible subvariety of codimension 1. The divisor Θ is called the **canonical divisor** of J. The Jacobian variety that is polarized by the divisor Θ is called the **canonically polarized Jacobian variety** (→ 3 Abelian Varieties G). If two curves Γ and Γ' are birationally equivalent, the canonically polarized Jacobians of Γ and Γ' are isomorphic. Conversely, if the canonically polarized Jacobian varieties J of Γ and J' of Γ' are isomorphic, then Γ and Γ' are birationally equivalent **(Torelli's theorem).** Let r be any integer such that $1\leqslant r\leqslant g$, and let W_r be the set of points that are written in the form

$$\varphi(P_1)+\ldots+\varphi(P_r)$$

$(W_1=\varphi(\Gamma), W_{g-1}=\Theta, W_g=J)$. Then we have $\Theta^{(r)}=r!\,W_{g-r}$ (†numerically equivalent) and $(\Theta^{(g)})$ = y!, where $\Theta^{(r)}$ is the class of intersections of r copies of Θ. The existence of a divisor Θ is characteristic for Jacobian varieties. Actually, if A is an Abelian variety of dimension n that has an irreducible subvariety X^{n-1} of codimension 1 and a positive 1-cycle C such that $(X^{(n)})$ = n! and $X^{(n-1)}=(n-1)!$ C, then C is a nonsingular irreducible curve, A is the Jacobian variety of C, and X is the canonical divisor. The canonical divisor Θ is defined by a †theta function in the classical case.
For a divisor a = $\sum n_iP_i$, we define $\varphi(\mathfrak{a})$ to be $\sum n_i\varphi(P_i)\in J$. For fixed numbers $1\leqslant r\leqslant g$, $0<d$, we let W_r^d denote $\{\varphi(\mathfrak{a})\in J\mid l(\mathfrak{a})\geqslant d+1, \deg\mathfrak{a}=$ Y$\}$. Then $W_r^d\subseteq W_r$ and is a †subscheme of J. We call the number $\rho=g-(d+1)(g-r+d)$ the **Brill-Noether number.** Then dim $W_r^d\geqslant\rho$. Moreover, if Γ is a general curve, then dim W_r^d $=\rho$. In particular, if $\rho<0$, then $W_r^d=\varnothing$. This result has been verified recently by S. Kleiman, D. Laksov, P. Griffiths, and J. Harris [9, 14].

Let Γ be a nonsingular curve, and let ω be a differential form on Γ. If the divisor (w) is $\succ 0$, the ω is called a regular 1-form or **differential form of the first kind.** Let Ω be the †sheaf of germs of regular differential forms. A differential form of the first kind is an element of

$H^0(\Gamma, \Omega)$, and vice versa. Let $\mathfrak{k}$ be a canonical divisor. Then we have a natural isomorphism $H^0(\Gamma, \Omega) \cong L(\mathfrak{f})$, and the number of linearly independent differential forms of the first kind is equal to the genus g of Γ. The †residue of a differential can be defined as in the classical case. A differential that has nonzero residues is called a **differential of the third kind.** The **residue theorem** $\sum \mathrm{Res}_P \omega = 0$ holds for any differential ω. The form ω is called a **differential form of the second kind** if for any $P \in \Gamma$ there exists a rational function f_P such that $\omega - df_P$ is regular at P. The set of differential forms of the second kind forms a linear space G_2 over the universal domain and contains the subspace G, consisting of the differential forms of the first kind. The quotient space $G_2/dk(\Gamma)$ has dimension 29 or y according as the characteristic of the universal domain is 0 or not.

When the characteristic p of the universal domain is positive, we have what is called the Cartier operator. Let Γ be a curve defined over a perfect field k, let L = $k(\Gamma)$, and let t be an element of L that is transcendental over k and such that $L/k(t)$ is separable. Then any differential ω of L/k is written uniquely as $\omega = (f_0^p + f_1^p t + \ldots + f_{p-1}^p t^{p-1})dt$, where $f_i \in L$. Then the **Cartier operator C** given by $C\omega = f_{p-1}\ dt$ is well defined and independent of the choice of t and leaves G_1 invariant. Hence given a basis $\omega_1, , \omega_g$ of G_1, we obtain a matrix (a,) with coefficient in L by Co, = $\sum a_{ij}\omega_j(1 \leqslant i \leqslant$ y). This g x g matrix A is called the **Hasse-Witt matrix** of Γ. The class of A modulo the transformations of the form $S^{-p}AS$ is a birational invariant of Γ and plays an important role in the theory of unramified cyclic p-extensions of the algebraic function field.

F. Generalized Jacobian Varieties

The notion of linear equivalence of divisors on a nonsingular curve can be extended to a more general situation. Such attempts have been made by M. Noether, F. Severi, and M. Rosenlicht, who succeeded in obtaining such a generalization [22].

Let Γ be an algebraic curve, and let $P_1, , P_t$ be singular points of Γ. Let $\mathfrak{O}_{P_i}$ be the †local ring of P_i. We set $\mathfrak{O} = \bigcap_{i=1}^t \mathfrak{O}_{P_i}$ and $\Gamma' = \Gamma - \{P_1, , P_t\}$. An element of the free Abelian group $G(\Gamma)$ generated by points of Γ' is called a Γ-divisor. Let a be a Γ-divisor and set $L(\mathfrak{a}) = \{f \in \mathfrak{O} \mid (f) + \mathfrak{a} \succ 0\} \cup \{0\}$. Then $L(a)$ is a finite-dimensional linear space (over the universal domain). The dimension of $L(a)$ is denoted by l(a), and we set $\dim|\mathfrak{a}| = \bar{l}(\mathfrak{a}) - 1$. The upper bound π of $\deg(\mathfrak{a}) - \overline{\dim}|\mathfrak{a}|$ is a nonnegative integer and is called the **$\mathfrak{O}$-genus** of Γ. We call $\bar{i}(\mathfrak{a}) = \pi - \deg \mathfrak{a} + \overline{\dim}|\mathfrak{a}|$ the **$\mathfrak{O}$-speciality index** of the divisor a. Let C be a nonsingular curve birationally equivalent to Γ, and let $Q_1, \ldots, Q_s$ be points of C that correspond to singular points of Γ. An **$\mathfrak{O}$-differential** ω is a differential form on C (of $K(\Gamma)$ = **K(C)**) such that $\sum_{i=1}^s \mathrm{Res}_{Q_i} f\omega = 0$ for any $f \in \mathfrak{O}$. Then i(a) is equal to the number of linearly independent $\mathfrak{O}$-differentials ω such that $(\omega) \succ \mathfrak{a}$ in Γ'. The equality $l(\mathfrak{a}) = \deg \mathfrak{a} - \pi + 1 + i(\mathfrak{a})$ is called the **generalized Riemann-Roch theorem.** An $\mathfrak{O}$-differential ω is called an $\mathfrak{O}$-differential of the first kind if ω is regular everywhere on Γ'. The number of linearly independent $\mathfrak{O}$-differentials of the first kind is equal to the $\mathfrak{O}$-genus π. Let g be the effective genus of Γ, i.e., the genus of C. Then we have the equality $\pi - g = \dim_K(\bar{\mathfrak{O}}/\mathfrak{O}) = \delta$, where $\bar{\mathfrak{O}}$ is the integral closure of 8 in **K(T).** The set of $\mathfrak{O}$-differentials forms an D-module that is in general not of rank 1. Hence in this case, we do not have the “canonical divisor.” Let c be the conductor of $\bar{\mathfrak{O}}/\mathfrak{O}$. Then c determines a divisor in a natural way. If we denote the degree of this divisor by d, we have the inequality $\delta + 1 \leqslant d \leqslant 2\delta$. We have $d = 2\delta$ if and only if the set of $\mathfrak{O}$-differentials forms an D-module of rank 1. This case occurs, for example, if Γ is a curve on a nonsingular surface or a complete intersection. Two Γ-divisors a and b are said to be **$\mathfrak{O}$-linearly equivalent** if there exists a unit f of $\mathfrak{O}$ such that a-b=(f). The set of Γ-divisors that are $\mathfrak{O}$-linearly equivalent to zero forms a subgroup $\bar{G}_l(\Gamma)$ of G(T). There exists a group variety $J_{\mathfrak{O}}$, unique up to isomorphism, that satisfies the four conditions required for Jacobian varieties (- Section E) with respect to the class group $\bar{G}_0(\Gamma)/\bar{G}_l(\Gamma)$. The variety J_0 is called the **generalized Jacobian variety.** The generalized Jacobian variety is not complete, in general. If J is the Jacobian variety of C, then $J_{\mathfrak{O}}$ is an extension of J by a connected †linear algebraic group $I_{\mathfrak{O}}$. Any Abelian extension of the function field of Γ can be obtained by the †isogenies of the generalized Jacobian variety of Γ [22]. This fact plays an important role in class field theory over algebraic function fields (→ 59 Class Field Theory G). The theory for nonsingular curves is considered as the special case in which $\mathfrak{O}$ = $K(\Gamma)$.

Suppose that Γ is situated in a projective space of dimension n. Let p be the prime ideal in $k[X_0, X_1, , X_n]$ defining Γ and χ(p, m) be the number of linearly independent homogeneous polynomials of degree m modulo p. Then χ(p, m) is a polynomial in m for large m. This polynomial is called the **Hilbert polynomial** of p (or Γ). Let c be the constant term of the Hilbert polynomial. The number $p_a(\Gamma) = 1 - c$ is called the **arithmetic genus** of Γ and is equal to the $\mathfrak{O}$-genus of Γ. Let Γ be a nonsingular

irreducible curve in $\mathbf{P}^3$ of degree d. If Γ is contained in a plane, then $p_a(\Gamma) = (d \quad 1)\cdot(d\text{-}2)/2$. Otherwise, $p_a(\Gamma) \leqslant d^2/4 - d + 1$ when d is even, and $p_a(\Gamma) \leqslant (d^2 - 1)/4\text{-}d + 1$ when d is odd [11].

G. Sheaf Theory

Let Γ be an irreducible curve and $\mathcal{O}_p$ be the local ring of p. Then $\mathcal{O}_\Gamma = \bigcup \mathcal{O}_p$ is an talgebraic coherent sheaf, which is called the structure sheaf of Γ, and dim, $H^1(\Gamma, \mathcal{O}_\Gamma)$ is equal to the arithmetic genus π of Γ. Let a be a Γ-divisor, and let $\mathcal{O}_\Gamma(\mathfrak{a})$ be the †sheaf of germs of rational functions f such that $(f) + \mathfrak{a} \succ 0$ and $f \in \mathfrak{O}_Q$ for every singular point Q of Γ (- 383 Sheaves D). Then dim, $H^1(\Gamma$, O,(a)) is equal to the speciality index $\bar{i}(\mathfrak{a})$, and dim, $H^0(\Gamma, \mathcal{O}_\Gamma(\mathfrak{a}))$ is equal to I(a). When Γ has no singular point, the Riemann-Roch theorem is deduced naturally from †Serre's duality theorem: $H^1(\Gamma, \mathcal{O}_\Gamma(\mathfrak{a})) \simeq H^0(\Gamma, \mathcal{O}_\Gamma(\mathfrak{k} - \mathfrak{a}))$.

H. Algebraic Correspondence

Let Γ be a nonsingular curve. A divisor of the product variety Γ x Γ is called an **algebraic correspondence** of Γ [26, 27]. Let D_0 be the subgroup consisting of divisors that are linearly equivalent to degenerate divisors a x $\Gamma + \Gamma$ x b. Then the class group $\mathscr{C}(\Gamma)$ = $G(\Gamma$ x $\Gamma)/D_0$ is called the **group of classes of algebraic correspondences.** We write X $\equiv 0$ if X is an element of D_0. Let X be an algebraic correspondence, k a field of definition for Γ over which X is rational, and P a generic point of Γ over k. Then $X(P) = \mathrm{pr}_2[X(P$ x $\Gamma)]$ is rational over $k(P)$. The composite X_1 o X_2 of two correspondences X, and X_2 is defined by $(X_1 \circ X_2)(P) = X_1(X_2(P))$ whenever they have meaning. The composite X, o X_2 determines an element of $\mathscr{C}(\Gamma)$ that depends only on the classes of X, and X,. This multiplication supplies the group $\mathscr{C}(\Gamma)$ with the structure of an associative ring. This ring is called the **correspondeuce ring** of Γ. The correspondence ring $\mathscr{C}(\Gamma)$ and the ring $\mathscr{A}$ of endomorphisms of the Jacobian variety J are isomorphic, and the isomorphism is given by the following rule: Let ξ be an element of $\mathscr{C}(\Gamma)$, and let X be a divisor in ξ. Let P be a generic point of Γ with reference to k over which X is rational. Let P_0 be a k-rational point of Γ. Then the class of $X(P) - X(P_0)$ modulo $G_l(\Gamma)$ is independent of the choice of a divisor X in the given class. We set $\Psi(P) = \Phi(X(P) - X(P_0))$ andlet λ be the linear extension of Y. The correspondence $\xi \rightarrow \lambda$ is an anti-isomorphism of $\mathscr{C}(\Gamma)$ and $\mathscr{A}$. Now we set $\mathscr{A}_0 = \mathscr{A} \otimes Q$. Then $\mathscr{A}_0$ contains an automorphism ι of order 2 called an **involution.** Let 1 be a rational prime different from the characteristic p. Then A has a faithful representation by 2g x $2g$ matrices with coefficients in l-adic integers. The †trace σ of this representation has the property that $\sigma(\beta\ 0\ \beta') > 0$ if $\beta \neq 0$ **(Castelnuovo's lemma).** $\mathscr{A}_0$ is an algebra of finite tank over Q, and $\mathscr{A}$ is a finitely generated Abelian group. Based on these results A. Weil proved the Riemann hypothesis for congruent ζ-functions on a nonsingular curve (- 450 Zeta Functions P).

I. Coverings

Let Γ and C be nonsingular curves such that there exists a regular mapping $\pi: \Gamma \rightarrow C$. Then there is an injection of the function field **K(C)** into **K(T).** If **K(f)** is separably algebraic over **K(C),** then Γ is called a **covering (curve)** of C. The integer **[K(I): K(C)]** = n is called the **degree of covering.** Let P be a point of Γ and let Q=n(P). Let t, s be local parameters at P on Γ and at Q on C, respectively. The nonnegative integer $v_P(ds/dt)$ is called the **differential index** at P and is denoted by m_P. The index m_P is zero except for a finite number of points. The divisor $\sum m_P P$ is called the **branch divisor.** The covering Γ is called an **unramified covering** if the branch divisor is zero. If we denote the branch divisor by a, we have the formula $2g(\Gamma) - 2 = n(2g(C) - 2) + \deg \mathfrak{a}$, where $g(\Gamma)$ and g(C) are genera of Γ and C, respectively. This is called the **Riemann-Hurwitz formula.** This formula yields at once that a rational curve has no nontrivial unramified covering and that Γ can be an unramified covering of itself if and only if Γ is an elliptic curve.

J. Theory of Moduli

Let $\mathscr{M}_g(k)$ be the set of isomorphy classes of complete nonsingular irreducible curves (here simply called "curves") of genus g defined over a field k. We can endow $\mathscr{M}_g(k)$ with a structure of an algebraic variety over k with the property that for any smooth family (over k) $\pi: \mathscr{C} \rightarrow S$ of curves of genus y the map $T: S \rightarrow \mathscr{M}_g(k)$ sending s to the isomorphy class of the fiber $\pi^{-1}(s)$ is a †morphism. This algebraic variety is called the **(coarse) moduli space of curves of genus** g. Furthermore, the moduli space over Spec(**Z**) exists (D. Mumford [16]). $\mathscr{M}_g(k)$ is †normal, irreducible and †quasiprojective (but not complete for $g > 0$) of dimension 0 (g = 0), 1 $(g = 1)$, $3g - 3$ $(g \geqslant 2)$ $(= 3g - 3 + \dim \mathrm{Aut}(C))$ with only quotient singularity [5].

Since when g = 1 any elliptic curve C over an algebraically closed field k is isomorphic to a plane curve with the Weierstrass canonical

form $y^2=4x^3-g_2x-g_3$ (if $\mathrm{char}(k)\neq 2$; the case of char(k) = 2 needs a slight modification), the correspondence $C\to j(g_2, g_3)$ (→ Section D) defines an isomorphism $\mathscr{M}_1 \simeq \mathbf{A}_k^1$ (affine line). In the case of g $\geqslant 2$ and k = C we have another function-theoretic construction of $\mathscr{M}_g$ due to Teichmüller (- 11 Algebraic Functions F).

Let $\mathscr{A}_g$ be the coarse moduli space (over **Z)** of principally polarized Abelian varieties of dimension g (→ 3 Abelian Varieties 1). For a curve C we denote the Jacobian variety of C by J(C) (- Section E). The correspondence $C\to J(C)$ defines a morphism $i:\mathscr{M}_g\to\mathscr{A}_g$ which is injective **(Theorem of Torelli).** It is even an †immersion (F. Oort, J. Steenbrinck). If g= 1, 2, 3, then i is an open immersion whose image we can describe.

For k =C we can express the above map i by using periods of curves. Namely, let $\alpha_1, , \alpha_g, \beta_1, , \beta_g$ be a (canonical) basis of H_1 (C, **Z)** detïned by †normal sections of C (considered as a real surface). Let $\omega_1, . . . , \omega_g$ be a basis of †differential forms of the first kind on C with $\int_{\beta_i}\omega_j = \delta_{ij}$. Then the matrix Ω = $(\int_{\alpha_i}\omega_j)$ is symmetric and has a positive definite imaginary part, i.e., is an element of the †Siegel upper half-space $\mathfrak{S}_g$ of degree y. With the identification $\mathscr{A}_g \simeq \mathfrak{S}_g/\mathrm{Sp}(g, \mathbf{Z})$ (- 3 Abelian Varieties 1) the map $i:\mathscr{M}_g\to\mathscr{A}_g$ is nothing but the one sending the isomorphy class of C to Rmod $\mathrm{Sp}(g, \mathbf{Z})$. For g $=4$ the closure of $i(\mathscr{M}_4)$ is a principal divisor detïned explicitly with theta constants (Schottky, J.-I. Igusa).

K. Stable Curves

A reduced connected complete curve C over k is called a **stable curve** of genus g (> (0) if(i) C has only ordinary double points as possible singularities; (ii) when Γ is a smooth rational irreducible component of C, then Γ intersects the other components in more than 2 points; (iii) dim, $H^1(C, \mathcal{O}_C) = g$.

There exists a †coarse moduli space $\mathscr{S}_g$ of stable curves of genus g which contains $\mathscr{M}_g$ as a Zariski open subset. The space $\mathscr{S}_g$ detïned over Spec(**Z**) is †complete and even †projective; hence $\mathscr{L}_g$ gives a compactification of $\mathscr{M}_g$. The completeness of $\mathscr{S}_g$ follows from the **stable reduction theorem** as follows: let R be a discrete valuation ring with quotient field K and C be a smooth connected curve over K. Then there exists a finite separable algebraic extension L of K such that the curve C x $_K L$ extends to a flat family of stable curves over $\mathrm{Spec}(R_L)$, where R_L denotes the integral closure of R in L. In this case we say that the curve **C x** $_K L$ has a **stable reduction** in R,. With the above notation a curve C over K has a stable reduction in R if and only if its Jacobian variety J(C) has a †stable reduction in R (- 3 Abelian Varieties N).

Over C, $\mathscr{A}_g$ has the †Satake compactification $\bar{\mathscr{A}}_g$, which is a disjoint union of $\mathscr{A}_{g'}$, $0 \leqslant g' \leqslant g$, as a set. The injection $i:\mathscr{M}_g\to\mathscr{A}_g$ (- Section J) extends to a morphism $j:\mathscr{S}_g\to\bar{\mathscr{A}}_g$ that sends the isomorphy class of a stable curve C = $C_1 \cup \cup C_k$ (C_i irreducible) to that of $J(\tilde{C}_1)$ x x $J(\tilde{C}_k)\in\mathscr{A}_{g'}$, $g'=\sum_i \mathrm{genus}(\tilde{C}_i)$, where $\tilde{C}_i$ is the †normalization of C_i. In particular the closure of the image of i in $\bar{\mathscr{A}}_g$ is the set of products of Jacobian varieties.

References

[1] S. Arakelov, Families of algebraic curves with fixed degeneracies, Math. USSR-I~V., 5 (1971), 127771302.

[2] H. F. Baker, Analytical principles of the theory of curves, Principles of Geometry 5, Cambridge Univ. Press, 1933.

[3] A. Brill and M. Noether, Über die algebraischen Functionen und ihre Anwendungen in der Geometrie, Math. Ann., 7 (1874), 269–310.

[4] G. Castelnuovo, Numero delle involutione rationali giancenti sopra uno curva di dato genere, Rend. Acad. Lincei, 4, 5 (1889).

[5] P. Deligne and D. Mumford, The irreducibility of the space of curves of given genus, Publ. Math. Inst. HES, 36 (1969), 75– 109.

[6] F. Enriques and 0. Chisini, Teoria geometrica delle equazioni dele funzioni algebriche, Zanichelli, Bolgna, 1934.

[7] T. Fujita, Detïning equations for certain types of polarized varieties, Complex Analysis and Algebraic Geometry, Iwanami, Tokyo, 1977, 165–173.

[8] W. Fulton, Algebraic curves, Benjamin, 1969.

[9] P. Griffiths.and J. Harris, Principles of algebraic geometry, Wiley, 1978.

[10] P. Griffiths and J. Harris, On the variety of special linear systems on a general algebraic curve, Duke Math. J., 47 (1980), 233-272.

[1 1] R. Hartshorne, Algebraic geometry, Graduate text 52, Springer, 1977.

[12] R. Hartshorne, On the classification of algebraic space curves, Vector Bundles and Differential Equations, Birkhauser, 1980.

[13] G. Kempf, Schubert methods with an application to algebraic curves, Publications of Mathematisch Centrum, Amsterdam, 1971.

[14] S. Kleiman and D. Laksov, On the existence of special divisors, Amer. J. Math., 94 (1972), 43 1-436.

[15] S. Kleiman, r-Special subschemes and an argument of Severi's, Advances in Math., 22 (1976), l-31.

[16] D. Mumford, Geometric invariant theory, Erg. Math., Springer, 1965.
[17] D. Mumford, Curves and their Jacobians, Univ. of Michigan Press, 1975.
[18] M. Namba, Families of meromorphic functions on compact Riemann surfaces, Lecture notes in math. 767, Springer, 1979.
[19] M. Noether, Zur Grundlegung der Theorie der algebraischen Raumcurven, J. Reine Angew. Math., 93 (1882), 271-318 (Berlin, 1883).
[20] B. Saint-Donat, On Petri's analysis of the linear system of quadrics through a canonical curve, Math. Ann., 206 (1973), 157-175.
[21] A. Seidenberg, Elements of the theory of algebraic curves, Addison-Wesley, 1968.
[22] J.-P. Serre, Groupes algébriques et corps de classes, Actualités Sci. Ind., Hermann, 1959.
[23] F. Severi, Vorlesungen über algebraische Geometrie, Teubner, 1921.
[24] J. M. Wahl, Deformations of plane curves with nodes and cusps, Am. J. Math., 96 (4) (1974), 529-577.
[25] R. J. Walker, Algebraic curves, Princeton Univ. Press, 19.50.
[26] A. Weil, Sur les courbes algébriques et les variétés qui s'en déduisent, Actualités Sci. Ind., Hermann, 1948.
[27] A. Weil, Variétés abéliennes et courbes algébriques, Actualités Sci. Ind., Hermann, 1948.
[28] A. Weil, Zum Beweis des Torellischen Satzes, Nachr. Akad. Wiss. Göttingen (1957), 33-53.
[29] H. Weyl, Die Idee der Riemannschen Flache, Teubner, third edition, 1955.
[30] G. B. Winters, On the existence of certain families of curves, Amer. J. Math., 96 (1974), 215-228.

10 (111.5) Algebraic Equations

A. General Remarks

Let $F_1(X_1,\ldots,X_m),\ldots,F_r(X_1,\ldots,X_m)$ be r †polynomials in m variables $X_1,\ldots,X_m$ over a †field k. Then the equations

$$F_1=0,\ \ldots,\ F_r=0$$

are called **algebraic equations in m unknowns.** When we consider these equations simultaneously, where $r\geqslant 2$, we call them a **system of r equations** or **simultaneous equations.** (For $r=1$, a system of one equation means the single equation $F_1=0$.) Coefficients of $F_1,\ldots,F_r$ are called **coefficients** of the system, and the greatest of the degrees of $F_1,\ldots,F_r$ is called the **degree** of the system.

To **solve** a system of equations (henceforth in this article we shall omit the word "algebraic") means to find the common †zero points (in an †algebraically closed field containing k) of elements $F_1,\ldots,F_r$ of the †polynomial ring $k[X_1,\ldots,X_m]$. If there exist no common zero points, the system is said to be **inconsistent;** if there exists a finite number of such points, it is said to be **regular;** and if there are an infinite number of such points, it is called **indeterminate**. The †elimination method allows us to reduce the problem of solving a system of r equations to the case r = 1. In particular, any regular system of equations can be reduced to the case $m=r=1$.

B. Equations in One Unknown

For the above reason, it is important to consider an equation of the form $f(X)=0$, where

$$f(X)=a_0X^n+a_1X^{n-1}+\ldots+a_n,\quad a_0\neq 0. \tag{1}$$

This gives the general form of an algebraic equation in one unknown.

According as $f(X)$ is reducible or not in the †polynomial ring $k[X]$, the equation $f(X)=0$ is called **reducible** or **irreducible (→ 337 Polynomials)**. In some †algebraic extension field K of k, $f(X)$ can be factored:

$$f(X)=a_0(X-\alpha_1)(X-\alpha_2)\ldots(X-\alpha_n). \tag{2}$$

$\alpha_1,\ldots,\alpha_n$ are called the **roots** of the equation $f(X)=0$. Hence, any algebraic equation of degree n has exactly n roots **(Kronecker's theorem).** Now, $(-1)^i a_i/a_0$ is equal to the †elementary symmetric function of degree i of $\alpha_1,\ldots,\alpha_n$. Some of the roots $\alpha_1,\ldots,\alpha_n$ may be identical. If α appears ρ times in $\alpha_1,\ldots,\alpha_n$, we say that α is a ρ-tuple root, and ρ is called the **multiplicity** of the root α. When $\rho=1$, α is called a **simple root**, and when $\rho\geqslant 2$, α is called a **multiple root.** Let $\beta_1,\ldots,\beta_\nu$ be all the distinct roots among $\alpha_1,\ldots,\alpha_n$, and let ρ_i be the multiplicity of $\beta_i (i=1,\ldots,\nu)$. Then

$$f(X)=a_0(X-\beta_1)^{\rho_1}\ldots(X-\beta_\nu)^{\rho_\nu}, \tag{2'}$$

$$\rho_1+\ldots+\rho_\nu=n.$$

If $\rho_1,\ldots,\rho_\nu$ are not divisible by the †characteristic of k, the greatest common divisor g of f and

$$f'=na_0X^{n-1}+(n-1)a_1X^{n-2}+\ldots+a_{n-1}$$

is $(X-\beta_1)^{\rho_1-1}\ldots(X-\beta_\nu)^{\rho_\nu-1}$. Thus we can reduce the multiplicity of every root to 1 by dividing f by g. Any irreducible equation over a field of characteristic 0 has no multiple roots. Equation (1) has multiple roots if and only if

its †discriminant D is equal to 0 (- 149 Fields; 172 Galois Theory).

C. Equations of Special Type

In Sections C and D we assume that the characteristic of k is zero.

Binomial Equations. An equation of the type $X^m - a = 0$ is called a **binomial equation.** It is solved by **root extraction.** Let $\sqrt[m]{a}$ **(mtb root** of u) be one of the roots (if a is a positive real number, $\sqrt[m]{a}$ usually denotes a positive real root). Then $\sqrt[m]{a}$ multiplied by 1, $\zeta, \zeta^2, , \zeta^{m-1}$ are the roots of $X^m - a = 0$, where ζ is a †primitive mth root of unity.

Reciprocal Equations. An equation a, $X^n + a_1 X^{n-1} + \ldots + a_n = 0$ is called a **reciprocal equation** if $a_0 = a_n$, $a_1 = a_{n-1}$, $a_2 = a_{n-2}, \ldots$. A reciprocal equation of an odd degree $n = 2m+1$ has a root X = -1, and dividing the left side by X + 1 we get a reciprocal equation of degree $2m$. A reciprocal equation of degree $n = 2m$ is reduced to an equation of degree m in Y= X + X $^{-1}$ and the quadratic equation $X^2 - XY + 1 = 0$.

D. Equations of Lower Degree
(- Appendix A, Table 1)

(1) A **linear equation** $a_0 X + a_1 = 0$ has a single root $-a_1/a_0$. **(2)** the roots of a **quadratic equation** $a_0 X^2 + a_1 X + a_2 = 0$ are given by $(-a_1 \pm \sqrt{a_1^2 - 4a_0 a_2})/2a_0$. **(3)** To solve a **cubic equation** $a_0 X^3 + a_1 X^2 + a_2 X + a_3 = 0$, **we** set $A_1 = 9a_0 a_1 a_2 - 2a_1^3 - 27a_0^2 a_3$, $A_2 = a_1^2 - 3a_0 a_2$, and solve the quadratic equation T^2 **A,** $T + A_2^3 = 0$. Let t_1 and t_2 be the roots of this quadratic equation, and let ω be any cube root of 1. Then (-a, $+\omega\sqrt[3]{t_1} + \omega^2\sqrt[3]{t_2})/3a_0$ is a root of the original cubic equation. **(Cardano's formula).** If we apply this method to a cubic equation $aX^3 + bX^2 + cX + d = 0$ with real coefficients, we need to use complex cube roots even if the roots of the equation are real. In fact, it has been proved that it is not possible to solve this equation within the real numbers in this case; i.e., if the cubic equation is irreducible over the extension $\mathbf{Q}(a, b, c, d)$ of the rational number field **Q**, and if all of its roots are real, it is impossible to find the roots only by rational operations and with real radicals. This is called the **casus irreducibilis.** **(4) A quartic equation** $a_0 X^4 + a_1 X^3 + a_2 X^2 + a_3 X + a_4 = 0$ can be solved by means of reduction to a cubic equation (L. Ferrari) (→ Appendix A, Table 1). Generally, the procedure of solving an algebraic equation, i.e., finding the roots of a given equation from its coefficients by means of a finite number of rational operations and extractions of radicals, is called a **solution by radicals** (or **algebraic solution).** The †general algebraic equation whose degree is $\geqslant 5$ cannot be solved by radicals (N. H. Abel) (→ 172 Galois Theory).

E. Analytic Tbeory

In this section, k denotes the field **R** of real numbers or the field C of complex numbers. These cases have been studied for a long time, for practical reasons.

Concerning the case $k = C$, the field C is †algebraically closed; i.e., every equation with coefficients from C has a root in C **(Gauss's tbeorem,** called the **fundamental tbeorem of algebra).** Accordingly, in the fïeld C, we always have equations (2) and (2').

Let $\alpha_1, , \alpha_n$ be the roots of equation (1). Then each α_i is a continuous function of coefficients a,, a,, , a,. Concerning the location of roots of $f(X) = 0$ and $f'(X) = 0$ on the complex plane, we have the following theorems:

(1) Any convex polygon on the complex plane containing the roots of $f(X) = 0$ also contains the roots off'(X) = 0 (Gauss).

(2) Let C be a rectifiable †Jordan curve not passing through a root off(X) = 0. Then the number (C, f) of the roots of $f(X) = 0$ lying in the region enclosed by C is equal to $(1/2\pi i)\int_C (f'(z)/f(x))\,dz$, where the multiplicity of the roots is taken into account.

(3) Let C be a Jordan curve on the complex plane. If $|f(z)| > |g(z)|$ at every point z on C, then the equations $f = 0$ and $f + g = 0$ have the same number of roots (counting multiplicity) within the region enclosed by **C (Roucbé's theorem).**

(4) The absolute value of a root of equation (1) is less than

$$M = \max(|a_1/a_0|, \ldots, |a_n/a_0|) + 1.$$

(5) Let D be the †discriminant of f, and assume that $|\alpha_i| \leqslant M$ $(i = 1, \ldots, n)$. Then $|\alpha_i - \alpha_j|^2 \geqslant D/(2M)^{n(n-1)-2} = E$. Since the value of $|D|$ is known fromf'and one value of M is given by theorem (4), we have one value of E. If we draw a circle on the complex plane with center at the origin and with radius M, and if we cover it with a net whose meshes have diameters less than $\sqrt{E}/2$, then the interior of each mesh contains at most one root of $f = 0$.

When $k = \mathbf{R}$, i.e., $f \in \mathbf{R}[X]$, let $\beta_1, \ldots, \beta_\nu$ denote the distinct roots off= 0, and recall equation (2'). Suppose that $\beta_1, , \beta_\lambda \in$ R and the others $\notin \mathbf{R}$. Then $\nu - \lambda$ is an even integer 2κ, and we can renumber $\beta_{\lambda+1}, , \beta_\nu$ so that $\beta_{\lambda+1} = \beta_{\lambda+\kappa+1}, \ldots, \overline{\beta}_{\lambda+\kappa} = \beta_\nu$ (β denotes the con-

jugate of β) and $\rho_{\lambda+1}=\rho_{\lambda+\kappa+1},\ldots,\rho_{\lambda+\kappa}=\rho_{\nu}$. In this case, $\beta_1,\ldots,\beta_\lambda$ are the **real roots** of equation (2), and the other β's are called the **imaginary roots.**

(6) If $f\in\mathbf{R}[X]$ and $a_0>a_1>\ldots>a_n>0$, then the absolute value of any root of equation (1) is less than 1 **(Kakeya-Eneström theorem).**

Concerning the real roots of an equation $f = 0$, where $f\in\mathbf{R}[X]$, we have the following theorems: Let $N(a, b)(a, b\in\mathbf{R})$ denote the number of real roots in the interval (a, b). Furthermore, let $V(c_1, c_2, \ldots, c_p)$ denote the number of changes of sign in the sequence $c_1, c_2, \ldots, c_p$ of real numbers, which is defined as follows: Suppose that we have the sequence $c_{\nu_1},\ldots,c_{\nu_q}$ after deleting the terms $c_i = 0$ from the sequence $c_1, c_2, \ldots, c_p$. Then

$$V(c_1,c_2,\ldots,c_p)=\frac{1}{2}\sum_{j=1}^{q-1}(1-\operatorname{sgn} c_{\nu_j}c_{\nu_{j+1}}).$$

(7) $N(0, \infty)\equiv V(a_0, a_1, \ldots, a_n)$ (mod 2) and $N \leqslant V$ **(Descartes's theorem).**

(8) Let $V(c)= V(f(c),f'(c),\ldots,f^{(n)}(c))$. Then $N(a, b) \equiv V(a) - V(b)$ (mod 2) and $N \leqslant V(a) - V(b)$ **(Fourier's theorem).**

(9) We may assume that $f=0$ has no multiple roots. Construct a finite series $f_0=f, f_1=f', \ldots, f_l$ of polynomials over $\mathbf{R}$ such that $f_{i-1}=f_iq_i-f_{i+1}$ for $i=1,2,\ldots,l-1$ and $f_l\in\mathbf{R}$, by successive application of the †division algorithm. Let $V(c)= V(f_0(c),f_1(c),\ldots,f_l(c))$. Then $N(a, b) = V(a) - V(b)$ **(Sturm's theorem).** By means of this theorem we can determine the location of real roots as precisely as we wish.

(10) In order that every root α_i of an equation $f=0$ with $a_0 > 0$ lies on the left side of the imaginary axis, i.e., Re $\alpha_i<0$, it is necessary and sufficient that in the following matrix the †principal minors composed of the first r rows and first r columns be positive for all $r=1,2,\ldots,n$ **(Hurwitz's theorem):**

$$\begin{pmatrix} a_1 & a_3 & a_5 & a_7 & \ldots & \ldots \\ a_0 & a_2 & a_4 & a_6 & \ldots & \ldots \\ 0 & a_1 & a_3 & a_5 & \ldots & \ldots \\ 0 & a_0 & a_2 & a_4 & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & & 0 & a_{n-2} & & a_n \end{pmatrix}.$$

Also, for $f\in\mathbf{C}[X]$, various results have been obtained about under what conditions all the roots of $f = 0$ lie on one side of a given straight line or inside a given circle (e.g., the unit circle) (→ 301 Numerical Solution of Algebraic Equations).

References

[1] L. E. Dickson, Elementary theory of equations, Wiley, 1914.

[2] J. Dieudonné, La théorie analytique des polynômes, Mémor. Sci. Math., Gauthier-Villars, 1938.

[3] M. Marden, Geometry of polynomials, Amer. Math. Soc. Math. Surveys, second edition, 1966.

[4] W. Specht, Algebraische Gleichungen mit reelen oder komplexen Koeffizienten, Teubner, 1958.

11 (XI.10) Algebraic Functions

A. Definition

An **algebraic function** is a multiple-valued †analytic function $w = w(z)$ defined by an †irreducible algebraic equation $P(z, w) = 0$ with complex coefficients.

B. History and Methods

The theory of algebraic functions evolved from the works of C. F. Gauss, N. H. Abel, and C. G. J. Jacobi on †elliptic functions in the early 19th Century. Stimulated by their works, B. Riemann and K. Weierstrass established the foundations of the theory of complex functions and developed the important theory of algebraic functions.

The equation $P(z, w) = 0$ defines a curve in the 2-dimensional complex †projective space with inhomogeneous coordinates z, w. Investigations from this point of view were initiated by Riemann, A. Clebsch, and P. Gordan. This approach was followed by A. Brill, M. Noether, and the Italian school (F. Severi, C. Segre, etc.) and has developed into contemporary algebraic geometry (→ 9 Algebraic Curves, 12 Algebraic Geometry).

The set of †function elements w(z) satisfying $P(z, w) = 0$ is a †complex manifold $\mathfrak{R}$, a closed (= compact) †Riemann surface, on which z and w are †meromorphic functions. The field $K_{\mathfrak{R}}$ consisting of the meromorphic functions on $\mathfrak{R}$ is an †algebraic function field $\mathbf{C}(z, w)$. Conversely, for any closed Riemann surface $\mathfrak{R}$, the field $K_{\mathfrak{R}}$ is an †algebraic function field in one variable over C, and any pair of functions z and w with $K_{\mathfrak{R}} = \mathbf{C}(z, w)$ has the property that $\mathfrak{R}$ is †conformally equivalent to the Riemann surface determined in the above fashion by the irreducible algebraic equation $P(z, w) = 0$ satisfied by z and w. Two Riemann surfaces $\mathfrak{R}_1, \mathfrak{R}_2$ determined by the equations $P_1 = 0$, $P_2 = 0$ are conformally equivalent if and only if the fields $K_{\mathfrak{R}_1}$ and $K_{\mathfrak{R}_2}$ are C-isomorphic. This condition is equivalent to the existence of a †birational

transformation between the algebraic curves $P_1 = 0$, $P_2 = 0$. The "analytic method" (the method of studying algebraic functions as functions on Riemann surfaces) is the creation of Riemann. It was extended by F. Klein and D. Hilbert, and later by H. Weyl, who established in his monograph [6] a rigorous foundation of the analytic method for the theory of algebraic functions.

Given an arbitrary algebraic function field K in one variable over C, the set $\mathfrak{R}$ of its †prime divisors with a suitable topology and analytic structure is a closed Riemann surface whose function field $K_{\mathfrak{R}}$ coincides with K. The "algebraic method" (the method of studying algebraic functions as elements of an algebraic function tïeld) was founded by J. W. Dedekind and H. Weber at the end of the 19th Century. In the 20th Century, the algebraic method has made remarkable progress, owing to the development of abstract algebra. It covers the case of an arbitrary ground lïeld as well as that of more than one variable. The theory of algebraic functions has had considerable influence on the development of number theory because of a basic analogy between the two subjects.

The †universal covering spaces (surface) $\tilde{\mathfrak{R}}$ of a closed Riemann surface $\mathfrak{R}$ can be regarded, by conforma1 mapping, as the Riemann sphere, the plane, or the unit disk (or, equivalently, to the upper half-plane) if the †genus g of $\mathfrak{R}$ is 0, 1, or $\geqslant 2$, respectively. Then the tcovering transformation group G, consisting of †linear fractional transformations without fixed points in $\tilde{\mathfrak{R}}$, is †properly discontinuous and has a compact †fundamental domain. Conversely, if D is one of the three domains just mentioned and if G is the group just described, then $\mathfrak{R} = D/G$ is a closed Riemann surface such that D and G are its universal covering space and covering transformation group. A meromorphic function on $\mathfrak{R}$ is represented as an tautomorphic function on $\tilde{\mathfrak{R}}$ with respect to G. If y = 0, then G = $\{1\}$, $\tilde{\mathfrak{R}} = \mathfrak{R}$, and $K_{\mathfrak{R}}$ is the field of rational functions. If $g = 1$, then $K_{\mathfrak{R}}$ is the field of †elliptic functions. The study of algebraic functions as automorphic functions was initiated by H. Poincaré and Klein. Recently, C. L. Siegel made a remarkable contribution to the investigation of the case of several variables. The theory of automorphic functions is also related to number theory. Works of E. Hecke, M. Eichler, and G. Shimura on this domain are noteworthy (- 32 Automorphic Functions, 73 Complex Multiplication).

Another important topic concerning algebraic functions (closed Riemann surfaces) is the problem of moduli. Riemann stated, without rigorous proof, that the set of conformal equivalence classes of closed Riemann surfaces of genus g ($\geqslant 2$) depends on 3g − 3 complex parameters, called **moduli.** This has led to the modern theory of Teichmüller spaces, which is developing into an extensive new field (- 234 Kleinian Groups, 416 Teichmüller Spaces).

In the rest of this article, we deal mainly with the analytic method (for the case of two variables → [1]).

C. Abelian Differentials

An **Abelian differential** on a closed Riemann surface $\mathfrak{R}$ is, by definition, a complex †differential form $\omega = a(z)dz$, where a(z) is a meromorphic function of a local parameter z. Such a differential is said to be of the **first kind** if a(z) is holomorphic, of the **second kind** if the residue vanishes everywhere, and of the **tbird kind** otherwise.

The indelïnite integral W(p) $= \int_{p_0}^{p} \omega$ of an Abelian differential ω, where p_0 is assumed not to be a pole of ω, is called an **Abelian integral.** It is said to be of the **first, second,** or **tbird kind** if the same holds for ω. If y is a 1-†cycle on $\mathfrak{R}$, the quantity $\int_\gamma \omega$ is referred to as the **period** of ω along y. An **elliptic integral** is defined to be an Abelian integral on a closed Riemann surface of genus 1. For example, this is the case if the equation $P(z, \text{w}) = 0$ defining the surface is of degree 2 with respect to w and of degree 3 or 4 with respect to z. More generally, a closed Riemann surface is called **byperelliptic** if $P(z, \text{w})$ is of degree 2 with respect to w, or, equivalently, if $\mathfrak{R}$ carries a meromorphic function with exactly two poles. An Abelian integral on such a surface is called a **hyperelliptic integral.**

On a closed surface $\mathfrak{R}$, let V_a be the linear space over C of the Abelian differentials of the first kind. Given a l-cycle α of $\mathfrak{R}$, there exists a unique $\omega_\alpha \in V_a$ such that $\mathrm{Re}\int_\gamma \omega_\alpha$ is equal to the tintersection number (α, y) for every 1 -cycle y. This differential is also characterized by the property $((\omega, \omega_\alpha)_{\mathfrak{R}} \equiv) \int_{\mathfrak{R}} \omega \wedge * \bar{\omega}_\alpha = -2\sqrt{-1}$ $\int_\alpha \omega$ for every $\omega \in V_a$. If $\{\alpha_1, \ldots$ ..a.,$\}$ form a basis of the 1-dimensional thomology group with integral coefficients, then Re ω_{α_i} ($i = 1, .\,.$, $2g$) form a basis of the linear space V_h over **R** of the †harmonic differentials on $\mathfrak{R}$ as well as that of the space $\{$ Re ω $\omega \in V_a\}$. Accordingly,

$$\dim_{\mathbf{C}} V_a = g, \qquad \dim, V_h = 2g.$$

These identities show a close relationship between the topological structure of $\mathfrak{R}$ and the space of the Abelian differentials on $\mathfrak{R}$ (→ 194 Harmonic Integrals).

One can choose a 1-dimensional homology basis $\{\alpha_i, \alpha_{g+i}\}_{i=1}^{g}$ so that $(\alpha_i, \alpha_j) = (\alpha_{g+i}, \alpha_{g+j}) = 0$, $(\alpha_i, \alpha_{g+i}) = 1$, and $(\alpha_i, \alpha_{g+j}) = 0$ $(i \neq j)$, $(i, j = 1, 2, , \; g)$. Such a basis is called a **canonical**

homology basis. If ω and σ are Abelian differentials of the first kind, then we have

$$(\omega,\bar{\sigma})_{\mathfrak{R}} = -\sqrt{-1}\sum_{j=1}\left(\int_{\alpha_j}\omega\int_{\alpha_{g+j}}\sigma - \int_{\alpha_{g+j}}\omega\int_{\alpha_j}\sigma\right)=0,$$

$$(\omega,\omega)_{\mathfrak{R}} = \sqrt{-1}\sum_{j=1}^{g}\left(\int_{\alpha_j}\omega\int_{\alpha_{g+j}}\bar{\omega} - \int_{\alpha_{g+j}}\omega\int_{\alpha_j}\bar{\omega}\right)\geqslant 0.$$

Together, these are called the (Riemann) **period relation.** The second formula implies that ω vanishes identically if the periods of ω along all α_j (j = 1, . . , y) are zero.

Let $\alpha_1, . . , \alpha_{2g}$ be a 1-dimensional homology basis, and let $\omega_1, . . , \omega_g$ form a basis of V_a over C. The g x $2g$ matrix Ω with $\int_{\alpha_j}\omega_i$ as its (i,j)-component is called a **period matrix.** Corresponding to the change of bases *(a)* and (w), it is subject to transformation into the form $A\Omega M$, where A is a g x g invertible complex matrix and M is a 2g x $2g$ integral square matrix with determinant $\pm$ 1. Conversely, two Riemann surfaces are conformally equivalent if they possess period matrices transformable to each other in this manner **(Torelli's theorem).** We can choose $\omega_1, \ldots, \omega_g$ so that the corresponding period matrix with respect to a canonical homology basis becomes (I_g, T) with the g x g unit matrix I_g. Then from the period relation, T is symmetric and Im T is positive definite.

On the complex linear space $\mathbf{C}^g$, consider the subgroup generated by the $2g$ column vectors of a period matrix Ω (the subgroup is also denoted by Ω). Since it is of rank $2g$ and properly discontinuous, a group manifold $\mathbf{C}^g/\Omega$ is obtained. It is determined by $\mathfrak{R}$ uniquely up to analytic isomorphism and is called the **Jacobian variety** of $\mathfrak{R}$. The **generalized Jacobian variety** is introduced in a similar fashion by means of Abelian integrals of the second and third kinds (- 9 Algebraic Curves).

D. The Riemann-Roch Theorem

In the present context, a 0-tchain with integral coefficients on a Riemann surface $\mathfrak{R}$ is referred to as a **divisor.** A divisor $d = \sum n_i p_i$ ($n_i \in Z$, $p_i \in \mathfrak{R}$) is an **integral divisor** (or **positive divisor)** if $n_i \geqslant 0$ in the reduced expression; d is a **prime** divisor if it consists of a single point p_1 and $n_1 = 1$. A divisor of a meromorphic function for an Abelian differential ω is defined by taking the p_i as the zeros (poles) off or ω and $n_i(-n_i)$ as the multiplicity of the zero (pole) at p_i. The divisors on $\mathfrak{R}$ constitute an Abelian group $\mathfrak{D}$ in which **principal divisors,** i.e., divisors of meromorphic functions, constitute a subgroup $\mathfrak{P}$. The factor group $\mathfrak{D}/\mathfrak{P}$ is called the **divisor class group;** an element of it is called a **divisor class.** The divisors of Abelian differential constitute a single divisor class, which is referred to as the **canonical divisor class** (or **differential divisor class**). The **degree** and the **dimension** of a divisor class D are detïned as follows, independent of the choice of the representative $d = \sum n_i p_i \in D$: $\deg D = \sum n_i$, $\dim D = \dim_{\mathbf{C}}\{ f \mid f$ is meromorphic, (divisor of $f) + d$ is a positive divisor$\}$. For example, the degree of the principal divisor class is zero.

In terms of these concepts, the **Riemann-Roch theorem** is stated as follows: For a divisor class D on a closed Riemann surface $\mathfrak{R}$ of genus g and for an integer n, we have

$$\dim(D+nW)-\dim(-D-(1-n)W)$$
$$=\deg D+(2n-1)(g-1),$$

where W is the canonical divisor class (- 9 Algebraic Curves).

This theorem implies the following properties of $\mathfrak{R}$: (i) deg $W = 2g\quad 2$. (ii) The holomorphic invariant forms φdz^2 (i.e., analytic tensors of order 2), referred to as **quadratic differentials,** constitute a linear space over C of dimension 0 (if $g=0$), 1 (if g = 1), or of dimension $3g-3$ (if $g \geqslant 2$). The quadratic differentials have close connection with textremal quasiconforma1 mappings and play an important role in the theory of Teichmüller spaces (→ 352 Quasiconformal Mappings, 416 Teichmüller Spaces). (iii) For a point $p \in \mathfrak{R}$, a positive integer m is called a **gap value** if $\mathfrak{R}$ carries no meromorphic function having a pole only at p with multiplicity m. Then if $g = 0$, no point has gap values; and if g $\geqslant$ 1, every point p has exactly g gap values; in this case, p also has a nongap value $m \leqslant g+1$. A point p is called an **ordinary point** if the gap values at p are $1,2,\,, g$; otherwise p is called a **Weierstrass point.** If y $\geqslant$ 2, then the total number N of Weierstrass points is not less than 2g + 2 and not greater than (g $-$ 1)$g(g+1)$ (A. Hurwitz). Moreover, the case N = $2g$ + 2 occurs if and only if $\mathfrak{R}$ is hyperelliptic, and then the gap values at Weierstrass points are 1, 3, . . , $2g\quad$ 1. This implies that every closed Riemann surface of genus 2 is hyperelliptic. (iv) Suppose the genus g of $\mathfrak{R}$ is $\geqslant$ 2. A conforma1 mapping f of $\mathfrak{R}$ onto itself with the property that every 1-cycle y is always homologous to f(y) is necessarily the identity transformation. Also, $\mathfrak{R}$ is known to admit only a finite number of conforma1 mappings onto itself (H. Schwarz); the total number does not exceed $84(g-1)$ (Hurwitz).

E. Abel's Tbeorem

Abel's tbeorem is stated as follows: A divisor d of degree zero is a principal divisor if and only if it is expressed as $d=\partial\gamma$ by means of a 1-chain y that has the property that $\int_\gamma \omega=0$ for every $\omega\in V_a$.

Given a divisor class D of degree zero, consider a 1-chain y with $\partial\gamma\in D$. For every l-cycle α there corresponds the quantity

$$\chi_\alpha(D)=\exp\left(2\pi\sqrt{-1}\ \mathrm{Re}\int_\gamma \omega_\alpha\right),$$

independent of the choice of y. Thus D determines a †character on the 1-dimensional homology group, called the **integral character.** Conversely, every character on the homology group is shown to be the integral character of some D. In terms of this notion, Abel's theorem can be stated as follows: D is the principal divisor class if and only if $\chi_\alpha(D)=1$ for every α. This result shows that the 1-dimensional homology group with integral coefficients and the group of the divisor classes of degree zero (with compact topology) are, with respect to integral characters, mutually dual (in the sense of Pontryagin) topological Abelian groups (→ 422 Topological Abelian Groups). For the relationship between Abelian integrals and Jacobian varieties, in particular the †Jacobi inverse problem, †Abelian functions, and †Riemann theta functions, → 3 Abelian Varieties L. Also → references to 234 Kleinian Groups, 352 Quasiconformal Mappings, 367 Riemann Surfaces, 416 Teichmüller Spaces.

References

[1] B. Riemann, Gesammelte mathematische Werke und wissenschaftlicher Nachlass, Teubner, 1876 (Collected works of B. Riemann, Dover, 1953).
[2] P. E. Appel1 and E. Goursat, Théorie des fonctions algébriques et de leurs intégrales, Gauthier-Villars, second edition, 1929.
[3] R. Dedekind and H. Weber, Theorie der algebraischen Funktionen einer Veränderlichen, J. Reine Angew. Math., 92 (1882), 181-290. (Gesammelte math. Werke, F. Vieweg, 1930, vol. 1, p. 2388350.)
[4] K. Hensel and G. Landsberg, Theorie der algebraischen Funktionen einer Variablen und ihre Anwendung auf algebraische Kurven und Abelsche lntegrale, Teubner, 1902 (Chelsea, 1965).
[5] W. F. Osgood, Lehrbuch der Funktionentheorie Il, Teubner, 1932.
[6] H. Weyl, Die ldee der Riemannschen Flache, Teubner, 1913, third edition, 1955; English translation, The concept of a Riemann surface, Addison-Wesley, 1964.
[7] M. Eichler, Einführung in die Theorie der algebraischen Zahlen und Funktionen, Birkhäuser, 1963.
[8] C. Chevalley, Introduction to the theory of algebraic functions of one variable, Amer. Math. Soc. Math. Surveys, 1951.
[9] C. Siegel, Topics in complex function theory Il, Wiley-Interscience, 1971.
For history and further references,
[10] A. Brill and M. Noether, Die Entwicklung der Theorie der algebraischen Funktionen, Jber. Deutsch. Math. Verein., 3 (1894), 107-566.
For the classical theory of algebraic functions of two variables,
[11] E. Picard and G. Simart, Théorie des fonctions algébriques de deux variables indépendentes, Gauthier-Villars, I, 1897; II, 1906.

12 (VIII.1) Algebraic Geometry

A. Introduction

Algebraic geometry is the branch of mathematics that deals with †algebraic varieties, that is, point sets defined by several algebraic equations in a space of any dimension or those derived from these sets by means of certain constructions (→ 16 Algebraic Varieties). It may also be considered to be a theory of the †field of algebraic functions in several variables in geometric language, and it is closely related to the theories of complex analytic manifolds, commutative algebra, and homological algebra. It also has an important connection with number theory through the theories of automorphic functions, Diophantine equations, and zeta functions.

To investigate local properties of algebraic varieties we consider varieties embedded in an †affine space; to study global properties we usually consider varieties contained in †projective spaces. A quantity (or property) that is invariant under †projective transformations, isomorphisms, i.e., †biregular and †birational transformations, or birational transformations is called a †projective invariant, a **relative invariant**, or an **absolute invariant (birational invariant)**, respectively. The study of projective invariants is a part of projective geometry, whose methods are important in algebraic geometry. The notions of relative invariant and absolute invariant are used, for example, in the classification of algebraic varieties.

We usually assume that the coordinates of each point of the variety belong to a certain fixed †field K. In the classical case, namely, when the field K is the field C of complex numbers, the algebraic varieties are considered as complex spaces and are studied by applying the theories of partial differential equations, †differential geometry, etc. Topological methods may also be applied. Algebraic geometry originated from such studies, but, for the study of properties such as †rational mappings or †algebraic systems, it became necessary to consider as well the case where the ground field K is not talgebraically closed. Furthermore, to apply this to number theory, it is necessary to establish the theory over the fïeld of any tcharacteristic p. For this purpose it is necessary to establish a theory for varieties having ground domains as general as possible.

B. History

Analytic geometry began with the study of lines and quadratic curves (surfaces) and later came to include the study of cubic and quartic curves (surfaces), and so on. These subjects originally belonged exclusively to analytic (or projective) geometry. At that time, the study could not have been described by so specific a title as algebraic geometry.

The study of such theories as the construction of an algebraic plane curve by families of curves of lower degree or the †algebraic *m-n* correspondence on a straight line probably began with research such as that by M. Chasles. The most outstanding event in the history of algebraic geometry was the introduction and development of the theory of algebraic functions (- 11 Algebraic Functions) by B. Riemann (1857). Before that time the degree of an algebraic curve (surface) was the only quantity known to be a projective invariant of the curve (surface).

With the theory of algebraic functions, Riemann gathered into one family all the curves that can be transformed onto each other by birational transformations. As the basis for his study, Riemann examined birational transformations in place of projective transformations. This idea led to the notion of the so-called †Riemann surface. The †genus of the surface was obtained as the characteristic number of the family of curves. The concept of genus was the first absolute invariant to appear in the history of algebraic geometry. Riemann based his theory on †Abelian integrals using †Dirichlet's principle, under the assumption that any algebraic curve reduces to one without tsingularities.

After Riemann many mathematicians tried to reconstruct the theory more precisely without using transcendental methods. M. Noether attempted this reconstruction by using geometric methods. Using the †Cremona transformation, he confirmed Riemann's assumption for curves: that any algebraic curve on a plane can be transformed by a birational transformation to a plane curve without singularities except for simple †nodes. He also contributed in making more precise the basis conditions for the †Riemann-Roch theorem, which is considered to be one of the most important theorems in the field. His results on space curves and surfaces are also noteworthy. J. Plücker detïned the concept of genus in geometric terms and introduced the †Plücker coordinates. A. Cayley and A. Brill worked along similar lines. Cayley's idea was developed later by B. L. van der Waerden and W. L. Chow, who introduced the †associated form of an algebraic variety and its †Chow coordinates.

Around 1890 the Italian school of algebraic geometry appeared. Following the tradition established by Noether, they employed algebrogeometric methods and uncovered many new facts concerning algebraic surfaces. Among those who belonged to this school were G. Castelnuovo, F. Enriques, and F. Severi.

In France, H. Poincaré and E. Picard initiated their study of algebraic functions of two complex variables. After them S. Lefschetz investigated the theory of complex algebraic surfaces [11, 123. The results attained by the Italian and French schools were very suggestive but lacked rigorous foundations.

On the other hand, rigorous number-theoretic theories of algebraic curves appeared in Germany. R. Dedekind and H. Weber developed the theory of algebraic function fields parallel to that of †algebraic number fields. K. Hensel introduced the concept of †p-adic numbers in analogy to †power series expansions of analytic functions. E. Noether constructed an abstract theory of †polynomial ideals from a forma1 theory by E. Lasker and F. S. Macauley. Under her influence there appeared the arithmetic algebraic geometry (of curves) over an abstract tïeld as developed by F. K. Schmidt and others.

In the higer-dimensional case, van der Waerden attempted to create a more rigorous foundation for algebraic geometry under the influence of Noether's abstract ideal theory (c. 1930) [14]. He introduced the concept of †generic points and tspecialization, and specifically defined the †multiplicity of intersections of two varieties in a projective space. He succeeded in getting a rigorous proof of **Bezout's theorem:** In n-dimensional projective

space, the number of intersections of an r-dimensional algebraic subvariety of degree l with an $(n-r)$-dimensional subvariety of degree m is always lm if they intersect in only a finite number of points.

The problem of intersections was taken up by C. Chevalley and A. Weil in the 1940s. Chevalley developed the ideal theory of †local rings (studied initially by W. Krull); he introduced topological concepts and applied them to the problem of intersections. The theory in this direction was later extended further by P. Samuel, M. Nagata, and J.-P. Serre.

Weil gave foundations of algebraic geometry over an abstract field and reconstructed the theory by introducing geometric language to designate objects of abstract algebra [15]. He thus gave quite a new aspect to the theory and extended H. Hasse's arithmetization of the theory of algebraic functions in one variable to the case of several variables. Reconstructing Severi's theory of algebraic correspondence over abstract fields, he succeeded in proving an analogy of the †Riemann hypothesis on †congruent zeta functions (→ 450 Zeta Functions N). He also constructed, purely algebraically, the entire theory of †Abelian varieties independent of characteristic.

Around 1930, O. Zariski gave another foundation to algebraic geometry by applying the generalized †valuation theory that had been introduced by Krull. Zariski clarified especially the properties of birational transformations by using valuation theory. Zariski's main theorem states that if a birational mapping is not †regular at a †normal point P (→ 16 Algebraic Varieties 1), each component of the image of P by the mapping is of dimension $\geqslant 1$.

Zariski also solved the problem of †resolution of singularities in the affirmative in the case of characteristic 0 for varieties of dimension $\leqslant 3$. The affirmative resolution of this problem (which Riemann assumed) says that any algebraic variety in a projective space can be transformed birationally to a projective algebraic variety without singularities. In 1964, H. Hironaka gave an affirmative answer for any dimension in the case of characteristic 0.

Along with the achievements in algebraic methods, great development took place in analytic methods. Unification of the concepts of Riemann surfaces and †Riemannian manifolds led to the concept of †complex analytic manifolds. Furthermore, G. de Rham's theorem on the duality of topologically defined homology and cohomology based on differential forms was proved; also, W. V. D. Hodge's theory of †harmonic integrals was developed. In the case of the complex dimension 1, any compact Riemann surface is derived from a certain projective algebraic curve. However, the situation is not so simple in the case of higher dimensions. Weil's concept of †abstract complete algebraic varieties can be considered as an analog of compact complex manifolds. If a compact complex analytic variety is projective, then it must be an algebraic variety (†Chow's theorem). K. Kodaira proved that a necessary and sufficient condition for a compact complex analytic manifold to be biholomorphically equivalent to a projective complex analytic manifold is that the manifold is a †Hodge variety.

Using results on harmonic integrals, J. Igusa and Weil established the theory of †Picard and †Albanese varieties associated with algebraic manifolds of arbitrary dimensions as a generalization of the theory of †Jacobian varieties associated with algebraic curves (→ 9 Algebraic Curves E). Thus many ambiguities in the theory as developed by the Italian school were clarified. Later the theory was generalized to the case of characteristic p by T. Matsusaka, Chow, and S. Lang. The duality theorem in this case was later proved by M. Nishi and P. Cartier (→ 3 Abelian Varieties D).

The concept of †sheaves (→ 383 Sheaves) had already been used in Kodaira's theory. Serre defined an abstract algebraic variety as a †ringed space by using an analogy to the concept of complex analytic spaces; he considered it as a topological space with respect to the †Zariski topology. By introducing †coherent algebraic sheaves, Serre clarified the idea that classical invariants (such as †arithmetic genus) may be considered cohomological quantities (→ 16 Algebraic Varieties E).

A. Grothendieck invented the concept of a †scheme, which is far more general than that of an algebraic variety, by admitting the existence of †nilpotent elements in structure sheaves and taking as a coordinate ring a general commutative ring with unity element. By the device of taking into account nilpotent elements, an analog of the method of successive approximation in analysis is now applicable. By masterful use of cohomological techniques, Grothendieck derived many results, including Zariski's important theorems.

In the classical case, F. Hirzebruch generalized the Riemann-Roch theorem to higher-dimensional manifolds. He made use of the language of sheaves and some topological results of A. Borel and R. Thom [S]. Later Grothendieck generalized the theorem for the abstract case as well. His idea in this work is recognized as the origin of †K-theory.

Every nonsingular complete curve of genus 0 is isomorphic to the projective line, and any

nonsingular complete curve of genus 1 is isomorphic to a projective curve delïned by the equation $X_2^2 X_0 = X_1(X_1$ -X,)(X, $-\lambda X_0)$ for some $\lambda \neq 0$, 1. On the other hand, the set of all isomorphism classes of nonsingular complete curves of genus $g > 1$ is parametrized by a normal quasiprojective variety of dimension 3g - 3. Such facts were first discussed by Riemann as the problem of moduli.

Concerning the moduli of manifolds of higher dimensions in the classical case, Kodaira and D. C. Spencer developed their theory of tdeformations of complex structures (→ 72 Complex Manifolds G). The meaning of number of †moduli is clarified by deformation theory. Deformation theory has been extended in various ways, and deformation is considered as one of the fundamental concepts in algebraic geometry.

TO investigate the global structure of the moduli varieties, D. Mumford introduced geometric invariant theory (- 16 Algebraic Varieties W).

Etale and crystalline cohomologies initiated by Grothendieck and others are useful for the study of algebraic varieties of positive characteristic. In particular, the conjecture made by Weil concerning †congruent zeta functions has been solved affirmatively by P. Deligne with the help of étale cohomology [2].

Many important questions have been answered by means of the geometric theory. For example, (1) every tvector bundle on $\mathbf{A}_k^n$ is trivial (- 16 Algebraic Varieties Z); (2) there exist †unirational but nonrational fields over C (- 16 Algebraic Varieties J); (3) the fundamental groups of the complements of node-curves on $\mathbf{P}_C^2$ are commutative (- 16 Algebraic Varieties I); (4) the cancellation theorem holds for $\mathbf{A}_k^2$ (→ 15 Algebraic Surfaces H); (5) whenever the tangent vector bundles are ample, the varieties are $\mathbf{P}_k^n$ (→ 16 Algebraic Varieties R).

References

[1] G. Castelnuovo, Sulle funzioni Abeliane, Memorie scelte, Zanichelli, 1937, 5299549.
[2] P. Deligne, La conjecture de Weil 1, Publ. Math. Inst. HES, 43 (1974), 2733307.
[3] J. Dieudonné, Cours de Géométrie Algébrique, I. Aperçu Historique sur le Développement de la Géométrie Algébrique, Presses Univ. France, Collection Sup., 1974.
[4] F. Enriques, Memorie scelte di geometria, Zanichelli, I, 1956; II, 1959; III, 1966.
[5] A. Grothendieck, The cohomology theory of abstract algebraic varieties, Proc. Intern. Congr. Math., Edinburgh, 1958, Cambridge Univ. Press, 1958, 103-l 18.
[6] A. Grothendieck, Eléments de géométrie algébrique, Publ. Math. Inst. HES, 4 (1960); 8 (1961); 11 (1961); 17 (1963); 20 (1964); 24 (1965); 28 (1966); 32 (1967).
[7] H. Hironaka, Resolution of singularities of an algebraic variety over a field of characteristic zero, Ann. Math., (2) 79 (1964), 1099326.
[8] F. Hirzebruch, Topological methods in algebraic geometry, Erg. Math., Springer, third edition, 1966.
[9] K. Kodaira, Collected Works I, II, III, Iwanami, 1975.
[10] S. Lefschetz, L'analysis situs et la géométrie algébrique, Gauthier-Villars, 1924.
[11] E. Picard and G. Simart, Théorie des fonctions algébriques de deux variables indépendantes, Gauthier-Villars, I, 1897; II, 1906.
[12] D. Quillen, Projective modules over polynomial rings, Inventiones Math., 36 (1976), 1677171.
[13] J. G. Semple and L. Roth, Introduction to algebraic geometry, Clarendon Press, 1949.
[14] B. L. van der Waerden, Einführung in die algebraischen Geometrie, Springer, 1939.
[15] A. Weil, Foundations of algebraic geometry, Amer. Math. Soc. Colloq. Publ., 1946, 1962.
[16] A. Weil, Oeuvres Scientifiques, Collected Papers I, II, III. Springer, 1980.
[17] 0. Zariski, Collected papers, MIT Press, I, 1972; II, 1973; III, 1978; IV, 1979.

13 (IV.1 1) Algebraic Groups

A. Definitions and General Remarks [1,2,8]

Let k be a field and Ω a tuniversal domain containing it. An **affine algebraic group G** defined over k is, by definition, a group G which has the structure of a (not necessarily irreducible) †algebraic variety defined over k in an affine space Ω^N such that the group operation $(x, y) \mapsto x^{-1} y$ on G is an everywhere †regular trational mapping defined over k. For such a group G, the set G_k of all k-trational points on G is an abstract group. The †irreducible component G_0 of G (viewed as an algebraic set) containing the identity element e is unique and is a normal subgroup defined over k with finite index in G; the decomposition of G into (absolute) irreducible components coincides with that into the cosets of G by G,. When G = G_0, the group G is called **connected**. It should be noted that for an algebraic group G defined over k, a coset gG_0 is not necessarily delïned over k; and if it is, a representative g cannot necessarily be taken to be k-rational.

However, if k is an infinite field and G is connected, and if moreover k is †perfect or G is †reductive, then G_k is †Zariski dense in G ([9], A. Borel and T. A. Springer [30]); hence G is uniquely determined by G_k. (In this case, G_k is sometimes referred to as a **k-group.)** When k is a †topological field, G_k is a †topological group with respect to the natural topology defined by that of k, which is generally stronger than the Zariski topology on G_k. For instance, when $k = \mathbf{R}$, $(G_0)_{\mathbf{R}}$ is a †Lie group with finitely many connected components.

A subgroup H of an affine algebraic group G, which is Zariski closed, is an affine algebraic group with respect to its natural induced structure and is called an **algebraic subgroup** of G. If H is defined over k, then H is k-closed; the converse is also true when k is †perfect. (An affine algebraic set A is called **k-closed** if A is a common zero of a set of polynomial equations with coefficients in k. The set A is k-closed if and only if all irreducible components of A are defined over the algebraic closure $\bar{k}$ of k, and for every †Galois automorphism σ of $\bar{k}/k$, $A^\sigma = A$.) The notions of homomorphism and isomorphism for algebraic groups can be defined in a natural manner. For instance, for affine algebraic groups G and G' defined over k, a **rational homomorphism** $\varphi: G \rightarrow G'$ defined over k, **or a k-morphism** for short, is a homomorphism of G into G' that is at the same time an (everywhere regular) rational mapping defined over k. For a k-morphism φ of G into G', the (set-theoretic) image $\varphi(G)$ is a closed subgroup of G' defined over k, the kernel $\varphi^{-1}(e')$ is a k-closed subgroup of G, and $\dim \varphi(G) = \dim G - \dim \varphi^{-1}(e')$. In particular, when $\dim G = \dim \varphi(G) = \dim G'$ (or equivalently, when $\varphi(G_0) = G_0'$ and $\varphi^{-1}(e')$ is finite), φ is called an **isogeny.** (Two groups G and G' are called **isogenous** if there exist a third group G" and isogenies $G'' \rightarrow G$, $G'' \rightarrow G'$.) When a k-morphism φ is bijective and φ^{-1} is also an (everywhere regular) rational mapping defined over k, φ is called a **birational isomorphism** defined over k, or a **k-isomorphism** for short. It should be noted that a rational homomorphism which is an isomorphism of abstract groups is not necessarily an isomorphism of algebraic groups (e.g., a Frobenius homomorphism); a similar statement holds for an injective continuous homomorphism of topological groups.

Given a connected affine algebraic group G and a closed subgroup H, both defined over k, the quotient space G/H has the uniquely determined structure of an algebraic variety defined over k such that the canonical mapping $G \rightarrow G/H$ is separable. The †function field of G/H is then identified with the subfield of the function field of G formed by all H-invariant elements. In particular, if H is a closed normal subgroup, then G/H has a natural structure of an affine algebraic group (defined over the same ground field k) [8, 10, 11].

As an example, we have the group $GL(n)$ of all $n \times n$ nonsingular matrices (x_{ij}). $GL(n)$ may be viewed as an algebraic set in Ω^{n^2+1}, defined by a single equation $\det(x_{ij})\, y = 1$, and as such is a connected algebraic group defined over the †prime field. In general, an algebraic group realized as a closed subgroup of $GL(n)$ is called a **linear algebraic group.** Since an affine algebraic group is always isomorphic to a linear algebraic group, these two terminologies are essentially synonymous [1, 2].

B. Generalization of the Definition

Replacing the term *affine algebraic set* (or *affine variety)* in the definition of an affine algebraic group by a more general term †*algebraic variety*, we obtain the notion of an **algebraic group (algebraic group variety,** or simply **group variety).** On this subject, the following facts are fundamental. A †complete connected algebraic group is an Abelian variety (→ 3 Abelian Varieties). More generally, given a connected algebraic group G defined over k, there always exists a (k-closed) largest linear connected closed normal subgroup L, and the factor group G/L is an Abelian variety. Furthermore, for a closed normal subgroup H of G, the factor group G/H is complete if and only if $H \supset L$ (**Chevalley's theorem** [8]). In particular, if a connected algebraic group G is complete and linear at the same time, then G reduces to the identity group. In view of these theorems, the study of algebraic groups can be reduced, in a sense, to the study of Abelian varieties and linear algebraic groups. For this reason, we henceforth restrict ourselves to linear algebraic groups, which are simply called algebraic groups. (The notion of †generalized Jacobian variety, introduced by M. Rosenlicht [12], is an example of an algebraic group in a general sense; → 9 Algebraic Curves.) The notion of algebraic groups has been generalized further to that of †group schemes by A. Grothendieck [3, 4].

C. Lie Algebras

Since an algebraic group G defined over k has no singularities, the †tangent space $\mathfrak{g}$ to G at the identity element e is defined and has the same dimension as G : $\dim \mathfrak{g} = \dim G$. The space $\mathfrak{g}$ can be identified in a natural manner with the space of all left-invariant †derivations of the function field of G_0 and thus has the struc-

ture of a Lie algebra defined over k (→ 248 Lie Algebras). We call $\mathfrak{g}$ (the Lie algebra $\mathfrak{g}_k$ over k of all k-rational points in $\mathfrak{g}$) the **Lie algebra** of G (of the k-group G_k). If G is a linear algebraic group contained in $GL(n)$, then $\mathfrak{g}_k$ is a Lie subalgebra of $\mathfrak{gl}(n, k)$ with the Lie product defined by $[x, y] = xy - yx$; a linear Lie algebra corresponding to a linear algebraic group is called an **algebraic Lie algebra.** When the characteristic of k is zero, conditions for a linear Lie algebra to be algebraic can be given in terms of the **replica** [5]. Also, in the case of characteristic zero, for $x \in \mathfrak{gl}(n, k)$, $x \in \mathfrak{g}_k$ if and only if $\exp(tx) \in G$, where t is a variable over k and $\exp(tx)$ is understood as a †formal power series in t (contained in Ω). From this, we can prove, exactly as in the theory of Lie groups (→ 249 Lie Croups), a one-to-one correspondence between k-closed subgroups H of G and algebraic Lie subalgebras $\mathfrak{h}_k$ of $\mathfrak{g}_k$, establishing a complete parallelism of the theories of algebraic groups and Lie algebras [5]. This parallelism breaks down when k has positive characteristic [1, 2]. On the other hand, over a field of characteristic $p > 0$, **we** have **formal groups,** an analog of local Lie groups introduced by J. Dieudonné [13], and also †hyperalgebras, which play the role of Lie algebras in characteristic 0.

D. Tori [1,2]

The group $\mathbf{G}_m = GL(1)$, the multiplicative group of nonzero elements in Ω, is a 1-dimensional connected algebraic group defined over the prime field. In general, an algebraic group G that is isomorphic to the direct product $(\mathbf{G}_m)^n$ is called an **(algebraic) torus.** When a torus G defined over k is isomorphic to $(\mathbf{G}_m)$ over an extension K of k, G is called **K-trivial** (or **K-split),** and the field K is called a **splitting field** for G. A torus G defined over k always has a splitting field K which is a finite separable extension of k.

In general, a rational homomorphism χ of an algebraic group G into $\mathbf{G}_m$ is called a **character** of G. If we define the sum of two characters χ_1 and χ_2 of G by $(\chi_1+\chi_2)(g)=\chi_1(g)\cdot\chi_2(g)$ ($g \in$ G), the totality of characters of G is an additive group, called the **character module** of G and denoted by X(G). Let G be a torus defined over k and X =X(G) its character module, and let K be a splitting field for G that is a finite †Galois extension of k. If a K-isomorphism $G \cong (\mathbf{G}_m)^n$ is given by the correspondence $G \ni g \rightarrow (\chi_1(g), \ldots, \chi_n(g))$, then the χ_i are characters of G, and X is a †free module of rank n generated by $\chi_1, \ldots, \chi_n$. Furthermore, if Γ denotes the Galois group of K/k, then for $\sigma \in \Gamma$ and $\chi \in X$, the conjugate χ^σ is also a character of G; under this action of Γ, X becomes a right f-module. We have complete duahty between a torus G and its character module X in the following sense. There exists a one-to-one correspondence between the closed subgroups G_1 (defined over k) of G and a (Γ-invariant) submodule X_1 of X for which X/X_1 has no p-†torsion (where p is the characteristic of k). This correspondence is determined by the relation of the annihilators $X_1 = G_1^{\perp}$, $G_1 = X_1^{\perp}$, and under the correspondence the character modules of G_1 and of G/G_1 are canonically identified with X/X_1 and X_1, respectively. Furthermore, let G' be another torus (defined over k and split over K) with the character module X', and suppose that we have a (k-) homomorphism $\varphi: G \rightarrow G'$. Then we can define a ($\Gamma$-) homomorphism ${}^t\varphi: X' \rightarrow X$, called the **dual homomorphism** of φ, by the relation ${}^t\varphi(\chi') = \chi' \circ \varphi$ for $\chi' \in$ X'; conversely, any (Γ-) homomorphism of X' into X is obtained uniquely in this manner. In particular, φ is a (k-) isomorphism if and only if its dual ${}^t\varphi$ is a (Γ-) isomorphism. Since for any free (Γ-) module X of finite rank there always exists a torus G (defined over k and split over K) such that X(G) ≅ X (as a special case of the existence theorem of k-forms), the †categories of all tori (defined over k and split over K) and that of all free (Γ-) modules of finite rank are mutually dual.

E. Semisimple Elements and Unipotent Elements

A matrix a is called †semisimple if it is diagonalizable, i.e., if the †minimal polynomial of a has only simple roots. A matrix a is called †unipotent if $a-1$ is nilpotent, i.e., if all characteristic roots of a are equal to 1. (When the characteristic of the ground field is zero, the unipotent elements u in $GL(n, k)$ and the nilpotent elements x in $\mathfrak{gl}(n, k)$ are in one-to-one correspondence by the relation $u = \exp x$.) Any nonsingular matrix a can be written uniquely as a product of a nonsingular semisimple matrix a' and a unipotent matrix a" of the same size which are mutually commutative: a = a'a" = $a''a'$ (†multiplicative Jordan decomposition); a' (a") is called the **semisimple (unipotent) part** of a and is denoted by a_s (a_u); a_s can be expressed as a polynomial of the matrix a with scalar coefficients. For an element a of a (linear) algebraic group G, the semisimplicity (unipotency) of a does not depend on the matrix representation of G. Moreover, these properties are preserved by homomorphisms of algebraic groups. Also, if $a \in G$, then a_s, $a_u \in$ G.

For an algebraic group G, we denote the

totality of semisimple (unipotent) elements contained in G by G_s (G_u) and call it the **semisimple (unipotent) part** of G. (Note that G_s and G_u are not necessarily subgroups.) A torus G is then characterized by the property that $G = G_0 = G_s$. On the other hand, an algebraic group G such that $G = G_u$ is called **unipotent.** For instance, the additive group of the universal domain,

$$\mathbf{G}_a \cong \left\{ \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \middle| x \in \Omega \right\},$$

is a 1-dimensional connected unipotent algebraic group.

F. Solvable Groups and Nilpotent Groups [1, 2, 7, 9]

For two closed normal subgroups H_1, H_2 (defined over k) of an algebraic group G, the †commutator group $[H_1, H_2]$ (in the sense of abstract group theory) is also a closed normal subgroup (defined over k) of G. In view of this fact, an algebraic group G is called **solvable (nilpotent),** when it is †solvable (†nilpotent) as an abstract group. For example, the totality T(n) of $n \times n$ nonsingular *upper unipotent* matrices, i.e., matrices of the form $\begin{pmatrix} * & \cdots & * \\ & \ddots & \vdots \\ 0 & & * \end{pmatrix}$, is a connected solvable algebraic group. A unipotent algebraic group is always nilpotent.

For any connected solvable algebraic group $G \subset GL(n)$, there always exists an element a in $GL(n)$ such that $a^{-1} Ga \subset T(n)$ **(Lie-Kolchin theorem** [1, 2]). A connected solvable algebraic group G has a †composition series $G = G_0 \supset G_1 \supset \cdots \supset G_r = \{e\}$ such that each G_i is a connected closed normal subgroup of G and G_{i-1}/G_i is isomorphic to either $\mathbf{G}_m$ or $\mathbf{G}_a$. If G is defined over k, the subgroup G_u is a connected k-closed normal subgroup of G, and for any maximal torus T in G, we have a decomposition into a †semidirect product $G = T \cdot G_u$ (in the sense of algebraic groups, i.e., the natural map $T \times G_u \to G$ is birational). It is known that for any algebraic group G defined over k there exists a maximal torus defined over k ([4], A. Borel and T. A. Springer [30]). G is nilpotent if and only if G has a unique maximal torus T; when that is so, $T = G_s$ and T is contained in the †center of G. For a connected solvable algebraic group G defined over k, we can take $a \in GL(n, k)$ such that $a^{-1} Ga \subset T(n)$ (see the Lie-Kolchin theorem) if and only if all characters $\chi \in X(G)$ are defined over k; when this condition is satisfied, G is called **k-solvable.** G_u is then defined over k, and G/G_u is a k-trivial torus.

When the characteristic of k is zero, any commutative unipotent algebraic group (defined over k) is (k-) isomorphic to the direct product $(\mathbf{G}_a)^n$. When k is an algebraically closed field of characteristic $p > 0$, any connected commutative unipotent algebraic group defined over k is k-isogenous to a direct product of a certain number of the groups $\mathbf{W}_m$ of †Witt vectors (of length m) (Chevalley-Chow theorem [12]). A 1-dimensional connected unipotent algebraic group defined over a perfect field k is k-isomorphic to $\mathbf{G}_a$ [1, 2].

G. Borel's Theory

Let G be an algebraic group and V an algebraic variety (both defined over k). We say that V is a **transformation space** of G (defined over k), or simply G acts on V, if there is given an everywhere regular rational mapping $G \times V \ni (g, v) \to gv \in V$ (defined over k) such that $g_1(g_2 v) = (g_1 g_2)v$, $ev = v$ ($g_1, g_2 \in G, v \in V$). When the action of G on V is †transitive, V is called a †homogeneous space of G. For a closed subgroup H of a connected algebraic group G (both defined over k), the quotient space G/H has the natural structure of a homogeneous space of G (defined over k). A. Borel [1, 2] proved the following theorems:

(1) If G is a connected solvable algebraic group and V a complete transformation space of G, then G has at least one fixed point in V. More precisely, in order that a connected algebraic group G defined over k be k-solvable, it is necessary and sufficient that for any complete transformation space V of G defined over k for which $V_k \neq \varnothing$, G have at least one k-rational fixed point in V [9].

(2) Let G be a connected algebraic group. A maximal connected solvable closed subgroup of **G** is called a **Borel subgroup** of G. Then (i) all pairs (T, B) formed by a maximal torus T in G and a Borel subgroup B containing it are conjugate to each other with respect to inner automorphisms of G. (ii) For a closed subgroup H of G, the quotient space G/H is complete if and only if H contains a Borel subgroup of G; and, when that is so, G/H is actually a †projective algebraic variety. (For instance, if $G = GL(n)$, $B = T(n)$, then G/B is a so-called †flag manifold.) (iii) The conjugates of B(T) cover the whole group G (G_s). A closed subgroup of G is called **parabolic** if it contains a Borel subgroup of G. A parabolic subgroup H coincides with its own †normalizer $N(H)$; in particular, H is always connected. Parabolic subgroups are significant in the theory of automorphic functions. (For the parabolic subgroups associated with BN-pairs → Sections Q, R.)

When G is a connected algebraic group defined over a perfect field k, proposition (i) can be sharpened: The pairs (A, H) formed by a maximal k-trivial torus A in G and a maximal connected k-solvable subgroup H containing it are conjugate to each other with respect to the inner automorphisms defined by elements in G_k. The normalizer $N(H)$ of such a k-solvable subgroup H is a minimal k-closed parabolic subgroup of G. When the maximal connected k-solvable subgroups of G are reduced to the identity group, G is called **k-compact or k-anisotropic.** (Otherwise, G is called **k-isotropic.)** For instance, the †orthogonal group $G = SO(n, f)$ of a †quadratic form f of n variables is k-compact if and only if the form f is **anisotropic,** i.e., the homogeneous equation $f = 0$ has no solution other than zero in k. Similar facts hold for other classical groups. When k is a †local field, G is k-compact if and only if G_k is compact as a topological group. In general, a k-compact group is †reductive.

H. The Weyl Group

Let G be a connected algebraic group and Q an arbitrary torus in G. The †centralizer Z(Q) of Q is then connected and coincides with the connected component of the normalizer N(Q). Hence the factor group $W = N(Q)/Z(Q)$ is finite and can be identified with a subgroup of the automorphism group of Q (or of its character module X(Q)) in a natural manner. The group W is called the **Weyl group** of G relative to Q. In particular, when $Q = T$ (a maximal torus), the order of W is equal to the number of Borel subgroups containing T. In this case, the centralizer $C = Z(T)$ is called a **Cartan subgroup** of G; it is characterized by the property that C is a (maximal) connected nilpotent closed subgroup of G which coincides with the connected component of its own normalizer N(C). The notions of Borel subgroups, Cartan subgroups, and maximal tori are preserved under rational homomorphisms of algebraic groups.

1. Semisimple Groups and Reductive Groups

In an algebraic group G defined over k, there exists a largest connected solvable closed normal subgroup R, called the **radical** of G. The unipotent part R_u of R is called the **unipotent radical** of G. When $R = \{e\}$, G is called **semisimple.** When R is a torus, namely, $R_u = \{e\}$, G is called **reductive.** Semisimplicity and reductiveness are preserved under forming a direct product and taking the image (or inverse image) of an isogeny. For a reductive group G, the †commutator subgroup D(G) is semisimple, and $G = D(G) \cdot R$, $D(G) \cap R =$ finite; in other words, G is isogenous to the direct product of a connected semisimple algebraic group and a torus. In general, if R is the radical of a connected algebraic group G and R_u is the unipotent radical of R, then the factor groups G/R, G/R_u are semisimple and reductive, respectively. Furthermore, if the characteristic of the field k is zero, there exists a reductive closed subgroup H of G such that G decomposes into a semidirect product $G = H \cdot R_u$ **(Chevalley decomposition** [5]). (In this case, R and R_u are k-closed, and H can be taken to be k-rational; such an H is unique up to inner automorphisms defined by elements in G_k.) Also in the case of characteristic zero, reductive algebraic groups are characterized by the property that all rational representations are completely reducible. But when k has the characteristic $p > 0$, this property characterizes tori (M. Nagata).

J. Root Systems [1, 2, 14]

Let G be a connected semisimple algebraic group, T a maximal torus, and $X = X(T)$ its character module. A character $\alpha \in X$ is called a **root** of G relative to T if there exists an isomorphism x_α of $\mathbf{G}_a$ onto its image in G such that

$$t^{-1}x_\alpha(\xi)t = x_\alpha(\alpha(t)\xi) \quad \text{for all } \xi \in \mathbf{G}_a, \ t \in T.$$

For a root α, such an isomorphism x_α is uniquely determined up to a scalar multiplication in $\mathbf{G}_a$; hence we put $P_\alpha = x_\alpha(\mathbf{G}_a)$.

If we denote by $\mathfrak{r}$ the totality of roots (relative to T), $\mathfrak{r}$ satisfies the following axioms, where $E = X \otimes Q$ and E^* is the †dual space of E with respect to the inner product $\langle\ \rangle$: (i) For each $\alpha \in \mathfrak{r}$, there corresponds $\alpha^* \in E^*$ such that $\langle \alpha^*, \alpha \rangle = 2$ and $\langle \alpha^*, \beta \rangle \in \mathbf{Z}$ for all $\beta \in \mathfrak{r}$. (ii) If we define a reflection w_α of E by

$$w_\alpha x = x - \langle \alpha^*, x \rangle \alpha \quad \text{for} \quad x \in E,$$

then $w_\alpha \beta \in \mathfrak{r}$ for all $\beta \in \mathfrak{r}$. (In particular, $w_\alpha \alpha = -\alpha \in \mathfrak{r}$.) (iii) If a, $\beta \in \mathfrak{r}$ are linearly dependent, then $\beta = \pm\alpha$. (iv) If $\dim E = r$, $\mathfrak{r}$ contains r linearly independent elements.

In general, a finite subset $\mathfrak{r}$ in a finite-dimensional vector space E over Q satisfying the axioms (i)–(iv) is called a **root system** in E. (This root system is sometimes said to be reduced, to distinguish it from the root system defined in Section Q, which does not satisfy axiom (iii).) For a root system $\mathfrak{r}$, the elements α^* of E^* corresponding to $\alpha \in \mathfrak{r}$ are uniquely determined by these conditions, and the set $\mathfrak{r}^* = \{\alpha^*\}$ is a root system in E^* ($a^* \in \mathfrak{r}^*$ is called a **coroot**). Also, the group W of linear transformations of E (E^*) generated by w_α

(w_{α^*}) with $\alpha \in r$ is finite and is called the **Weyl group** of the root system r. If we identify E^* with E by means of any W-invariant (positive definite) metric on E, then a* $=(2/\langle\alpha, \alpha\rangle)\alpha$. When r is a root system of a semisimple algebraic group,

$$\langle\alpha^*, \chi\rangle \in \mathbf{Z} \quad \text{for all} \quad \alpha \in r, \quad \chi \in X, \qquad (1)$$

so that X is W-invariant, and the Weyl group of the root system r can be identified with the group $N(T)/Z(T)$ of Section H. (In general, a maximal torus in a connected reductive algebraic group coincides with its own centralizer, so that the Weyl group W can be identified with N(T)/T.)

When a †linear ordering (compatible with the addition) is given in E, we denote by r_+ the set of all positive roots in r. An element $\alpha \in r_+$ is called a **simple root** if it cannot be written as $\alpha = \alpha' + \alpha''$ with $\alpha', \alpha'' \in r_+$. If $\Delta = \{\alpha_1, \ldots, \alpha_r\}$ is the totality of (distinct) simple roots in r_+, the elements α_1, , α_r are linearly independent, and any root $\alpha \in r$ can be written uniquely in the form a = $\pm \sum_{i=1}^{r} m_i \alpha_i$, with $m_i \in \mathbf{Z}$, $m_i \geqslant 0$. In general, a subset A of r having this property is called a **fundamental system;** a fundamental system is always obtained in the manner explained from a linear ordering on E. For a fundamental system A, the cone A, in E^*, defined as the set of x in E^* satisfying the inequalities $\langle\alpha_i$, x) > 0 ($1 \leqslant i \leqslant r$), is called a **Weyl chamber**. If we denote by L_α the hyperplane defined by the linear equation $\langle\alpha$, x) = 0 for a root α, then $E^* - \bigcup_{\alpha \in r} L_\alpha = \bigcup_\Delta \Lambda_\Delta$, and W acts †simply transitively on the set of all Weyl chambers $\{\Lambda_\Delta\}$. The Weyl group W is generated by r reflections w_{α_i} ($1 \leqslant i \leqslant r$).

In a semisimple algebraic group G, Borel subgroups B containing a (fixed) maximal torus Tare in one-to-one correspondence with the fundamental systems A (or r+) relative to T by the relation $B_u = \prod_{\alpha \in r_+} P_\alpha$, where $P_\alpha = x_\alpha(\mathbf{G}_a)$. (More precisely, every element in B_u can be written uniquely as a product of the elements in P_α, where the ordering of the P_α is taken arbitrarily.)

K. Bruhat Decomposition

If we take a representative s_w of $w \in W$ in N(T), there is a decomposition $G = \bigcup_{w \in W} Bs_w B$ (disjoint union). Furthermore, if for w $\in W$ we put $N_w = \prod_{\alpha \in r_+ \cap wr_+} P_\alpha$, and in particular N = $N_e = B_u$, and denote by w_0 the unique element in W such that $w_0 \Delta =$ -A, then the element in $Bs_w B$ can be written uniquely as a product of elements in N_{ww_0}, $s_w T$, N. Hence we have

$$\mathrm{G} = \bigcup_{w \in W} N_{ww_0} \cdot s_w T \cdot N,$$

which is called a **Bruhat decomposition** of G. In particular, if we put N' = $s_{w_0}{}^{-1} N s_{w_0}$ (which is the unipotent part of the Borel subgroup corresponding to -A), then $N'TN$ is a Zariski open set in G, and the natural map N' x T x $N \to G$ is birational. This implies that the function field of G is rational (*i.e.*, †purely transcendental over Ω).

L. Structure of Semisimple Groups

A subset r_1 of a root system r is called a **closed subsystem** if $r_{1\mathbf{Z}} \cap r = r_1$, where $r_{1\mathbf{Z}}$ denotes the submodule of X generated by r_1 A closed subsystem satisfies conditions (i), (ii), and (iii) of a root system. For a closed subsystem r_1 of a root system r of a semisimple algebraic group G, the subgroup $G(r_1)$ of G generated by the P_α ($\alpha \in r_1$) is a semisimple closed subgroup with a maximal torus $T_1 = (G(r_1) \cap T)_0$, of which the root system relative to T_1 coincides with the restriction of r_1 on T_1 and the coroot system can be identified with $r_1^* = \{\alpha^* \mid \alpha \in r_1\}$. The subgroup $G(r_1)$ is normal if and only if $r - r_1$ is also a closed subsystem; when this is so, $G = G(r_1) \cdot G(r - r_1), G(r_1) \cap G(r - r_1) =$ finite. All connected closed normal subgroups of G are obtained in this manner. In order that G be **simple** (sometimes called **ahsolutely simple** or **almost simple)** as an algebraic group (i.e., without proper connected normal subgroups), it is necessary and sufftcient that r be **irreducible** (i.e., r cannot be decomposed into a disjoint union of two proper closed subsystems). In general, a root system r can be decomposed uniquely into the disjoint union r = r_1 **U U** r_s of irreducible closed subsystems r_i such that r_1 **U U** r_i ($1 \leqslant i \leqslant$ s) are also closed subsystems; correspondingly, G is isogenous to the direct product G_1 x x G_s of (absolutely) simple algebraic groups $G_i = G(r_i)$. (G is actually a direct product if it is simply connected or an adjoint group.) The subgroups G_i are determined uniquely and only by G.

M. k-Forms [15]

Let K be an extension of k and G_1 an algebraic group defined over K. An algebraic group G defmed over k is called a **k-form** of G_1 if there is a K-isomorphism f of G onto G_1. Suppose further that K/k is finite separable, and for every Galois automorphism σ of k/k, put $\varphi_\sigma = f^\sigma \circ f^{-1}$. Then φ_σ is an isomorphism of G_1 onto G_1^σ, and the φ_σ satisfy the relation φ_σ^τ o $\varphi_\tau = \varphi_{\sigma\tau}$. Conversely, given a collection of isomorphisms $\{\varphi_\sigma\}$ satisfying these conditions, there always exists a k-form G (with a K-isomorphism f onto G_1 such that $\varphi_\sigma = f^\sigma \circ f^{-1}$), which is unique up to k-isomorphism (Weil). In particular, if K/k is a

finite Galois extension with Galois group Γ, then $\{\varphi_\sigma\}$ is a (continuous) †1-cocycle of Γ in $\mathrm{Aut}_K(G_1)$ (the group of all K-automorphisms of G_1), and by the above correspondence the k-isomorphism classes of k-forms G are in one-to-one correspondence with the (continuous) 1-†cohomology classes of the cocycle $\{\varphi_\sigma\}$ (in the cohomology set $H^1(\Gamma, \mathrm{Aut}_K(G_1))$) (→ 172 Galois Theory J)

To a given finite separable extension K/k of degree d and an algebraic group G_1 defined over K of dimension n, we can associate a certain algebraic group $\mathfrak{R}_{K/k}(G_1)$ defined over k of dimension dn, which is obtained from G_1 by restricting the ground field [19]. A more precise definition is as follows. Let $\{\sigma_1, \sigma_2, \ldots, \sigma_d\}$ ($\sigma_1 = 1$) be a set of automorphisms of $\bar{k}/k$ such that σ_i K ($1 \leqslant i \leqslant d$) are all distinct. Then one can find a k-form $\tilde{G}$ of $\tilde{G}_1 = \prod_{i=1}^d G_1^{\sigma_i}$ with an isomorphism $\tilde{f}: \tilde{G} \rightarrow \tilde{G}_1$ such that $\tilde{\varphi}_\sigma = \tilde{f}^\sigma \circ \tilde{f}^{-1}$ is given by $\tilde{\varphi}_\sigma((x_i)) = (x_{i^\sigma})$, where i^σ is defined by the relation $(\sigma_i\sigma)$ $K = \sigma_{i^\sigma}$ K. If we denote by p_1 the canonical projection of $\tilde{G}_1$ onto its first component G_1 and put $p = p_1 \circ \tilde{f}$, then the pair ($\tilde{G}, p$) is uniquely characterized (up to k-isomorphism) by the following universality property: If $\tilde{G}'$ is any algebraic group defined over k and φ is a K-morphism of $\tilde{G}'$ into G_1, then there exists a (uniquely determined) k-morphism $\tilde{\varphi}$ of $\tilde{G}'$ into $\tilde{G}$ such that $\varphi = p \circ \tilde{\varphi}$. The group $\tilde{G}$ (together with p) is denoted by $\mathfrak{R}_{K/k}(G_1)$. For the group of rational points, $\tilde{G}_k = G_{1K}$. When the algebraic group G_1 has some additional structure (such as that of †vector space, †algebra, etc.), then $\mathfrak{R}_{K/k}(G_1)$ automatically has the same kind of additional structure.

N. Chevalley's Fundamental Theorems

Let G, G' be connected semisimple algebraic groups, and let T (T') be a maximal torus in G (G'), X(X') its character module, $\mathfrak{r}$ ($\mathfrak{r}'$) a root system of G (G') relative to T (T'), etc. If we have an †isogeny φ of G onto G' such that $\varphi(T) = T'$, then there is a bijection $\alpha \rightarrow \alpha'$ of $\mathfrak{r}$ onto $\mathfrak{r}'$ such that $\psi(\alpha') = q_\alpha \alpha$, where ψ is the dual homomorphism of φ T and q_α is a positive integer, which equals 1 if the characteristic is zero and is a power of p if the characteristic is $p > 0$. Conversely, any injective homomorphism $\psi: X' \rightarrow X$ satisfying this condition (with respect to a certain bijection $\mathfrak{r} \rightarrow \mathfrak{r}'$ and q_α) comes from an isogeny $\varphi: G \rightarrow G'$ in the manner already stated. In particular, φ is an isomorphism if and only if ψ is an isomorphism such that $\psi(\mathfrak{r}') = \mathfrak{r}$ (i.e., $q_\alpha = 1$ for all $\alpha \in \mathfrak{r}$) [7]. The isomorphism class of G is thus completely determined by the pair (X, $\mathfrak{r}$), so that we sometimes write G = $G(X, \mathfrak{r})$. A connected semisimple algebraic group G defined over k is called **of Chevalley type** over k (or **k-split**) if there exists a k-trivial maximal torus T in G. If, in the above theorem, G and G' are of Chevalley type over k and T and T' are k-trivial, then the theorem remains true if we replace isogeny by k-isogeny. In particular, the k-isomorphism class of a connected semisimple algebraic group of Chevalley type over k is completely determined by (X, $\mathfrak{r}$). Chevalley also showed that, for any pair (X, $\mathfrak{r}$) satisfying condition (1) above, there exists a connected semisimple algebraic group $G(X, \mathfrak{r})$ of Chevalley type defined over the prime field. Therefore, since the classification of semisimple algebraic groups of Chevalley type is reduced essentially to that of root systems (X, $\mathfrak{r}$), it turns out that, over any ground field k, there exist as many connected simple algebraic groups of Chevalley type as connected simple complex Lie groups (→ 249 Lie Groups; Appendix A, Table 5.1).

For a given semisimple algebraic group G = $G(X, \mathfrak{r})$ defined over k, put $X_0 = \mathfrak{r}_{\mathbf{Z}}$ (= the submodule of X generated by $\mathfrak{r}$), $X^0 = \{x \in F \mid \langle \alpha^*, x \rangle \in \mathbf{Z}$ for all $\alpha \in \mathfrak{r}\}$. Then we have natural isogenies $G(X^0, \mathfrak{r}) \rightarrow G(X, \mathfrak{r}) \rightarrow G(X_0, \mathfrak{r})$ (with $q_\alpha = 1$), all of which can be taken to be defined over k. The group $G(X^0, \mathfrak{r})$ ($G(X_0, \mathfrak{r})$) is called the **simply connected group** (the **adjoint group**) isogenous to G. When the characteristic of k is zero, these isogenies (which are already known in the classical theory of complex Lie groups) are essentially the only possible isogenies among the semisimple algebraic groups. But when the characteristic is $p > 0$, there are, in addition to these, the Frobenius homomorphism (with $q_\alpha = p$) and the following "singular" isogenies (for which q_α = 1 or p depending on α): $B_n \rightleftarrows C_n$, $F_4 \rightarrow F_4$ ($p = 2$), $G_2 \rightarrow G_2$ (p = 3). In particular, when k is a finite field, taking the set of fixed points of the singular k-isogenies, we obtain the simple finite groups of M. Suzuki and R. Ree (→ 151 Finite Groups).

O. Classification Theory

A connected semisimple algebraic group G defined over k is called k- **(almost) simple** if there is no proper connected closed normal subgroup of G defined over k. (When G is k-simple and k-split, the factor group $D(G_k)$/center is an abstract simple group except for a few special cases [17]. For more general results → Section Q.) For a k-simple algebraic group G, let G_1 be any one of its absolutely simple components, and let k_1 be the smallest field of definition for G_1 containing k. Then k_1/k is a finite separable extension, and G is k-isogenous to $\mathfrak{R}_{k_1/k}(G_1)$. Hence the problem of

classifying all k-simple groups (up to isogeny) is equivalent to that of finding all k_s-forms of simple groups of Chevalley type. This latter problem can be reduced, in principle, to the classification of compact k_s-forms and that of certain diagrams (i.e., †Dynkin diagrams along with an action of the Galois group) [16], J. Tits [30] (- Appendix A, Table 5.1). For instance, when k is a finite field (or, more generally, a field of dimension $\leqslant 1$ [15]), there is no compact k-simple group; hence, using a simple classification theory of the diagrams, we can show that the only absolutely simple algebraic groups G defined over k are either of Chevalley type or of the types introduced by R. Steinberg (denoted by ${}^2A_n$, ${}^2D_n$, ${}^3D_4$, ${}^6D_4$, ${}^2E_6$). Connected semisimple algebraic groups composed of the groups of these types are characterized by the property that they have a Borel subgroup defined over k. Such groups are said to be of **Steinberg type** over k (or **k-quasi-split).** Absolutely simple algebraic groups over a †p-adic field have been classified by M. Kneser and J. Tits [30, 31]. When the characteristic of k is not equal to 2, the classification of simple groups of classical type (except for the type D_4) is known to be equivalent to that of semisimple †associative algebras with †involution [18]. A similar relation also holds between some of the exceptional simple groups and †Cayley algebras or †Jordan algebras (H. Hijikata, T. A. Springer, J. Tits).

The following is a list of absolutely simple algebraic groups of classical type.

1. k-forms of $SL(n)$ $(n \geqslant 2)$.
   1.1. $G_{1k}=SL(m,\mathfrak{K})=\{g\in M_m(\mathfrak{K})\mid N(g)=1\}$, where $\mathfrak{K}$ is a †central division algebra over k with $(\mathfrak{K}:k)=r^2$, $n=mr$, and N denotes the †reduced norm in $M_m(\mathfrak{K})$.
   1.2. $G_{1k}=SU(m,\mathfrak{K},f)=\{g\in SL(m,\mathfrak{K})\mid f(gx,gy)=f(x,y)$ for $x, y\in\mathfrak{K}^m\}$, where $\mathfrak{K}$ is a central division algebra over a quadratic extension k' of k with an involution ι of the second kind (which means that $\{\xi\in k'\mid\xi^\iota=\xi\}=k$), $(K:k')=r^2$, $n=mr$, and f is a (nondegenerate) †Hermitian form of m variables over $\mathfrak{K}$ with respect to the involution ι.

II. k-forms of $SO(n)$ $(n\geqslant 3,\ n\neq 4)$, $Sp(n)$ $(n$ even, $n\geqslant 2)$.

$G_{1k}=SU(n,\mathfrak{K},f)$, where $\mathfrak{K}$ is a central division algebra over k with an involution ι of the first kind (i.e., such that $\{\xi\in k\mid\xi^\iota=\xi\}=k$), $(\mathfrak{K}:k)=r^2$, $n=mr$, and f is a nondegenerate c-Hermitian form of m variables over $\mathfrak{K}$ with respect to the involution ι. In this case, $\dim\{\xi\in K\mid\xi^\iota=\xi\}=r(r+\varepsilon_0)/2$ with $\varepsilon_0=\pm 1$, and G_1 is a k-form of SO or Sp according as $\varepsilon\varepsilon_0=1$ or -1. ($SO(8)$ may have other k-forms coming from the so-called triality.) The case where the characteristic of k is 2 can also be discussed by a method given by J. Tits (*Inventiones Math.* 5, 1968).

When k is a local field or an algebraic number field, the only central division algebra with an involution of the first kind is a †quaternion algebra (and, if ι is the "canonical involution," then $\varepsilon_0=$ 1).

P. Algebraic Groups over an Algebraic Number Field

Let G be a connected algebraic group defined over an algebraic number field k of finite degree. Let $\{v\}$ be the totality of †prime divisors (i.e., equivalence classes of valuations) of k. Taking the †restricted direct product of a family of locally compact topological groups $\{G_{k_v}\}$, we obtain a locally compact topological group G_A, called the **adele group** of G[19] (- 6 Adeles and Ideles). In particular, when G = $\mathbf{G}_m$, the adele group $I=(\mathbf{G}_m)_A$ is exactly the †idele group introduced by Chevalley in class field theory. If we identify $x\in G_k$ with an adele whose components are all equal to x, G_k becomes a discrete subgroup of G_A.

Concerning the finiteness property of G_A/G_k, the following results have been obtained [21, 22]: A character $\chi\in X_k(G)$(= the module of all k-rational characters of G) gives rise to a (continuous) homomorphism $\chi_A: G_A\to I=(\mathbf{G}_m)_A$. Put $G_A^0=\{g\in G_A\mid|\chi_A(g)|=1$ forall $\chi\in$ X,(G)}, where || is the standard norm in I. Then G_A^0 is †unimodular, and the quotient space G_A^0/G_k is of finite volume with respect to the (unique) invariant measure on it. G_A^0/G_k (G_A/G_k) is compact if and only if the semisimple part G/R (the reductive part G/R_u) of G is k-compact. From the arithmetic point of view, it is important to determine explicitly the volume of G_A^0/G_k with respect to the invariant measure normalized in a certain manner; such a volume is called the **Tamagawa number** of G and is usually denoted by $\tau(G)$ [19, 23]. For instance, Siegel's formulas on the volume of the fundamental domain of the unit group of a quadratic form f over k are essentially equivalent to a theorem on the Tamagawa number stating that $\tau(SO(f))=2$.

Let $\mathfrak{o}$ be the †ring of integers in k and L an o-lattice in the vector space on which G is acting. We can define in a natural manner an action of G_A on the set of all o-lattices; then the orbit G_AL (G_kL) of L with respect to G_A (G_k) is called the **genus (class)** of L. The †stability subgroup $G_{A,L}$ of L in G_A is open, and the double coset space $G_{A,L}\backslash G_A/G_k$ is finite (finiteness of the class number). Moreover, let $\{v_1, , v_r\}$ be the totality of †infinite prime

divisors of k, and put $G_\infty = \prod_{i=1}^r G_{k_{v_i}}$. Then G_∞ is a Lie group, and the canonical projection G_L on G_∞ of $G_{A,L} \subset G_A$ is a discrete subgroup of finite type. (In general, (discrete) subgroups of G_∞ which are †commensurable with G_L are called **arithmetic subgroups.**) As in the adele case, $\chi \in X_k(G)$ gives rise to a (continuous) homomorphism $\chi_\infty : G_\infty \to (\mathbf{R}^\times)^r$, and if $G_\infty^0 = \{g \in G_\infty \mid |\chi_\infty(g)| = 1\}$, then the quotient space G_∞^0/G_L is of finite volume. Moreover, G_∞^0/G_L (G, $/G_L$) is compact if and only if G_A^0/G_k (G_A/G_k) is compact.

In addition to these, the †approximation theorem and the †Hasse principle are also extended to (classical, or general) algebraic groups (M. Eichler, M. Kneser, G. Shimura, Hijikata, Springer; → [30]).

Q. Structure of Reductive Groups [24]

Let G be a connected †reductive group defined over a field k. Then G has a †maximal torus defined over k and G is †split over a finite †separable extension K of k. The structure of the group G_K can be discussed as in Sections I, J, K, and L. Here, we discuss the structure of G_k, i.e., a †k-form of G,.

Let S be a **maximal k-split torus** of G, i.e., a k-subtorus of G that is k-split and maximal with respect to these properties. Any two such tori are conjugate over k, i.e., by an element of G_k. Their dimension is called the k-rank of G. To say that G is of k-rank zero is equivalent to saying that G is †k-anisotropic. The centralizer Z = Z,(S) of S in G is a reductive k-group, and its derived group is k-anisotropic. Let N = N,(S) be the normalizer of S in G. The †Weyl group ${}_kW = N/Z$ relative to S is called the **k-Weyl group** of G. A **k-root** of G with respect to S is a nontrivial character of S that appears when one diagonahzes the representation of S in the +Lie algebra $\mathfrak{g}$ of G, S operating via adjoint representation. Denote by ${}_k\mathfrak{r}$ the set of all k-roots of G with respect to S. There is a decomposition of the Lie algebra $\mathfrak{g}$ of $G: \mathfrak{g} = \mathfrak{g}_0 + \sum_{\alpha \in {}_k\mathfrak{r}} \mathfrak{g}_\alpha$ where $\mathfrak{g}_\alpha = \{X \in \mathfrak{g} \mid \mathrm{Ad}(s)X = \alpha(s)X$ for all $s \in S\}$. Then $\mathfrak{g}_0$ is the Lie algebra of Z, and there is a unique unipotent k-subgroup P_α of G normalized by S such that its Lie algebra is $\mathfrak{g}_\alpha$. The set ${}_k\mathfrak{r}$ is a †root system in a suitable Euclidean space whose Weyl group is isomorphic to ${}_kW$; if G is k-split, ${}_k\mathfrak{r}$ is the ordinary root system, and the P_α are as in Section J. In general, ${}_k\mathfrak{r}$ need not be reduced (i.e., axiom (iii) in Section J need not be satisfied), nor should dim $\mathfrak{g}_\alpha$ = dim P_α always be 1.

A closed subgroup of G defined over k which is minimal among the parabolic subgroups of G is called a **minimal paraholic k-subgroup** of G. (If G is k-split, a minimal parabolic k-subgroup is a †Borel subgroup of G.) Any two minimal parabohc k-subgroups are conjugate to each other over k. If P is one such subgroup, then there exists a maximal k-split torus S such that P is the semidirect product of the reductive k-subgroup Z = Z,(S) and the †unipotent radical $U = R,(P)$ of P. The expression $P = ZU$ is called a **Levi-decomposition** of P, and Z is called a **Levi-suhgroup** of P. Any two Levi-subgroups of P are conjugate by an element in U_k. There is an ordering of ${}_k\mathfrak{r}$ such that P is generated by Z and P_α with $\alpha > 0$. The minimal parabolic k-subgroups containing a given maximal k-split torus S correspond to the †Weyl chambers of ${}_k\mathfrak{r}$. They are permuted simply transitively by the Weyl group ${}_kW$.

Fix an ordering of ${}_k\mathfrak{r}$, and let ${}_k\Delta$ be the †fundamental system of ${}_k\mathfrak{r}$ with respect to the given order. For any subset θ of ${}_k\Delta$, denote by P_θ the subgroup generated by Z = Z,(S) and P_α, where α is a linear combination of the roots of ${}_k\mathfrak{r}$ in which all roots not in θ occur with a coefficient $\geqslant 0$. Then P_θ contains P and, in particular, $P_\varphi = P$. P_θ is called a **standard parabolic k-subgroup** of G containing P. Any parabolic k-subgroup of G is conjugate over k to a unique P_θ. If S_θ is the identity component of $\bigcap_{\alpha \in \theta}(\ker\alpha)$, then S_θ is a k-split torus of G and $P_\theta = Z(S_\theta)R_u(P_\theta)$. This shows that any parabohc k-subgroup of G has a Levi-decomposition and its Levi-subgroups are conjugate to each other over k. Let P be a minimal parabolic k-subgroup of G containing a maximal k-split torus S. Put $U = R,(P)$, Z = Z,(S), and N = N,(S). Then, N = N_kZ, so G_k is the disjoint union over ${}_kW$ of double cosets $U_k n_w P_k$ $(w \in {}_kW)$, where n_w is a representative in N_k of $w \in {}_kW$. More precisely, if $w \in {}_kW$, there exist two k-subgroups U'_w and U''_w such that $U = U'_w \times U''_w$ (product of k-varieties), the map $U'_w \times P \to U n_w P$ sending (x, y) onto $x n_w y$ is an isomorphism, and

$$(G/P)_k = G_k/P_k = \bigcup_{w \in {}_kW} \pi((U'_w)_k),$$

where π is the projection $G \to G/P$. This is called a **relative Bruhat decomposition.** If G is k-split, this gives an ordinary Bruhat decomposition (→ Section K). If θ is a subset of ${}_k\Delta$, let W_θ be the subgroup of ${}_kW$ generated by reflections defined by the α's in 0. If θ, θ' are subsets of ${}_k\Delta$, then there is a bijection of double cosets

$$(P_\theta)_k \backslash G_k/(P_{\theta'})_k = W_\theta \backslash {}_kW/W_{\theta'}.$$

(Note that all these properties follow from the fact that (P_k, N_k) is a BN-pair in G_k; → Section R.)

If G is k-isotropic, let G^+ be the subgroup of G_k generated by all U_k, where U runs over the unipotent radicals of the minimal parabolic k-

subgroups of G. Kneser and Tits conjectured that when G is semisimple and simply connected and k arbitrary, $G_k = G^+$. Platonov showed that this conjecture fails for some k-forms of SL_n ($n \geqslant 2$), but it is true in many cases, e.g., when G is ^{+}k-split or ^{+}k-quasi-split (J. Tits, *Sém. Bourbaki* 29e, no. 505, 1976677). In this case, if G is simple as an algebraic group and the cardinality of k is $\geqslant 4$, then $G_k/Z(G_k)$ is simple as an abstract group. The connected semisimple algebraic k-groups over a finite field k are k-split or k-quasi-split (→ Section 0). J. Tits has given a reduction process for classifying the groups over an arbitrary field k. He defined the index of the k-group (to a certain extent a generalization of Witt's theorem characterizing a quadratic form by means of its index and anisotropic kernel) and gave all possible indices of the k-groups and also a complete list in the case of local tïeld (J. Tits [30, 31]).

R. Buildings and BN-pairs [25]

The origin of the notations of buildings and BN-pairs lies in an attempt to give a systematic procedure for the geometric interpretation of the semisimple Lie groups and, in particular, the exceptional groups. The theory has various applications to the groups of Lie types. To describe this precisely, we must introduce a number of detïnitions.

A **complex** A is a set with an order relation $\subset$, read "is a **face** of" or "is contained in" such that for a given element *A*, the ordered subset *S(A)* of all faces of *A* is isomorphic to the set of all subsets of a set. The subset *S(A)* of A is called a **simplex** in A. A complex has a smallest element, which we denote by 0. For an element *A*, the number rk *A* of minimal nonzero faces of *A* is called the **rank** of *A*. Define rk A = sup(rk $A \mid A \in \Delta$). **A morpbism** $\alpha: \Delta \to \Delta'$ (where A and A' are complexes) is a mapping of the underlying sets such that for every $A \in \Delta$, $S(A) \cong S(\alpha(A))$. A subcomplex of A is a complex whose underlying set is a subset of A such that the inclusion is a morphism. If $A \in \Delta$, the **star** St(A) of *A* is the set of all elements of A containing *A*. With the order relation induced from A, *St(A)* is a complex. If $B \in \mathrm{St}(A)$, the rank of *B* in *St(A)* is called the **codimension** of *A* in *B* and is denoted by $\mathrm{codim}_B A$.

A complex A is called a **cbamber complex** if every element is contained in a maximal element, which is called a **cbamber,** and if given two chambers C, C', there exists a sequence of chambers C = C_0, C, , , C_m = C' such that $\mathrm{codim}_{C_{i-1}}(C_{i-1} \cap C_i) = \mathrm{codim}_{C_i}(C_{i-1} \cap C_i) \leqslant 1$ for all i= 1, 2, , *m.* A chamber complex is called **tbick** (resp. **tbin)** if every element of codimension 1 is contained in at least three (resp. exactly two) chambers. An endomorphism φ of a thin chamber complex A is called a **folding** if $\varphi^2 = \varphi$ and if any chamber on $\varphi(\Delta)$ is the image of exactly two chambers in A by φ. For any folding φ, $\varphi(\Delta)$ is called a **root. A** thin chamber complex is called a **Coxeter complex** if for any pair of **adjacent** chambers C, C', i.e., chambers such that $\mathrm{codim}_C(C \cap C') = 1$, there is a root containing C and not C'. Let Σ be a Coxeter complex. For any root $\varphi(\Sigma)$ of Σ, there is only one root $\varphi'(\Sigma)$, called the **opposite** of $\varphi(\Sigma)$, such that $\varphi(\Sigma) \cap \varphi'(\Sigma)$ does not contain any chamber and $\varphi(\Sigma) \cup \varphi'(\Sigma) = \Sigma$; there is also an involutive automorphism called the **reflection** associated with φ that transforms $\varphi(\Sigma)$ onto $\varphi'(\Sigma)$. The group W(C) generated by all reflections of Σ is called the **Weyl group** of Σ, which turns out to be a **Coxeter group,** i.e., there exist sets I and $\{m_{ij}\}$, $(i,j) \in I \times I$, where m_{ij} are integers or ∞ and $m_{ii} = 1$, such that the group is presented by the generators $\{r_i\}_{i \in I}$ and the fundamental relations $(r_i r_j)^{m_{ij}} = 1$, $i, j \in I$, $m_{ij} \neq \infty$. A Coxeter complex is called **irreducible** if it is not the join of two nonempty Coxeter subcomplexes.

A **building** is a thick chamber complex A with a system $\mathfrak{A}$ of Coxeter subcomplexes, called the **apartments** of A such that (i) every two simplexes of A belong to an apartment. (ii) If $\Sigma, \Sigma' \in \mathfrak{A}$, there exists an isomorphism of Σ onto Σ' that fixes $\Sigma \cap \Sigma'$ (elementwise). Since the apartments of a building are isomorphic to each other, we can define the Weyl group, rank, and irreducibility of the building to be those of its apartment. A building with finitely many chambers is said to be of **spberical type.** If a building can be realized as a simplicial decomposition of a Euclidean space, then one says it is of **Euclidean type.**

Examples of buildings are provided by *BN*-pairs. A **BN-pair** or **Tits system** in a group G is a system *(B,* N) consisting of two subgroups of G such that

(BNO) B and N generate G;
$(BN1)$ $B \cap N = H \lhd N$; and
(BN2) The group $W = N/H$ has a generating set R such that for any $r \in R$ and any $w \in W$,
(BN2') $rBw \subset BwB \cap BrwB$,
(BN2") $rBr \neq B$.

The group W is called the **Weyl group** of the BN-pair. For any subset S of *R,* let W_s denote the subgroup of W generated by S. Set P_S = $B\,W_S B$. Then P_S is a subgroup of G and the mapping $S \mapsto P_S$ is a bijection of the lattice of all subsets of R onto the lattice of all subgroups of G containing *B.* A subgroup of G is called **parabolic** if it is conjugate to some P_S. Let A be the set of all left cosets of all subgroups P_S, $S \subset R$, ordered by the opposite of the inclusion relation. Let G operate on A on

the left. Let Σ be the subset $\{nP_S \; n\in N, S\subset R\}$ of A, and $\mathfrak{A}$ be the set of G-translates of Σ. Then (A, $\mathfrak{A}$) is a building whose Weyl group is W.

Let G be a reductive algebraic group defined over a field k and P be a minimal parabolic k-subgroup of G containing a maximal k-split torus S and set N = N,(S). Then (P_k, N_k) *is* a BN-pair in G_k. Therefore there is a building of spherical type associated with the group G. Conversely, the buildings of rank $\geqslant 3$ and irreducible spherical type (roughly speaking) all turn out to be associated with simple algebraic or classical groups (J. Tits [25]). This result gives a complete and unified description of structures that were discovered previously in certain cases. For example, the building of type A, gives a †projective space (E. Abe, T. Tsuzuku), that of type C_n gives a polar space (Veldkamp), and that of a k-form of type E_6 gives a Cayley space (J. Tits). As an application, one can show that a finite building of rank $\geqslant 3$ and irreducible spherical type is isomorphic to the building of an †absolutely simple algebraic group over a finite field.

When k is local (i.e., endowed with a complete discrete valuation whose residue field is Perfect), the reductive group defined over k has another BN-pair such that the associated building is Euclidean. This theory was initiated by N. Iwahori and H. Matsumoto [26], who considered split semisimple groups. Later, quasi-split and classical groups were studied by H. Hijikata, and the theory for the general case was given by F. Bruhat and J. Tits [27, 28]. To distinguish from the usual BN-pair structure, the subgroups conjugate to B in this case are called **Iwabori subgroups,** and parabolic subgroups are called **paraholic subgroups.** The Euclidean buildings are the "ultrametric analogs" of the †Riemannian symmetric spaces. In other words, in the study of p-adic simple groups, they play a role similar to that of the symmetric spaces in the theory of simple groups.

References

[1] A. Borel, Linear algebraic groups, Benjamin, 1969.
[2] J. E. Humphreys, Linear algebraic groups, Springer, 1975.
[3] M. Demazure and P. Gabriel, Groupes algebriques 1, North-Holland, 1970.
[4] M. Demazure and A. Grothendieck, Schémas en groupes, Séminaire de géometrie algébrique, 196331964, Publ. Math. Inst. HES, 1964 (Lecture notes in math. 153, Springer, 1970).
[5] C. Chevalley, Theorie des groupes de Lie, Groupes algébriques, Théorèmes généraux sur les algebre de Lie, Actualités Sci. Ind., 1152 and 1226, Hermann, 1968.
[6] C. Chevalley, Sur certains groupes simples, Tôhoku Math. J., (2) 7 (1955), 14–66.
[7] C. Chevalley, Classification des groupes de Lie algébriques, Séminaire C. Chevalley, Inst. H. Poincaré, Univ. Paris, 195661958.
[8] M. Rosenlicht, Some basic theorems on algebraic groups, Amer. J. Math., 78 (1956), 401–443.
[9] M. Rosenlicht, Some rationality questions on algebraic groups, Ann. Mat. Pura Appl., 43 (1957), 25–50.
[10] A. Weil, On algebraic groups of transformations, Amer. J. Math., 77 (1955), 355–391.
[11] A. Weil, On algebraic groups and homogeneous spaces, Amer. J. Math., 77 (1955), 4933512.
[12] J.-P. Serre, Groupes algébriques et corps de classes, Actualités Sci. Ind., Hermann, 1959.
[13] J. Dieudonné, Introduction to formal groups, Dekker, 1973.
[14] N. Bourbaki, Groupes et algèbres de Lie, Hermann, 1968, chs. 446.
[15] J.-P. Serre, Cohomologie galoisienne, Lecture notes in math. 5, Springer, 1964.
[16] I. Satake, Classification theory of semisimple algebraic groups, Dekker, 1971.
[17] J. Tits, Algebraic and abstract simple groups, Ann. Math., (2) 80 (1964), 313-329.
[18] A. Weil, Algebras with involutions and the classical groups, J. Indian Math. Soc., 24 (1960), 5899623.
[19] A. Weil, Adeles and algebraic groups, Lecture notes, Institute for Advanced Study, Princeton, 1961.
[20] A. Borel, Introduction aux groupes arithmétiques, Hermann, 1969.
[21] A. Borel and Harish-Chandra, Arithmetic subgroups of algebraic groups, Ann. Math., (2) 75 (1962), 4855535.
[22] G. D. Mostow and T. Tamagawa, On the compactness of arithmetically defined homogeneous spaces, Ann. Math., (2) 76 (1962), 446-463.
[23] T. Ono, On the Tamagawa number of algebraic tori, Ann. Math., (2) 78 (1963), 44-73.
[24] A. Borel and J. Tits, Groupes réductifs, Publ. Math. Inst. HES, 27 (1965), 555151. Complements, Publ. Math. Inst HES, 41 (1972), 2533276.
[25] J. Tits, Buildings of spherical type and finite BN-pairs, Lecture notes in math. 386, Springer, 1974.
[26] N. Iwahori and H. Matsumoto, On some Bruhat decompositions and the structure of p-adic Chevalley groups, Publ. Math. Inst. HES. 25 (1965), 5-48.

[27] H. Hijikata, On the structure of semi-simple algebraic groups over valuation fields I, Japan. J. Math. 1 (1975), 2255300.
[28] F. Bruhat and J. Tits, Groupes réductifs sur un corps local I. Données radicielles valuées, Publ. Math. Inst. HES, 41 (1972), 55252. II. Schémas en groupes. Existence d'une donnée radicielle valuée, Publ. Math. Inst. HES, 60 (1984), 5- 184.
[29] Colloque sur la théorie des groupes algébriques, C. B. R. M., Brussels, 1962.
[30] Algebraic groups and discontinuous subgroups, Amer. Math. Soc. Proc. Symposia in Pure Math., 9, 1966.
[31] Automorphic forms, representations and L-functions, Amer. Math. Soc. Proc. Symposia in Pure Math., 33, 1979.
[32] T. A. Springer, Linear algebraic groups, Birkhäuser, 198 1.
[33] G. P. Hochschild, Basic theory of algebraic groups and Lie algebras, Springer, 198 1.

14 (V.13) Algebraic Number Fields

A. Introduction

A complex number that satisfies an algebraic equation with rational integral coefficients is said to be an **algebraic number.** If the coefficient of the term of highest degree of the equation is 1, this algebraic number is said to be an **algebraic integer.** The set A of all algebraic numbers is a field which is the †algebraic closure of the rational number tïeld Q in the complex number tïeld C. The set I of all algebraic integers is an †integral domain which contains the integral domain **Z** of all the rational integers. The †field of quotients of I is A.

B. Principal Order

An extension fïeld k of Q of finite degree (which we shall always suppose to be contained in C) is said to be an **algebraic number field** of finite degree, and k is a subfield of A. The intersection o = $k \cap I$ is an integral domain whose field of quotients is k; o is called the **principal order** of k. (More generally, a subring R of o containing 1 is said to be an **order** of k if the fïeld of quotients of R is k. The set $\mathfrak{f}$ of all elements y of o such that $\gamma\mathfrak{o} \subset R$ is an **ideal** of o; in addition $\mathfrak{f}$ is called the **conductor** of R.) Let n be the degree of k over Q. Then the additive group of the principal order o of k is a †free Abelian group of †rank n. A †basis $(\omega_1, , \omega_n)$ of o as a free Abelian group (or Z-module) is said to be a **minimal basis** of o (or of k). Let $\omega_j^{(i)}$ (i = 1, , n) be †conjugate elements of ω_j over **Q**, and let A = $|\omega_j^{(i)}|$ be the determinant whose (i,j) entry is $\omega_j^{(i)}$. Then $D_k = \Delta^2$ is a rational integer that is independent of the choice of a minimal basis of o. D_k is called the **discriminant** of k. If $k \neq \mathbf{Q}$, then $|D_k| > 1$ **(Minkowski's theorem,** 1891). For any given rational integer m there are only a finite number of algebraic number fields whose discriminants are equal to m (C. Hermite and H. Minkowski, 1896). The proof of these theorems depends on the methods of geometry of numbers (- 182 Geometry of Numbers).

C. Ideals of tbe Principal Order

An †ideal a of the principal order o is said to be an **integral ideal** of k. In particular, a prime ideal ($\neq 0$) of o is called simply a prime ideal of k. The domain o is not necessarily a †principal ideal ring but is always a †Dedekind domain. That is, every ideal a of o is uniquely expressed (up to the order of the factors) as a finite product of powers of prime ideals of o. This theorem is called the **fundamental theorem of tbe principal order o.**

The quotient ring $\mathfrak{o}/\mathfrak{a}$ of o by an ideal a $(\neq 0)$ of o is a finite ring. The number of elements of $\mathfrak{o}/\mathfrak{a}$ is called the **absolute norm** of a and is denoted by N(a). We have $N(\mathfrak{ab}) = N(\mathfrak{a})N(\mathfrak{b})$. Every prime ideal $\mathfrak{p}$ ($\neq 0$) of o is a †maximal ideal of o, and $\mathfrak{o}/\mathfrak{p}$ is a Imite field. Let the †characteristic of $\mathfrak{o}/\mathfrak{p}$ be p, where p is a prime number. Then $\mathfrak{o}/\mathfrak{p}$ is a Imite extension of the †prime field $\mathbf{Z}/p\mathbf{Z}$. Let the degree of $\mathfrak{o}/\mathfrak{p}$ over $\mathbf{Z}/p\mathbf{Z}$ be f. Then N(p) = p^f, and f is said to be the **degree** of the prime ideal p.

Let s be a complex variable. The (complex-valued) function

$$\zeta_k(s) = \sum_{\mathfrak{a}} 1/(N(\mathfrak{a}))^s = \prod_{\mathfrak{p}} 1/(1-(N(\mathfrak{p}))^{-s})$$

of s $\in$ C is called the **Dedekind zeta function** of k (R. Dedekind, 1871). Here the summation extends over all ideals a of o, and the product extends over all prime ideals p of o. This series converges absolutely for Re $s > 1$, and the function $\zeta_k(s)$ has a single-valued †analytic continuation to a +meromorphic function on the whole complex plane (- 450 Zeta Functions).

D. Units

An algebraic integer ε of k is said to be a **unit** of k if ε^{-1} is also an algebraic integer. Hence ε is a unit of k if and only if the +Principal ideal

(E) is o. The set E_k of all units of k forms an Abelian group under multiplication, which is called the **unit group** of k. The set of all elements of E_k of finite order coincides with the set of all the roots of unity contained in k and forms a cyclic group of a finite order w. Let n be the degree of k over Q. Then for each element $\alpha \in k$ there are n conjugate elements $\alpha^{(i)}$ over Q. Let $\alpha^{(i)}$ ($i = 1, , r_1$) be real for any $\alpha \in k$, and let $\alpha^{(r_1+j)}$ and $\alpha^{(r_1+r_2+j)}$ (j = 1, . , r_2) be pairs of complex conjugates for any $\alpha \in k$. Then we have $n = r_1 + 2r_2$. The unit group E_k of k is the direct product of a cyclic group of order w and the free Abelian multiplicative group of †rank $r = r_1 + r_2 - 1$. This theorem is called **Dirichlet's unit theorem** (1846). A basis $(\varepsilon_1, , \varepsilon_r)$ of this free group is called a system of **fundamental units** of k.

Let $l^{(i)}\alpha = \log|\alpha^{(i)}|$ $(i = 1, \ldots, r_1)$, $l^{(j)}\alpha = 2\log|\alpha^{(j)}|$ $(j = r_1 + 1, \ldots, r_1 + r_2)$ for $\alpha \in k$. For r elements $\eta_1, , \eta_r$ of E_k,

$$R[\eta_1, \ldots, \eta_r] = \begin{vmatrix} l^{(1)}\eta_1 & l^{(1)}\eta_2 & \ldots & l^{(1)}\eta_r \\ l^{(2)}\eta_1 & l^{(2)}\eta_2 & & l^{(2)}\eta_r \\ \cdot & & & \\ l^{(r)}\eta_1 & l^{(r)}\eta_2 & & l^{(r)}\eta_r \end{vmatrix}$$

is called the **regulator** of $(\eta_1, , \eta_r)$ (Dedekind). In order for $\eta_1, \ldots, \eta_r$ to be multiplicatively independent, it is necessary and sufficient that R $[\eta_1, , \eta_r] \neq 0$. The absolute value of R $[\eta_1, . , \eta_r]$ takes the minimum positive value R for fundamental units $(\varepsilon_1, . . , \varepsilon_r)$. R = R $[\varepsilon_1, , E,]$ is independent of the choice of fundamental units $(\varepsilon_1, . , \varepsilon_r)$ of k. R is called the **regulator** of k. In general, $R[\eta_1, , \eta_r]/R$ is equal to the index $[\![E_k : H]\!]$ of the group H generated by the roots of unity in k and $\eta_1, , \eta_r$. H. W. Leopoldt conjectured that units in k, which are multiplicatively independent over **Z**, remain multiplicatively independent over $\mathbf{Z}_p$ (the ring of p-adic integers) when they are considered as elements of the tensor product $k \otimes \mathbf{Q}_p$ over Q. This conjecture was affirmatively proved in some special cases by J. Ax (Illinois *J.* Math., 9 (1965)) and others.

If $k/\mathbf{Q}$ is a Galois extension, there exists a unit ε of k such that the conjugates of ε over Q contain r multiplicatively independent units **(Minkowski's theorem).**

E. Ideal Classes

An o-module contained in k (i.e., $\mathfrak{o}\mathfrak{a} \subset \mathfrak{a}$) such that $\alpha\mathfrak{a} \subset \mathfrak{o}$ holds for some element α $(\neq 0)$ of k is said to be a **fractional ideal** of k. For two fractional ideals a, b of k the "product" ab defined by $\{\sum \alpha_i\beta_i \text{ (finite sum)} \mid \alpha_i \in \mathfrak{a}, \beta_i \in \mathfrak{b}\}$ is also a fractional ideal. Thus the set of the fractional ideals of k forms a multiplicative commutative semigroup. For a fractional ideal a the set $\mathfrak{a}^{-1} = \{\alpha \in k \mid \alpha\mathfrak{a} \subset \mathfrak{o}\}$ is also a fractional ideal of k, and we have $\mathfrak{a}\mathfrak{a}^{-1} = \mathfrak{o}$. Thus the set of all nonzero fractional ideals of k forms an Abelian group $\mathfrak{I}_k$ under multiplication with o as identity. Each fractional ideal a $(\neq 0)$ is uniquely expressed as a finite product of powers of prime ideals, if we admit negative powers. Namely, $\mathfrak{I}_k$ is a free Abelian multiplicative group with the set of all prime ideals as basis. Given fractional ideals a and b, we say that a is **divisible** by b if $\mathfrak{a} \subset \mathfrak{b}$; in this case, we call b a **divisor** of a and a a **multiple** of b. Also, $\mathfrak{a} \subset \mathfrak{b}$ if and only if there exists an integral ideal c such that a = bc. Given fractional ideals $\mathfrak{a} = \prod \mathfrak{p}_i^{e_i}$ and $\mathfrak{b} = \prod \mathfrak{q}_j^{f_j}$ $(e_i \neq 0, f_j \neq 0)$, we say that a and b are **relatively prime** if $\{\mathfrak{p}_i\}$ and $\{\mathfrak{q}_j\}$ are disjoint. Usually a fractional ideal of k is simply called an ideal of k.

For an element α $(\neq 0)$ of k, $(\alpha) = \alpha\mathfrak{o}$ is a (fractional) ideal of k, and (α) is said to be a **principal ideal** of k. The set P_k of all principal ideals (α) $(\alpha \in k, \alpha \neq 0)$ is a subgroup of $\mathfrak{I}_k$. Since (α) = o is equivalent to $\alpha \in E_k$, we have $P_k \cong k^*/E_k$, where k^* is the multiplicative group of all nonzero elements of k.

Each coset of $\mathfrak{I}_k$ modulo P_k is called an **ideal class** of k, and the group $\mathfrak{C}_k = \mathfrak{I}_k/P_k$ is called the **ideal class group** of k. Each ideal class contains an integral ideal a with N(a) $\leqslant \sqrt{|D_k|}$ (more precisely, with $N(\mathfrak{a}) \leqslant (4/\pi)^{r_2}(n!/n^n)\sqrt{|D_k|}$). From this it follows that $\mathfrak{C}_k$ is a finite Abelian group. The order h of $\mathfrak{C}_k$ is called the **class number** of k. For the calculation of the class number the †residue at the pole $s = 1$ of the Dedekind zeta function is used. Namely,

$$\lim_{s \to 1+0} (s-1)\zeta_k(s) = gh,$$

$$g = 2^{r_1+r_2}\pi^{r_2}R_k/w_k\sqrt{|D_k|},$$

where R_k is the regulator of k and w_k is the number of roots of unity in k (Dedekind, 1877). This formula is used, in particular, for the computation of the class numbers of †quadratic fields and †cyclotomic fields (→ 347 Quadratic Fields). The class numbers of cubic and quartic (real) †cyclic fields over Q were computed by H. Hasse in the case where the †conductor of $k/\mathbf{Q}$ is less than 100 *(Abh. Deutsch. Akad. Wiss. Berlin, 2* (1948)). Hasse has also given a detailed computation of the class number of cyclotomic fields [15). In general, let the degree $n = [k : \mathbf{Q}]$ be fixed and let $|D_k| \to \infty$. Then

$$\lim(\log(h_kR_k)/\log\sqrt{|D_k|}) = 1.$$

(This formula was proved for $n = 2$ by C. L. Siegel, 1935, and for general n by R. Brauer, *Amer. J. Math., 69* (1947).)

F. Valuations

All the †Archimedean and †non-Archimedean valuations of an algebraic number field k can be obtained as follows (→ 439 Valuations):

Archimedean Valuations. Let $n=[k:\mathbf{Q}]$, and let n conjugates of $\alpha\in k$ be $\alpha^{(1)},\ldots,\alpha^{(n)}$ such that $\alpha^{(i)}$ $(i=1,\ldots,r_1)$ is real, and $\alpha^{(r_1+j)}$ and $\alpha^{(r_1+r_2+j)}$ $(j=1,\ldots,r_2)$ are pairs of complex conjugates. We write $|\alpha|_j=|\alpha^{(j)}|$ $(j=1,\ldots,r_1+r_2)$; these are Archimedean valuations of k that are not mutually equivalent. The equivalence classes of these valuations are denoted by $\mathfrak{p}_\infty^{(1)},\ldots,\mathfrak{p}_\infty^{(r_1+r_2)}$, respectively, and are called the †infinite prime divisors of k. The first r_1 infinite prime divisors are called †real and the remaining r_2 are called †imaginary (or complex). The valuations of k defined by

$$|\alpha|_{\mathfrak{p}_\infty^{(j)}}=|\alpha|_j,\quad j=1,\ldots,r_1,$$
$$=|\alpha|_j^2,\quad j=r_1+1,\ldots,r_1+r_2,$$

are called †normal valuations of k. Here $|\;|_{\mathfrak{p}_\infty^{(j)}}$ $(j=r_1+1,\ldots,r_1+r_2)$ are valuations in the wider sense. If $r_1=n$ we call k a **totally real field**, and if $r_1=0$ we call k a **totally imaginary field**.

Non-Archimedean Valuations. Let $\mathfrak{p}$ be a prime ideal of k and α an element of k. Let $(\alpha)=\mathfrak{p}^a\mathfrak{b}$, where $\mathfrak{p}$ and $\mathfrak{b}$ are relatively prime. Put $v_\mathfrak{p}(\alpha)=a$. Then for any constant ρ $(0<\rho<1)$,

$$|\alpha|_\mathfrak{p}=\rho^{v_\mathfrak{p}(\alpha)}$$

is a non-Archimedean valuation of k. This valuation of k is called the †$\mathfrak{p}$-adic valuation of k; $\mathfrak{p}$-adic valuations for different prime ideals are mutually inequivalent. The valuation $|\alpha|_\mathfrak{p}$ with $\rho=(N(\mathfrak{p}))^{-1}$ is called a †normal valuation of k. The equivalence class of valuations containing $|\;|_\mathfrak{p}$ is denoted by the same letter $\mathfrak{p}$ and is called a †finite prime divisor of k.

A formal finite product of powers of finite or infinite prime divisors $\mathfrak{m}^*=\prod\mathfrak{p}_i^{e_i}$ is called a **divisor** of k. If all $e_i\geqslant0$, then $\mathfrak{m}^*$ is called an **integral divisor** of k. Given divisors $\mathfrak{m}^*=\prod\mathfrak{p}_i^{e_i}$ and $\mathfrak{n}^*=\prod\mathfrak{p}_i^{f_i}$, we write $\mathfrak{m}^*\mid\mathfrak{n}^*$ if $e_i\leqslant f_i$ $(i=1,2,\ldots)$.

Any valuation of k is equivalent to one of the valuations defined previously (A. Ostrowski, 1918; E. Artin, 1932). For any element α $(\neq0)$ of k the †product formula $\prod_\mathfrak{p}|\alpha|_\mathfrak{p}=1$ holds, where $\mathfrak{p}$ runs over all finite and infinite prime divisors of k and $|\;|_\mathfrak{p}$ are the normal valuations of k. Conversely, let k be a field, and let $V=\{|\;|_\mathfrak{p}\}$ be a set of inequivalent valuations of k such that (i) for any $\alpha\in k$ $(\alpha\neq0)$ $|\alpha|_\mathfrak{p}\neq1$ holds only for a finite number of $\mathfrak{p}$ in V; (ii) the product formula $\prod_\mathfrak{p}|\alpha|_\mathfrak{p}=1$ $(\alpha\in k,\ \alpha\neq0)$ holds; and (iii) there is at least one Archimedean valuation in V. Then k is an algebraic number field and V is the set of all the prime divisors of k (Artin and G. Whaples, *Bull. Amer. Math. Soc.*, 51 (1945)).

G. Ideal Classes in the Narrow Sense

For $\alpha,\beta\in k$, the expression $\alpha\equiv\beta\pmod{\mathfrak{p}_\infty^{(i)}}$ means $\alpha^{(i)}\beta^{(i)}>0$ for a real infinite prime divisor $\mathfrak{p}_\infty^{(i)}$ and $\alpha^{(i)}\beta^{(i)}\neq0$ for an imaginary infinite prime divisor $\mathfrak{p}_\infty^{(i)}$. We call an element $\alpha\in k$ **totally positive** if all real conjugates $\alpha^{(i)}$ $(i=1,\ldots,r_1)$ are positive. In the notation just given, this means $\alpha\equiv1\pmod{\mathfrak{p}_\infty^{(i)}}$ $(i=1,\ldots,r_1)$. The set of all principal ideals (α) generated by totally positive elements $\alpha\in k$ is a multiplicative subgroup P_k^+ of P_k. Each coset of $\mathfrak{J}_k$ modulo P_k^+ is called an **ideal class of k in the narrow sense.** Let E_k^+ be the group of all totally positive units of k. Then we have $(\mathfrak{J}_k:P_k^+)=h2^{r_1}/(E_k:E_k^+)$.

H. Multiplicative Congruence

Let $\mathfrak{m}$ be an integral ideal of k, and let $k^*(\mathfrak{m})$ be the multiplicative group of all elements α in k such that (α) is relatively prime to $\mathfrak{m}$. Any element $\alpha\in k^*(\mathfrak{m})$ can be expressed in the form $\alpha=\beta/\gamma$ such that $\beta,\gamma\in\mathfrak{o}$ and $(\beta),(\gamma)$ are relatively prime to $\mathfrak{m}$.

Consider an integral divisor $\mathfrak{m}^*=\mathfrak{m}\prod\mathfrak{p}_\infty^{(i)}$ which is a formal product of $\mathfrak{m}$ and infinite prime divisors $\mathfrak{p}_\infty^{(i)}$ of k. We call $\mathfrak{m}$ the finite part of $\mathfrak{m}^*$. Given an element $\alpha\in k^*(\mathfrak{m})$ and elements $\beta,\gamma\in k^*(\mathfrak{m})\cap\mathfrak{o}$ such that $\alpha=\beta/\gamma$, we set $\alpha\equiv1\pmod{^\times\mathfrak{m}^*}$ if $\beta\equiv\gamma\pmod{\mathfrak{m}}$, and $\alpha\equiv1\pmod{\mathfrak{p}_\infty^{(i)}}$. The set of all α in $k^*(\mathfrak{m})$ such that $\alpha\equiv1\pmod{^\times\mathfrak{m}^*}$ forms a multiplicative group. We write $\alpha\equiv\beta\pmod{^\times\mathfrak{m}^*}$ for $\alpha,\beta\in k$ if $\alpha/\beta\in k^*(\mathfrak{m})$ and $\alpha/\beta\equiv1\pmod{^\times\mathfrak{m}^*}$. This congruence is called the **multiplicative congruence.** In the following discussion we shall write $\operatorname{mod}\mathfrak{m}^*$ for $\operatorname{mod}^\times\mathfrak{m}^*$.

We denote by $\mathfrak{J}_k(\mathfrak{m})$ the group of all ideals of k that are relatively prime to an integral ideal $\mathfrak{m}$, and by $S(\mathfrak{m}^*)$ the group of all principal ideals (α) such that $\alpha\in k^*(\mathfrak{m})$, $\alpha\equiv1$ $(\operatorname{mod}\mathfrak{m}^*)$; $S(\mathfrak{m}^*)$ is known as the **ray** modulo $\mathfrak{m}^*$. Any subgroup H of $\mathfrak{J}_k(\mathfrak{m})$ which contains $S(\mathfrak{m}^*)$ is called an **ideal group** modulo $\mathfrak{m}^*$, and the factor group $\mathfrak{J}_k(\mathfrak{m})/H$ is called a **group of congruence classes** of ideals modulo $\mathfrak{m}^*$.

If $\mathfrak{n}^*\mid\mathfrak{m}^*$ for integral divisors $\mathfrak{m}^*$ and $\mathfrak{n}^*$ of k, then $\mathfrak{J}_k(\mathfrak{m})\subset\mathfrak{J}_k(\mathfrak{n})$ and $S(\mathfrak{m}^*)\subset S(\mathfrak{n}^*)$. If H is an ideal group modulo $\mathfrak{n}^*$, then $Q(H)=H\cap\mathfrak{J}_k(\mathfrak{m})$ is an ideal group modulo $\mathfrak{m}^*$, and we have $\mathfrak{J}_k(\mathfrak{n})/H\cong\mathfrak{J}_k(\mathfrak{m})/Q(H)$. For any given ideal group H_0 modulo $\mathfrak{m}^*$ there is a smallest

integral divisor $\mathfrak{f}^*$ such that $\mathfrak{f}^*$ $\mathfrak{m}^*$, and there exists an ideal group H modulo $\mathfrak{f}^*$ with $\Phi(H) = H_0$ (i.e., if there is an ideal group H' modulo $\mathfrak{n}^*$ with $\Phi(H') = H_0$, then $\mathfrak{f}^*$ $\mathfrak{n}^*$). We call $\mathfrak{f}^*$ the **conductor** of the ideal group H. The notion of multiplicative congruence is used in †class field theory and in the theory of †norm-residue symbols.

I. Ideal Tbeory for Relative Extensions

If an algebraic number lïeld K has a subfield k, we say that K/k is a **relative algebraic number field.** Let $\mathfrak{O}$ be the principal order of K. For a (fractional) ideal a of k, $\mathfrak{O}\mathfrak{a}$ is an ideal of K. We write $\mathfrak{O}\mathfrak{a}=E(\mathfrak{a})$ and call E(a) the **extension** of a to K. For ideals a, b of k, we have $E(\mathfrak{a}\mathfrak{b})=E(\mathfrak{a})E(\mathfrak{b})$ and $E(\mathfrak{a})\cap k=\mathfrak{a}$.

Let $\Psi_i : K \to$ C be k-isomorphisms ($i = 1,\ldots,n$), where $n=[K:k]$. We write $K^{(i)}=\Psi_i(K)$ and $A^{(i)}=Y_i(A)$ for $A\in K$. For an ideal $\mathfrak{A}$ of K, $\mathfrak{A}^{(i)}=\{A^{(i)}$ $A\in\mathfrak{A}\}$ is an ideal of $K^{(i)}$, and $\mathfrak{A}^{(i)}$ is called the **conjugate ideal** of $\mathfrak{A}$ in $K^{(i)}$. Let L be the composite field of $K^{(1)}$, , $K^{(n)}$. Then the ideal generated by $\mathfrak{A}^{(1)}\ldots\mathfrak{A}^{(n)}$ in L is the extension of an ideal a of k. We write a = $N_{K/k}(\mathfrak{A})$ and call a the **relative norm** of $\mathfrak{A}$ over k. We have $N_{K/k}(\mathfrak{A}\mathfrak{B}) = N_{K/k}(\mathfrak{A})N_{K/k}(\mathfrak{B})$ and $N_{K/k}(E(\mathfrak{a}))$ = a" (for an ideal a of k). In particular, for k = $\mathbf{Q}$, $N_{K/\mathbf{Q}}(\mathfrak{A})=(N(\mathfrak{A}))$.

Let p be a prime ideal of k. Then E(p) = $\mathfrak{P}_1^{e_1}$ $\mathfrak{P}_2^{e_2}$ $\mathfrak{P}_g^{e_g}$ in $\mathfrak{O}$, where $\mathfrak{P}_1$, , $\mathfrak{P}_g$ are prime ideals of K. Let f_i be the degree of the finite field $\mathfrak{O}/\mathfrak{P}_i$ over $\mathfrak{o}/\mathfrak{p}$. Then $N_{K/k}(\mathfrak{P}_i) = \mathfrak{p}^{f_i}$; f_i is called the **relative degree** of $\mathfrak{P}_i$ over k, and e_i is called the **(relative) ramification index** of $\mathfrak{P}_i$ over k. We have the relation $n = \sum_{i=1}^{g} e_i f_i$ between these numbers. If e, = $=e_g=$ 1, the prime ideal p is said to be **unramified** for K/k. Otherwise, p is said to be **ramified** for K/k. If every prime ideal of k is unramified for K/k, we call K/k an **unramified extension.** (For an infinite prime divisor p$_\chi$ of k we write $\mathfrak{p}_\infty = \prod_{i=1}^{g} \mathfrak{P}_\infty^{(i)e_i}$ if the Archimedean valuation $|_{\mathfrak{p}_\infty}$ of k can be extended to g Archimedean valuations $|_{\mathfrak{P}_\infty^{(i)}}$ $(i=1,\ldots,g)$ of K, where $e_i=$ 2 if $\mathfrak{P}_\infty^{(i)}$ is imaginary and $\mathfrak{p}_\infty$ is real, e_i = 1 otherwise.)

J. Relative Differents and Relative Discriminants

Let K/k be a relative algebraic number field and $\mathfrak{o}$, 8 be the principal orders of k, K, respectively. Put $\mathfrak{M}=$ {A $\in K$ $Tr_{K/k}(AD)\subset\mathfrak{o}\}$, where $Tr_{K/k}$ is the †trace (- 149 Fields J). Then $\mathfrak{M}$ is a (fractional) ideal of K and $\mathfrak{M}^{-1}$ = $\mathfrak{D}_{K/k}$ is an integral ideal of K; $\mathfrak{D}_{K/k}$ is called the **relative different** of K over k. When k = $\mathbf{Q}$, $\mathfrak{D}_{K/\mathbf{Q}}$ is simply called the **different** of K. For $L\supset K\supset k$, **we** have the **chain theorem:** $\mathfrak{D}_{L/k}=\mathfrak{D}_{L/K}\mathfrak{D}_{K/k}$.

Let the conjugates of $A\in K$ over k be $A^{(1)}$, , A'"', and assume that $A^{(1)}=A$. Put $\delta_{K/k}(A)=\prod_{i=2}^{n}(A-A^{(i)})$ for $A\in K$. If $A\in\mathfrak{O}$, thenà$_{K/k}(A)\in\mathfrak{D}_{K/k}$ $\mathfrak{D}_{K/k}$ is generated by $\{\delta_{K/k}(A)\mid A\in\mathfrak{O}\}$. The integral ideal $\mathfrak{E}^{(i)}$ generated by {A $-A^{(i)}$ $A\in\mathfrak{O}\}$ in the field $L=K^{(1)}K^{(2)}$ $K^{(n)}$ was called an *element* by D. Hilbert. We also have $\mathfrak{D}_{K/k}$ = $\mathfrak{E}^{(2)}\mathfrak{E}^{(3)}$. $\mathfrak{E}^{(n)}$. The integral ideal $\mathfrak{d}_{K/k}=N_{K/k}(\mathfrak{D}_{K/k})$ of k is called the **relative discriminant** of K/k. If $k = \mathbf{Q}$, $\mathfrak{d}_{K/\mathbf{Q}}=(D_K)$.

For the relative different $\mathfrak{D}_{K/k}$ to be divisible by a prime ideal $\mathfrak{P}$ of K, it is necessary and sufficient that E(p) = $\mathfrak{P}^e\mathfrak{P}_2^{e_2}$ $\mathfrak{P}_g^{e_g}$ with e >1, where p = $\mathfrak{P}\cap k$ **(Dedekind's discriminant theorem,** 1882). Hence a prime ideal p of k is ramilïed for K/k if and only if p divides the relative discriminant $\mathfrak{d}_{K/k}$; there are thus only a lïnite number of prime ideals of k which ramify for K/k. In particular, K/k is unramilïed if and only if $\mathfrak{d}_{K/k}$ = 0.

K. Arithmetic of Galois Extensions

Let K/k be a relative algebraic number field such that K is a †Galois extension of k of degree n, and let G be the †Galois group of K/k. Let o, $\mathfrak{O}$ be the principal order of k, K, respectively. The conjugate ideals of an ideal $\mathfrak{A}$ of K are given by $\mathfrak{A}^\sigma=\{A^\sigma\mid A\in\mathfrak{A}\}$ $(\sigma\in G)$. If $N_{K/k}(\mathfrak{A})$ = a, then E(a) = $\prod_{\sigma\in G}\mathfrak{A}^\sigma$. For a prime ideal p of K, $E(\mathfrak{p})=(\mathfrak{P}_1\mathfrak{P}_2\ldots\mathfrak{P}_g)^e$, where $N_{K/k}(\mathfrak{P}_i)=\mathfrak{p}^f$ $(i=1,\ldots.g)$, $n=efg$, and $\mathfrak{P}_1,\ldots,\mathfrak{P}_g$ are mutually conjugate prime ideals of K over k.

Hilbert (1894) developed the decomposition theory of a prime ideal p of k for a Galois extension K/k in terms of the Galois group G as follows: Let $\mathfrak{P}$ be a prime ideal of $\mathfrak{O}$. Then

$$Z=\{\sigma\in G\mid\mathfrak{P}^\sigma=\mathfrak{P}\}$$

is a subgroup of the Galois group G of K/k. Z is called the **decomposition group** of $\mathfrak{P}$ over k. Let G = $\bigcup_i Z\tau_i$ be the left coset decomposition of G. Then $\mathfrak{P}_i$ = $\mathfrak{P}^{\tau_i}$ (i = 1, , g) are all the conjugate ideals of $\mathfrak{P}$ over k.

The subgroup

$$T=\{\sigma\in Z\mid A^\sigma\equiv A(\mathrm{mod}\,\mathfrak{P}), A\in\mathfrak{O}\}$$

of the decomposition group Z is normal, and T is called the **inertia group** of $\mathfrak{P}$ over k. The quotient group Z/T is a cyclic group of order f (the relative degree of $\mathfrak{P}$). There exists an element σ of Z such that

$$A^\sigma\equiv A^{N(\mathfrak{p})}(\mathrm{mod}\,\mathfrak{P}),\quad A\in\mathfrak{O},$$

and σ is uniquely determined mod T; σT gen-

erates the cyclic group Z/T. This σ is called the **Frobenius substitution** (or **Frobenius automorphism)** of $\mathfrak{P}$ over k. For $m = 1, 2, $,

$$V^{(m)} = \{\sigma \in Z \mid A^{\sigma} \equiv A (\bmod \mathfrak{P}^{m+1}), A \in \mathfrak{O}\}$$

are normal subgroups of Z; the group $V^{(m)}$ is called the mth **ramification group** of $\mathfrak{P}$ over k. Let

$$V^{(0)} = \ldots = V^{(v_1)} \supsetneqq V^{(v_1+1)} = \ldots$$

$$= V^{(v_2)} \supsetneqq \ldots \supsetneqq V^{(v_{r-1}+1)} = \ldots$$

$$= V^{(v_r+1)} \supsetneqq V^{(v_r+1)} = 1$$

Let $V_\rho - V^{(v_\rho+1)}$ $(\rho = 0, 1, \ldots, r)$, where $v_0 = -1$. In particular, $V_0 = V^{(0)} = T$. The integers v_1, v_2, .. are called the **ramification numbers** of $\mathfrak{P}$. The group T/V_1 is isomorphic to a subgroup of the multiplicative group of the finite field $\mathfrak{O}/\mathfrak{P}$. Hence T/V_1 is a cyclic group whose order e_0 is a divisor of $N(\mathfrak{P}) - 1$. The group V_m/V_{m+1} $(m \geqslant 1)$ is isomorphic to a subgroup of the additive group of the finite field $\mathfrak{O}/\mathfrak{P}$. Hence V_m/V_{m+1} is an Abelian group of type $(p, p, $, p) whose order divides $N(\mathfrak{P})$. From $e = T| =$ (T: V_1)$|V_1$ it follows that $e = e_0 p^{\mu}, (e_0, p) = 1$. Here $G|$ denotes the order of a finite group G. Hence the decomposition group of $\mathfrak{P}$ is a †solvable group. The relation between the ramification numbers for K/k and those for an intermediate Galois extension F/k was completely determined by J. Herbrand (*J.* Math. *Pures Appl.*, 10 (1931)) [ll].

Let $\mathfrak{P}^d$ be the $\mathfrak{P}$-component of the relative different $\mathfrak{D}_{K/k}$ of a Galois extension K/k. Then

$$d = \sum_{\rho=0}^{r-1} (v_{\rho+1} - v_\rho)(|V_\rho| - 1) = \sum_{i=0}^{v_r} (|V^{(i)}| - 1).$$

In particular, $d = 0$ if $T = 1$, and $d = e$ 1 if $V^{(1)} = 1$.

Let k_Z, k_T, and $k_{V^{(m)}}$ be the intermediate fields which correspond to the subgroups Z, T, and $V^{(m)}$, respectively, in the sense of †Galois theory; the fields k_Z, k_T, and $k_{V^{(m)}}$ are called the **decomposition field,** the **inertia field,** and the mth **ramification field** of $\mathfrak{P}$, respectively. Let $\mathfrak{P}$ be a prime ideal of K containing $\mathfrak{p}$, and let $\mathfrak{p}_Z$ and $\mathfrak{p}_T$ be prime ideals in k_Z and k_T such that $\mathfrak{p}_Z = \mathfrak{P} \cap k_Z$ and $\mathfrak{p}_T = \mathfrak{P} \cap k_T$. Then we have $E(\mathfrak{p}) = \mathfrak{p}_Z \mathfrak{p}_Z^{(2)}$ $\mathfrak{p}_Z^{(g)}$ for k_Z/k; $E(\mathfrak{p}_Z) = \mathfrak{p}_T$ and $N_{k_T/k_Z}(\mathfrak{p}_T) = \mathfrak{p}_Z^f$ for k_T/k_Z; and $E(\mathfrak{p}_T) = \mathfrak{P}^e$ for K/k_T.

If a prime ideal p of k is unramified for a Galois extension K/k then we have E(p) = $\mathfrak{P}_1 \mathfrak{P}_2$ $\mathfrak{P}_g$, $\mathfrak{P}_i = \mathfrak{P}^{\tau_i}$ (i = 1, , g), $N_{N/k}(\mathfrak{P}_i)$ = $\mathfrak{p}^f$, and $n = fg$. The Frobenius automorphism $\sigma_i: A^{\sigma_i} \equiv A^{N(\mathfrak{p})} (\bmod \mathfrak{P}_i)$ (A $\in \mathfrak{O}$) for the prime ideal $\mathfrak{P}_i$ is uniquely determined, and its order is f. Since $\mathfrak{P}_i = \mathfrak{P}^{\tau_i}$, we have $\sigma_i = \tau_i^{-1} \sigma_1 \tau_i$. Hence σ_1, , σ_g belong to the same †conjugate class of G. In particular, if G is an Abelian group, then $\sigma_1 = .. = \sigma_g$ and

$$A^{\sigma_1} \equiv A^{N(\mathfrak{p})} (\bmod \mathfrak{p}), \quad A \in \mathfrak{O}.$$

We then write

$$\sigma_1 = \left(\frac{K/k}{\mathfrak{p}}\right) (\in G)$$

and call this symbol the **Artin symbol** for p for the Abelian extension K/k. For an ideal $\mathfrak{a} = \prod \mathfrak{p}^e$ of k that is relatively prime to the relative discriminant of K/k, *we* define

$$\left(\frac{K/k}{\mathfrak{a}}\right) = \prod \left(\frac{K/k}{\mathfrak{p}}\right)^e \quad (\in G)$$

Evidently, we have

$$\left(\frac{K/k}{\mathfrak{ab}}\right) = \left(\frac{K/k}{\mathfrak{a}}\right)\left(\frac{K/k}{\mathfrak{b}}\right).$$

The arithmetic of quadratic fields (→ 347 Quadratic Fields) and the arithmetic of cyclotomic fields (→ Section L) have been developed since the 19th Century.

L. Arithmetic of Cyclotomic Fields

A complex number ζ whose mth power is 1 but whose m'th power is not 1 for m' $< m$ is called an mth **primitive root of unity**. There are $\varphi(m)$ primitive roots of unity: $\exp(2\pi i r/m)$ ((r, m) = 1), where φ is †Euler's function. These $\varphi(m)$ primitive roots of unity are the zeros of an irreducible polynomial over Q of degree $\varphi(m)$:

$$F_m(X) = \prod_{d|m} (X^{m/d} - 1)^{\mu(d)},$$

where μ is the †Möbius function. The coefficient of the highest term of $F_m(X)$ is 1, and the other coefficients are all rational integers. $F_m(X)$ is called a **cyclotomic polynomial.** An example is

$$F_{12}(X) = (X^{12} - 1)(X^2 - 1)/(X^6 - 1)(X^4 - 1)$$

$$= X^4 - X^2 + 1.$$

The algebraic number field $K_m = \mathbf{Q}(\zeta_m)$ $(\zeta_m = \exp(2\pi i/m))$ obtained by adjoining an mth primitive root of unity to Q is a Galois extension over Q of degree $\varphi(m)$ whose Galois group G is isomorphic to the multiplicative Abelian group of †reduced residue classes of **Z** modulo $m: G = \{\sigma_r \mid \zeta_m^{\sigma_r} = \zeta_m^r, (r, m) = 1\}$. K_m is called the mth **cyclotomic field.** Cyclotomic fields are †Abelian extensions of Q. Conversely, every Abelian extension of Q is a subfield of a cyclotomic field **(Kronecker's theorem, 18.53,** 1877).

We can choose $(1, \zeta_m, \zeta_m^2, \ldots, \zeta_m^{\varphi(m)-1})$ as a minimal basis of K_m. Let $m = l_1^{h_1} l_2^{h_2}$ $l_t^{h_t}$ be the decompositions of m in powers of prime numbers l_1, , l_t. Put $K^{(i)} = K_{l_i^{h_i}}$. Then K_m is the composite field $K_m = K^{(1)} K^{(2)} \ldots K^{(t)}$. The different of K_m is given by $\mathfrak{D}_{K_m/\mathbf{Q}}$ =

$\mathfrak{D}_{K^{(1)}/\mathbf{Q}}\mathfrak{D}_{K^{(2)}/\mathbf{Q}}\ldots\mathfrak{D}_{K^{(t)}/\mathbf{Q}}$, and the discriminant of K_m is given by $D_{K_m}=D_{K^{(1)}}^{n_1}\ldots D_{K^{(t)}}^{n_t}$ (n, = $\varphi(m)/\varphi(l_i^{h_i})$). If $m=l^h$, the discriminant of K_m is $D_{K_m}=\varepsilon l^a$ $(a=l^{h-1}(hl-h-1))$, where $\varepsilon=-1$ if $l^h=4$ or $l\equiv 3\ (\mathrm{mod}\ 4)$, and $\varepsilon=1$ otherwise. Hence the discriminant of K_m is $D_{K_m}=(\sqrt{-1}\ m/\prod_{p|m}p^{1/(p-1)})^{\varphi(m)}$. Suppose that $m\not\equiv 2$ (mod 4). Then a prime number p ramifies for $K_m/\mathbf{Q}$ if and only if p divides m. In particular, if $m=l^h$ $(m>2)$, then only 1 ramifies for $K_m/\mathbf{Q}$ and $(l)=\mathfrak{l}^{\varphi(m)}$, N(I) = l ([is explicitly given by $\mathfrak{l}=$ (l-i,,,)). For $l\neq 2$ the ramification numbers for l are $v_i=l^i$ 1 (i = 1,2,), and the ramification fields are K_l, K_{l^2}, For $l=2$ the ramification numbers are $1,3,7$, and the ramification fields are $\mathbf{Q}$, K_4, K_8,

In $K_m/\mathbf{Q}$ a prime number p $(p\nmid m)$ is decomposed as $(p)=\mathfrak{p}_1\mathfrak{p}_g$, $N(\mathfrak{p}_i)=p^f(i=1,\ldots.$ y) and $fg=qn(m)$. Here the degree f of $\mathfrak{p}_i$ is determined as the minimal positive integer f such that $p^f\equiv 1$ (mod m). Hence the decomposition law of a prime number p in $K_m/\mathbf{Q}$ is derermined by its residue class modulo m. This is a prototype form of class tield theory (- 59 Class Field Theory).

The class number of the cyclotomic tield K_m can be calculated by Dedekind's formula (- Section E; see also Hilbert [4]). Here we shall give the result for $m=l$ (a prime number).
Let r be a +Primitive root modulo l. For $\zeta=\exp(2\pi i/l)$, we put

$$\varepsilon=\varepsilon(\zeta)=\left(\frac{1-\zeta^r}{1-\zeta}\frac{1-\zeta^{-r}}{1-\zeta^{-1}}\right)^{1/2}.$$

Then ε is a unit in K_l. Define an element σ of the Galois group of $K_l/\mathbf{Q}$ by $\zeta^\sigma=\zeta^r$, and put $\varepsilon_i=\varepsilon^{\sigma^i}$ $(i=0,1,\ldots)$. Then $\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{\rho-1}$ $(\rho=(l-3)/2)$ are multiplicatively independent units. That is, the regulator $R[\varepsilon_0,\varepsilon_1,,\varepsilon_{\rho-1}]=E\neq 0$. The units $\varepsilon_0,\varepsilon_1,\ldots.\ \varepsilon_{\rho-1}$ are called **circular units.** The class number h of K_l is the product of two factors, $h=h_1h_2$. Here h_2 is the class number of the real subfield $K_l'=\mathbf{Q}(\zeta+\zeta^{-1})$. E. E. Kummer called h_1 the **first factor** and h_2 the **second factor** of the class number h. Let $\chi_1,\chi_2,,\chi_{l-1}$ be the multiplicative characters of the treduced residue classes of $\mathbf{Z}$ modulo 1, and let $\chi_i(i=1,,\rho+1)$ be the characters among them such that $\chi_i(-1)=-1$. Then

$$h_1=\frac{(-1)^{\rho+1}}{(2l)^\rho}\prod_{i=1}^{\rho+1}\left(\sum_{a=1}^{l-1}a\chi_i(a)\right),\quad h_2=\frac{|E|}{R_0}$$

(Kummer, 1850). Here R_0 is the regulator of K_l'. Since circular units belong to K_l', the class number h_2 of K_l' is equal to the index of the subgroup generated by ± 1, $\varepsilon_0,,\varepsilon_{\rho-1}$ in the group of units of K_l'. The class number h of K_l is equal to 1 for $l\leqslant 19$ and it has been conjectured that there exist no more fields K_l with $h=1$. H. L. Montgomery and K. Uchida solved this conjecture by proving that the first factor $h_1>1$ for $l>19$ (*Tôhoku Math. J.*, **23** (1971)). J. M. Masley and Montgomery proved that there are precisely 29 distinct K_m with class number 1 (*J. Reine Angew. Math.*, 286/287 (1976)).

According to Kummer, l divides h if and only if l divides h,. Since h_1 can be computed explicitly, we can readily determine whether $l|h$ or not (- 145 Fermat's Problem; Appendix B, Table 4.111). A prime l is called **regular** if $l\nmid h$, otherwise it is called **irregular.** Thus an odd prime $l\geqslant 5$ is irregular if and only if there exists an even integer j with $2\leqslant j\leqslant l-3$ such that l divides the numerator of the jth Bernoulli number B_j. This criterion of Kummer can be strengthened as follows. Let S denote the l-Sylow subgroup of the ideal class group $\mathfrak{C}_{K_l}$ of the prime cyclotomic tield $K_l=\mathbf{Q}(\zeta_l)$. Because the Galois group G of $\mathbf{Q}(\zeta_l)/\mathbf{Q}$ operates on S, the Abelian group S becomes naturally a $\mathbf{Z}_l[G]$-module. Choosing the canonical character $\theta:G\to\mathbf{Z}_l^*$ defined by $\zeta_l^\sigma=\zeta_l^{\theta(\sigma)}$ for any $\sigma\in$ G, we have a direct decomposition $S=\prod_{i=0}^{l-2}S^{(i)}$, where $S^{(i)}=\{s\in S\,|\,s^\sigma=s^{\theta^i(\sigma)},\ \sigma\in\mathrm{G}\}$. Then it is necessary and sufficient for $l|$ (the numerator of B_j) that $S^{(l-j)}\neq 0$ (K. A. Ribet, *Inventiones Math.*, *34* (1976)). Moreover, because G is naturally isomorphic to the multiplicative group of the reduced residue classes of $\mathbf{Z}/l\mathbf{Z}$, each θ^i becomes a †Dirichlet character. Let

$$B_{1,\theta^{-i}}=\frac{1}{l}\sum_{a=1}^{l-1}\theta^{-i}(a)a;$$

then this number, regarded as an element in $\mathbf{Q}_l$, is equal to a factor of the product (up to ± 1 and 2) appearing in the class number formula for h,. Let m_i be the l-exponent of $B_{1,\theta^{-i}}$ for each odd i $(2<i\leqslant l-3)$. Then it is conjectured that the order of $S^{(i)}$ is precisely equal to l^{m_i} for each odd i with $2<i\leqslant l-3$. In particular, if $S'=\prod_{j;\,\mathrm{even}}S^{(j)}$ is trivial, then this is known to be true. In general, it is also conjectured that $l\nmid h_2$ holds for any prime l **(Vandiver's conjecture).** If the group $S^{(i)}$ is cychc, then $S^{(i)}$ is of order l^{m_i} (A. Wiles, *Inventiones Math.*, ***58*** (1980)).

When a Galois extension K over a finite algebraic number field k has the Galois group isomorphic to the additive group of the l-adic integer ring $\mathbf{Z}_l$, the extension K/k is called a **$\mathbf{Z}_l$-extension** or **r-extension.** Then for each integer $n\geqslant 0$ there exists a unique subfield k_n of K with degree l^n over k. Let $k=K_l=\mathbf{Q}(\zeta_l)$ $(l>2)$, and let K be the union of all l^{n+1} th cyclotomic fields $K_{l^{n+1}}$ $(n\geqslant 0)$. Then K/k is a typical example of a $\mathbf{Z}_l$-extension with subfields $k_n=K_{l^{n+1}}$. Let K/k be a $\mathbf{Z}_l$-extension with sub-

fields k_n, and let l^{e_n} denote the l-component of the class number of k_n. Then there exist integers $\lambda \geqslant 0$, $\mu \geqslant 0$, v independently of n such that $e_n = \lambda n + \mu l^n + \nu$ for all suffciently large numbers n (K. Iwasawa, Bull. Amer. Math. Soc., 65 (1959)). These numbers λ, μ, **v** are called the **Iwasawa invariants** for the Z,-extension K/k. There exists a unique $\mathbf{Z}_l$-extension K° over the rationals **Q**. An extension $K^\circ k$ over k obtained by taking the composite of K° and k is called the **basic $\mathbf{Z}_l$-extension** of k or the **cyclotomic $\mathbf{Z}_l$-extension** of k. We denote its Iwasawa invariants by $\lambda_l(k)$, $\mu_l(k)$, $\nu_l(k)$. It is known that there are $\mathbf{Z}_l$-extensions with arbitrarily large μ. But it is conjectured that $\mu_l(k) = 0$ holds for the basic Z,-extension of any k. In particular, in the case $k = K_l = \mathbf{Q}(\zeta_l)$ it has been computed that $\mu_l(k) = 0$ for $l < 125{,}000$ (for $l < 4001$ by K. Iwasawa and C. C. Sims, *J.* Math. Soc. *Japan*, 18 (1966); for $l < 30{,}000$ by W. Johnson, *Math. Comp., 29* (1975); for $l <$ 125,000 by S. Wagstaff, *Math. Comp.*, 32 (1978)). By applying the theory of uniform distributions to l-adic situations, B. Ferrero and L. C. Washington (Ann. *Math.*, 109 (1978)) proved that $\mu_l(k) = 0$ when $k/\mathbf{Q}$ is Abelian. In particular, when $k = K_l$ and $k_n = K_{l^{n+1}}$, $e_n > 0$ if and only if the class number of K_l is divisible by l (P. Furtwangler, 1911).

Since any quadratic field is a subfield of a cyclotomic field (by a †Gaussian sum formula we have $\mathbf{Q}(\sqrt{m}) \subset \mathbf{Q}(\zeta_{|d|})$, where d is the discriminant of $\mathbf{Q}(\sqrt{m})$), the computation of the class number of quadratic fields and the proof of the law of reciprocity for the †Legendre symbol follow from the arithmetic of cyclotomic fields.

M. Aritbmetic of Kummer Extensions

Assume that an algebraic number field k contains an nth primitive root of unity. Then a †Kummer extension $K = k(\sqrt[n]{\mu})$ $(\mu \in k)$ is a †cyclic extension of k. Assume that $[K: k] = n$. In order that a prime ideal p of k ramify for K/k, it is necessary that $\mathfrak{p}$ be a divisor appearing in (n) or (μ). If $\mathfrak{p} \nmid (n)$ and $\nu_\mathfrak{p}(\mu) \not\equiv 0$ (modn), then p ramifies for K/k. A prime ideal $\mathfrak{p}$ which is relatively prime to (μ) has the decomposition E(p) = $\mathfrak{P}_1 \ \mathfrak{P}_n$ with distinct primes $\mathfrak{P}_i$ in K if and only if the equation $\mu \equiv \xi^n$ (mod p") is satisfied by some $\xi \in$ o for any positive integer m. In particular, if $p \nmid (n)$ and $\nu_\mathfrak{p}(\mu) = 0$, we have E(p) = $\mathfrak{P}_1 \ \mathfrak{P}_n$ if and only if $\mu \equiv \xi^n$ (mod p) is solvable in o.

If for an element μ of 0

$$\mu \equiv \xi^n \pmod{\mathfrak{p}}$$

is solvable by some $\xi \in$ o, μ is said to be a residue of tbe ntb power modulo p. Assume that $\mathfrak{p} \nmid (n)$ and $\nu_\mathfrak{p}(\mu) = 0$. Let f be the minimal positive integer such that μ^f is a residue of the nth power modulo p. Then p is decomposed in K as $E(\mathfrak{p}) = \mathfrak{P}_1 \ \mathfrak{P}_g$, and $N_{K/k}(\mathfrak{P}_i) = \mathfrak{p}^f$ $(i = 1, \ldots, g)$.

N. Power-Residue Symbol

Let $\zeta_n = \exp(2\pi i/n) \in k$, and let p be relatively prime to (n) and (α) $(\alpha \in k)$. Then for some r we have

$$\alpha^{(N(\mathfrak{p})-1)/n} \equiv \zeta_n^r \pmod{\mathfrak{p}},$$

and we write

$$\zeta_n^r = \left(\frac{\alpha}{\mathfrak{p}}\right)_n$$

This symbol is called the nth **power-residue symbol** (Kummer). Generalizing this definition, we can define the nth power-residue symbol $(\alpha/\mathfrak{b})_n$ for an ideal b of k which is relatively prime to the relative discriminant of $k(\sqrt[n]{\alpha})/k$ by using the †Artin symbol $((K/k)/\mathfrak{b})$:

$$(\sqrt[n]{\alpha})^\sigma = \left(\frac{\alpha}{\mathfrak{b}}\right)_n \sqrt[n]{\alpha}, \quad \sigma = \left(\frac{K/k}{\mathfrak{b}}\right).$$

This symbol satisfies

$$\left(\frac{\alpha}{\mathfrak{b}_1\mathfrak{b}_2}\right)_n = \left(\frac{\alpha}{\mathfrak{b}_1}\right)_n \left(\frac{\alpha}{\mathfrak{b}_2}\right)_n,$$

$$\left(\frac{\alpha_1\alpha_2}{\mathfrak{b}}\right)_n = \left(\frac{\alpha_1}{\mathfrak{b}}\right)_n \left(\frac{\alpha_2}{\mathfrak{b}}\right)_n$$

if all the symbols are well delïned. In particular, α is a residue of the nth power modulo $\mathfrak{p}$ if and only if $(\alpha/\mathfrak{p})_n = 1$. This symbol coincides with the †quadratic residue symbol for $n = 2$, $k = \mathbf{Q}$, and $p \neq 2$.

0. Law of Reciprocity for tbe Power-Residue Symbol

Several formulas concerning the power-residue symbol are known which are similar to that for the quadratic residue symbol (F. G. M. Eisenstein, Kummer, Furtwangler, Takagi, Artin, Hasse). These can be proved by means of Artin's †general law of reciprocity in class field theory (→ 59 Class Field Theory).

There are many formulas concerning the reciprocity of the power-residue symbol. One of them is as follows: Let $n = l$ be a prime number. Let $\alpha, \beta \in k$ and assume that (i) α is totally positive; (ii) $\nu_\mathfrak{p}(\alpha) \equiv 0 \pmod{l}$ if $\nu_\mathfrak{p}(\beta) \not\equiv 0 \pmod{l}$, and $\nu_\mathfrak{p}(\beta) \equiv 0 \pmod{l}$ if $\nu_\mathfrak{p}(\alpha) \not\equiv 0 \pmod{l}$, for any prime ideal p; and (iii) $a \equiv 1 \pmod{l}$ and $\beta \equiv 1 \pmod{(1 - \zeta_l)}$. Then

$$\left(\frac{\alpha}{\beta}\right)_l\left(\frac{\beta}{\alpha}\right)_l^{-1}=\zeta_l^a,$$

$$a=Tr_{k/\mathbf{Q}}\left(\frac{\alpha-1}{l}\frac{\beta-1}{1-\zeta_l}\right)$$

(**general law of reciprocity,** Hasse, 1924). This result is a generalization of the formula of Eisenstein (1850). If $\alpha(\in k)$ is totally positive and $\alpha\equiv 1 \pmod{l}$, then

$$\left(\frac{\zeta_l}{\alpha}\right)_l=\zeta_l^b,\quad b=Tr_{k/\mathbf{Q}}\left(\frac{\alpha-1}{l}\right).$$

If $\alpha\equiv 1 \pmod{l(1-\zeta_l)}$, then

$$\left(\frac{l}{\alpha}\right)_l=\zeta_l^c,\quad c=Tr_{k/\mathbf{Q}}\left(\frac{\alpha-1}{l(1-\zeta_l)}\right);$$

$$\left(\frac{1-\zeta_l}{\alpha}\right)_l=\zeta_l^d,\quad d=-Tr_{k/\mathbf{Q}}\left(\frac{\alpha-1}{l(1-\zeta_l)}\right)$$

(**complementary law of reciprocity,** Hasse, 1924).

P. Norm Residue

Let $\mathfrak{m}$ be an integral divisor of k such that $\mathfrak{m}=\prod_i\mathfrak{p}_i^{e_i}\prod_j\mathfrak{p}_\infty^{(j)}$ $(e_i>0)$ with finite prime divisors $\{\mathfrak{p}_i\}$ and infinite prime divisors $\{\mathfrak{p}_\infty^{(j)}\}$, and let β be an element of k that is relatively prime to $\mathfrak{m}$. For a relative algebraic number field K/k and an element B of K, we set $\beta\equiv N_{K/k}(B) \pmod{\mathfrak{m}}$ if the following two conditions are satisfied: (i) $\beta\equiv N_{K/k}(B) \pmod{\mathfrak{p}_i^{e_i}}$ for every finite $\mathfrak{p}_i$ and (ii) $\beta^{(j)}>0$ for every infinite prime $\mathfrak{p}_\infty^{(j)}$ such that $\mathfrak{p}_\infty^{(j)}$ is real and its extension to K is imaginary. $\beta\in k$ is then said to be a **norm residue** modulo $\mathfrak{m}$ for K/k if there exists a number B of K such that $\beta\equiv N_{K/k}(B) \pmod{\mathfrak{m}}$.

Let $\mathfrak{p}$ be a finite prime divisor of k. If β is a norm residue modulo $\mathfrak{p}^c$ for a sufficiently large c, then β is a norm residue modulo $\mathfrak{p}^e$ for any $e>c$. Let c be the smallest such integer $(c\geqslant 0)$. Then the ideal $\mathfrak{f}_\mathfrak{p}=\mathfrak{p}^c$ is said to be the $\mathfrak{p}$-**conductor** of norm residue for K/k. If $\mathfrak{p}$ is unramified for K/k, then $c=0$; i.e., every $\beta\in k$ which is relatively prime to $\mathfrak{p}$ is a norm residue modulo $\mathfrak{p}^e$ for any $e>0$. For a ramified $\mathfrak{p}$ put $\mathfrak{f}_\mathfrak{p}=\mathfrak{p}^c$. For a Galois extension K/k, c is not greater than

$$\sum_{\rho=0}^{r-1}\frac{|V_\rho|}{|V_0|}(v_{\rho+1}-v_\rho)=\sum_{i=0}^{v_r}\frac{|V^{(i)}|}{|V_0|}.$$

In particular, for an Abelian extension K/k this value is an integer and is equal to c (Hasse, *J. Fac. Sci. Univ. Tokyo,* 1934). For example, the l-conductor of the cyclotomic field $K_{l^h}/\mathbf{Q}$ is l^h. We define the $\mathfrak{p}_\infty$-conductor for K/k for an infinite prime divisor $\mathfrak{p}_\infty$ of k by $\mathfrak{f}_{\mathfrak{p}_\infty}=\mathfrak{p}_\infty$ if $\mathfrak{p}_\infty$ is real and its extension to K is imaginary, and $\mathfrak{f}_{\mathfrak{p}_\infty}=1$ otherwise (→ 257 Local Fields F).

Q. Norm Residue Symbol

For an Abelian extension K/k, the positive divisor

$$\mathfrak{f}=\prod_\mathfrak{p}\mathfrak{f}_\mathfrak{p}$$

(where $\mathfrak{p}$ runs over all finite and infinite prime divisors of k) is called the **conductor** of K/k (→ 59 Class Field Theory). For $a\in k$ $(a\neq 0)$ take α_0 such that $\alpha/\alpha_0\equiv 1 \pmod{\mathfrak{f}_\mathfrak{p}}$ and $\alpha_0\equiv 1 \pmod{\mathfrak{f}\mathfrak{f}_\mathfrak{p}^{-1}}$, and put $(\alpha_0)=\mathfrak{p}^a\mathfrak{b}$ with $\mathfrak{b}$ relatively prime to $\mathfrak{p}$. Then $\mathfrak{b}$ is relatively prime to the relative discriminant $\mathfrak{d}_{K/k}$. We define a new symbol by

$$\left(\frac{\alpha,K/k}{\mathfrak{p}}\right)=\left(\frac{K/k}{\mathfrak{b}}\right)\quad(\in G),$$

where $((K/k)/\mathfrak{b})$ is the Artin symbol. This value is independent of the choice of the auxiliary element α_0. The new symbol is called the **norm-residue symbol** (Hasse, *J. Reine Angew.* Math., 162 (1930)). In particular, for an infinite real prime divisor $\mathfrak{p}_\infty^{(j)}$ of k whose extension $\mathfrak{P}_\infty^{(j)}$ for K is imaginary, we have

$$\left(\frac{\alpha,K/k}{\mathfrak{p}_\infty^{(j)}}\right)=1\text{ or }=\sigma$$

according to whether the conjugate $\alpha^{(j)}$ is positive or negative, where σ is the automorphism of K/k induced from the complex conjugation of the completion C of K with respect to $\mathfrak{P}_\infty^{(j)}$.

The norm-residue symbol has the following properties:

(1) $\left(\frac{\alpha\alpha',K/k}{\mathfrak{p}}\right)=\left(\frac{\alpha,K/k}{\mathfrak{p}}\right)\left(\frac{\alpha',K/k}{\mathfrak{p}}\right)$;

(2) if $\mathfrak{p}$ is unramified for K/k, then

$$\left(\frac{\alpha,K/k}{\mathfrak{p}}\right)=\left(\frac{K/k}{\mathfrak{p}}\right)^{-v_\mathfrak{p}(\alpha)};$$

(3) in order that α be a norm residue modulo $\mathfrak{f}_\mathfrak{p}$ for K/k, it is necessary and sufficient that

$$\left(\frac{\alpha,K/k}{\mathfrak{p}}\right)=1;$$

(4) the **product formula** for the norm-residue symbol (Hasse) is

$$\prod_\mathfrak{p}\left(\frac{\alpha,K/k}{\mathfrak{p}}\right)=1,$$

where $\mathfrak{p}$ runs over all finite and infinite prime divisors of k; and (5) if the domain of the variable α is the whole k $(\neq 0)$, or the set of all α such that (α) is relatively prime to $\mathfrak{p}$, or the set of all α such that $\alpha\equiv 1 \pmod{\mathfrak{p}^m}$ $(v_\rho+1\leqslant m\leqslant v_{\rho+1})$, then the range of $((\alpha, K/k)/\mathfrak{p})$ is the decomposition group Z of $\mathfrak{p}$, the inertia group T of $\mathfrak{p}$, or the ramification group V_ρ

of $\mathfrak{p}$, respectively (Hasse, S. Iyanaga, 1933) (→ 257 Local Fields F).

R. Hilbert Norm-Residue Symbol

The symbol first introduced by Hilbert for quadratic fields can be defined in a general algebraic number field k containing an nth primitive root ζ_n of unity. Let $\alpha, \beta \in k$ ($\alpha \neq 0$, $\beta \neq 0$), and let $\mathfrak{p}$ be a prime divisor. Then the following nth root of unity $((\alpha, \beta)/\mathfrak{p})_n$ is defined by using the norm-residue symbol:

$$\left(\frac{\alpha,\beta}{\mathfrak{p}}\right)_n \sqrt[n]{\beta} = (\sqrt[n]{\beta})^{\sigma},$$

$$\sigma = \left(\frac{\alpha, k(\sqrt[n]{\beta})/k}{\mathfrak{p}}\right).$$

This symbol $((\alpha, \beta)/\mathfrak{p})_n$ is called the **Hilbert norm-residue symbol.** It is also called the **Hilbert-Hasse norm-residue symbol.** For α, α', $\beta \in k$, we have

(1) $$\left(\frac{\alpha\alpha',\beta}{\mathfrak{p}}\right)_n = \left(\frac{\alpha,\beta}{\mathfrak{p}}\right)_n \left(\frac{\alpha',\beta}{\mathfrak{p}}\right)_n;$$

(2) the **law of symmetry:**

$$\left(\frac{\alpha,\beta}{\mathfrak{p}}\right)_n = \left(\frac{\beta,\alpha}{\mathfrak{p}}\right)_n^{-1};$$

and **(3)** the **product formula** for the Hilbert norm-residue symbol:

$$\prod_{\mathfrak{p}} \left(\frac{\alpha,\beta}{\mathfrak{p}}\right)_n = 1$$

(Hilbert, Furtwängler, Takagi, Artin, Hasse). For detailed properties concerning the norm-residue symbol, power-residue symbol, and Hilbert norm-residue symbol and for references for them see Hasse [6].

In general, the problem of obtaining various laws of reciprocity for the power-residue symbol is reduced to the one of computing the Hilbert norm-residue symbol explicitly. Detailed formulas for these symbols are called **explicit reciprocity laws;** they are treated as a topic in the number theory of local fields (→ 257 Local Fields H).

S. Density Theorem

Let M be a set of prime ideals of k. If

$$\lim_{s \to 1+0} \sum_{\mathfrak{p} \in M} \frac{1}{(N(\mathfrak{p}))^s} \bigg/ \log\frac{1}{s-1}$$

exists, its value is said to be the **density** of M. The density of the set of all prime ideals of k is 1. Let H be an ideal group modulo an integral divisor $\mathfrak{m}$. Then the density of the set of all prime ideals contained in each coset of $\mathfrak{J}(\mathfrak{m})$ modulo H is $1/(\mathfrak{J}(\mathfrak{m}): H)$. In particular, let H be the ray $S(\mathfrak{m})$. Then this result implies that each coset of $\mathfrak{J}(\mathfrak{m})$ modulo $S(\mathfrak{m})$ contains infinitely many prime ideals (a generalization to algebraic number fields of the †prime number theorem for arithmetic progression).

Let K/k be a Galois extension, C be a conjugate class of the Galois group G of K/k, and M(C) be the set of all prime ideals $\mathfrak{p}$ of k such that the †Frobenius automorphism of each prime factor $\mathfrak{P}_i$ of $\mathfrak{p}$ in K belongs to C. Then the density of **M(C)** is $|C|/|G|$ **(Chebotarev's density theorem, Math.** *Ann.,* 95 (1926)).

Each element σ of the Galois group G of K/k can be expressed by the permutation z of the conjugate fields $K^{(1)}, \ldots, K^{(n)}$ of K over k. Let z be expressed as the product of γ †cycles of length $f_1, \ldots, f_r$. Hence $n = f_1 + \ldots + f_r$. Let $C(f_1, \ldots, f_r)$ be the set of all such σ in G, and let $M(f_1, \ldots, f_r)$ be the set of all prime ideals $\mathfrak{p}$ of k such that $\mathfrak{p}$ is decomposed in K/k as the product of r prime ideals of K with relative degree $f_1, \ldots, f_r$. Then the density of $M(f_1, \ldots, f_r)$ is $C(f_1, \ldots, f_r)|/|G|$ (Artin, *Math. Ann.,* 89 (1923)).

T. Relation to the Arithmetic of Local Fields

It is quite useful to investigate the relation between the arithmetic of algebraic number fields and that of local fields. For example, let a prime ideal $\mathfrak{p}$ of an algebraic number field k be decomposed as $E(\mathfrak{p}) = \mathfrak{P}_1^{e_1} \ldots \mathfrak{P}_g^{e_g}$, $N_{K/k}(\mathfrak{P}_i) = \mathfrak{p}^{f_i}$ $(i = 1, \ldots, g)$ in an extension K of k. Let $K_{\mathfrak{P}}$ and $k_{\mathfrak{p}}$ be the completion of K and k with respect to $\mathfrak{P}$-adic and $\mathfrak{p}$-adic valuations, respectively. Then we have $[K_{\mathfrak{P}_i}: k_{\mathfrak{p}}] = e_i f_i$ and $K \otimes_K k_{\mathfrak{p}} \cong K_{\mathfrak{P}_1} + \ldots + K_{\mathfrak{P}_g}$ (direct sum). The relative different $\mathfrak{D}_{K/k}$ is expressed as (the $\mathfrak{p}$-component of $\mathfrak{D}_{K/k}$) = $\prod_{i=1}^{g} \mathfrak{D}_{K_{\mathfrak{P}_i}/k_{\mathfrak{p}}}$. For a Galois extension K/k the $\mathfrak{p}$-conductor $\mathfrak{f}_{\mathfrak{p}} = \mathfrak{p}^c$ for the norm-residue and the conductor $\mathfrak{p}^c$ of local extension $K_{\mathfrak{P}}/k_{\mathfrak{p}}$ have the same exponent c. For a local field $K_{\mathfrak{P}}/k_{\mathfrak{p}}$, each norm-residue modulo $\mathfrak{f}_{\mathfrak{p}}$ is a norm of an element of $K_{\mathfrak{P}}$. Hence precise results concerning the norm-residue in local fields can be applied immediately to a global field K/k (→ 257 Local Fields).

We can also apply the method of the idele group of an algebraic number field k, and therefrom we can prove results concerning the ideal class group, unit group, and zeta function of k (→ 6 Adeles and and Ideles).

U. History of the Arithmetic of Algebraic Number Fields

C. F. Gauss (1832) was the first to generalize the notion of integers to algebraic number

fields in considering the elements of $\mathbf{Z}[\sqrt{-1}]$, now called **Gaussian integers** ($\mathbf{Z}[\sqrt{-1}]$ is the principal order of $\mathbf{Q}(\sqrt{-1})$). After investigations by G. L. Dirichlet and Kummer, the notion of ideals was introduced by Dedekind (1871) [2]. L. Kronecker gave another foundation for the arithmetic of algebraic number fields (1882) [3]. Dirichlet proved the unit theorem and, introducing the analytic method to number theory, gave the class number formula of quadratic fields (→ 347 Quadratic Fields). H. Minkowski first applied the theory of lattice points to number theory (→ 182 Geometry of Numbers), and K. Hensel introduced the p-adic method (→ 257 Local Fields). Hilbert (1897) [4] and Hasse (1926, 1927, 1930) [6] summarized the main results on the arithmetic of algebraic number fields known at that time. In particular, Hilbert's report centered around the arithmetic of Galois extensions, and Hasse's around the class field theory obtained by T. Takagi, E. Artin, and H. Hasse (→ 59 Class Field Theory). Since c. 1950, when the notions of ideles and adeles were introduced, cohomology-theoretic methods have been successfully applied to number theory (→ 6 Adeles and Ideles). Recently various local methods, for example, the Iwasawa theory of $\mathbf{Z}_l$-extensions, †formal groups, and †p-adic L-functions (T. Kubota and H. W. Leopoldt, *J. Reine Angew. Math.*, 214/215 (1964)) have been frequently applied to research in algebraic number fields.

References

[1] K. F. Gauss, Theoria residuorum biquadraticorum 1 (1825), II (1831), Werke 2, Göttingen, 1863, 65–148.
[2] R. Dedekind, Dirichlet's Vorlesungen über Zahlentheorie, fourth edition, 1894, Supplement XI (Gesammelte mathematische Werke 3, Braunschweig, 1932) (Chelsea, 1969).
[3] L. Kronecker, Grundzüge einer arithmetischen Theorie der algebraischen Grosse, J. Reine Angew. Math., 92 (1882), 1-122 (Werke 2, Teubner, 1897, 237–387) (Chelsea, 1968).
[4] D. Hilbert, Die Theorie der algebraischen Zahlkörper, Jber. Deutsch. Math. Verein. 4, (1897), 175-546 (Gesammelte Abhandlungen 1, Springer, 1932, 63-363) (Chelsea, 1967).
[5] E. Hecke, Vorlesungen über die Theorie der algebraischen Zahlen, Akademische Verlag, 1923 (Chelsea, 1970).
[6] H. Hasse, Bericht über neuere Untersuchungen und Probleme aus der Theorie der algebraischen Zahlkörper, Jber. Deutsch. Math. Verein., 1, 35 (1926), l-55; Ia, 36 (1927), 233–311; II (1930) (Physica Verlag, 1965).
[7] H. Weyl, Algebraic theory of numbers, Ann. Math. Studies, Princeton Univ. Press, 1940.
[8] H. Hasse, Zahlentheorie, Akademie-Verlag, third edition, 1969; English translation, Number theory, Springer, 1980.
[9] E. Weiss, Algebraic number theory, McGraw-Hill, 1963.
[10] S. Lang, Algebraic number theory, Addison-Wesley, 1970.
[11] E. Artin, Algebraic numbers and algebraic functions, Lecture notes, Princeton Univ., 1950–1951 (Gordon & Breach, 1967).
[12] S. Iyanaga (ed.), Theory of numbers, North-Holland, 1975. (Original in Japanese, 1969.)
[13] A. Weil, Basic number theory, Springer, second edition, 1973.
[14] J. W. S. Cassels and A. Fröhlich (eds.), Algebraic number theory, Academic Press, 1967.
[15] H. Hasse, Über die Klassenzahl abelscher Zahlkörper, Akademie-Verlag, 1952.
[16] Z. 1. Borevich and 1. R. Shafarevich, Number theory, Academic Press, 1966. (Original in Russian, 1964.)
[17] S. Lang, Cyclotomic fields, Springer, 1978, II, 1980.

15 (VIII.3) Algebraic Surfaces

A. Definition

An algebraic variety of dimension 2 is called an **algebraic surface.** In this article, by a surface we mean a complete irreducible algebraic surface defined over an algebraically closed field K.

B. History

The history of algebraic surfaces originated with the study of algebraic functions of two variables. In the case of algebraic functions of one variable, the introduction of †Riemann surfaces attached to such functions played an essential role in the development of the theory. The study of algebraic functions of two variables led naturally to the consideration of the surfaces defined by a suitable polynomial equation. H. Poincaré and E. Picard are among those who studied the homological structure of the surface defined by the equation $P(x, y, z) = 0$. The theory of †Abelian integrals (Picard integrals) is one of the consequences of

such topological investigations. S. Lefschetz obtained further results in this direction.

M. Noether and geometers of the Italian school, such as F. Enriques, G. Castelnuovo, and F. Severi, studied algebrogeometric properties of algebraic surfaces. In particular, the Italian school geometers recognized the importance of irregularity and thoroughly investigated its geometric meaning. In the early 20th Century they succeeded in constructing the great edifice of the theory of algebraic surfaces. Though some of their results lack rigorous proof, efforts to build a foundation for those results have led to the recent developments in algebraic geometry. A significant contribution to the modernization of the theory was made by 0. Zariski and K. Kodaira.

The resolution of singularities of an algebraic surface is one of the most fundamental problems in the field. When the †universal domain is the complex number field, function-theoretic methods were used by Italian-school geometers and R. J. Walker (Ann. Math., 36 (1935)). Zariski introduced the †valuation-theoretic method to deal with the problem when the characteristic of the universal domain is zero. S. Abhyankar (1966) succeeded in resolving the case of positive characteristics.

C. Divisors and Linear Systems

In what follows, let S denote a nonsingular surface. S can be embedded into some projective space. Let Σ be a linear system of divisors on S and $f_0, , f_n$ be a basis of the defining module for Σ over **K.** Associated with Σ we have a rational mapping $\Phi_\Sigma : S \to \mathbf{P}^n$ defined by $\Phi_\Sigma(P) = (f_0(P), \ldots, f_n(P))$ for general points P on S. Pullbacks of hyperplanes by Φ_Σ are called **variable components** of Σ. Any member of Σ is a sum of a variable component and a fixed component of Σ. Let S' denote the closure of the image of S by Φ_Σ. If dim S' = 2, a general variable component is irreducible. If dim S' = 1, then a general variable component is composed of an algebraic system of dimension 1, which is called an **algebraic pencil.** These result from Bertini's theorems. For any divisors D and D' on S, the **intersection number** (or the †Kronecker index) $I(D \cdot D')$ is defmed; this number is a symmetric bilinear form such that $I(D \cdot D) = I(D_1 \cdot D_1')$ for any divisors D_1 and D_1' linearly equivalent to D and D', respectively. If C is a nonsingular curve on S, $I(C \cdot D)$ coincides with the degree of the restriction $C \cdot D$ of D to C. $I(D \cdot D) = (D^2)$ issaidtobethe **self-intersection number** of D. If D is an ample divisor on S, then $(D^2) > 0$ and $I(D \cdot C) > 0$ for any irreducible curve C on S. These properties characterize an ample divisor (Y. Nakai; → 16 Algebraic Varieties E).

Let Σ be an irreducible linear system of dimension $r(\geqslant 1)$, and let C be a generic component of Σ. Let C' be a member of Σ different from C. Then the set of C-divisors $C \cdot C'$ forms a linear system of dimension Y- 1 on C. This is called the **trace** of Σ on C and is denoted by $\mathrm{Tr}_C \Sigma$. The trace is, in general, not complete. The integer $\dim |\mathrm{Tr}_C \Sigma| - \dim \mathrm{Tr}_C \Sigma = \delta(\Sigma)$ is called the **deficiency** of Σ. The deficiency of the complete linear system $|D|$ is denoted by $\delta(D)$.

Let x be a point of S, and let $\mathcal{O}_x$ be the local ring of x. Then $\mathcal{O}_S = \bigcup_{x \in S} \mathcal{O}_x$ is an †algebraic coherent sheaf, called the †structure sheaf of S. Let D be a divisor on S. The sheaf of germs of rational functions f such that $(f) + D \succ 0$ is denoted by $\mathcal{O}(D)$. Then $H^0(S, \mathcal{O}_S(D))$ is a defining module for the complete linear system $|D|$. $D > 0$ if and only if $\mathcal{O}_S(-D)$ is a sheaf of $\mathcal{O}_S$-ideals. The quotient sheaf $\mathcal{O}_S/\mathcal{O}_S(-D)$ will be denoted by 0,. If D is a †prime divisor, then $\mathcal{O}_D$ is the structure sheaf of the algebraic curve D (→ 9 Algebraic Curves). Let $\mathcal{F}$ be a sheaf on S. We set

$$\chi(S, \mathcal{F}) = \textstyle\sum_{q=0}^{2} (-1)^q \dim H^q(S, \mathcal{F}).$$

$\chi(S, \mathcal{O}_S)$ will be denoted simply by $\chi(S)$. We call $p_a(S) = \chi(S) - 1$ the **aritbmetic genus** of the algebraic surface S. Sometimes $\chi(S)$ is referred to as the arithmetic genus of S. We set $\chi_S(D) = \chi(S) - \chi(S, \mathcal{O}_S(-D))$. The integer $p_a(D) = 1 - \chi_S(D)$ is, by definition, the **arithmetic genus of the divisor** D. If D is a prime divisor, then $p_a(D)$ coincides with the arithmetic genus of the algebraic curve, i.e., $p_a(D) = \dim H^1(D, \mathcal{O}_D)$.

D. Riemann-Roch Theorem

Let S be a nonsingular surface and let K denote a canonical divisor on S, i.e., $K = (\omega)$ for some nonzero rational 2-form ω on S. If C is a nonsingular irreducible curve on S, $(K + C) \cdot C$ becomes a canonical divisor on C; hence, $\deg((K + C) \cdot C) = 2p_a(C) - 2$ byacorollaryto the Riemann-Roch theorem on C (→ 9 Algebraic Curves C). Since $I((K + C) \cdot C) = \deg((K + C) \cdot C)$, it follows that $p_a(C) = I((K + C) \cdot C)/2 + 1$. Moreover, the formula

$$p_a(D) = I((K + D) \cdot D)/2 + 1$$

holds for an arbitrary curve D on S; this is called the **adjunction formula.** For any divisor, we define $p_a(D)$ to be $I(D \cdot (D + K))/2 + 1$. Then

$$\chi(S, \mathcal{O}(D)) = \chi(S) - \chi_S(-D) = \chi(S) + p_a(-D) - 1$$
$$= I(D \cdot (D - K))/2 + \chi(S).$$

This formula is called the **Riemann-Rocb theorem** on S (→ 366 Riemann-Roch Theorems C). Applying Serre's duality theorem to D (→ 16 Algebraic Varieties E) we have

$$\dim H^i(S, \mathcal{O}(D)) = \dim H^{2-i}(S, \mathcal{O}(K-D)) \quad \text{for } i=0, 1, 2.$$

In particular, dim $H^2(S, \mathcal{O}(D)) = l(K-D)$, which is called the **index of speciality** of D; dim $H^1(S, \mathcal{O}(D)) = \dim H^1(S, \mathcal{O}(K-D))$ is called the **superabundance** of D, denoted by $h^1(D)$. The inequality

$$l(D)+l(K-D) \geqslant \chi(S, \mathcal{O}(D)) = I(D\cdot(D-K))/2+\chi(S)$$

is called the **Riemann-Roch inequality,** where equality holds if and only if $h^1(D) = 0$. If D is a curve with s connected components, then

$$l(D+K) = h^1(-D)+p_a(D)-s+\chi(S).$$

In addition, if $h^1(0) = 0$, then $h^1(D) = 0$ and thus $l(D+K) = p_a(D)-s+\chi(S)$. This is called the **Riemann-Roch theorem for the adjoint system.** Note that $|D+K|$ is called the **adjoint system** of D. The Noether formula, $(K^2)+c_2(S)=12\chi(S)$, is a special case of Hirzebruch's theorem of Riemann-Roch type (→ 366 Riemann-Roch Theorems B). Here, $c_2(S)$ denotes the second Chern number of S, which coincides with the Euler number of S if **K** is the field of complex numbers.

Let Div(S) denote the group of all divisors on S; by linearity we can define the bilinear form $I(D\cdot D')$ on $\mathrm{Div}(S)_0 = \mathrm{Div}(S)\otimes_{\mathbf{Z}}\mathbf{Q}$. $J = \{D \in \mathrm{Div}(S)_0 \mid I(D\cdot D') = 0 \text{ for all } D'\}$ is a subgroup and X = $\mathrm{Div}(S)_0/J$ is a finite-dimensional vector space over $\mathbf{Q}$, on which the nondegenerate bilinear form I is induced. I has a unique positive eigenvalue; thus the other eigenvalues are all negative. This is called the **index theorem of Hodge;** it is derived from the Riemann-Roch theorem on S. From this, we inferthatif $D^2 = I(D\cdot D)>0$, then $I(D\cdot D')^2 \geqslant D^2\cdot D'^2$ for any D'. X is said to be the **Neron-Severi group** of S and dim X is said to be the **Picard number** of S (→ 16 Algebraic Varieties P).

To study $l(mD)$ as a function of $m \gg 0$, O. Zariski writes an arbitrary effective divisor D as a sum of $D^{(+)}$ and $D^{(-)} \in \mathrm{Div}(S)_0$ with nonnegative rational coefficients such that (1) $D^{(+)}$ is **arithmetically effective** (or, numerically semipositive), i.e., $I(D^{(+)}\cdot C) \geqslant 0$ for any curve C on S; (2) $D^{(-)}=0$ or the intersection matrix of the support of $D^{(-)}$ is negative definite; and (3) $I(D^{(+)}\cdot D^{(-)}) = 0$. Such a decomposition is unique and is called the **Zariski decomposition** of D [7]. If $mD^{(+)}$ is a divisor for some $m>0$, then $l(mD) = l(mD^{(+)})$.

E. Invariants of Algebraic Surfaces

There are many invariants besides the arithmetic genus discussed earlier. We set $h^{q,p} = \dim H^p(S, \Omega^q)$. Then $h^{2,0}$ is equal to the number of linearly independent holomorphic 2-forms; it is called the **geometric genus** and is usually denoted by p_g. Since $h^{0,1}$ gives the maximum among the deficiencies $\delta(\Sigma)$ of linear systems on S, it is called the **maximal deficiency** of S. For a divisor C such that $h^1(C) = 0$, we have $\delta(C) = h^{0,1}$. The number $h^{0,1}$ was formerly called the **irregularity** of S, because $h^{0,1}$ was considered to be a correction term in the equality $p_a(S) = p_g - h^{0,1}$. The study of higher-dimensional varieties showed, however, that it was unnatural to regard $h^{0,1}$ as a correction term. At present, by **irregularity** we mean the dimension of the Picard variety of S (→ 16 Algebraic Varieties P), and we denote this number by q.

When S is defined over the complex number field, we have $h^{p,q} = h^{q,p}$. In particular, $q = h^{0,1} = h^{1,0}$. This number is equal to the number of linearly independent †Abelian simple integrals of the first kind; it is also equal to one-half the first Betti number of S. In cases with positive characteristic, these equalities do not hold in general. J.-P. Serre gave an example of an algebraic surface S such that $h^{0,1} \neq h^{1,0}$, and J. Igusa gave an example such that $q < h^{0,1} = h^{1,0}$. Let K be a canonical divisor of S. The number $P_i = l(iK)$ is called the **i-genus**, and P_i $(i = 2, 3, \ldots)$ are generally called **plurigenera.** If $P_n = 0$ and $d \mid n$, then P_d is also zero. The numbers $p_a(S) = h^{2,0} - h^{0,1}$, $p_g(S) = h^{2,0}$, $h^{1,0}$, $h^{0,1}$, P_i $(i=2,3,\ldots)$ $(P_1 = p_g)$ are †absolute invariants of S'; i.e., they take the same values for any nonsingular surface S' that is birationally equivalent to S. However, $h^{1,1}$ is not an absolute invariant. For a projective plane, all plurigenera P_n vanish and the irregularity $q = 0$. Thus if S is a **rational surface** (→ 16 Algebraic Varieties J), i.e., a surface which is birationally equivalent to $\mathbf{P}^2$, all $P_n = q = 0$. Conversely, any surface with $q = P_2 = 0$ is a rational surface. This is called **Castelnuovo's criterion.** A **ruled surface** is defined to be a surface that is birationally equivalent to a product of the projective line and a curve. All P_n of a ruled surface equal 0 and any surface with $P_4 = P_6 = 0$ is a ruled surface. This is called the **criterion of ruled surfaces** (Enriques).

F. Characteristic Linear Systems of Algebraic Families

One of the central problems considered by the Italian school was to prove that the irregular-

ity q is equal to the maximal deficiency $h^{0,1}$. For that purpose, Severi introduced the notion of characteristic linear systems of algebraic families. Let Σ be an irreducible algebraic family of positive divisors on S such that a generic member C of Σ is an irreducible nonsingular curve, and let r be the dimension of Σ. Let Σ_1 be a 1-dimensional subfamily of Σ containing C as a simple member, and let C be a generic member of Σ_1. Then the specialization of $C' \cdot C$ over the specialization $C' \rightarrow C$ is a well-defined C-divisor of degree $n = I(C \cdot C')$. The set of C-divisors thus obtained is called the **characteristic set.** The characteristic set forms an (Y − 1)-dimensional linear system and contains Tr, $C|$ as a subfamily. This linear system is called the **characteristic linear system** of Σ. For any algebraic family of dimension r, we have $r \leqslant$ dim Cl + q. In particular, there exists an algebraic family Σ that contains ∞^q linear systems and such that for a generic curve C we have $h^1(C) = 0$. For such an algebraic family, we have the equality r = dim $C|$ + q; hence the inequality $q \leqslant h^{0,1}$ follows. Moreover, if the characteristic linear system is complete, we have $q = h^{0,1}$. The proof of the completeness of characteristic linear systems given by Severi is valid only in some special cases (e.g., the case $p_g = 0$). For a complex algebraic manifold, a rigorous proof was given later. When the characteristic is positive, the completeness does not hold in general (Igusa); however, for the surface with $p_g = 0$, the completeness holds (Y. Nakai). The completeness holds if and only if the Picard scheme of S is reduced [14].

G. Birational Transformations of Algebraic Surfaces

Let S and S' denote nonsingular surfaces. If there exists a birational morphism $T: S \rightarrow S'$, we say that S dominates S', and we write S $\geqslant S'$. In addition, if T is not an isomorphism, we write S > S'. In case there does not exist an S' with S > S', S is said to be **relatively minimal.** On the other hand, if we have $S' > S$ for any S' which is birationally equivalent to S, S is said to be **minimal.** Any minimal surface is, by definition, relatively minimal. If a minimal (resp. relatively minimal) surface S is birationally equivalent to S', we say that S is a **minimal** (resp. **relatively minimal) model** of S' or of the field **K(S).** A necessary and sufficient condition for S to have a minimal mode1 is that S not be a ruled surface (Castelnuovo and Enriques).

Let S be a nonsingular surface and P be a point on S. Replacing P by a projective line $\mathbf{P}^1$, we have a nonsingular surface S' and a birational morphism $T: S' \rightarrow S$ such that $T^{-1}(P) = E \cong \mathbf{P}^1$ and $S' - E \cong S - \{P\}$ by T. $T: S' \rightarrow S$ (or T^{-1} in some references) is said to be a locally quadratic transformation. Any birational morphism between nonsingular surfaces is a composition of locally quadratic transformations and an isomorphism. Given a birational mapping $T: S \rightarrow S'$ and P on S', $T^{-1}\{P\}$ is called an **exceptional curve** whenever it is not a point. Moreover, it is called an **exceptional curve of the first kind** if T is regular along $T^{-1}\{P\}$. Otherwise, it is of **the second kind.** Exceptional curves consist of irreducible rational curves. An irreducible curve E is an exceptional curve of the first kind if and only if $(E^2) = -1$ and $E \cong \mathbf{P}^1$. S is relatively minimal if and only if S has no exceptional curves of the first kind; S is minimal if and only if S has no exceptional curves at all. A relatively minimal surface that is not minimal is a ruled surface. Such a surface is either $\mathbf{P}^2$ or a $\mathbf{P}^1$-bundle over a curve. In particular, a relatively minimal rational surface is either $\mathbf{P}^2$ or a $\mathbf{P}^1$-bundle over **P'.** Any surface of the latter type is occasionally called a **Hirzebruch surface** and its (K^2) is 8, where K denotes a canonical divisor. However, (K^2) of $\mathbf{P}^2$ is 9. Define the **linear genus** of a rational surface to be 10. If S is not a rational surface, taking a relatively minimal surface S', **we** define the **linear genus** $p^{(1)}$ of S to be $(K^2)+1$ of S'.

H. Examples of Algehraic Surfaces

Let S_m denote a nonsingular surface in $\mathbf{P}^3$ defined by a homogeneous polynomial of degree m. Let H denote a divisor on S_m induced from a (hyper)plane on $\mathbf{P}^3$. Then the canonical divisor K is linearly equivalent to $(m-4)H$, i.e. ***K-(m-4)H.*** Hence, $p_g = (m-1)$ $(m\ \ 2)$ $(m\ \ 3)/6$. Moreover, $q = 0$; if K is the field of complex numbers, S_m is simply connected. S_2 is isomorphic to the product of two copies of $\mathbf{P}^1$, i.e. $S_2 \cong \mathbf{P}^1 \times \mathbf{P}^1$; hence, it is a rational surface. S_3 is also rational. There exist 27 lines on S_3. Contracting 6 mutually disjoint lines among these 27 lines, we obtain a projective plane. Conversely, given 6 points on $\mathbf{P}^2$ in general position, by performing locally quadratic transformations with these points as centers we get a cubic surface S_3 and a birational morphism $T: S_3 \rightarrow \mathbf{P}^2$. The inverse images by T of the 6 given points, the proper transforms of 15 lines connecting every pair of points chosen from the 6, and 6 conics passing every 5 of the 6 points by T^{-1}, exhaust 27 lines on S,. If $m = 4$, then $K \sim 0$. In general, a nonsingular surface with $K \sim 0$ and $q = 0$ is said to be a ***K3* surface** (- 72 Complex Manifolds J). K3 surfaces have certain properties similar to **Abelian surfaces** that are defined to be 2-

dimensional Abelian varieties (→ 3 Abelian Varieties). For $K = C$, the period mapping defined by integrating regular 2-forms has been studied extensively in connection with moduli theory.

In general, surfaces with the bigenus $P_2 = p_g = 1$ and $q = 0$ are birationally equivalent to $K3$ surfaces; surfaces with $P_4 = p_g = 1$ and $q = 2$ are birationally equivalent to Abelian surfaces. Every Abelian surface has the involution i defined by $i(\alpha) = -a$ and the quotient surface by i is a singular surface with 16 ordinary double points. By performing 16 locally quadratic transformations successively with these singular points as centers we obtain a $K3$ surface. Such a $K3$ surface is called a **Kummer surface.** The original singular surface is also called a Kummer surface. A quartic surface in $\mathbf{P}^3$ with 16 double points is an example of a Kummer surface.

The theory of birational classification of surfaces was developed by Castelnuovo, Enriques, and others of the Ttalian school. This theory has been extensively enriched and generalized in various ways. Kodaira's theory of analytic surfaces includes classification of algebraic surfaces (→ 72 Complex Manifolds I, J), and classification of surfaces in the positive-characteristic case has been recently studied in detail [4, 5]. When the field $K(S)$ is a subfield of a purely transcendental extension $K(X, Y)$, S is said to be a unirational surface. If the extension $K(X, Y)/K(S)$ is separable, S is a rational surface. However, if it is inseparable, S may become nonrational, a $K3$ surface, an elliptic surface, or a surface of general type [5, 12]. Even for noncomplete surfaces, we have a classification theory [10] similar to the previous ones. The following result is one of the applications: whenever $S \times A' \cong A^3$, S is isomorphic to A^2. Here A" means the affine n-space (M. Miyanishi, T. Sugie, T. Fujita; [7, 9, 13]).

Let I be the ring of integers of a real quadratic field of discriminant d. The Hilbert modular group $G = SL(2, I)/\{\pm 1\}$ acts on the product $\mathscr{H}^2$, $\mathscr{H}$ being the complex upper half-plane. The normal complex space $\mathscr{H}^2/G$ can be compactified by adding a finite number of points and thus a compact normal surface is obtained. Resolving these singularities in the canonical minimal way, we have a nonsingular surface $Y(d)$ over C, which is called the **Hilbert modular surface** with discriminant d. $Y(d)$ is simply connected; hence, $q(Y(d)) = 0$. If $d = 5, 8, 12, 13, 17, 21, 24, 28, 33, 60$, then $Y(d)$ is a rational surface. If $d = 29, 37, 40, 41, 44, 56, 57, 69, 105$, then $Y(d)$ is birationally equivalent to a $K3$ surface. If $d = 53, 61, 65, 73, 76, 77, 85, 88, 92, 93, 120, 140, 165$, then $Y(d)$ is an elliptic surface with $\kappa(Y(d)) = 1$. Otherwise, it is a surface of general type (→ [19, 20] in 72 Complex Manifolds).

For 0-cycles on a surface → 16 Algebraic Varieties J.

References

[1] M. Artin, Some numerical criteria for contractibility of curves on algebraic surfaces, Amer. J. Math., 84 (1962), 4855496.

[2] H. F. Baker, Introduction to the theory of algebraic surfaces and higher loci, Principles of Geometry 6, Cambridge Univ. Press, 1933.

[3] A. Beauville, Surfaces algebriques complexes, C.I.M.E. (1977), Astérique 54 (1978).

[4] E. Bombieri and D. Husemoller, Classification and embeddings of surfaces, Algebraic Geometry, Arcata, 1974, Amer. Math. Soc. Proc. Symp. Pure Math., 29 (1975), 3299420.

[5] E. Bombieri and D. Mumford, Enriques classification of surfaces in characteristic p: II, Complex Analysis and Algebraic Geometry (In Honour of Kodaira), Iwanami Shoten and Cambridge Univ. Press (1977), 23342, III, Invent. Math., 35 (1976), 1977232.

[6] F. Enriques, Le superficie algebriche, Zanichelli, 1949.

[7] T. Fujita, On Zariski problem, Proc. Japan Acad., 55 (1979), 106–110.

[8] H. W. E. Jung, Algebraische Flächen, Helwingsche, 1925.

[9] T. Kambayashi, On Fujita's strong cancellation theorem for the affine plane, J. Fac. Sci. Univ. of Tokyo, 27 (1980), 5355548.

[10] Y. Kawamata, On the classification of non-complete algebraic surfaces, Proc. Algebraic Geometry, Copenhagen, 1978, Lecture notes in math. 732, Springer, 1979, 215–232.

[11] K. Kodaira, Collected Works III, Iwanami, 1975.

[12] M. Miyanishi, Unirational quasi-elliptic surfaces, Japan. J. Math., 3 (1977), 395-416.

[13] M. Miyanishi and T. Sugie, Affine surfaces containing cylinderlike open sets, J. Math. Kyoto, 20 (1980), 11–42.

[14] D. Mumford, Lectures on curves on an algebraic surface, Ann. Math. Studies, Princeton Univ. Press, 1966.

[15] D. Mumford, Enriques' classification of surfaces in char p: I, Global Analysis (In Honor of Kodaira), Univ. of Tokyo Press and Princeton Univ. Press, 1969.

[16] M. Nagata, On rational surfaces I, 11, Mem. Coll. Sci. Univ. Kyôto, (A) 32 (1960), 351-370; 33 (1961), 271–293.

[17] E. Pascal, Repertorium der höheren Mathematik, II,, Teubner, second edition, 1922.

[18] E. Picard and G. Simart, Théorie des

fonctions algébriques de deux variables indépendentes 1, II, Gauthier-Villars, 1897, 1906.
[19] A. N. Rudakov and I. R. Safarevich, Inseparable morphisms of algebraic surfaces, Math. USSR-I~V., 10 (1976), 1205-1237.
[20] 1. R. Safarevich, Algebraic surfaces, Proc. Steklov Inst. Math., Amer. Math. Soc., 1967.
[21] 0. Zariski, Collected Papers II, III, MIT Press, 1973, 1978.
[22] 0. Zariski, Algebraic surfaces, Erg. Math., Springer, second edition, 1971.

16 (VIII.4) Algebraic Varieties

A. Affine Algebraic Varieties and Projective Algebraic Varieties

Fix a field k. A subset of the n-dimensional †affine space k^n over k is called an **affine algebraic variety** (or simply **affine variety**) if it can be expressed as the set of the common zeros of a (finite or infinite) set of polynomials $F_i(X_1, \ldots, X_n)$ with coefficients in k. Similarly, a subset of the n-dimensional †projective space $P^n(k)$ over k that can be expressed as the set of the common zeros of a set of homogeneous polynomials $G_j(Y_0, \ldots, Y_n)$ is called a **projective algebraic variety** (or simply **projective variety**). In this section, **variety** means either an affine or projective variety. A variety which is a subset of another variety is called a subvariety. These varieties are the forerunners of the modern, more general versions of algebraic varieties, which we will discuss later.

When V is an affine variety in k^n, the set of the polynomials in $k[X] = k[X_1, \ldots, X_n]$ that vanish at every point of V form an ideal I(V) of $k[X]$. The residue class ring $A_V = k[X]/I(V)$ is called the **coordinate ring** (or **affine ring**) of V. We can regard A_V as the ring of k-valued functions on V that can be expressed as polynomials of the coordinates of k^n. When V is a projective variety, the †homogeneous ideal generated by the homogeneous polynomials in $k[Y] = k[Y_0, \ldots, Y_n]$ that vanish at every point of V is denoted by I(V), and the ring $A_V = k[Y]/I(V)$ is called the **homogeneous coordinate ring** of V.

A variety V is said to be **reducible** or **irreducible** according as it is the union of two proper subvarieties or not. A maximal irreducible subvariety of V is called an **irreducible component** of V. Any variety can be written uniquely as the union of a finite number of irreducible components. A variety V is irreducible if and only if $I(V)$ is a †prime ideal. When that is the case, the field of quotients of A_V (when V is affine) or the subfield of the field of quotients of A_V consisting of the homogeneous elements of degree 0 (when V is projective) is called the **function field** of V and is denoted by $k(V)$. Elements of $k(V)$ are called **rational functions** (or simply **functions**) **on the variety** V. The field $k(V)$ is †finitely generated over k. The transcendence degree of $k(V)$ over k is called the **dimension** of V. When V is reducible, the maximum of the dimensions of its irreducible components is called its dimension. If W is a subvariety of an irreducible variety V, then $\dim V - \dim W$ is called the **codimension** of W on V. A subvariety of pure codimension 1 of an affine or projective space can be defined by a single equation and is called a **hypersurface.** If the ideal I(V) of a variety V of dimension r in a projective space $P^n(k)$ is generated by $n-r$ homogeneous polynomials, then V is called a **complete intersection.** Compared with general varieties, complete intersections have some simpler properties. On the other hand, many important varieties are not complete intersections, e.g., †Abelian varieties of dimension $\geqslant 2$.

The intersections and finite unions of subvarieties on a variety V are also subvarieties. Thus the subvarieties can be taken as the †system of closed sets of a topology on V (→ 426 Topology), which is called the **Zariski topology of the variety** V. When k is the field of complex numbers, V can be viewed as an †analytic space, and the topology of V as such (the "usual" topology) is much stronger than the Zariski topology. For the rest of this article, varieties will be considered as having Zariski topologies unless stated otherwise. Terms such as **Zariski open**, **Zariski closed**, and **Zariski dense** are used to mean open, closed, or dense in a Zariski topology. Suppose a condition (P) concerning the points of an irreducible variety V (concerning the elements of a set M parametrized by the points of V) is satisfied in a nonempty Zariski open set of V. Then we say that the condition (P) holds at **almost all points of the variety** V or at general points of V (almost all elements of the set M).

Let U and V be affine varieties in k^n and k^m, respectively. Then the product set $U \times V$ is an affine variety in k^{n+m} and is called the **product algebraic variety** (or simply the **product**) of U and V. Note that the Zariski topology on $U \times V$ is stronger than the product of the topologies of U and V. When k is †algebraically closed, $U \times V$ is irreducible if U and V are irreducible.

Suppose that k is algebraically closed. Let $\mathfrak{P}$ be a †prime ideal of $k[X] = k[X_1, \ldots, X_n]$, and let V be the affine variety in k^n defined as the zero points of $\mathfrak{P}$. Then I(V) = $\mathfrak{P}$ (†Hilbert zero point theorem) (→ 369 Rings of Polynomials D). Therefore there exists a one-to-one correspondence between the set of prime ideals of

$k[X]$ and the set of irreducible varieties in k^n. In particular, the †maximal ideals correspond to the points of k^n. Similarly, there exists a one-to-one correspondence between the set of homogeneous prime ideals of $k[Y]$ other than $\sum_{i=0}^{n} Y_i k[Y]$ and the set of irreducible subvarieties in $P^n(k)$.

When we deal with nonlinear algebraic equations, we cannot expect a simple, clearcut theory without assuming that k is algebraically closed. Hence we take an algebraically closed field K containing k and regard a variety V in k^n as a subset of the variety V_K in K^n defined by the same equations. From now on, we suppose that k is algebraically closed. If the ideal $I(V)$ of $k[X]$ or $k[Y]$ determined by a variety V is generated by polynomials with coefficients in a subfield k' of k, we say that V is **defined over** k' or that k' is a **field of definition** for V. Any variety has the smallest field of definition, which is finitely generated over the prime field. In the theory of A. Weil [92], we fix an algebraically closed field K that has an infinite transcendence degree over the prime field. This K is called the **universal domain**. A point of V is called a **k'-rational point** of V if all of its coordinates belong to a subfield k' of K. Let K_1, K_2 be two extension fields of a field L, and let $(x)\in K_1^n$, $(y)\in K_2^k$. We say (y) is a **specialization** of the point (x) over L (notation: $(x)\underset{L}{\to}(y)$) if all polynomials $f(X)\in L[X_1,\ldots,X_n]$ satisfying $f(x)=0$ also satisfy $f(y)=0$; in other words, if there exists a homomorphism of L-algebras $L[x_1,\ldots,x_n]\to L[y_1,\ldots,y_n]$ mapping x_i to y_i. Let K be the universal domain, V an irreducible variety in K^n, and k' $(\subset K)$ a field of definition for V having a finite transcendence degree over the prime field. Then there exists a point (x) of V such that all points of V are specializations of (x) over k'. Such a point (x) (in general not uniquely determined) is called a **generic point** of V over k'. The ring $k'[x]$ is isomorphic to $k'[X]/I(V)\cap k'[X]$ over k'. (Some authors use the term *generic point* to mean *almost all points* as defined earlier.)

B. Local Rings

Let V be an affine variety and let W be an irreducible subvariety of V. Let $\mathfrak{P}_W$ be the subset of A_V consisting of the elements that vanish identically on W. Then $\mathfrak{P}_W$ is a prime ideal of A_V. The ring of †quotients of A_V with respect to $\mathfrak{P}_W$ is denoted by $\mathfrak{O}_{V,W}$ or by $\mathfrak{O}_W$ and is called the **local ring** of W on V (or of V at W). Suppose for simplicity that V is irreducible. Then $\mathfrak{O}_W$ is the subring $\{f/g \mid f, g\in A_V, g\notin\mathfrak{P}_V\}$ of $k(V)$, and the †residue field of $\mathfrak{O}_W$ modulo the maximal ideal $\mathfrak{P}_W\mathfrak{O}_W$ can be identified with $k(W)$. When a function φ on V $(\varphi\in k(V))$ belongs to $\mathfrak{O}_W$, it is said to be **regular** at W. For a given function $\varphi\in k(V)$, the set of the points of V where φ is regular is Zariski open. In the case of a projective variety, the local ring $\mathfrak{O}_{V,W}$ is defined as the subring of the ring of quotients of A_V with respect to $\mathfrak{P}_W$, consisting of the homogeneous elements of degree 0.

A mapping from an open set U of a variety V to k that is regular at every point of U is called a **regular function** on U. The ring of regular functions on U is denoted by A_U. By assigning A_U to each open set U, we can define a †sheaf of rings $\mathcal{O}_V$ on V, of which the †stalk $\mathcal{O}_{V,x}$ at a point $x\in V$ is the local ring $\mathfrak{O}_{V,x}$. The sheaf $\mathcal{O}_V$ is called the **sheaf of germs of regular functions** on V (or the **structure sheaf** of V) (→ 383 Sheaves H).

C. General Definition

Consider a pair $(V, \mathcal{O})$ of a topological space V and a sheaf $\mathcal{O}$ of germs of mappings from V to k. If V has a finite open covering (U_i) such that each $(U_i, \mathcal{O}|U_i)$ is isomorphic to an affine variety V_i (in the sense that there exists a homeomorphism from U_i to V_i that transforms $\mathcal{O}|U_i$ to the structure sheaf of V_i), the pair $(V, \mathcal{O})$ is called a **prealgebraic variety over** k, and $\mathcal{O}$ is called its **structure sheaf.** Usually $(V, \mathcal{O})$ is denoted simply by V.

A regular mapping between prealgebraic varieties is defined as a continuous mapping $g: V\to V'$ satisfying $\varphi\circ g\in\mathcal{O}_{V,x}$ for any $x\in V$ and $\varphi\in\mathcal{O}_{V',g(x)}$. Furthermore, if g is a homeomorphism and g^{-1} is also regular, then g is called an **isomorphism** or a **biregular mapping.** The Cartesian product $X\times Y$ of prealgebraic varieties X and Y is locally a product of affine varieties. Therefore $X\times Y$ has the structure of a prealgebraic variety. A prealgebraic variety X is called an **algebraic variety** if the image of the diagonal mapping $X\to X\times X$ is closed in the Zariski topology of the product variety $X\times X$ ("separation condition"). (This definition is due to J.-P. Serre [81].) The separation condition corresponds to †Hausdorff's separation axiom. If W is a locally closed subset (i.e., the intersection of an open set and a closed set) of an algebraic variety V, then it becomes an algebraic variety in a natural manner (the germs of regular functions at $P\in W$ are taken to be the germs of functions induced on W by the functions in $\mathcal{O}_{V,P}$). Locally closed subvarieties of k^n or $P^n(k)$ are called **quasi-affine** or **quasiprojective algebraic varieties,** respectively. Definitions of irreducibility and local rings for general algebraic varieties are given in the same manner

as before. In this article, algebraic varieties will often be referred to simply as varieties.

The notion of an irreducible algebraic variety was developed from that of **abstract algebraic variety** (or simply **abstract variety)** defined by Weil.

D. Scbemes

The set of prime ideals ($\neq$ (1)) of a commutative ring A with unity element 1 is denoted by $\mathrm{Spec}(A)$ and is called the **spectrum** of A. For any subset a of A, we denote by V(a) the set of the prime ideals containing a. We define a topology on $\mathrm{Spec}(A)$ in which the closed sets are V(a). This, again, is called the **Zariski topology** of $\mathrm{Spec}(A)$. For an element f of A, the open set $D(f)=\mathrm{Spec}(A)-V(f)$ is called an elementary open set. The elementary open sets form a base of open sets in the Zariski topology of $\mathrm{Spec}(A)$. The set of closed points is nothing but the set of maximal ideals of A. Assigning to each point $\mathfrak{P}$ of $\mathrm{Spec}(A)$ the †ring of quotients $A_{\mathfrak{P}}$, we obtain a sheaf of rings $\tilde{A}$ on $\mathrm{Spec}(A)$. We have the equality $\Gamma(D(f), \tilde{A})=A_f$, where A_f is the †ring of quotients by the multiplicative system $\{f^n \mid n \geqslant 0\}$. In particular, we have $\Gamma(\mathrm{Spec}(A), \tilde{A}) = A$. Regarded as a †local-ringed space with $\tilde{A}$ as the structure sheaf, $\mathrm{Spec}(A)$ is called an **affine scbeme.**

A local-ringed space X which is locally isomorphic to an affine scheme is called a **scbeme.** A **morphism of schemes** is, by definition, a †morphism between them as local-ringed spaces. Thus we obtain a †category whose objects are schemes. We denote it by (Sch). Giving a morphism $f: X\rightarrow\mathrm{Spec}(A)$ is equivalent to giving a ring homomorphism $\Gamma(f)$: $A\rightarrow\Gamma(X, \mathcal{O}_X)$. Hence the category of affine schemes (which is a †full subcategory of (Sch)) is contravariantly equivalent to the category of commutative rings. If there is given a morphism of schemes $f: X\rightarrow S$, X is said to be an **S-scheme** or a **scheme over** S, and f is called the structure morphism and S the base scheme. For two S-schemes $f: X\rightarrow S$, g: $Y\rightarrow S$, a morphism of S-schemes is defined to be a morphism of schemes $h: X\rightarrow$ Y with $f=g\circ h$. Thus we obtain the category of S-schemes denoted by (Sch/S). $\mathrm{Spec}(\mathbf{Z})$ is the unique †final object in (Sch), hence (Sch) is nothing but (Sch/Spec(Z)).

The †fiber product always exists in (Sch). In fact in the case of affine S-schemes X = $\mathrm{Spec}(B)$ and Y= $\mathrm{Spec}(C)$ with S = $\mathrm{Spec}(A)$ *we* have $X \times_S Y = \mathrm{Spec}(B \otimes_A C)$, and in the general case we construct $X \times_S Y$ by patching together fiber products of affine schemes.

A morphism $f: X\rightarrow S$ is called **separated** if the image of the diagonal morphism $\Delta_{X/S}$: $X\rightarrow X\times_S X$ is closed. We also say that X is separated over S or **X** is a **separated S-scheme.** A scheme X is said to be **separated** if it is separated over $\mathrm{Spec}(\mathbf{Z})$. All affine schemes are separated.

When K is a field, $\mathrm{Spec}(K)$ is a space having only one point and equipped with K as the stalk of the structure sheaf. For a point x of a scheme X, denote by k(x) the residue field of $\mathcal{O}_{X,x}$. For $f\in\mathcal{O}_{X,x}$ we call the residue class off in k(x) the value off at x, denoted by f(x). We have a natural morphism i_x: $\mathrm{Spec}(k(x))\rightarrow X$ whose image is {x}. More generally, we call a morphism i of a spectrum $\mathrm{Spec}(K)$ of a field K to X a point of X with values in K. Such a point is determined by a point x in X and an embedding of $k(x)$ in K. A point of X with values in an algebraically closed field is called a **geometric point.** For a morphism f: X $\rightarrow$ S and a point s in S, the fiber product $X \times_S \mathrm{Spec}(k(s))$ is called the **fiber** off over s and denoted by $f^{-1}(s)$. For a geometric point $\mathrm{Spec}(K)\rightarrow S$, $X \times_S \mathrm{Spec}(K)$ is called a **geometric liber.**

A scheme **X** is called **reduced** if the local ring at each point of X has no †nilpotent elements. A scheme is said to be **irreducible** if its underlying topological space is not a union of two proper closed subsets. A scheme is called **integral** if it is reduced and irreducible. Every local ring of an integral scheme is an †integral domain. If a scheme X has an affine open covering $\{U_i = \mathrm{Spec}(A_i)\}$ such that every A_i is a †Noetherian ring, X is said to be **locally Noetherian.** A locally Noetherian scheme is called **Noetherian** if its underlying topological space is †(quasi-)Compact.

A morphism $f: X\rightarrow$ Y = $\mathrm{Spec}(A)$ is said to be **locally of finite type (of finite type)** if X has an open affine covering (a finite open affme covering) $\{U_i = \mathrm{Spec}(A_i)\}$ such that each A_i is a finitely generated A-algebra. A general morphism f: X $\rightarrow$ Y is said to be **locally of finite type (of finite type)** if there is an open affine covering $\{V_i\}$ of Y such that every restriction of $f: f^{-1}(V_i)\rightarrow V_i$ is locally of finite type (of finite type). If f: $X\rightarrow$ Y is (locally) of finite type we say that X is (locally) of finite type over Y.

A scheme of finite type over a field K (i.e., over Spec(K)) is called an **algebraic scheme** over K. There is a †natural equivalence of categories between the category of reduced separated algebraic K-schemes (as a full subcategory of (Sch/K)) and the category of algebraic varieties over K (defined in Section C) equipped with regular mappings as morphisms. Hence we identify these categories from now on. Occasionally, algebraic variety means irreducible variety. Nonalgebraic schemes are also important as tools for the

study of algebraic varieties. For example, for a point x in a scheme X there is a canonical monomorphism j,: $\mathrm{Spec}(\mathcal{O}_{X,x})\rightarrow X$ by which the unique closed point of $\mathrm{Spec}(\mathcal{O}_{X,x})$ is mapped to x. If for two algebraic K-schemes X, Y and for two points $x\in X$, $y\in$ Y there is a K-isomorphism $\mathcal{O}_{X,x}\cong\mathcal{O}_{Y,y}$, then suitable neighborhoods of x and y are isomorphic over K.

Many concepts concerning varieties, e.g., dimension, generic points, specialization, can be naturally extended to the case of schemes by virtue of commutative ring theory.

A morphism of schemes $f: X\rightarrow$ Y is called **proper** if it satisfies the following two conditions: (1) f is separated and of finite type, (2) for every scheme T and for every morphism $T\rightarrow$ Y, the morphism X x $_Y T\rightarrow T$ obtained from f by the "change of base" is a closed mapping. We also say that X is a proper Y-scheme or X is proper over Y. A proper algebraic K-scheme is called **complete.** A projective variety is complete, while an affine variety over K is complete only when it is of dimension zero. Every algebraic variety can be embedded in a complete variety (M. Nagata [63]).

A morphism of schemes $f: X\rightarrow$ Y is called **affine** if every inverse image by f of an open affine subset of Y is again an affine scheme.

A morphism of schemes $f: X\rightarrow$ Y is called **finite** if it is of finite type and there is an affine open covering $\{U_i = \mathrm{Spec}(A_i)\}$ of Y such that $f^{-1}(U_i) = \mathrm{Spec}(B_i)$, where B_i is tintegral over A_i. For a locally Noetherian scheme Y and a morphism of schemes $f: X\rightarrow$ Y the following three conditions are equivalent: (i) f is finite; (ii) f is affine and proper; (iii) f is proper and every fiber off is a finite set. For a finite surjective morphism of Noetherian schemes f: $X\rightarrow$ Y, X is an affine scheme if and only if Y is an affine scheme.

A morphism of schemesf: $X\rightarrow$ Y is said to be **flat** if for each point x $\in$ X, $\mathcal{O}_{X,x}$ is a †flat $\mathcal{O}_{Y,f(x)}$-module. If, moreover, fis surjective, then f is called **faithfully flat.** Assume that g: Y'+ Y is a faithfully flat morphism of finite type of locally Noetherian schemes and $f: X\rightarrow$ Y is a morphism of schemes. Then for many important properties of morphisms, f has these properties if and only if the pull-back $f_{Y'}$: X x $_Y$ Y'+ Y' has the same properties (theory of descent [29, 30]).

E. Cohomology Theory

Let (X, 0) be a ringed space. An $\mathcal{O}$-Module (i.e., a sheaf of O-modules) F is said to be **quasicoherent** if for each point x of X there exist a neighborhood U of x and an †exact sequence $M\rightarrow N\rightarrow F_{|U}\rightarrow 0$, where M and N are free $\mathcal{O}_{|U}$-Modules. An $\mathcal{O}$-Module F is said to be **of finite type** if F is locally generated by a finite number of sections over $\mathcal{O}$; F is **of finite presentation** if, locally, there exists an exact sequence $\mathcal{O}^p\rightarrow\mathcal{O}^q\rightarrow F\rightarrow 0$ where p and q are positive integers (they need not be globally constant); F is **coherent** if(i) F is of finite type and (ii) the kernel of any homomorphism $\mathcal{O}^n_{|U}\rightarrow F_{|U}$ (where n is an arbitrary positive integer, and U is an open set) is of finite type. Obviously, if F is coherent, then F is of finite presentation, which implies that F is quasicoherenet and of finite type. In the category of U-Modules, the full subcategory of coherent sheaves is closed under almost all operations of sheaves. If $\mathcal{O}$ itself is coherent as an O-Module, $\mathcal{O}$ is said to be a **coherent sheaf of rings.** In this case every $\mathcal{O}$-Module of finite presentation is coherent.

The structure sheaf of a locally Noetherian scheme is a coherent sheaf of rings. On a locally Noetherian scheme X, every quasicoherent sub-@,-Module or quotient @,-Module of a coherent $\mathcal{O}_X$-Module is coherent. A coherent $\mathcal{O}_V$-Module on an algebraic variety V is called a **coherent algebraic sheaf.**

Let X = $\mathrm{Spec}(A)$ be an affine scheme. Then every quasicoherent $\mathcal{O}_X$-Module F on X is generated by its global sections. The correspondence $F\rightarrow\Gamma(X, F)$ defines an equivalence beeween the category of quasicoherent sheaves on X and the category of A-modules; if A is Noetherian, then the coherent sheaves and the finite A-modules correspond to each other under this equivalence.

Let X be a separated scheme, and $\mathfrak{U} = \{U_\lambda\}$ an affine open covering of X. For each quasicoherent @,-Module F, the cohomology group $H^q(X$, F) is canonically isomorphic to the †Čech cohomology $H^q(\mathfrak{U}, F)$ (- 383 Sheaves F). If X is of dimension d, then $H^q(X, F) =0$ for every sheaf F of Abelian groups on X and $q>d$.

For a scheme X we define the **cohomological dimension** cd(X) to be the largest integer q such that $H^q(X, F)\neq 0$ for a quasicoherent $\mathcal{O}_X$-Module F on X [35]. The cohomological dimension cd(X) does not exceed the dimension of X. If X is an affine scheme, then cd(X) = 0. The converse is true under the assumption that X is Noetherian (Serre's criterion [29, III]). For an algebraic scheme X of dimension n, cd(X) = n if and only if X is complete (S. L. Kleiman).

Let f: X $\rightarrow$ Y be a proper morphism of Noetherian schemes. Then for every coherent $\mathcal{O}_X$-Module F and for every $q\geqslant 0$, $R^qf_*(F)$ (→ 200 Homological Algebra) is also coherent. In the special case of Y = $\mathrm{Spec}(k)$ with a field k this means that for an algebraic coherent sheaf F on a complete variety X the cohomology group $H^q(X, F)$ is a finite-dimensional vector space over k.

Let X be a scheme over k and let F be a locally free $\mathcal{O}_X$-Module of rank n (i.e., an $\mathcal{O}_X$-Module which is locally isomorphic to $\mathcal{O}_X^n$). If we take an open covering $\{U_i\}$ of X and isomorphisms $\varphi_i: F_{|U_i} \cong \mathcal{O}^n_{X|U_i}$, then $\varphi_i \circ \varphi_j^{-1}$ defines a morphism $g_{ij}: U_i \cap U_j \rightarrow GL(n, k)$, which is called the coordinate transformation of F. If we construct a †vector bundle B on X by the same coordinate transformations g_{ij}, then F can be regarded as the sheaf $B(B)$ of germs of sections of B. By means of the canonical homomorphism $GL(n,k) \rightarrow PGL(n-1, k)$ we can construct a projective bundle **P(F)** on X (which is said to be associated with F). (Note that in [29], **P(F)** is defined to be a projective bundle with coordinate transformations ${}^t g_{ij}^{-1}$, i.e., associated with the dual of F in our sense.) This procedure of associating **P(F)** with a locally free $\mathcal{O}_X$-Module F can be generalized for any quasicoherent Os-Module of finite type on an arbitrary scheme S.

A closed (locally closed) S-subscheme $f: X \rightarrow S$ of p: $\mathbf{P}(F) \rightarrow S$ is called a **projective scheme (quasiprojective scheme)** over S, or f is said to be a **projective morphism (quasiprojective morphism).** A projective morphism is proper. A reduced projective scheme over a field k is nothing but a projective variety over k. We can develop the theory of projective schemes by means of †graded rings in a way similar to affine schemes.

A locally free $\mathcal{O}_X$-Module of rank 1 is called an **invertihle sheaf.** Invertible sheaves correspond to line bundles up to isomorphisms. Let $P = \mathbf{P}^N(k)$ be a projective space, (y,, y,, . . . , y_N) a system of homogeneous coordinates of P, and U_i the open subset of P defined by $y_i \neq 0$. Denote by $\mathcal{O}(n)$ the invertible sheaf on P defined by the coordinate transformation $g_{ij} = (y_j/y_i)^n$. More generally, let p: $P = \mathbf{P}(F) \rightarrow S$ be the projective bundle associated with a locally free Us-Module F of rank N + 1 on a scheme S. Then there is an invertible sheaf O(n) = $\mathcal{O}(1)^{\otimes n}$ on P with the properties: (i) for each SES its restriction to the fiber $p^{-1}(s) = \mathbf{P}^N(k(s))$ is $\mathcal{O}(n)$ defined above; (ii) $p_*(\mathcal{O}(n)) = \mathbf{S}^n(F)$ for $n>0$ where $S^n(F)$ denotes the nth symmetric product of F. The invertible sheaf $\mathcal{O}(1)$ is called the **tautological line hundle** on P. (Note that $\mathcal{O}(1)$ in the sense of [29] is $\mathcal{O}(-1)$ in our sense, but since the definition of **P(E)** is also different, the above property (ii) holds without modification.)

For a quasiprojective S-scheme $f: X \rightarrow \mathbf{P}(E) \rightarrow S$, the restriction of $\mathcal{O}(1)$ to X is denoted by $\mathcal{O}_X(1)$ (or simply $\mathcal{O}(1)$). An invertible sheaf L on X is called **very ample over** S if there exist a locally free $\mathcal{O}_S$-Module of finite type E on S and an S-immersion i: X $\rightarrow P(F)$ such that $\mathcal{O}_X(1) = L$; L is called **relatively ample** over S or S-ample if $L^{\otimes n}$ is very ample over S for a certain $n>0$ (- Section N). When S is an affine scheme, an ample (very ample) sheaf over S is simply called **ample (very ample).** There is the following cohomological criterion of ampleness (generalized **Serre's theorem).**

Let Y be a Noetherian scheme, $f: X \rightarrow$ Y a proper morphism, and L an invertible $\mathcal{O}_X$-Module. Then the following four conditions are equivalent: (i) L is f-ample; (ii) for each coherent @,-Module F there is an integer N such that $R^q f_*(F \otimes L^{\otimes n}) = 0$ for all $n \geqslant N$ and $q>0$; (iii) for each coherent sheaf of ideals $\mathcal{I}$ of $\mathcal{O}_X$ there is an integer N such that $R^1 f_*(\mathcal{I} \otimes L^{\otimes n}) = 0$ for all $n \geqslant N$. They imply the condition (iv): for each coherent $\mathcal{O}_X$-Module F there is an integer N such that the canonical homomorphisms $f^*f_*(F \otimes L^{\otimes n}) \rightarrow F \otimes L^{\otimes n}$ are surjective for all $n \geqslant N$.

Let X be a scheme proper over a field k (such an X is called a **k-complete scheme),** and let F be a coherent Q-module. Then all $H^q(X$, F) are finite-dimensional k-vector spaces (finiteness theorem), and $\chi(F) = \sum(-1)^q \dim H^q(X$, F) is a finite number, called the **Euler-Poincaré characteristic** of F over X. For every invertible sheaf L, $\chi(F \otimes L^{\otimes m})$ is a polynomial in m, which is said to be the **Snapper polynomial** [48]. Supp F, defined to be the set {x $\in$ X $F_x \neq 0$}, is a closed subset of X and is called the **support** of F. The degree of the polynomial $\chi(F \otimes L^{\otimes m})$ is at most Y = dim supp F, and we have $\chi(F \otimes L^{\otimes m}) = e \cdot m^r/r! +$ lower terms in m for some e $\in$ **Z.** e is said to be the **intersection number** of L^r with F [48]. When W is a closed subscheme of X defined by an $\mathcal{O}_X$-ideal I, $(L^r \cdot W)$ is defined to be the intersection number of L^r with $\mathcal{O}_X/\mathcal{I}$, where r = dim W. If L is ample, then $(L^r \cdot W) > 0$ for any W. In particular, when $W = X$, it follows that $(L^n) > 0$, where n = dim X. The converse of this fact is the **Nakai-Moishezon criterion,** saying that if $(L^r\ W) > 0$ for any closed subvariety W of X where $r = \dim W$, then L is ample [56, 64].

When L is ample, $H^q(X, F \otimes L^{\otimes m}) = 0$ for any $q>0$ and for sufficiently large m; hence $\chi(F \otimes L^{\otimes m}) = \dim H^0(X, F \otimes L^{\otimes m})$ and the Snapper polynomial $\chi(F \otimes L^{\otimes m})$ is also called the **Hilbert polynomial** or the **Hilbert characteristic function** of F. If L is the invertible sheaf defined by a hyperplane section of X in $\mathbf{P}_k^N$, i.e., $L = \mathcal{O}_X(1)$, then $\chi(L^{\otimes m}) = \dim H^0(X, L^{\otimes m}) = \dim R_m$ for sufficiently large m where $\bigoplus R_m$ denotes the homogeneous coordinate ring of X in $\mathbf{P}_k^N$.

In general, for a complete irreducible variety V of dimension Y, we put

$$\chi(V) = \chi(\mathcal{O}_V) = \sum_{q=0}^{r} (-1)^q h^{0,q}$$

($h^{0,q} = \dim H^q(V, \mathcal{O}_V)$) and call it the **arithmetic**

genus of V. Classically, the number p,(V) defined by $p_a(V)=(-1)^r(\chi(V)-1)=h^{0,r}-h^{0,r-1}+\ldots\pm h^{0,1}$ was called the arithmetic genus of V, instead of $\chi(V)$. When V is a nonsingular irreducible curve, $p,(V)$ is the usual †genus. If V is a projective variety, the constant term of the Hilbert polynomial is $\chi(V)$ and the coefficient of its highest term is (deg $V)/r!$.

Let V be a †normal variety, and D a †divisor on V. If, for each point $x \in V$, we denote by $\mathcal{O}(D)_x$ the set of the functions $f \in k(V)$ that satisfy (f) + $D \geqslant 0$ on some neighborhood of x, we obtain a coherent algebraic sheaf $\mathcal{O}(D)$, and we have dim $H^0(V, O(D))$ = $I(D)$. If, moreover, V is complete, we put $\chi_V(D)=\chi(V)-\chi(\mathcal{O}(-D))$ and call it the **virtual arithmetic genus** of D. Classically, the number $p,(D) =(-1)^{r-1}(\chi_V(D)$ − 1) was called by that name. When D is effective and has no multiple components, $\chi_V(D)$ coincides with the arithmetic genus of D regarded as a variety. In general, $\chi_V(D)$ stays invariant if D is replaced by a divisor that is algebraically equivalent to D (→ Section N).

If D is a Cartier divisor, $O(D)$ is an invertible sheaf. For two Cartier divisors D_1, D_2, $\mathcal{O}(D_1+D_2)=\mathcal{O}(D_1)\otimes\mathcal{O}(D_2)$, and D_1 and D_2 are linearly equivalent if and only if $\mathcal{O}(D_1)\cong\mathcal{O}(D_2)$.

Let V be an irreducible nonsingular variety and let Ω^p denote the sheaf of germs of regular differential forms of degree p ($\Omega^0=\mathcal{O}_V$). If V is complete, then we denote dim $H^q(V,\Omega^p)$ by $h^{p,q}$.

Serre's duality theorem: Let V be a nonsingular complete irreducible variety of dimension r, B an algebraic vector bundle over V, and B^* the dual vector bundle of B. Denote by $\mathscr{B}$ and $\mathscr{B}^*$ the sheaves of germs of sections of B and B^*, respectively. Then (i) $H^r(V,\Omega^r)$ is canonically isomorphic to k, and (ii) $H^q(V,\mathscr{B})$ and $H^{r-q}(V,\mathscr{B}^*\otimes\Omega^r)$ are dual to each other as linear spaces by means of the †cup product of the above spaces with $H^r(V,\Omega^r)\cong k$. In particular, $H^q(V,\Omega^p)$ is dual to $H^{r-q}(V,\Omega^{r-p})$; hence we have $h^{p,q}=h^{r-p,r-q}$.

This theorem was extensively generalized by Grothendieck in the category of schemes [1, 331.

When the field k is of characteristic 0, we furthermore have $h^{p,q}=h^{q,p}$ by complex conjugation (→ 232 Kähler Manifolds C); but in characteristic p, there are examples for which this symmetry does not hold [61]. In general, $h^{p,q}$ is a †relative invariant but not an †absolute invariant; however, as $h^{p,0}$ is the dimension of the hnear space of the differential forms of the first kind with degree p, $h^{p,0}$ is an absolute invariant. Hence $h^{0,p}$ is also an absolute invariant in the case of characteristic 0. When the characteristic is positive, the absolute invariance of $h^{0,p}$ has not yet been proved.

F. Simple Points and Singular Points

Let V be a variety over an algebraically closed field k. We say that a point P of $V$is **simple** or that $V$is **nonsingular** or **smooth** at P if the local ring $\mathfrak{O}_P$ is a †regular local ring. Since the problem is local, we may assume that $V$is an affine variety in k". Then the simplicity of P on $V$is equivalent to the following condition: P is contained in only one irreducible component of V, and if that component has dimension r there exist n-r polynomials $F_i(X)$ in I(V) such that rank $(\partial F_i/\partial X_j)_{(X)=P}=n-r$. A point of V that is not simple is called a **singular point** or a **multiple point.** The set of singular points on V (called the **singular** locus of V) is a proper closed subset of V. A variety with no multiple points is called **smooth** or **nonsingular.**

This notion can be made relative. A morphism $f: X\rightarrow Y$ of a locally Noetherian scheme is called **smooth** if f is flat and locally of finite type and all the geometric fibers off are nonsingular. In the case of an affine morphism $f: X=\mathrm{Spec}(R[X_1,\ldots,X_{r+s}]/(f_1,\ldots,f_m))\rightarrow Y=\mathrm{Spec}(R)$ of relative dimension r (by which we mean the dimension of the general fiber) with a Noetherian ring R, the smoothness off amounts to a condition that rank $((\partial f_i/\partial X_j)(x))=s$ at each point x of X.

When for a point P of a variety V the local ring $\mathfrak{O}_P$ is †normal, P is called a **normal** point. A simple point is normal. The set of normal points is a nonempty open subset of V. An irreducible variety whose points are all normal is called a **normal algebraic variety** (or simply **normal variety).** The singular locus of a normal variety has codimension $\geqslant 2$. For an irreducible variety V, there exists a pair (V', f) of a normal variety V' and a birational finite morphism $f: V'\rightarrow V$; V' is unique up to isomorphisms and is called the **derived normal model** or **normalization** of V.

Simplicity and normality for V at a subvariety W are defined in the same way as at a point by using the local ring $\mathfrak{O}_{V,W}$.

For a morphism $f: X\rightarrow Y$ of locally Noetherian schemes, locally of finite type, the following three conditions are equivalent: (i) f is smooth and every fiber off is a discrete set; (ii) fis flat and every geometric fiber over Spec(K) off is a union of spectra of fields isomorphic to K; (iii) f is flat and every fiber off over $y \in Y$ is a union of spectra of fields that are finite †separable extensions of $k(y)$. These conditions are local with respect to X. If a morphism f satisfies these equivalent conditions, we say that f *is* **étale** or X is étale over Y. A morphism

$$f:\ X=\mathrm{Spec}(R[X_1,\ ,\ X_n]/(f_1,\ .\ .\ ,f_n))$$
$$\rightarrow Y=\mathrm{Spec}(R)$$

is étale if and only if $\det((\partial f_i/\partial X_j)(x))\neq 0$ for all $x\in X$. Hence étale morphisms correspond to local isomorphisms in the analytic category. For a surjective étale morphism $f: X\rightarrow Y$ many important geometric properties (reduced, integral, normal, nonsingular, etc.) hold on X if and only if they hold on Y (theory of descent).

G. Dimension Theorems

Let V be an irreducible variety, and let U and W be irreducible subvarieties of V. Then any irreducible component of $U\cap W$ that is simple on V has dimension $\geqslant \dim U + \dim W - \dim V$. When the equality holds, the component is called a **proper component** of the intersection $U\cap W$. If every component of $U\cap W$ that is simple on V is proper, we say that U and W **properly intersect** on V. Any two subvarieties U and W of $\mathbf{P}^n(k)$ with $\dim U+\dim W\geqslant n$ intersect each other. When V is an irreducible r-dimensional variety in $\mathbf{P}^n(k)$, the number of points of intersection $V\cap L$ of V with an $(n-r)$-dimensional linear variety L is independent of the choice of L as long as L is in a "general position." This number is called the **degree** of V and is denoted by deg(V). Letting $\mathcal{O}(1)$ be the fundamental sheaf of $\mathbf{P}^n(k)$, we have $\deg V=(\mathcal{O}(1)^r\cdot V)$ (- Section E).

H. Group Varieties

An algebraic variety G is called an †algebraic group if it has a group structure and if the mapping $G\times G\rightarrow G$ sending (x, y) to xy^{-1} is a morphism. Every algebraic group is quasi-projective (Chow). If G is irreducible, then it is also called a **group variety;** a complete group variety is called an †Abelian variety (- 3 Abelian Varieties B, 13 Algebraic Groups B). A scheme G over another scheme S equipped with morphisms over $S: G\times_S G\rightarrow G$, $G\rightarrow G$, and $S\rightarrow G$, called multiplication, inverse, and unit section, respectively, which satisfy the relations corresponding to the usual axioms of group, is called a **group scheme** (over S). As a point set, G is not a group, while, for any scheme T over S the set $G(T)=\mathrm{Hom}_S(T,G)$ of the morphisms from T to G is a group (- 52 Categories and Functors M). Consider an algebraic group scheme G over $S=\mathrm{Spec}(k)$. If the characteristic of k is zero, then G is necessarily reduced, so an algebraic group scheme over k is essentially the same as an algebraic group; if k has characteristic p, there exist algebraic group schemes over k that are not reduced.

I. Rational Mappings

Let $f: V\rightarrow V'$ be a morphism of varieties. If V is not complete, the image f(V) is not always closed; the closure of f(V) (in V') is called the **closed image** of V. The image f(V) contains an open dense subset of the closed image.

Let V and W be irreducible varieties. A closed subset T of $V\times W$ is called an **algebraic correspondence** of V and W. We say that points $P\in V$ and $Q\in W$ correspond to each other by T if $(P,Q)\in T$. If T is irreducible and the closed image of the projection $T\rightarrow V$ coincides with V, then the function field $k(V)$ can be identified with a subfield of $k(T)$; if we have $k(V)=k(T)$ with this identification, then T is called a **rational mapping** from V to W. Moreover, if the same conditions are satisfied for W, then T is called a **birational mapping** (of **birational correspondence** or **birational transformation),** and in this case we have $k(V)=k(W)$. A morphism can be considered a special kind of rational mapping by taking the graph. If T is a rational mapping from V to W and W_1 is the closed image of T in W, then $k(W_1)$ can be regarded as a subfield of $k(T)=k(V)$. If $k(V)$ is †separably generated (†purely inseparable) over $k(W_1)$, then T is said to be **separable (purely inseparable).**

Let T be a rational mapping from V to W, and let V' and W' be irreducible subvarieties of V and W, respectively. If there exists an irreducible subvariety T' of T whose projections have the closed images V' and W', then we say V' and W' correspond to each other by T. The union of irreducible subvarieties of W that correspond to V' by T is a closed subset of W; it is called the **proper transform** of V' by T and is denoted by $T[V']$. Note that $V'\supset V''$ does not imply $T[V']\supset T[V'']$. The set of points of W that correspond to the points of V' is called the **total transform** of V' by T and is denoted by $T\{V'\}$. Identifying $k(T)$ with $k(V)$, we have $\mathfrak{O}_{T,T'}\supset\mathfrak{O}_{V,V'}$ in general; if the equality holds, we say that T is **regular** (or **defined) along** V'. In that case, W' is the unique irreducible subvariety of W corresponding to V' by T. If V' is simple and of codimension 1 in V, then T is always regular along V'. The set U of points of V at which T is regular is a nonempty open subset, and the restriction of T to U defines a morphism from U to W. A rational mapping can be defined as the closure of the graph of a morphism defined on an open subset of V. The study of a rational mapping can be reduced to that of the birational morphism $T\rightarrow V$. **Zariski's main theorem:** Let $S: X\rightarrow Y$ be a birational mapping, and assume that the inverse $S^{-1}: Y\rightarrow X$ is regular, and that X is normal along an irreducible subvariety X'. If there exists an irre-

ducible component Y' of $S[X']$ with $\dim X' \geqslant$ dim Y', then S is regular along X'. It follows from the above main theorem that, if $T: V \rightarrow W$ is a rational mapping and if P is a normal point of V such that $T[P]$ contains an isolated point, then T is regular at P.

For a birational mapping $T: V \rightarrow W$ between complete irreducible varieties, a subvariety V of V is said to be **fundamental** when dim $T[V']$ > dim V'. When V' is a point (curve) we say V' is a **fundamental point (fundamental curve)** with respect to T. The most classical example of a birational correspondence with fundamental points is the **quadratic transformation** T of a projective plane onto itself given by $(x_0 : x_1 : x_2) \rightarrow (x_1 x_2 : x_2 x_0 : x_0 x_1)$ with a suitable coordinate system. In this case, T^2 = identity and the points P_i defined by $x_j = 0$ for $j \neq i$ correspond to the lines $x_i = 0$ by T. Let Cr, denote the group of birational transformations of **P"(k)** into itself. These transformations are called **Cremona transformations** when $n \geqslant 2$. Cr, is generated by linear transformations and quadratic transformations (M. Noether). Recently, Cr, has been studied in detail by M. Demazure [21] and H. Umemura [91]. Let V be a complete nonsingular irreducible variety over k. It is called **relatively minimal** if every birational morphism from V to a complete nonsingular variety V' is an isomorphism. It is called **minimal** if every birational mapping from a nonsingular variety V' to Vis a morphism. Replacing birational mapping by rational mapping, we can define **strong (absolute) minimality.** Abelian varieties and complete nonsingular curves with positive genera are strongly minimal. In general, minimality implies relative minimality.

Zariski's main theorem is closely related to the **general connectedness theorem due to W. Fulton and J. Hansen** stated as follows: Let P be the product of r copies of $\mathbf{P}_k^m$ and A be the diagonal subvariety of P. If X is an irreducible projective variety and $f: X \rightarrow P$ is a morphism with $\dim f(X) > (r-1)m$, then $f^{-1}(\Delta)$ is connected [22]. The following results are derived from this theorem: (1) if X is singular with only normal crossings (- Section L), then X cannot be imbedded in $\mathbf{P}^{2n-1}$, where n = dim X. (2) Let X be a nonsingular subvariety of $\mathbf{P}^N$ which is not contained in any hyperplane. Suppose that by the projection $\pi_x: \mathbf{P}^N \rightarrow \mathbf{P}^{N-1}$ with center x, x being a general point of $\mathbf{P}^N$, **X** is isomorphic to $\pi_x(X)$. Then $3 \cdot \dim X \leqslant 2(N-2)$ (Zak's theorem). (3) Let X be a nonsingular subvariety of $\mathbf{P}^N$ and H be an arbitrary hyperplane section of X. Then dim(Sing H) < codim X (J. Roberts).

By a similar connectedness theorem, it was proved that if C is a (reducible) curve with normal crossings on $\mathbf{P}^2$ over C, then the fundamental group of $\mathbf{P}^2$ -C is commutative, which was conjectured by 0. Zariski [20].

J. Rational Varieties

An irreducible algebraic variety V over k whose function field is purely transcendental over k is called a **rational variety.** A complete smooth surface S over an algebraically closed field is rational if and only if $P_2(S) = q(S) = 0$ (the Castelnuovo-Zariski criterion, → 15 Algebraic Surfaces E).

If the function field of V has a finite algebraic extension which is purely transcendental over k, then Vis called **unirational.** A unirational curve is in fact rational. More generally, if the function field $k(C)$ of a curve C over k is contained in a field finitely generated and purely transcendental over k, then C is rational **(Lüroth's theorem).** A unirational surface over an algebraically closed field of characteristic zero is rational by virtue of the above criterion, but in the case of positive characteristic there are unirational surfaces which are not rational (Zariski). There are nonrational unirational threefolds even of characteristic zero; for example, all smooth cubic hypersurfaces in $\mathbf{P}^4$ (C. H. Clemens and P. A. Griffiths [16]), and some smooth quartic hypersurfaces in $\mathbf{P}^4$ (V. A. Iskovskiĭ and Yu. 1. Manin [45]). See [6].

K. Monoidal Transformations

Let V be an irreducible variety and $\mathscr{I}$ be a sheaf of ideals of $\mathscr{O}_V$. For any affine open set U of V, $\mathscr{I}|U$ is determined by an ideal a of the coordinate ring A of U. Let a,, a,, , a, be a system of generators of a, and let U' be the graph of a rational mapping from U to $\mathbf{P}^m(k)$ such that the points $P \in U$ and $(a_0(P) : a_1(P) : \ldots : a_m(P)) \in \mathbf{P}^m(k)$ correspond to each other by U'. Then U' is uniquely determined (up to isomorphisms) by U and a only. Suppose V has a covering by affine open sets U_i. We obtain U_i' over U_i as before. By patching them together, we get a birational morphism $T: V' \rightarrow V$, which is unique up to isomorphisms. This T with V' is called the **monoidal transformation** or **blowing up** of V by the ideal sheaf $\mathscr{I}$. Note that in some references, the inverse transformation T^{-1} is said to be the monoidal transformation. The inverse image ideal $T^{-1}(\mathscr{I})$ is an invertible sheaf, that is relatively ample over V. If W is the support of $\mathscr{O}_V/\mathscr{I}$ and $W \neq V$, then $T^{-1}(W)$ has codimension 1 and T gives rise to the isomorphism from $V' - T^{-1}(W)$ onto $V - W$. Thus one can say that V' is obtained from V by replacing W with $T^{-1}(W)$, which is locally defined by prin-

cipal ideals. If every point of W is a nonsingular point of V and W is itself nonsingular, $T^{-1}(W)$ is the projective bundle over W whose fiber is the projective r-space where r = dim V – dim $W-1$. In general, if W is a subvariety defined by a sheaf of ideals $\mathscr{I}$, then T is said to be the **monoidal transformation with center** W. In particular, if W is a point, T is called the **(locally) quadratic transformation.**

L. Resolution of Singularities

Given an arbitrary irreducible variety V, we have the problem of finding out a nonsingular projective variety V' birationally equivalent to V. This is called the problem of **resolution of singularities.** In the case of characteristic zero, this problem was solved by Zariski (1944) for dimension $\leqslant 3$ and by H. Hironaka (1964) [37] for any dimension. In the case of characteristic p, S. Abhyankar solved the 2-dimensional case (1956) and the 3-dimensional case (1966). Hironaka's theorem of resolution of singularities is stated as follows: Let V be a variety over a field of characteristic zero. Then there exists a finite sequence of morphisms of varieties: $V_r \rightarrow V_{r-1} \rightarrow . V_1 \rightarrow V_0 =$ V such that (1) V_r has no singular points, (2) each $V_{i+1} \rightarrow V_i$ is a monoidal transformation of V_i with center D_i, (3) each D_i is a nonsingular subvariety, (4) each V_i is normally flat along D_i. Here, $V$is said to be **normally flat** along a subscheme D of V defined by the sheaf of ideals $\mathscr{J}$, if the quotient modules $\mathscr{J}_x^p/\mathscr{J}_x^{p+1}$ are flat $\mathcal{O}_{D,x}$-modules for all p and all points x of D.

Let V be a nonsingular variety and D an effective divisor on V. D is said to be a divisor with (only) **normal crossings** at $x \in V$, if D is defined by $f_1 . f_s$ such that $(f_1, , f_s)$ is a part of a †system of local coordinates around x. D is said to be a divisor with only normal crossings, if it is so everywhere. For any subvariety W on V, there exist $h: \tilde{V} \rightarrow V$ that is a composition of monoidal transformations with nonsingular centers, such that $h^{-1}(W)$ has only normal crossings. This results from Hironaka's main theorem II [37]. The normal crossing divisor is also defined for complex manifolds, and similar results hold.

M. Cycles and Divisors

Let V be an irreducible variety. We denote by $\mathfrak{B}_r$ the set of r-dimensional irreducible subvarieties of V that are simple on V (i.e., are not contained in the singular locus of V), and by $\mathfrak{Z}_r(V)$ the free Abelian group with basis 23,. Elements of $\mathfrak{Z}_r(V)$ are called cycles of dimension Y (or r-cycles) on V. Let A and B be r-cycles; $A=\sum n_i A_i$, $B=\sum m_i A_i$ $(A_i \in \mathfrak{B}_r$, $A_i \neq A_j$ if $i \neq j)$. If $n_i \geqslant m_i$ for all i, then we write $A \geqslant B$. If $A \geqslant 0$, then A is said to be a **positive cycle.** For a O-cycle $A=\sum n_i P_i$ the integer $\deg(A)$ = $\sum n_i$ is called the **degree** of A.

A cycle of codimension 1 is called a **divisor.** A divisor $\geqslant 0$ is usually called **effective** instead of positive. If $V$is of dimension d and if W ● 23,-, , the local ring $\mathcal{O}_{V,W}$ is a †discrete valuation ring. The †normalized valuation defined by it is denoted by $v_W(\ )$. For a function $f \in k(V)$, we say that W is a **zero** of order n if $v_W(f)=n>0$, and that W is a **pole** of order $-n$ if $v_W(f)=n<0$. Any function $f \in k(V)$, other than the constant 0, has at most a finite number of zeros and poles. We denote by $(f)_0$ the sum $\sum v_W(f)$ W extended over all the zeros W off, and put $(f^{-1})_0=(f)_\infty$ and $(f)_0-$ (f),=(f). We call $(f)_0$, $(f)_\infty$, and (f) the **zero divisor,** the **pole divisor,** and the **divisor** off; respectively. The divisor (f) is equal to $\sum v_W(f)\, W$, where the summation is taken over $W \in \mathfrak{B}_{d-1}$, and we have $(fg)=(f)+(g)$. When V is complete and the singular locus of V has codimension >1, then fis constant if and only if(f), =0 (or $(f)_0=0$). Let D_1 and D_2 be divisors; if there exists a function $f(\neq 0) \in k(V)$ such that $D_1-D_2=(f)$, then D_1 and D_2 are called **linearly equivalent** to each other and we write $D_1 \sim D_2$. The **linear equivalence class** containing a divisor D is denoted by cl(D). A divisor which is linearly equivalent to 0 on a neighborhood of each point of V is called a **Cartier divisor** (some authors call a Cartier divisor simply a divisor). If V is smooth, then any divisor is a Cartier divisor. If a divisor D can be written as $D=(f)$ on an open set U, then the function fis called a **local equation** of D on U. Let $T: V' \rightarrow V$ be a rational mapping from a normal variety V' to a complete variety V, let D be a Cartier divisor on V, and assume that the closed image of T is not contained in D. Since T is regular on some open set U such that codim(V' $U)>1$, we have a morphism $\varphi = T_{|U}$ and the pullback $\varphi^*(D)$ defined by composing the local equations of D with φ. Taking the closure of this divisor in V', we obtain a divisor on V', which is denoted by $T^*(D)$.

N. Divisors and Linear Systems

Let V be a complete irreducible variety, f_0, $f_1, \ldots, f_n$ be elements of the function field $k(V)$ of V, and D be a divisor on V satisfying (fi) + $D \geqslant 0$ for each i. Then the set Σ of the divisors of the form $(\sum a_i f_i)+D$, where the a_i are elements of k and not all zero, is called a **linear system.** The linear space $kf_0+kf_1+ . +kf_n$ is called a **defming module** of Σ. The divisors in

Σ are positive and are linearly equivalent to each other; if every positive divisor that is linearly equivalent to a member of Σ belongs to Σ, then Σ is said to be a **complete linear system.** For any linear system Σ, there exists a unique complete linear system containing it, which is denoted by $|\Sigma|$. The maximal positive divisor D_0 that is contained in all divisors of Σ is called the **fixed component** of Σ, and for each $D \in C$ we call $D - D_0$ the **variable component** of D (or of Σ). A point P of V is called a **base point** of a linear system Σ if P is on each variable component of Σ. A linear system Σ is called **irreducible** if its generic member is irreducible; otherwise it is called **reducible.** The dimension of a defining module of a linear system Σ is denoted by 1(C); we call I(C)- 1 the **dimension** of Σ and denote it by dim Σ. A linear system of dimension 1 is called a **linear pencil**.

A defining module of a linear system Σ is determined uniquely up to k-isomorphisms. Let L be a defming module, and let $f_0, f_1, \ldots, f_n$ be a linearly independent basis of L over k. If we associate to each point P of V the point $Q=(f_0(P):f_1(P):\ldots:f_n(P))$ of the n-dimensional projective space, then we obtain a rational mapping Φ_Σ from V to another variety V'. Outside the base points of Σ, the rational mapping Φ_Σ is regular; and the base points are the fundamental points of Φ_Σ. We say that Φ_Σ is the rational mapping defined by the linear system Σ. When Chas no fixed components and Φ_Σ is a closed immersion, Σ is said to be **very ample** (or **ample).** For a divisor D, the set of positive divisors that are linearly equivalent to D is a linear system, which is called the **complete linear system** and is denoted by $|D|$. We usually write $l(D)$ instead of $l(|D|)$. If $|D|$ is very ample, we say that D is **very ample.** We say that D is **ample** (or **nondegenerate**) if mD is very ample for some $m>0$.

0. Differential Forms

Let V be an n-dimensional irreducible variety, and let A = $k(V)$ be its function field. We denote by $\mathfrak{D}^*$ the set of **derivations** of $\mathfrak{K}$ over k, i.e., the k-linear mappings $D: \mathfrak{K} \to \mathfrak{K}$ satisfying $D(fg)=D(f)g+fD(g)$. Then $\mathfrak{D}^*$ is an n-dimensional linear space over $\mathfrak{K}$. Let $\mathfrak{D}$ denote the †dual space of $\mathfrak{D}^*$ over R. For each $f \in \mathfrak{K}$, let df be an element of $\mathfrak{D}$ defined by $\langle df, D\rangle = D(f)$ $(D \in \mathfrak{D}^*)$. Let $x_1, x_2, \ldots, x_n$ be a separating transcendence basis of $\mathfrak{K}$ over k, in the sense that $x_1, \ldots, x_n$ are algebraically independent over k and $\mathfrak{K}$ is a †separable algebraic extension over $k(x_1, \ldots, x_n)$. (Such a basis exists under the weaker hypothesis that k is †perfect.) Then $dx_1, \ldots, dx_n$ form a basis of $\mathfrak{D}$ over R. The homogeneous elements of degree r of the †Grassmann algebra of $\mathfrak{D}$ over $\mathfrak{K}$ are called **differential forms of degree** r on V (or rational r-forms). The set of the differential forms of degree n is a 1-dimensional linear space over $\mathfrak{K}$ spanned by $dx_1 \wedge dx_2 \wedge \ldots \wedge dx_n$.

A set of n functions $f_1, \ldots, f_n$ in $\mathfrak{K}$ is called a system of **local coordinates** on an open set U of V if $f_1-f_1(P), \ldots, f_n-f_n(P)$ is a †regular system of parameters of the local ring $\mathfrak{O}_P$ for each $P \in U$. In that case, $f_1, \ldots, f_n$ is also a separating transcendence basis of $\mathfrak{K}$. If P is a simple point of V, then there exists a system of local coordinates on a suitable neighborhood of P. Let ω be a differential form of degree r on V, and write $\omega = \sum_{i_1 < \ldots < i_r} \varphi_{(i)} df_{i_1} \wedge \ldots \wedge df_{i_r}$, where $(f_1, \ldots, f_n)$ is a system of local coordinates around P. If the coefficients $\varphi_{(i)}$ are regular at P, then ω is said to be **regular** at P.

When V is a complete variety without singular points, a differential form that is everywhere regular on V is called a **differential form of tbe first kind** (or a **regular form);** the differential forms of the first kind are determined by the function field A and are independent of the choice of the nonsingular model V. The number of linearly independent differential forms of the first kind, of degree n, is denoted by p_g and is called the **geometric genus of** V.

Let V be a complete variety, W an irreducible subvariety of V of codimension 1, and P a point of W that is simple on V. Choose a system of local coordinates (fi). Given a differential form ω on V, we write it as a "polynomial" in the df_i, and denote by $v_W(\omega)$ the minimum of the values of the coefficients for the †valuation $v_W(\ )$. The number $v_W(\omega)$ is determined by ω and W, and it is independent of the choice of P and of the local coordinates. Then ω defines a divisor (w) = $\sum_W v_W(\omega)\, W$ on V, which is called the **divisor of a differential form** ω. The divisor of a differential form of degree n (= dim V) is called a **canonical divisor** and is usually denoted by K. The canonical divisors form a linear equivalence class of divisors.

P. Albanese Variety, Picard Variety, Néron-Severi Group

Let V be an irreducible variety. Then we can construct a couple (A, f) consisting of an †Abelian variety called the **Albanese variety** of V and a rational mapping $f: V \to A$ (called an Albanese mapping) such that: (i) the image of f generates A, i.e., the sum off with itself n times, F: $V^n \to A$, is generically surjective for sufficiently large n; (ii) for every rational mapping g: $V \to B$ of V into an Abelian variety B,

there exist a homomorphism h: A $\to B$ and a point $b \in B$ such that $g = h \cdot f + b$. The Albanese variety is uniquely determined up to isomorphisms and f is determined up to translations.

In the case of k = C, if V is a complete nonsingular variety and if q is the dimension of the linear space of differential 1-forms of the first kind on V, then the first †Betti number i?, is equal to 2q. Let $\omega_1, , \omega_q$ be a basis of the linear space and let $\gamma_1, , \gamma_{2q}$ be a basis of the first homology group modulo torsion. Put $\alpha_{ji} = \int_{\gamma_j} \omega_i$ and $\alpha_j = (\alpha_{j1}, , \alpha_{jq})$. Then the period vectors α_j $(1 \leqslant j \leqslant 29)$ are linearly independent over **R** in $\mathbf{C}^q$. If Γ denotes the discrete subgroup of $\mathbf{C}^q$ generated by the α_j, then the quotient group $\mathbf{C}^q/\Gamma$ is the Albanese variety of V. The Albanese mapping is given by the mapping $P \to (\int_Q^P \omega_1, . . , \int_Q^P \omega_q) \pmod \Gamma$, where P is a variable point on V and Q is a fixed point on V (→ 232 Kähler Manifolds C).

Replacing the term "rational mapping" by "morphism" in the definition, we can define the **strict Albanese variety** of V and prove its existence. It is a quotient Abelian variety of the Albanese variety of V. If V is nonsingular, both coincide by virtue of the †strong minimality of an Abelian variety (- Section I).

Let V be a complete normal variety, U the set of the simple points of V, and D a divisor on V. Then D is said to be **algebraically equivalent to** 0 if there exist a nonsingular curve C, a divisor Γ on U x C, and two points P and Q on C such that D can be written as D = $\varphi_P^*(\Gamma) - \varphi_Q^*(\Gamma)$, where φ_P and φ_Q are the morphisms $\varphi_P: U \to U \times P \to U \times C$ and $\varphi_Q: U \to$ U x Q-U x C. We denote by $\mathfrak{G}(V)$, $\mathfrak{G}_a(V)$, and $\mathfrak{G}_l(V)$ the set of all divisors on V, the set of divisors that are algebraically equivalent to 0, and the set of divisors that are linearly equivalent to 0, respectively. We can introduce a canonical structure of an Abelian variety into $\mathfrak{G}_a(V)/\mathfrak{G}_l(V)$, which is called the **Picard variety** of V. The dimension q of the Picard variety is called the **number of irregularity** of V; if q = 0 we say that $V$is **regular.**

The Albanese variety and the Picard variety of V are †isogeneous to each other, and each one is the Picard variety of the other. If $V$is a curve, they are isomorphic and called the **Jacobian variety** (→ 9 Algebraic Curves E).

Using Cartier divisors instead of divisors, we get an analogous theory to construct another kind of Picard variety which turns out to be isomorphic to the Picard variety of the strict Albanese variety of V. The group of the linear equivalence classes of Cartier divisors can be identified with $H^1(V, \mathcal{O}_V^*)$, where $\mathcal{O}_V^*$ is the sheaf of multiplicative groups of the invertible elements in 0,. From this point of view, we can generalize the theory of Picard variety to the case of schemes also. The theory thus obtained is called the theory of **Picard schemes** [60]. The quotient group NS(V) = $\mathfrak{G}(V)/\mathfrak{G}_a(V)$ is finitely generated [66] and is called the **Neron-Severi group** of V. We call the rank of NS(V) the **Picard number** of V and denote it by p(V). In the case of a nonsingular projective variety over k = C we have an inequality p(V) $\leqslant$ $h^{1,1}(V)$ (= dim, $H^1(V, \Omega_V^1)$) and the **Lefschetz number** $B_2(V) - \rho(V)$ is a birational invariant (where B,(V) is the second Betti number of V) [39]. For the positive characteristic case, however, the above inequality does not hold in general [61].

The torsion part of NS(V) is $\mathfrak{G}_n(V)/\mathfrak{G}_a(V)$, where $\mathfrak{G}_n(V)$ denotes the group of divisors numerically equivalent to zero (- Section Q) (T. Matsusaka). The last fact cannot be generalized for higher codimensional cycles [25].

Q. General Intersection Theory

Let V be an irreducible variety of dimension n, and let A and B be irreducible subvarieties of V of dimension r and s, respectively. If C is a proper component of A $\cap B$, we can define the **intersection multiplicity** $i(A\,B$, C; $V)$ of A and B along C on V, which has properties consistent with our geometric intuitions. In particular, it is invariant under biregular mappings. If A and B intersect properly on V and if C,, , C_N are the proper components of A $\cap B$, we define an $(r+s-n)$-cycle $A \cdot B$ by $A \cdot B = \sum_v i(A\,.\,B, C_v; V)\,C_v$ and call it the **intersection product** of A and B. If each component X_α of an r-cycle X $= \sum_\alpha n_\alpha X_\alpha$ and each component Y_β of an s-cycle Y = $\sum_\beta m_\beta$ Y, intersect properly, we define

$$X \cdot Y = \sum_\alpha \sum_\beta n_\alpha m_\beta X_\alpha \cdot Y_\beta.$$

Then we have the associative law (X $Y) \cdot$ Z = $X \cdot (Y \cdot$ Z) for cycles X, Y, and Z whenever both sides are defined. Two r-cycles X_1 and X_2 are said to be **numerically equivalent** if for every (n − r)-cycle Y that intersects them properly we have deg(X, Y) = deg(X_2 Y).

The theory of intersection is one of the most basis theories in algebraic geometry, for the other theories can be constructed from it [92].

R. Chow Rings

Let U and T be nonsingular irreducible varieties. If Z is a cycle on U x T such that $Z \cdot$ $(U$ x t) is defined for every point t of T, we put $Z \cdot (U$ x $t) = X(t) \times t$, and we obtain a family {X(t)} of cycles on U parametrized by the points of T. Such a family is called an **algebraic family** of cycles. Two cycles X_1 and X_2 whose difference is equal to the difference of two cycles in an algebraic family are said to

be **algebraically equivalent.** In particular, if $X_1 - X_2$ can be expressed as the difference of two cycles of an algebraic family parametrized by the points of the affine line, then X, and X_2 are said to be **rationally equivalent.** The set of the cycles that are algebraically (rationally) equivalent to 0 is a subgroup $\mathfrak{Z}_a(U)(\mathfrak{Z}_{rat}(U))$ of J(U). For divisors, rational equivalence coincides with linear equivalence.

Let $f: V \rightarrow U$ be a morphism between nonsingular irreducible varieties. For an irreducible subvariety W of V, let W' denote the closure of $f(W)$, and put $f_{\mathfrak{Z}}(W) = 0$ if dim $W >$ dim W' and $f_{\mathfrak{Z}}(W) =$ m W' if dim $W =$ dim W', where $m= [k(W): k(W')]$ is the degree of the morphism $W \rightarrow W'$. Extending $f_{\mathfrak{Z}}$ by linearity, we obtain a module homomorphism $f_{\mathfrak{Z}}: \mathfrak{Z}(V) \rightarrow \mathfrak{Z}(U)$. If f is proper, then $f_{\mathfrak{Z}}$ induces a module homomorphism f_* from $\mathfrak{Z}(V)/\mathfrak{Z}_{rat}(V)$ =A(V) to A(U).

Let U and V be nonsingular irreducible projective varieties, and f: V+U be a morphism with graph Γ. If $Y \in \mathfrak{Z}(U)$ is such that $\Gamma \cdot (V \times Y)$ is defined, we denote by $f_{\mathfrak{Z}}^*(Y)$ the image of $\Gamma \cdot (V \times Y)$ under the induced isomorphism $\Gamma \rightarrow V$. Each class of the rational equivalence class group A(U) = $\mathfrak{Z}(U)/\mathfrak{Z}_{rat}(U)$ contains a cycle for which $f_{\mathfrak{Z}}^*$ is defined. Hence we can define $f^*: A(U) \rightarrow A(V)$. Let $\Delta: U \rightarrow U \times U$ be the diagonal morphism, and define $x \cdot y$ = $\Delta^*(x \times y)$ for x, $y \in$ A(U). Then A(U) is a ring with respect to this product; moreover, it is a †graded ring with the grading by codimension. This graded ring $A(U)$ is called the **Chow ring** of U, and the mappingf*: $A(U)$+ $A(V)$ is a ring homomorphism. If f is proper, we have $f_*(y \cdot f^*(x)) = f_*(y) \cdot x$ for $x \in A(U)$, $y \in A(V)$.

Zero Cycles. Let V be a nonsingular irreducible projective variety over an uncountable algebraically closed field of characteristic zero (say C). Denote by A,,(V) the group of classes of O-cycles of degree 0 on V modulo rational equivalence. We say A,(V) is finite-dimensional if the mapping $V^n \times V$"+A,(V) sending $(a_1, \ldots, a_n, b_1, \ldots, b_n)$ to $\sum a_i - \sum b_i$ is surjective for a certain **n.**

If A,(V) is finite-dimensional, then the canonical mapping A,$(V) \rightarrow$ Alb(V) induced from $V \rightarrow$ Alb(V) is bijective (A. A. Rojtman). In general, the torsion part of A,(V) is isomorphic to that of Alb(V) (Rojtman [75]). If $h^{p,0}(V) > 0$ for some $p > 1$, then A,(V) is not finite-dimensional (Mumford, Rojtman [74]). If $V$is a surface which is not of general type and with $h^{2,0}(V) = 0$, then $A_0(V)$ is isomorphic to Alb(V) (S. Bloch, A. S. Kas, D. Lieberman [13]). There exist surfaces of general type with $h^{2,0}$ =0 such that A,(V) =O. For instance, Godeaux surfaces are such surfaces (H. Inose [43]).

One Cycles. Let V be a nonsingular projective irreducible variety and $N(V)$ denote {numerical equivalence classes of l-cycles on V} $\otimes_{\mathbf{Z}} \mathbf{R}$. Via the intersection pairing, N(V) is dual to the Neron-Severi group tensored by R (- Section P); and let NE(V) denote the smallest cone in N(V) containing all effective l-cycles. S. Mori studied the structure of $NE(V)$ in detail. One of his results is stated as follows: If the anticanonical divisor is ample, then there exist rational irreducible curves l_1, , l_r such that (1) $-(l_i \cdot K_V) < n+1$ and (2) $NE(V)$= $\mathbf{R}_+[l_1] + \ldots + \mathbf{R}_+[l_r]$, where $n = \dim V$, K_V is the canonical divisor on V, and $[l_i]$ denotes the class represented by 1,. Moreover, Mori has proved that (1) if K_V is not numerically effective, e.g., if $-K_V$ is ample, then V contains a rational curve, and (2) if the tangent bundle of V is †ample, then $V$is isomorphic to the projective space [59]. (This is called the **Hartshorne conjecture.)**

S. Chow Coordinates, Hilbert Schemes

Consider an irreducible algebraic correspondence T between irreducible varieties V and W. Let V' and W' be the closed images of the projections of T to V and W, and let a and c be their dimensions, respectively. Take generic points P, Q of V', W', and consider the total transforms $T\{P\}$ of P to W and $T\{Q\}$ of Q to V. Denoting the dimensions of $T\{P\}$ and $T\{Q\}$ by b and d, respectively, we have a + $b = c + d$, where both sides are equal to the dimension of T. This property is called the **principle of counting constants.**

This simple principle has wide application. For instance, let V be an r-dimensional variety in P"(k), and let $\sum_j u_{ij} X_j = 0$ $(0 \leqslant i \leqslant r)$ be the equations of $r+1$ hyperplanes H_i. The condition $V \cap (H_0 \cap \cap H_r) \neq \varnothing$ defines an irreducible algebraic correspondence T between V and the multiprojective space W = F'"(k) x x P"(k) with u_{ij} as coordinates: $T = \{(x, u) \mid x \in V, \sum u_{ij} x_j = 0\}$. In this case, we have a = r, b = (n l)(r + 1), and d =0 in the notation introduced previously, so that $c = n(r+1) - 1$. This implies that W' is of codimension 1 in W; hence W' is defined by a single equation $F(u_{ij})$ = 0. This form F is the **associated form** of V of B. L. van der Waerden and W. L. Chow. It is a homogeneous form of degree d $(d$ = deg$(V))$ in each $(u_{i0}, , u_{in})$ and is symmetric in the indices i. More generally, for a positive cycle $X = \sum n_\alpha V_\alpha$ of dimension r and of degree d $(= \sum n_\alpha \deg(V_\alpha))$ in $\mathbf{P}^n(k)$, the product $\prod F_\alpha^{n_\alpha}$ of the associated forms F_α of V_α is called the associated form of X. The coefficients of F, arranged in a fixed order and regarded as the homogeneous coordinates of a point of a pro-

jective space, are called the **Chow coordinates** of X. This is a natural generalization of the †Plücker coordinates. Given r, d and a projective variety $U(\subset P^n(k))$, the set of Chow coordinates of the positive cycles that are contained in U, whose dimension is r and whose degree is d, is a projective variety called a **Chow variety.**

In scheme theory, the Hilbert scheme is introduced in the following way [60]. Let $\mathbf{P}^n_{\mathbf{Z}}$ be the n-dimensional projective space over Spec **Z.** For a locally Noetherian scheme S, and for a closed subscheme Z of $\mathbf{P}^n_{\mathbf{Z}} \times S$, we have the composition of the immersion of Z and the projection $\mathbf{P}^n_{\mathbf{Z}} \times S$ to S; this is denoted by $f_Z: Z \rightarrow S$. For any point s of S, $f_Z^{-1}(s)$ is a closed subscheme of $\mathbf{P}^n_{k(s)}$ and the restriction of $\mathcal{O}(1)$ to $f_Z^{-1}(s)$ is denoted by L_s. Put $\mathcal{M}(S) = \{Z \subset \mathbf{P}^n_{\mathbf{Z}} \times S\} \mid f_Z$ is flat$\}$. Then M becomes a contravariant functor to the category of sets, which is representable, i.e., there exists a locally Noetherian scheme M such that $\mathcal{M}$ is naturally isomorphic to the functor Hom($-$, M). M is written as $\mathrm{Hilb}(\mathbf{P}^n_{\mathbf{Z}})$ and is called the **Hilhert scheme.** There exists a closed subscheme W of $\mathbf{P}^n_{\mathbf{Z}} \times \mathrm{Hilb}(\mathbf{P}^n_{\mathbf{Z}})$ with flat f_W such that for any $Z \in M(S)$, there exists a unique $\varphi: S \rightarrow M = \mathrm{Hilb}(\mathbf{P}^n_{\mathbf{Z}})$ in such a way that $Z = W \times_M S$. In particular, if X is a closed subscheme of $\mathbf{P}^n_{\mathbf{Z}}$ over a field k, there exist SE $\mathrm{Hilb}(\mathbf{P}^n_{\mathbf{Z}})$ and a field extension $k/k(s)$ such that $X = f_W^{-1}(s) \otimes_{k(s)} k$. Thus $\mathrm{Hilb}(\mathbf{P}^n_{\mathbf{Z}})$ parametrizes all closed subschemes of $\mathbf{P}^n_k$. Fix a polynomial P and define $\mathcal{M}^P(S)$ to be $\{Z \in M(S) \mid \chi(L_s^{\otimes m}) = P(m)$ for all s$\}$. $\mathcal{M}^P$ is also a contravariant functor, represented by a scheme $\mathrm{Hilb}^P(\mathbf{P}^n_{\mathbf{Z}})$ which is projective over Spec **Z**. The direct sum of all $\mathrm{Hilb}^P(\mathbf{P}^n_{\mathbf{Z}})$ is just $\mathrm{Hilb}(\mathbf{P}^n_{\mathbf{Z}})$.

T. Algehraic Geometry and Complex Analytic Geometry

When k = C, an algebraic variety is called a **complex algehraic variety,** and it has the structure of a †complex analytic manifold or (if it has singular points) of an †analytic space. If we denote by $\mathcal{O}^h_{V,x}$ the ring of holomorphic functions at a point x of V, then $\mathcal{O}_{V,x} \subset \mathcal{O}^h_{V,x}$, and their †completions coincide. If x is a simple point of V, then $\mathcal{O}^h_{V,x}$ is the †ring of convergent power series, and its completion is the †ring of formal power series. The prime ideals of $\mathcal{O}^h_{V,x}$ remain prime under completion (M. Nagata). Because of this, the analytic behavior of V in a neighborhood of x can be investigated algebraically through the completion of $\mathcal{O}_{V,x}$. If V is complete, then the analytic coherent sheaves on V and the algebraic coherent sheaves on V correspond to each other bijectively; consequently, for propositions that can be stated in terms of coherent sheaves, the results in the analytic sense remain valid in the algebraic sense also, and vice versa (J.-P. Serre [82]) (- 72 Complex Manifolds E).

U. Topology of Algebraic Varieties

Every algebraic variety defined over R (or C) can be triangulated by real analytic cells [46]. Let V be a nonsingular connected algebraic variety defined over C. For an algebraic automorphism σ of C we can define V^σ by letting σ operate on the coefficients of the defining equations of open affine coverings of V. V and V^σ are not necessarily homeomorphic or even of the same homotopy type [83]. Grothendieck has shown that there is a †spectral sequence $E_1^{q,p} = H^p(V, \Omega^q_V) \Rightarrow H^{p+q}(V, \mathbf{C})$, called a **Hodge spectral sequence**. This spectral sequence can be defined in a purely algebraic way and $H^{p+q}(V, \mathbf{C})$ can be considered as the †hypercohomology of the de Rham complex $\{\Gamma(V, \Omega^{\cdot}_V), d\}$ of V. If V is projective, V carries a Kähler metric and by the theory of tharmonic integrals the Hodge spectral sequence degenerates and $E_1^{p,q}$ is C-isomorphic to the complex conjugate of $E_1^{q,p}$ (- 232 Kähler Manifolds B). This is also the case if V is complete.

The topology of a nonsingular projective surface was studied by Lefschetz using the method of Lefschetz pencils. For a projective nonsingular irreducible variety V of dimension n defined over **C,** a **Lefschetz pencil** $\{W_t\}_{t \in \mathbf{P}^1}$ of V is, by definition, a linear pencil consisting of hyperplane sections W_t of V such that: (i) for all $t \in U = \mathbf{P}^1_{\mathbf{C}} - \{t_1, t_2, , t_d\}$, W_t is nonsingular; (ii) each W_{t_i} has only one singular point that is an ordinary double point; and (iii) $W_0 \cap W_\infty$ is nonsingular, where we assume $0, \infty \in U$. Embed V into $\mathbf{P}^N$ by a high multiple of a hyperplane of V and take a generic linear pencil $\{H_t\}$ of hyperplanes in $\mathbf{P}^N$. Then $\{W_t = H_t \cap V\}$ is a Lefschetz pencil of V. By blowing up V along $W_0 \cap W_\infty$ we obtain a smooth variety $\tilde{V}$ and a surjective morphism $\pi: \tilde{V} \rightarrow \mathbf{P}^1$. Let $W = \pi^{-1}(0)$, $\pi_U = \pi \mid \pi^{-1}(U)$. $R^p \pi_{U*} \mathbf{Q}$ is a local system attached to the monodromy representation $\varphi_p: \pi_1(U, 0) \rightarrow GL(H^p(W, \mathbf{Q}))$. φ_p is trivial if $p \neq n-1$. For each point t_i there corresponds a cocycle δ_i of $H^{n-1}(W, Q)$ called a **vanishing cocycle** such that if γ_i is a loop based at 0 going once (counterclockwise) around t_i, we have, for each $x \in H^{n-1}(W, \mathbf{Q})$, $\varphi_p(\gamma_i)(x) = x \pm \langle x, \delta_i \rangle \delta_i$, where $\langle\ \rangle$ is the intersection pairing of $H^{n-1}(W, \mathbf{Q})$. $\varphi_p(\gamma_i)$ is called a **Picard-Lefschetz transformation.** The main results due to Lefschetz are restated as follows. (1) **(Weak Lefschetz theorem).** The natural homomorphism $H^i(V, \mathbf{Q}) \rightarrow H^i(W, \mathbf{Q})$ is an isomorphism for $0 \leqslant i \leqslant n-2$ and is an injection for

$i = n-1$, or equivalently $H_i(V, W, Q) = 0$ for $0 \leqslant i \leqslant n-1$. **(2) (Strong Lefschetz theorem).** Let ξ be the cohomology class of $H^2(V, Q)$ corresponding to the hyperplane section W, and let $L: H^*(V, \mathbf{Q}) \to H^{*+2}(V, \mathbf{Q})$ be the homomorphism defined by the cup product with ξ. Then for each $i \leqslant n$, $L^{n-i}: H^i(V, \mathbf{Q}) \to H^{2n-i}(V, Q)$ is an isomorphism. The weak Lefschetz theorem is true for a cohomology with integral coefficients. In fact, V-W has the homotopy type of a real n-dimensional finite CW complex, and $\pi_r(V, W) = 0$ for $r < n$. The strong Lefschetz theorem is equivalent to the statement that $H^{n-1}(W, Q)$ is the direct sum of the vector space spanned by the vanishing cocycles δ_i and the vector space spanned by the invariant cocyles (i.e., $\varphi_{n-1}(\gamma_i)x = x, i = 1, \ldots, d$). Lefschetz's original proof of this statement is incomplete, and no direct topological proof is known. The transcendental proof of (2) is given by the theory of harmonic integrals. A version of Lefschetz pencils is a proper morphism $f: X \to D = \{z \mid |z| < \varepsilon\}$ of a complex manifold X onto a disk D such that $f^* = f \mid f^{-1}(D^*)$, $D^* = D - \{0\}$, is of maximal rank at every point of $f^{-1}(D^*)$. Fix a point $s \in D^*$. $\pi_1(D^*, s)$ operates on $H^j(W, \mathbf{Z})$, $W = f^{-1}(s)$, and we have a representation $\varphi_j: \pi_1(D^*, s) \to GL(H^j(W, \mathbf{Z}))$. For a loop γ based at s and going once around 0 the Picard-Lefschetz transformation $\varphi_j(\gamma)$ is quasiunipotent (i.e., for a certain integer m, $\varphi_j(\gamma)^m$ is unipotent).

For a nonsingular projective variety defined over a field k with characteristic $p > 0$, the above Lefschetz theorems hold for an l-adic cohomology $(l \neq p)$ [19, 30 (SGA 7)] (→ Section AA). Using the theory of finite étale coverings of an algebraic variety defined over a field k, we can define the **algebraic fundamental group** and the **algebraic homotopy groups**, which are profinite completions of the topological fundamental group and the topological homotopy groups, respectively, where $k = C$ [5, 30 (SGA 1)]. Let (X,x) be a germ of a complex space with isolated singular point x. (X,x) is always algebraizable, i.e., the completion $\hat{\mathcal{O}}_{X,x}$ of the analytic local ring $\mathcal{O}_{X,x}$ is isomorphic to the completion of the local ring of a closed point of an algebraic variety defined over C [3]. For topological type of algebraic surfaces → Moishezon [57].

V. Hodge Theory

Let $H_{\mathbf{R}}$ be a finite-dimensional real vector space containing a lattice $H_{\mathbf{Z}}$, and let $H = H_{\mathbf{R}} \otimes_{\mathbf{R}} \mathbf{C}$ be its complexification. A **Hodge structure** of weight m on H (or $H_{\mathbf{R}}$) is, by definition, a direct sum decomposion $H = \bigoplus_{p+q=m} H^{p,q}$, $\bar{H}^{p,q} \cong H^{q,p}$, where $H^{p,q}$ is a complex vector subspace and the overbar denotes complex conjugation. If H and H' carry Hodges structures of weight m and m', respectively, then $H \otimes H'$, $\mathrm{Hom}_{\mathbf{C}}(H, H')$, $\bigwedge^p H$, and H^* carry Hodge structures of weight $m+m'$, $m'-m$, pm, and $-m$, respectively. For a Hodge structure H of weight m, $F^pH = \bigoplus_{k \geqslant p} H^{k,m-k}$, $p = 0, 1, \ldots, m$, induces a decreasing filtration. Let H be a Hodge structure of weight m, and let Q be a bilinear form on H. If the following three conditions are satisfied, the Hodge structure H is said to be polarized by Q. (i) Q is defined over Q and is symmetric (skew-symmetric) if m is even (odd). (ii) $Q(H^{p,q}, H^{p',q'}) = 0$ unless $p = p', q = q'$. (iii) $(\sqrt{-1})^{p-q} Q(v, \bar{v}) > 0$ for nonzero $v \in H^{p,q}$. Let V be a compact Kähler manifold. Then $H = H^n(V, \mathbf{C})$ carries the Hodge structure induced by the type (p, q)-decomposition (→ 232 Kähler manifolds B). This is also the case if V is a compact complex manifold which is the image of a holomorphic mapping from a compact Kähler manifold of the same dimension. Moreover, if V is projective, the Hodge-Riemann bilinear relations define a natural polarization on the subspace P of H consisting of all primitive cohomology classes.

Each algebraic cycle W of codimension s determines a cohomology class $[W] \in H^{2s}(V, \mathbf{Q})$, which belongs to $H^{s,s}(V)$. (Such a class is called an algebraic cycle.) The converse of this fact is called the Hodge conjecture, which says that $H^{2s}(V, \mathbf{Q}) \cap H^{s,s}(V)$ is spanned by algebraic cycles. The case $s = 1$ has been verified by Hodge., Lefschetz, and Kodaira.

Let $V(W)$ be a smooth irreducible algebraic variety defined over C (complex manifold), and let $\varphi: V \to W$ be a projective smooth morphism with connected fibers. Then $\mathbf{H} = R^m\varphi_* \mathbf{C}$ is a flat vector bundle on W with the flat connection ∇. ∇ is often called the **Gauss-Manin connection**, and it can be defined algebraically if W is also algebraic.

For each fiber $V_s = \varphi^{-1}(s)$, $s \in W$, the filtration $F^pH^m(V_s, \mathbf{C}) = \bigoplus_{k \geqslant p} H^{k,m-k}(V_s)$ induces a complex subbundle $\mathbf{F}^p$ and the connection ∇ has the property $\nabla(\mathcal{O}(\mathbf{F}^p)) \subset \mathcal{O}(\mathbf{F}^{p-1} \otimes \mathbf{T}^*)$, where T is the tangent bundle of W [26]. Moreover, if W is algebraic, ∇ is a differential equation with regular singular points on $\bar{W}$ where $\bar{W}$ is a smooth compactification of W, such that $\bar{W} - W$ is a divisor with normal crossings (→ Section L). If we consider the subbundle **P** of **H** consisting of all primitive cohomology classes, the polarization on each fiber induces a Hermitian pseudometric on **P**. Curvatures of bundles $\mathbf{P} \cap \mathbf{F}^p$ have been studied by Griffiths [26]. There exists a classifying space D for polarized Hodge structures and there exists a holomorphic mapping of the universal covering $\tilde{W}$ of W into D, usually

called a period mapping. D may not be a bounded symmetric domain but has several interesting properties [24, 27]. In some cases D/Γ with a suitable discrete subgroup Γ is the moduli space of polarized algebraic varieties (e.g., curves, Abelian varieties).

P. Deligne [18] has generalized Hodge theory to arbitrary algebraic varieties (more generally schemes of finite type over C). The simplest case is the Hodge theory of a smooth noncomplete irreducible variety X. By Nagata's embedding theorem [63] there exists a complete algebraic variety X such that Y = X-X is a subvariety. By virtue of Hironaka's resolution theorem we can assume that $\bar{X}$ is nonsingular and that Y is a divisor with normal crossings. Let $\Omega_X^1(\log Y)$ be a sheaf of germs of meromorphic 1-forms with logarithmic pole along Y, i.e., locally written as $\sum_{i=1}^k a_i(x)(dx_i/x_i)+\sum_{j=k+1}^n a_j(x)dx_j$, where $(x_1, , x_n)$ is a system of local coordinates with center $p\in$ Y in X such that x x_k = 0 is a local equation of Y and $a_i(x)$, $a_j(x)$ are holomorphic at p. Using the complex $\{\Omega_X^p(\log Y)= \bigwedge^p\Omega_X^1(\log Y), d\}$, with a suitable filtration, Deligne has shown that $H = H^m(X, C)$ carries a mixed Hodge structure and this structure is independent of the choice of X. The mixed Hodge structure on H consists of two finite filtrations, i.e., Oc.. $\subset W_{n-1}\subset W_n \subset . \subset H$, the weight filtration which is detined over $\mathbf{Q}$, and 0 $\subset . \subset F^p \subset F^{p+1} \subset . \subset H$, the Hodge filtration such that F^p induces on W_n/W_{n-1} a Hodge structure of weight n. As a corollary he has shown that a meromorphic p-form on $\bar{X}$ with logarithmic pole along Y (i.e., a section of $\Omega_X^p(\log Y)$ is d-closed on X, and $\omega = 0$ if and only if $\omega_{|X}$ is zero in $H^p(X, C)$. An important application of the theory of this mixed Hodge structure on $H^m(X, C)$ is the following. Let V and W be smooth irreducible varieties, and let $\varphi: V\rightarrow W$ be a smooth projective morphism. If $\bar{V}$ is a smooth compactification of V, the canonical homomorphism $H^m(V, \mathbf{Q})\rightarrow H^0(W, R^m\varphi_*\mathbf{Q})$ is surjective. Fix a point SE W. $\pi_1(W, s)$ operates on $H^m(V_s, \mathbf{Q})$. Then this action is semisimple [18].

W. Deformations, Moduli, Algebraic Spaces

In this section for simplicity the field k is assumed to be algebraically closed. Let X be an algebraic scheme over k. A (flat) **deformation of X over a connected scheme** S over k with base point s_0 consists of the following data: (1) A morphism $p: \mathfrak{X}\rightarrow S$ that is flat and of finite type. If X is complete, p is also proper. (2) A closed point $s_0 \in$ S such that the fiber $\mathfrak{X} \times_S k(s_0)$ is isomorphic to X. For any closed point $s \in$ S, the liber X,=X $\times_S$ $k(s)$ is called a **flat deformation** of X. If X is smooth and complete, we assume further that p is smooth. Similarly, we can define a deformation of a polarized algebraic manifold, an embedding deformation of X in an algebraic scheme Y over k, a deformation of an affine scheme with isolated singular points, and a deformation of vector bundles on a fixed algebraic scheme over k, etc. The theory has two aspects: local theory and global theory.

Let $(R, \mathfrak{m})$ be a complete Noetherian local ring such that $R/\mathfrak{m} = k$. Set $R_n = R/\mathfrak{m}^n$. A formal deformation X_R of X is a sequence $\{X_n\}$ such that (i) X_n is a deformation of X over $\mathrm{Spec}(R_n)$ and (ii) there is a compatible sequence of isomorphisms $X_n \otimes_{R_n} R_{n-1} \cong X_{n-1}$ for any n. Let (FLA/k) be the category of finite-dimensional commutative local k-algebras. The local theory of deformation is the study of the covariant functor F of (FLA/k) to (Set), where, for $A\in$(FLA/k), $F(A)$ is the set of isomorphism classes of deformations of X over $\mathrm{Spec}(A)$. The functor is in general neither representable nor prorepresentable (i.e., there exists a formal deformation X_R of X such that $F(A) = \mathrm{Hom}_{k\text{-alg}}(R, A)$). But, under reasonably mild conditions on F, F has the hull R [78]. That is, there is a formal deformation X_R of X and a natural transformation j: G-F, where $G(A) = \mathrm{Hom}_{k\text{-alg}}(R, A)$, such that j is formally smooth (i.e., for any surjection A'+ A in (FLA/k), $G(A')\rightarrow G(A)\times_{F(A)}F(A')$ is surjective) and $G(k[\varepsilon])\rightarrow F(k[\varepsilon])$ is bijective for the ring of dual numbers $k[\varepsilon]$. The formal deformation X_R is called a versal deformation of X. The hull R is unique up to noncanonical isomorphism. The deformation functor F has the hull R if X is (i) a complete algebraic scheme over k, (ii) an affine scheme with isolated singular points, (iii) a polarized algebraic variety over k, or (iv) a vector bundle on a complete algebraic scheme over k, etc. If there exists a deformation $\pi: \mathfrak{X}\rightarrow S$ of X over a scheme S with base point s_0 over k such that $R = \hat{\mathcal{O}}_{S,s_0}$, $X_n \cong 3 \otimes R_n$, the formal deformation $\{\pi_n : X_n\rightarrow \mathrm{Spec}(R_n)\}$ is called algebraizable. Algebraizability of the versal deformation has been studied by M. Artin. Since the assumption that S is a scheme is rather restrictive, Artin has introduced the notion of algebraic spaces and has considered algebraizability in the category of algebraic spaces [2-4]. For a complete algebraic variety, the versal deformation is not necessarily algebraizable and we need to consider deformations of polarized algebraic varieties. The versa1 deformation of an affine variety with an isolated singularity is algebraizable in the category of algebraic spaces. For the global theory of deformations, we need the projectivity assumption, and the theory is essentially reduced to the theory of

Hilbert schemes (or Chow varieties) (→ Section S). The problem of moduli is considered as the study of the set A4 of all isomorphism classes of deformations of X. Usually we consider the moduli of polarized varieties.

Let (Sch) be the category of Noetherian schemes and $\mathscr{M}$ a contravariant functor from (Sch) to (Set) defined by $\mathscr{M}(X)=\{$isomorphism classes of families of polarized varieties parametrized by X $\in$Ob(Sch), e.g., families of polarized Abelian varieties with(out) additional conditions$\}$. The functor $\mathscr{M}$ is called a moduli functor. If it is represented by a scheme M, M is called a **fine moduli scheme** [60]. In this case there is the universal family $\pi: \mathscr{V} \rightarrow M$. In many cases the moduli functor is not representable. A **coarse moduli scheme** for a given moduli problem is a scheme A4 together with a natural transformation ϕ: $\mathscr{M} \rightarrow \mathrm{Hom}(-, M)$ such that (1) $\phi(\mathrm{Spec}(k))$: $M(\mathrm{Spec}(k)) \rightarrow \mathrm{Hom}(\mathrm{Spec}(k), M)$ is bijective for any algebrically closed field k, and (2) for any scheme N and any natural transformation $\psi: \mathscr{M} \rightarrow \mathrm{Hom}(-, \mathrm{N})$, there is a unique natural transformation λ: Hom$(-, M) \rightarrow \mathrm{Hom}(-, \mathrm{N})$ with $\psi = \lambda \cdot \phi$. A coarse moduli scheme is called a **moduli space** or **a moduli scheme.**

In many cases, the moduli space can be obtained as the quotient space of a certain (locally closed) subset H of a Hilbert scheme by the following equivalence relation: s-s' $\in$H if and only if $X_s \cong X_{s'}$ as polarized varieties, where $\pi: \mathfrak{X} \rightarrow H$, $\pi^{-1}(s) = X_s$ (T. Matsusaka [53]). This equivalence relation is often induced by an action of a †reductive algebraic group G. Suppose that a reductive algebraic group G operates on an algebraic k-scheme Z. A G-invariant morphism $f: Z \rightarrow$ Y (i.e., for the trivial action of G on Y, f is a G-equivariant morphism) is called a **geometric quotient** if (1) f is a surjective affine morphism and $f_*(\mathcal{O}_Z)^G = \mathcal{O}_Y$, (2) if X is a G-stable closed subset of Z, then $f(X)$ is closed in Y, and (3) for $x_1, x_2 \in$ Z, $f(x_1) = f(x_2)$ if and only if the G-orbits of x_1 and x_2 are the same. Let G be a reductive algebraic group, χ: $G \rightarrow \mathrm{Aut}(V)$ a rational representation on a finite-dimensional vector space V over k, and $v_0 \neq 0$ a G-invariant point. Then there exists a G-invariant homogeneous polynomial F of degree $\geqslant 1$ on V such that $F(v_0) \neq 0$ (W. J. Haboush [32]). This implies that if a reductive group G operates on an algebraic k-affine scheme Spec(A), then the invariant ring A^G is a finitely generated k-algebra and, moreover, if any G-orbit in Spec(A) is closed, the natural morphism Spec$(A) \rightarrow$ Spec(A^G) is a geometric quotient. For a quasiprojective scheme Z over k with an action of a reductive group G we need a notion of stable points [60, 62]. The subset of stable points of Z consists of all geometric points of a G-stable open subscheme Z^s of Z, and there exists a geometric quotient $f: Z^s \rightarrow$ Y where Y is quasiprojective (Mumford [60], C. S. Seshadri [79], Haboush [32]). In this way Mumford has shown the existence of coarse moduli schemes of nonsingular complete irreducible algebraic curves and polarized Abelian varieties. But, in general, analysis of stable points is very difficult and it is desirable to extend the category of schemes so that it becomes easier to obtain a quotient. Matsusaka has introduced the notion of a Q-variety [52]. M. Artin has introduced the notion of an algebraic space (D. Knutson [49]).

An **algebraic space** X of **finite type** consists of an affine scheme U and a closed subscheme $R \subset U \times U$ such that (1) R is an †equivalence relation, and (2) the projections p_i: $R \rightarrow U$ $(i = 1, 2)$ are étale. (These are often written as $R \rightrightarrows U \rightarrow X$.) A morphism g: $V \rightarrow X$ of an affine scheme V to an algebraic space X consists of a closed subscheme $W \subset U \times V$ such that (1) the projection $W \rightarrow V$ is étale and surjective, and (2) the two closed subschemes $R \times_U W$, $W \times_V W$ of $U \times U \times V$ are equal. Let $S \rightrightarrows V \rightarrow Y$ be an algebraic space. Then Hom(Y, X) is defined as the kernel of $\mathrm{Hom}(V, X) \rightrightarrows \mathrm{Hom}(S, X)$. If Y is an affine scheme, this definition of Hom(Y, X) is equivalent to the previous definition by virtue of the étaie descent. Thus algebraic spaces form a category which contains the category of schemes. We can define the structure sheaf of an algebraic space and construct a cohomology theory. Many important notions and theorems for schemes can be generalized to those for algebraic spaces. Every algebraic space has a dense open subset that is an affine scheme. A group algebraic space is a group scheme (J. P. Murre). Suppose k = C. If an algebraic group G operates on an algebraic k-scheme properly with a finite stabilizer group, the quotient space exists as an algebraic space. In this way H. Popp has shown the existence of moduli spaces of algebraic surfaces of general type as algebraic spaces [70, 71] (→ 15 Algebraic Surfaces; also [54]). Moreover, every separated algebraic space X of finite type over C carries a natural structure X"" of an analytic space. If there exists a proper modification morphism f: X"" → Y of a separated algebraic space X onto an analytic space Y, then Y carries a structure of an algebraic space and f becomes a morphism of algebraic spaces (Artin [2]). For any algebraically closed field k, Artin has introduced the notion of formal algebraic space and formal contraction and has obtained results similar to those for algebraic spaces [49].

An irreducible compact complex space X whose algebraic dimension (→ 72 Complex Manifolds F) is equal to dim X is called a

Moishezon space. As a corollary of the above theorem, any Moishezon space X carries a structure M of a compact algebraic space such that $X \simeq M^{an}$.

X. Formal Schemes

Let A be a ring which we assume to be Noetherian, for simplicity, and I an ideal of A. Taking $\{I^n\}_{n>0}$ as a fundamental system of neighborhoods of 0, we can introduce a structure of a topological ring into A called **I-adic topology.** The †completion of A with I-adic topology is isomorphic to the projective limit $\hat{A} = \varprojlim_{n>0} A/I^n$ (here A/I^n are regarded as discrete topological rings) and called the **completion of A along I.** If A is Noetherian, then $\hat{A}$ is again Noetherian. There is a canonical continuous homomorphism $i: A \rightarrow \hat{A}$ whose kernel comprises the zero divisors a with $a - 1 \in I$ (intersection theorem of Krull; → 284 Noetherian Rings B). If i is an isomorphism, we say A is **complete** with respect to I. The topology of $\hat{A}$ is the $\hat{I}$-adic topology where $\hat{I} = i(I)\hat{A}$ and $\hat{A}$ is complete with respect to $\hat{I}$. Take a Noetherian ring A complete with respect to I which we consider as an I-adic topological ring by identifying its completion along I with A. On $\mathfrak{X} = V(I) \subset \operatorname{Spec}(A)$ we can define a sheaf of topological rings $\mathcal{O}_{\mathfrak{X}}$ by $\Gamma(\mathcal{D}(f), \mathcal{O}_{\mathfrak{X}}) = \varprojlim_{n>0} A_f/I^nA_f$ for $\mathcal{D}(f) = D(f) \cap \mathfrak{X}$ with $f \in A$. We call **$(\mathfrak{X}, \mathcal{O}_{\mathfrak{X}})$** the **formal spectrum** of A and write Spf(A). I is called a **defining ideal** of Spf(A). A (locally Noetherian) **formal scheme** is by definition a topological local ringed space which is locally isomorphic to a formal spectrum (of a Noetherian ring). If we define morphisms between two formal schemes by those in the category of topological local ringed spaces, the formal schemes form a category.

For two formal spectra Spf(A) and Spf(B) with defining ideals I and J, respectively, the direct product Spf(A) x Spf(B) in the category of formal schemes *is* the formal spectrum of the completion of $A \otimes B$ along the image of $I \otimes B + A \otimes J$. Similarly, we can construct a fiber product of formal schemes. A formal scheme $\mathfrak{X}$ is called **separated** if the image of the diagonal morphism $\Delta_{\mathfrak{X}}: \mathfrak{X} \rightarrow \mathfrak{X} \times \mathfrak{X}$ is closed (→ Section D).

For a Noetherian ring A with an ideal I, the formal spectrum Spf($\hat{A}$) (with a defining ideal $\hat{I}$) is called the **completion of** Spec(A) **along** P'(I). Similarly for a Noetherian scheme X and a closed subscheme X' we can define the **completion** $X_{|X'}$ **of** X **along** X'. Every completion of a separated scheme *is* separated. For a coherent sheaf F on X one can define its **completion** $F_{|X'}$ **along** X', which is again coherent (under the assumption that X is locally Noetherian).

Thus we can develop a theory of formal schemes in a way similar to that of schemes, which **we call "formal geometry"** (for the more general definitions and further discussions see [2; 29, I, III; 303). Roughly speaking, a function on $X_{|X'}$ is a formal Taylor series with respect to the direction normal to X' whose coefficients are regular functions on X'. The method of formal completion enables us to introduce "analytic" or "infinitesimal" methods in algebraic geometry. Among many important theorems, we state here the following two theorems. (1) **The fundamental theorem of proper mapping:** Let $f: X \rightarrow Y$ be a proper morphism of locally Noetherian schemes, Y' a closed subscheme of Y, and $X' = X \times_Y Y'$ the inverse image of Y'. Denote the respective completions of X and Y along X' and Y' by $\hat{X}$ and $\hat{Y}$, respectively. We have the induced proper morphism of formal schemes $\hat{f}: \hat{X} \rightarrow \hat{Y}$. Then we have canonical isomorphisms, $(R^nf_*(F))_{|Y'} \cong R^n\hat{f}_*(F_{|X'}), n \geqslant 0$, for every coherent, $\mathcal{O}_X$-Module F on X. This theorem can be applied to prove **Zariski's connectedness theorem:** for a proper morphism $f: X \rightarrow Y$ of locally Noetherian schemes with $f_*(\mathcal{O}_X) = \mathcal{O}_Y$, every fiber f^{-1}(y) off is connected and nonempty for $y \in Y$. (2) We use the same notation as in (1) and assume, moreover, that Y = Spec(A) for a Noetherian ring A, complete with respect to an ideal I, and Y' = V(I). Then the correspondence $F \rightarrow F_{|X'}$ gives an equivalence between the category of coherent $\mathcal{O}_X$-Modules with proper support over Y and the category of coherent $\mathcal{O}_{\hat{X}}$-Modules with proper support over $\hat{Y}$. This theorem plays an important role in the theory of †deformations of algebraic varieties.

Y. Algebraic Vector Bundles

In this section a **vector bundle** is a locally free sheaf of finite constant rank (→ Section E). A quotient sheaf of a vector bundle is called a **quotient bundle** if it is a vector bundle. A subsheaf F of a vector bundle E is a **subbundle** when both F and E/F are vector bundles. A vector bundle is said to be **indecomposable** unless it is a direct sum of proper subbundles. Every vector bundle E on $\mathbf{P}^1$ is a direct sum of line bundles, that is, $E \simeq \bigoplus \mathcal{O}(a_i)$ (Grothendieck). This property characterizes $\mathbf{P}^1$ in the category of nonsingular projective varieties. In fact, if X is a nonsingular projective variety with dim X > 0 and $X \not\cong \mathbf{P}^1$ and if r is an integer with $r \geqslant$ dim X, there are stable (see below), a fortiori, indecomposable vector bundles on X of rank r (J. Simonis and M. Maruyama). Vector bundles are closely related to subschemes of the base variety. Let E be a

vector bundle of rank 2 on a nonsingular quasiprojective variety X and $\mathcal{O}(1)$ an †ample invertible sheaf on X. For $n \gg 0$ and general $s \in H^0(X, E(n))$, $E(n)/s\mathcal{O}_X \cong L \otimes I$ with L a line bundle and I an ideal in $\mathcal{O}_X$ which defines a smooth subscheme Y= $(s)_0$ of codimension 2. When X is, for example, P" (n $\geqslant$ 3) or affine, the converse holds. Let Y be a subscheme of X purely of codimension 2 and locally of complete intersection. If $\omega_Y \cong$ A4 $\otimes \mathcal{O}_Y$ for an invertible sheaf M on X, there is a vector bundle E of rank 2 and $s \in H^0(X, E)$ with $(s)_0$ = Y. Moreover, E is decomposable if and only if **Y** is globally a complete intersection. An Abelian surface can be embedded in $\mathbf{P}^4(\mathbf{C})$, and hence we have an indecomposable vector bundle of rank 2 on $\mathbf{P}^4(\mathbf{C})$ [40]. For $n \geqslant 5$, we have no examples of indecomposable vector bundles of rank 2 on **P"(C)** [36]. Every vector bundle on **A"** = k^n is trivial (D. Quillen [72], A. Suslin). From this and the fact stated above several results can be deduced; every nonsingular curve in $\mathbf{A}^3$ is set-theoretically a complete intersection (L. Szpiro).

For a vector bundle E on a complete scheme, the following are equivalent: (i) for every coherent $\mathcal{O}_X$-module F, $F \otimes$ S"(E) is generated by its global sections for $n \gg 0$, (ii) for every coherent sheaf F on X, $H^i(X, F \otimes$ S"(E)) = 0 for all $i > 0$ and $n \gg 0$, (iii) the †tautological line bundle $\mathcal{O}(1)$ on **P(E)** is ample. A vector bundle having these properties is said to be **ample.** This is a generalization of the notion of ampleness of invertible sheaves (- Section E). The set of ample vector bundles is closed under several operations [34]. No good criterion such as Nakai-Moishezon's for the ampleness of invertible sheaves (→ Section E) is yet known. The tangent bundle of a nonsingular complete variety X is ample if and only if X $\cong \mathbf{P}^n$ (→ Section R; S. Mori [59]).

Let Y be a nonsingular projective variety over an algebraically closed field and $\mathcal{O}_Y(1)$ an ample invertible sheaf on Y. A coherent sheaf E on Y is said to be **stable** (or, **semistable)** with respect to $\mathcal{O}_Y(1)$ if E *is* torsion-free and if for every coherent subsheaf F of E with $0 \neq F \neq E$, $\chi(F(m))/r(F) < (\text{or} \leqslant) \chi(E(m))/r(E)$ for all $m \gg 0$, where *r(*)* denotes the rank. If E is stable (or semistable) and locally free, it is called a **stable** (resp. **semistable) vector bundle.** Let $f: X \to S$ be a †smooth, projective, geometrically integral morphism and $\mathcal{O}_X(1)$ an f-ample invertible sheaf on X. For a numerical polynomial H and an S-scheme T, set $\Sigma^H_{X/S}(T) = \{E$ E is a coherent sheaf on X_T with the properties (a) and (b)$\}/\sim$; (a) E is T-flat, (b) for every geometric fiber X_t, $E_t = E_{X_t}$ is stable with respect to $\mathcal{O}_X(1)_{X_t}$ and $\chi(E_t(m)) = H(m)$, and if $E_1 \cong E_2 \otimes f_T^*(L)$ with L an invertible sheaf on T, $E_1 \sim E_2$. $\Sigma^H_{X/S}$ is a †contravariant functor of the category of †locally Noetherian S-schemes to the category of sets. If S is of †finite type over a †universally Japanese ring A, $\Sigma^H_{X/S}$ has a coarse moduli scheme $M_{X/S}(H)$ (→ Section W) and it is †locally of finite type over S (D. Mumford, Seshadri, D. Gieseker, Maruyama [23, 51]). The set of the classes of semistable sheaves under a suitable equivalence relation (S-equivalence) on geometric libers of X over S also has a coarse moduli scheme $\bar{M}_{X/S}(H)$ and *M,,,(H)* is its open subscheme. Moreover, it is known that *M,,,(H)* is projective over S in some cases, for example, when A is a field of characteristic zero. When $\dim X/S = 1$, $X \cong \mathbf{P}^2$ or $\mathbf{P}^3$, the structures of *M,,,(H)* have been extensively studied [11, 69]. Theories of vector bundles are used in theoretical physics [9].

Z. Torus Embeddings

Over an algebraically closed field *k,* a **torus embedding,** or a **toric variety,** is a normal scheme X locally of finite type over *k* on which an algebraic torus T acts with a dense open orbit isomorphic to T. Such X's, as well as many of their algebrogeometric properties, can be described very simply in terms of cones in real affine spaces, as Demazure [21] first saw in the nonsingular case in connection with algebraic subgroups of the Cremona transformation group, and then as D. Mumford et al. [47] as well as K. Miyake and T. Oda [67] saw in the general case immediately after H. Sumihiro [SS] proved a basic theorem on linear algebraic group actions.

The group N, written additively, of one-parameter subgroups of an r-dimensional algebraic torus T is a free Abelian group of rank *r.* A **convex rational polyhedral** cone σ in $N_{\mathbf{R}} = \mathbf{R} \otimes_{\mathbf{Z}}$ N with 0 as the vertex is the set of nonnegative linear combinations of a finite number of elements of N such that $\sigma \cap (-\sigma) = \{O\}$. A subset τ of σ is called a face, and is denoted by $\tau < \sigma$, if there exists a linear functional m on $N_{\mathbf{R}}$ having nonnegative values on σ and $\tau = \{y \in \sigma \mid m(y) = 0\}$. A **fan** Δ in $N_{\mathbf{R}}$ is a collection of such σ's satisfying the conditions (i) $\Delta \ni \sigma, \sigma > \tau \Rightarrow \Delta \ni \tau$ and (ii) $\Delta \ni \sigma, \sigma' \Rightarrow \sigma > \sigma \cap \sigma' < \sigma'$.

Each $\sigma \in \Delta$ gives rise to an affine torus embedding U_σ as follows: Let M be the Z-module dual to N; hence A4 is the group of †characters of T. For $\sigma \in \Delta$, the set $M \cap \check{\sigma} = \{m \in M \mid m(y) \geqslant 0$ for all $y \in \sigma\}$ is seen to be a lïnitely generated additive subsemigroup of M containing 0 and generating M as a group. U_σ is then the spectrum of the semigroup algebra $k[M \cap \check{\sigma}]$.

Thanks to the condition (ii) above, the U_σ's can be naturally pasted together to produce a torus embedding X = $\bigcup_{\sigma\in\Delta} U_\sigma$.

Every torus embedding is obtained in this way. Equivariant dominant morphisms between torus embeddings can be described in terms of Z-linear maps between N's which send a fan to another.

Many algebrogeometric properties of X can be described in terms of A, e.g., X is nonsingular if and only if every $\sigma \in$ A can be spanned by a part of a Z-basis of N. X is complete (proper over *k)* if and only if the union of σ's in A coincides with N,. X has at worst cyclic quotient singularities if and only if every $\sigma \in$ A is simplicial, i.e., if it is spanned by R-linearly independent elements of N. An equivariant resolution of singularities of X exists and is obtained by a suitable subdivision of A (Mumford et al. [47]). X has only rational singularities, hence is Cohen-Macaulay (M. Hochster [38] and Mumford et al. [47]). The set of T-orbits in X is in one-to-one correspondence with $\sigma\in\Delta$. A reduced T-invariant subscheme Y of X is the union of T-orbits, and hence corresponds to a subset Σ of A. M.-N. Ishida [44] determined when Y is Cohen-Macaulay or Gorenstein in terms of the combinatorics of Σ.

A T-invariant Cartier divisor D on X corresponds to a support function h, which is a continuous R-valued function on the union $\bigcup_{\sigma\in\Delta}\sigma$ which is (i) positively homogeneous, i.e., $h(\lambda y)=\lambda h(y)$ for $\lambda>0$ and $y\in\bigcup_{\sigma\in\Delta}\sigma$, (ii) Z-valued on N $\cap$ ($\bigcup_{\sigma\in\Delta}\sigma$) and (iii) linear when restricted to each $\sigma\in\Delta$. Various properties of the invertible sheaf $\mathcal{O}_X(D)$ (→ Section E) can be described in terms of h. For instance, when X is complete, $\mathcal{O}_X(D)$ is †ample if and only if h is upper convex, i.e., h(y) + $h(y') \leqslant h(y+y')$, and moreover, A is the coarest fan with the property (iii) above. The cohomology of $\mathcal{O}_X(D)$ can be calculated by means of h (Demazure [21] and V. Danilov [17]).

A support function h, on the other hand, gives rise to a convex polyhedron in $M_{\mathbf{R}}$ with vertices in M. In this way, certain aspects of the geometry of convex sets can be thought of as a part of the theory of projective varieties (R. Stanley [87], B. Teissier [89]).

Mumford et al. [47] introduced a more general concept of **toroidal embedding:** A normal algebraic variety Y and a nonsingular Zariski open subset U such that Y $\supset U$ is formally isomorphic at each point to a torus embedding X $\supset T$. This concept has been used very effectively to prove important theorems: (1) Systematic nice compactitïcations of arithmetic quotients of bounded symmetric domains (D. Mumford et al. [7], 1. Satake [77]). Y. Namikawa [65] worked out the details in the case of the Siegel upper planes, improving their earlier Satake compactifications studied by I. Satake, W. Baily, J.-I. Igusa, and A. Bore1 [12]. **(2) Semistable reduction theorem** (Mumford et al. [47]). Let $f: V\rightarrow C$ be a flat morphism from a nonsingular variety V to a nonsingular curve C in characteristic 0. After a finite base extension $C'\rightarrow$ C and a modification $V'\rightarrow V\times_C C'$, we can get $f':V'\rightarrow C'$ whose singular fibers are reduced with only nonsingular components crossing normally with each other (- Section L). Without the reducedness requirement, the existence is derived from a result of Hironaka [37].

AA. Etale Topology

Let $\mathscr{S}$ be a category. We say that a **Grothendieck topology** on Y is given if, for each S$\in$ Ob(S), families of morphisms **(covering families** of S) are given and satisfy the following conditions: (1) If $\varphi: T\rightarrow S$ is an isomorphism, $\{\varphi: T\rightarrow S\}$ is a covering family, (2) if $\{\varphi_i: R_i\rightarrow S\}_{i\in I}$ is a covering family, for any morphism $\varphi': S'\rightarrow S$, the fiber product R_i' = $R_i \times_S S'$ exists and the induced family $\{\varphi_i'$: $R_i'\rightarrow S'\}$ is a covering family of S', (3) if $\{\varphi_i$: $R_i\rightarrow S\}_{i\in I}$ is a covering family and if for each $i\in I$, $\{s_{i,a}: S_{i,a}\rightarrow R_i\}_{a\in A_i}$ is a covering family, then $\{\varphi_i^0 s_i, a: S_{i,a}\rightarrow S\}_{i\in I, a\in A_i}$ is a covering family. In general, it is more convenient to use a notion of sieves to define a Grothendieck topology [30 (SGA 4)].) A category with Grothendieck topology is called a **site.** Let X be a scheme and Et/X the category whose objects consists of schemes étale over X. If we choose a family of morphisms $\{\phi_i: Y_i\rightarrow Y\}_{i\in I}$, with $\bigcup_{i\in I}\phi_i(Y_i)$ = Y, as a covering family of Y $\in$ Ob(Et/X), this defines the **étale topology** on X and the **étale site** $X_{\text{ét}}$. Similarly, one can define the **Zariski site** X_{Zar} (resp. the **flat site** X_{fl}), using open immersions (resp. flat morphisms locally of finite type). A **presheaf on a site** Y is a contravariant functor from Y to (Set). A presheaf F is a **sheaf** if, for any covering family $\{\phi_i: R_i\rightarrow R\}_{i\in I}$, $F(R)\rightarrow\prod_{i\in I}F(R_i)\rightrightarrows\prod_{i,j\in I}F(R_i\times_R R_j)$ is exact. If Y has a final object X, the functor $F\mapsto F(X)$ is left exact on the category of Abelian sheaves on $\mathscr{S}$ and the cohomology groups $H\cdot(X, F)$ are defined as the right derived functors. Using covering families, one can define the Čech cohomology $\check{H}\cdot(X, F)$ as usual. In the following, a sheaf means an Abelian sheaf. For a †geometric point $i: x\rightarrow X$ of a scheme, an étale neighborhood U of x consists of an étale morphism $f: U\rightarrow X$ and a morphism j: $x\rightarrow U$ such that $f\circ j = i$. For a sheaf F on $X_{\text{ét}}$, the stalk F_x at a geometric point x is defined by F_x =

$\varinjlim \Gamma(U, F)$, U being the étale neighborhoods of x. On $X_{\text{ét}}$ the sheaf $\mathbf{G}_m$ is defined by $\mathbf{G}_m(S) = \Gamma(S, \mathcal{O}_S)^*$ for each $S \in \text{Ob}(\text{Et}/X)$. The kernel of the nth power homomorphism $\mathbf{G}_m \xrightarrow{f_n} \mathbf{G}_m$ is denoted by μ_n. The sequence $1 \to \mu_n \to \mathbf{G}_m \xrightarrow{f_n} \mathbf{G}_m \to 1$ is exact in $X_{\text{ét}}$ if p is prime to residual characteristics of X, but not necessarily exact in X_{Zar}. There are canonical isomorphisms $\text{Pic}(X) = H^1(X_{\text{Zar}}, \mathcal{O}_X^*) \simeq H^1(X_{\text{ét}}, \mathbf{G}_m) \simeq H^1(X_{\text{fl}}, \mathbf{G}_m)$. For X = Spec($k$), k a field, the étale cohomology theory of sheaves on $X_{\text{ét}}$ is equivalent to the †Galois cohomology theory over k; hence $H^i(X_{\text{ét}}, F) \simeq H^i(G, F_{\bar{k}})$, where $G = \text{Gal}(\bar{k}/k)$, $\bar{k}$ the †separable closure of k and $F_{\bar{k}}$ is the stalk of F at the geometric point $\otimes_k$ of X. For a morphism f: X → Y of schemes and a sheaf F on $X_{\text{ét}}$, the direct image sheaf $f_* F$ of F is defined by $f_* F(S) = F(X \times_Y S)$, $S \in \text{Ob}(\text{Et}/Y)$, and higher direct-image sheaves $R^i f_* F$ are defined by the right-derived functor. Let X be a separated scheme of finite type over a field k. By a theorem of Nagata [63], there exists a scheme X proper over k and an open immersion j: $X \to \bar{X}$. For a torsion sheaf F on $X_{\text{ét}}$, let $j_! F$ be a sheaf on $\bar{X}_{\text{ét}}$ extended by 0 outside X. The cohomology with compact support $H_c^q(X_{\text{ét}}, F)$ is defined by $H_c^q(X_{\text{ét}}, F) = H^q(\bar{X}_{\text{ét}}, j_! F)$. This is independent of compactifications. Similarly, for a separated morphism f: X → S of finite type of schemes and a torsion sheaf F on $X_{\text{ét}}$, one can define higher direct image sheaf with compact support $R^q f_! F$. For the étale topology, torsion sheaves are important. All torsion sheaves on $X_{\text{ét}}$ on a Noetherian scheme X are inductive limits of constructible sheaves. A sheaf F on $X_{\text{ét}}$ is called **locally constructible (constant)** if F is represented by an étale covering of X. A sheaf F on $X_{\text{ét}}$ is called **constructible** if there exists a finite surjective family of subscheme X_i of X such that the restriction of F to X_i is locally constructible (constant). This is equivalent to saying that every irreducible closed subscheme Z of X contains a nonempty open subscheme U such that the restriction F to U is locally constant and has finite stalks, i.e., there is a covering family $\{\phi_i: U_i \to U\}$ such that F U_i is constant and the stalk F_x is a finite Abelian group for all geometric points x of U. The cohomology of a torsion sheaf or a constructible sheaf has properties similar to those of the classical cohomology. Let $f: X \to S$ be a proper morphism and F a torsion sheaf on $X_{\text{ét}}$. Then the stalk $R^q f_* F$ at a geometric point s of S is isomorphic to $H^q(X_{s,\text{ét}}, F)$, where $X_s = X \times_S \text{Spec}\, k(s)$. If f is a separated morphism of finite type of Noetherian schemes, a similar fact holds for the cohomology with compact support. Moreover, if F is a constructible sheaf, $R^q f_! F$ is also constructible **(finiteness theorem)**. For an affine scheme X of finite type over a separably closed field and F a torsion sheaf on $X_{\text{ét}}$, one has $H^q(X_{\text{ét}}, F) = 0$ for $q >$ dim X. Let $f: X \to S$ be a separated morphism of schemes of finite type on C, F a torsion sheaf on $X_{\text{ét}}$. One has a canonical isomorphism $(R^q f_! F)^{an} \simeq R^q f_!^{an} F^{an}$. In particular, if X is proper over C, one has $H^q(X_{\text{ét}}, \mathbf{Z}/(n)) \simeq H^q(X^{an}, \mathbf{Z}/(n))$. On the other hand, for a nonsingular complete curve C over an algebraically closed field, one always has $H^1(C_{\text{ét}}, \mathbf{Z}) = 0$. For a geometric point s of S the **strict localization** of S at s is the ring $\mathcal{O}_{S,s} = \varinjlim \Gamma(U, \mathcal{O}_U)$, U being étale neighborhoods of s. A geometric point t of S is called a **generalization** of s if t is defined by an algebraic closure of the residue field of a point of Spec $\mathcal{O}_{S,s}$. In this case, s is called a **specialization** of t. If f: X → S is a proper smooth morphism the sheaf $R^q f_* \mathbf{Z}/(n)$ is locally constructible (constant), and ifs is a specialization of t one can define a **cospecialization** of $H^q(X_{t,\text{ét}}, \mathbf{Z}/(n))$ to $H^q(X_{s,\text{ét}}, \mathbf{Z}/(n))$ which is bijective. Let X be a scheme of finite type over a field k. If l #ch(k) is prime, one can define the l-adic cohomology $H^q(X, \mathbf{Q}_l) = \lim_n H^q(X, \mathbf{Z}/(l^n)) \otimes_{\mathbf{Z}_l} \mathbf{Q}_l$. The l-adic cohomology has as many good properties as the classical cohomology (→ Section U).

References

[1] A. Altman and S. L. Kleiman, Introduction to Grothendieck duality theory, Lecture notes in math. 146, Springer, 1970.
[2] M. Artin, Algebraization of formal moduli; I, Global analysis, Univ. Tokyo Press and Princeton Univ. Press, 1969, 21-71; II, Existence of Modifications, Ann. Math., (2) 91 (1970), 88-135.
[3] M. Artin, Algebraic approximation of structures over complete local rings, Publ. Math. Inst. HES, 36 (1969), 23358.
[4] M. Artin, Théorèmes de représentabilité pour les espaces algébriques, Les Presses de l'Université de Montréal, 44, 1973.
[5] M. Artin and B. Mazur, Etale homotopy, Lecture notes in math. 100, Springer, 1969.
[6] M. Artin and D. Mumford, Some elementary examples of unirational varieties which are not rational, Proc. London Math. Soc., (3) 25 (1972), 75595.
[7] A. Ash, D. Mumford, M. Rapoport, and Y. Tai, Smooth compactification of locally symmetric varieties, Math.-Sci. Press, 1975.
[8] M. F. Atiyah, Vector bundles over an elliptic curve, Proc. London Math. Soc., (3) 7 (1957), 414-452.
[9] M. F. Atiyah and R. S. Ward, Instantons and algebraic geometry, Comm. Math. Phys., 55 (1977), 117–124.

[10] P. Berthelot, Cohomologie cristalline des schémas de caractéristique $p > 0$, Lecture notes in math. 407, Springer, 1974.
[11] W. Barth, Some properties of stable rank-2 vector bundles on P_n, Math. Ann., 226 (1977), 1255150.
[12] W. L. Baily and A. Borel, Compactification of arithmetic quotients of bounded symmetric domains, Ann. Math., (2) 84 (1966), 442-528.
[13] S. Bloch, D. Lieberman, and A. Kas, Zero cycles on surfaces with $p_g = 0$, Compositio Math., 33 (1976), 135–145.
[14] E. Brieskorn, Rationale Singularitaten Komplexer Flächen, Inventiones Math., 4 (1968), 336-358.
[15] Séminaire C. Chevalley, Anneaux de Chow et applications, Ecole Norm. Sup., 1958.
[16] C. H. Clemens and P. A. Griffiths, The intermediate Jacobian of the cubic threefold, Ann. Math., (2) 95 (1972), 281–356.
[17] V. 1. Danilov, The geometry of toric varieties, Russian Math. Surveys, 33 : 2, (1978), 977154.
[18] P. Deligne, Théorie de Hodge: I, Actes Congr. Intern. Math., 1970, Nice, Gauthier-Villars, 1971, vol. 1, 425–430; II, Publ. Math. Inst. HES, 40 (1971), 5–57; III, 44 (1974), 5577.
[19] P. Deligne, La conjecture de Weil I, Publ. Math. Inst. HES, 43 (1943), 273–307.
[20] P. Deligne, Le groupe fondamental du complément d'une courbe plane n'ayant que des points doubles ordinaires est abélien (d'après W. Fulton), Séminaire Bourbaki, exp. 543, Nov. 1979, Lecture notes in math. 842, Springer, 1980.
[21] M. Demazure, Sous-groupes algébriques de rang maximum du groupe de Cremona, Ann. Sci. Ecole Norm. Sup., 3 (1970), 507–588.
[22] W. Fulton and J. Hansen, A connectedness theorem for projective varieties, with applications to intersections and singularities of mappings, Ann. Math., (2) 110 (1979), 159–166.
[23] D. Gieseker, On the moduli of vector bundles on an algebraic surface, Ann. Math., (2) 106 (1977), 45–60.
[24] P. A. Griffiths, Periods of integrals on algebraic manifolds; summary of main results and discussion of open problems, Bull. Amer. Math. Soc., 76 (1970), 228-296.
[25] P. A. Griffiths, On the periods of certain rational integrals I, II, Ann. Math., (2) 90 (1969), 460–495, 496–541.
[26] P. A. Griffiths, Periods of integrals on algebraic manifolds III, Publ. Math. Inst. HES, 38 (1970), 1255180.
[27] P. A. Griffiths and W. Schmid, Recent developments in Hodge theory, Discrete Subgroups of Lie Groups and Application to Moduli, Oxford Univ. Press, 1975, 3 l-127.
[28] P. A. Griffiths and J. Harris, Principles of algebraic geometry, Wiley, 1978.
[29] A. Grothendieck, Eléments de géométrie algébriques, Publ. Math. Inst. HES: 1, 4 (1960) (Springer, 1971); II, 8 (1961); III, 11 (1961), 17 (1963); IV, 20 (1964), 24 (1965), 28 (1966), 32 (1967).
[30] Cited as SGA. A. Grothendieck and others, Séminaire de géométrie algébrique: SGA 1, Revêtements étales et groupe fondamental, Lecture notes in math. 224, Springer, 1971; SGA 2, Cohomologie locale des faisceaux et théorèmes de Lefschetz locaux et globaux, North-Holland, 1962; SGA 3 (with M. Demazure), Schémas en groupes I, II, III, Lecture notes in math. 151, 152, 153, Springer, 1970; SGA 4 (with M. Artin and J. L. Verdier), Théorie des topos et cohomologie étale des schémas, Lecture notes in math. 269, 270, 305, Springer, 1972-1973; SGA 5, Cohomologie l-adique et fonctions L; SGA 6, Théorie des intersections et théorème de Riemann-Roch, Lecture notes in math. 225, Springer, 1971; SGA 7, Groupes de monodromie en géométrie algébrique, pt. I, Lecture notes in math. 288; pt. II (by P. Deligne and N. Katz), 340, Springer, 197221973.
[31] B. Grünbaum, Polytopes, graphs and complexes, Bull. Amer. Math. Soc., 76 (1970), 1131–1201.
[32] W. J. Haboush, Reductive groups are geometrically reductive, Ann. Math., (2) 102 (1975), 67783.
[33] R. Hartshorne, Residues and duality, Lecture notes in math. 20, Springer, 1966.
[34] R. Hartshorne, Ample subvarieties of algebraic varieties, Lecture notes in math. 156, Springer, 1970.
[35] R. Harshorne, Algebraic geometry, Graduate texts in math. 52, Springer, 1977.
[36] R. Hartshorne, Algebraic vector bundles on projective spaces: A problem list, Topology, 18 (1979), 117-128.
[37] H. Hironaka, Resolution of singularities of an algebraic variety over a field of characteristic zero, Ann. Math., (2) 79 (1964), 109-326.
[38] M. Hochster, Rings of invariants of tori, Cohen-Macaulay rings generated by monomials and polytopes, Ann. Math., (2) 96, (1972), 3188337.
[39] W. V. D. Hodge and M. F. Atiyah, Integrals of the second kind on an algebraic variety, Ann. Math., (2) 62 (1955), 56-91.
[40] G. Horrocks and D. Mumford, A rank-2 vector bundle on $\mathbf{P}^4$ with 15,000 symmetries, Topology, 12 (1973), 63-81.
[41] J. Igusa, Arithmetic variety of moduli for genus two, Ann. Math., (2) 72 (1960), 612-649.
[42] J. Igusa, Betti and Picard numbers of

abstract algebraic surfaces, Proc. Nat. Acad. Sci. US, 46 (1960), 7244726.
[43] H. Inose and M. Mizukami, Rational equivalence of O-cycles on some surfaces of general type with $p_g = 0$, Math. Ann., 244 (1979), 205-217.
[44] M.-N. Ishida, Torus embeddings and dualizing complexes, Tôhoku Math. J., 32, (1980), 111-146.
[45] V. A. Iskovskiï and Yu. 1. Manin, Three-dimensional quartics and counterexamples to the Lüroth problem, Math. USSR-Sb., 15 (1971), 141-166. (Original in Russian, 1971.)
[46] S. Lojasiewicz, Triangulation of semianalytic sets, Ann. Scuola Norm. Sup. Pisa, 18 (1964), 449-474.
[47] G. Kempf, F. Knudsen, D. Mumford, and B. Saint-Donat, Toroidal embeddings I, Lecture notes in math. 339, Springer, 1973.
[48] S. L. Kleiman, Toward a numerical theory of ampleness, Ann. Math., 84 (1966), 293-344.
[49] D. Knutson, Algebraic spaces, Lecture notes in math. 203, Springer, 197 1.
[SO] K. Kodaira, Collected works I–III, Iwanami, 1975.
[51] M. Maruyama, Moduli of stable sheaves, I, II, J. Math. Kyoto Univ., 17 (1977), 91–126 and 18 (1978), 5577614.
[52] T. Matsusaka, Theorey of Q-varieties, Publ. Math. Soc. Japan, no. 8, Tokyo, 1965.
[53] T. Matsusaka, Polarized varieties with a given Hilbert polynomial, Amer. J. Math., 94 (1972), 102771077.
[54] T. Matsusaka and D. Mumford, Two fundamental theorems on deformations of polarized varieties, Amer. J. Math., 86 (1964), 668-684.
[55] M. Miyanishi, On curves on rational and unirational surfaces, Tata lecture notes 60, Springer, 1978.
[56] B. G. Moishezon, A criterion for projectivity of complete algebraic abstract varieties, Amer. Math. Soc. Transl., 63 (1967), 1–50.
[57] B. G. Moishezon, Complex surfaces and connected sums of complex projective planes, Lecture notes in math. 603, Springer, 1977.
[SS] S. Mori, On a generalization of complete intersections, J. Math. Kyoto Univ., 15 (1975), 619–646.
[59] S. Mori, Projective manifolds with ample tangent bundles, Ann. Math., (2) 110 (1979), 593-606.
[60] D. Mumford, Geometric invariant theory, Springer, 1965; second edition, 1982.
[61] D. Mumford, Pathologies of modular algebraic surfaces, Amer. J. Math., 83 (1961), 3399342; Further pathologies in algebraic geometry, ibid., 84 (1962), 642-648; Pathologies III, ibid., 89 (1967), 944104.
[62] D. Mumford, Stability of projective varieties, Enseignement Math., 23 (1977), 39-110.
[63] M. Nagata, Imbedding of an abstract variety in a complete variety, J. Math. Kyoto Univ., 2 (1962), 1–10.
[64] Y. Nakai, A criterion of an ample sheaf on a projective scheme, Amer. J. Math., 85 (1963), 14-26.
[65] Y. Namikawa, A new compactilïcation of the Siegel space and the degeneration of Abelian varieties, I, II, Math. Ann., 221 (1976), 97-141 and 201–241.
[66] A. Néron, Problèmes arithmétiques et géométriques attachés à la notion du rang d'une courbe algébrique, Bull. Soc. Math. France, 80 (1952), 101–166.
[67] T. Oda, On torus embeddings and applications (based on joint work with K. Miyake), Tata lecture notes 57, Springer, 1978.
[68] T. Oda and C. S. Seshadri, Compactifications of the generalized Jacobian variety, Trans. Amer. Math. Soc., 253 (1979), 1–90.
[69] C. Okonek, M. Schneider, and H. Spindler, Vector bundles on complex projective spaces, Progress in math. 3, Birkhäuser, 1980.
[70] H. Popp, On moduli of algebraic varieties II, Compositio Math. 28 (1974), 51-81; III, Fine moduli spaces, ibid., 31 (1975), 237-258.
[71] H. Popp, Moduli theory and classification theory of algebraic varieties, Lecture notes in math. 620, Springer, 1977.
[72] D. Quillen, Projective modules over polynomial rings, Inventiones Math., 36 (1976), 167–171.
[73] M. A. Reid, Canonical 3-folds, Journées de Géométrie Algébrique, July 1979, 273-310, Sijthoff and Noordhoff, 1980.
[74] A. A. Rojtman, Rational equivalence of zero cycles, Math. USSR-Sb., 18 (1972), 571–588. (Original in Russian, 1972.)
[75] A. A. Rojtman, The torsion of the group of O-cycles modulo rational equivalence, Ann. Math., (2) 111 (1980), 553-569.
[76] L. Roth, Algebraic threefolds, Springer, 1955.
[77] I. Satake, On the arithmetic of tube domains, Bull. Amer. Math. Soc., 79 (1973), 1076- 1094.
[78] M. Schlessinger, Functors of Artin rings, Trans. Amer. Math. Soc., 130 (1968), 208–222.
[79] C. S. Seshadri, Quotient spaces modulo reductive algebraic groups, Ann. Math., (2) 95 (1972), 511–556.
[SO] J.-P. Serre, Un théorème de dualité, Comm. Math. Helv., 29 (1955), 9-26.
[81] J.-P. Serre, Faisceaux algébriques cohérents, Ann. Math., (2) 61 (1955), 197-278.
[82] J.-P. Serre, Géométrie algébrique et géométrie analytique, Ann. Inst. Fourier, 6 (1956), 1-42.
[83] J.-P. Serre, Exemple de variété projectives

conjuguées non homéomorphes, C. R. Acad. Sci. Paris, 258 (1964), 419444196.
[84] I. R. Shafarevich, Basic algebraic geometry, Springer, 1974.
[SS] G. Shimura, Moduli and fibre systems of Abelian varieties, Ann. Math., (2) 83 (1966), 294-338.
[86] G. Shimura, Algebraic varieties without deformation and the Chow variety, J. Math. Soc. Japan, 20 (1968), 336-341.
[87] R. Stanlay, The number of faces of a simplicial convex polytope, Advances in math., 35 (1980), 2366238.
[88] H. Sumihiro, Equivariant completion I, J. Math. Kyoto Univ. 14 (1974), l-28; II, 15 (1975), 573-605.
[89] B. Teissier, Variétés toriques et polytopes, Sém. Bourbaki, 1980/81, exp. 565, Lecture notes in math. 901, Springer, 1981.
[90] K. Ueno, Classification theory of algebraic varieties and compact complex spaces, Lecture notes in math. 439, Springer, 1975.
[91] H. Umemura, Sur les sous-groupes algébriques primitifs du groupe de Cremona a trois variables, Nagoya Math. J., 79 (1980), 47767.
[92] A. Weil, Foundations of algebraic geometry, Amer. Math. Soc. Colloq. Publ., second edition, 1962.
[93] 0. Zariski, Collected papers, I, II, III, IV, MIT Press, 1972, 1973, 1978, 1979.

17 (X1.11) Algebroidal Functions

A. General Remarks

If an analytic function f satisfies an †irreducible algebraic equation

$$A_0(z)f^k + A_1(z)f^{k-1} + \ldots + A_k(z) = 0 \qquad (1)$$

with single-valued †meromorphic functions $A_j(z)$ in a domain G in the complex z-plane, then f is called a **k-valued algebroidal function** in G. With no loss of generality, we can assume that there is no common zero among the $A_j(z)$ and that all the $A_j(z)$ are †holomorphic in G. When $k = 1$, the solution of (1) is a single-valued meromorphic function in G. If all the $A_j(z)$ are polynomials, then f is an †algebraic function. Thus algebroidal functions can be regarded as extensions of single-valued to multiple-valued functions and also as extensions of algebraic to †transcendental functions. Since (1) is irreducible, its discriminant D(z) does not vanish identically. For all the points a satisfying $D(a)\neq 0$, $A_0(a)\neq 0$, $a\in G$, equation (1) determines k holomorphic function elements $f_1(z)$, $f_k(z)$ in a suitable neighborhood of a that determine the analytic function f. They can be †prolonged analytically in G in the wider sense. At any point satisfying $A_0(z) = 0$, at least one element has a pole; and at any point satisfying $D(z) = 0$, there may appear ramified elements. Therefore an algebroidal function can be defined as a finitely multiple-valued analytic function in G with the exception of poles and †algebraic branch points. Every algebroidal function f(z) determines a †Riemann surface, which may be considered a †covering surface 3 of G. This surface 3 is a k-sheeted †covering surface over G with no singular point except for algebraic branch points. Also, f reduces to a single-valued meromorphic function on 3, and all the function elements over a point z are different. A k-valued algebroidal function can also be characterized by this property.

These two (equivalent) delïnitions of algebroidal functions give rise to two distinct methods of studying these functions. In adopting the first definition, we can make use of results in the single-valued case, as did G. Rémoundos and G. Valiron. When the latter definition is adopted, we can use several methods that are also applicable in the single-valued case, as did H. Selberg and E. Ullrich.

Research on algebroidal functions has been carried out mainly for the case where G is the finite plane $|z|<\infty$ or the unit disk $|z|<1$. Almost all the known results for algebroidal functions are extensions of those on single-valued functions, but several results particularly relevant to algebroidal functions have been discovered. The existence of branch points makes it difficult to investigate algebroidal functions in some cases.

B. Absolute Value

Among several results on the absolute values of algebroidal functions, the †maximum principle, one of the basic principles, holds on the Riemann surface 3. The following relation holds among the $A_j(z)$ and $|f_\nu(z)|$. Assume that (1) has the form

$$f^k + A_1(z)f^{k-1} + \ldots + A_k(z) = 0. \qquad (1')$$

Then $\log(1+A(z))/\log(1+F(z))$ is bounded, where $A(z)=\max|A_j(z)|$ and $F(z)=\max|f_\nu(z)|$.

An algebroidal function that has no pole in $|z|<\infty$ is called an **entire algebroidal function.** The successive derivatives of an entire algebroidal function may have poles at every branch point, which is a departure from the case of single-valued integral functions. An algebroidal function defined by (1) or (1') with entire functions $A_j(z)$ and zero-free $A_0(z)$ is

entire. Dividing both sides of (1) by $A_0(z)$, we obtain equation (1′), and all its coefficients are entire. In equation (1′), the †order, type, and class of an entire algebroidal function coincide with those of the largest $A_j(z)$ (→ 429 Transcendental Entire Functions).

For an entire algebroidal function of order less than 1/2, F(z) tends to infinity uniformly along a sequence of concentric circles $|z|=r_n$ ($r_n\to\infty$). However, it can be shown that not all the branches of $f(z)$ necessarily tend to infinity on the Riemann surface $\mathfrak{Z}$; in this sense, †Wiman's theorem does not remain true. But, using $\min_{|z|=r}\max_\nu|f_\nu(z)|$, one can obtain some extensions of the generalized Wiman theorem.

C. Picard's Theorem and Its Extension

Rémoundos first extended †Picard's theorem and †Borel's theorem to an algebroidal function. Every k-valued transcendental algebroidal function in the finite plane takes on every value infinitely often with at most $2k$ exceptional values. There are examples where 2k values are actually omitted. Hence the theorem is the best one possible in this sense. T. Varopoulos (*Bull. Soc. Math. France*, 53 (1925), 23-34) introduced the **degeneracy index**

$$\lambda=\dim\{(c_0,c_1,\dots,c_k)\in C^{k+1}\,|\,c_0A_0+c_1A_1+\dots+c_kA_k=0\}$$

($0\leqslant\lambda\leqslant k-1$) into equation (1) and showed that the number of Picard's exceptional values of f is at most $k+\lambda+1$. There is no single-valued meromorphic function with $\lambda>0$, so this result is relevant only for algebroidal functions. Extending this idea further, J. Dufresnoy and others obtained more precise results. The situation is the same for the Borel exceptional values. That is, the †convergence exponent of $f(z)-w=0$ coincides with the †order of f except for at most $2k$ values, for which the convergence exponents of f are less than the order of f (E. Borel). There are at most $2k$ polynomials P(z) for which $f(z)-P(z)=0$ has at most a finite number of roots (Borel), and furthermore, using the degeneracy index, we can give more precise results.

Selberg was the first to extend the †Nevanlinna theory of meromorphic functions to algebroidal functions (→ 272 Meromorphic Functions). Almost simultaneously, Valiron obtained the same results starting from the coefficients of (1); then Ullrich improved the results by considering the effect of branch points of $\mathfrak{Z}$.

Let $\mathfrak{Z}_r$ be the part of $\mathfrak{Z}$ over $|z|<r$, let $n(r,w)$ be the number of roots of $f(z)-w=0$ in $\mathfrak{Z}_r$, and let $n(r,\mathfrak{Z})$ be the number of branch points of $\mathfrak{Z}_r$. With these notations we write

$$N(r,w)=\frac{1}{k}\int_0^r(n(t,w)-n(0,w))\frac{dt}{t}+\frac{n(0,w)}{k}\log r,$$

$$m(r,w)=\frac{1}{2k\pi}\int_{|z|=r}\log^+\frac{1}{|f(re^{i\varphi})-w|}\,d\varphi,\quad w\neq\infty,$$

$$T(r,w)=m(r,w)+N(r,w).$$

Let $T(r,f)$ be the †logarithmic integral of the spherical area of the image of $\mathfrak{Z}_r$ under $w=f(z)$,

$$T(r,f)=\frac{1}{k}\int^r\frac{dt}{t}\iint_{\mathfrak{Z}_t}\frac{|f'(te^{i\varphi})|^2}{(1+|f(te^{i\varphi})|^2)^2}t\,dt\,d\varphi,$$

and let $N(r,\mathfrak{Z})$ be the logarithmic integral of $n(r,\mathfrak{Z})$. Then we have $T(r,w)=T(r,f)+O(1)$ and the **ramification theorem:**

$$N(r,\mathfrak{Z})<(2k-2)\,T(r,f)+O(1).$$

Also, $T(r,f)=O(\log r)$ holds if and only if f is algebraic. Let $A(z)$ be the maximum of $|A_j(z)|$, and let

$$\mu(r)=\frac{1}{2k\pi}\int_0^{2\pi}\log A(re^{i\varphi})\,d\varphi.$$

Then $T(r,f)=\mu(r)+O(1)$. As the second fundamental theorem we have

$$\sum_{\nu=1}^{q}N(r,w_\nu)>(q-2)T(r,f)-N(r,\mathfrak{Z})+\sum_{\nu=1}^{q}N_1(r,w_\nu)+O(\log rT)$$
$$>(q-2k)\,T(r,f)+\sum_{\nu=1}^{q}N_1(r,w_\nu)+O(\log rT),$$

where $N_1(r,w)$ is the logarithmic integral of $n_1(r,w)$ which is the sum of the †multiplicity minus one of all the roots of $f(x)-w=0$ in $\mathfrak{Z}_r$. Furthermore, the **deficiency, ramification index** of f(z), and **ramification index** of the surface $\mathfrak{Z}$ are defined by

$$\delta(w)=1-\limsup N(r,w)/T(r,f)$$
$$=\liminf m(r,w)/T(r,f),$$

$$\mathfrak{D}(w)=\liminf N_1(r,w)/T(r,f),$$

$$\xi=\liminf N(r,\mathfrak{Z})/T(r,f).$$

With these notations, we have

$$\sum\delta(w_\nu)+\sum\mathfrak{D}(w_\nu)\leqslant2+\xi\leqslant2k.$$

These results contain the Picard theorem and the Borel theorem. Furthermore, by considering the effect of branch points, the Ahlfors theory of covering surfaces can be extended to algebroidal functions (Y. Tumura). By using this result, †Bloch's theorem can be obtained very simply.

But these results do not contain the Varopoulos result. Taking into consideration the degeneracy index λ, H. Cartan proved the following inequality when $\lambda=0$:

$$(q-k-1)T(r,f)<\sum_{\nu=1}^{q}N_k(r,w_\nu)+S(r),$$

and for general λ he conjectured that the number $q-k-1$ may be changed to the number $q-k-\lambda-1$. This conjecture is still unsolved except for some special cases where for example, $q=k+\lambda+2$ (N. Toda, *Nagoya* Math. *J.*, 91 (1983), 37–47).

Interesting results follow upon a study of this degeneracy index. For example, when f is an entire algebroidal function, if $\sum_{\nu=1}^{2k-1}\delta(w_\nu)>2k-2$ $(w_\nu\neq\infty)$, there are at least $k-1$ †Picard's exceptional values in $\{w_\nu\}$ (Niino, Ozawa, and Toda).

With respect to the relation between the number of exceptional values and the order of fin $|z|<\infty$, there are some results similar to those for single-valued functions. For example, if f has $k+1$ Picards exceptional values, the order off must be a positive integer or ∞.

D. Asymptotic Values and Other Results

In the single-valued case, Valiron, L. Ahlfors, W. H. J. Fuchs, A. Edrei, W. K. Hayman, and others studied the †Borel direction, the number of †asymptotic values, the relation between the deficiency values and the asymptotic values, etc. However, almost none of the corresponding results holds for the algebroidal case as shown by several counterexamples.

There is no relationship between the order of an entire algebroidal function and the number of its finite †asymptotic values, which is quite different from the single-valued case. Furthermore, it is possible to have an infinite number of asymptotic values even if the order is equal to zero. If an algebroidal function f satisfies $\liminf T(r,f)/(\log r)^2<+\infty$, then it has at most k asymptotic values (Valiron-Tumura). The Ahlfors theorem, which is concerned with the number of †direct transcendental singular points of the inverse function and the order of a meromorphic function in the single-valued case, was extended to the algebroidal function by Lü Yinian (*Scientia Sinica*, 23 (1980)).

The †Julia direction or the Borel direction for an algebroidal function is defined not on 3 but on $|z|<$ cc because of the appearance of branch points. With respect to the Julia direction, there are results similar to those for the single-valued case, but it is unknown in general whether the Bore1 direction exists. A. Rauch proved that if $\int^\infty T(r,f)/r^{\rho+1}\,dr=\infty$, then there is a sector with an angle of at least π/ρ in which $L(\varphi)=\int^\infty\log^+F(re^{i\varphi})/r^{\rho+1}\,dr$ diverges. Apart from the theory of distribution of values, Selberg obtained some conditions under which the inverse functions of †Abelian integrals of a special kind reduce to algebroidal functions.

References

[1] G. Rémoundos, Extensions aux fonctions algébroïdes multiformes du théorème de M. Picard et de ses généralisations, Gauthier-Villars, 1927.
[2] H. Selberg, Algebroide Funktionen und Umkehrfunktionen Abelscher Integrale, Avhandl. Norske Videnskaps-Akad., Oslo, 8 (1934), 1-72.
[3] N. Toda, Surrounding algebroidal functions (in Japanese), Bull. Math. Soc. Japan, 24 (1972), 2006209.
[4] H. Cartan, Sur les zéros des combinaisons linéaires de p fonctions holomorphes données, Mathematica, 7 (1933), 5-31.

18 (X.25) Almost Periodic Functions

A. History

The theory of almost periodic functions was originated by H. Bohr in 1924 as a result of his study of †Dirichlet series. The theory provides a method of studying a wide class of trigonometric series (→ 159 Fourier Series) of general type. Further generalizations were made by N. Wiener, V. V. Stepanov, A. S. Besikovich, S. Bochner, and others. H. Weyl, J. von Neumann, and others clarified the relations between this theory and †representation of groups, specifically, the relations between almost periodic functions in a †topological group and representation theory of a †compact group.

B. Almost Periodic Functions in the Sense of Bohr

Let f(x) be a complex-valued continuous function defined for all real values of x. A number τ is called a **translation numher** of f(x) belonging to $\varepsilon>0$ if

$$\sup_{-\infty<x<\infty}|f(x+\tau)-f(x)|\leqslant\varepsilon.$$

If for any $\varepsilon>0$ there exists a number $l(\varepsilon)>0$

such that any interval of length $l(\varepsilon)$ contains a translation number off belonging to ε, then f(x) is called **almost periodic in the sense of Bohr.** We denote by **B** the set of all almost periodic functions in the sense of Bohr.

If f(x) is †periodic with a period p, then $f \in$ **B,** because each number $l \geqslant p$ plays the role of $l(\varepsilon)$ for any $\varepsilon > 0$. Any $f \in$ **B** is bounded and uniformly continuous. A necessary and sufficient condition for a bounded continuous function on $(-\infty, \infty)$ to belong to **B** is that for any given sequence $\{h_n\}$ of real numbers, there exist a subsequence $\{h_{n_\nu}\}$ such that the sequence of functions $\{f(x+h_{n_\nu})\}$ is uniformly convergent in $(-\infty, \infty)$; i.e., the set $\{f(x+h) \mid h \in (-\infty, \infty)\}$ is †totally bounded with respect to the uniform norm $\|f\| = \sup |f(x)|$ in the space of bounded continuous functions in $(-\infty, \infty)$.

If f(x) $\in$ **B,** then $f(-x)$, $\overline{f(x)}$, $\alpha f(x)$ (where α is a complex number), and $f(x+h)$ (where h is a real number)$\in$**B**. If $f(x)$, $g(x) \in$ **B**, then $f(x) \pm g(x)$ and $f(x)g(x) \in$ **B**. If $f_n(x) \in$ **B** and $\{f_n(x)\}$ converges uniformly to $f(x)$, then $f(x) \in$ **B**. For any real number A, $\exp i\lambda x$ (where i is the †imaginary unit) is continuous and periodic. Hence the polynomial function $\sum_{n=1}^{m} \alpha_n \exp i\lambda_n x \in$ **B.** Moreover, if the latter function converges uniformly to $\sum \alpha_n \exp i\lambda_n x$ as m tends to ∞, then the limit function is also an element of **B.** The polynomial $\sum_{n=1}^{m} \alpha_n \exp i\lambda_n x$ and the series $\sum_{n=1}^{\infty} \alpha_n \exp i\lambda_n x$ are called a **generalized trigonometric polynomial** and a **generalized trigonometric series**, respectively.

For any $f \in$ **B**, its mean exists:

$$M[f] = \lim_{T \to \infty} \frac{1}{T} \int_a^{a+T} f(x)\, dx.$$

The convergence of the right-hand formula is uniform in $a \in (-\infty, \infty)$, and the limit is independent of the choice of a. Thus $M[f]$ is a †linear functional defined on **B.** Since $M[\exp i\lambda x] = 1$ for $\lambda = 0$ and $=0$ for $\lambda \neq 0$, the family $\{\exp i\lambda x \mid -\infty < \lambda < \infty\}$ is an †orthonormal system with respect to the †inner product $(f, g) = M[f(x)\overline{g(x)}]$ defined on **B.** Let $\alpha(\lambda) = M[f(x)\exp(-i\lambda x)]$ for any $f \in$ **B;** then there exist countably many values of λ for which $\alpha(\lambda)$ differs from zero. Denote these values of λ by $\lambda_1, \lambda_2, \ldots$ and write $\alpha(\lambda_n) = \alpha_n$. We call the numbers $\alpha_1, \alpha_2, \ldots, \alpha_n, \ldots$ the **Fourier coefficients** of f(x). The formal series $\sum_{n=1}^{\infty} \alpha_n \exp i\lambda_n x$ is called the **Fourier series** of $f(x)$. Moreover, the **Parseval equality** $M[|f(x)|^2] = \sum_{n=1}^{\infty} |\alpha_n|^2$ is valid for any $f \in$ **B**. For every periodic function, these definitions coincide with the ordinary Fourier coefficients and the Fourier series (→ 159 Fourier Series). Any almost periodic function in the sense of Bohr is uniquely determined by its Fourier coefficients; i.e., if two almost periodic functions have the same Fourier series, then they are identical. For any $f \in$ **B,** its Fourier series does not always converge uniformly, but $f(x) \in$ **B** can be approximated uniformly by a sequence of trigonometric polynomials. Hence the almost periodic functions in the sense of Bohr are also called **uniformly almost periodic functions.**

C. Generalizations of Almost Periodic Functions

Let $C(-\infty, \infty)$ be the space (→ 168 Function Spaces) of all bounded continuous functions on $(-\infty, \infty)$ with distance $\rho(f, g) = \sup_{-\infty < x < \infty} |f(x) - g(x)|$. Then a uniformly almost periodic function is the limit of a sequence of trigonometric polynomials with respect to this distance. Generally, let ρ be a †distance function introduced in a function space (whose elements are not necessarily continuous in $(-\infty, \infty)$). Then the limit of a sequence of generalized trigonometric polynomials with respect to the distance ρ is called an **almost periodic function with respect to** ρ. For example, for $p \geqslant 1$, we set

$$D_{S^p}[f, g] = \sup_{-\infty < a < \infty} \left\{ \int_a^{a+1} |f(x) - g(x)|^p dx \right\}^{1/p},$$

$$D_{W^p}[f, g] = \lim_{l \to \infty} \sup_{-\infty < a < \infty} \left\{ \frac{1}{l} \int_a^{a+l} |f(x) - g(x)|^p dx \right\}^{1/p}$$

These are distance functions. The properties of the corresponding almost periodic functions and their relations to other classes of almost periodic functions have been studied by Besikovich [1].

D. Analytic Almost Periodic Functions

Let D be a strip domain, $a < \mathrm{Re}\, z < b$, defined in a complex plane. For any †holomorphic function $f(z)$ in D and $\varepsilon > 0$, a real number τ is called a **translation number** off belonging to $\varepsilon > 0$ if $\sup_{z \in D} |f(z+i\tau) - f(z)| \leqslant \varepsilon$. If for any $\varepsilon > 0$ there exists a number $l(\varepsilon) > 0$ such that any interval of length $l(\varepsilon)$ contains a translation number off belonging to ε, then f(x) is called an **analytic almost periodic function** in D. We denote the set of all analytic almost periodic functions in $D = \{a < \mathrm{Re}\, z < b\}$ by $\mathbf{A}(a, b)$. If we fix an x in $a < x < b$, then $g(y) = f(x + iy)$ for any $f(z) \in \mathbf{A}(a, b)$ belongs to **B.**

For any $f \in \mathbf{A}(a, b)$ there corresponds a †Dirichlet series $\sum_{n=1}^{\infty} \alpha_n \exp \lambda_n z$ such that two analytic almost periodic functions are identical if the corresponding Dirichlet series are

identical. Here the coefficients

$$\alpha_n = M_y[f(x+iy)\exp(-\text{ii, y})]$$

are determined independently of x (a $<x<b$), and Parseval's identity

$$M_y[|f(x+iy)|^2] = \sum_{n=1}^{\infty} |\alpha_n|^2 \exp 2\lambda_n x$$

holds (Bohr [3]). If the series

$$\sum \alpha_n \exp \lambda_n x \exp i\lambda_n y$$

at x = a and x = **b** represent the Fourier series of $f_a(y)$ and $f_b(y) \in \mathbf{B}$, respectively, then there exists $f \in \mathbf{A}(a, b)$ such that f(z) is continuous on $\bar{D}$ and

$$f(a+iy) = f_a(y), \qquad f(b+iy) = f_b(y).$$

The behavior of $f \in \mathbf{A}(a,$ **b)** at the boundary or exterior points of the domain $D = \{a <$ Re $z <$ **b}** has also been investigated by Besikovich [1].

E. Almost Periodic Functions on Groups

Von Neumann defined almost periodic functions on any group, generalizing the characterization of uniformly almost periodic functions on $(-\infty, \infty)$. Let B(G) be the set of all complex-valued bounded functions on a group G. Then B(G) is a metric space with the †distance $\rho(f,g) = \sup_{x \in G} |f(x) - g(x)|$. If for any $f \in B(G)$ the set A,= $\{f_{a,b}(x) = f(axb) \mid a, b \in G\}$ is totally bounded in the metric space B(G), we call f an **almost periodic function on the group** G. This condition is equivalent to the total boundedness of $B_f = \{f_a(x) = f(xa) \ \ a \in \mathrm{G}\}$ or $C_f = \{{}_af(x) = f(ax) \mid a \in G\}$. We denote the set of almost periodic functions on G by $\mathscr{A}(G)$.

For f(x), $g(x) \in \mathscr{A}(G)$, the linear combinations $a \cdot f(x) + b \cdot g(x)$ (a, $b \in \mathbf{C}$) and the product $f(x)g(x)$ are both contained in $\mathscr{A}(G)$. If $f_n \in \mathscr{A}(G)$ and $\{f_n\}$ converges to f uniformly on G, then $f \in \mathscr{A}(G)$. If $f \in \mathscr{A}(G)$, then $f_{a,b}, f_a, {}_af \in$ d(G) also. Hence $\mathscr{A}(G)$ is a closed subalgebra invariant under two-sided translation in the †Banach algebra i?(G). For any $f \in \mathscr{A}(G)$ there exists only one number $M[f]$ in the closure (with respect to the distance **p** in B(G)) of

$$A'_f = \left\{ \sum_{i=1}^{n} c_i f(a_i x b_i) \,\middle|\, c_i > 0, \sum c_i = 1, a_i, b_i \in G \right\}$$

(= the least closed †convex set including A,). We call $M[f]$ the **mean** off on G. The mapping $f \to M[f]$ is a linear functional on $\mathscr{A}(G)$, and we have $M[f] \geqslant 0$ if $f \geqslant 0$.

F. Relation to Bounded Representation

Suppose that we are given a finite-dimensional matrix representation **D(x)** = (d,(x)) of a group G. Then the following three conditions are equivalent: (i) All the $d_{ij}(x)$ are bounded on G. (ii) All the d,(x) are almost periodic in G. (iii) The representation D is †equivalent to a representation by unitary matrices. The inner product $(f,$ g) = $M[f(x)g(x)]$ provides the algebra $\mathscr{A}(G)$ with the structure of a †pre-Hilbert space. Let H(G) be the Hilbert space that is the completion of $\mathscr{A}(G)$. If we select D"(x) = $(d_{ij}^{\lambda}(x))$ from each L, where L is an equivalence class of bounded irreducible representations of G, and if n_λ is the order of D^λ, then $\{(1/\sqrt{n_\lambda}) d_{ij}^{\lambda}(x) \mid 1 \leqslant$ i, j $\leqslant n_\lambda$, $\lambda \in L\}$ is a †complete orthonormal system in the Hilbert space H(G). Any f(x) $\in$ $\mathscr{A}(G)$ can be approximated uniformly in G by a finite linear combination of the $d_{ij}(x)$.

G. Almost Periodic Functions on Topological Groups

When G is a †separated topological group, we denote the set of all continuous functions on G contained in $\mathscr{A}(G)$ by $\mathscr{A}_*(G)$. The statements of the theorems in the previous section concerning $\mathscr{A}(G)$ and the representation D remain valid if we replace $\mathscr{A}(G)$ by d,(G) and replace D by a continuous representation of G. In particular, if G is the additive group of real numbers **R,** then **d,(R)** is exactly **B.**

H. Relation to Compact Groups

Every continuous function on a compact group G is almost periodic; i.e., $\mathscr{A}_*(G)$ = C(G). The mean value $M[f]$ of $f \in \mathscr{A}_*(G)$ is identical to $\int_G f(x)dx$, where the †Haar measure **dx** is normalized so that $\int_G dx = 1$. In this case, the theory of bounded representations discussed above is the Peter-Weyl theory (- 69 Compact Groups).

In general, let G be a separated topological group. There exists a continuous homomorphism φ of G onto a compact group K = K(G) with the following two properties: (i) For any compact group K' and a continuous homomorphism $\varphi': G \to K'$, there exists a continuous homomorphism $\psi: \boldsymbol{K} \to K'$ such that $\varphi' = \psi$ o φ. (ii) Such a pair $\boldsymbol{K} = (K, \varphi)$ is unique up to isomorphism. K is called the **Bohr compactification** of G, and φ is called the canonical mapping. In particular, suppose that G is a locally compact Abelian group and G* is its †character group. We denote by G' the group G* with discrete topology. Let K be the character group of G', and let φ^* be the identity mapping $G' \to G^*$ and φ be its conjugate mapping $G \to K$, which is a continuous homomorphism. Then K is the Bohr compactification of G with the canonical mapping φ. A necessary and sufficient condition for f on G to be con-

tinuous almost periodic is that a continuous function f' on K such that $f = f' \circ \varphi$ must exist. If this condition is satisfied, then the mean M [f] is identical to $\int_K f'(x)\, dx$. For any finite-dimensional continuous unitary representation D' of K, $D = D' \circ \varphi$ is a finite-dimensional continuous unitary representation of G, and vice versa. Hence there exists a canonical isomorphism (determined by $D = D' \circ \varphi$) between the equivalence classes of finite-dimensional unitary representations of a separated topological group G and the equivalence classes of finite-dimensional unitary representations of its Bohr compactification K. The †kernel of the canonical mapping $\varphi: G \rightarrow K$ is identical to the intersection of all the kernels of finite-dimensional continuous unitary representations of G.

1. Maximally Almost Periodic Croups

Let G be a topological group. If for each pair a, b of distinct elements of G there exists a continuous almost periodic function f on G such that $f(a) \neq f(b)$, then G is called a **maximally almost periodic group.** This is the case if and only if G has sufficiently many finite-dimensional unitary representations. For a connected locally compact group G, the following six conditions are equivalent: (1) G is a maximally almost periodic group. (2) There is a one-to-one continuous homomorphism from G into a compact group. (3) G is the direct product of a compact group and a vector group $\mathbf{R}^n$. (4) G is the †projective limit of †Lie groups that are locally isomorphic to compact groups. (5) The quotient group G/Z is compact, where Z is the center of G. (6) The system of all neighborhoods that are invariant under the †inner automorphisms constitutes a basis for the neighborhood system of the unit [7].

Moreover, any discrete free group is maximally almost periodic. If there is no continuous almost periodic function except constant functions, the topological group is called **minimally almost periodic.** Any noncompact connected †simple Lie group is minimally almost periodic.

References

[1] A. S. Besicovich (Besikovitch), Almost periodic functions, Cambridge Univ. Press, 1932.
[2] S. Bochner, Bertrager zur Theorie der fastperiodischen Funktionen I, II, Math. Ann., 96 (1926), 119–147, 383–409.
[3] H. Bohr, Fastperiodischen Funktionen, Erg. Math., Springer, 1932.
[4] V. Stepanov, über einige Verallgemeinerungen der fastperiodischen Funktionen, Math. Ann., 95 (1926) 473-498.
[5] N. Wiener, On the representation of functions by trigonometrical integrals, Math. Z., 24 (1926), 5755616.
[6] J. von Neumann, Almost periodic functions in a group I, Trans. Amer. Math. Soc., 36 (1934), 4455492.
[7] J. Dixmier, Les C*-algèbres et leurs représentations, Gauthier-Villars, 1964.
[8] W. Maak, Fastperiodische Funktionen, Springer, 1950.

19 (XV.1 0) Analog Computation

A. History

The term **analog computation** is a generic term describing various techniques of computation employing diagrams, or physical systems whose equations are similar to the mathematical problems in question. The history of analog computation is probably as old as that of digital computation; for example, the ancient Greeks tried to solve cubic equations using diagrams, and the astrolabe widely used by astronomers through the medieval period is also a kind of analog computer. Soon after the discovery of logarithms, the **slide rule** was invented. In the 18th Century, the **planimeter,** used to measure plane areas, was introduced, and in the 19th Century, the **nomogram** appeared (- Section D). In the first half of 20th Century, a large electronic analog computer was developed, thus predating the first practical digital computer.

However, analog computation has an essential defect, namely, the limitation of accuracy. Today it is useful for simple calculation, but is rapidly becoming less important as the development of digital computers, including pocket calculators, advances.

B. Graphical Calculation

Graphical calculation is a method of computation by means of geometric constructions using common drawing tools. Some typical examples of practical graphical calculation are the following: evaluation of linear functions of several variables (J. Massau, 1887), of systems of linear equations (F. J. van den Berg, 1888), of polynomials (J. A. Segner, 1761), of algebraic equations (Lill, 1867), **graphical integration, graphical differentiation,** and the solu-

tion of ordinary differential equations of the first order (J. Massau, 1878), of linear differential equations (Czuber), and of ordinary differential equations of the second order (Lord Kelvin, 1892).

Using bisquare-root graph paper, sold under the names of **binomial probability paper** or **stochastic paper;** we can handle †*F*-distributions or other probability distributions reducible to F-distributions, such as binomial or normal distributions, to a fairly good degree of approximation. This **grapbical method of statistical inference** is, even now, a powerful method of statistical quality control.

C. Graphical Mechanics

Graphical mechanics is the graphical treatment of mechanical problems, especially problems of equilibrium. In this method, the fundamental constructions are the composition and resolution of forces by means of force polygons (Fig. 1). This method is also applicable to problems in dynamics when they can be reduced to problems of equilibrium by means of d'Alembert's principle.

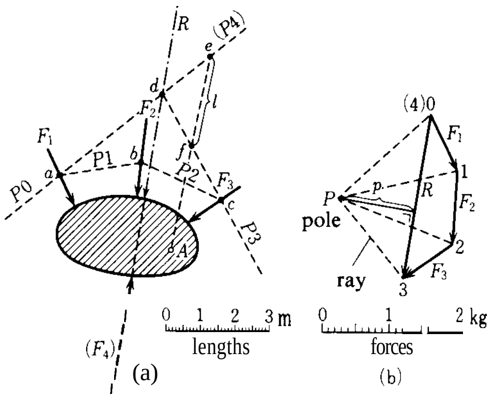


Fig. 1
An example of the graphical method of constructing the composite and the line of action of three given forces F_1, F_2, and F_3. In the force polygon in (b), the vector $\overline{03}$ represents the composite *R*. On the link polygon in (a), the point *d* is a point on the line of action of the force. If the fourth force F_4 has the same magnitude as *R* with opposite direction, both polygons are closed, and the four forces F_1, F_2, F_3, and F_4 are in equilibrium.

Graphical mechanics is most convenient when applied to problems reducible to 2-dimensional structural mechanics. But we can also use it for 3-dimensional problems if we work on projections, e.g., on the plan and the elevation, of the original body. In recent years, however, the amount of work involved in geometric construction has been deemed non-negligible, and the technique of computer graphics is now considered to be much more convenient. Here the computation is done by a digital computer, and the result is displayed in a convenient graphical form.

D. Nomograms

Nomograms are charts in which we can easily read off the corresponding value u_n from given values $u_1, , u_{n-1}$ when there is a relation $F(u_1, \ldots, u_n) = 0$ among *n* real variables u_1, ... u_n. The construction of nomograms has been thoroughly investigated by M. d'Ocagne [5].

A function of two variables $F(u_1, u_2) = 0$ can be represented by a **two-sided scale** or by **functional paper.** The most useful nomograms are those for functions of three variables, $F(u_1, u_2, u_3) = 0$. Changing this into $f_i(x, y, u_i) = 0$ $(i = 1, 2, 3)$ and drawing the curves $f_i(x, y, u_i) = 0$ for fixed values of u_i, **we** have an **intersection chart**, where three curves for the corresponding values u_1, u_2, u_3 meet at a point. In practice, however, it is much more convenient to use the **alignment chart,** which is a dual of the intersection chart. It is especially useful when the relation $F(u_1, u_2, u_3) = 0$ can be decomposed into an equation of the form

$$\begin{vmatrix} f_1(u_1) & g_1(u_1) & h_1(u_1) \\ f_2(u_2) & g_2(u_2) & h_2(u_2) \\ f_3(u_3) & g_3(u_3) & h_3(u_3) \end{vmatrix} = 0. \tag{1}$$

Putting $x_i = f_i/h_i$, $y_i = g_i/h_i$ $(i = 1, 2, 3)$ and drawing the curves (x_i, y_i) with the parameter u_i scaled on it (**u_i-scaled**), the corresponding values u_1, u_2, and u_3 satisfying relation (1) lie on a straight line (Fig. 2). Using this property we can easily read off the corresponding values (by laying down a ruler). In the strict sense, the term "nomogram" usually refers specifically to an alignment chart.

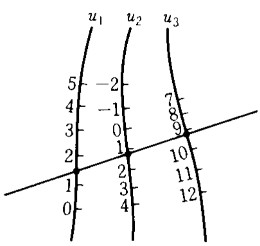


Fig. 2

For four or more variables, we can apply similar techniques if the function is **separable** into functions of three variables. If it is not separable, we can sometimes apply such techniques such as two-functional scales with functional networks, cocircular charts, coplanar charts, or moving charts. Using func-

tional approximations, an approximate nomogram has recently been constructed with errors considered to be admissible in practical use.

Historically, the origin of Hilbert's 13th problem, which asks if it is possible to represent a function of many variables as the composite of functions of two variables, cornes from a study of nomograms with many variables.

E. Analog Computers

The **analog computer** is a special purpose machine designed for a specified analog computation. In a wider sense it includes special devices that perform †analog simulation. The following mechanical devices are well known: the **pantograph** for copying plans, the **harmonic analyzer** for obtaining the Fourier expansion of a periodic function, and V. Bush's **differential analyzer.** Since about 1940 large **electronic analog computers** for solving differential equations have been extensively developed.

Analog computers have up to recently had the advantages of ease of construction, simplicity, and inexpensive operation, and they were also considered to be fast enough for use in **real-time** computation. To compensate for the limitations of analog computers in comparison to digital ones, several **hyhrid computers** have also been used. However, with the rapid progress of digital computers and digital-analog converters, the analog computers are now considered less important than in the past.

F. Curve Fitting

Curve fitting is a method of finding a simple curve $y = f(x)$ supplying the best possible approximation to the values y,, y,, for discrete values x_1, x_2, of the independent variable. A polynomial passing through all the given points is constructed by means of tinterpolation. For experimental data we usually construct a curve by using methods, such as the †method of least squares, that take the errors due to observation into account.

A function constructed to best fit the observed values of the function $y = f(x)$ expressing some physical law is called an **empirical formula,** in contrast to the **theoretical formula.** Usually, we first assume that the function contains several **empirical constants,** which are then determined to fit the experimental data.

The most common case is a linear approximation (using a linear function) with a suitable change of variables. **Semilogarithmic paper, logarithmic paper,** or **probability paper** are used to facilitate transformations in terms of logarithms or normal distributions.

We also frequently use polynomials or trigonometric polynomials of lower degree. To determine the empirical constants the following methods are used: the graphing method, the selection of certain points, the mean value method, or the method of least squares.

G. Orthogonal Polynomials

Let there be given the values $y_0, y_1, , y_{n-1}$ for discrete values x = 0, 1, . , $n-1$ with equal differences. A polynomial f(x) of degree k $(\leqslant n)$ that minimizes the sum of the squares G,(n) = $\Sigma(y_m - f(m))^2$ is given by

$$f(x) = \sum_{v=0}^{k} a_v q_v(n,x),$$

$$a_v = \sum_{m=0}^{n-1} y_m q_v(n,m)/S_v(n),$$

$$S_v(n) = \sum_{m=0}^{n-1} (q_v(n,m))^2,$$

and we have $G_k(n) = \sum_{m=0}^{n-1} y_m^2 - \sum_{v=0}^{k} a_v^2 S_v(n)$. Here the function $q_v(n$, x), called a **Chebyshev q-function,** is defined by

$$q_v(n,x) = \sum_{m=0}^{v} \binom{v+m}{m}\binom{v-n}{v-m}\binom{x}{m}.$$

In practical applications, it is better to replace $q_v(n$, x) by the function

$$q_v^*(n,x) = q_v(n,x)/2^{-v} v! M_v(n),$$

where M,(n) is the greatest common divisor of

$$\binom{v+m}{m}\binom{n-1-m}{v-m} \quad (m = 0, 1, \ldots, v).$$

The values $q_v^*(n,m)$ $(m = 0, 1, \ldots, n-1)$ are mutually coprime integers. The function $q_v^*(n, m)$ is called the **simplest Chebyshev q-function** or **simplest orthogonal polynomial,** and is sometimes denoted by X,,(x), $\xi'_{n,m}(x)$, or $\varphi_{n,m}(x)$.

When two functions satisfy the condition

$$(f,g)_n = \sum_{m=0}^{n} f(m)g(m) = 0,$$

they are called **orthogonal for a finite sum.** If from $1, x, x^2, \ldots, x^v$ we construct a system orthogonal with respect to this definition, then we have the polynomial

$$P_{v,n}(x) = \sum_{m=0}^{v} (-1)^m \binom{v}{m}\binom{v+m}{m}\binom{x}{m}\bigg/\binom{n}{m},$$

which has the following connection with q_v:

$$q_v(n,x)$$
$$= ((-1)^v (n-1)!/2^v (n-v-1)!) P_{v,n-1}(x).$$

The polynomial $P_{v,n}(x)$ is called a **Chebyshev orthogonal polynomial** or sometimes simply an **orthogonal polynomial.** The polynomials $P_{v,n}(x)$ are orthogonal for the finite sum. If $n\to\infty$ and $0\leqslant x\leqslant 1$, the function $P_{v,n}(x)$ tends to $P_v(1-2x)$, where P_v is the †Legendre polynomial. These functions may be conveniently used in least-squares curve fitting.

References

[1] G. W. Smith and R. C. Wood, Principles of analog computation, McGraw-Hill, 1959.
For graphical calculation,
[2] C. Runge and F. A. Willers, Numberische und graphische Quadratur und Integration gewohnlicher und partieller Differentialgleichungen, Enzykl. der Math., Bd. II, Teil 3, Heft 2 (1915), 477176.
For graphical mechanics,
[3] C. B. Biezeno and R. Grammel, Technische Dynamik, Springer, 1939.
[4] H. F. B. Müller-Breslau, Die graphische Statik der Baukonstruktionen, Kröner, I, sixth edition, 1927; II, pt. 1, fifth edition, 1922; II, pt. 2, second edition, 1925.
For nomograms,
[5] M. d'Ocagne, Traité de nomographie, Gauthier-Villars, second edition, 1921.
[6] R. Soreau, Nomographie, E. Chiron, second edition, 1921.
[7] H. Schwerdt, Lehrbuch der Nomographie, Springer, 1924.
[S] D. P. Adams, An index of nomograms, MIT Press, 1950.
For analog computers,
[9] G. A. Korn and T. M. Korn, Electronic analog computers, McGraw-Hill, 1952; second edition, 1956.
For curve tïtting and empirical formulas,
[10] T. R. Running, Empirical formulas, Wiley, 1917.
[11] P. G. Guest, Numerical methods of curve fitting, Cambridge Univ. Press, 1961.
For tables of Chebyshev q functions,
[12] R. A. Fisher and F. Yates, Statistical tables for biological, agricultural, and medical research, Oliver & Boyd, 1938 [$n\leqslant 75, v\leqslant 5$].
[13] R. L. Anderson and E. E. Houseman, Tables of orthogonal polynomial values extended to n = 104, Agricultural Exp. Station, Iowa State College Stat. Sec. Res. Bull., 297 (1942), 595-672 [$n\leqslant 104, v\leqslant 5$].
[14] J. Yamanouchi, On the Chebyshev q functions. I. Simplest Chebyshev q functions and their tables (in Japanese), Report of the Electrotechnical Laboratory, 454 (1942) [$n\leqslant 15, v\leqslant n-1; 16\leqslant n\leqslant 21, v\leqslant 9; 22\leqslant n\leqslant 31, v\leqslant 7; 32\leqslant n\leqslant 51, v\leqslant 5$].

20 (X.1) Analysis

The origin of analysis can be traced back to the time when Eudoxus (4th century B.C.) and Archimedes (3rd century B.C.) devised the so-called method of exhaustion for calculating the area of a plane figure and the volume of a solid. Their objects of investigation were restricted, however, to particular types of figures or solids. In the 16th and 17th centuries, F. Viète, J. Kepler, and B. Cavalieri again took up this problem. In the 17th Century, the problem of drawing a tangent to a given curve was studied by R. Descartes, P. de Fermat, B. Pascal, and J. Wallis. Fermat, in particular, applied the result to find the maxima and minima of certain functions. It is worth noting that a certain type of mathematics powerful enough to produce similar results was independently developed in Japan around that time. In 1684, G. Leibniz in Germany introduced the symbols ***dx*** and ***dy*** in treating the same problem. He proved that dy/dx represents the slope of the tangent to the curve at a given point and discovered a new operation to calculate it. In 1686, he established the "inverse tangent method," which is what we now call integral calculus. He also introduced the notation $\int$. On the other hand, I. Newton in England developed his "method of fluxions," corresponding to our differential and integral calculus, from the viewpoint of mechanics. But neither Leibniz nor Newton formulated the fundamental concepts rigorously, and therefore they were criticized severely by many contemporary scholars. The new calculus gained ground in Great Britain rather slowly. B. Taylor in England and C. MacLaurin in Scotland demonstrated its usefulness in 1715 and in 1745, respectively. On the continent, however, Leibniz's symbolic calculus was taken up by mathematicians of the Bernoulli family, G. F. A. de l'Hôpital, G. Fagnano, and many others; with it, they solved many scientific problems which until then had remained intractable. This motivated subsequent researchers to pose new problems in the form of differential equations.

One of these problems, treated by J. le Rond d'Alembert in relation to the vibration of a chord, concerns the †partial differential equation

$$\partial^2 y/\partial t^2 = a^2 \partial^2 y/\partial x^2 \qquad (1)$$

for $y = y(t, x)$ with the boundary conditions $y = 0$ for $x = 0$ and $x = l$. He obtained the solution $y=f(at+x)-f(at-x)$, where f is an arbitrary function of period 21. In 1753, D.

Bernoulli showed that solutions of equation (1) are given by functions of the form

$$y=\frac{a_0}{2}+\sum_{k=1}^{\infty}\left(a_k\cos\frac{k\pi x}{l}+b_k\sin\frac{k\pi x}{l}\right)$$

These two kinds of solutions gave rise to the question of whether an arbitrary function can be expressed by a †trigonometric series. This problem was studied by A. C. Clairaut, J. L. Lagrange, and L. Euler. In 1807, J. Fourier in France, in treating a problem on the conduction of heat, claimed that an arbitrary function of period 2π can be expressed as

$$y=\frac{a_0}{2}+\sum_{k=1}^{\infty}(a_k\cos kx+b_k\sin kx), \tag{2}$$

where the coefficients a_k and b_k are given by

$$a_k=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\cos kx\,dx,$$
$$b_k=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\sin kx\,dx. \tag{3}$$

This series is now called the †Fourier series, but Fourier never verified the fact that the series (2) with coefficients (3) converges and represents f(x). It was only as late as 1820 that A. L. Cauchy in France first noted that to treat a series properly, one must examine its convergence.

In the 19th Century, the concept of †functions, which had been taken in the sense of "analytic expressions," came to be defined by the correspondence relation. Cauchy clarified the ideas of †limit and †continuity, †differentiability and †integrability. He showed that a function that is continuous in a bounded closed interval is integrable in that interval. But his proof was not rigorous, as he lacked the notion of †uniform continuity. In his 1854 paper on the trigonometric series, B. Riemann in Germany considered the integrability of functions that might be discontinuous and introduced the concept of what we now call the †Riemann integral.

The theory of †sets, initiated by G. Cantor in Germany in his paper of 1874, revolutionized analysis. R. Baire, E. Borel, and H. Lebesgue in France contributed to the establishment of analysis on the basis of set theory. Baire made a classification of discontinuous functions. Generalizing his results, Lebesgue gave a definition of analytic expressions, thus clarifying the term that had been used vaguely since the time of Euler. Lebesgue also tried to define the concepts of the integral of a function, the length of a curve, and the area of a surface from the most general viewpoint. In generalizing the notion of †measure introduced by Borel, he established in 1902 the theory of †Lebesgue measure with which he laid the foundations of the theory of †Lebesgue integrals. The introduction of this theory gave to the theory of Fourier series a new turn in the direction of functional analysis. Measure theory was also employed by A. N. Kolmogorov of the Soviet Union to lay a solid foundation for †probability theory in 1933.

The study of functions of a complex variable was originated by Cauchy in the first half of the 19th Century. He began his research by introducing the notion of "monogenic functions"; a function which is monogenic at every point of a domain is what we now call a †holomorphic function. He established †Cauchy's integral theorem and integral formulas for these functions, and deduced from these theorems the †residue theorem for functions with †poles. Making use of the integral formula, Cauchy proved that a function that is holomorphic at a point a can be expanded in a power series of the form $\sum_{k=0}^{\infty}a_k(z-a)^k$ in a neighborhood of this point.

Riemann considered a complex variable w as a function of another complex variable z when dw/dz is independent of the value of the differential dz. This amounts to the same thing as a "monogenic function" of Cauchy. Riemann's mapping theorem became a model for subsequent developments. Riemann introduced the concept of †Riemann surfaces in order to †uniformize multivalued functions. This important idea was basic to the progress of analysis and †topology in the 20th Century.

K. Weierstrass, who was a contemporary of Riemann, developed the theory of functions of a complex variable from a purely analytic viewpoint. He defined an element of a function to be a power series $\sum_{k=0}^{\infty}a_k(z\quad a)^k$ of z -a, representing a holomorphic function in the interior of its †circle of convergence, and defined an analytic function to be an aggregate of such elements that are derived from one of them by means of †analytic continuations, along all curves having the point a as the initial point.

Riemann and Weierstrass constructed their theory to complete the theory of †elliptic functions and †Abelian functions that was initiated by N. H. Abel and C. G. J. Jacobi. Their results were a high point of 19th-century classical mathematics. H. Poincaré in France built upon their work. Another high point in the theory of functions of a complex variable was reached when J. Hadamard and C. de La Vallée Poussin in France made use of it to prove the †prime number theorem in 1896.

Weierstrass also initiated the study of functions of several complex variables and was succeeded by Poincaré, P. Cousin, and E. Picard in France. They tried to extend the theory of functions of one complex variable to

that of many variables. F. Hartogs, however, discovered a phenomenon quite different from the case of one variable. Cousin posed in 1895 the problem of constructing a function of several complex variables that has assigned poles. This problem was pursued by A. Weil in France and solved by K. Oka in Japan in 1936. Also, the problem of characterizing †domains of holomorphy was investigated by E. E. Levi in Italy, H. Cartan in France, and P. Thullen and K. Stein in Germany and was finally solved by Oka in 1953. H. Cartan and J.-P. Serre reformulated these results in terms of †cohomology with coefficients in †sheaves. This considerably influenced the formulations of mathematics thereafter. These results were further generalized to †analytic space by H. Grauert, R. Remmert, and Stein in Germany. The theory of †complex manifolds, which are generalizations of Riemann surfaces to several variables, was initiated by W. V. D. Hodge in England and K. Kodaira in Japan and continued by F. Hirzebruch in Germany, M. F. Atiyah in England, and I. M. Singer in the United States.

Differential calculus gives a general method of finding extreme values of a given function. Likewise, in order to find a function that produces an extremal of the given †functional, the calculus of variations was created in the 18th Century. For example, Euler considered the problem of finding a particular function y(x) that renders the functional $\int_a^b F(x, y, y') dx$ an extreme value among all those functions y(X) for which the plane curve y = y(x) passes through two given points (a, A) and *(b, B)* of the plane; he showed that this y(x) must satisfy the differential equation $dF_{y'}/dx - F_y = 0$ (1744). Lagrange, W. R. Hamilton, and others developed this result into a general †variational principle that governs not only classical †mechanics but also †quantum mechanics.

From research on the continuity or differentiability of the functional with respect to y emerged the idea of considering a function as a "point" in a †function space. This gave rise to †functional analysis, a branch of analysis that treats functions as elements of certain spaces and utilizes the methods of algebra and topology. The first result in this regard was the theory of †integral equations of V. Volterra in Italy and E. I. Fredholm in Sweden in the beginning of the 20th Century. Fredholm's work was motivated by his desire to solve the †Dirichlet problem, the solution of which had been used by Riemann in the proof of his mapping theorem, etc. However, Riemann's own proof of existence of the solution, called the †Dirichlet principle, was not rigorous, and attempts to save the proof provided one of the central problems of analysis for some time. Fredholm's solutions were different from Riemann% but D. Hilbert in Germany was able to justify Riemann's original proof. Hilbert's proof was later simplified and generalized by R. Courant, H. Weyl, and others. Hilbert also introduced the †function spaces l_2 and L_2 to study the †eigenvalue problem of integral equations with †symmetric kernels. Later J. von Neumann in Germany established †spectral theory in abstract †Hilbert spaces and applied it to the mathematical foundations of quantum mechanics (1929). S. Banach in Poland created the theory of †linear operators in Banach spaces (1932). This theory was further generalized to that of †linear topological spaces and was applied to the theory of distributions.

In their study of partial differential equations and Fourier analysis, Hadamard, J. Leray, S. L. Sobolev, T. Carleman, and many others had to extend the notion of functions; they also enlarged the notion of derivatives. L. Schwartz in France introduced †distributions and defined derivatives in the sense of distributions to unify these generalizations (1945). M. Sato in Japan defined more general generalized functions, called †hyperfunctions (1958). It has become evident that both distributions and hyperfunctions have provided the most powerful tools in recent research in the general theory of †partial differential equations, to which L. Hormander in Sweden has made outstanding contributions.

References

[1] M. B. Cantor, Vorlesungen über Geschichte der Mathematik, Teubner, 189441908.
Also → references to 265 Mathematics in the 17th Century, 266 Mathematics in the 18th Century, and 267 Mathematics in the 19th Century.

[2] N. Bourbaki, Eléments de mathématique, Fonctions d'une variable réele, Integration, Espace vectoriels topologiques, Théorie spectrales, Actualités Sci. Ind., Hermann, 1955–*1969.*

[3] C. Carathéodory, Vorlesungen über reelle Funktionen, second edition, Teubner, 1927 (Chelsea, 1968).

[4] H. Cartan, 1. Calcul différentiel, 2. Formes différentielles, Hermann, 1967.

[5] G. Choquet, Lectures on analysis I, II, III, Benjamin, 1969.

[6] R. Courant and D. Hilbert, Methods of mathematical physics I, II, Interscience, 1953, 1962.

[7] R. Courant, Differential- und Integralrechnung I, II, Springer, third edition, 1955; English translation, Differential and integral calculus, Nordemann, 1938.

[8] C. J. de La Vallée-Poussin, Cours d'analyse infinitésimale I, II, Gauthier-Villars, seventh edition, 1930.
[9] J. Dieudonné, Treatise on analysis I–VII, Academic Press, 1960–1978.
[10] J. Dieudonné, Eléments d'analyse I-VIII, Hermann, 1971-1978.
[11] E. Goursat, Cours d'analyse mathematique I, II, III, Gauthier-Villars, fifth edition, 1933-1956.
[12] J. Hadamard, Cours d'analyse I, II, Hermann, 1927, 1930.
[13] G. H. Hardy, A course of pure mathematics, Cambridge Univ. Press, seventh edition, 1938.
[14] D. Hilbert, Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen, Teubner, second edition, 1924.
[15] C. Jordan, Cours d'analyse de l'école polytechnique: I. Calcul différentiel; II. Calcul intégral; III. Calcul intégral (équations différentielles), Gauthier-Villars, third edition, I, 1909; II, 1913; III, 1915.
[16] H. L. Lebesgue, Leçons sur l'intégration et la recherche des fonctions primitives, Gauthier-Villars, second edition, 1928.
[17] É. Picard, Traité d'analyse, Gauthier-Villars, I, 1922; II, 1926; III, 1928.
[18] G. Pólya and G. Szegö, Aufgaben und Lehrsätze aus der Analysis I, 11, Springer, second edition, 1954.
[19] W. Rudin, Real and complex analysis, McGraw-Hill, 1966.
[20] L. Schwartz, Cours d'analyse mathématique I, II, Hermann, 1967.
[20A] L. Schwartz, Analyse-Topologie générale et analyse fonctionelle, Hermann, 1971.
[21] L. Schwartz, Théorie des distributions, Actualites Sci. Ind., Hermann, 1950, second edition, 1966.
[22] V. I. Smirnov, A course of higher mathematics: I. Elementary calculus; II. Advanced calculus; III-1. Linear algebra; III-2. Complex variables, special functions; IV. Integral equations and partial differential equations; V. Integration and functional analysis, Addison-Wesley, 1964. (Original in Russian, 1961.)
Also → references to 106 Differential Calculus, 107 Differential Equations, 162 Functional Analysis, and 198 Holomorphic Functions.

21 (XI.20) Analytic Functions of Several Complex Variables

A. Holomorphic Functions

As in the case of †holomorphic functions of one complex variable, the definition of holomorphic functions can be given in two ways: the first definition utilizes differentiability, following the approach of B. Riemann; and the second method utilizes the notion of power series expansion as developed by K. Weierstrass. In this article, N = {0, 1, 2, . . .}.

B. Power Series

Let z be an n-tuple of complex variables $z_1, \ldots, z_n$, and $c=(c_1, \ldots, c_n)$ a point of $\mathbf{C}^n$. An infinite sum P of monomials $a_k(z-c)^k = a_{k_1,\ldots,k_n}(z_1-c_1)^{k_1}\ldots(z_n-c_n)^{k_n}$ $(k=(k_1,\ldots,k_n)\in \mathbf{N}^n)$, where $a_k \in$ C, is called a **power series** with center c and coefficients a_k. If, for a bijection φ of N onto $\mathbf{N}^n$, the simple series $\sum_{p\in\mathbf{N}}|a_{\varphi(p)}(z-c)^{\varphi(p)}|$ is convergent at z = z^0, we say that P is **absolutely convergent** at z^0. Its sum at z^0, denoted by $\sum a_k(z^0-c)^k$, is defined as the sum $\sum a_{\varphi(p)}(z^0-c)^{\varphi(p)}$, which is independent of the choice of φ. If P is uniformly bounded at z^0, then P is absolutely convergent at every point of the open **polydisk** $S=\{z \mid |z_j-c_j|<|z_j^0-c_j|, j=1,\ldots,n\}$. Furthermore, in this case P converges absolutely and uniformly on every compact set in S (N. H. Abel).

The **convergence domain** of a power series P is the set D of points z^0 such that P is absolutely convergent at every point in a neighborhood of z^0. The interior of the set B of points at which the infinite sum P is uniformly bounded is equal to D. A **(complete) Reinhardt domain** with center c is a domain D in $\mathbf{C}^n$ such that whenever D contains z_0, the domain D also contains the torus $\{z \mid |z_j-c_j|=|z_j^0-c_j|, j=1,\ldots,n\}$ (the closed polydisk $\{z \mid |z_j-c_j|\leqslant|z_j^0-c_j|, j=1,\ldots,n\}$). If the convergence domain D of the power series P is not empty, it is a complete Reinhardt domain and is also **logarithmically convex**; that is, the set $D-\bigcup_j\{z \mid z_j=c_j\}$ is mapped onto a convex domain in $\mathbf{R}^n$ by the mapping $z_j\to\log|z_j-c_j|$ (j= 1, . . . , n). The set $\tilde{D}$ of points at which P is absolutely convergent is, in general, greater than D, and it is possible that $\tilde{D}$ contains exterior points of D. A **thorn** of D is the set of exterior points of D contained in $\tilde{D}$ that are located on the planes $\{z \mid z_j=c_j\}$ (j = 1, . . , n). An n-tuple $r\in\mathbf{R}_+^n$ is called a set of **associated convergence radii** if P is absolutely convergent at every point of $\{z \mid |z_j-c_j|<r_j, j=1,\ldots,n\}$ but not of $\{z \mid |z_j-c_j|>r_j, j=1,\ldots,n\}$. An n-tuple of associated convergence radii may not be uniquely determined, but it satisfies

$$\limsup_{|k|\to+\infty}(|a_k|r^k)^{1/|k|}=1, \qquad |k|=k_1+\ldots+k_n$$

(E. Lemaire).

Let f be a complex-valued function defined in a neighborhood of $z^0\in\mathbf{C}^n$. If there exists a

convergent power series P with center z^0 such that at every point of a neighborhood of z^0 the value off' and the sum of P coincide, then f is called **analytic** at z^0 in the sense of Weierstrass, and P is the **Taylor expansion** of f at z^0.

C. Differentiability

Let f be a complex-valued function detïned in a neighborhood of $z^0 \in$ C". If in a neighborhood of z^0 we have

$$f(z)-f(z^0)=\alpha_1(z_1-z_1^0)+\ldots+\alpha_n(z_n-z_n^0)+\varepsilon,$$

with $\alpha_1,\ldots,\alpha_n \in \mathbf{C}$ and

$$\lim_{z\to z^0} \varepsilon/(|z_1-z_1^0|+\ldots+|z_n-z_n^0|)=0, \qquad (1)$$

then we say that f is **(totally) differentiable** at z^0. The function f is then continuous at z^0, and the partial derivatives $\partial f/\partial z_j$ (j = 1, , n) exist. Furthermore, the Cauchy-Riemann differential equations $\partial f/\partial \bar{z}_j=0$ (j= 1, , n) hold, where $\partial f/\partial z_j=(1/2)(\partial f/\partial x_j - i\partial f/\partial y_j)$ and $\partial f/\partial \bar{z}_j=(1/2)(\partial f/\partial x_j + i\partial f/\partial y_j)$ with $z_j = x_j + iy_j$ We say that f is **holomorphic** at z^0 in the sense of Riemann if f is differentiable at every point in a neighborhood of z^0. Analyticity in the sense of Weierstrass is equivalent to holomorphy in the sense of Riemann. Furthermore, if the partial derivatives $\partial f/\partial z_j$ $(j = 1, \ldots, n)$ exist at every point in a neighborhood of z^0, then f is, without assuming continuity, proved to be holomorphic. Thus the holomorphy off in each variable z_j implies the holomorphy off in $z=(z_1, \ldots) z_,)$ **(Hartogs's theorem of holomorphy,** 1906).

A complex-valued function in a domain G $\subset$ $\mathbf{C}^n$ is called **holomorphic** or **analytic** in G if it is holomorphic at every point of G. Let H(G) be the ring of holomorphic functions in G. For f = u $+ iv \in$ N(G), u and v satisfy in G the differential equations $\partial^2 u(z,\bar{z})/\partial z_j \partial \bar{z}_k$ = 0; that is,

$$\frac{\partial^2 u}{\partial x_j \partial x_k}+\frac{\partial^2 u}{\partial y_j \partial y_k}=0, \qquad (2)$$

$$\frac{\partial^2 u}{\partial x_j \partial y_k}-\frac{\partial^2 u}{\partial x_k \partial y_j}=0, \qquad j,k=1,\ldots,n.$$

A †distribution TED'(G) is called **pluriharmonic** in G if it satisfies (2) in G. Then T is harmonic and hence a real analytic function by †Weyl's lemma.

Let G_j be a domain in the z_j-plane with piecewise smooth boundary C_j. If $f \in$ H(G) (G = $\prod_{j=1}^n G_j$) is continuous on $\bar{G}$, then

$$\frac{1}{(2\pi i)^n}\int_{C_1\times\ldots\times C_n}\frac{f(\zeta)}{(\zeta_1-z_1)\ldots(\zeta_n-z_n)} d\zeta_1\wedge\ldots\wedge d\zeta_n=\begin{cases} f(z), z\in G,\\ 0, z\notin \bar{G}\end{cases} \qquad (3)$$

(Cauchy's integral representation). Thus if $n \geqslant 2$, then f is determined by its values only on the proper subset C = Ci x x C_n of ∂G, which is called the **skeleton** (or **determining set)** of G. For (pluriharmonic) functions of several complex variables, the boundary value problem, not necessarily solvable in its classical form, is not so effective as the Dirichlet problem in one complex variable.

As in the case of one variable, the †Laurent expansion is valid for every holomorphic function in a domain of the form G = $\prod_{j=1}^n G_j$, where the G_j are circular annuli $\subset$ C. Suppose that we are given $f_1 \in H(G_1)$ and $f_2 \in H(G_2)$, where G_1, G_2 are domains in C" such that $G_1 \cap G_2$ is nonempty and connected. If $f_1 = f_2$ on $\{z \,|\, |z_j - z_j^0| < r_j, y = y^0, j=1,\ldots,n\}$, where $z^0 = x^0 + iy^0 \in G_1 \cap G_2$, then there exists a unique $f \in H(G_1 \cup G_2)$ such that $f|G_1 = f_1$ and $f|G_2 = f_2$ **(tbeorem of identity).** Thus †analytic continuation proceeds as in the case of one variable. Similarly, some fundamental theorems in one variable, such as †Liouville's theorem on †entire functions and the †maximum principle, hold also for several variables. However, there are some properties that reveal the differences between the cases of one and several variables. For instance, the set of zeros of a holomorphic function (→ 23 Analytic Spaces B) has no isolated point for $n \geqslant 2$. The investigation of these remarkable differences is one of the purposes of the theory of analytic functions of several complex variables.

D. Shilov Boundaries

While the maximum principle holds for a holomorphic function in a domain G, the set of points where the maximum is attained may be a proper subset S of ôG. For instance, if G is the product of annuli as before, then the skeleton of G can be taken as S. In connection with the theory of †normed rings, G. E. Shilov proved that there exists a unique smallest member S_0 (called the **Shilov boundary** of G) in the family of closed subsets S such that $\sup\{|f(z)| \,|\, z\in S\} = \sup\{|f(z)| \,|\, z \in G\}$ for every $f \in$ H(G) continuous on G. The structure of S_0 is investigated in detail together with the pseudoconvexity of G connected with it. Applying †Perron's method for the Dirichlet problem to †plurisubharmonic functions and the Shilov boundary, H. J. Bremermann solved one type of boundary problem.

E. Local Theory

Let f and g be two functions defined in neighborhoods of a set Sc $\mathbf{C}^n$. If f = g in a neighborhood of S, then f and g are called **equiva-**

lent with respect to S. The germ off on S, denoted by f_S, is the equivalence class off: A **germ of a bolomorpbic function** on S is the germ on S of a holomorphic function defined in a neighborhood of S, and H(S) denotes the ring of germs of holomorphic functions on S. Given a point 0 in C”, H(0) = $H(\{0\})$ is isomorphic to the ring H_n of convergent power series at 0, i.e., the power series that are absolutely convergent in some neighborhoods of 0. For every nonzero function $f \in H(O)$, there exists a system of coordinates $(z_1,, z_n)$ centered at 0 such that $f(0, , 0, z_n) \neq 0$ for every $z_n \neq 0$ in a neighborhood of $z_n = 0$. In a neighborhood of 0, f is then equal to the product of an invertible element of H_n and a **distinguished pseudopolynomial**

$$P(z_n)=z_n^p+a_1(z_1, \ldots z_{n-1})z_n^{p-1}+$$
$$+a_p(z_1, \ldots, z_{n-1}) \in H_{n-1}[z_n],$$

with $a_1(0,\ldots,0)=\ldots=a_p(0,\ldots,0)=0$, and $P(z_n)$ is uniquely determined by f and the coordinates $z_1,, z_n$ **(Weierstrass's preparation theorem).** It follows from this that H_n is an n-dimensional †regular local ring. Considering $H(0)$ as the †inductive limit of †locally convex rings $H(U)$, where U ranges over a base for a neighborhood system of 0, H. Cartan proved the preparation theorem in a more precise form in which the association $f \rightarrow a_j$ is continuous with respect to the supremum norm. Based on a deep consideration of this situation, K. Oka proved a theorem of fundamental importance: The †sheaf $\mathcal{O}_{\mathbf{C}^n}$ defined by $\mathcal{O}_{\mathbf{C}^n,z} = H(z)$ $(z \in \mathbf{C}^n)$ is †coherent.

F. Domains of Holomorphy

Given a domain $G \subset \mathbf{G}^n$ for $n \geqslant 2$, it may be that there exists a domain G’ strictly greater than G such that all the functions that are holomorphic in G extend to holomorphic functions in G’. For instance, let S = S’ x σ, where S’ and σ are open polydisks in $(z_1, . ,$ $z_{n-1})$-space and z_n-space, respectively, and let $T \subset$ C” be an open set. If there exists an open set $U\ (\neq \varnothing) \subset S'$ such that $(U \times \sigma) \cup (S' \times \partial\sigma) \subset T$ and if $S \cap T$ is connected, then all the functions that are holomorphic in T extend uniquely to holomorphic functions in SU T **(Hartogs's continuation theorem).** In particular, if A is an †analytic set in a domain $G \subset \mathbf{C}^n$ with dim $A \leqslant n$ 2, then all the functions that are holomorphic in G-A extend uniquely to holomorphic functions in G. Furthermore, if A is an analytic set in G with $A \neq G$, then every $f \in H(G-A)$ that is locally bounded at the points of A extends uniquely to a holomorphic function in **G (Riemann% continuation theorem** for $n \geqslant 2$). The domain $\tilde{G}_f$ of holomorphy for f is defined to be the maximal domain to which f may be continued analytically. A domain G is called a **domain of holomorphy** if $G = \tilde{G}_f$ for some $f \in H(G)$. However, $\tilde{G}_f$ is, in general, not a subdomain of $\mathbf{C}^n$. $\tilde{G}_f$ is, generally, a manifold spread over C”; i.e., $\tilde{G}_f$ is a connected n-dimensional †complex analytic manifold with a holomorphic mapping $\varphi: \tilde{G}_f \rightarrow \mathbf{C}^n$ of maximum Jacobian rank (φ is then an open mapping). The same is true for the common existence domain of functions in a subfamily of $H(G)$. The common existence domain $\tilde{G}$ of all the functions in $H(G)$ is called the **envelope of holomorphy.** A **holomorphically complete domain** is a domain G such that $G = \tilde{G}$. These notions carry over to the case where G is a manifold spread over C”. The (general) Levi problem of determining the conditions for a given domain to be holomorphically complete is fundamental to the theory of analytic functions of several complex variables (→ Section I). In connection with this problem, various notions of pseudoconvexity of holomorphically complete domains are defined.

G. Pseudoconvexity

An upper semicontinuous real-valued function $u\ (-\infty \leqslant u < +\infty)$ in a domain $G \subset \mathbf{C}^n$ is said to be **plurisubharmonic** if for every $z^0 \in$ G and every $a \in \mathbf{C}^n$ the function $u(z^0$ + ta) of t is †subharmonic (including the constant $-\infty$) in all the connected components of {t $z^0 + ta \in$ G}. A domain G is said to be **pseudoconvex** (or ***d*-pseudoconvex)** if $u = -$ logd, is plurisubharmonic in G, where $d,(z)$ is the distance from z $\in$ G to ∂G with respect to any norm in C”. Every connected component of the interior of the intersection of a family of pseudoconvex domains is pseudoconvex, and the union of an increasing sequence of pseudoconvex domains is pseudoconvex. Suppose that we are given a domain G and a function u of class C^2 in a neighborhood of $\bar{G}$ such that G = $\{z \mid$ u(z) < 0} and, for some $\varepsilon > 0$, $\sum_{jk}(\partial^2 u/\partial z_j \partial \bar{z}_k) a_j \bar{a}_k \geqslant \varepsilon |a|^2$ for every $a \in \mathbf{C}^n$. Then the domain G is said to be **strongly pseudoconvex.** Strong pseudoconvexity implies pseudoconvexity. Every pseudoconvex domain is exhausted by an increasing sequence of strongly pseudoconvex domains. An open set **Pc C”** is called an **analytic polyhedron** if $P = \{z \mid |\chi_\alpha(z)| < 1, \alpha =$ 1, , N}, $\chi_\alpha \in H(\bar{P})$ (a= 1, . . . N). Then every connected component of P is pseudoconvex. A **Weil domain** is a connected and bounded analytic polyhedron P defined by χ_α(α = 1, , N) with $N \geqslant n$, such that for every k $(1 \leqslant k \leqslant$ n) the intersection of the hypersurfaces $|\chi_{\alpha_i}(z)| = 1$ $(1 \leqslant i \leqslant k)$ is of dimension $\leqslant 2n-k$.

H. Holomorphic Convexity

A domain $G \subset \mathbf{C}^n$ is called **holomorphically convex** if for every compact set $K \subset G$, $\hat{K} = \bigcap_{f \in H(G)} \{z \mid |f(z)| \leqslant \sup_{\omega \in K} |f(\omega)|\}$ (the **holomorphic hull** of K) is a compact set contained in G. (For a domain G contained in an †analytic set we can similarly define holomorphic convexity of G.) Every connected component of the intersection of a family of holomorphically convex domains is holomorphically convex, and a holomorphically convex domain is exhausted by an increasing sequence of Weil domains. Holomorphic completeness implies holomorphic convexity. The converse is true for domains in $\mathbf{C}^n$. If G is holomorphically convex, then for every point ζ of ∂G there exists an $f \in H(G)$ such that f is not locally bounded at ζ (H. Cartan and P. Thullen, 1932). Hence a holomorphically convex domain is a domain of holomorphy. Thus a domain in C is holomorphically convex if and only if it is a domain of holomorphy. (The same is true for unramified covering domains over $\mathbf{C}^n$ (Oka, 1953).) The union of an increasing sequence of domains of holomorphy is a domain of holomorphy (H. Behnke and K. Stein [2], 1939). Suppose that we are given a domain G and domains S_α, T_α (a = 1, 2,) such that $S_\alpha \cup$ 7" $\subset$ G and $\sup_{T_\alpha} |f| = \sup_{S_\alpha \cup T_\alpha} |f|$ for every $f \in H(G)$. Suppose also that $S_0 = \lim S_\alpha$ is bounded. We say that the **continuity principle** holds in G if $\bar{T}_0 \subset G$ ($T_0 = \lim T_\alpha$) implies $S_0 \subset G$. The continuity principle holds in a domain of holomorphy **(Hartogs's theorem of continuity).** This implies that if G is a bounded domain $\subset \mathbf{C}^n$ ($n \geqslant 2$) with connected boundary ∂G, then every function holomorphic in a neighborhood of ∂G extends to a holomorphic function in G **(Hartogs-Osgood theorem).** In particular, for $n \geqslant 2$, the set of singular points of a holomorphic function has no isolated point. A domain is pseudoconvex if the continuity principle holds there. Hence a domain of holomorphy is pseudoconvex.

I. The Levi Problem

Let G be a domain in $\mathbf{C}^n$ and $z^0 \in \partial G$. If there exists an open neighborhood U of z^0 such that every connected component of $G \cap U$ is a domain of holomorphy, then G is called **Cartan pseudoconvex** at z^0. On the other hand, if every 1-dimensional analytic set that has z^0 as an ordinary point contains points not belonging to G U $\{z^0\}$ in the neighborhoods of z^0, then G is called **Levi pseudoconvex** at z^0. Furthermore, G is called **locally Cartan (Levi) pseudoconvex** if G is Cartan (Levi) pseudoconvex at every point of ∂G. Every domain of holomorphy is locally Cartan pseudoconvex. If G is pseudoconvex and there exists a neighborhood U of z^0 such that $G \cap U = \{z \mid \varphi(z) < 0\}$, where $\varphi \in C^1(U)$, then G is Levi pseudoconvex at z^0.

The (proper) **Levi problem** of whether every pseudoconvex domain is a domain of holomorphy was proposed by E. E. Levi (1911). After unsuccessful efforts by various mathematicians to solve the problem, it was affirmatively solved by Oka (1942 for $n = 2$ and 1953 for manifolds spread over $\mathbf{C}^n$ for $n \geqslant 2$), H. Bremermann, and F. Norguet. The problem was solved also by H. Grauert [14] (1958) in a more general form (- Section L) using results on linear topological spaces and by L. Hörmander [18] (1965) using methods of the theory of partial differential equations. A fundamental step in Oka's solution is his **gluing theorem:** Let G be a bounded domain $\subset \mathbf{C}^n$. If every connected component of $G_1 = \{z \mid x_1 > a\} \cap G$ and $G_2 = \{z \mid x_1 < b\} \cap G$ $(a < b)$ is a domain of holomorphy, then G is a domain of holomorphy. Indeed, by virtue of the Behnke-Stein theorem and the fact that every pseudoconvex domain is the union of an increasing sequence of bounded locally Cartan pseudoconvex domains, it suffices to solve the Levi problem in the case of a bounded locally Cartan pseudoconvex domain. The Levi problem in this case is solved by the gluing theorem. Various integral representations of holomorphic functions are known besides the Cauchy representation. The Bergmann-Weil integral representation in a Weil domain was used as an important means of solving the Levi problem.

J. Holomorphic Mappings

Holomorphic functions with values in a †quasi-complete †locally convex complex vector space E have also been investigated. The classical theory described above has been generalized, to some extent, to this case. In this way, many applications of the theory have been discovered. An E-valued function in a domain G $\subset$ C" is holomorphic if and only if the mapping uof: $G \rightarrow C$ is holomorphic for every continuous linear form u on E. By this theorem, most problems concerning E-valued holomorphic functions can be reduced to those of ordinary holomorphic functions. Note that the vector space H(G) of ordinary holomorphic functions in G is a †Fréchet space. The spaces $\mathbf{C}^p$ and complex †Banach spaces belong to the above category of E. A. $\mathbf{C}^p$-valued holomorphic function in a domain $G \subset \mathbf{C}^n$ is called a **holomorphic** (or **analytic) mapping** of G into $\mathbf{C}^p$. An isomorphism in the category of do-

mains $G \subset \mathbf{C}^n$ and holomorphic mappings is called an **analytic isomorphism** (or **biholomorphic mapping).** An automorphism in the category is called an **analytic** (or **holomorphic) automorphism.** With every domain $G \subset C$ is associated the sheaf $\mathcal{O}_G$ of germs of holomorphic functions over G. Thus we have the notion of a †ringed space (G, 0,). A complex analytic manifold can be defined as a (Hausdorff) ringed space that is locally isomorphic to some (G, 0,).

A **meromorphic function** in G is a function that is locally the quotient of two holomorphic functions with denominator $\neq 0$. It may be defined more rigorously as a meromorphic mapping of G into **P,(C)** (- 23 Analytic Spaces D).

K. The Cousin Problems

The Cousin problems are those of constructing meromorphic functions with given zeros or poles. In terms of sheaves the problems are stated as follows: Let $\mathcal{K}_G$ be the sheaf of germs of meromorphic functions over a domain $G \subset$ C". The **first Cousin problem** asks whether the mapping $\Gamma(G,\mathcal{K}_G)\rightarrow\Gamma(G,\mathcal{P}_G)$ induced by the exact sequence $0\rightarrow\mathcal{O}_G\rightarrow\mathcal{K}_G\rightarrow\mathcal{P}_G\rightarrow 0$ ($\mathcal{P}_G = \mathcal{K}_G/\mathcal{O}_G$) is surjective, where $\Gamma(G,\mathcal{F})$ is the module of †sections of $\mathcal{F}$ over G (- 383 Sheaves C). Let $\mathcal{K}_G{}^*$ be the sheaf of multiplicative groups of germs of meromorphic functions not identically 0 and $\mathcal{O}_G{}^*$ be the subsheaf of $\mathcal{K}_G{}^*$ formed by germs of nonzero holomorphic functions. The **second Cousin problem** asks whether the mapping $\Gamma(G, \mathcal{K}_G{}^*)\rightarrow P(G,\mathcal{D}_G)$ ($\mathcal{D}_G = \mathcal{K}_G{}^*/\mathcal{O}_G{}^*$) is surjective. P. Cousin (1895) solved the first problem for $G = \mathbf{C}^n$ or $\prod_{j=1}^n G_j$ and the second problem for $G = \mathbf{C}^n$. Oka (1935) proved that the first problem is solvable in every domain of holomorphy. In solving the second problem in a domain of holomorphy, Oka established the notion of †fiber bundles and proved that the problem for any domain is reduced to holomorphic triviality of a holomorphic principal fiber bundle over the domain and that holomorphic triviality is equivalent to topological triviality when the domain is of holomorphy **(Oka's principle**). Using the solutions of the Cousin problems, Oka proved his gluing theorem, described in Section 1.

L. Stein Manifolds

Abstracting certain important properties of a domain of holomorphy, Stein [33] introduced the following category of complex analytic manifolds (X, $\mathcal{O}_X$): (1) X is paracompact (i.e., each connected component of X has a countable open base). (2) Functions in $\Gamma(X,\mathcal{O}_X)$ separate the points of X. (3) For every point $x \in X$ there exists a system of local coordinates around x that is formed by functions in $\Gamma(X,\mathcal{O}_X)$. (4) X is holomorphically convex. (X, 0,) is then called a **Stein manifold.** It was later discovered by Grauert [12] that conditions (2) and (4) imply (1) and (3).

Applying the theory of †cohomology with coefficients in sheaves, H. Cartan and J.-P. Serre obtained for an †analytic coherent sheaf $\mathcal{F}$ on a Stein manifold X, the following **fundamental theorems of Stein manifolds. Theorem A:** $H^0(X,\mathcal{F})$ generates the stalk $\mathcal{F}_X$ (as an $\mathcal{O}_X$-module) at every point **x** of **X. Theorem B:** $H^q(X,\mathcal{F})=0$ for all $q>0$ [7, 30]. Conversely, for a complex analytic manifold X, if for every analytic coherent sheaf $\mathcal{I}$ of ideals defined by a 0-dimensional analytic set in X (i.e., a discrete subset of X), H^1 (X, $\mathcal{I}$) = 0, then X is a Stein manifold. Furthermore, if $\Gamma(X,\mathcal{O}_X) =$ $\Gamma($ Y, $\mathcal{O}_Y)$ for a Stein manifold Y (as in the case where $X \subset \mathbf{C}^n$), then the fundamental theorem A for every coherent sheaf of ideals implies that X is a Stein manifold (I. Wakabayashi).

Due to the fundamental theorems, most results on domains of holomorphy hold unchanged for Stein manifolds. For instance, the first Cousin problem is always solvable. The second Cousin problem is solvable if and only if $H^2(X,$ **Z)=O.** An n-dimensional Stein manifold can be realized as a (ramified) covering domain of holomorphy over C". Furthermore, some theorems on differentiable manifolds have analogs on Stein manifolds. For instance, the cohomology groups of the complex of holomorphic differential forms over a Stein manifold X are isomorphic to the cohomology groups $H^*(X,$ **C) (analog of †de Rham's theorem).** Every n-dimensional Stein manifold X is realized as a closed complex analytic submanifold in $\mathbf{C}^{2n+1}$; that is, there exists an injective †proper holomorphic mapping f: $X\rightarrow\mathbf{C}^{2n+1}$ with $df\neq 0$. Consider all the holomorphic †principle fiber bundles over a Stein manifold X whose fibers are isomorphic to a complex Lie group G. The analytic isomorphism classes of the bundles and the elements in $H^1(X,$ G") (where G^a is the sheaf of germs of holomorphic mappings of X into G) are in one-to-one correspondence. The same is true for the topological isomorphism classes of the bundles and the elements in $H^1(X, G^c)$ (where G^c is the sheaf of germs of continuous mappings of X into G). The mapping $H^1(X, G^a)\rightarrow$ $H^1(X, G')$ induced by the canonical injection $G^a \rightarrow G^c$ is bijective (Grauert [13]). Every relatively compact domain in a complex analytic manifold is holomorphically convex if it is strongly pseudoconvex (Grauert [14]). Hence

such a domain is a Stein manifold. It follows from this that every real analytic manifold with countable base for open sets is realized as a closed real analytic submanifold of some **R”.**

The notion of Stein manifolds is generalized to that of weakly 1-complete manifolds: a complex manifold X is called a **weakly 1-complete manifold** if there exists a plurisubharmonic function u of class C^∞ on X, such that for any $c\in\mathbf{R}$, $X_c=\{x\in X\mid u(x)<c\}$ is relatively compact. The family of such manifolds includes compact complex manifolds too. For a weakly 1-complete manifold also, vanishing theorems of cohomology have been established by H. Hironaka, S. Nakano [26], and H. Kazama [2] (- 232 Kähler Manifolds D).

The theory of entire functions of two variables has been developed from a new viewpoint established by T. Nishino. An entire function f of two variables defines a family of Riemann surfaces $\{f=c\ \ c\in C\}$ on $\mathbf{C}^2$, and investigations of the structure of such a family play an important role. For instance the following is proved in this way: If every irreducible component off= c (CE C) is biholomorphic to $\mathbf{C}^1$, then there exists an entire function g such that (f, y): $\mathbf{C}^2\to\mathbf{C}^2$ is a biholomorphic mapping (Nishino [28]). In this theorem, the analyticity of g is obtained from the fact that $\mathbf{C}^2$ is a Stein manifold.

M. Continuation of Analytic Sets

The application of the theory of cohomology with coefficients in sheaves is not restricted to problems concerning Stein manifolds. Given $G_0=\{z\,\|\,|z_j|<1, 1\leqslant j\leqslant n\}$ $(n\geqslant 3)$, $G_1=\{z\,\|\,|z_1|<\varepsilon, |z_j|<1, 2\leqslant j\leqslant n\}$, and $G^{(m)}=G_1\cup(G_0-\{z\mid z_2=\ldots=z_m=0\})$ $(3\leqslant m\leqslant n)$, we have $H^p(G^{(m)}, \mathcal{O}_{G(m)})=0$ $(1\leqslant p\leqslant m-2)$ **(Scheja’s theorem** [32]). Let $\mathcal{F}$ be a coherent analytic sheaf over a domain $G\subset\mathbf{C}^n$. If, for every point z of an analytic set $A\subsetneqq G$, $\mathcal{F}_z=\{0\}$ or $p\leqslant n-\dim A-2$ hd, $\mathcal{F}$ (where hd, $\mathcal{F}$ is the †homological dimension of the $\mathcal{O}_{G,z}$-module $\mathcal{F}_z$), then it follows from Scheja’s theorem that the mapping $H^p(G,\mathcal{F})\to H^p(G-A,\mathcal{F})$ induced by the canonical injection $G-A\to G$ is bijective. This generalizes Hartogs’s continuation theorem for holomorphic functions, which corresponds to the case $p=0$.

Besides the continuation of holomorphic functions, we can consider the continuation of analytic sets. Let A be an analytic set in a domain $G\subset\mathbf{C}^n$ and S an analytic set in G A. A point $z\in G$ is said to be **regular (essentially singular)** with respect to S if the closure $\bar{S}$ of S in G is (is not) analytic at z. If dim $A<$ dim, S for every point $z\in S$, then $\bar{S}$ is analytic in G. If S is purely d-dimensional and S is analytic at a point of a d-dimensional irreducible component A' of A, then S is analytic at every point of A' that is not located in any other irreducible component of A. Furthermore, if dim $A\leqslant$ dim S and S is purely d-dimensional, then the following hold: (1) The set E of essential singularities of S is, if not empty, a purely d-dimensional analytic set in G, formed by irreducible components of A. (2) If every irreducible component of A contains points of E not located in any other irreducible component of A, then $A\subset E$, and A is, if not empty, purely d-dimensional. (3) If every d-dimensional irreducible component of A contains points that are regular with respect to S, then S is a purely d-dimensional analytic set in G (Thullen, Remmert, and Stein). By these results it is possible to give a proof for †Chow’s theorem that every analytic set in $\mathbf{P}_n(\mathbf{C})$ is algebraic.

The continuation of holomorphic functions is related to the continuation of their graphs. W. Rothstein investigated the continuation of analytic sets to obtain the following analog to Hartogs’s theorem of continuity: If $G=G_1\cup G_2$, $G_1=\{z\,\|\,|z_1|<1/2, \sum_{j=2}^n|z_j|^2<1\}$, $G_2=\{z\,\|\,|z_1|<1, 1/2<\sum_{j=2}^n|z_j|^2<1\}$, and $\tilde{G}=\{z\,\|\,|z_1|<1, \sum_{j=2}^n|z_j|^2<1\}$ (the envelope of holomorphy of G) with $n\geqslant 3$, then every purely $(n-1)$-dimensional analytic set A in G extends to an analytic set in $\tilde{G}$; that is, there exists a purely $(n-1)$-dimensional analytic set $\tilde{A}$ in $\tilde{G}$ such that $A=\tilde{A}\cap G$. K. Kasahara and H. Fujimoto generalized this theorem to the case of analytic spaces.

N. Nevanlinna Theory for Several Complex Variables

The Nevanlinna theory investigates holomorphic mappings between complex manifolds. In function theory of one variable, for a holomorphic mapping $f:\mathbf{C}\to\mathbf{P}^1(\mathbf{C})$ from C into the 1-dimensional †complex projective space, the famous +Picard theorem states that if f omits three values, then f must be constant. R. Nevanlinna developed the quantitative theory of value distributions off: L. Ahlfors [1], introduced the geometric approach, and enunciated the principle that the negative curvature of the image manifold restricts a holomorphic mapping. In higher-dimensional situations, this principle has been realized in many cases. The Nevanlinna theory for a holomorphic mapping f: $\mathbf{C}\to\mathbf{P}^n(\mathbf{C})$ with respect to †hyperplanes was established by Ahlfors and H. Weyl and J. Weyl [34]. Holomorphic mappings into projective spaces have been studied in detail by W. Stoll, H. Fujimoto, and M. Green. Following the work of S. S.

Chern [9] of around 1970, the equidimensional case was investigated (P. A. Griffiths, K. Kodaira, S. Kobayashi and T. Ochiai). J. A. Carlson and Griffiths [6] succeeded in obtaining a †defect relation for equidimensional holomorphic mappings and †hypersurfaces of the image manifold. This yields the following Picard-type theorem (after the formulation by F. Sakai, *Inventiones Math.*, 16 (1974)): Let X be a †projective algebraic manifold of dimension n and $f: \mathbf{C}^n \rightarrow X$ a holomorphic mapping. Let D be a hypersurface of X. We take a †desingularization $\pi: X^* \rightarrow X$ so that the inverse image $D^* = \pi^{-1}(D)$ has at most normal crossings as singularities. We denote by K^* a †canonical divisor of X*. We assume that

$$\lim_{m\to\infty} \sup \frac{\dim H^0(X^*, \mathcal{O}(m(K^*+D^*)))}{m^n} > 0.$$

If f omits D, then the Jacobian of f vanishes everywhere. As a corollary, it follows that the †universal covering manifold of X - D cannot be $\mathbf{C}^n$. When $X = \mathbf{P}^n(\mathbf{C})$ and D is a nonsingular hypersurface of degree d, the above assumption is satisfied if $d \geqslant n+2$. In the nonequidimensional case, Ochiai [29] (with supplementary works by M. Green, Y. Kawamata, and P. Wong) verified the following assertion, which was first stated by A. Bloch with rough arguments: Let X be a projective algebraic manifold of dimension n, and suppose $\dim H^1(X, \mathcal{O}_X) > n$; then the image of a holomorphic mapping $f: \mathbf{C} \rightarrow X$ is contained in a proper subvariety of X (→ 124 Distributions of Values of Functions of a Complex Variable; 272 Meromorphic Functions).

0. Hyperbolic Manifolds

Every complex analytic space X has two invariant †pseudodistances: the **Carathéodory pseudodistance** c_X and the **Kobayashi pseudodistance** d_X, both of which generalize the Poincaré distance ρ of the unit disk $D = \{|z| < 1\}$. These pseudodistances can be defined by the property that d_X is the largest pseudodistance among all pseudodistances δ_X on X for which all holomorphic mappings $(X, \delta_X) \rightarrow (D, \rho)$ are distance-decreasing, while c_X is the smallest among all pseudodistances δ_X for which all holomorphic mappings $(D, \rho) \rightarrow (X, \delta_X)$ are distance-decreasing. Then $c_X \leqslant d_X$. If Y is another complex analytic space, every holomorphic mapping $f: X \rightarrow Y$ is distance-decreasing with respect to either its Carathéodory or Kobayashi pseudodistance. This may be considered to be a generalization of the Schwarz-Pick lemma. (The †Schwarz lemma is often referred to by this name, since G. Pick also investigated the distance-decreasing property.) The Kobayashi pseudodistance d_X can also be obtained by integrating an infinitesimal differential pseudometric called the Kobayashi pseudometric in the same manner as the Riemannian distance is obtained from the Riemannian metric (H. L. Royden).

If d_X is a (complete) distance on X, then X is said to be **(complete) hyperbolic.** A Riemann surface is hyperbolic in this sense if and only if its universal covering is biholomorphic to the unit disk. If X is open in Y and if, for every pair of sequences of points $\{p_n\}$ and $\{q_n\}$ in X converging to distinct points of X, $\lim d_X(p_n, q_n)$ is positive, then X is said to be **hyperbolically embedded** in Y. The Riemann sphere $\mathbf{C} \cup \{\infty\}$ minus three points is not only complete hyperbolic but also hyperbolically embedded in the Riemann sphere. Every holomorphic mapping of C into a hyperbolic complex analytic space is constant, while every holomorphic mapping of the punctured disk $D^* = \{0 < |z| < 1\}$ into a hyperbolically embedded space $X \subset Y$ extends to a holomorphic mapping of D into Y. Thus the classical little Picard theorem reduces to the statement that $\mathbf{C} - \{0, 1\}$ is hyperbolic, while the great Picard theorem reduces to showing that $\mathbf{C} - \{0, 1\}$ is hyperbolically embedded in the Riemann sphere. These classical theorems can be generalized in two ways. If M is a †symmetric bounded domain and Γ is a discrete arithmetic group acting freely on M, then M/Γ is not only complete hyperbolic but also hyperbolically embedded in its Satake compactification (Kobayashi and Ochiai). If X is the complement of $2n+1$ hyperplanes in general position in the complex projective space $\mathbf{P}^n(\mathbf{C})$, then X is complete hyperbolic and hyperbolically embedded in $\mathbf{P}^n(\mathbf{C})$ (a restatement of a result going back to E. Borel, A. Bloch, and H. Cartan).

Although there are some noncompact nonhyperbolic complex manifolds X for which every holomorphic mapping $f: \mathbf{C} \rightarrow X$ is constant, a compact complex manifold X is hyperbolic if and only if every holomorphic mapping $f: \mathbf{C} \rightarrow X$ is constant (R. Brody).

For the †Teichmüller space $X = \mathbf{T}_g$ of compact Riemann surfaces of genus g, the Kobayashi distance d_X agrees with the Teichmüller distance which had been introduced before the complex structure of $\mathbf{T}_g$ was defined (H. L. Royden).

In the study of pseudoconvex domains, the †Bergman metric, the Carathéodory distance, and the Kobayashi distance serve as useful tools. Their behavior at the boundary has been studied extensively.

Given an n-dimensional complex manifold X, an invariant measure Ψ_X can be defined as the largest measure such that every holomorphic mapping f from the polydisk D^n with invariant measure into X is measure-decreasing. For an algebraic manifold of general type, this measure is everywhere positive.

P. Bounded Domains in C^n

For any bounded domain in C^n there is naturally assigned a †Kähler metric called the Bergman metric. Using the invariance of this metric, E. Cartan proved that all the Hermitian symmetric spaces of noncompact type are realized as bounded domains. In view of the fact that some important period matrix domains (e.g., the 19-dimensional bounded symmetric domain of type IV, the Siegel upper space, etc.) are of this type, it is obviously of great significance in algebraic geometry to study discontinuous subgroups of the automorphism groups of such domains (Pyatetskiĭ-Shapiro).

On the other hand, it is almost impossible for general bounded domains to determine the explicit form of their Bergman metrics. But Cartan also investigated strongly pseudoconvex real hypersurfaces in $\mathbf{C}^2$, and he solved completely the (local) equivalence problem for them, introducing a definite type of †Cartan connections over them in a functorial way. Thereafter N. Tanaka (and for hypersurfaces also Chern and J. Moser) generalized this result to all pseudoconvex real submanifolds in higher-dimensional complex manifolds. One can further apply this result to the equivalence problem of bounded domains with smooth strongly pseudoconvex boundaries, for C. Fefferman [10] proved by analyzing the boundary behavior of the Bergman metric that every biholomorphic mapping between two such domains is extended smoothly up to their boundaries (I. Naruki [27] gave an alternative proof).

Q. History

In connection with †Abelian functions, analytic functions of several complex variables have been studied sporadically since the time of Riemann and Weierstrass (H. Poincaré, Cousin). A series of investigations by Hartogs ([17] (1906), etc.) that revealed the distinctive properties of several complex variables initiated a new epoch in complex analysis. Levi (1910–1911) generalized Hartogs's results to the case of meromorphic functions, introduced the notion of pseudoconvexity, and proposed the so-called Levi problem. After a lapse of time, many contributions to this new area of complex analysis have been made since 1920. The study by K. Reinhardt (1921) of analytic automorphisms was further developed by C. Carathéodory and Behnke. The †kernel function introduced by S. Bochner and S. Bergmann (1922) produced many remarkable results. In contrast with †Picard's theorem in one variable, P. Fatou found a holomorphic mapping $f: \mathbf{C}^2 \to \mathbf{C}^2$ with nonvanishing Jacobian such that the image $f(\mathbf{C}^2)$ has an exterior point.

The theory of analytic functions of several complex variables has flourished since 1926. Behnke and Thullen in Münster, together with G. Julia and H. Cartan in Paris, were the most active investigators. The results on †normal families of analytic functions of several complex variables (Julia [20], 1926), the uniqueness theorem of holomorphic mappings (H. Cartan, 1930), and a characterization of a domain of holomorphy by holomorphic convexity (Cartan and Thullen, 1932) are their most remarkable achievements. Behnke and Thullen [3] systematized the results obtained since the discovery of the theory by providing a complete bibliography of articles up to 1934.

The three major unsolved problems at that time-those of Cousin, Levi, and the approximation of holomorphic functions-were intensively studied by Oka from 1936, who has given complete solutions [30]. The investigation of ideals of holomorphic functions by H. Cartan (1944), together with that of ideals with undetermined domains by Oka, has developed into the theory of coherent analytic sheaves. The notion of analytic spaces, first introduced by Behnke and Stein (1951), extended the field of investigation in the theory of analytic functions of several complex variables. The theory of cohomology with coefficients in sheaves has been effectively applied by H. Cartan and Serre (1951–1952). The introduction of the notion of Stein manifolds (195 1) came at the same time. Grauert's deep investigations since 1955, together with those of Stein and Remmert, have contributed greatly to the development of the theory of analytic spaces. In the 1960s, active investigations took place also in the United States [16]. The theory of automorphic functions of several complex variables has been developed by C. L. Siegel, I. Satake, and others in connection with the theory of numbers. Entire functions of two variables have been investigated from a new point of view by Nishino and others since 1968 (→ last paragraph of Section N). The notion of a hyperbolic manifold introduced by Kobayashi (1967) enables us to obtain many results on

complex manifolds by means of the methods of differential geometry. The theory of automorphisms of bounded pseudoconvex domains was developed extensively in the 1970s by Fefferman, Naruki, and others. The theory of an envelope of holomorphy has also been successfully applied to the theory of elementary particles in physics.

References

[1] L. Ahlfors, An extension of Schwarz's lemma, Trans. Amer. Math. Soc., 43 (1938), 359–364.
[2] H. Benhke and K. Stein, Konvergentenfolgen von Regularitätsbereichen und der Meromorphiekonvexität, Math. Ann., 116 (1939), 204–216.
[3] H. Behnke and P. Thullen, Theorie der Funktionen mehrerer komplexer Veränderlichen, Springer, 1934; second edition, 1970.
[4] L. Bers, Introduction to several complex variables, Courant Institute, 1964.
[5] S. Bochner and W. T. Martin, Several complex variables, Princeton Univ. Press, 1948.
[6] J. A. Carlson and P. A. Griffiths, A defect relation for equidimensional holomorphic mappings between algebraic varieties, Ann. Math., (2) 95 (1972), 557–584.
[7] Sém. H. Cartan, Ecole Norm. Sup., 1951–1952.
[8] H. Cartan, Oeuvres, Springer, 1979.
[9] S. S. Chern, On holomorphic mappings of Hermitian manifolds of the same dimension, Amer. Math. Soc. Proc. Symposia in Pure Math., 11 (1968), 157–170.
[10] C. Fefferman, The Bergman kernel and biholomorphic mappings of pseudoconvex domains, Inventiones Math., 26 (1974), 1–65.
[1 1] B. A. Fuks, Special chapters in the theory of analytic functions of several complex variables, Amer. Math. Soc. Transl. of Math. Monographs, 1965.
[123 H. Grauert, Charakterisierung der holomorph-vollständiger komplexen Räume, Math. Ann., 129 (1955), 233-259.
[13] H. Grauert, Holomorphe Funktionen mit Werten in komplexen Lieschen Gruppen, Math. Ann., 133 (1957), 450–472.
[14] H. Grauert, On Levi's problem and the imbedding of real-analytic manifolds, Ann. Math., (2) 68 (1958), 460–472.
[15] H. Grauert and R. Remmert, Theory of Stein spaces, Springer, 1979.
[16] R. C. Gunning and H. Rossi, Analytic functions of several complex variables, Prentice-Hall, 1965.
[17] F. Hartogs, Zur Theorie der analytischen Funktionen mehrerer unabhängiger Veränderlichen, insbesondere über die Darstellung derselben durch Reihen, welche nach Potenzen einer Veränderlichen fortschreiten, Math. Ann., 62 (1906), 1–88.
[1 8] L. Hörmander, L^2 estimates and existence theorems for the $\bar{\partial}$ operator, Acta Math., 113 (1965), 89-152.
[193 L. Hörmander, An introduction to complex analysis in several complex variables, Van Nostrand, 1966.
[20] G. Julia, Sur les familles de fonctions analytiques de plusieurs variables, Acta Math., 47 (1926), 53-115.
[21] H. Kazama, Approximation theorem and application to Nakano's vanishing theorem for weakly 1-complete manifolds, Mem. Fac. Sci. Kyushu Univ., 27 (1973), 221–240.
[22] S. Kobayashi, Hyperbolic manifolds and holomorphic mappings, Dekker, 1970.
[23] S. Kobayashi, Intrinsic distances, measures and geometric function theory, Bull. Amer. Math. Soc., 82 (1976), 357–416.
[24] P. Lelong, Fonctions plurisousharmoniques et formes différentielles positives, Cordon & Breach, 1968.
[25] L. Nachbin, Holomorphic functions, domains of holomorphy and local properties, North-Holland, 1970.
[26] S. Nakano, Vanishing theorems for weakly 1-complete manifolds, II, Publ. Res. Inst. Math. Sci., 10 (1974), 101-l 10.
[27] 1. Naruki, On extendibility of isomorphisms of Cartan connections and biholomorphic mappings of bounded domains, Tôhoku Math. J., 28 (1976), 117-122.
[28] T. Nishino, Nouvelles recherches sur les fonctions entières de plusieurs variables complexes (II), Fonctions entières qui se réduisent à celles d'une variables, J. Math. Kyoto Univ., 9 (1969), 221-274.
[29] T. Ochiai, On holomorphic curves in algebraic varieties with ample irregularity, Inventiones Math., 43 (1977), 83–96.
[30] K. Oka, Sur les fonctions analytiques de plusieurs variables (collected papers), Iwanami, 1961.
[31] W. F. Osgood, Lehrbuch der Funktionentheorie II, Teubner, 1924; revised edition, 1929 (Chelsea, 1965).
[32] G. Scheja, Riemannsche Hebbarkeitssätze für Cohomologieklassen, Math. Ann., 144 (1961), 345–360.
[33] K. Stein, Analytische Functionen mehrerer komplexer Veränderlichen zu vorgegebenen Periodizitätsmoduln und das zweite Cousinsche Problem, Math. Ann., 123 (1951), 201–222.
[34] H. Weyl and J. Weyl, Meromorphic functions and analytic curves, Ann. Math.

Studies 12, Princeton Univ. Press, 1943.
Also → references to 23 Analytic Spaces.

22 (I.12) Analytic Sets

A. General Remarks

The notion of analytic sets was first defined by N. N. Luzin and M. Ya. Suslin in 1916, and it was extended to that of projective sets by operations such as complementation and projection (Luzin, 1924). Most mathematicians, including Luzin and W. Sierpinski, who worked in this field, were in agreement with †French empiricism (or †semi-intuitionism), which defended the standpoint of R. Baire, E. Borel, H. Lebesgue, and others. An object is said to be **effectively given** if it can be uniquely, individually, and unambiguously determined in finite terms so that anyone can reach the same object by following the defining procedure. Semi-intuitionists claim that only effectively given objects have mathematical existence, and they do not recognize as a mathematical object something that needs the axiom of choice for its definition. From this point of view, †Borel sets were "well-defined" sets to which classical analysis had to be restricted. Thus the question was raised whether it is possible to extend the class of Borel sets to a wider class of sets with the same certainty. Lebesgue (*J. Math. Pures Appl.*, 1 (1905)) defined a function not belonging to any class of Baire functions by using the totality of †ordinals of the second class. (Later, this method was systematically developed as the †theory of sieves by Luzin.) However, it did not satisfy Borel as being effective. Can we, then, extend the Borel sets without any use of ordinals of the second class? The discovery of analytic sets gave an affirmative answer.

In this article (except in Section 1), we treat a space (denoted by X, Y, .) that is †homeomorphic to a †complete †separable †metric space and its subspace. Denote by $\mathfrak{N}$ the **space of irrational numbers** (a metric space consisting of the irrational numbers $\in \mathbf{R}$ with the metric $|x-y|$ of x and y). The following properties of a subset S of a space X are equivalent: (i) S is a continuous image of $\mathfrak{N}$; (ii) S is a continuous image of a Borel set in X; (iii) S is the projection of a closed set in a product space $X \times \mathfrak{N}$; (iv) S is the projection of a Borel set in X x Y. We call a set satisfying one of these properties an **analytic set,** an **A set,** or a Σ_1^1 **set** (in X). The complement of an analytic set is called a **complementary analytic set** (or simply **coanalytic set),** a **CA set, or** a Π_1^1 **set.**

B. The Operation A and Sieves

When to each $(n_1, , n_k)$ of finite sequences of natural numbers there corresponds a unique element E(n,, , n_k) of a family F of sets, this correspondence {E(n,, . . , n,)} is called a **schema of Suslin** (or **system of Suslin)** consisting of sets in F. Denoting an infinite sequence of natural numbers by $\{n_i\}$, the set given by $\bigcup_{\{n_j\}} \bigcap_k E(n_1, , n_k)$ is called the **kernel** of a system of Suslin, and the operation of taking the kernel is called the **operation A (analytic operation).**

Let Q be the set of all rational numbers between 0 and 1 and F be a family of sets. Take a family $\{C_r\}_{r\in Q}$ of sets belonging to F with the index set Q (or more geometrically, a subset C= $\bigcup_{r\in Q} C_r \times (r)$ of X x Q when F is a family of subsets of a space X), and call it a **sieve** consisting of sets in F. Denoting by $\{r_i\}$ a (strictly) monotone decreasing sequence of elements of **Q**, we call the set $\bigcup_{\{r_i\}} \bigcap_k C_{r_k}$ (namely, the set of all x such that $C^{(x)} =$ $\{r \mid (x, r) \in C\}$ is not well-ordered by the order $\leqslant$ of rational numbers) the set obtained by a sieve C or the **sieved set** obtained by C. If the family F is closed with respect to countable intersection, then the family of all sets obtained by sieves consisting of sets in F is identical to the family of all sets obtained by applying the operation A to F. When F consists of all the closed sets in a given space, this is the family of all analytic sets. In particular, it is sufficient to take the family of closed intervals as F when X is the space of real numbers. Note that we can define sieves and sieved sets more generally by using the space of real numbers **R** instead of the set Q of rationals.

C. Properties of Analytic Sets

It is evident from the definition that every †Borel set is analytic. If a Borel set is uncountable, then it is the union of a countable set and a one-to-one continuous image of $\mathfrak{N}$. The analyticity of sets is invariant under countable unions, intersections, and Cartesian products and the operation A and †Borel-measurable transformations. An uncountable analytic set contains a †perfect subset (Suslin). Therefore the possible cardinality of an analytic set is at most countable or is that of the continuum. Every analytic set enjoys the †Baire property, and in Euclidean space every analytic set is †Lebesgue measurable (Luzin, Sierpinski). If a set E in the Euclidean plane is analytic (coana-

lytic), then I(E) is also analytic (coanalytic), where I(E) is the set of all x such that the section $E^{(x)}$ of E that is parallel to the y-axis has a positive measure (M. Kondô and T. Tugué). Concerning the Baire property, the similar result for a set E of X x Y is obtained by replacing "to have a positive measure" by "to be the †second category (nonmeager) set" (A. S. Kechris). The Lebesgue measure of an analytic set is effectively calculable (in the sense that the measure of a Σ_1^1 (or $\Sigma_1^1(\alpha)$) set is a Σ_1^1 (resp. $\Sigma_1^1(\alpha)$) real number (- Section G and 356 Recursive Functions) (H. Tanaka). Every analytic (coanalytic) set E can be decomposed into $\aleph_1$ Borel sets. This decomposition is called a decomposition of E into **constituents.** An analytic (coanalytic) set is a Borel set if and only if it is decomposable into a countable number of constituents (Luzin, Sierpinski). In a space with the cardinality of the continuum, there exist analytic sets that are not Borelian. For example, in the space C([0, 1]) of continuous functions on the interval [0, 1] (→ 168 Function Spaces) the set of all differentiable functions is coanalytic but not Borelian (S. Mazurkiewicz).

The following theorems are especially important in analytic set theory. **Luzin's first principle** (the first separation theorem): For every pair of disjoint analytic sets A_1, A_2, there exists a Borel set B such that $A_1 \subset B$ and $B \cap A_2 = \varnothing$. An immediate corollary of Luzin's first principle is **Suslin's theorem:** If both A and X − A are analytic,' then A is a Borel set. **Luzin's second principle** (the second separation theorem): For every pair of analytic sets A and B, there exist complementary analytic sets C and D such that $A-B\subset C$, $B-A\subset D$, and $C \cap D = \varnothing$. A one-to-one continuous image of a Borel set is Borelian (Suslin). More generally, for a given B-measurable function f defined on a Borel set B, the set A $(\subset f(B))$ of all points y whose inverse images f^{-1} (y) are singletons is a complementary analytic set **(Luzin's unicity theorem).** In this theorem, we can replace "a singleton" by "a †σ-compact set" (V. Ya. Arsenin, K. Kunugui). Therefore, if a set is the image of a Borel set by a continuous function such that the inverse image of each point is a σ-compact set, then it is a Borel set.

D. Generalization to Projective Sets

A **projective set of class** n is inductively defined as follows: (i) the Borel sets are the projective sets of class 0; (ii) the projective sets of class 2n + 1 are the continuous images of the sets of class $2n$; (iii) the projective sets of class 2n are the complements of the sets of class 2n − 1.

The projective sets of class 1 are exactly the analytic sets, and those of class 2 are the complementary analytic sets. The following are fundamental properties of projective sets. Denote by L_n the family of the projective sets of class n. Then (1) $L_{2n}\subset L_{2n+k}$ and $L_{2n+1}\subset L_{2n+2+k}$ $(k=1,2,\ldots)$. (2) the property of being a set of class n is invariant under countable unions, intersections, and Cartesian products and homeomorphisms; (3) a continuous image of a projective set of class $2n+1$ is of the same class; (4) the projection on X of a set of class $2n+1$ in X x Y is a set of class $2n+1$ in X; (5) the family of the projective sets of class 2n + 1 in a space X is the family of the projections of all sets of class 2n in X x X (or X x $\mathfrak{N}$); (6) the kernel of a system of Suslin consisting of sets of class n is a projective set of the same class, where $n\neq 0, 2$.

We frequently call a projective set of class 2n 1 a P_n **set** or a Σ_n^1 **set,** and that of class $2n$ a C_n **set** or a Π_n^1 **set. A** B_n **set** or a Δ_n^1 **set** is a set that is both P_n and C_n. The respective families of these sets are also denoted by P_n (Σ_n^1), C_n (Π_n^1), and B_n (Δ_n^1), respectively. In general, for a family $\mathfrak{F}$ of sets in a space X, we denote by $C\mathfrak{F}$ the family of the complements X-E of all sets E in $\mathfrak{F}$. We write $\mathrm{Sep}_{\mathrm{I}}(\mathfrak{F})$ and $\mathrm{Sep}_{\mathrm{II}}(\mathfrak{F})$ for the propositions obtained by substituting "set in $\mathfrak{F}$," "set in $\mathfrak{F}$ and in $C\mathfrak{F}$," and "set of $C\mathfrak{F}$" for "analytic set," "Borel set," and "coanalytic set," respectively, in Luzin's first principle and Luzin's second principle, respectively. Furthermore, we say that the reduction principle holds for $\mathfrak{F}$, and denote it by $\mathrm{Red}(\mathfrak{F})$, when for any sets A, $B\in\mathfrak{F}$ there exist A_1, $B_1\in\mathfrak{F}$ such that $A_1\subset A$, $B_1\subset B$, $A_1\cup B_1=A\cup B$, and $A_1\cap B_1=\varnothing$. C. Kuratowski introduced the latter principle $\mathrm{Red}(\mathfrak{F})$ which implies $\mathrm{Sep}_{\mathrm{I}}(C\mathfrak{F})$ and $\mathrm{Sep}_{\mathrm{II}}(C\mathfrak{F})$, and proved $\mathrm{Red}(\Pi_1^1)$. Classically, reduction or separation principles for the projective sets of higher classes were not settled except for $\mathrm{Sep}_{\mathrm{I}}(\Pi_2^1)$, $\mathrm{Sep}_{\mathrm{II}}(\Pi_2^1)$ (P. S. Novikov), and $\mathrm{Red}(\Sigma_2^1)$ (Kuratowski). Nowadays, it is known that these principles for $\{\Sigma_n^1,\Pi_n^1\}_{n\geqslant 3}$ are undecidable from the axiomatic set theory ZFC. If we assume †the axiom of constructibility $V=L$, then $\mathrm{Red}(\Sigma_n^1)$ holds for $n\geqslant 3$ (J. W. Addison). On the other hand, under the assumption of †projective determinacy PD (→ Section H), it follows that $\mathrm{Red}(\Pi_n^1)$ $(\mathrm{Red}(\Sigma_n^1))$ holds when n is odd (even) (A. Martin, Addison and Y. N. Moschovakis).

E. Universal Sets

A set U in $\mathfrak{N}$ x X is called the **universal set** for the projective sets of class n in X if for any projective set P of class n in X, there exists $z_0\in\mathfrak{N}$ such that $P=\{x\mid(z_0,x)\in U\}$. Concern-

ing universal sets, we have the following result: for every $n>0$, there exists a universal set for the projective sets of class n in X that is of the same class in $\mathfrak{N}$ x X. Hence in a space with the cardinality of the continuum, there exists a projective set of class $n+1$ which is not of class n. In general, if a class $\mathfrak{F}$ of sets is closed under taking a universal set and continuous preimages, then $\mathfrak{F}$ and $C\mathfrak{F}$ cannot both have the reduction property. Therefore Red(Π_n^1) or Red(Σ_n^1) fails for each $n \geqslant 1$.

Any two universal sets of analytic sets are †Borel isomorphic. On the other hand, an analytic set E is not Borel isomorphic to any universal analytic set if the complement of E is an uncountable set without a perfect subset (A. Maitra and C. Ryll-Nardzewski). Hence, there are at least two equivalence classes of uncountable analytic sets with respect to the Borel isomorphism ≅ under the assumption of $V=L$, since the latter assumption implies the existence of such a set E (- Section H).

F. Uniformization Principle

The uniformization problem arose during investigations of implicit functions. For a set E in a space X x Y, **uniformization of** E is the finding of a subset V of E such that

$$\forall x(\exists y((x,y)\in E)\Leftrightarrow \exists! y((x,y)\in V)),$$

where $\exists! y$ is the †quantifier which means "there exists exactly one y." A Borel set can be uniformized by choosing a suitable coanalytic set (Luzin). **Kondô's uniformization theorem** (*Japan. J.* Math., 15(1938)) is the most important result in descriptive set theory: every coanalytic set is uniformizable by a coanalytic set. As a corollary to this, any Σ_2^1 set is uniformizable by a Σ_2^1 set, and every Σ_2^1 set is obtained as a one-to-one continuous image of a coanalytic set.

Kondô's original proof was very difficult to understand. Since †effective descriptive set theory was introduced by Kleene and Addison, the proof has been simplified by Addison, and a more elegant one has been given by J. R. Shoenfield [7, p. 188]. Nowadays, the theorem is also called the Novikov-Kondô-Addison theorem, and is one of the most powerful and fundamental theorems not only in descriptive set theory, but also in the †foundations of mathematics.

The uniformization of an analytic set is, in general, not to be found among analytic or coanalytic sets. There was a conjecture that any analytic set is uniformizable by specifying an A, set (difference of two analytic sets). Recently, this conjecture was negatively settled (J. Steel, Martin). Assuming that $V=L$, the uniformization of a Σ_n^1 ($n \geqslant 3$) set is determined by specifying a Σ_n^1 set, and that of a Π_n^1 ($n \geqslant 2$) set by specifying a $\Pi_{n\rho}^1$ set. On the other hand, if an axiom system of set theory (e.g., ZF; → 33 Axiomatic Set Theory) is consistent, then it is still consistent even if we add to it the following proposition: There exists a Π_2^1 set whose uniformization is impossible by any choice of a definable set in the system (P. J. Cohen, A. Lévy). However, if we assume the existence of a †measurable cardinal (MC), it is known that every Π_2^1 set is uniformizable by a Π_2^1 set (Martin and R. M. Solovay). On the other hand, PD implies "every Π_n^1 (Σ_n^1) set is uniformizable by a Π_n^1 (Σ_n^1) set for each odd (resp. even) n" (Moschovakis).

There are sufficient conditions on sections $E^{(x)}$ of a Borel set E in the Cartesian product X x Y for E to be uniformizable by the Borel set as follows: all sections $E^{(x)}$ are (i) countable (Luzin, Novikov), (ii) a-compact (Arsenin, Kunugui); or, as "large section property," (iii) in the second category (S. K. Thomason, P. G. Hinman), or (iv) of positive measure (Tanaka, G. E. Sacks). For applications of descriptive set theory to analysis, an important uniformization result is **von Neumann's selection theorem:** for any **A** set $E\subset$ X x Y, there exists a †Baire measurable and †absolutely measurable function $f: X\rightarrow$ Y (- 270 Measure Theory L (vi)) such that

$$\forall x(\exists y((x,y)\in E)\Leftrightarrow (x,f(x))\in E).$$

Concerning implicit functions, any Borel set E ⊂ X x Y such that all sections $E^{(x)}$ are at most countable is expressed by a union of intersections of E with graphs of some Baire functions $f_n: X\rightarrow Y, n\in \mathbf{N}$ (Luzin). Recently, it was shown known as a generalization of this that any Borel set $E\subset X$ x Y such that all sections are a-compact is a union of countably many Borel sets E_n for each of which all sections $E_n^{(x)}$ are compact (J. Saint-Raymond).

G. Kleene's Hierarchy and Effectiveness

First, projective set theory in any space is reducible to the theory in the space of irrational numbers. Second, if we introduce a †weak topology in the set $\mathbf{N}^{\mathbf{N}} = {}^{\omega}\omega$ of †number-theoretic functions α with one argument, the resulting topological space $\mathbf{N}^{\mathbf{N}}$ is homeomorphic to the †Baire zero-dimensional space of irrational numbers. Third, any subset B of $\mathbf{N}^{\mathbf{N}}$ is open and closed in this topology if and only if there exist a function $\xi(\in\mathbf{N}^{\mathbf{N}})$ and a predicate $A^{\xi}(\alpha)$ that is †general recursive in ξ such that $\alpha\in B\Leftrightarrow A^{\xi}(\alpha)$. Fourth, logical operations such as ¬, v , A, 3x (where x is a variable ranging over the natural numbers), and $\exists\alpha$

exactly correspond to the operations (on sets) complementation, union, intersection, countable union, and projection, respectively. On the basis of these facts, projective set theory is regarded as the theory of the N^N †analytic hierarchy of Kleene. Here, the following example is remarkable: We can construct a Σ_1^1 set which is universal for the analytic sets (namely, a Σ_1^1 [N”] set) in the space of irrational numbers (→ 356 Recursive Functions H).

The connection between projective set theory and logic has been discussed by C. Kuratowski and A. Tarski. From their point of view, semi-intuitionists such as Borel regard the set of natural numbers to be precisely clear in itself and also the continuum to be immediately recognizable by our geometric intuition. In their argument rational numbers do not play such an important role. They take, a priori, the set of irrational numbers as the fundamental domain, and intervals with rational extremities as the simplest sets of points among the subsets of the fundamental domain as the starting point of their argument. Here, the fundamental domain or each interval is not conceived as a totality of its elements, but recognized as a “uniform extent.” In contrast to this, singletons and individual irrational numbers are not so simple. For this reason Borel introduced the notion of calculable numbers to study definable real numbers. Following Luzin, we say that a **calculable number** is a constructible real number in the sense that we can give it by an arithmetical approximation as precisely as we want. Now, this notion is nearly identical to the notion of an effectively calculable real number given by A. Church or A. M. Turing.

In the mathematics of the semi-intuitionists, the word “effective” has played an especially important role. Although these mathematicians have always agreed not to accept the †axiom of choice, the exact meaning of “effective” has differed slightly among different members of their group or in different stages of the development of the theory. Such differences mainly arose in connection with the question: How can we tell whether given entities are finitary or individual? One way to guess the original intention held by Borel and others when they used the term “effective” is to replace the term by “recursive.” Nowadays, the concept of “effectiveness” is used in this sense (→ 356 Recursive Functions C), and all classical results in descriptive set theory essentially have effective versions (or refinements via relativization). For example, the Novikov-Kondô-Addison theorem is described as follows: Any Π_1^1 ($\Pi_1^1(\alpha)$ for an $\alpha \in {}^\omega\omega$) set is uniformizable by a Π_1^1 (resp. $\Pi_1^1(\alpha)$) set. Similarly, Suslin’s theorem is that every analytic and coanalytic (i.e., $\Delta_1^1(\alpha)$) set A is a Borel set of class v(x) which is effective relative to the “definability α” of given A, where v(a) is the +Constructive ordinal relative to α (Kleene, Tugué and Tanaka, A. Louveau, etc.).

H. Further Results in Axiomatic Set Theory and Strong Axioms

The recent development of axiomatic set theory has yielded the following propositions. Under the assumption that $V = L$ (K. Gödel, Novikov, Addison): (1) there exists an uncountable Π_1^1 set that does not contain any perfect subset; (2) there exists a nonmeasurable Δ_2^1 set; (3) there exists a Δ_2^1 set that does not have the Baire property. On the other hand, if the axioms of set theory ZFC plus “there exists an †inaccessible cardinal number” are consistent, then so is “every projective set is Lebesgue measurable, has the Baire property, and has a perfect subset when it is uncountable” with them (Solovay). Concerning these properties, †Martin’s axiom MA and “$2^{\aleph_0} > \aleph_1$” (MA + ¬CH is consistent with ZFC; → 33 Axiomatic Set Theory E) implies that every Σ_2^1 is measurable, and has the Baire property (Martin and Solovay). The possible cardinality of a Σ_2^1 set is at most $\aleph_1$ or that of the continuum, which is implied from “every Σ_2^1 set is a union of $\aleph_1$ Borel sets” (Sierpinski). It is known that one cannot prove or refute from ZFC the converse of the latter statement. If one assumes MC (the existence of a measurable cardinal), then every Σ_3^1 set is a union of $\aleph_2$ Bore1 sets. Therefore MC implies that the possible cardinality of a Σ_3^1 set is at most $\aleph_2$ or that of the continuum.

With each set $A \subset {}^\omega\omega$ we associate the following infinite game of perfect information $G(A)$, played by two players I and II. First player I chooses $n_0 \in \omega$, then player II chooses $n_1 \in \omega$, then I chooses $n_2 \in \omega$, and so on. The game ends after ω steps. Let r(k) = n_k; if XE A, then I wins $G(A)$; otherwise II wins. A strategy (for 1 or II) is a rule that tells the player what move to make, depending on the previous moves of both players. A winning strategy is one such that the player who follows it always wins. The game $G(A)$ is determined if one of the players has a winning strategy. The axiom of **projective determinacy** PD is the assertion that for every projective set $A(\subset {}^\omega\omega)$ the game $G(A)$ is determined. PD is widely used to solve problems on projective sets that are not decidable from ZFC, and is a plausible hypothesis that does not seem to contradict the axiom of choice at present; and it has pleasing consequences in descriptive set theory. For n 3 3,

reductions and uniformization results of Σ_n^1 (Π_n^1) sets mentioned earlier are such examples. Furthermore, PD implies that every projective set is Lebesgue measurable, has the Baire property, and contains a perfect subset if it is uncountable.

Now, write Determinacy (Σ_n^1) (Determinacy (Δ_n^1)), or simply, Det(Σ_n^1) (Det(Δ_n^1)), when G(A) is determined for every Σ_n^1 (Δ_n^1) set *A*. Using the fact that the open game is determined, namely Det(Σ_1^0) (D. Gale and F. M. Stewart), D. Blackwell has given a short proof of $\mathrm{Sep}_I(\Sigma_1^1)$. The extremely difficult problem was to show Det(Δ_1^1) in ZF (indeed, one cannot prove Det(Δ_1^1) only from the Zermelo axioms z; → 33 Axiomatic Set Theory B). In 1975, Martin solved this problem. Concerning Σ_1^1, Det(Σ_1^1) cannot be proved from only the ZF axioms, but MC implies Det(Σ_1^1) (Martin). However, it is known that Det(Σ_2^1) cannot be proved even if we assume MC.

By a pre-well-ordering, we mean a relation $\leqslant$ having all the properties of a well-ordering except for †antisymmetry. Denote by δ_n^1 the least ordinal that cannot be the length of a preewell-ordering on ${}^\omega\omega$ belonging to Δ_n^1. It is a classical result that $\delta_1^1=\omega_1$ (= K,). The following results are mainly due to Martin: $\delta_2^1 \leqslant \aleph_2$, MC $\Rightarrow \delta_3^1 \leqslant \aleph_3$, and also, PD $\Rightarrow$ $\delta_4^1 \leqslant \aleph_4$. Let κ be an infinite cardinal. A set $A \subset X$ is called **κ-Suslin** when there exists a closed set $C \subset X \times {}^\omega\kappa$ (${}^\omega\kappa=\{f \mid f:\omega\rightarrow\kappa\}$) such that

$$x \in A \Leftrightarrow \exists f[f \in {}^\omega\kappa \wedge (x,f) \in C].$$

A is $\aleph_0$-Suslin iff $A \in \Sigma_1^1$ If *A* is a Σ_2^2 set, then *A* is $\aleph_1$-Suslin (Shoenfield). Under the assumptions of MC, if *A* is a Σ_3^1 set, then it is $\aleph_2$-Suslin (Martin). Furthermore, Det(Δ_{2n}^1) implies the following facts (Kechris, Moschovakis): every Σ_{2n+2}^1 set is κ-Suslin, where κ is the cardinal of δ_{2n+1}^1; every Σ_{2n+1}^1 set in κ-Suslin for some $\kappa < \delta_{2n+1}^1$

The **axiom of determinacy AD** is the assertion that the game *G(A)* is determined for every set A. AD implies that every set is measurable (J. Mycielski and S. Swierczkowski), has the Baire property, and has a perfect subset if it is uncountable (Morton Davis). Though AD contradicts the axiom of choice, some consequences of AD in the area of projective sets are more desirable than the consequences of the axiom of choice. The following are examples of consequences of AD. (1) *A* is a Σ_2^1 set iff *A* is a union of $\aleph_1$ Bore1 sets (Moschovakis). (2) For each *n*, δ_n^1 is a cardinal (Moschovakis). (3) Generalization of Suslin's theorem holds as follows: $\Delta_n^1 = \mathbb{B}_{\delta_n^1}$ (= the smallest †Boolean algebra containing the open sets and closed under complementation and unions of length $<\delta_n^1$) (Martin, Moschovakis).

I. Polish Spaces, Luzin Spaces, and Suslin Spaces

A topological space homeomorphic to a complete separable metric space is called a **Polish space**. A subspace *E* of a Polish space X is Polish if and only if it is a G_δ-subset of X, i.e., a countable intersection of open subsets of X (Aleksandrov and Uryson). A Hausdorff topological space X is called a **Luzin space** (resp. **Suslin space**) if we can find a Polish space S and a continuous bijective (resp. surjective) mapping $f: S \rightarrow X$. Every Polish space is a Luzin space and every Luzin space is a Suslin space.

Let X be a Hausdorff topological space. A subset *E* of X is called a **standard set** (resp. **analytic set)** if the set *E* with the relative topology is a Luzin (resp. Suslin) space. The analytic subsets are closed under analytic operations. Every analytic set is obtained from closed sets by applying the analytic operation. Every analytic set is †universally measurable. Every standard set is a Borel set. Every countable union or intersection of standard sets is standard.

A subset *E* of a Suslin space is a Borel set if and only if both *E* and E^c are analytic. A subset *E* of a Luzin space is a Bore1 set if and only if *E* is standard.

Let *f* be a Borel measurable mapping from a Suslin space X into another Suslin space Y. Then the image $f(X)$ is an analytic subset of Y. Furthermore, if *f* is injective, then f'gives a †Borel isomorphism between X and $f(X)$. If both X and Y are Luzin spaces and if *f* is injective, then f(X) is a Borel subset of Y. Every Suslin space is †Borel isomorphic to an analytic subset of **R,** and every Luzin space is Borel isomorphic to one of the following spaces: (1) $\mathbf{N}_n=\{1,2,\ldots,n\}$, (2) $\mathbf{N}=\{1,2,\ldots\}$, (3) **R**.

The selection theorem due to von Neumann (- Section F) holds when X and Y are Suslin spaces. This fact and its ramifications [15,161 are useful in nonlinear functional analysis and control theory.

Practically all useful spaces appearing in functional analysis and probability theory are Polish or Luzin spaces. Examples: (i) Every locally compact Hausdorff space with a countable open base is Polish. (ii) Every separable Banach space is Polish. (iii) The set C of all continuous functions on [0, 1] with the †topology of uniform convergence is a Polish space. (iv) The set *D* of all right continuous functions on [0, 1] with left limits is a Polish space when it is endowed with the Skorokhod topology (- 250 Limit Theorems in Probability Theory E). (v) The space $\mathscr{D}'$ of distributions and the space $\mathscr{S}'$ of tempered distri-

butions (→ 125 Distributions and Hyperfunctions) are Luzin spaces.

References

[1] C. Kuratowski, Topologie I, Warsaw, revised edition, 1948.
[2] A. A. Lyapunov (Liapunov), E. A. Shchegol'kov (Stschegolkow), and V. J. Arsenin, Arbeiten zur deskriptiven Mengenlehre, Deutscher Verlag der Wissenschaften, 1955.
[3] N. N. Luzin, Sur les ensembles analytiques, Fund. Math., 10 (1927), 1–95.
[4] N. N. Luzin, Leçons sur les ensembles analytiques et leurs applications, Gauthier-Villars, 1930.
[5] W. Sierpiński, Les ensembles projectifs et analytiques, Mémor. Sci. Math., Gauthier-Villars, 1950.
[6] H. Rogers, Jr., Theory of recursive functions and effective computability, McGraw-Hill, 1967, ch. 16.
[7] J. R. Shoenfield, Mathematical logic, Addison-Wesley, 1967, ch. 7.
[S] D. A. Martin, Descriptive set theory: Projective sets, Handbook of Mathematical Logic, North-Holland, 1977, ch. C.8.
[9] T. Jech, Set theory, Academic Press, 1978, pt. IV.
[10] P. G. Hinman, Recursion-theoretic hierarchies, Springer, 1978, pt. B.
[11] Y. N. Moschovakis, Descriptive set theory, North-Holland, 1980.
[12] C. A. Rogers, J. E. Jayne, C. Dellacherie, F. Topsøe, J. Hoffman-Jørgensen, D. A. Martin, A. S. Kechris, and A. H. Stone, Analytic sets, Academic Press, 1980.
[133 N. Bourbaki, Éléments de mathématique, Topologie générale, Hermann, 1974, ch. X, sec. 6.
[14] J. Hoffmann-Jørgensen, The theory of analytic spaces, Various publication series 10, Math. Institute, Aarhus Univ., 1970.
[15] M.-F. Sainte-Beuve, On the extension of von Neumann-Aumann's theorem, J. Functional Anal., 17 (1974), 112–129.
[16] A. F. Filippov, On certain questions in the theory of optimal control, SIAM J. Control, 1 (1962), 76-84.

23 (XI.21) Analytic Spaces

A. General Remarks

An †analytic function of a complex variable has as its natural domain of definition a †Riemann surface, i.e., a 1-dimensional complex †analytic manifold. In the case of several complex variables, the set of zeros of an analytic function, the quotient space of a domain by a †properly discontinuous group of analytic automorphisms, the existence domain of an †algebroidal function, etc., are, strictly speaking, not necessarily complex analytic manifolds. It is necessary to consider a more general category of complex analytic manifolds with singularities, and the notion of analytic spaces is drawn from these examples. Many of the properties of complex analytic manifolds are extended to analytic spaces; on the other hand, theories specific to analytic spaces have also been developed.

B. Analytic Sets

We say that a subset A of a complex analytic manifold G is an **analytic set** in G if it is a closed subset and each point of A has a neighborhood U such that $U \cap A$ is the set of common zeros of a finite number of holomorphic functions in U. Specifically, if A is locally the set of zeros of a single holomorphic function that does not vanish identically, then A is called **principal.** Two subsets S_1 and S_2 of G are called equivalent at $z^0 \in$ G if there exists a neighborhood U of z^0 such that $S_1 \cap U = S_2 \cap U$. By this equivalence relation, every subsets S of G defines its germ S_{z^0} at z^0. A **germ of an analytic set** at z^0 is the germ at z^0 of an analytic set in a neighborhood of z^0. Each germ A_0 of an analytic set at $z^0 = 0 \in G$ is associated with an ideal $I(A_0) = \{f \mid f \in H(0), f \mid A_0 = 0\}$ in the ring H(0) of †germs of holomorphic functions at **0.** We call A_0 **reducible** if A_0 is the union of two germs of analytic sets A_0' and A_0'' with $A_0' \neq A_0$, $A_0'' \neq A_0$; otherwise, A_0 is called **irreducible.** An analytic set A is called **irreducible at 0** if the germ at 0 of A is irreducible. Properties of A_0 and $I(A_0)$ correspond to each other. Thus A_0 is irreducible if and only if $I(A_0)$ is prime.

As the ring H(0) is †Noetherian, in a neighborhood of every point z^0 an analytic set A is represented as a union of a finite number of analytic sets A_i that are irreducible at z^0. These A_i are essentially unique. If an analytic set A is irreducible at z^0, then there exists a system of local coordinates $(z_1, \ldots, z_n)$ centered at z^0 and a pair of natural numbers $d \leqslant n$ and k such that, in a neighborhood of z^0, A is a k-sheeted **ramified covering space** with covering mapping $\varphi:(z_1, \ldots, z_n) \rightarrow (z_1, \ldots, z_d)$; i.e., for an analytic set R in a neighborhood of $0 \in \mathbf{C}^d$, $\varphi: A-\varphi^{-1}(R) \rightarrow \mathbf{C}^d - R$ is, in a neighborhood of z^0, a k-sheeted covering mapping, where $A - \varphi^{-1}(R)$ is a connected d-dimensional com-

plex analytic manifold in the neighborhood. The coordinates of the k points of $A-\varphi^{-1}(R)$ over each point $(z_1, , z_d)$ are holomorphic in these variables. The number d is the **(local) dimension** of A at z^0 and is denoted by $\dim_{z^0} A$. From this local representation, we obtain **Riickert's zero-point theorem:** Every prime ideal $\mathfrak{P}$ in H(0) is equal to $I(A_0)$, with A, an irreducible germ of an analytic set at 0. In this case the dimension at 0 of an analytic set A that defines A, is equal to the †Krull dimension of the †local ring $H(0)/\mathfrak{P}$. The theory of local rings is very important in the study of germs of analytic sets. The dimension of a general analytic set A at z^0 is detïned by $\dim_{z^0} A = \sup_i \dim_{z^0} A_i$, where $A = \bigcup_i A_i$ in a neighborhood of z^0, with the A_i irreducible at z^0. If $\dim_{z^0} A_i$ is equal to d for all i, then A is called **purely d-dimensional at** z^0. The **(global) dimension** of A is defined by dim A = $\sup_{z\in A}$ dim, A. A **purely d-dimensional analytic set** is defined to be an analytic set that is purely d-dimensional at every one of its points.

A point z^0 of an analytic set A is called **ordinary (regular** or **simple)** if A has the structure of a complex analytic submanifold in a neighborhood of z^0. The set A' of ordinary points of A is dense and open in A. The set A* = A **-A'** of **singular** (not ordinary) points is an analytic set in G. If A is purely d-dimensional, then A' is a d-dimensional complex analytic manifold and A* is an analytic set of dimension $\leqslant d-1$.

Let A be an analytic set of dimension d in G, and B a purely d'-dimensional analytic set in G -A with $d' > d$. Then the closure $\bar{B}$ of B in G is a purely d'-dimensional analytic set **(Remmert-Stein continuation theorem** [17]).

For every analytic set A in G, the †analytic sheaf $\mathscr{I}(A)$ of germs of holomorphic functions over G that vanish on A is †coherent (H. Cartan). We call $\mathscr{I}(A)$ the sheaf of ideals defined by an analytic set A. Let $\mathcal{O}_G$ be the sheaf of germs of holomorphic functions over G. Then $\mathcal{O}_A = (\mathcal{O}_G/\mathscr{I}(A))|A$ is a coherent sheaf of rings over A. $\mathcal{O}_A$ is called the sheaf of germs of holomorphic functions on an analytic set A.

C. Analytic Spaces

A †ringed space (X, 0,) with Hausdorff base space X is called an **analytic space** if for every point $x \in X$, there exists an open neighborhood U of x such that the ringed space $(U, \mathcal{O}_X|U)$ is isomorphic to a ringed space $(A, \mathcal{O}_A)$, where A is an analytic set in an open set G of some C". The structure sheaf $\mathcal{O}_X$ is then called a **sheaf of germs of holomorphic functions.** The notion of †holomorphic mapping from one open set in $\mathbf{C}^n$ into another is generalized to the case of mappings from one analytic set into another. An analytic set Y in an analytic space X and the sheaf $\mathcal{O}_Y$ of germs of holomorphic functions on Y are detïned as in the case where X is a complex manifold. The ringed space (Y, 0,) is an analytic space and is called an **analytic suhspace** of X. For an analytic space X, the notions of dim, X, dim X, irreducibility, and pure dimensionality are detïned as for an analytic set $A \subset G \subset$ C". Every analytic space X is the union of a locally finite family of irreducible analytic subspaces X, called the **irreducihle components** of X.

Let φ : X → Y be a holomorphic mapping of an analytic space X into another, Y. Its **rank at** $x \in$ X is detïned by $r_\varphi(x)$ = dim, X − $\dim_X \varphi^{-1}(\varphi(x))$. The number $r_\varphi = \sup_{x\in X} r_\varphi(x)$ is called the **rank** of φ. **The set of degeneracy** E_φ of φ is the set of points $x \in X$ such that $r_{\varphi|X'}(x) < r_{\varphi|X'}$ for an irreducible component X' of X through x. The mapping φ is **nondegenerate** if $E_\varphi = \varnothing$. For any $k \in N$, $\{x \in X$ $r_\varphi(x) \leqslant k\}$ is an analytic set (R. Remmert [16]). In particular, E_φ is analytic. For a holomorphic mapping φ :X → Y, the inverse image of an analytic set in Y is an analytic set in X. However, the image of an analytic set is not necessarily analytic. If φ is †proper, then the image $\varphi(X')$ of an analytic set X' in X is an analytic set in Y of dimension $r_{\varphi|X'}$ and is irreducible if X' is irreducible **(Remmert's theorem** [16]).

D. Modifications and Resolution of Singularities

Let M be a subset of an analytic space X. If, for every point x $\in$ X, there exists an open neighborhood U of x and and an analytic set M* in U, containing $U \cap M$ such that U-M* is dense in U, then M is called **analytically thin.** Let $\varphi: X \to$ Y be a holomorphic mapping. Suppose that there exist two analytically thin sets A4 ⊂ X, N ⊂ Y such that φ induces an isomorphism between X − M and Y − N. Then X is called a **holomorphic modification** of Y. If furthermore φ is proper, then X is called a **proper modification** of Y. A **monoidal transformation** of an analytic space X with respect to a coherent sheaf of ideals $\mathscr{I}$ is detïned as in the case of a complex manifold or an algebraic variety (→ 16 Algebraic Varieties L; 72 Complex Manifolds H). It is a proper modification f: X* → X such that the inverse image ideal sheaf $f^{-1}\mathscr{I} \cdot \mathcal{O}_{X^*}$ is invertible, and is universal among all proper holomorphic mappings $h: Z \to X$ with the property that $h^{-1}\mathscr{I} \cdot \mathcal{O}_{Z^*}$ is invertible, where $f^{-1}\mathscr{I}$ is the †inverse image of $\mathscr{I}$. If $\mathscr{I}$ is the sheaf of ideals detïned by an analytic set Y in X, the mon-

oidal transformation of X with respect to $\mathscr{I}$ is often called the **blowing-up** of X with center Y.

H. Hironaka [14] proved that, if X is an analytic space which is countable at infinity (i.e., a countable union of compact sets), then there is a proper modification $\pi: X' \rightarrow X$ with X' smooth (i.e., free from singular points). Such a modification is called a **desingularization** (or **resolution of singularities**) of X. Moreover, over any relatively compact open set U of X, π is the product of a finite sequence of blowing-ups $\pi_i: X'_i \rightarrow X'_{i-1}$ $(X'_0 = X)$, with smooth centers Y'_{i-1} along which X'_{i-1} is †normally flat. This deep result enables one to derive properties of analytic spaces from those of complex manifolds.

Let X and Y be two analytic spaces and G an analytic set in X x Y. If the canonical projection $\pi: G \rightarrow X$ is a holomorphic (or proper) modification, then we say that a **meromorphic mapping** (a proper meromorphic mapping) μ of X into Y is defined. The set G is then called the **graph** of μ. A holomorphic mapping $\varphi: X \rightarrow$ Y can be viewed as a proper meromorphic mapping. Let μ : X → Y be a proper meromorphic mapping. Then $\mu(x)$ (the projection of $\pi^{-1}(x)$ into Y) is a nonempty analytic set in Y for every point x ∈ X. Moreover, there exists an analytic set N, with X-N dense in X, such that μ maps X-N into Y holomorphically. The smallest set N with this property is called the **set of points of indeterminacy** or the **singularity set** of μ. A meromorphic mapping $f: X \rightarrow \mathbf{P}^1(\mathbf{C})$ is called a **meromorphic function** on X if none of the irreducible components of X is mapped to $\{\infty\}$ by f. The set $f^{-1}(0) = \pi((X \times \{0\}) \cap G)$ is called the set of zero points of X, and the set $f^{-1}(\infty)$ is called the set of poles. These are analytic sets in X. Let $f_1, \ldots, f_k$ be meromorphic functions on X. Then, by a suitable proper modification of X, one can eliminate the points of indeterminacy of the meromorphic mapping $f: X \rightarrow (\mathbf{P}^1(\mathbf{C}))^k$ defined by $x \rightarrow (f_1(x), \ldots, f_k(x))$, i.e., one can modify f to be holomorphic. The ring of meromorphic functions on X is invariant under proper modifications of X. If X is irreducible and compact, then the field of meromorphic functions on X is a simple algebraic extension of the †field of rational functions of k $(\leqslant \dim X)$ variables.

Let (X, 0,) be an analytic space. A point x ∈ X is called **normal** for X if $\mathscr{O}_{X,x}$ is a †normal local ring. The set of nonnormal points for X is an analytically thin analytic set in X (K. Oka). Every ordinary point of X is normal. We call X **normal** if every one of its points is normal. Every nondegenerate holomorphic mapping of an irreducible X into an irreducible and normal Y is an open mapping if its rank is equal to the dimension of Y (Remmert).

For every analytic space X, there exists a proper modification $\nu: \tilde{X} \rightarrow X$ with $\tilde{X}$ normal such that v is nondegenerate and v $\tilde{X} - \nu^{-1}(S)$ is an isomorphism, where S is the set of nonnormal points. Such a proper modification of X is unique up to isomorphisms. We call $\tilde{X}$ a **normalization** of X with normalizing mapping v.

Let φ : X → Y be a holomorphic modification. Suppose that Y is normal at $\varphi(x^0)$ $(x^0 \in X)$. If the set $\varphi^{-1}(\varphi(x^0))$ contains an isolated point, then $\varphi^{-1}(\varphi(x^0)) = x^0$, and φ is an isomorphism in a neighborhood of x^0 (an analog of †Zariski's main theorem). In particular, if $\varphi: X \rightarrow$ Y is a holomorphic modification and Y is normal, then φ maps X $-E_\varphi$ isomorphically onto the dense open set $\varphi(X$ $E_\varphi)$. Furthermore, if a holomorphic mapping φ : X → Y is injective and X and Y are irreducible, normal, and n-dimensional, then $\varphi(X)$ is an open set in Y and $\varphi^{-1}: \varphi(X) \rightarrow X$ is holomorphic.

E. Analytic Spaces in the Sense of Behnke and Stein

Let $\varphi: \tilde{G} \rightarrow G$ be a proper continuous mapping of a connected †locally compact space $\tilde{G}$ onto a domain $G \subset \mathbf{C}^n$. The triple $\mathfrak{G} = (\tilde{G}, \varphi, G)$ is an **analytic covering space** over G if the following conditions are satisfied: (i) $\varphi^{-1}(z^0)$ is a finite set for every point $z^0 \in$ G. (ii) There exists an analytic set A $\subset G$ of dimension $\leqslant n-1$ such that φ $\tilde{G} - \varphi^{-1}(A)$ is a local homeomorphism and every point of $\varphi^{-1}(A)$ has a fundamental system of neighborhoods U such that both U and $U - \varphi^{-1}(A) \neq \varnothing$ are connected. As $\mathfrak{G}$ is unramified over G-A, the number of points in $\varphi^{-1}(z^0)$ is constant for $z^0 \in$ G-A and is called the **number of sheets** of $\mathfrak{G}$. A point $\tilde{z} \in \tilde{G}$ is called a **ramification point** of $\mathfrak{G}$ if the restriction of φ to any neighborhood of $\tilde{z}$ is not a homeomorphism. Denote by B the set of ramification points of $\mathfrak{G}$. Then $\varphi(B) \subset A$ is an analytic set of dimension $n-1$. Let f be a continuous complex-valued function in an open set D in $\tilde{G}$. We call f **holomorphic** in D if for every point $\tilde{z}^0 \in D$ B and for every open neighborhood V of $z^0 = \varphi(\tilde{z}^0)$ over which φ is a homeomorphism, $f \circ \varphi^{-1}$ is holomorphic in V. Denote by $\mathscr{O}_{\tilde{G}}$ the sheaf of germs of holomorphic functions over $\tilde{G}$. Then $(\tilde{G}, \mathscr{O}_{\tilde{G}})$ is a ringed space. An **analytic space in the sense of Behnke and Stein** is a Hausdorff ringed space (X, 0,) that is locally isomorphic to a ringed space of the form $(\tilde{G}, \mathscr{O}_{\tilde{G}})$ [3]. †Riemann's theorem on removable singularities holds for such spaces. Every normal analytic space is an analytic space in the sense of Behnke and Stein. An analytic covering space $\mathfrak{G} = (\tilde{G}, \varphi, G)$

is a **C-covering space** (covering space in the sense of Cartan) if for every point $z^0 \in G$ there exist an open neighborhood V of z" and a holomorphic function g in $U = \varphi^{-1}(V)$ which can be defined by an irreducible polynomial of degree k such that its coefficients are holomorphic functions on V and the coefficient of its highest term is 1, where k is the number of sheets of $\mathfrak{G}$. A **C-analytic space** is an analytic space in the sense of Behnke and Stein that is locally isomorphic to a C-covering space. The category of C-analytic spaces coincides with that of normal analytic spaces. According to H. Grauert and Remmert [10], every analytic covering space is a C-analytic covering space. Therefore every analytic space in the sense of Behnke and Stein is a normal analytic space.

Let R be an equivalence relation in an analytic space X. Given a subset A of X, denote by $R[A]$ the set of points of X which are R-equivalent to points of A. We call R **proper** if $R[K]$ is compact for every compact set K in X. Let φ be a proper holomorphic mapping of an analytic space X into another Y. For x, x'E X, let x =x'(R) be defined by $\varphi(x) = \varphi(x')$. The equivalence relation R is then proper. We consider the quotient space X/R and the canonical projection $p: X \rightarrow X/R$. With each open set U in the quotient space X/R *we* can associate the ring of holomorphic functions in $p^{-1}(U)$ that are constant on $p^{-1}(\bar{x})$ for every $\bar{x} \in U$. This leads to a ringed space $(X/R, \mathcal{O}_X/R)$, which is proved to be an analytic space by **Grauert's theorem** [9]: All the †direct images of a coherent analytic sheaf over X by a proper holomorphic mapping $\varphi: X \rightarrow$ Y are coherent. For every proper equivalence relation R in X, the ringed space $(X/R, \mathcal{O}_X/R)$ is an analytic space if and only if for every point $\bar{x} \in X/R$ there exists an open neighborhood V of $\bar{x}$ such that functions in $\Gamma(V, \mathcal{O}_X/R)$ separate the points of V (H. Cartan).

F. Stein Spaces

For an analytic space (X, $\mathcal{O}_X$) let us consider the following conditions: (i) Functions in I(X, $\mathcal{O}_X$) separate the points of X. (ii) X is K-**complete**; i.e., for every point $x \in X$ there exist a finite number of $f_i \in$ I(X, $\mathcal{O}_X$) (i = 1, , k) such that the holomorphic mapping $f = (f_i)$: X $\rightarrow \mathbf{C}^k$ is nondegenerate at x. (iii) Every compact analytic set in X is a finite set. Condition (i) implies (ii), and (ii) implies (iii). If an irreducible analytic space X is K-complete, then X is a countable union of compact sets (Grauert [S]). In fact, if $n =$ dim X, there exist functions $f_i \in \Gamma(X, \mathcal{O}_X)$ $(i = 1, \ldots, n)$ such that the holomorphic mapping $f = (f_i)$: X $\rightarrow \mathbf{C}^n$ is nondegenerate. The notion of holomorphic convexity (→ 21 Analytic Functions of Several Complex Variables H) is carried over to analytic spaces. For a holomorphically convex analytic space, conditions (i), (ii), and (iii) are equivalent (Grauer [8]). A **Stein space** (or **holomorphically complete space)** is a holomorphically convex analytic space that satisfies one of the conditions (i), (ii), or (iii). In a holomorphically convex analytic space $(X, \mathcal{O}_X)$, let R be the equivalence relation defined by I(X, 6,); i.e., for x, x' $\in$ X, $x \equiv$ x'(R) if and only if f(x) $= f(x')$ for every $f \in \Gamma(X, \mathcal{O}_X)$. Then R is proper. The analytic space $(X/R, \mathcal{O}_X/R)$ is a Stein space. A Stein space is a generalization of the notion of a Stein manifold. Fundamental theorems A and B on Stein manifolds hold Verbatim for Stein spaces (→ 21 Analytic Functions of Several Complex Variables L). Therefore the main properties of Stein manifolds are inherited by Stein spaces. Let φ: X → Y be a holomorphic mapping of an analytic space X into another, Y. If for every $x \in X$ all the connected components of the fibers $\varphi^{-1}(\varphi(x))$ are compact, then the equivalence relation R' defined by those components (i.e., for x and x' in X, x $\equiv x'(R')$ if and only if x and x' belong to the same component of $\varphi^{-1}(\varphi(x))$) is proper, and the ringed space $(X/R', \mathcal{O}_X/R')$ is an analytic space. In particular, if X is a holomorphically convex irreducible analytic space and R is the equivalence relation defined by $\Gamma(X, \mathcal{O}_X)$, then all the fibers of the canonical projection p: X $\rightarrow X/R$ are connected.

G. Further Topics

The notion of analytic space can be generalized as follows (Grauert [9]). A ringed space **(X,** $\mathcal{O}_X$) is a **general analytic space** if it is locally isomorphic to a ringed space $(A, \mathcal{H}_A)$, where A is an analytic set in a domain G $\subsetneqq \mathbf{C}^n$, and $\mathcal{H}_A = (\mathcal{O}_G/\mathcal{J})|_A$ for some coherent analytic subsheaf $\mathcal{J}$ of $\mathcal{I}(A)$ such that Supp$(\mathcal{O}_G/\mathcal{J})$ (= $\{z \in$ G $(\mathcal{O}_G/\mathcal{J})_z \neq 0\}$) = A. An **analytic subspace** of (X, $\mathcal{O}_X$) is a ringed space (Y, $\mathcal{O}_Y$) where Y = Supp$(\mathcal{O}_X/\mathcal{I})$ and $\mathcal{O}_Y = \mathcal{O}_X/\mathcal{I}$ for some coherent sheaf of ideals $\mathcal{I}$ of $\mathcal{O}_X$. (Y, $\mathcal{O}_Y$) is also a general analytic space. A. Douady [6] showed that for any general analytic space (X, $\mathcal{O}_X$) there exists a natural structure of a general analytic space on the totality of all the compact analytic subspaces of (X, $\mathcal{O}_X$). The resulting analytic space is called the **Douady space** of (X, $\mathcal{O}_X$). For the proof, the notion of a **Banach analytic space** is used, which is obtained by first defining analytic subspaces in an open subset of a complex †Banach space and then patching them. A Douady space can

be used, for example, to show that the identity component of the analytic automorphism group of a compact Kähler manifold is a †complex Lie group which is naturally an extension of a †complex torus by a †linear algebraic group (A. Fujiki, D. Lieberman; → 232 Kähler Manifolds C).

References

[1] S. S. Abhyankar, Local analytic geometry, Academic Press, 1964.
[2] C. Banica and 0. Stanasila, Algebraic methods in the global theory of complex spaces, Wiley, 1976.
[3] H. Behnke and K. Stein, Modifikation komplexer Mannigfaltigkeiten und Riemannscher Gebiete, Math. Ann., 124 (1951), l-1 6.
[4] L. Bers, Introduction to several complex variables, Courant Institute, 1963.
[5] Sém. H. Cartan, Ecole Norm. Sup., 1953-1954, 1957-1958, 1960-1961.
[6] A. Douady, Le problème de modules pour les sous-espaces analytiques complexes d'un espace analytique donné, Ann. Inst. Fourier, 16 (1966), 1-95.
[7] G. Fischer, Complex analytic geometry, Springer, 1976.
[8] H. Grauert, Charakterisierung der holomorph-vollstandiger komplexen Räume, Math. Ann., 129 (1955), 233-259.
[9] H. Grauert, Ein Theorem der analytischen Garbentheorie und die Modulraume komplexer Strukuren, Publ. Math. Inst. HES, 5 (1960).
[10] H. Grauert and R. Remmert, Komplexe Räume, Math. Ann., 136 (1958), 245-318.
[11] H. Grauert and R. Remmert, Analytische Stellenalgebren, Springer, 1971.
[12] R. C. Gunning and H. Rossi, Analytic functions of several complex variables, Prentice-Hall, 1965.
[13] M. Hervé, Several complex variables, local theory, Tata Inst. Studies in Math., Oxford Univ. Press, 1963.
[14] H. Hironaka, Desingularization of complex-analytic varieties, Act. Congr. Intern. Math., Nice, 1972, Gauthier-Villars, vol. 2, 6277632.
[15] R. Narashimhan, Introduction to the theory of analytic spaces, Lecture notes in math. 25, Springer, 1966.
[16] R. Remmert, Holomorphe und meromorphe Abbildungen komplexer Räume, Math. Ann., 133 (1957), 3288370.
[17] R. Remmert and K. Stein, Über die wesentlichen Singularitaten analytischer Mengen, Math. Ann., 126 (1953), 2633306.
Also → references to 21 Analytic Functions of Several Complex Variables.

24 (Xx1.1) Ancient Mathematics

A. General Remarks

To determine the beginning of the history of mathematics, one must define the term "mathematics." Only speculation based on the observation of primitive peoples today can be made regarding the development of the number concept among prehistoric peoples. The prehistoric period ended in Egypt and Mesopotamia c. 3000 B.C., and a little later in the valleys of the large rivers in India and China.

Since the basis of the civilizations in the river valleys of the ancient world was agriculture, the administrators first had to control watering systems through irrigation, drainage, pumping, and canalization; second, they had to measure land and harvests for tax collection; and third, they had to establish a calendar by observation of the heavenly bodies. All these tasks demanded some knowledge of mathematics.

Additional knowledge of mathematics was certainly needed for construction of the palaces and tombs. We know something of the development of mathematical knowledge during these ages from some recovered artifacts, but there remains the possibility of new finds that will bring about a basic change in our knowledge of the history of mathematics during this period.

B. Mathematics in Egypt

The main sources for our understanding of the history of mathematics in Egypt are the Moscow papyrus and the more important Rhind papyrus, both discovered in the 19th Century. The Greeks place the origin of their mathematics in Egypt, but it seems that Egyptian mathematics was limited to practical mathematics. The Egyptians had a decimal numeration system, but the place value was not clear; they used fractions, which they always decomposed into the sums of unit fractions (i.e., fractions with 1 as numerator); they solved the problems of everyday arithmetic that were reducible to linear equations; they computed approximate areas and volumes of some figures for the purpose of measurement of farmland or granaries and for construction work; they had exact formulas for the computation of areas of triangles and of trapezoids; and they used $(16/9)^2 = 3.1605$ as the value of π; but no trace has been found to prove the existence of demonstrative mathematics in ancient Egypt.

C. Mathematics in Mesopotamia

Sources abound for the study of Mesopotamian mathematics, and these sources may very well increase in the future. The Mesopotamians kept exact records of astronomical observations for long periods of time. Their more advanced mathematics was not limited to practical use, as was that of the Egyptians. They used a sexagesimal system of numeration with place value, and also used sexagesimal fractions; however, they lacked a cipher to denote zero until the 4th Century B.C., and they did not have a symbol corresponding to our decimal point, so that exact place value had to be determined from the context of each expression. They had a multiplication table and tables of inverses, squares, and cubes of numbers, and they used these tables to solve equations, even some simple equations of the third degree, as well as simultaneous equations of the second degree for two unknowns. They had accurate solutions for quadratic equations (expressed in words); they discarded negative roots, but they admitted both positive roots when two existed. They studied integral solutions of $a^2+b^2=c^2$ (the largest of their solutions were 12,709, 13,500, and 18,541) and approximate computation of quadratic roots. which suggests some relation to Greek mathematics. We have evidence that some of the geometric algebra in Euclid's *Elements* can be traced to Mesopotamian algebra. Some historians also affirm that the concept of demonstration in Greek mathematics originated with the Mesopotamians, but this theory lacks suffcient proof.

By the 7th Century, the Mayas in Central America also possessed a numeration system, with the base 20. As far as we know, the Mesopotamians and the Mayas were the earliest people to possess numeration systems with place value (- 57 Chinese Mathematics, 187 Greek Mathematics, 209 lndian Mathematics).

References

[1] 0. Neugebauer, Vorlesungen über Geschichte der antiken mathematischen Wissenschaften I, Springer, 1934.
[2] 0. Neugebauer, Mathematische Keilschrift-Texte, Springer, 1935.
[3] A. B. Chace, The Rhind mathematical papyrus II, Mathematical Association of America, 1929.
[4] B. L. van der Waerden, Science awakening, Noordhoff, 1954.
[5] 0. Neugebauer, The exact sciences in antiquity, Brown Univ. Press, second edition, 1957.
[6] K. Vogel, Vorgriechische Mathematik I. Vorgeschichte und Ägypten II. Die Mathematik der Babylonier, Schroedel, 1958–1959.

25 (XX.35) Approximation Methods in Physics

A. Introduction

It is often possible to solve a differential equation of mathematical physics analytically by expanding the solution in a series with respect to a small (or large) parameter involved in the equation. In general, such an approach is called a **perturbation method** (- Section B). For an equation which is difficult to treat analytically, it is possible to get a numerical solution either by replacing the derivatives by difference quotients (†difference method), or by expanding the solution in terms of suitable functions and determining the expansion coefficients numerically from the equation rewritten in a weak form (†variational method, †method of weighted residuals, †Galerkin's method, †finite-element method, etc.). For such numerical methods → 46 Calculus of Variations, 301 Numerical Solution of Algebraic Equations, 302 Numerical Solution of Linear Equations, 303 Numerical Solution of Ordinary Differential Equations, 304 Numerical Solution of Partial Differential Equations, 441 Variational Principles.

In Section C we describe a method for deriving an asymptotic expression for a function written as an integral in a complex plane.

B. Perturbation Method

(1) Regular Perturbation. (a) Initial-value problem. Consider an initial-value problem of an ordinary differential equation involving a small parameter ε:

$$\frac{du}{dt}=f(t,u;\varepsilon), \tag{1}$$

$$u=b(\varepsilon)\quad\text{at}\quad t=0. \tag{2}$$

We assume that f is sufficiently smooth as a function of t and is regular as a function of u and ε, and that b is a regular function of ε.

Substituting the power series in ε for f, u, and b,

$$f(t,u;\varepsilon)=f^{(0)}(t,u)+\varepsilon f^{(1)}(t,u)+,$$

$$u(t;\varepsilon)=u^{(0)}(t)+\varepsilon u^{(1)}(t)+\ldots)$$

$$b(\varepsilon)=b^{(0)}+\varepsilon b^{(1)}+,$$

into equations (1) and (2) and equating terms of equal power, we get

$$\frac{du^{(0)}}{dt}=f^{(0)}(t,u^{(0)}),\qquad u^{(0)}(0)=b^{(0)},$$

$$\frac{du^{(1)}}{dt}=f_u^{(0)}(t,u^{(0)})u^{(1)}+f^{(1)}(t,u^{(0)}),\quad u^{(1)}(0)=b^{(1)},$$

$$\frac{du^{(2)}}{dt}=f_u^{(0)}(t,u^{(0)})u^{(2)}+\frac{1}{2}f_{uu}^{(0)}(t,u^{(0)})u^{(1)2}$$
$$+f_u^{(1)}(t,u^{(0)})u^{(1)}+f^{(2)}(t,u^{(0)}),$$
$$u^{(2)}(0)=b^{(2)},$$

the solutions of which determine the perturbation series for u. The equation for $u^{(0)}$ (the unperturbed equation) is nonlinear in general, but those for $u^{(1)}$, $u^{(2)}$, are inhomogeneous linear equations with identical principal parts.

The system of differential equations

$$\frac{du_i}{dt}=f_i(t,u_1,\ldots,u_n;\varepsilon)\qquad(i=1,2,\ldots,n),$$

$$u_i=b_i(\varepsilon)\ \text{at}\ t=0,$$

or a differential equation of higher order,

$$\frac{d^n u}{dt^n}=f\left(t,u,\ldots,\frac{d^{n-1}u}{dt^{n-1}};\varepsilon\right),$$

$$u=b^{(0)}(\varepsilon)\ ,\ldots,\ \frac{d^{n-1}u}{dt^{n-1}}=b^{(n-1)}(\varepsilon)\ \text{at}\ t=0,$$

can be treated in a similar way.

(b) Boundary-value problem. Given a linear differential operator $H = H_0 + \varepsilon V$ involving a small parameter ε linearly (H_0 is a linear differential operator and V is a function of $(x_1,x_2,\ldots,x_n)$), consider the boundary-value problem

$$(H_0+\varepsilon V)u=f(x_1,\ldots,x_n)\qquad(x_1,\ldots,x_n)\in\Omega,\ (3)$$

$$Bu=0\qquad(x_1,\ldots,x_n)\in\partial\Omega,\qquad(4)$$

where B is a linear and homogeneous operator.

Substitution of the Taylor expansion

$$u=u_0+\varepsilon u_1\quad+$$

into equations (3) and (4) gives

$$H_0u_0=f,\qquad Bu_0=0,$$

$$H_0u_1+Vu_0=0,\qquad Bu_1=0,$$

$$\ldots.$$

of which the solutions are

$$u_0=Kf,\qquad u_1=-KVKf,$$

$$u_2=KVKVKf,\ldots.$$

Here $K\equiv H_0^{-1}$ is an integral operator such that, for any function $v = v(x_1, x_2, \ldots, x_n)$, we have

$$(Kv)(x_1,\ldots,x_n)=\int\cdots\int G(x_1,\ldots,x_n|\xi_1,\ldots,\xi_n)$$
$$\times\ v(\xi_1,\ldots,\xi_n)d\xi_1\cdots d\xi_n,$$

where $G(x_1,\ldots,x_n|\xi_1,\ldots,\xi_n)$ is the Green's function for the unperturbed problem. The perturbation series

$$u=Kf-\varepsilon KVKf+\varepsilon^2KVKVKf-\ldots$$

converges if VK is bounded and $|\varepsilon|<1/\|VK\|$.

(c) Eigenvalue problem. The solution of an eigenvalue problem

$$(H_0+\varepsilon V)u=\lambda u\qquad(5)$$

with the boundary condition $Bu = 0$ for the operators considered in (b) can be obtained in power series in ε:

$$\lambda=\lambda_0+\varepsilon\lambda^{(1)}+\varepsilon^2\lambda^{(2)}+\ldots,$$

$$u=u_0+\varepsilon u^{(1)}+\varepsilon^2u^{(2)}+\ldots,$$

where u_0 and λ_0 are the eigenfunction and nondegenerate eigenvalue, respectively, of the unperturbed operator H_0, which is assumed to be self-adjoint. Substitution of the series into (5) gives

$$(H_0-\lambda_0)u^{(1)}=-Vu_0+\lambda^{(1)}u_0,\qquad(6)$$

$$(H_0-\lambda_0)u^{(2)}=-Vu^{(1)}+\lambda^{(2)}u_0+\lambda^{(1)}u^{(1)}$$

$$(H_0-\lambda_0)u^{(3)}=-Vu^{(2)}+\lambda^{(3)}u_0+\lambda^{(2)}u^{(1)}$$
$$+\lambda^{(1)}u^{(2)},$$

If we normalize u_0 and u by the conditions $(u_0,u_0) = 1$ and $(u,u_0) = 1$, we get from equation (6)

$$\lambda^{(1)}=(Vu_0,u_0)\equiv\alpha.$$

Let S be the inverse of $H_0-\lambda_0 I$ in the subspace perpendicular to u_0 (we put $Su_0 = 0$), then each term of the perturbation series for u and λ is obtained in the following way from the system of the equations given above:

$$u^{(1)}=-SVu_0,$$

$$\lambda^{(2)}=-(VSVu_0,u_0),$$

$$u^{(2)}=S(V-\alpha)SVu_0,$$

$$\lambda^{(3)}=(VS(V-\alpha)SVu_0,u_0),$$

$$u^{(3)}=-S(V-\alpha)S(V-\alpha)SVu_0$$
$$+(VSVu_0,u_0)S^2Vu_0,$$

(2) Singular Perturbation. Formal application of the procedure described in (1) often fails when the term including the highest-order derivative is multiplied by a small parameter (or a lower-order term is multiplied by a large parameter), i.e., when we deal with singular

perturbation problems. Typical methods for treating such cases are described below.

(a) **WKB method.** The method for getting an asymptotic solution of the second-order differential equation

$$\frac{d^2u}{dx^2}+k^2P(x)u=0 \tag{7}$$

for large values of k was developed for problems arising in classical wave motion (Jeffreys, 1924 [1]) and for problems of quantum mechanics (Wentzel, Kramers, Brillouin, 1926), and is called the **Jeffreys method, WKB method,** or **WKBJ method.**

If the function P in (7) does not vary rapidly, the solution is expected to have the form

$$u\propto\exp\{ik\varphi(x)\},$$

with $\varphi(x)$ not very different from a linear function of x, since the exact solution for $P = P_0 =$ const is $\exp(\pm ik\sqrt{P_0}\,x)$.

Substitution of this expression for u in equation (7) gives

$$ik\varphi''-k^2\varphi'^2+k^2P=0.$$

Neglecting the first term, we get

$$\varphi'^2=P,\quad \text{i.e.,}\quad \varphi(x)=\int^x P^{1/2}\,dx.$$

This result suggests a transformation of the form

$$w=P^{1/4}u,\qquad z=\int^x P^{1/2}\,dx,$$

which gives

$$\frac{d^2w}{dz^2}+(k^2-Q)w=0, \tag{8}$$

where

$$Q=P^{-1/4}\frac{d^2}{dz^2}(P^{1/4})=-P^{-3/4}\frac{d^2}{dx^2}(P^{-1/4}).$$

Equation (8) can be transformed further into a Volterra's integral equation of the second kind:

$$w(z;k)=Ae^{ikz}+Be^{-ikz}+\frac{1}{k}\int_{z_0}^{z}\sin\{k(z-\zeta)\}Q(\zeta)w(\zeta;k)\,d\zeta.$$

Therefore, if we expand w as

$$w(z;k)=w_0(z)+\frac{1}{k}w_1(z)+\ldots, \tag{9}$$

we have

$$w_0(z)=Ae^{ikz}+Be^{-ikz},$$

$$w_n(z)=\int_{z_0}^{z}\sin\{k(z-\zeta)\}Q(\zeta)w_{n-1}(\zeta)\,d\zeta \qquad (n=1,2,\ldots).$$

If P is positive for the interval $z_0\leqslant z\leqslant z_1$ in question and has continuous derivatives up to second order, the foregoing series converges to the solution of (8).

On the other hand, when kz takes complex values, the series (9) does not converge in general. The series expansion appropriate for this case is

$$w(z;k)=\exp\{\psi(z;k)\},\quad \psi=\pm ikz+\sum_{1}^{\infty}\frac{1}{(ik)^n}\psi_n(z),$$

and we get the asymptotic expansion for w:

$$w\sim\exp\left\{\pm ikz\mp\frac{i}{2k}\int Q\,d\zeta+\frac{1}{4k^2}Q(z)\pm\frac{i}{8k^3}\left[Q'(z)-\int^z Q^2\,d\zeta\right]+\frac{1}{16k^4}\{Q''(z)-2Q(z)^2\}+\cdots\right\}$$

Now, in many cases, P takes both positive and negative values within the interval in question, e.g., in the problem of transonic flow of a gas, in the problem of determining the reflectivity and transmissivity of a wave propagating in a nonuniform medium, or in the quantum-mechanical problem of a material wave passing through a potential wall. The point at which P vanishes is called a **turning point.** In this case there arises the problem of analytic continuation of the solution obtained in one side into the other through the turning point. For the special case when $P\propto x^n$ (the turning point is chosen as $x = 0$), it can be shown that the exact solution of equation (7) is explicitly given by

$$w\propto\sqrt{z}\,Z_\nu(kz),$$

where $\nu = 1/(n+2)$, and Z_ν is a cylindrical function of order ν. Based on this fact, the continuation formula for the case when P is approximated by x^n near $x = 0$ has been obtained.

(b) **Lighthill's method.** With nonlinear equations, formal application of a perturbation procedure sometimes leads to the difficulty that the solution of the unperturbed equation has a singularity not exhibited by the exact solution of the original equation and whose order increases in higher approximations, so that the perturbation series ceases to have any meaning near this point. Such a difficulty can often be overcome by so-called coordinate straining. This method has been applied to a number of fluid-dynamical problems, such as those arising in theories of aircraft wings, boundary layers, and shock waves.

Before describing the method, it is convenient to sketch **Poincaré% method** of getting the perturbation series for the period of a nonlinear oscillation. In order to solve the

nonlinear equation

$$\frac{du}{dt}=f(u;T_0,\varepsilon),$$

where T_0 is the period of oscillation when the small parameter ε is taken to be zero, Poincaré wrote series expansions for both u and the period T:

$$u=u_0(t)+\varepsilon u_1(t)+\ldots,\qquad T=T_0+\varepsilon T_1+\ldots,$$

and determined $T_1, T_2, \ldots$ in such a manner as to avoid resonance at each step in the determination of u,(t), u,(t), . .

Based on this idea, Lighthill(l949) [2] developed coordinate straining in order to overcome the difficulty stated above. We illustrate the method here by means of the following simple example.

Consider a boundary-value problem of a nonlinear differential equation

$$(x+\varepsilon u)\frac{du}{dx}+u=0\ \ (0<x<1): \tag{10}$$

$$u(1)=1. \tag{11}$$

Application of the usual perturbation method would give the series solution

$$u=\frac{1}{x}+\frac{\varepsilon}{2}\left(\frac{1}{x}-\frac{1}{x^3}\right)-\frac{\varepsilon^2}{2}\left(\frac{1}{x^3}-\frac{1}{x^5}\right)+\ldots.$$

As is seen, however, the order of an apparent singularity occurring at x = 0 increases as we proceed to higher approximations so that it is not possible to see the true behavior of the solution near x = 0.

In Lighthill's method, we assume that series expansion is valid not only for u but also for the independent variable x, i.e.,

$$u=u_0(\xi)+\varepsilon u_1(\xi)+\ldots,\qquad x=\xi+\varepsilon x_1(\xi)+\ldots.$$

Substitution of these series into equations (10) and (11) gives

$$\xi u_0'+u_0=0,\qquad u_0(1)=1,$$

$$\xi u_1'+u_1=-(u_0x_1'+u_0'x_1+u_0u_0'),$$

$$u_1(1)=0,\qquad x_1(1)=0,$$

From the first equation we have $u_0=1/\xi$. Before solving the second equation for u_1, we determine $x_1(\xi)$ so as to avoid the increase of the order of singularity in $u_0(\xi)$ at ξ =O. This requirement is satislïed by taking $x_1 = (1/2)(\xi\quad 1/\xi)$, and this gives $u_1\equiv 0$; thus, at this stage of approximation, we have

$$u=u_0(\xi)+\varepsilon u_1(\xi)=\frac{1}{\xi},$$

$$x=\xi+\varepsilon x_1(\xi)=\xi+\frac{\varepsilon}{2}\left(\xi-\frac{1}{\xi}\right)$$

The guiding principle is the same for higher approximations. (The foregoing expression happens to be the exact solution, but this a fortuitous result of our choice of this particular example.)

The convergence of the series obtained by Lighthill's method has been proved by Wasow (1955) [3] for the more general problem

$$(x+\varepsilon u)\frac{du}{dx}+q(x)u=r(x),$$

$$u(a)=b,$$

where ε, a, and b are positive constants and $q(x)$ and r(x) are functions regular in $|x|\leqslant a$.

Lighthill's method can also be applied to partial differential equations. Sometimes it is called the **PLK method (Poincaré-Lighthill-Kuo method),** after its successful application to the problem of the boundary layer of a thin flat plate by Kuo.

(c) Method of matched asymptotic expansions. In some cases it is possible to get a uniformly valid series solution if we divide the domain into two or more subdomains, solve the equation by use of a suitable independent variable for each subdomain, and then determine the coefficients in each solution by the process of matching neighboring solutions on their common boundary. Such a perturbation technique was developed in treating the boundary-layer equation in fluid dynamics and has been systematized into the **method of matched asymptotic expansions.**

The idea of the method can be shown by the following simple example. Consider the boundary-value problem

$$\varepsilon\frac{d^2u}{dx^2}+\frac{du}{dx}=a\qquad(0<x<1),$$

$$u(0)=0,\qquad u(1)=1, \tag{12}$$

where ε is a small positive parameter and a is a constant such that $0<a<1$. In order to solve this problem, we first take x itself to be an independent variable (the **outer variable)** in the domain $x\gg\varepsilon$ and expand the solution as

$$u=\quad U_0(x)+\varepsilon U_1(x)+$$

Because the highest derivative is multiplied by a small parameter ε in equation (12), the perturbation equations to be solved are all of first order, of which the first one is

$$\frac{dU_0}{dx}=a,\qquad U_0(1)=1.$$

Solving this, we get an approximation for the **outer solution:**

$$U_0(x)=1-a(1-x).$$

On the other hand, we introduce a new independent variable ξ (the **inner variable)** by

putting $x = \varepsilon\xi$ for $x \lesssim \varepsilon$. This transforms equation (12) into

$$\frac{d^2v}{d\xi^2}+\frac{dv}{d\xi}=\varepsilon a, \qquad v(\xi)\equiv u(x).$$

If we assume the expansion

$$v=V_0(\xi)+\varepsilon V_1(\xi)+\ldots,$$

we have the equation for V_0:

$$\frac{d^2V_0}{d\xi^2}+\frac{dV_0}{d\xi}=0, \qquad V_0(0)=0.$$

This gives an approximation for the **inner solution:**

$$V_0(\xi)=c(1-e^{-\xi}).$$

Finally, in order to match the inner and the outer solutions, we equate their values at some point within the interval $\varepsilon \ll x \ll 1$ ($\xi\to\infty$ for the inner solution, $x\to 0$ for the outer solution):

$$V_0(\infty)=U_0(0)=1-a,$$

from which the coefficient in V_0 is determined to be $c = 1-a$. Therefore we have an approximate solution valid in the whole interval $0\leqslant x\leqslant 1$:

$$u\cong U_0(x)+V_0(\xi)-(1-a)$$
$$=ax+(1-a)(1-e^{-x/\varepsilon}).$$

C. Method of Steepest Descent

If an analytic function f(z) of a complex variable z is expressed in terms of an analytic function g(t) by an integral

$$f(z)=\int_C \exp\{zg(t)\}\,dt, \tag{13}$$

where C is a curve on the complex t-plane, then an asymptotic expression for f(z) for large values of $|z|$ can be derived by the **method of steepest descent** (the **saddle-point method).** The idea of the method may be traced back to Riemann, and various asymptotic expressions for cylindrical functions were obtained by Debye (1909) [5].

The point at which the first derivative of g(t) in the integral (13) vanishes is called a **saddle point.** The function g(t) is expanded in a power series near this point as

$$g(t)=g(t_0)+\frac{1}{2}g''(t_0)(t-t_0)^2+\ldots$$

We have an inequality

$$zg''(t_0)(t-t_0)^2\leqslant 0$$

along the line

$$L:\ \arg(t-t_0)=\frac{\pi}{2}-\frac{1}{2}\arg\{zg''(t_0)\}. \tag{14}$$

Therefore the absolute value of the integrand exp{zg(t)} reaches a maximum at t_0 on the line L and decreases along it more rapidly than along any other direction. Hence if we deform the integration path in such a way that it passes through the saddle point t_0 in the direction of steepest descent, the value of the integrand is practically zero on the new path except very near to t_0 when $|z|$ is large, whereas the value of the integral remains the same. Therefore we can get an asymptotic expression of f(z) for large $|z|$ by truncating the Taylor series for $g(t)$ up to the second term and taking the line L as the integration path:

$$f(z)\approx\exp\{zg(t_0)\}\int_L \exp\left\{\frac{1}{2}zg''(t_0)(t-t_0)^2\right\}dt$$
$$=\sqrt{\frac{2\pi}{-zg''(t_0)}}\exp\{zg(t_0)\}.$$

As an example, **Stirling's formula**

$$n!\approx\sqrt{2\pi n}\;n^ne^{-n}$$

for a large positive integer is derived if we start with the formula

$$n!=\int_0^\infty e^{-t}t^n\,dt=\int_0^\infty e^{-ns}(ns)^n\,d(ns)$$
$$=n^{n+1}\int_0^\infty \exp\{n(\log s-s)\}\,ds$$

and make use of the method of steepest descent.

References

[1] H. Jeffreys, On certain approximate solutions of linear differential equations of the second order, London Math. Soc., (2) 23 (1924), 428-436.
[2] M. J. Lighthill, A technique for rendering approximate solutions to physical problems uniformly valid, Phil. Mag., (7) 40 (1949), 117991201.
[3] W. Wasow, On the convergence of an approximation method of M. J. Lighthill, J. Rational Mech. Anal., 4 (1955), 751–767.
[4] H. S. Tsien, The Poincaré-Lighthill-Kuo method, Advances in Applied Mechanics, Academic Press, 1956, vol. 4, 281–349.
[5] P. Debye, Naherungsformeln für die Zylinderfunktionen für grosse Werte des Arguments und unbeschrankt veranderliche Werte des Index, Math. Ann., 67 (1909), 5355558.
[6] P. M. Morse and H. Feshback, Methods of theoretical physics, McGraw-Hill, 1953.
[7] H. Jeffreys and B. S. Jeffreys, Methods of mathematical physics, Cambridge Univ. Press, third edition, 1962.
[8] J. D. Cole, Perturbation methods in applied mathematics, Blaisdell, 1968.

[9] A. H. Nayfeh, Perturbation methods, Wiley, 1973.
[10] A. Erdélyi, Asymptotic expansions, Dover, 1956.

26 (XX1.4) Arab Mathematics

The role of the Arabs in cultural history has been partly that of cultural transmitter. Between the 7th and 13th centuries, they established a religious empire that extended from India to Spain; later it was divided into the eastern and western empires. The caliphs of these empires encouraged research in the sciences, so the capitals Baghdad and Cordova became centers of culture where scholars from different countries gathered.

Arabic scholarship is sometimes called the stepfather of European culture. During the 13th Century, Alphonso X (1252–1284) invited Islamic and Hebrew scholars to the Spanish court to translate their writings on algebra, medicine, and astronomy into Spanish. This accomplishment earned him the title of Alphonso the Wise.

The first contact between Greek and Indian mathematics took place in Baghdad under Caliph Al-Mansūr (754-775); Euclid's *Elements* was introduced by way of the Byzantine Empire, while Brahmagupta's *Brahmasphuta-siddhânta* came directly from India. Many mathematical texts found in the Eastern Roman Empire and Syria, including some Greek works, were translated into Arabic. Though it is difficult to discern essential scientific advances in Arabian works, the diffusion of these translations was instrumental in the development of European mathematics.

The Arabs did not use written numerals until Mohammed's time (570–632). Signs representing numbers had been introduced into Arabia when its influence encompassed Egypt and Greece. Indian numerals were imported with Brahmagupta's book and became our present **Arabic numerals** after a series of modifications.

Among all the branches of Arab mathematics, algebra was the most advanced. It started with Alkwarizmi's (820) Al *Gebr W'al Muquabala*, the origin of the word "algebra." It was the first mathematical book written in Arabic. Its content was essentially a variety of methods of solving algebraic equations. Al *gebr* means "transposition of negative terms on one side of the equation to the other side and changing their signs." and *al muquabala* means "simplification of the equation by gathering similar terms." For example, quadratic equations can be brought by these methods into one of the following three types: $x^2=px+q$; $x^2+q=px$; $x^2+px=q$, where p, q are positive numbers. The Arabs expressed the rule of solving these equations verbally. They apparently knew that quadratic equations have two roots, but they adopted only positive roots; when the equation had two positive roots they adopted the smaller root. The proof was given geometrically. It is possible that they learned geometric proofs from the Greeks.

For the Arabs, geometry was secondary to algebra. They did not appreciate proof as seen in Euclid's *Elements.* The book on conic sections by Apollonius was also translated into Arabic, but no essential progress was made in this area. The only remarkable contribution was that of Omar Khayyám, author of *The Rubáiyát*, who applied conic sections to the solution of the cubic equation $x^3+bx=a$.

In trigonometry, Al Battani (c. 858-929) left a notable contribution. He studied *The Almagest,* the Arabic translation of Ptolemy's astronomical work. He added nothing outstanding to plane trigonometry, but obtained such formulas as $\cos a = \cos b \cos c + \sin b \sin c \cos \alpha$ for spherical triangles, which were not mentioned in *The Almagest.*

References

[1] M. Cantor, Vorlesungen über Geschichte der Mathematik I, Teubner, third edition, 1907.
[2] G. Sarton, Introduction to the history of science I. From Homer to Omar Khayyam, Carnegie Institution of Washington, 1927.

27 (V.18) Arithmetic of Associative Algebras

A. General Theory

Let $\mathfrak{g}$ be a †Dedekind domain (i.e., an integral domain in which every ideal is uniquely decomposed into a product of prime ideals), let F be the field of quotients of $\mathfrak{g}$, and let A be a †separable algebra of finite degree over F. A $\mathfrak{g}$-**lattice** a of A is a $\mathfrak{g}$-submodule of A that is finitely generated over $\mathfrak{g}$ and satisfies $A = Fa.$ If a subring o of A is a g-lattice containing $\mathfrak{g}$, then o is called an **order.** A **maximal order** is an order that is not contained in any other order. A maximal order always exists although

it may not be unique; in particular, if A is commutative, it has only one maximal order. If we put $\mathfrak{o}_l=\{x\in A\,|\,x\mathfrak{a}\subset\mathfrak{a}\}$ $(\mathfrak{o}_r=\{x\in A\,|\,\mathfrak{a}x\subset\mathfrak{a}\})$ for a g-lattice a of A, then $\mathfrak{o}_l$ $(\mathfrak{o}_r)$ is an order of *A* and is called the **left (right) order** of a. If the left order of a is maximal, then so is the right order; the converse is also true. If $\mathfrak{o}_l$ and $\mathfrak{o}_r$ are maximal, **we** call a a **normal g-lattice.** To describe the same situation we say that a is a **left $\mathfrak{o}_l$-ideal** or a **right $\mathfrak{o}_r$-ideal**. If $\mathfrak{o}_l=\mathfrak{o}_r=\mathfrak{o}$, then we say that a is a **two-sided o-ideal.** A lattice a with $\mathfrak{a}\subset\mathfrak{o}_l$ (or equivalently with $\mathfrak{a}\subset\mathfrak{o}_r$) is called an **integral g-lattice** or **integral left (right) ideal.** The product ab of two normal g-lattices a, b is called a **proper product** if the right order of a coincides with the left order of b. The proper product defines the structure of a †groupoid on the set of all normal g-lattices of *A*. In particular, the inverse $\mathfrak{a}^{-1}$ of a normal g-lattice a satisfying $\mathfrak{a}\mathfrak{a}^{-1}=\mathfrak{o}_l$, $\mathfrak{a}^{-1}\mathfrak{a}=\mathfrak{o}_r$ is given by $\mathfrak{a}^{-1}=\{x\in A\,|\,x\mathfrak{a}\subset\mathfrak{o}_r\}=\{x\in A\,|\,\mathfrak{a}x\subset\mathfrak{o}_l\}$. For a fixed maximal order o, a maximal **integral two-sided o-ideal** $\mathfrak{p}$ different from **o** is called a **prime ideal** of o. If p is prime, then $\mathfrak{o}/\mathfrak{p}$ is the matrix algebra of degree κ over a division algebra, and κ is called the **capacity** of the prime ideal $\mathfrak{p}$. The set of all two-sided o-ideals forms a multiplicative group, of which prime ideals are independent generators.

B. Maximal Orders of a Simple Ring

In the rest of this article, *A* is a †simple ring, F is the †center of *A*, and o is a maximal order of *A*. The prime ideals q of o and the prime ideals $\mathfrak{p}$ of $\mathfrak{g}$ are in one-to-one correspondence by the relation $\mathfrak{q}\cap\mathfrak{g}=\mathfrak{p}$; $\mathfrak{o}/\mathfrak{q}$ is a simple algebra over $\mathfrak{g}/\mathfrak{p}$, and $\mathfrak{q}^e=\mathfrak{p}\mathfrak{o}$ for some natural number **e.** The **different** $\mathfrak{d}$ of o is defined by $\mathfrak{d}^{-1}=\{x\in A\,|\,Tr(x\mathfrak{o})\subset\mathfrak{g}\}$. (*Tr* is the †reduced trace from A to *F*; $\mathfrak{d}$ is an integral two-sided o-ideal, and is divisible by $\mathfrak{q}^{e-1}$ if $\mathfrak{q}^e=\mathfrak{p}\mathfrak{o}$.) For q to divide $\mathfrak{d}$, it is necessary and sufficient that either $e>1$ or $\mathfrak{o}/\mathfrak{q}$ not be separable over $\mathfrak{g}/\mathfrak{p}$. In particular, if *A* is a †total matrix algebra over *F*, then $\mathfrak{d}=\mathfrak{o}$, $\mathfrak{q}=\mathfrak{p}\mathfrak{o}$. The ideal of $\mathfrak{g}$ generated by the †reduced norms (to F) of the elements in a normal g-lattice a of *A* is denoted by $N_{A/F}(\mathfrak{a})$. If ab is a proper product, then we have $N_{A/F}(\mathfrak{a}\mathfrak{b})=N_{A/F}(\mathfrak{a})N_{A/F}(\mathfrak{b})$, where $N_{A/F}(\mathfrak{d})$ does not depend on the choice of o; this is called the **discriminant** of *A*. If $[A:F]=n^2$ and $\mathfrak{q}^e=\mathfrak{p}\mathfrak{o}$, then $N_{A/F}(\mathfrak{q})=\mathfrak{p}^f$, $ef=n$.

C. Simple Rings over a Local Field

Let *F* be a field that is complete with respect to a †discrete valuation whose field of residue classes is finite. Let $\mathfrak{g}$ be the †valuation ring of *F*, p be the maximal ideal of $\mathfrak{g}$, and *A* be a simple algebra with *F* as its center. If *A* is a †division algebra and $[A:F]=n^2$, then *A* has only one maximal order o, and o has only one prime ideal q such that $\mathfrak{q}^n=\mathfrak{p}\mathfrak{o}$; $\mathfrak{o}/\mathfrak{q}$ is an extension of degree *n* of $\mathfrak{g}/\mathfrak{p}$. If *A* is not necessarily a division algebra, the relation $\xi\mathfrak{o}\xi^{-1}=\mathfrak{o}'$ with an element ξ of *A* holds between two maximal orders o and o' of *A*. Furthermore, for any left (right) o-ideal a, there exists an element α such that $\mathfrak{a}=\mathfrak{o}\alpha$ $(\mathfrak{a}=\alpha\mathfrak{o})$. By using the notation of †cyclic algebra, we can express *A* in the form (K_n,σ,π^r). Here K_n is the unramified extension of degree n over *F*, σ is the †Frobenius substitution of K_n/F, π is a prime element of *F*, and $0\leqslant r<n$. The element of $\mathbf{Q}/\mathbf{Z}$ determined by r/n (mod **Z**) is denoted by $\{A\}$ and is called the †Hasse invariant of the algebra class containing *A*. The mapping $A\to\{A\}$ gives an isomorphism of the †Brauer group of *F* onto the additive group $\mathbf{Q}/\mathbf{Z}$. If *M* is an extension of *F* of finite degree, then $\{A^M\}=[M:F]\{A\}$ holds for the algebra A^M obtained from *A* by scalar extension (→ 29 Associative Algebras).

D. Simple Rings over an Algebraic Number Field

Let *F* be an algebraic number field of finite degree, and let *A* be a simple algebra with center F. Then *A* is a cyclic algebra and is isomorphic to a total matrix algebra over a division algebra *D*. The order *n* of the algebra class of *A* over *F* is determined by $n^2=[D:F]$ (H. Hasse, R. Brauer, and E. Noether).

Denote by $F_\mathfrak{p}$ the completion of *F* with respect to a †prime divisor p of *F*, and let $A_\mathfrak{p}$ be the algebra obtained from *A* by the scalar extension $F_\mathfrak{p}$ over F. For a finite prime divisor $\mathfrak{p}$, the meaning of $\{A_\mathfrak{p}\}$ is as before; for an infinite prime divisor $\mathfrak{p}$, put $\{A_\mathfrak{p}\}=0$ or $1/2$ (mod **Z)** according as $A_\mathfrak{p}$ is a total matrix algebra over $F_\mathfrak{p}$ or not. Furthermore, define the subgroup $J_\mathfrak{p}$ of $\mathbf{Q}/\mathbf{Z}$ by

$J_\mathfrak{p}=\mathbf{Q}/\mathbf{Z}$,	p a finite prime divisor,
$=\{0,1/2\ (\mathrm{mod}\ \mathbf{Z})\}$	p a real infinite prime divisor,
$=\{0\}$,	p a complex infinite prime divisor.

Now let *J* be the subgroup of the direct product $\prod_\mathfrak{p}J_\mathfrak{p}$ consisting of all elements of the form $(\alpha_\mathfrak{p})$ $(\alpha_\mathfrak{p}\in J_\mathfrak{p})$ such that $\alpha_\mathfrak{p}=0$ except for a finite number of prime divisors and $\sum_\mathfrak{p}\alpha_\mathfrak{p}=0$. Then $A\to(\{A_\mathfrak{p}\})$ gives rise to an isomorphism of the Brauer group over *F* onto J (Hasse). Each $\{A_\mathfrak{p}\}$ is called the †p-invariant of *A*. In particular, *A* is a total matrix algebra over *F* if

and only if A_p is a total matrix algebra over F_p for all p (→ 29 Associative Algebras). These theorems are closely related to †class field theory.

If o is a maximal order of A and a, b are left o-ideals, then $\mathfrak{a}\xi$ = b with an element ξ of A defines an equivalence relation between a and b. The number of equivalence classes of left ideals with respect to this equivalence is called the **class number** of *A;* it is independent of the choice of o and is equal to the class number defined by using right ideals. The product of all real infinite prime divisors p with $\{A_p\} = 1/2$ is denoted by P_∞. If P_∞ is the product of all infinite prime divisors and [A : *F]* = 4, then *A* is called a **totally definite quaternion algebra.**

If o is a maximal order of *A* and *A* is not a totally definite quaternion algebra, then $\mathfrak{a} \rightarrow N_{A/F}(\mathfrak{a})$ gives a one-to-one correspondence between the classes of left o-ideals and the congruence classes of ideals of *F* modulo P_∞ (→ 14 Algebraic Number Fields H) (Eichler's theorem).

In particular, if *A* is a total matrix algebra over *F*, then the class number of *A* is equal to the class number of F. The class number of a totally definite quaternion algebra was determined by M. Eichler by using the zeta function (→ Section F) of *A* [4].

Let o be a maximal order of A, a be an integral two-sided o-ideal, *b* be an integer of *F,* and ξ be an element of o. Furthermore, assume $b \equiv 1 \pmod{P_\infty}$, $N_{A/F}(\xi) \equiv b \pmod{\mathfrak{a} \cap F}$ (†multiplicative congruence). Then there exists an element β of o such that $N_{A/F}(\beta) = b$, $\beta \equiv \xi \pmod{\mathfrak{a}}$ ($N_{A/F}$ is the reduced norm), provided that *A* is not a totally definite quaternion algebra [5]. This theorem, which is called **Eichler's approximation theorem,** is widely applicable; e.g., it yields the previous theorem on the class number and can be generalized to the case of semisimple †algebraic groups (→ 13 Algebraic Groups).

E. Algebras over a Function Field

The Hasse-Brauer-Noether and Hasse theorems also hold for †normal simple algebras over a †field of algebraic functions of one variable over a finite field. On the other hand, a normal simple algebra over a field *K* of algebraic functions of one variable over an algebraically closed field is a total matrix algebra over *K* **(Tsen's theorem).**

F. Other Notions

Adeles and ideles for a simple algebra *A* over an algebraic number field of finite degree can be introduced as in the case of number fields (→ 6 Adeles and Ideles). If N(a) stands for the number of elements in $\mathfrak{o}/\mathfrak{a}$, where o is a maximal order of *A* and a is an integral left o-ideal, then the **zeta function of the simple algebra** *A,* called the **Hey zeta function**, is defined by $\zeta_A(s) = \sum N(\mathfrak{a})^{-s}$ (the sum over all integral left o-ideals). This function has properties similar to those of Dedekind zeta functions (→ 450 Zeta Functions L). Let p be an infinite prime divisor of the center *F* of *A,* and let G_p be the group of elements in A_p with the reduced norm 1. Put $G = \prod_p G_p$ (the product over all infinite prime divisors of *F).* Then the group Γ of units with the reduced norm 1 in o is naturally regarded as a subgroup of G. To be more precise, Γ is a discrete subgroup of G, the volume of G/Γ is finite with respect to an invariant measure, and G/Γ is compact if and only if *A* is a division algebra (→ 122 Discontinuous Groups). This result can be viewed as a special case of more general facts about semisimple algebraic groups (→ 13 Algebraic Groups). If *K* is a †maximal compact subgroup of G, then Γ gives rise to a †discontinuous group operating on the homogeneous space G/K, and we obtain †automorphic forms with respect to Γ. The case where *A* is a †quaternion algebra has been studied extensively (→ 32 Automorphic Functions).

References

[1] C. Chevalley, L'arithmétique dans les algèbres de matrices, Actualités Sci. Ind., Hermann, 1936.

[2] M. Deuring, Algebren, Erg. Math., Springer, second edition, 1968.

[3] M. Eichler, Über die Idealklassenzahl hyperkomplexer Systeme, Math. Z., 43 (1938), 481–494.

[4] M. Eichler, Über die Idealklassenzahl total definiter Quaternionen Algebren, Math. Z., 43 (1938), 102–109.

[5] M. Eichler, Allgemeine Kongruenzklasseneinteilungen der Ideale einfacher Algebren über algebraischen Zahlkorpern und ihre *L*-Reihen, J. Reine Angew. Math., 179 (1938), 227–251.

[6] N. Jacobson, The theory of rings, Amer. Math. Soc. Math. Surveys, 1943.

[7] C. L. Siegel, Discontinuous groups, Ann. Math., (2) 44 (1943), 674-689. (Gesammelte Abhandlungen, Springer, 1966, vol. 2, 390-405.)

[8] A. Weil, Adeles and algebraic groups, Lecture notes, Institute for Advanced Study, Princeton, 1961.

28 (XXI.13)
Artin, Emil

Emil Artin (March 3, 1898–December 20, 1962) was born in Vienna. After he studied at the Universities of Vienna, Leipzig, and Göttingen, he taught at the University of Hamburg from 1923 to 1937. In 1937 he left Germany under the Nazi regime for America, where he taught at Notre Dame, Indiana, and Princeton universities. He returned to Germany in 1948 and taught again at the University of Hamburg until his fatal heart attack.

In his thesis (1923), he proved affirmatively the †Riemann hypothesis in function fields over finite constant fields in the hyperelliptic case, and conjectured that this should be valid in general. This was eventually verified by A. Weil in 1941. Also in 1923, he introduced the †L-function for Galois extension and was led to the †general law of reciprocity, which he stated as a conjecture and proved four years later, thus bringing †class field theory to completion. Around the same period, he established in collaboration with 0. Schreier the theory of †formally real fields and solved the 17th †problem of Hilbert. He also initiated the theory of †braids, which he later developed with F. Bohnenblust in the 1940s. In view of these and other ingenious works as well as his inspiring teaching in all areas of mathematics, Artin is considered to have been one of the most influential mathematicians of this Century.

References

[1] S. Lang and J. T. Tate (eds.), The collected papers of Emil Artin, Addison-Wesley, 1965.
[2] H. Cartan, Emil Artin, Hamb. Abh., 29 (1965).

29 (III.10)
Associative Algebras

A. Fundamental Concepts

Let K be a commutative ring with unity element 1 (- 368 Rings A), and let A be a ring which is a †unitary K-module (- 277 Modules). Such a ring A is called an **associative algebra over** K (or simply **algebra over** K) if it satisfies the condition $\lambda(ab)=(\lambda a)b=a(\lambda b)$ $(\lambda\in K;\ a,b\in A)$. An (associative) algebra A over K is often written A/K, and K is called the **coefficient ring** (or **ground ring**) of the algebra $A=A/K$. In particular, if K is a †field, then it is called the **coefficient field** (or **ground field**) of A. Algebras over fields have been studied in detail. Notions such as **zero algebra, unitary algebra, commutative algebra, (semi) simple algebra, and division algebra** are replicas of the respective ones for rings (- 368 Rings). Considering both structures as rings and as K-modules, homomorphisms and isomorphisms are defined in a natural manner, and are called **algebra homomorphism** and **algebra isomorphism,** respectively. In this connection, **subalgebra, quotient algebra** (or **residue class algebra),** and **direct product** of algebras are also defined as in the case of rings.

An †**ideal** of an algebra A is defined as an ideal of the ring A, which is at the same time a submodule of the module A over the coefficient ring. The †radical of an algebra A considered as a ring is then an ideal of A in this sense. In fact, the existence of a unity in an algebra A implies that any ideal of A considered as a ring is necessarily an ideal of the algebra A.

In the rest of this article, we assume that all rings have a unity element and that all homomorphisms are unitary. Hence, when we consider subalgebras of an algebra A, we require that they share the unity element with A. If e is the unity element of an algebra A over K, then the mapping $\lambda\to\lambda e=\lambda'$ $(\lambda\in K)$ is a homomorphism $K\to A$ whose image Ke is contained in the †center of A, and the scalar multiplication λa is equal to the ring multiplication $\lambda' a$ $(\lambda\in K,$ $a\in A)$. Conversely, given a homomorphism of K into a ring A whose image is contained in the center of A, we can regard A as an algebra over K in an obvious way. Hence we are given an algebra A over K if and only if there exists a pair (A,ρ) of a ring A and a homomorphism $\rho: K\to A$ whose image is contained in the center of A. There exists a uniquely determined (unitary) homomorphism of the ring **Z** of rational integers into any ring; hence any ring can be regarded as an algebra over **Z.** If the coefficient ring K of a nonzero algebra A is a field, then K can be regarded as a subfield contained in the center of A, and the unity element 1 of K coincides with the unity element of A.

Let A and B be algebras over K. Then the †tensor product $A\otimes_K B$ of K-modules is an algebra over K under the multiplication $(a\otimes b)(a'\otimes b')=aa'\otimes bb'$ $(a,a'\in A,\ b,b'\in B)$. This algebra is called the **tensor product** of algebras A and B. Moreover, the mapping $a\to a\otimes 1$ (resp. $b\to 1\otimes b$) $(a\in A,\ b\in B)$ gives an algebra homomorphism $A\to A\otimes_K B$ (resp. $B\to$

$A \otimes_K B$, which is called the **canonical homomorphism.** In particular, if A is a commutative algebra, then $A \otimes_K B$ can be regarded as an algebra over A by this canonical homomorphism; and in this case $A \otimes_K B$ is called the algebra obtained by **extension of the coefficient ring** of B to A (or a **scalar extension** of B by A), and is often denoted by B^A. The algebra $K \otimes_K B$ is canonically isomorphic to B. Furthermore, $(A \otimes_K B) \otimes_K C$ and $A \otimes_K (B \otimes_K C)$ are canonically isomorphic and are written $A \otimes B \otimes C$. Let A and B be commutative algebras over K. Then, for any commutative algebra C and homomorphisms $\alpha : A \rightarrow C$, $\beta : B \rightarrow C$, there exists one and only one homomorphism $\gamma : A \otimes_K B \rightarrow C$ such that $\alpha(a) = \gamma(a \otimes 1)$, $\beta(b) = \gamma(1 \otimes b)$ $(a \in A, b \in B)$. This property characterizes the tensor product $A \otimes_K B$ of commutative algebras A and B. In this sense, $A \otimes_K B$ is sometimes called the **coproduct** of A and B (→ 52 Categories and Functors E).

B. Examples of Associative Algebras

As we mentioned, any ring can be regarded as an algebra over the ring **Z** of rational integers. But it is often useful to deal with algebras over "large" or more "efficient" coefficient rings. For instance, we have many rings which are algebras over a commutative ring K, such as the †ring of polynomials, the †ring of formal power series in n variables with coefficients in K, the †endomorphism ring of a K-module, and the †full matrix ring of degree n over K (→ 368 Rings C). There are other important classes of algebras, such as (semi) group algebras, Hecke algebras, and crossed-product algebras, which will be explained later. These algebras are defined by a canonical basis connected directly with a (semi) group structure. On the other hand, the †tensor algebra and the †exterior algebra of linear spaces and the †Clifford algebra associated with a given quadratic form are also important (→ 61 Clifford Algebras, 256 Linear Spaces).

The most frequently used example of a division algebra is the **quaternion field H** (often called **Hamilton's quaternion algebra,** W. R. Hamilton, 1858). This is a 4-dimensional linear space over the real number field **R** with basis $\{1, i, j, k\}$, with the following laws of multiplication: 1 is a unity element, $i^2 = j^2 = k^2 = -1$, $ij = -ji = k$, $jk = -kj = i$, and $ki = -ik = j$. An element of **H** is called a **quaternion.** The only finite-dimensional division algebras over the real number field **R** are the real number field **R,** the complex number field C, and the quaternion field **H.**

C. Group Algebras and Hecke Algebras

Let K be a commutative ring and $K^{(G)}$ be the direct sum $\sum_{s \in G} K_s$ of modules K_s, where each K_s is isomorphic to K with a group G as the index set (→ 277 Modules F). The elements of $K^{(G)}$ are the families $(\lambda_s)_{s \in G}$ of elements of K whose components are all zero except for a finite number of them. Let $\{u_s\}_{s \in G}$ be the canonical basis of $K^{(G)}$, namely, u_s is an element of $K^{(G)}$ of which the sth component is 1 and the others are 0. The module $K^{(G)}$ has the structure of an algebra over K, where the law of multiplication is determined by $u_s u_t = u_{st}$ $(s, t \in G)$. This algebra is called the **group algebra** of G over K. The product $\lambda * \mu$ of elements $\lambda = (\lambda_s)$ and $\mu = (\mu_s)$ of $K^{(G)}$ is then given by

$$(\lambda * \mu)_s = \sum_{s = rl} \lambda_r \mu_l, \quad s \in G. \tag{1}$$

Each basis element u_s is often identified with the group element s, and in this manner, the group G is regarded as a basis of $K^{(G)}$, which is usually written KG or $K[G]$.

In this definition, the group G can be replaced by a semigroup G, and then the algebra $K^{(G)}$ is called a **semigroup algebra.** As an example, let N be the additive semigroup of all nonnegative rational integers and $\mathbf{N}^n$ be the direct product of n copies of N. If we denote the elements $(i_1, \ldots, i_n)$ of $\mathbf{N}^n$ by $X_1^{i_1} \ldots X_n^{i_n}$ and use multiplication instead of addition, then the semigroup algebra of $\mathbf{N}^n$ over K is exactly the †ring of polynomials $K[X_1, \ldots, X_n]$. On the other hand, if G is a semigroup, then even when it is infinite, it may occur that for any $s \in G$, there exists only a finite number of pairs (r, l) of elements of G such that $s = rl$. In this case, formula (1) also defines the law of multiplication on the †Cartesian product K^G. This algebra is called a **large semigroup algebra** and contains $K^{(G)}$ as a subalgebra. In particular, the large semigroup algebra of $\mathbf{N}^n$ is exactly the †ring of formal power series $K[[X_1, \ldots, X_n]]$.

Let H be a subgroup of a group G, and assume that the index of $H \cap sHs^{-1}$ in H is finite for any $s \in G$. This assumption is equivalent to the condition that any double coset of G by H is a union of a finite number of left as well as right cosets. Let $H \backslash G$, G/H, and $H \backslash G/H$ be the set of all †right cosets, †left cosets, and †double cosets, respectively. Then each element of the direct sum $K^{(H \backslash G)}$, $K^{(G/H)}$, or $K^{(H \backslash G/H)}$ can be regarded as a function defined on G taking a constant value on each right, left, or double coset, respectively. Conversely, any function defined on G can be regarded as an element of $K^{(H \backslash G)}$, $K^{(G/H)}$, or

$K^{(H\backslash G/H)}$ if it takes a constant value on each left, right, or double coset, respectively, and if it vanishes everywhere except on a finite number of cosets of respective type. Under such identification, let λ_s, λ_l, λ_r, and λ_t denote the values of a function $\lambda: G \rightarrow K$ on $s \in G$, $l \in H\backslash G$, $r \in G/H$, and $t \in H\backslash G/H$, respectively. For any $\lambda, \mu \in K^{(H\backslash G/H)}$ we define a function $\lambda * \mu$ by

$$(\lambda * \mu)_s = \sum_{s \in rl} \lambda_r \mu_l, \quad s \in G, \tag{2}$$

where the right-hand side is the sum taken over all pairs (r, l) such that $s \in rl$, $r \in G/H$, $l \in H\backslash G$. It can be shown that this sum is a finite sum and that $\lambda * \mu \in K^{(H\backslash G/H)}$. Hence the module $K^{(H\backslash G/H)}$ has the structure of an algebra over K, which is called the **Hecke algebra** of (G, H) over K and often denoted by $\mathscr{H}_K(G, H)$. If $H = \{e\}$, the $\mathscr{H}_K(G, H)$ is exactly the group algebra of G. In general, $\mathscr{H}_K(G, H)$ can be regarded as an algebra obtained by extending the coefficient ring of $\mathscr{H}_{\mathbf{Z}}(G, H)$ to K. Furthermore, $K^{(H\backslash G)}$ ($K^{(G/H)}$) can be regarded as a right (left) module over the group algebra $K^{(G)}$, and the endomorphism ring $\mathscr{E}_{K^{(G)}}(K^{(H\backslash G)})$ (resp. $\mathscr{E}_{K^{(G)}}(K^{(G/H)})$) is canonically isomorphic (anti-isomorphic) to the Hecke algebra $K^{(H\backslash G/H)}$.

D. General Crossed Product

Let G be a group that operates on a commutative ring L, and denote the operation by $(s, \lambda) \rightarrow s(\lambda)$ $(s \in G, \lambda \in L)$; thus for any $s \in G$ the mapping $\lambda \rightarrow s(\lambda)$ $(\lambda \in L)$ is an automorphism of L satisfying $s(t(\lambda)) = st(\lambda)$ $(s, t \in G)$. For any λ, $\mu \in L^{(G)}$, we define the product $\lambda * \mu \in L^{(G)}$ by

$$(\lambda * \mu)_s = \sum_{s=rl} \lambda_r r(\mu_l) f(r, l), \quad s \in G, \tag{3}$$

where $\{f(r, l)\}_{r, l \in G}$ is a given family of elements of L. If this family satisfies the equations

$$f(s, r) f(sr, l) = s(f(r, l)) f(s, rl), \quad s, r, l \in G,$$

then $L^{(G)}$ forms a ring. In terms of the canonical basis $\{u_s\}_{s \in G}$, this ring structure is defined by the formulas $u_r u_l = f(r, l) u_{rl}$, $u_s \lambda = s(\lambda) u_s$ $(\lambda \in L)$. If K is the subring of L consisting of all $\lambda \in L$ such that $s(\lambda) = \lambda$ $(s \in G)$, then the ring $L^{(G)}$ is an algebra over K, called the **crossed product** of L and G with respect to the given operation and the given **factor set** f of G. In a narrower sense of the term, we consider only the case when L is a field, G is a finite group, $f(r, l) \neq 0$, and G operates on L faithfully. In this case, G can be identified with the †Galois group of a finite †Galois extension L/K, and so the crossed product is written $(L/K, f)$. This is a †central simple algebra over K (- Sections E, F).

If the operation of G in the abovementioned general crossed product is trivial, namely, $L = K$, the crossed product of K and G is called an **algebra extension** of G over K with respect to f. Usually, we assume that K is a field and $f(r, l) \neq 0$. If $f(r, l) = 1$, the algebra extension is the group algebra. If G is a finite group whose order is prime to the †characteristic of K, then any algebra extension of G over K (in particular, the group algebra) is always †semisimple and †separable. If G is a finite Abelian group and fis a factor set of G, then a bihomomorphism φ: G x $G \rightarrow K^*$ (a mapping which is a homomorphism in each variable) is defined by $\varphi(s, t) = f(s, t) \cdot f(t, s)^{-1}$. An algebra extension of G over K with respect to f is a central simple algebra if and only if φ is nondegenerate. In particular, if G is a direct product of n copies of a group of order 2, then by choosing n elements $s_1, s_2, \ldots, s_n$ of G suitably, any element of G can be uniquely expressed as $s_a s_b \ldots s_z$, where $\{a, b, \ldots, z\}$ is a subset of $\{1, \ldots, n\}$ in the natural order. Hence, if we take $u_{(a, \ldots, z)} = u_a u_b \ldots u_z$ as a basis of an algebra extension, then any factor set f is determined by $f(s_i, s_j) = \lambda_{ij}$, where $\lambda_{ij} = 1$ $(i < j)$, $\lambda_{ij} = \pm 1$ $(i > j)$, and λ_{ii} are arbitrary. When $\lambda_{ij} = -1$ $(i > j)$, the corresponding algebra extension is a †Clifford algebra. Furthermore, if $\lambda_{ii} = 0$, then it is a †Grassmann algebra. If $\lambda_{ii} \neq 0$ and n is even, then it is a central simple algebra whenever the characteristic of K is not 2 (- 61 Clifford Algebras).

Now let K be the real number field $\mathbf{R}$ and A, be the Clifford algebra with $\lambda_{ii} =$ 1. Then A, is the quaternion field. Elements of A, are called **sedenions,** and are important in †spinor theory and †Dirac's equation. In general, let K be an arbitrary field whose characteristic is not 2 and put $n = 2$, $\lambda_{12} = 1$, $\lambda_{21} = -1$, $\lambda_{11} = \lambda \neq 0$, and $\lambda_{22} = \mu \neq 0$, in the previous notation. Then the corresponding central simple algebra Q is called a **(generalized) quaternion algebra.** Thus Q has a basis $\{1, u, v, w\}$ satisfying the following laws: 1 is the unity element, $w = uv = -vu$, $u^2 = \lambda$, and $v^2 = \mu$ $(\lambda, \mu \in K)$. Any central simple algebra of dimension 4 is isomorphic to a certain quaternion algebra. (In particular, if $K = \mathbf{R}$ and $\lambda = \mu = -1$, then Q coincides with the quaternion field **H.**) For any element $x = \alpha + \beta u + \gamma v + \delta w$ of Q, the element $\bar{x} = \alpha - \beta u - \gamma v - \delta w$ is said to be **conjugate** to x, and $N(x) = x\bar{x} \in K$ is called the **norm** of x. An element x of Q is invertible in Q if and only if $N(x) \neq 0$.

E. Finite-Dimensional Associative Algebras over a Field

For the rest of this article, we assume that the algebras considered are unitary and finite

dimensional over a field K. Then by general properties of left (right) †Artinian rings, any associative algebra A has the following structure: the †radical N of A is the greatest †nilpotent ideal, and the quotient algebra $A/N = \bar{A}$ is †semisimple and decomposed into a direct sum of ideals which are simple algebras:

$$\bar{A} = \bar{A}_1 + \ldots + \bar{A}_n.$$

Each simple component $\bar{A}_i$ is a full matrix ring of degree r_i over a certain division algebra D_i, and $\bar{A}_i$ is decomposed into a direct sum of r_i minimal left ideals which are mutually A-isomorphic:

$$\bar{A}_i = \bar{A}\bar{e}_i^{(1)} + \ldots + \bar{A}\bar{e}_i^{(r_i)}, \quad 1 \leqslant i \leqslant n,$$

where $\bar{e}_i^{(1)}, \ldots, \bar{e}_i^{(r_i)}$ are orthogonal †idempotent elements of $\bar{A}_i$ whose sum is equal to the unity element of $\bar{A}_i$. On the other hand,

$$\bar{A}_i = \bar{e}_i^{(1)}\bar{A} + \ldots + \bar{e}_i^{(r_i)}\bar{A}$$

gives a decomposition of A_i into the direct sum of minimal right ideals that are mutually A-isomorphic. Moreover, we can choose an idempotent element $e_i^{(s)}$ of A from each residue class $\bar{e}_i^{(s)}$ such that $\{e_i^{(s)}\}$ forms a system of orthogonal idempotent elements, whose sum is equal to the unity element 1 of A, and

$$A = \sum_{i=1}^{n} \sum_{s=1}^{r_i} Ae_i^{(s)} = \sum_{i=1}^{n} \sum_{s=1}^{r_i} e_i^{(s)}A$$

gives a decomposition of A into a direct sum of †direct indecomposable left (right) ideals. Here, $Ae_i^{(s)}$ and $Ae_j^{(t)}$ ($e_i^{(s)}A$ and $e_j^{(t)}A$) are A-isomorphic if and only if $i = j$. Conversely, every decomposition of A into indecomposable ideals is obtained as above. The A-submodule $Ne_i^{(s)}$ of $Ae_i^{(s)}$ is the unique maximal proper submodule, and $Ae_i^{(s)}/Ne_i^{(s)}$ and $Ae_j^{(t)}/Ne_j^{(t)}$ are A-isomorphic if and only if $i = j$. Any simple A-module is A-isomorphic to a certain $Ae_i^{(s)}/Ne_i^{(s)}$ (- Sections H, I).

Any simple algebra A over K is isomorphic to a full matrix ring M,(D) over a certain division algebra D. This is called **Wedderburn's tbeorem.** Here n is determined uniquely by A, and D is also determined uniquely by A up to isomorphism. Moreover, the center of A is isomorphic to the center of D. If the center of A coincides with K, then A is called a **central simple algebra** (or **normal simple algebra)** over K. In this case, any isomorphism of two simple subalgebras of A can be extended to an †inner automorphism of A. Let V(B) denote the †commutor of a simple subalgebra B of A. Then V(B) is also a simple subalgebra and V(I'(B)) $= B$, $\dim A = \dim B \cdot \dim V(B)$. In particular, if B is central over K, then there is a canonical isomorphism $A \cong B \otimes_K V(B)$. If D is a central division algebra, then any maximal commutative subalgebra L of D is a field and satisfies $(\dim L)^2 = \dim D$. Moreover, there are †separable extensions over K among such L. In general, the dimension of a central simple algebra A is a square number r^2, where r is called the **degree** of A.

Two central simple algebras are said to be **similar** if they are isomorphic to full matrix rings over the same division algebra. This is an equivalence relation, and each equivalence class is called an **algebra class.** On the other hand, if A is a central simple algebra and B is a simple algebra, then $A \otimes_K B$ is a simple algebra. Moreover, if B is central, then $A \otimes_K B$ is also a central simple algebra. If A and B are similar to A' and B', respectively, then $A \otimes_K B$ is similar to $A' \otimes_K B'$. Hence, the tensor product $\otimes$ defines multiplication in the set $\mathscr{B}(K)$ of the algebra classes over K. If A is a central simple algebra and A° is an algebra anti-isomorphic to A, then A° is also central simple and $A \otimes_K A^\circ$ is isomorphic to a full matrix ring over K. This shows that $\mathscr{B}(K)$ forms a group, which is called the **Brauer group,** after R. Brauer, who introduced this concept, or the **algebra class group** over K. For a central simple algebra A, the degree of a division algebra similar to it is called the **Schur index** of A (or of the algebra class of A), and the order of the algebra class of A in the Brauer group is called the **exponent** of A. The Schur index is divisible by the exponent; and conversely, the exponent is divisible by every prime divisor of the Schur index.

F. Extensions of Coefficient Fields

Let L be an †extension field over a field K. Then for any algebra A over K, $L \otimes_K A$ can be regarded as an algebra over L. This algebra is denoted by A^L and is called an **algebra obtained by extending the coefficient field** to L. Let us denote the radical of a ring A by $\mathfrak{R}(A)$. An algebra A over K is called a **separable algebra** if it satisfies $\mathfrak{R}(A^L) = \{0\}$ for any extension field L over K. In the special case when A is an †algebraic extension field over K, A is a separable algebra if and only if every element of A (or every element of a subset which generates A) is †separable over K (- 149 Fields). A (finite-dimensional) algebra A over K is separable if and only if A is semisimple and the center of every simple component of A is a separable extension over K. If the quotient algebra $A/\mathfrak{R}(A)$ of an algebra A is separable, then there exists a subalgebra S such that $A = S + \mathfrak{R}(A)$ and $S \cap \mathfrak{R}(A) = \{0\}$, and S is uniquely determined up to inner automorphisms **(Wedderburn-Mal'tsev tbeorem).**

In order that an algebra A over K be central simple, it is necessary and sufficient that A^L be

simple for any extension field L over K. This latter statement holds if and only if A^L is isomorphic to a full matrix algebra over L for a certain extension field L over K. Such an extension field is a †splitting field of A (→ 362 Representations F). For a central simple algebra A, a (finite) extension field L of degree r over K is a splitting field of A if and only if there exists a central simple algebra B of degree r which is similar to A having a subfield which is K-isomorphic to L. In this case, r is divisible by the Schur index of A. Furthermore, A has a separable splitting field whose extension degree is equal to the Schur index of A. This shows that A has a splitting field which is a finite Galois extension field over K.

Let L be a finite †Galois extension over a field K and G be its †Galois group. If f is a factor set with respect to the operation of G on L^*, then the crossed product $(L/K,f)$ is a central simple algebra over K (→ Section D). The product fg of two factor sets f and g is also a factor set, and the set of all factor sets forms an †Abelian group. Thus $(L/K,fg)$ is similar to $(L/K,f)\otimes_K(L/K,g)$. On the other hand, factor sets f and g are said to be **associated** with each other if there exists a family $\{\lambda_s\}_{s\in G}$ of elements of L such that

$$f(r,l)=g(r,l)r(\lambda_l)\lambda_{rl}^{-1}\lambda_r, \quad r,l\in G.$$

Hence, f and g are associated with each other if and only if $(L/K,f)$ and $(L/K,g)$ are similar. Therefore, the mapping $f\rightarrow(L/K,f)$ gives a monomorphism of the group $H^2(G,L^*)$ of all associated classes of factor sets (which can be identified with the 2-dimensional †cohomology group of G with coefficients in L^*) into the Brauer group $\mathscr{B}(K)$ over K. Its image coincides with the subgroup of all algebra classes which have L as a splitting field. In particular, any algebra class is similar to the crossed product of a certain finite Galois extension L and its Galois group G (R. Brauer, E. Noether, A. Albert, K. Shoda, and others).

G. Cyclic Algebras

Let Z be a †cyclic extension field of degree n over a field K. Then the crossed product of Z and its Galois group G is called a **cyclic algebra** over K. For a fixed generator s of G and any element $\alpha\neq 0$ of K, a factor set $f(s^i,s^j)$ $(0\leqslant i,j<n)$ can be defined by $f(s^i,s^j)=1$ $(i+j<n)$ and $f(s^i,s^j)=\alpha$ $(i+j\geqslant n)$. Let (Z, s,α) denote the corresponding crossed product. Then (Z, s, a) and (Z, s,β) are similar if and only if α/β is a norm of a certain element of Z into K. On the other hand, any crossed product of Z and G is similar to a certain (Z, s, cc), and the correspondence $\alpha\rightarrow(Z,s,a)$ gives an isomorphism of $K^*/N_{Z/K}(Z^*)$ (the norm class group of Z/K) to the group consisting of the algebra classes over K which have Z as a splitting field. If K is a †p-adic field or a finite †algebraic number field, then any central simple algebra over K is isomorphic to a certain cyclic algebra. In this section, we describe this situation in detail.

Let K be a p-adic field and q the number of residue classes modulo $\mathfrak{p}$. If A is a central simple algebra over K, then its Schur index coincides with the exponent, which is called simply the **index.** A finite extension field L over K is a splitting field for A if and only if the degree of L is a multiple of the index of A. If n is the degree of A, then the field $W=K(w)$ obtained by adjoining to K a †primitive (q^n-1)st root of unity ω is a cyclic (and †unramified) extension of degree n over K, and its Galois group is generated by the automorphism σ of W determined by $\omega^\sigma=\omega^q$, $\sigma|K$ = identity. Moreover, we have $A\cong(W,\sigma,\alpha)$ for a certain $\alpha\in K^*$. Let v be the †exponential p-adic valuation $v(\alpha)$ of α. Then v/n (mod **Z**) is uniquely determined by the algebra class of A, which is called the **Hasse invariant** of A (or of the algebra class of A). By assigning to each of the algebra classes its Hasse invariant, we get an isomorphism of the Brauer group $B(K)$ over K to the group $\mathbf{Q}/\mathbf{Z}$, the additive group of the rational numbers mod **Z** (H. Hasse, 1931).

Let K be a finite algebraic number field and A be a central simple algebra over K. Let p be a (finite or infinite) †prime divisor of K and $K_\mathfrak{p}$ the †p-adic extension field over K. The algebra A, which is obtained from A by extending the coefficient field to $K_\mathfrak{p}$ is a central simple algebra over K,. Except for a finite number of $\mathfrak{p}$, A, is isomorphic to a full matrix ring over $K_\mathfrak{p}$, and A itself is isomorphic to a full matrix ring over K if and only if A, is isomorphic to a full matrix ring over $K_\mathfrak{p}$ for all p. The index $m_\mathfrak{p}$ of A, is called the p-index of A, and the Hasse invariant of A, is called the **p-invariant of** A, which is denoted by $(A/\mathfrak{p})$. If **p** is an †infinite prime divisor, then $m_\mathfrak{p}$ is equal to 1 or 2, and in each case, we define the p-invariant by setting $(A/\mathfrak{p})=0$ or $1/2$ (mod **Z**) correspondingly. The Schur index of A is the L.C.M. of the p-indices $m_\mathfrak{p}$ for all **p** and coincides with the exponent of A. This is called simply the **index** of A. On the other hand, the p-invariants satisfy $(A/\mathfrak{p})\equiv 0$ (mod **Z**) except for a finite number of $\mathfrak{p}$, and

$$\sum_{\mathfrak{p}}(A/\mathfrak{p})\equiv 0 \pmod{\mathbf{Z}}.$$

Conversely, given a rational number $\rho_\mathfrak{p}$ for each p such that (i) $\rho_\mathfrak{p}\equiv 0$ (mod **Z**) except for a finite number of p; (ii) $\rho_\mathfrak{p}\equiv 0$ (mod **Z**) if p is infinite and imaginary, $\rho_\mathfrak{p}\equiv 0$ or $1/2$ (mod **Z**) if

$\mathfrak{p}$ is infinite and real; (iii) $\sum_{\mathfrak{p}} \rho_{\mathfrak{p}} \equiv 0$ (mod **Z**), then there is a uniquely determined algebra class of central simple algebras A over K such that $(A/\mathfrak{p}) \equiv \rho_{\mathfrak{p}}$ (mod **Z**) for each $\mathfrak{p}$. In this way the structure of the Brauer group over a finite algebraic number field is completely determined (Hasse, 1933).

H. Frobenius Algebras

Let A be an algebra over a field K. A is called a **Frobenius algebra** if its †regular representation and †coregular representation (→ 362 Representations E) are similar. Thus A is a Frobenius algebra if the left A-module A and the dual module A^* of the right A-module A are isomorphic as left A-modules. Let

$$A = \sum_{i=1}^{n} \sum_{s=1}^{r_i} Ae_i^{(s)} = \sum_{i=1}^{n} \sum_{s=1}^{r_i} e_i^{(s)} A$$

be direct decompositions of A into indecomposable left (right) ideals (→ Section E). We denote $e_i^{(1)}$ by e_i. Then A is a Frobenius algebra if and only if there exists a permutation π on $1, \ldots, n$ such that (i) $Ae_i \cong (e_{\pi(i)}A)^*$; (ii) $r_i = r_{\pi(i)}$. When there exists a permutation satisfying only condition (i), A is called a **quasi-Frobenius algebra.**

For a subset S of A, the ideals $l(S) = \{a \in A \mid aS = 0\}$ and $r(S) = \{a \in A \mid Sa = 0\}$ are called the **left annihilator** and **right annihilator** of S, respectively. Then A is a quasi-Frobenius algebra if and only if (ii) $l(r(\mathfrak{l})) = \mathfrak{l}$ and $r(l(\mathfrak{r})) = \mathfrak{r}$ for any left ideal $\mathfrak{l}$ and any right ideal $\mathfrak{r}$. In general, if we are given a left (right) A-module M, we denote the right (left) A-module $\mathrm{Hom}_A(M, A)$ by $\hat{M}$. Then if A is a quasi-Frobenius algebra, there is a canonical isomorphism $\hat{\hat{M}} \cong M$, and the annihilator relation gives a one-to-one and dual correspondence between the set of the submodules of M and the set of submodules of $\hat{\hat{M}}$ (M. Hall). If a quasi-Frobenius algebra A satisfies (iv) $\dim \mathfrak{r} + \dim l(\mathfrak{r}) = \dim \mathfrak{l} + \dim r(\mathfrak{l}) = \dim A$ for any left ideal $\mathfrak{l}$ and any right ideal $\mathfrak{r}$, then A is a Frobenius algebra; the converse is also true.

A criterion for an algebra A to be a Frobenius algebra is that there is a linear form $x \rightarrow \lambda(x)$ on A such that if $\lambda(xa) = 0$ for all $x \in A$, then $a = 0$. Moreover, if λ satisfies $\lambda(xy) = \lambda(yx)$ $(x, y \in A)$, then A is called a **symmetric algebra.** For example, semisimple algebras and group algebras are symmetric algebras. If A is a symmetric algebra, then for any left (right) A-module M the right (left) A-modules $M^* = \mathrm{Hom}_K(M, K)$ and $\hat{M} = \mathrm{Hom}_A(M, A)$ are canonically A-isomorphic.

If A is a Frobenius algebra, the radical N of A satisfies $l(N) = r(N)$, and the annihilator of N is a principal left and principal right ideal; the converse is also true. For a two-sided ideal $\mathfrak{z}$ of a Frobenius algebra A, the quotient algebra $A/\mathfrak{z}$ is a Frobenius algebra if and only if $l(\mathfrak{z})$ and $r(\mathfrak{z})$ are a principal left and a principal right ideal, respectively. If A is a symmetric algebra, then any two-sided ideal $\mathfrak{z}$ satisfies $l(\mathfrak{z}) = r(\mathfrak{z})$, and $A/\mathfrak{z}$ is also a symmetric algebra if and only if $l(\mathfrak{z}) = r(\mathfrak{z})$ is a principal ideal generated by an element in the center.

Furthermore, for any extension L of the coefficient field K, A^L is a quasi-Frobenius algebra (resp. Frobenius algebra, symmetric algebra) if and only if A is (T. Nakayama, 1939, 1941). The concept of Frobenius algebras has been extended to algebras B over a ring A (F. Kasch, 1954).

I. Uniserial Algebras

In the notation of the preceding section, if each indecomposable left ideal Ae_i (right ideal e_iA) of an algebra A over K has a unique composition series, then A is called a **generalized uniserial algebra.** If an algebra A is decomposed into a direct sum of ideals which are primary rings, then it is called a **uniserial algebra.** Any left module over a generalized uniserial algebra A is decomposed into a direct sum of submodules which are A-homomorphic images of Ae_i: An algebra whose radical N is a principal left and principal right ideal is a generalized uniserial algebra. For an algebra A to be uniserial, it is necessary and sufficient that every two-sided ideal of A be a principal left and principal right ideal. Hence A is uniserial if and only if every quotient algebra of A is a Frobenius algebra. If A^L is uniserial for an extension field L of K, then A itself is uniserial. The converse, however, is not always true. If A^L is uniserial for any extension field L of K, then A is called an **absolutely uniserial algebra.** For A to be absolutely uniserial it is necessary and sufficient that its radical N be a principal ideal generated by an element in the center Z and Z be decomposed into a direct sum of simple extensions of K (i.e., ideals of the form $K[a]$) (K. Asano, G. Köthe, Nakayama, G. Azumaya).

J. Algebraic Algebras

Here, we consider general (not necessarily finite-dimensional) algebras A over a field K. We say that A is an **algebraic algebra** if every element of A is algebraic over K, i.e., every element of A is a root of a certain polynomial with coefficients in K. We say that A satisfies a **polynomial identity** $p(X_1, \ldots, X_n) = 0$ or that A is a **PI-algebra** with an identity $p(X_1, \ldots, X_n) =$

0 if there exists a nonzero (noncommutative) polynomial $p(X_1, , X_n)$ in $X_1, , X_n$ with coefficients in K such that $p(a_1, , a_n) = 0$ for all $a_i \in A$. A PI-algebra satisfies an identity which is homogeneous linear in each variable, and also an identity of the form $[x_1, , x_n]^m = 0$ (where [ ] is the sum $\sum \pm x_{i_1} \ldots x_{i_n}$ taken over all permutations of $1, 2, , n$ and $\pm$ is the sign of the permutation). An algebra is said to be **locally finite** if any finite number of elements of A generate a finite-dimensional subalgebra. For PI-algebras, an affirmative answer was found for Kurosh's problem, which asks whether an algebraic algebra is locally finite if the degree of any element α of A (i.e., dim $K[a]$) is bounded.

K. Brauer Group of a Ring

Let R be a commutative ring. An R-algebra is called a **separable algebra** if A is †projective as a two-sided A-module (- 200 Homological Algebra F). When the base ring is a field, this agrees with the classical notion of separability; and A is separable over R if and only if $A/\mathfrak{m}A$ is separable over the residue field $R/\mathfrak{m}$ for every maximal ideal m of R. **A central separable algebra** is also called an **Azumaya algebra.** If P is a finitely generated faithful projective R-module (briefly: **R-progenerator),** the endomorphism ring $\mathrm{End}_R(P)$ is an Azumaya R-algebra. Azumaya algebras A, and A, are said to be in the same class (similar) if there exist R-progenerators P_1 and P_2 such that $A_1 \otimes \mathrm{End}_R(P_1) \cong A_2 \otimes \mathrm{End}_R(P_2)$. The set of similarity classes forms an Abelian group with respect to $\otimes$. This is called the **Brauer group** $B(R)$ of R (Auslander and Goldman [10]). Every element of $B(R)$ is of finite order [12, 13].

$B(R)$ is a covariant functor from commutative rings to Abelian groups. If R is a field, $B(R)$ coincides with the classically defined one (- Section E). If R is a †Henselian local ring with the residue field k, the mapping $B(R) \rightarrow B(k)$ is an isomorphism [9]. If R is a †regular ring with the quotient field K, $B(R) \rightarrow B(K)$ is injective. If further dim $R \leqslant 2$, we have $B(R) = \bigcap_{\mathfrak{p}} B(R_{\mathfrak{p}})$, $\mathfrak{p}$ running over all primes of height 1 of R, where $R_{\mathfrak{p}}$ is the localization of R at $\mathfrak{p}$ and $B(R)$ and $B(R_{\mathfrak{p}})$ are considered as embedded in $B(K)$ [10]. As an example, putting these facts together with the structure of $B(K)$ of the algebraic number field K (→ Section G), we have the structure of $B(R)$ of the ring of integers of K: $B(R) = 0$ if K is totally imaginary, and $\cong (\mathbf{Z}/2\mathbf{Z})^{r-1}$ if K has $r(>0)$ real infinite places.

A commutative R-algebra S is called a **splitting ring** of A if the S-algebra $S \otimes A$ is isomorphic to $\mathrm{End}_S(P)$ for some S-progenerator P. We denote by $B(S/R)$ the subgroup of $B(R)$ consisting of all algebra classes split by S. Since an Azumaya algebra over a ring need not have a Galois extension as splitting ring, the description of the Brauer groups by means of Galois cohomology ceases to have full generality. Instead we have the following exact sequence of †Amitsur cohomology, assuming S is an R-progenerator (Chase and Rosenberg [11]): $0 \rightarrow H^1(S/R, U) \rightarrow \mathrm{Pic}(R) \rightarrow H^0(S/R, \mathrm{Pic}) \rightarrow H^2(S/R, U) \rightarrow B(S/R) \rightarrow H^1(S/R, \mathrm{Pic}) \rightarrow H^3(S/R, U) \rightarrow$, where we denote by $U(T)$ and Pic(T) of a commutative ring T the unit group and the †Picard group of rank 1 projective modules of T, respectively. The full Brauer group $B(R)$ is mapped monomorphically into $H^2(R, U) = \varinjlim H^2(S/R, U)$, the limit over the †faithfully flat R-algebras S.

Grothendieck and others studied the Brauer groups in a more general geometrical context [12].

References

[1] A. A. Albert, Structure of algebras, Amer. Math. Soc. Colloq. Publ., 1939.

[2] E. Artin, C. J. Nesbitt, and R. Thrall, Rings with minimum condition, Univ. of Michigan Press, 1944.

[3] M. Deuring, Algebren, Erg. Math., Springer, second edition, 1968 (Chelsea, 1948).

[4] N. Jacobson, The theory of rings, Amer. Math. Soc. Math. Surveys, 1943.

[5] N. Jacobson, Structure of rings, Amer. Math. Soc. Colloq. Publ., 1956.

[6] B. L. van der Waerden, Algebra II, Springer, 1967.

[7] C. W. Curtis and 1. Reiner, Representation theory of finite groups and associative algebras, Interscience, 1962.

[8] J.-P. Serre, Corps locaux, Actualités Sci. Ind., Hermann, 1962.

[9] G. Azumaya, On maximally central algebras, Nagoya Math. J., 2 (1951), 119–150.

[10] M. Auslander and 0. Goldman, the Brauer group of a commutative ring, Trans. Amer. Math. Soc., 97 (1960), 3677409.

[11] S. Chase and A. Rosenberg, Amitsur cohomology and the Brauer group, Mem. Amer. Math. Soc., 52 (1965), 34-79.

[12] A. Grothendieck, Le groupe de Brauer, 1, II, III, Dix exposés sur la cohomologie des schémas, North-Holland, 1968, 46–188.

[13] M.-A. Knus and M. Ojanguren, Théorie de la descente et algèbres d'Azumaya, Lecture notes in math. 389, Springer, 1974.

[14] M. Orzech and C. Small, the Brauer group of commutative rings, Lecture notes in pure and appl. math. 11, Dekker, 1975.

30 (X.19) Asymptotic Series

A. Asymptotic Series and Asymptotic Expansions

Let $\varphi_n(x)$, $n=0$, 1, 2, be functions defined in $\mathbf{R}_+$. The sequence of functions $\{\varphi_n(x)\}_{n=0}^{\infty}$ is called an **asymptotic sequence** for $x\to\infty$ if

$$\varphi_{n+1}(x)=o(\varphi_n(x)) \text{ as } x\to\infty \tag{1}$$

for each n. When $\{\varphi_n(x)\}_{n=0}^{\infty}$ is an asymptotic sequence, the formal series $\sum a_n\varphi_n(x)$ is called an **asymptotic series**. In most cases, the functions $\varphi_n(x)$ have the form $\psi(x)\varphi(x)^n$. When $\varphi_n(x)=x^{-n}$ an asymptotic series is called an **asymptotic power series.** A function f(x) defined on R_+ is said to have the **asymptotic expansion** (or be **asymptotically developable** in the form)

$$f(x)\sim a_0\varphi_0(x)+a_1\varphi_1(x)+\ldots+a_n\varphi_n(x)+\ldots \tag{2}$$

as $x\to\infty$ if f(x) satisfies

$$f(x)-a_0\varphi_0(x)-a_1\varphi_1(x)-\ldots-a_n\varphi_n(x)$$

$$=O(\varphi_{n+1}(x)) \tag{3}$$

for any integer $n\geqslant 0$ as $x\to\infty$. The coefficients a, (n = 0, 1,) appearing in (2) are uniquely determined. This fact immediately follows from the formulas

$$a_0=\lim_{x\to\infty} f(x)/\varphi_0(x),\ldots,$$

$$a_n=\lim_{x\to\infty}(f(x)-a_0\varphi_0(x)-\ldots$$

$$-a_{n-1}\varphi_{n-1}(x))/\varphi_n(x).$$

For example, the †Hankel function of the first kind $H_\nu^{(1)}(x)$ has an asymptotic expansion

$$H_\nu^{(1)}(x)$$

$$\sim\frac{1}{\sqrt{\pi x}}\exp\left[i\left(x-\frac{\pi\nu}{2}-\frac{\pi}{4}\right)\right]\sum_{m=0}^{\infty}\frac{(\nu,m)}{(-2ix)^m}, \tag{4}$$

$$(\nu\ m)=\frac{(4\nu^2-1)(4\nu^2-3^2)\ldots[4\nu^2-(2m-1)^2]}{2^{2m}m!}$$

If $\nu+\frac{1}{2}$ is not a positive integer, $\sum_{m=0}^{\infty}(\nu,m)(-2ix)^{-m}$ is not convergent, and (4) means that for each n,

$$H_\nu^{(1)}(x)-\frac{1}{\sqrt{\pi x}}\exp\left[i\left(x-\frac{\nu\pi}{2}-\frac{\pi}{4}\right)\right]\sum_{m=0}^{n-1}\frac{(\nu,m)}{(-2ix)^m}$$

$$=O(x^{-n-1/2})$$

for $x\to\infty$.

Assume that f(x) and g(x) admit asymptotic expansions in power series for $x\to\infty$. Then all functions defined by $f(x)+g(x)$, $f(x)-g(x)$, $f(x)g(x)$, $f(x)/g(x)$ ($\lim_{x\to\infty}g(x)\neq 0$) admit asymptotic expansions in power series for $x\to\infty$. Furthermore, if f'(x) has an asymptotic expansion for $x\to\infty$, its expansion is obtained by termwise differentiation of the $f(x)$'s.

For any asymptotic power series $\sum_{n=0}^{\infty}a_nx^{-n}$ there always exists a smooth function f(x) defined in R_+ such that

$$f(x)\sim a_0+a_1x^{-1}+a_2x^{-2}+\ldots$$

for $x\to\infty$, and all the derivatives of f(x) are asymptotically developable in power series. Indeed, a function defined by

$$f(x)=\sum_{n=0}^{\infty}a_n\theta(x/t_n)x^{-n}$$

satisfies the required properties if $\theta(t)$ is a smooth function satisfying

$$\theta(t)=\begin{cases}0, & t\leqslant 1,\\ 1, & t\geqslant 2,\end{cases}$$

and the sequence $\{t_n\}_{n=0}^{\infty}$ satisfies $1\leqslant t_0\leqslant t_1\leqslant$ ... and

$$|n(n+1)\ldots(2n-1)2na_nt_n^{-1}|\leqslant 2^{-n}$$

for all n.

Until now only asymptotic series for $x\to\infty$ have been considered, but the notion of asymptotic series has been adapted in various forms to the spaces of functions under consideration.

(a) Asymptotic series in a complex domain. Let α be a boundary point of an open connected domain D in the complex z-plane. When the asymptotic behavior for $z\to\alpha$ of holomorphic functions in D is considered, we adopt as an asymptotic sequence $\{\varphi_n(z)\}_{n=0}^{\infty}$ of holomorphic functions in D satisfying for each n

$$\varphi_{n+1}(z)=o(\varphi_n(z))$$

as z tends to α through D. Let D be an angular domain with vertex at α. For any asymptotic power series $\sum_{n=0}^{\infty}a_n(z-\alpha)^n$ there always exists a function f(z) that is holomorphic in D and that admits an asymptotic expansion

$$f(z)\sim a_0+a_1(z-\alpha)+\ldots+a_n(z-\alpha)^n+\ldots \tag{5}$$

for $z\to\alpha$. Concerning the uniform convergence of asymptotic expansions, **Carleman's theorem** is well known [2]: If the asymptotic expansion (5) is valid as z tends to α through an arbitrary angular domain in the Riemann surface of $\log(z-a)$ having α as its vertex, then the asymptotic expansion (5) is uniformly convergent.

(b) Asymptotic expansions by a large parameter. In the theory of differential equations, functions with a large parameter are often considered. Let Ω be a domain in $\mathbf{R}^n$ and k a parameter $\geqslant 1$, and let $\varphi_n(x,k)$, $n=0$, 1, 2,. , be functions defined for $(x,k)\in\Omega\times[1,\infty)$. If

they satisfy for each $x \in \Omega$ and n

$$\varphi_{n+1}(x,k)=o(\varphi_n(x,k)) \qquad (6)$$

for $k \to \infty$, the sequence $\{\varphi_n(x,k)\}_{n=0}^{\infty}$ is called an asymptotic sequence, and the formal series $\sum_{n=0}^{\infty} a_n(x)\varphi_n(x,k)$ is called an asymptotic series. A function $f(x,k)$ is said to be asymptotically developable in the form

$$f(x,k) \sim a_0(x)\varphi_0(x,k)+a_1(x)\varphi_1(x,k)+\ldots + a_n(x)\varphi_n(x,k) + \ldots$$

for $k \to \infty$ if for every $x \in \Omega$ and every n

$$f(x,k)-a_0(x)\varphi_0(x,k)-\ldots-a_{n-1}(x)\varphi_{n-1}(x,k) = O(\varphi_n(x,k)) \qquad (7)$$

holds as $k \to \infty$. Especially, if the order relations (6) and (7) hold for each n uniformly in $x \in \Omega$, we say that $f(x,k)$ is uniformly asymptotically developable. In most cases, the $\varphi_n(x,k)$ are of the form $\psi(x,k)k^{-n}$.

B. Methods of Asymptotic Expansion

There are several general methods of obtaining asymptotic expansions of functions represented by integrals with a large parameter. These methods are useful in applications because most †special functions are represented by integrals (→ 39 Bessel Functions, 389 Special Functions).

(a) **Laplace's method.** Let h(s) be a smooth real-valued function in $[a-\delta, a+\delta]$ $(\delta>0)$ satisfying

$$h(a)<h(s) \text{ for all } s\in[a-\delta,a+\delta]-\{a\}$$

and

$$h''(a)>0,$$

and let $g(s)\in C^{\infty}([a-\delta,a+\delta])$. Then

$$I(x)=\int_{a-\delta}^{a+\delta} e^{-xh(s)}g(s)\,ds$$

has an asymptotic expansion

$$I(x) \sim e^{-xh(a)}\frac{\sqrt{2\pi}}{\sqrt{x}}\left\{g(a)\frac{1}{\sqrt{h''(a)}}+c_1x^{-1}+\ldots + c_nx^{-n}+\ldots\right\}$$

for $x \to \infty$, where the c_n are determined by derivatives of h(s) at $s=a$ of order $\leqslant 2n+2$ and those of g(s) at $s=a$ of order $\leqslant 2n$.

As an application of this method, consider the Γ-function $\Gamma(x)=\int_0^{\infty} e^{-t}t^{x-1}\,dt$. By a change of variable $t=xs$, we have $\Gamma(x)= x^x\int_0^{\infty} e^{-x(s-\log s)}s^{-1}\,ds$. Since for any $\delta>0$

$$\left|\int_0^{1-\delta} e^{-x(s-\log s)}s^{-1}\,ds\right| + \left|\int_{1+\delta}^{\infty} e^{-x(s-\log s)}s^{-1}\,ds\right| \leqq Ce^{-x(1+\alpha)} \qquad (\alpha>0),$$

the asymptotic expansion by the Laplace method

$$\int_{1-\delta}^{1+\delta} e^{-x(s-\log s)}s^{-1}\,ds \sim e^{-x}\sqrt{\frac{2\pi}{x}}(1+c_1x^{-1}+\ldots)$$

implies †Stirling's formula

$$\Gamma(x)=\sqrt{2\pi}\,x^{x-1/2}e^{-x}(1+O(x^{-1})).$$

(b) **Stationary phase method.** Let $h(s)$ be a real-valued smooth function defined in $[a-\delta, a+\delta]$ such that

$$h'(a)=0,\ h''(a)\neq 0 \text{ and } h'(s)\neq 0 \text{ for } s\neq a,$$

and let g(s) be a function in $C_0^{\infty}(a-\delta, a+\delta)$. Then

$$I(x)=\int_{a-\delta}^{a+\delta} e^{ixh(s)}g(s)\,ds$$

has an asymptotic expansion

$$I(x) \sim e^{ixh(a)}e^{i\sigma\pi/4}\frac{1}{\sqrt{x}}\left\{\frac{g(a)}{\sqrt{|h''(a)|}}+c_1x^{-1}+\ldots + c_nx^{-n}+\ldots\right\},$$

where $\sigma=h''(a)/|h''(a)|$ and the c_n are determined by derivatives of $h(s)$ at $s=a$ of order $\leqslant 2n+2$ and those of g(s) at $s=a$ of order $\leqslant 2n$.

Further, †integration by parts and the †method of steepest decent are often used to derive asymptotic expansions of integrals with a large parameter.

C. Application to Differential Equations

For a system of linear ordinary differential equations with an †irregular singular point at $z=0$, even when there exists a formal power series solution, the power series is generally divergent. Taking the hint given by Stirling's formula for the Γ-function, Poincaré introduced the notion of asymptotic expansions and succeeded in giving an analytic meaning to formal solutions of divergent type. Actual solutions for †difference equations, †difference-differential equations, and †ordinary differential equations of canonical form (including second-order linear †differential equations of confluent type) always have asymptotic expansions.

In the study of certain ordinary differential equations containing a large parameter k, an asymptotic solution is often obtained in the form

$$u(x,k) \sim e^{ik\varphi(x)}(a_0(x)+a_1(x)k^{-1}+\ldots + a_n(x)k^{-n}+\ldots)$$

(†WKB method). Also, in the study of linear

partial differential equations, an asymptotic solution of the above form plays an important role. For example, consider the Schrodinger equation

$$L_\lambda[u]=\frac{1}{i\lambda}\frac{\partial u}{\partial t}-\frac{1}{(i\lambda)^2}\Delta u-v(x)u=0.$$

An asymptotic series

$$u(x,t;\lambda)\sim e^{i\lambda\varphi(x,t)}\left(a_0(x,t)+a_1(x,t)\frac{1}{i\lambda}+\dots+a_n(x,t)\left(\frac{1}{i\lambda}\right)^n+.\right) \qquad (8)$$

satisfies $L_\lambda[u]\to 0$ for $\lambda\to\infty$ if

$$\frac{\partial\varphi}{\partial t}\quad(\nabla\varphi)^2\quad v(x)\varphi=0 \qquad \text{(eikonal equation)},$$

$$\frac{\partial a_n}{\partial t}-2\nabla\varphi\cdot\nabla a_n-\Delta\varphi a_n=\Delta a_{n-1},\quad n=0,\ 1,\ 2, \qquad \text{(transport equation)},$$

where we set $a_{-1}\equiv 0$. For each asymptotic solution of the form (8) there exists an actual solution of $L_\lambda[f(x,t;\lambda)]=0$ which is asymptotically developable in the form

$$f(x,t;\lambda)\sim\sum_{n=0}^{\infty}e^{i\lambda\varphi(x,t)}a_n(x,t)(i\lambda)^{-n}$$

Then asymptotic solutions of the form (8) can describe the fundamental properties of the phenomenon governed by the equation $L_\lambda[u]=0$.

†Geometric optics can also be illustrated by asymptotic solutions of the form (8) (→ 325 Partial Differential Equations of Hyperbolic Type L). The ideas of asymptotic series, asymptotic expansions, and asymptotic solutions are essential to the theory of †pseudodifferential operators and †Fourier integral operators (→ 345 Pseudodifferential Operators).

References

[1] A. Erdélyi, Asymptotic expansions, Dover, 1956.
[2] T. Carleman, Les fonctions quasi-analytiques, Gauthier-Villars, 1926.
[3] H. Poincaré, Sur les intégrales irrégulières des équations linéaires, Acta Math., 8 (1886), 295–344.
[4] W. R. Wasow, Asymptotic expansions for ordinary differential equations, Interscience, 1965.
[5] W. B. Ford, The asymptotic developments of functions defined by Maclaurin series, Scientific Series, Univ. of Michigan, 1936 (Chelsea, 1960).
[6] M. Hukuhara, T. Kimura, and T. Matuda, Equations différentielles ordinaires du premier ordre dans le champ complex, Publ. Math. Soc. Japan, 1961.
[7] F. W. J. Olver, Asymptotic and special functions, Academic Press, 1974.
[8] G. D. Birkhoff, Collected mathematical papers, Dover, 1968.
[9] V. P. Maslov, Théorie des perturbations et méthodes asymptotiques, Dunord, Gauthier-Villars, 1972. (Original in Russian, 1965.)

31 (XVI.1) Automata

A. General Remarks

Automaton is a generic term for the mathematical models for automatic machines having memory devices and capable of performing definite acts within each unit of time. There are several types of automata, namely, Turing machines for the theory of computability [6], linear bounded automata, pushdown automata, and finite automata for mathematical language theory [10–13]. These models are also used in the theory of †complexity of computation of algorithms [14].

B. Turing Machines

Since the †arithmetization of metamathematics was used in the proof of †Gödel's incompleteness theorem, mathematical formulations of the notion of †effectively computable functions have been attempted by many mathematicians, including S. C. Kleene, A. Church, and others, as well as by Gödel himself (→ 356 Recursive Functions). With this end in mind, A. M. Turing and E. L. Post independently introduced a sort of ideal computer [1,2], called a **Turing machine.**

Such a machine is capable of being in one of a finite number of **internal states** (possible contents of a memory device) $q_0, q_1, \dots, q_n$ fixed for the machine; and the machine is supplied with a tape divided into squares. The tape is infinite in both directions. Each square can be blank or can have printed on it any one of the previously specified symbols $s_0, s_1, \dots, s_m$, where s_0 stands for the blank. In any situation the number of nonblank squares is finite. Let $Q=\{q_0,q_1,\dots,q_n\}$, $M=\{s_0,s_1,\dots,s_m\}$, and consider the mapping $t: Q\times M\to Q\times M\times\{R, L, 0\}$. A Turing machine T is specified by (Q, M, t, q_0, F), where F is a subset of Q called a set of final states and q_0 is the specified **initial state.** If $t(q,s)=(q',s',x)$, then the 5-tuple (q,s,q',s',x) is called an **instruction** for T. The

machine acts as follows: (1) At the start, the machine is in the state q_0. Some sequence of symbols in M are written on the tape as an input, and the machine is on the square next to the leftmost nonblank symbol. (2) With respect to the present state q and the scanned symbol, the machine selects an instruction (q, s, q', s', x); then the next state becomes q', s is replaced by s', and the machine moves one square right or left or stands still according as x is R, L, or 0. (3) The machine stops when the state becomes an element of F, and then the sequence of symbols on the tape is the result of the computation and considered to be the output. A Turing machine is called **deterministic** if there is always at most one instruction associated with a given pair (q, **s);** otherwise it is called **nondeterministic.**

We represent a natural number y on the tape by y + 1 squares, each of which has the symbol s_1 printed on it. A pair (y_1, y_2) is represented on the tape by the following sequence of symbols of length y, + $y_2 + 5$: one s_0, the representation of y,; one s_0, the representation of y_2; and one s_0. Similarly, in general we can represent a k-tuple (y,, . . , y_k) of natural numbers on the tape. Let Φ be a partial function of k variables, $k \geqslant 0$. We say that a Turing machine T **computes Φ** if T starts with the tape in which the representation of $(x_1, , x_k)$ is printed and stops with the tape having the representation of $(x_1, , x_k, x)$ when $\Phi(x_1,$ . ..> $x_k)$ is defined and equal to x. A partial function is **computable** (in the sense of Turing) if there exists a Turing machine that computes the function. It is known that a partial function is computable if and only if it is partial recursive. The equivalence of computability (in the sense of Turing) and recursiveness for number-theoretic functions is strong evidence for the validity of Church's thesis (- 356 Recursive Functions).

We can apply a Turing machine not only to the computation of number-theoretic functions but also to the transformation of a finite sequence of letters in any language with a finite list of letters. A finite nonempty list of distinct symbols is called an **alphabet,** and a finite sequence of symbols in the alphabet is called a **word** in that alphabet. All the words in an alphabet constitute a †free monoid if we add the empty sequence and think of the concatenation of two words as its multiplication. We call a finite nonempty list (A $_1$, B_1), , (A,, B_n) of pairs of words a set of **production rules.** By an **application** of a rule (A_i, B_i) to a word W, we mean that an occurrence of A, exists in W, and by the application, A_i is replaced by B_i. We say that two words R and S are **equivalent** if they are transformable into each other by a finite number of applications of the production rules. The following problem is called **Tbue's (general) problem:** Is there an algorithm for deciding, for any given alphabet and a set of production rules, whether or not any given two words are equivalent? E. L. Post and A. A. Markov proved by applying Turing machines that this decision problem is unsolvable [4, 3] (a negative solution for the so-called word problem for semigroups).

C. Universal Turing Machine and Generalizations

Turing showed that it is possible to construct a machine U with the following property [1]: If we supply U with a tape on which a suitably coded list of instructions of any given machine T is printed then U successively prints on it the behavior of T which results from the execution of the instructions of T. In other words, U simulates the operations performed by any machine. Such a machine U is called a **universal Turing machine.** We can construct a universal Turing machine with only two internal states and also one having only two symbols, including the blank s_0 (C. E. Shannon and J. McCarthy [7]).

Let $\psi_1, . . , \psi_k$ be given functions that are all defined everywhere on the domain considered. Then we can relativize (→ 356 Recursive Functions) the notion of computability of functions (in the sense of Turing) to $\psi_1, , \psi_k$. Indeed, we can consider a machine acting as follows: When the machine is in a state q_{i_j} ($j =$ **1, ,,, k** and we assume that these states are specified correspondingly to $\psi_1, . . , \psi_k$ and are not in F), then the machine prints the representation of $\psi_j(y_1, ,$ Y,,) to the right of $(y_1, ,$ y,,) and scans the resulting $(n_j + 1)$-tuple (y,, . , y_{n_j}, $\psi_j(y_1, ,$ y,,)). In this process, the symbol a that is written to the right of (y,, . . , y,,) is moved one step to the right so that a is preserved on the right of the resulting (n + 1)-tuple. Extending the preceding, Kleene defined a machine that computes a functional of finite types [S]. A functional of variables of any finite types is computable by such a generalized Turing machine if and only if it is partial recursive.

D. Language and Automata

Let M and N be mutually disjoint alphabets and $V =$ MU N. Then G $=(N, M, P,$ S) is called a **Chomsky grammar** if P is a set of production rules consisting of finite number of pairs of elements of V^* and S is specified element of N, where V^* means the free monoid generated by V. For a word $u \in V^*$, if the result of a finite

number of applications of rules in P is v, then we write $u \to v$, and v is said to be **derived** from u by **G.** The **language** generated by G is L(G) = $\{u \ S \to u$ and $u \in M^*\}$. N. Chomsky classified the grammars into three families with respect to the form of production rules: (1) **Context-sensitive grammar:** For each $(u, v) \in P$, the length of u is less than or equal to the length of v. **(2) Context-free grammar:** For each $(u, v) \in P$, u is an element of N and v is not a empty word. **(3) Regular grammar:** For each $(u, v) \in P$, $u \in$ N and v is either an element of M or of the form ***ax,*** where $a \in M$ and $x \in$ N. Let $F_1(M)$, $F_2(M)$, and $F_3(M)$ be the families of languages generated by the grammars in the corresponding families classified as above. Then $F_1(M) \supsetneqq F_2(M) \supsetneqq F_3(M)$ holds and this is called a **Chomsky hierarchy** [10]. Formal language theory studies various structures of these families of languages.

Corresponding to the three families of languages, we have the following families of automata: (1) A **deterministic (nondeterministic) linear bounded automaton** is a Turing machine with input tape, namely, $T = (Q, M \cup N, t, q_0, F)$, where Q is the set of states, M an input alphabet, N a tape alphabet, t is a mapping from $Q \times (M \cup \{e\}) \times (N \cup \{e\})$ into $Q \times N \times \{L, R\}$ (the power set of Q x N x $\{L, R\}$), $q_0 \in Q$ is the initial state, and F is a subset of Q. A word $w \in M^*$ is placed on the input tape, and T starts its computation from the initial state with blank tape and by reading one symbol from the input tape, and with respect to the mapping t, we decide the configuration of the next step. If at a certain stage, T reaches a state in F when T reads up w and if the length of the tape used is a constant multiple of the length of the input tape, then w is said to be **accepted** by T. (The special symbol e makes it possible to change the configuration without moving the input tape.) (2) A **push-down automaton** is a Turing machine with an input tape. The working tape is of infinite length and moves in only one direction, say to the right. The computation with input word $w \in M^*$ starts with the position at the left end of the work tape. It uses the tape according to a first-in-first-out principle; namely, the machine may write a symbol to the blank square, but if it is on the rightmost nonblank square, it may not go left without erasing the nonblank symbol. If it reaches a state in F, then w is said to be **accepted** by T. **(3)** A **finite automaton** is a machine without its infinite tape; a word w is accepted if T reaches a state in F (without using the symbol e).

Let L(T) be the set of words over M accepted by T; then $L(T)$ is called the **language accepted by** T. Corresponding to the three families of automata, $DF_i(M)$ and $NF_i(M)$, $i =$ 1, 2, 3, denote the family of languages accepted by deterministic or nondeterministic automata. It is known that $DF_i(M) = NF_i(M) = F_i(M)$ for $i = 2$ and 3, and $NF_1(M) = F_1(M)$; however, whether $DF_1(M) = NF_1(M)$ holds or not is still open, and this problem is called the **LBA problem.** These automata are used to study the closure properties under various operations within the families $F_i(M)$. In particular, the LBA problem is equivalent to the problem of whether $F_1(M)$ is closed under complementation.

E. Applications of Automata Theory

Finite automata are formulated as models for digital circuits with memory. Therefore there exist structural studies of finite automata; for example, methods to decompose a given finite automaton to smaller automata corresponding to a circuit and its subcircuits are studied using tsemigroup theory [14, 17].

Context-free grammars are useful in describing the syntax of programming languages. For a context-free grammar G, a **derivation tree** of $x \in$ L(G) is a tree whose terminal nodes are labeled by each symbol of x, its root is labeled by S, and other nodes are labeled by some elements of N in such a way that if for a node with $A \in$ N the nodes connected by an outgoing edge from the former have labels $x_1, \ldots, x_k$, then $(A, x_1, \ldots, x_k)$ is a production rule of **G. Parsing** is an algorithm to construct a derivation tree of each $x \in$ L(G). Various parsing algorithms have been studied within subfamilies of context-free grammars [16] and they are used in the construction of compilers.

References

[1] A. M. Turing, On computable numbers, with an application to the Entscheidungsproblem, Proc. London Math. Soc., (2) 42 (1936), 230–265; 43 (1937), 544-546.
[2] E. L. Post, Finite combinatory processes—formulation 1, J. Symbolic Logic, 1 (1936), 103-105.
[3] A. A. Markov, On the impossibility of certain algorithms in the theory of associative systems I (in Russian), Dokl. Akad. Nauk SSSR, 55 (1947), 583-586; II, Dokl. Akad. Nauk SSSR, 58 (1947), 3533356.
[4] E. L. Post, Recursive unsolvability of a problem of Thue, J. Symbolic Logic, 12 (1947), 1–11.
[5] S. C. Kleene, Introduction to metamathematics, Van Nostrand, 1952.
[6] M. Davis, Computability and unsolvability, McGraw-Hill, 1958.

[7] C. E. Shannon and J. McCarthy, Automata studies, Ann. Math. Studies, Princeton Univ. Press, 1956.
[8] S. C. Kleene, Turing-machine computable functionals of finite types I, Logic, methodology and philosophy of science, Proc. 1960 Intern. Congr. Logic, Standford Univ. Press, 1962, 38845: II, Proc. London Math. Soc., (3) *12* (1962), *245-258.*
[9] H. Hermes, Enumerability, decidability, computability, Springer, 1965.
[10] N. Chomsky, Three models for the description of language, IRE Trans. on Information Theory, 2 (1956), 1133124.
[11] J. E. Hopcroft and J. D. Ullman, Introduction to automata theory, languages, and computation, Addison-Wesley, 1979.
[12] S. Eilenberg, Automata, languages and machines, 2 vols., Academic Press, 1974, 1976.
[13] J. Berstel, Transductions and context-free languages, Teubner, 1979.
[14] A. V. Aho, J. E. Hopcroft, and J. D. Ullman, The design and analysis of computer algorithms, Addison-Wesley, 1974.
[15] A. V. Aho and J. D. Ullman, Principles of compiler design, Addison-Wesley, 1977.
[16] K. Krohn and J. L. Rhodes, Algebraic theory of machines I, Prime decomposition theorem for finite semi-groups and machines, Trans. Amer. Math. Soc., 116 (1965), 450–464.
[17] A. Nozaki, Practical decomposition of automata, Information and Control, 36 (1978), *275-291.*

32 (X1.22) Automorphic Functions

A. General Remarks [1]

Let X be a (complex) †analytic manifold and Γ a †discontinuous group of (complex) analytic automorphisms of X. A set of †holomorphic functions (without zero) on X, {j,(z)} ($\gamma\in\Gamma$), is called a **factor of automorphy** if it satisfies the condition j,,.(z) $=j_\gamma(\gamma'(z))j_{\gamma'}(z)$ for all y, $\gamma'\in\Gamma$, $z\in X$. A †meromorphic function f on X is called a **(multiplicative) automorphic function** with respect to the factor of automorphy {j,(z)} if it satisfies the condition $f(\gamma(z))=f(z)j_\gamma(z)$ for all $\gamma\in\Gamma$, $z\in X$. When all j,(z) are identically equal to 1, i.e., when f(z) is Γ-invariant, f is simply called an **automorphic function with respect to** Γ. When all j_γ are constant (and hence y-j, is a "quasicharacter" of Γ), f is called a **multiplicative function.** If we denote by J,(z) the functional determinant of the transformation y, then for an integer m, an automorphic function f with respect to a factor of automorphy of the form j,,(z) = J,(z))" is called an **automorphic form of weight** m. (In this case, it is customary to assume that fis holomorphic. Also, in this and most other cases, we add a suitable condition on the behavior off at the "points at infinity," when $\Gamma\backslash X$ is not compact (- 234 Kleinian Groups).)

B. The Case of One Variable [2-6]

Except for the cases where Γ is a finite group or its extension by a free Abelian group of rank $\leqslant 2$ (→ 134 Elliptic Functions E), it is essentially enough to consider the case where X is the upper half-plane $\mathfrak{H}=\{z\in C \quad \mathrm{Im}\, z>0\}$. In this case, Γ may be obtained as a †discrete subgroup of the †special linear group G = $SL(2,$ **R).** In the following, we restrict ourselves to the case where Γ is a †Fuchsian group of the first kind (i.e., the case where the †Haar measure $\mu(\Gamma\backslash\mathfrak{H})$ of the quotient space $\Gamma\backslash\mathfrak{H}$ is finite; → 122 Discontinuous Groups). An **automorphic function** (or **Fuchsian function**) f with respect to Γ is, by definition, a meromorphic function f on $\mathfrak{H}$ satisfying the following conditions: (i) f is Γ-invariant, i.e., $f(\gamma z)=f(z)$ for all $\gamma\in\Gamma$; (ii) f is also meromorphic around any †cusp x_0 of Γ, i.e., if φ_0 is a real linear fractional transformation mapping x_0 to ∞ (e.g., $\varphi_0(z)=-1/(z-x_0)$) and if the transformation $z\to z+h$ $(h>0)$ is a generator of $\varphi_0\Gamma_{x_0}\varphi_0^{-1}$ (where Γ_{x_0} is the †stabilizer of x_0), then $f(\varphi_0^{-1}(z))$ can be expanded into a †Laurent series of $q_h=\exp((2\pi i/h)z)$ (which has only finitely many terms with negative exponent) in a neighborhood Im z > y, of ∞ (q_h is called the **local parameter** around the cusp x_0). If we denote by $\mathfrak{R}_\Gamma$ the compact †Riemann surface obtained from the quotient space $\Gamma\backslash\mathfrak{H}$ by adjoining a certain (finite) number of "points at infinity" corresponding to the equivalence classes of cusps of Γ, then conditions (i) and (ii) amount to saying that f gives rise in a natural manner to a meromorphic function on $\mathfrak{R}_\Gamma$. Thus the field $\mathfrak{K}_\Gamma$ of all automorphic functions with respect to Γ can be identified with the †algebraic function field belonging to the Riemann surface $\mathfrak{R}_\Gamma$ (→ 11 Algebraic Functions A). (While any (nonconstant) automorphic function with respect to a Fuchsian group of the first kind Γ has the real axis as its †natural boundary, an automorphic function with respect to a Fuchsian group of the second kind can always be analytically extended to the lower half-plane through a neighborhood of any "ordinary point" on the real axis (→ 234 Kleinian Groups).)

Let k be an even integer. A (holomorphic) **automorphic form** (or **Fuchsian form**) f **of**

weight k or **of dimension** -k with respect to Γ is, by definition, a holomorphic function f on $\mathfrak{H}$ satisfying the following conditions: (i) For every $\sigma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$, $f(\sigma z) = f(z) \cdot (cz+d)^k$. (In other words, $J_\sigma(z)^{-k/2} = (d\sigma(z)/dz)^{-k/2} = (cz+d)^k$ is the factor of automorphy.) (ii) f is also holomorphic around any cusp x_0, i.e., in the above notation, we have an integral "Fourier expansion" $f(\varphi_0^{-1}(z))(d\varphi_0^{-1}(z)/dz)^{k/2} = \sum_{\nu=0}^{\infty} a_\nu q_h^\nu$. Note that this definition of weight is different from that given in Section A (- Sections F, G). If, moreover, the constant term a, in this Fourier expansion vanishes at all cusps of Γ, f is called a **cusp form.** In particular, fis a cusp form of weight 2 if and only if $f(z)dz$ gives rise in a natural manner to a †differential form of the first kind on the Riemann surface $\mathfrak{R}_\Gamma$. (By a slight modification of condition (ii), we can also define an automorphic form of odd weight k. When $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \in \Gamma$, condition (i) (for k odd) implies that f is identically equal to zero. So assume that $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \notin \Gamma$. Then cusps x_0 are classified into two categories, according as $\begin{pmatrix} -1 & h \\ 0 & -1 \end{pmatrix} \in \varphi_0 \Gamma_{x_0} \varphi_0^{-1}$ or not, and in the first case we should replace the power series in q_h in condition (ii) by $q_h^{1/2}$ x (power series in q_h).) (Note that the above definition of weight is slightly different from the general one given in Section A. See also the Siegel and Hilbert modular cases (- Sections F, G).)

We denote by $\mathfrak{M}_k(\Gamma)$ (resp. $\mathfrak{S}_k(\Gamma)$) the linear space of all automorphic forms (cusp forms) of weight k with respect to Γ. Since clearly the relations $\mathfrak{M}_k \mathfrak{M}_{k'} \subset \mathfrak{M}_{k+k'}$, $\mathfrak{M}_k \mathfrak{S}_{k'} \subset \mathfrak{S}_{k+k'}$ hold, the direct sum $\mathfrak{M}(\Gamma) = \sum_k \mathfrak{M}_k(\Gamma)$ is a (commutative) †graded algebra, of which $\mathfrak{S}(\Gamma) = \sum_k \mathfrak{S}_k(\Gamma)$ is an ideal. It is known that $\mathfrak{M}_k(\Gamma)$ is of finite dimension, and $d_k = \dim \mathfrak{M}_k(\Gamma)$, $d_k^0 = \dim \mathfrak{S}_k(\Gamma)$ can be determined (from the †Riemann-Roch theorem) as follows:

$$d_k = 0 \quad \text{for} \quad k<0;$$

$$d_0 = 1, \quad d_0^0 = \begin{cases} 1 & \text{for} \quad t = 0, \\ 0 & \text{for} \quad t > 0, \end{cases}$$

$$d_2 = \begin{cases} g & \text{for} \quad t = 0, \\ g+t-1 & \text{for} \quad t > 0, \end{cases} \qquad d_2^0 = g;$$

$$d_k = (k-1)(g-1) + \sum_{i=1}^{s} \left(\frac{k}{2}\left(1 - \frac{1}{e_i}\right)\right) + \frac{k}{2} t,$$

$$d_k^0 = d_k - t \quad \text{for} \quad k \text{ even}, \geqslant 4,$$

where s is the number of the equivalence classes of †elliptic fixed points of Γ, e_i $(1 \leqslant i \leqslant s)$ is the order of the stabilizers of these elliptic points, t is the number of the equivalence classes of cusps of Γ, and g is the †genus of the Riemann surface $\mathfrak{R}_\Gamma$. (When $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \notin \Gamma$, we can obtain a similar formula for k odd except for d_1, d_1^0, which, in general, cannot be determined from the Riemann-Roch theorem.)

One method of constructing automorphic forms is provided by Poincaré series as follows: Let σ_0 be a †linear fractional transformation that maps the unit disk to the upper half-plane (e.g., the †Cayley transformation), and put $\Gamma' = \sigma_0^{-1} \Gamma \sigma_0$; let φ be a function holomorphic on the unit disk including its boundary (e.g., a polynomial). Then for $k \geqslant 4$, the series

$$P_\varphi(z) = \sum_{\sigma' \in \Gamma'} \varphi(\sigma'(z))(d\sigma'(z)/dz)^{k/2}$$

is †uniformly convergent in the wide sense (i.e., uniformly convergent on every compact set) in the unit disk and expresses an automorphic cusp form of weight k with respect to Γ', i.e., we have $P_\varphi(\sigma_0^{-1}(z))(d\sigma_0^{-1}(z)/dz)^{k/2} \in \mathfrak{S}_k(\Gamma)$. Conversely, every element in $\mathfrak{S}_k(\Gamma)$ $(k \geqslant 4)$ can be expressed in this form. A series of this type is called a **theta-Fuchsian series of Poincaré** (or simply **Poincaré series**).

For $k \geqslant 3$, we can define a (positive definite Hermitian) inner product on $\mathfrak{S}_k(\Gamma)$ as follows:

$$(f, g) = \int_F f(z)\overline{g(z)}\, y^{k-2}\, dx\, dy,$$

where $z = x + iy$ and F is a †fundamental region of Γ in $\mathfrak{H}$. This is called the **Petersson metric.** Since this integral converges also for $f, g \in \mathfrak{M}_k(\Gamma)$ if one of them belongs to $\mathfrak{S}_k(\Gamma)$, we can define the orthogonal complement $\mathfrak{E}_k(\Gamma)$ of $\mathfrak{S}_k(\Gamma)$ in $\mathfrak{M}_k(\Gamma)$. As we shall see, $\mathfrak{E}_k(\Gamma)$ is generated by †Eisenstein series.

Any automorphic function f with respect to Γ can be expressed as a quotient of two automorphic forms $f_1, f_2 \in \mathfrak{M}_k(\Gamma)$ for a sufficiently large k. If we put $w = f(z)$, the inverse function $z = f^{-1}(w)$ can be expressed as the quotient of two linearly independent solutions of a †linear differential equation of the form $d^2z/dw^2 = \varphi(w)z$, where $\varphi(w)$ is an †algebraic function belonging to the Riemann surface $\mathfrak{R}_\Gamma$.

C. Modular Functions and Modular Forms [4, 7, 8]

The (elliptic) †modular group

$$\Gamma = \Gamma(1)$$
$$= SL(2, \mathbf{Z})$$
$$= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \middle| \, a, b, c, d \in \mathbf{Z}, ad - bc = 1 \right\}$$

is a Fuchsian group of the first kind acting on

the upper half-plane $\mathfrak{H}$. The quotient space $\Gamma\backslash\mathfrak{H}$ can be compactified to get a compact Riemann surface $\mathfrak{R}_\Gamma = \Gamma\backslash\mathfrak{H} \cup \{\infty\}$ of genus zero by joining one point at infinity corresponding to the cusp ∞, around which we take $q = e^{2\pi i z}$ as a local parameter. The dimension $d_k = \dim \mathfrak{M}_k(\Gamma)$ (k even, $\geqslant 2$) is given as follows:

$$d_k = \begin{cases} [k/12] & \text{for } k \equiv 2 \pmod{12}, \\ [k/12]+1 & \text{for } k \not\equiv 2 \pmod{12}, \end{cases}$$

where [ ] is the †Gauss symbol.

More generally, an automorphic function (automorphic form) with respect to the principal congruence subgroup

$$\Gamma(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \pmod N \right\}$$

is called a **modular function (modular form) of** level N. (For the number of cusps of $\Gamma(N)$, denoted by t(N), and the genus of $\mathfrak{R}_{\Gamma(N)}$ → 122 Discontinuous Groups.) At any cusp of T(N), the local parameter q_h in (ii) is given by $q_N = \exp(2\pi i z/N)$. For N, $k \geqslant 3$, the dimensions d_k, d_k^0 are given by the general formula of the preceding paragraph (including odd k). For $k \geqslant 3$, we defme the (extended) **Eisenstein series** by

$$G_k(z; c_1, c_2, N) = \sum_{m_i \equiv c_i (\text{mod } N)}' \frac{1}{(m_1 + m_2 z)^k},$$

where c_1, c_2 are integers such that $(c_1, c_2, N) = 1$ and the symbol $\sum'$ denotes the summation excepting the pair $(m_1, m_2) = (0,0)$. Then $G_k(z; c_1, c_2, N)$ depends only on c_1, c_2 (mod N), and if we take a set $\{(c_1, c_2)\}$ suchthat $\{(c_1, c_2), (-c_1, -c_2)\}$ forms a complete set of representatives of the (primitive) residue classes (mod N), then the corresponding set of Eisenstein series forms a basis of the space $\mathfrak{E}_k(\Gamma(N))$ $(k \geqslant 3)$, whence we obtain $\dim \mathfrak{E}_k = d_k - d_k^0 =$ t(N) $(k \geqslant 3)$ (for details, including the case k= 1, 2, *see* E. Hecke, *Abh. Math. Sem. Univ. Hamburg,* 5 (1927)).

The Fourier coefficient a, of the Eisenstein series can be calculated easily, and we get an estimate a, = $O(n^{k-1+\varepsilon})$ for $k \geqslant 2$, where ε is an arbitrary positive number. On the other hand, for cusp forms, Hecke *(Abh. Math. Sem. Univ. Hamburg,* 5 (1927)) gave $a_n = O(n^{k/2})$. In an attempt to improve this estimate, H. D. Kloosterman was led to consider the sum

$$K(u, v, q) = \sum_{\substack{x (\text{mod } q) \\ (x,q)=1}} \exp\left(\frac{2\pi i}{q}\left(ux + \frac{v}{x}\right)\right)$$

$$(u, v, 4 \in \mathbf{Z}),$$

called the **Kloosterman sum,** which is also related to the arithmetic of quadratic forms. Using A. Weil's estimate $|K(u, v, p)| \leqslant 2\sqrt{p}$ (where p is an odd prime, $(u, p) = (v, p) = 1$), based on the analog of the Riemann hypothesis, we obtain a, = $O(n^{k/2-1/4+\varepsilon})$. The estimate $a_n = O(n^{k/2-1/2+\varepsilon})$ is a weak form of what is known as the Ramanujan-Petersson conjecture (→ Section D). In the case N = 1, we obtain the classical Eisenstein series $G_k(z) = G_k(z; 0, 0, 1)$ (k even, $\geqslant 4$), from which we define the modular forms $g_2(z) = 60G_4(z)$, $g_3(z) = 140\ G_6(z)$, and $\Delta(z) = g_2^3 - 27g_3^2$ of weight 4, 6, and 12, respectively. If we denote by $\wp = \wp(u; 1, z)$ the Weierstrass †$\wp$-function with the fundamental period (1, z) we have the relation $\wp'^2 = 4\wp^3 - g_2\wp - g_3$, and A(z) is the †discriminant of the cubic polynomial appearing in this relation. It is known that every modular form can be expressed uniquely as a polynomial in g_2 and g_3, or in other words, we have $\mathfrak{M}(\Gamma(1)) \cong \mathbf{C}[g_2, g_3]$. The polynomial in g_2, g_3 expressing a modular form is, moreover, **isobaric,** i.e., consists of terms of the form $c g_2^\mu g_3^\nu$, $c \in \mathbf{C}$, where $2\mu + 3\nu$ is a constant called the **weight** of the isobaric polynomial. Also, A(z) is a cusp form of the smallest weight, and the ideal of all cusp forms, $\mathfrak{S}(\Gamma(1))$, is a principal ideal in $\mathfrak{M}(\Gamma(1))$ generated by A(z). The Fourier expansion of the Eisenstein series $G_k(z)$ is given as follows:

$$G_k(z) = (2\pi)^k \frac{1}{k!}\left(B_k + (-1)^{k/2} 2k \sum_{\nu=1}^{\infty} \nu^{k-1} \frac{q^\nu}{1-q^\nu}\right),$$

where $q = e^{2\pi i z}$ and B_k is the †Bernoulli number. The discriminant A(z) is expressed as an infinite product as follows:

$$\Delta(z) = (2\pi)^{12} q \prod_{\nu=1}^{\infty} (1-q^\nu)^{24}.$$

If we put $J(z) = g_2^3/\Delta$, the function $J(z)$ is a modular function and gives an †analytic isomorphism of the Riemann surface $\mathfrak{R}_{\Gamma(1)} = \Gamma(1)\backslash\mathfrak{H} \cup \{\infty\}$ onto the Riemann sphere $\mathbf{C} \cup \{\infty\}$ (which maps ζ_3, ζ_4, ∞ to 0, 1, ∞, respectively). Hence the field of modular functions $\mathfrak{K}_{\Gamma(1)}$ is a rational function field C(J) generated by the function J. The analytic isomorphism class of †complex tori of dimension 1 (= †elliptic curves) $E_{(\omega_1, \omega_2)} = \mathbf{C}/(\mathbf{Z}\omega_1 + \mathbf{Z}\omega_2)$ with the fundamental period (ω_1, ω_2) is uniquely determined by the $\Gamma(1)$-equivalence class of the **modulus** $\tau = \omega_2/\omega_1$ $(\in \mathfrak{H})$, and hence by the value $J(\tau)$ of the function J. This is the historical origin of the name of tmodular function [4].

As an example of a modular function of level **2, we** have the **λ-function**:

$$\lambda(z) = \frac{\wp((1+z)/2) - \wp(z/2)}{\wp(1/2) - \wp(z/2)},$$

which gives an analytic isomorphism of the Riemann surface $\mathfrak{R}_{\Gamma(2)} = \Gamma(2)\backslash\mathfrak{H} \cup$ (three

points) to the Riemann sphere C ∪ {∞} (mapping 0, 1, ∞ to 1, ∞, 0, respectively; this property is used in a proof of the †Picard theorem). Hence we again have $\mathfrak{R}_{\Gamma(2)} = \mathbf{C}(\lambda)$. The following relation holds between J and λ:

$$J(z)=\frac{4}{27}\frac{(1-\lambda+\lambda^2)^3}{\lambda^2(1-\lambda)^2}$$

In general, the field $\mathfrak{R}_{\Gamma(N)}$ of modular functions of level N is generated by $(g_2/g_3)\cdot\wp((a_1+a_2z)/N;1,z)$ $(a_1,a_2\in\mathbf{Z})$.

D. The Hecke Ring and Its Representation [S-1 13

In general, let $\tilde\Gamma$ be a group, Γ a subgroup of $\tilde\Gamma$, and suppose that, for every $\sigma\in\tilde\Gamma$, $\sigma\Gamma\sigma^{-1}$ and Γ are †commensurable, which amounts to saying that every double coset $\Gamma\sigma\Gamma$ is a union of a finite number of left or right cosets of Γ. Then in the free module generated by all double cosets $\Gamma\sigma\Gamma$ $(\sigma\in\tilde\Gamma)$, we can define a bilinear associative product as follows:

$$\Gamma\sigma\Gamma\cdot\Gamma\tau\Gamma=\sum_{\Gamma\rho\Gamma}m(\sigma,\tau;\rho)\Gamma\rho\Gamma,$$

where, writing $\Gamma\sigma\Gamma=\bigcup_i\Gamma\sigma_i$, $\Gamma\tau\Gamma=\bigcup_j\Gamma\tau_j$, we denote by $m(\sigma,\tau;\rho)$ the number of pairs (i,j) such that $\Gamma\sigma_i\tau_j=\Gamma\rho$. The associative ring thus obtained is called the **Hecke ring** (or **Hecke algebra**) and is denoted by $\mathscr{H}(\tilde\Gamma,\Gamma)$. (If $\tilde\Gamma$ is a †topological group and Γ is an open compact subgroup, the Hecke ring $\mathscr{H}(\tilde\Gamma,\Gamma)$ can also be interpreted as the ring of all Z-valued continuous functions on $\tilde\Gamma$ with compact support provided with a †convolution product; → 29 Associative Algebras C.)

As an example, setting $\tilde\Gamma=GL^+(2,\mathbf{Q})$, $\Gamma=SL(2,\mathbf{Z})$, we obtain a Hecke ring $\mathscr{H}(\tilde\Gamma,\Gamma)$. In virtue of the theory of †elementary divisors, a complete set of representatives of $\Gamma\backslash\tilde\Gamma/\Gamma$ is given by

$$\left\{\begin{pmatrix}a_1&0\\0&a_2\end{pmatrix}\middle|\,a_1,a_2\in\mathbf{Q},a_1,a_2>0,a_1\,|\,a_2\right\}.$$

We write $T(a_1,a_2)=\Gamma\begin{pmatrix}a_1&0\\0&a_2\end{pmatrix}\Gamma$. From the relation $T(a_1a_2,a_1a_2)T(a_1^{-1},a_2^{-1})=T(a_1,a_2)$ follows the commutativity of $\mathscr{H}(\tilde\Gamma,\Gamma)$. Furthermore, if we put, for a positive integer n,

$$T_n=T(n)\sum_{\substack{a_1,a_2\in Z\\ a_1a_2=n}}T(a_1,a_2),$$

then the following multiplication formula holds:

$$T(n)\cdot T(m)=\sum_{d|n,m}dT(d,d)\,.\,T(nm/d^2).$$

We obtain a representation of $\mathscr{H}(\tilde\Gamma,\Gamma)$ in $\mathfrak{M}_k(\Gamma)$ as follows. For $T=\Gamma\sigma\Gamma=\cup\Gamma\sigma_i\in\mathscr{H}(\tilde\Gamma,\Gamma)$ and $f\in\mathfrak{M}_k(\Gamma)$ we put

$$(f\,|\,T)(z)=\sum_i f(\sigma_i(z))(c_iz+d_i)^{-k}\det(\sigma_i)^l,$$

where $\sigma_i=\begin{pmatrix}a_i&b_i\\c_i&d_i\end{pmatrix}$ and l is a fixed integer. This representation leaves $\mathfrak{S}_k(\Gamma)$ and $\mathfrak{E}_k(\Gamma)$ invariant. In particular, the representation of T_n (with $l=k$ 1), called the **Hecke operator** (Hecke, Math. *Ann.*, 114 (1937)), is †Hermitian with respect to the Petersson metric.

Following Hecke, we associate with every modular form $f(z)=\sum_{n=0}^{\infty}a_nq^n$ $(q=e^{2\pi iz})$ a †Dirichlet series $\varphi(s)=\sum_{n=1}^{\infty}a_nn^{-s}$. Since $a_n=O(n^{k-1+\varepsilon})$ for $f\in\mathfrak{M}_k$ $(k\geqslant2)$, $\varphi(s)$ is absolutely convergent in the half-plane Re $s>k$. The conditions for f to be a modular form of weight k are equivalent to the following conditions for φ: (s-k) φ (s) can be extended to an †entire function of fmite †genus (actually, of genus 1) and, if we put $R(s)=(2\pi)^{-s}\Gamma(s)\varphi(s)$, R(s) satisfies a functional equation of the form $R(k-s)=(-1)^{k/2}R(s)$. A correspondence between f and φ that satisfies these conditions is one-to-one; in fact we have $a_n=(-1)^{k/2}\operatorname{Res}_{s=k}R(s)$, and the function $g(x)=f(ix)-a_0$ $(x>0)$ and $R(s)$ are related by the †Mellin transform as follows:

$$R(s)=\int_0^{\infty}g(x)x^{s-1}\,dx,$$

$$g(x)=\frac{1}{2\pi i}\int_{\operatorname{Re}s=\sigma_0}R(s)x^{-s}ds.$$

This correspondence between Dirichlet series and automorphic forms can be generalized to the case of a certain discrete subgroup of $SL(2,\mathbf{R})$, which is not necessarily of the first kind (Hecke, *Math. Ann.*, 112 (1936); [10]). It is also known that if the functional equations for $\sum_{n=1}^{\infty}\chi(n)a_nn^{-s}$ hold for sufficiently many characters χ, then $f=\sum_{n=1}^{\infty}a_nq^n$ is a cusp form with a character for some congruence subgroup (Weil, *Math. Ann.*, 168 (1967); [12]).

Now, suppose that a linear subspace $\mathfrak{M}$ of $\mathfrak{M}_k(\Gamma)$ is invariant under all T_n $(n=1,2,\ldots)$, and let $(f_1,\ldots,f_\kappa)$ $(\kappa=\dim\mathfrak{M})$ be a basis of $\mathfrak{M}$. If we denote by $\lambda(n)$ the $\kappa\times\kappa$ matrix representing T_n in this basis, it can be shown that there exist $\kappa\times\kappa$ (complex) matrices $B^{(i)}$ $(1\leqslant i\leqslant\kappa)$ and $\lambda(0)$ such that we have

$$F(z)=\sum_{n=0}^{\infty}q^n\lambda(n)=\sum_{i=1}^{\kappa}f_i(z)B^{(i)}.$$

Similarly we associate with $F(z)$ a matrix-valued Dirichlet series

$$\Phi(s)=\sum_{n=1}^{\infty}n^{-s}\lambda(n).$$

Then from the multiplicative property of the $\lambda(n)$, we obtain the following †Euler product

expression of $\Phi(s)$:

$$\Phi(s)=\prod_p (I_\kappa-\lambda(p)p^{-s}+p^{\kappa-1-2s}I_\kappa)^{-1},$$

where I_κ denotes the unit matrix of degree κ. In particular, when $\dim\mathfrak{M}=1$, i.e., when $\mathfrak{M}=\mathbf{C}f$, where f is an †eigenfunction of all Hecke operators T_n, the corresponding Dirichlet series $\varphi(s)$ has an Euler product of the following form:

$$\varphi(s)=\prod_p (1-\lambda_p p^{-s}+p^{k-1-2s})^{-1}.$$

For instance, the Dirichlet series corresponding to the Eisenstein series G_k is $(2(2\pi i)^k(k-1)!)\zeta(s)\zeta(s-k+1)$ (where ζ is the †Riemann zeta function), which has a well-known Euler product. Since $\mathfrak{S}_{12}$ = CA, the Dirichlet series corresponding to $\Delta(z)$ also has an Euler product. (This is part of the **Ramanujan conjecture.** Ramanujan also,conjectured that the quadratic polynomial $1-\lambda_p X+p^{11}X^2$ appearing in this Euler product has imaginary roots, i.e., $|\lambda_p|<2p^{11/2}$.) Generalizing the Ramanujan conjecture, Petersson conjectured that all eigenvalues of the Hecke operator T_p in $\mathfrak{S}_k$ have absolute value $\leqslant 2p^{(k-1)/2}$ **(Ramanujan-Petersson conjecture).** P. Deligne proved this conjecture for $k\geqslant 2$, and his method extends to the case when Γ is a congruence subgroup (*Publ. Math. Inst. HES*, 53 (1974)). For $k=1$, see Deligne and J.-P. Serre (Ann. *Sci. Ecole Norm. Sup.*, 7 (1974)). It may be noted that this conjecture was proved earlier for $k=2$ (and for almost all p) by M. Eichler (1954) and by G. Shimura (1958) (→ 450 Zeta Functions M).

E. The Case of Many Variables [1, 15,161

Let X be a bounded domain in $\mathbf{C}^N$ and Γ a †discontinuous group of analytic automorphisms of X. The Poincaré series of weight (→ Section A) $m\geqslant 2$, defined similarly to the case of the unit disk, converges normally in X and expresses an automorphic form of weight m. If $\Gamma\backslash X$ is compact, let $(f_1, ,f_\kappa)$ be a basis of the space of Poincaré series of weight m. Then for a sufficiently large m (which is a multiple of the order of Γ_x for all x ∈ X), the map X ∋ $z\to$ (f,(z), $,f_\kappa(z)$) defines in a natural manner a projective embedding of the quotient space $\Gamma\backslash X$ into $P^{\kappa-1}$ (C), which actually gives an analytic isomorphism of $\Gamma\backslash X$ onto a tnormal projective algebraic variety. It follows that the field of automorphic functions with respect to Γ is an †algebraic function field of dimension N, and that every automorphic function can be written as the quotient of two Poincaré series of the same weight. In the case where $\Gamma\backslash X$ is not compact, we first discuss some examples.

F. Siegel Modular Functions [15, 17, 18]

The **Siegel upper half-space** $\mathfrak{H}_n$ **of degree** n (or **Siegel space of degree** n) is, by definition, the space of all $n\times n$ complex symmetric matrices $Z=X+iY$ with the imaginary part Y > 0 (positive definite). ($\mathfrak{H}_n$ is analytically equivalent to a †symmetric bounded domain.) The group of all (complex) analytic automorphisms of $\mathfrak{H}_n$ is given by the real (projective) †symplectic group $Sp(n,\mathbf{R})/\{\pm I_{2n}\}$, which acts transitively on $\mathfrak{H}_n$ by $Z\to\sigma(Z)=(AZ+B)/(CZ+D)$ for $\sigma=\begin{pmatrix}A & B\\ C & D\end{pmatrix}\in Sp(n,\mathbf{R})$. The subgroup $\Gamma_n=Sp(n,\mathbf{Z})$ consisting of integral matrices, or the corresponding group of linear fractional transformations, is called the **Siegel modular group of degree** n. Γ_n is a †discontinuous group of the first kind acting on $\mathfrak{H}_n$.

A **Siegel modular form** f **of weight** k or **of dimension** -k is, by definition, a holomorphic function on $\mathfrak{H}_n$ satisfying the following conditions: (i) for $\sigma=\begin{pmatrix}A & B\\ C & D\end{pmatrix}\in\Gamma_n$, $f(\sigma(Z))=$ f(Z)det$(CZ+D)^k$; (ii) f has an integral Fourier expansion of the form f(Z) = $\sum_{T\geqslant 0} a_T e^{2\pi i \operatorname{tr}(TZ)}$, where T runs over all half-integral †positive semidefinite symmetric matrices of degree n. (Note that the Jacobian determinant J,(Z) of $\sigma=\begin{pmatrix}A & B\\ c & D\end{pmatrix}$ is given by $J_\sigma(Z)=\det(CZ+D)^{-k-1}$.) For $n\geqslant 2$, condition (ii) is superfluous (M. Koecher, Math. Z., 59 (1954)). We denote by $\mathfrak{M}_k^{(n)}=\mathfrak{M}_k(\Gamma_n)$ the space of all Siegel modular forms of degree n and weight k. When we write $\mathfrak{H}_n$ 3 Z = $\begin{pmatrix}Z_1 & 3\\ 3 & z\end{pmatrix}$ with $Z_1\in\mathfrak{H}_{n-1}$, $\mathfrak{z}\in\mathbf{C}^{n-1}$, $z\in\mathfrak{H}_1$, and

$$\mathfrak{M}_k^{(n)}\ni f(Z)=\sum_{n=0}^{\infty} a_n(Z_1,\mathfrak{z})e^{2\pi i n z},$$

it turns out that $a_0(Z_1, 3)$ depends only on Z_1 and, writing it as $f_1(Z_1)$, we have $f_1\in\mathfrak{M}_k^{(n-1)}$. The mapping $\Phi: f\to f_1$ thus defined is a linear mapping from $\mathfrak{M}_k^{(n)}$ into $\mathfrak{M}_k^{(n-1)}$, which is surjective if k is even and > 2n (H. Maass, Math. *Ann.*, 123 (1951)). We denote the kernel of Φ by $\mathfrak{S}_k^{(n)}$ and call $f\in\mathfrak{S}_k^{(n)}$ a **cusp form** (viewing $\mathfrak{H}_{n-1}$ as a cusp of $\mathfrak{H}_n$). For $f\in\mathfrak{M}_k^{(n)}$, the following three conditions are equivalent: (a) $f\in\mathfrak{S}_k^{(n)}$; (b) f(Z) = $\sum_{T>0} a_T e^{2\pi i\operatorname{tr}(TZ)}$; (c) $|f(Z)\det(Y)^{k/2}$ is bounded. We have $\mathfrak{M}_k^{(n)}=\{0\}$ if $k<0$ or if both k and n are odd, and $\mathfrak{M}_0^{(n)}$ = C. In general, dim $\mathfrak{M}_k^{(n)}$ is finite and $= O(k^{n(n+1)/2})$ (k-t ∞). The graded algebra $\mathfrak{M}(\Gamma_n)=\sum_{k=0}^{\infty}\mathfrak{M}_k^{(n)}$ is finitely generated, and its †Krull dimension is $n(n+1)/2+1$. In particular, for n= 2 the structure of the graded algebra $\mathfrak{M}(\Gamma_2)$ is determined explicitly. (The even part of this algebra is isomorphic to the

polynomial ring of four variables.) Also, $\dim \mathfrak{M}_k^{(2)}$ and the singularities of the quotient variety $\Gamma_2 \backslash \mathfrak{H}_2$ are known explicitly (J. 1. Igusa, *Amer. J. Math.*, 84 (1962), 86 (1964); U. Christian).

As an example of Siegel modular forms, we have the **Eisenstein-Poincaré series**, defined as follows:

$$E_{S,k}(Z)=\sum_{\sigma\in\Gamma_n(S)\backslash\Gamma_n} e^{2\pi i\,\mathrm{tr}(S\sigma(z))}\det(CZ+D)^{-k},$$

where S is an $n \times n$ rational symmetric matrix $\geqslant 0$, and we put

$$\Gamma_n(S)=\left\{\begin{pmatrix} U & T\,{}^tU^{-1} \\ 0 & {}^tU^{-1}\end{pmatrix}\in\Gamma_n \,\middle|\, {}^tUSU=S,\ e^{2\pi i\,\mathrm{tr}(ST)}=\det(U)^k\right\}$$

This series is convergent for $k>n+\operatorname{rank} S+1$ if $\operatorname{rank} S<n$ and for $k>2n$ if $S>0$, and the totality of $E_{S,k}$ (those series with S > 0) spans $\mathfrak{M}_k^{(n)}$ ($\mathfrak{S}_k^{(n)}$) (Maass, *Math. Ann.*, 123 (1951)). †Theta series defined by integral quadratic forms are examples of Siegel modular forms (with a certain level) and are of great significance in the arithmetic of quadratic forms (C. L. Siegel, *Ann. Math.*, *(2) 36* (1935)).

The quotient space $\Gamma_n \backslash \mathfrak{H}_n$ can be compactified as follows (I. Satake and W. Baily [15]): There exists a positive integer k_0 such that, for any multiple k of k_0, any basis of $\mathfrak{M}_k^{(n)}$ defines in a natural manner a one-to-one biholomorphic projective embedding of $\Gamma_n \backslash \mathfrak{H}_n$, of which the image is a Zariski open set of a normal projective algebraic variety. Since the structure of this projective variety is independent of *k*, we denote it simply by $\overline{\Gamma_n \backslash \mathfrak{H}_n}$. In other words, $\overline{\Gamma_n \backslash \mathfrak{H}_n}$ is the projective variety Proj $\mathfrak{M}(\Gamma_n)$ associated with the graded algebra $\mathfrak{M}(\Gamma_n)$. Then we have $\overline{\Gamma_n \backslash \mathfrak{H}_n}=\bigcup_{r=0}^{n}\Gamma_r \backslash \mathfrak{H}_r$. When $n\geqslant 2$ for $r<n$, $\Gamma_r \backslash \mathfrak{H}_r$ is of codimension $\geqslant 2$ in $\overline{\Gamma_n \backslash \mathfrak{H}_n}$, so that (by virtue of †Hartogs's continuation theorem) the conditions at the points at infinity in the definitions of modular functions and forms become superfluous. In fact, for $n \geqslant 2$, if we define a **Siegel modular function of degree** n simply as a Γ_n-invariant meromorphic function on $\mathfrak{H}_n$, then it can automatically be extended to a meromorphic function on the compactification $\overline{\Gamma_n \backslash \mathfrak{H}_n}$, and hence be expressed as the quotient of two modular forms of the same weight. It also follows that the field of all Siegel modular functions of degree n is an algebraic function field of dimension $n(n+1)/2$. (These results can also be obtained from the †pseudoconcavity of $\Gamma_n \backslash \mathfrak{H}_n$ without using compactification (A. Andreotti and H. Grauert, *Nachr. Akud. Wiss. Göttingen*).) For $n=2$, the field of Siegel modular functions is a rational function field. On the other hand, for $n \equiv 1 \bmod 8$, $n>9$, the field of Siegel modular functions is not rational (E. Freitag, *Abh. Math. Sem. Univ. Hamburg* (1978)).

Let Γ be a subgroup of Γ_n of finite index. Then $\Gamma \backslash \mathfrak{H}_n$ can be compactified in the same way as above [23]. However the singularities of $\overline{\Gamma \backslash \mathfrak{H}_n}$ are extraordinarily complicated (Igusa [24]). When Γ is the principal congruence subgroup $\Gamma_n(N)$ of level N > 2, the blowing-up of $\overline{\Gamma_n(N) \backslash \mathfrak{H}_n}$ along the cusps is nonsingular for $n=2, 3$ (Igusa, *Math. Ann.*, 168 (1967)). Following Igusa's idea, it is possible to construct an explicit nonsingular compactification of $\Gamma \backslash D$, in which D is a bounded symmetric domain and Γ is a suitable arithmetic subgroup of $\mathrm{Aut}(D)$ ([19]; see also Satake, Bull. *Amer. Math. Soc., 79* (1973)). For $D=\mathfrak{H}\times\mathfrak{H}$ and Γ the Hilbert modular group (→ Section G), F. Hirzebruch has given explicit nonsingular compactifications of $\Gamma \backslash D$ *(Enseignement* Math., 19 (1973)).

The Hecke theory (→ Section D) can also be extended to the Siegel modular case. For any integer $m>0$, let $S_m=S_m^{(n)}$ be the set of $2n \times 2n$ integral matrices M such that ${}^tMJM=mJ$, where $J=\begin{pmatrix} 0 & -I_n \\ I_n & 0\end{pmatrix}$ and I_n is the identity matrix of degree n. We define the Hecke operator $T(m)$ on $\mathfrak{M}_k^{(n)}$ by

$$T(m)f=m^{nk-n(n+1)/2}\sum_{\left(\begin{smallmatrix} A & B \\ C & D\end{smallmatrix}\right)=M\in V(m)}\det(CZ+D)^{-k}f(MZ),$$

where $f\in\mathfrak{M}_k^{(n)}$, $S_m=\bigcup_{M\in V(m)}\Gamma_n M$, and $V(m)$ is a complete system of representatives of $\Gamma_n \backslash S_m$. Then for any prime number p, the formal operator series $D_p(X)=\sum_{n=0}^{\infty}T(p^n)X^n$ is formally equal to a rational function in X:

$$D_p(X)=\frac{P_p(X)}{Q_p(X)},$$

in which $P_p(X)$ and $Q_p(X)$ are polynomials in X of degree 2^n-2 and 2^n, respectively, with coefficients in the ring of Hecke operators (Satake, *Publ. Math. Inst. HES*, 18 (1964)). For $n=2$ we have (Shimura, *Proc. Nat. Acud. Sci.* US, 49 (1963))

$$D_p(X)=(1-p^{2k-4}X^2)\{1-T(p)X+(T(p)^2-T(p^2)-p^{2k-4})X^2-p^{2k-3}T(p)X^3+p^{4k-6}X^4\}^{-1}.$$

Let $f(Z)\in\mathfrak{M}_k^{(2)}$ be a Siegel modular form of degree 2 such that $T(m)f=\lambda(m)f$ for all $m\geqslant 1$. We define $L_f(s)$ by

$$L_f(s)=\zeta(2s-2k+4)\sum_{m=1}^{\infty}\frac{\lambda(m)}{m^s},$$

in which $\zeta(s)$ is the Riemann zeta function. It is known that $L_f(s)$ can be meromorphically continued to the entire plane and satisfies the functional equation

$$\Psi_f(2k-2-s)=\Psi_f(s),$$

where $\Psi_f(s)=(2\pi)^{-2s}\Gamma(s)\Gamma(s-k+2)L_f(s)$ and the function $(s-k)(s-k+2)\Psi_f(s)$ is entire (A. Andrianov, *Trudy Math. Inst. Stekloo.*, 112 (1971)).

G. Hilbert Modular Functions

Following a suggestion of Hilbert, 0. Blumenthal studied a generalization of modular functions of the following type: Let K be a totally real †algebraic number field of Imite degree, and let $K^{(1)}, \ldots, K^{(n)}$ be the conjugates of K, where $n=[K:\mathbf{Q}]$. ("Totally real" means that all the $K^{(i)}$ are real.) For $\alpha\in K$, we denote by $\alpha^{(i)}\in K^{(i)}$ the ith conjugate of a. Let $\mathfrak{D}$ be the †Principal order of K, and consider the group

$$\Gamma_{\mathfrak{D}}=SL(2,\mathfrak{D})=\left\{\sigma=\begin{pmatrix}\alpha&\beta\\ \gamma&\delta\end{pmatrix}\middle|\,\alpha,\beta,\gamma,\delta\in\mathfrak{D},\alpha\delta-\beta\gamma=1\right\}.$$

We define the action of $\Gamma_{\mathfrak{D}}$ on the n-fold product of the upper half-plane $\mathfrak{H}^n=\{z=(z_1,\ldots,z_n)\mid z_i\in\mathfrak{H}\}$ by

$$\sigma(z)=\left(\frac{\alpha^{(1)}z_1+\beta^{(1)}}{\gamma^{(1)}z_1+\delta^{(1)}},\ldots,\frac{\alpha^{(n)}z_n+\beta^{(n)}}{\gamma^{(n)}z_n+\delta^{(n)}}\right)$$

for $\sigma=\begin{pmatrix}\alpha&\beta\\ \gamma&\delta\end{pmatrix}$; then $\Gamma_{\mathfrak{D}}$ becomes a discontinuous group of the first kind ($\Gamma_{\mathfrak{D}}$ can also be considered as an †irreducible discrete subgroup of $SL(2,\mathbf{R})^n$). The group $\Gamma_{\mathfrak{D}}$ is called the **Hilbert modular group** of K (in the strict sense). (The defmition is sometimes modilïed by replacing the condition $a6-\beta\gamma=1$ by, say, $\alpha\delta-\beta\gamma$ =(totally positive †unit).) If the †class number of K is h, the quotient space $\Gamma_{\mathfrak{D}}\backslash\mathfrak{H}^n$ can be compactified by adjoining h points at infinity. Therefore, if $n\geqslant 2$, a **Hilbert modular function** can be defined as a meromorphic function on $\mathfrak{H}^n$ invariant under $\Gamma_{\mathfrak{D}}$; and similarly a **Hilbert modular form** f **of weigbt** k or **of dimension** $-k$ as a holomorphic function f on $\mathfrak{H}^n$ such that $(f|\sigma)(z)=f(\sigma(z))\prod_{i=1}^n(\gamma^{(i)}z_i+\delta^{(i)})^{-k}=f(z)$ for all $\sigma\in\Gamma_{\mathfrak{D}}$. (In the latter case, f is holomorphic at all cusps, i.e., for every $\sigma\in SL(2,K)$, $f|\sigma$ has an integral Fourier expansion.) For Hilbert modular functions and forms, results quite similar to those for the case $n=1$ or the case of Siegel modular groups have been obtained (Kloosterman, Maass, K. B. Gundlach, H. Klingen).

H. Furtber Generalizations

As further generalizations of the notion of modular function, we have Hilbert-Siegel modular functions (1.1. Pyatetskiĭ-Shapiro, Baily, Christian [20], where we can tïnd some 150 references), Hermitian modular functions (H. Braun, Klingen), etc. For the most general case, i.e., for an arithmetically defined discontinuous group acting on a †symmetric bounded domain, a unified theory of automorphic functions has been established in the works of Pyatetskiĭ-Shapiro [21, 22] and Baily and A. Borel [23, 24].

On the other hand, exactly as in the classical theory, where the elliptic modular function gave an invariant of 1-dimensional complex tori, generalized modular functions can be viewed as giving an analytic invariant of a certain family of (polarized) †Abelian varieties. From this point of *view*, a deep number-theoretic (and algebrogeometric) study of automorphic functions (initiated by Hecke and Eichler) has been substantially carried forward by the work of Shimura (see the series of his papers in *Ann. Math.* starting from vol. 70 (1959); see also [9, 24, 25]).

From the analytic point of view, the theory of automorphic functions is closely connected with the unitary representation of G in the space $L_2(\Gamma\backslash G)$ (or its adelic analog) [26, 27] (for adelic analogs → [12,131). In this respect, the †trace formula of A. Selberg [28], generalizing the †Poisson summation formula, is of fundamental importance; and actually it can be used effectively for calculations of the dimension of the space of automorphic forms and of the trace of Hecke operators (R. P. Langlands, *Amer. J. Math.*, *85* (1963); H. Shimizu, *Ann. Math.*, *(2) 77* (1963), *J. Fac. Sci. Univ. Tokyo*, 10 (1963); also [24]). When $X=G/K$ is a symmetric domain, we can define, for any representation ρ of K, a (matrix-valued) canonical automorphy factor, by which we defme (vector-valued) automorphic forms with respect to a discrete subgroup Γ of G, and under a further condition (say, Γ free, $\Gamma\backslash X$ compact, and the †highest weight of ρ sufftciently large) we obtain a formula for the dimension of the space of such automorphic forms in terms of the †arithmetic genus of $\Gamma\backslash X$ and certain numbers related to the "dual" $X_u=G_u/K$ and the representation ρ [15, 29, 30].

References

[1] Sém. H. Cartan 6, Théorie des fonctions de plusieurs variables; Fonctions automorphes et espaces analytiques, Ecole Norm. Sup., 1953–1954.

[2] M. Eichler, Einführung in die Theorie der algebraischen Zahlen und Funktionen, Birkhäuser, 1963; English translation, Introduction to the theory of algebraic numbers and functions, Academic Press, 1966.
[3] R. Fricke and F. Klein, Vorlesungen über die Theorie der automorphen Funktionen, Teubner, 1, 1897; II, 1901; second edition, 1926.
[4] F. Klein and R. Fricke, Vorlesungen über die Theorie der elliptischen Modulfunktionen, Teubner, I, 1890; II, 1892.
[5] J. Lehner, Discontinuous groups and automorphic functions, Amer. Math. Soc. Math. Surveys, 1964.
[6] H. Poincaré, Mémoire sur les fonctions fuchsiennes, Acta Math., 1 (1882), 193-294 (Oeuvres, Gauthier-Villars, 1916, vol. 2, 167-257).
[7] A. P. Ogg, Modular forms and Dirichlet series, Benjamin, 1969.
[8] G. Shimura, Introduction to the arithmetic theory of automorphic functions, Publ. Math. Soc. Japan, Iwanami Shoten and Princeton Univ. Press, 1971.
[9] M. Eichler, Quadratische Formen und Modulfunktionen, Acta Arith., 4 (1958), 217-239.
[10] E. Hecke, Mathematische Werke, Vandenhoeck & Ruprecht, 1959.
[11] E. Hecke, Dirichlet series, modular forms and quadratic forms, Lecture notes, Institute for Advanced Study, Princeton, 1938.
[12] H. Jacquet and R. P. Langlands, Automorphic forms on $GL(2)$, Lecture notes in math. 114, Springer, 1970.
[13] Automorphic forms, representations and L-functions, Amer. Math. Soc. Proc. Symposia in Pure Math., 33 (1979).
[14] Modular functions of one variable I-VI, Lecture notes in math., Springer, 320 (1973); 349 (1973); 350 (1973); 476 (1975); 601 (1977); 627 (1977).
[15] Sém. H. Cartan 10, Fonctions automorphes, Ecole Norm. Sup., 1957-1958.
[16] C. L. Siegel, Analytic functions of several complex variables, Lecture notes, Institute for Advanced Study, Princeton, 1950.
[17] H. Maass, Siegel's modular forms and Dirichlet series, Lecture notes in math. 216, Springer, 1971.
[18] C. L. Siegel, Einführung in die Theorie der Modulfunktionen n-ten Grades, Math. Ann., 116 (1939), 617-657 (Gesammelte Abhandlungen, Springer, 1966, vol. 2, 97-137).
[19] D. Mumford, et al., Smooth compactification of locally symmetric varieties, Math. Sci. Press, 1975.
[20] U. Christian, Zur Theorie der Hilbert-Siegelschen Modulfunktionen, Math. Ann., 152 (1963), 275-341.
[21] 1.1. Pyatetskii-Shapiro (Pjateckiĭ-Šapiro), Automorphic functions and the geometry of classical domains, Gordon & Breach, 1969. (Original in Russian, 1961.)
[22] I. I. Pyatetskiĭ-Shapiro (Pjateckiĭ-Šapiro), Arithmetic groups in complex domains, Russian Math. Surveys, 19 (1964), 83–109. (Original in Russian, 1964.)
[23] W. L. Baily and A. Borel, Compactification of arithmetic quotients of bounded symmetric domains, Ann. Math., (2) 84 (1966), 442-528.
[24] Algebraic groups and discontinuous subgroups, Amer. Math. Soc. Proc. Symposia in Pure Math., 9 (1966).
[25] G. Shimura, Automorphic functions and number theory, Lecture notes in math. 54, Springer, 1968.
[26] I. M. Gel'fand, M. I. Graev, and I. I. Pyatetskiĭ-Shapiro (Pjateckiĭ-Šapiro), Generalized functions 6, Representation theory and automorphic functions, Saunders, 1969. (Original in Russian, 1966.)
[27] Harish-Chandra, Automorphic forms on semisimple Lie groups, Lecture notes in math. 62, Springer, 1968.
[28] A. Selberg, Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series, J. Indian Math. Soc., 20 (1956), 47–87.
[29] Y. Matsushima and S. Murakami, On vector bundle valued harmonic forms and automorphic forms on symmetric Riemannian manifolds, Ann. Math., (2) 78 (1963), 365–416.
[30] Y. Matsushima and S. Murakami, On certain cohomology groups attached to Hermitian symmetric spaces, Osaka J. Math., 2 (1965), 1–35.
[31] R. Tsushima, A formula for the dimensions of spaces of Siegel cusp forms of degree three, Amer. J. Math., 102 (1980), 933-977.

33 (1.5) Axiomatic Set Theory

A. General Remarks

Axiomatic set theory pursues the goal of reestablishing the essentials of G. Cantor's rather intuitive †set theory by axiomatic constructions consistent with modern theories of the foundations of mathematics.

A system of axioms for set theory was first given by E. Zermelo [47], and was completed by A. Fraenkel [8]. J. von Neumann [30] expressed it in †symbolic logic, gave a formal generalization, and eliminated ambiguous concepts. P. Bernays and K. Gödel [2, 11]

further refined and simplified von Neumann's formulation. The theories based on the systems before and after the formal generalization are called **Zermelo-Fraenkel set theory** (ZF) and **Bernays-Gödel set theory** (BG), respectively.

B. Zermelo-Fraenkel Set Theory

ZF is a formal system expressed in †first-order predicate logic with the predicate symbol = (equality) and based on the following axioms 1-9 (→ 411 Symbolic Logic). These axioms do not contain any predicate symbol other than ∈, where $x \in y$ is read "x is an element of y." Any formula containing only ∈ as a predicate symbol is called a **set-theoretic formula.** Following the usual convention, we omit the outermost universal quantifier, and use **restricted quantifiers** such as $\exists x \in a$, $\forall x \in a$.

Axiom 1 **(axiom of extensionality):**

$$\forall x(x \in a \equiv x \in b) \rightarrow a = b.$$

This asserts that sets formed by the same elements are equal. The formula $\forall x \in a(x \in b)$ is denoted by $a \subset b$. This means "a is a subset of b." Then Axiom 1 can be expressed by

$$a \subset b \wedge b \subset a \rightarrow a = b.$$

Axiom 2 **(axiom of the unordered pair):**

$$\exists x \forall y(y \in x \equiv y = a \vee y = b).$$

This asserts the existence, for any sets a and b, of a set x having a and b as its only elements. This **x** is called the **unordered pair** of a and b and is denoted by {a, b}. The notion of **ordered pair** is characterized by

$$(a, b) = (c, d) \equiv a = c \wedge b = d.$$

An example of such a set is (a, b) = {{a, a}, {a, b}}.

Axiom 3 **(axiom of the sum set):**

$$\exists x \forall y(y \in x \equiv \exists z \in a(y \in z)).$$

This asserts the existence for any set a of the sum (or union) x of all the sets that are elements of a. This x is denoted by $\cup a$ or S(a). We write $a \cup b$ for $\cup\{a, b\}$ and a' for $a \cup \{a, a\}$.

Axiom 4 **(axiom of the power set):**

$$\forall x \exists y(y \in x \equiv \forall z \in y(z \in a)).$$

This asserts the existence for any set a of the power set x consisting of all the subsets of a. This x is denoted by P(a). We have $S(P(a)) = a$, so S is a left inverse of P, and the products SP and PS are idempotent.

Axiom 5 **(axiom of the empty set):**

$$\exists x \forall y(\neg y \in x).$$

This asserts the existence of the empty set. The empty set is denoted by ∅ or 0.

Axiom 6 **(axiom of infinity):**

$$\exists x(0 \in x \vee \forall y \in x(y' \in x)).$$

This asserts the existence of a set containing all the natural numbers, where $0, 1 = 0' = \{0\}$, $2 = 1' = \{0, 1\}$, $3 = 2' = \{0, 1, 2\}$, This definition of natural number is due to von Neumann.

Axiom 7 **(axiom of separation):**

$$\exists x \forall y(y \in x \equiv y \in a \wedge A(y)).$$

This asserts that the existence for any set a and a formula A(y) of a set x consisting of all elements of a satisfying A(y). This x is denoted by $\{y \in a \mid A(y)\}$. This is rather a schema for an infinite number of axioms, for there are an infinite number of A(y). This axiom, also called the **axiom of comprehension** or **axiom of subsets,** was introduced by Zermelo.

For example, the set of all natural numbers is introduced by

$$\{x \in a \mid \forall y(0 \in y \wedge \forall z \in y(z' \in y) \rightarrow x \in y)\},$$

where a is a set satisfying Axiom 6. The set of all natural numbers is denoted by ω or N.

Axiom 8 **(axiom of replacement):**

$$\exists x \forall y \in a(\exists z A(y, z) \rightarrow \exists z \in x A(y, z)).$$

This asserts the existence for any set a of a set x such that for any y of a, if there exists a z satisfying A(y, z) then such z exists in x. If the relation A(y, z) determines a function, then the image of a set by the relation is included in a set, so by Axiom 7, the image is a set. This axiom was introduced by Fraenkel.

Axiom 9 **(axiom of regularity):**

$$\exists x A(x) \rightarrow \exists x(A(x) \wedge \forall y \in x(\neg A(y))).$$

This asserts that if there exists a set satisfying A(x), then there is a set x satisfying A(x) but every element y of x does not satisfy the property $A(y)$. This axiom is also called the **axiom of foundation.** Its dual expression, called ∈-induction, is of course equivalent to Axiom 9.

Axiom 9' **(axiom of ∈-induction):**

$$\forall x(\forall y \in x A(y) \rightarrow A(x)) \rightarrow \forall x A(x).$$

Using this we can show that $\neg x \in x$, $\neg(x \in y \wedge y \in x)$, etc. If we assume the axiom of choice stated below, then this is equivalent to the nonexistence of an infinite descending sequence

$$x_n \in \ldots \in x_2 \in x_1.$$

If a model of set theory satisfies the axiom of regularity and has an infinite descending sequence that is not in the model, then such a model is called a nonstandard model.

Axiom 10 **(axiom of choice):**

$$\forall x \in a \exists y A(x, y) \rightarrow \exists y \forall x \in a A(x, y(x)).$$

This asserts that if for any element x of a there is a set y such that $A(x, y)$, then there is a **choice function** y for the formula, i.e., $A(x, y(x))$ for all x in a. Usually a function is represented by its graph. A set f is called a function detïned on a if

$$\forall x \forall y((x, y) \in f \rightarrow x \in a), \qquad \forall x \in a \exists y((x, y) \in f),$$

$$\forall x \in a \forall y \forall z((x, y) \in f \wedge (x, z) \in f \rightarrow y = z).$$

This formula is denoted by $\mathrm{Fnc}(f)$; then the formula $A(x, f(x))$ is an abbreviation of

$$\mathrm{Fnc}(f) \wedge \exists y((x, y) \in f \wedge A(x, y)).$$

The axiom of choice is equivalent to many properties, such as the well-ordering theorem, Zorn's lemma, and Tikhonov's theorem on the product of compact spaces, and it is used widely and essentially in mathematics.

The system of axioms 1 to 9 is called **Zermelo-Fraenkel set theory** and is denoted by **ZF;** the system ZF minus the axiom of replacement is called **Zermelo set theory,** denoted by **Z;** and the system ZF plus the axiom of choice is denoted by **ZFC.**

The system Z is weaker than ZF. Indeed, the existence of the set ω of all the natural numbers and of P(ro), $P(P(\omega))$, can be proved in Z, but the existence of the set $\{\omega, p(\omega), P(P(\omega)), .\}$ cannot be proved in Z. However, we can prove its existence in ZF.

The theory ZF minus the axiom of intïnity is called **general set theory.** Its consistency can be reduced to the consistency of the theory of natural numbers as follows. Let n be any natural number and $n = 2^{n_1} + 2^{n_2} + \ldots + 2^{n_k} (n_1 < n_2 < \ldots < n_k)$ be its binary expansion. We can define a model of general set theory by identifying a set with the natural number n and defining $m \in n$ by "m appears as one of the n_i." In general set theory we can show the existence of 0, $P(O)$, $P(P(0))$, but the existence of the set $\{0, P(O), P(P(0)), .\}$ cannot be proved. However, we can prove its existence in Z. The set consisting of all **hereditary** finite sets is the smallest model of general set theory.

C. Bernays-Gödel Set Theory

The existence of $\{u \mid A(u)\}$ for an arbitrary set-theoretic formula .4(u) cannot be deduced from the axioms of ZF. We call this object a **class** to distinguish it from sets. We introduce a generalized logical system of first-order predicate logic by adding to the logical system used in ZF class variables, class constants, and inference rules with respect to quantifiers for classes. For any set-theoretic formula $A(u)$ in which no †bound class variable occurs, we adopt

$$\exists X \forall u(u \in X \leftrightarrows A(u))$$

as an axiom, where capital letters X are class variables. The set theory thus obtained is equivalent to **Gödel set theory** [11]. Von Neumann axiomatized set theory by making use of the notion of functions instead of that of classes [30]. **In** refïning this theory and introducing the notion of classes, Bernays and Gode1 [2, 11] initiated Bernays-Gödel set theory, BG.

ZF and BG are related as follows: Any formula provable in ZF is provable in BG, and any set-theoretic formula provable in BG and having neither class variable nor class constant is provable in ZF. In this sense, the systems can be regarded as essentially equivalent, but as BG has class variables and class constants, it is more convenient for expressing set-theoretic notions.

Von Neumann detïned the following function R by †transfinite induction:

$$R(0) = \varnothing, \qquad R(\alpha) = \bigcup_{\beta < \alpha} \mathfrak{P}(R(\beta)),$$

where α and β are ordinal numbers and $\bigcup_{\beta<\alpha} \mathfrak{P}(R(\beta))$ denotes the set sum of $\mathfrak{P}(R(0))$, $\mathfrak{P}(R(1))$, ... $\mathfrak{P}(R(\beta))$, ... $(\beta < \alpha)$. The function R can be detïned by a set-theoretic formula, so it exists as a class. Now consider the mode1 $M = M(\alpha)$ for a fixed ordinal number α. We define sets of the model M as elements of $R(\alpha)$, and classes of the model as subsets of $R(\alpha)$. We denote the $\in$ relation of the model by $\in_M$. For classes X and Y of the model, we write $X \in_M Y \rightleftarrows X \in Y$ Then a necessary and sufficient condition for $R(\alpha)$ to be a mode1 of BG is that α is an †inaccessible ordinal number (→ 3 12 Ordinal Numbers). The existence of an inaccessible ordinal number cannot be deduced from the axioms of ZF. There is a series of studies of axiomatization of set theory in which any number of inaccessible ordinal numbers is assumed to exist [23]. When $R(\alpha)$ is a model of BG (ZF), it is called a **natural model** of BG (ZF). Furthermore, if H is $\bigcup_\alpha R(\alpha)$ then H satisfies all the axioms of BG (ZF). As we do not need the axiom of regularity for defining the class H, we see that BG (ZF) is consistent as long as BG (ZF) without the axiom of regularity is consistent.

D. Independence of the Continuum Hypothesis and the Axiom of Choice

These axiomatizations of set theory motivated a series of studies from the standpoint of the

†foundations of mathematics on problems that remained unsolved after the appearance of Cantor's primitive set theory. Among these, the problem of the relation between the continuum hypothesis and the axiom of choice was central.

Consistency of the Axiom of Choice and the Continuum Hypothesis. Gödel [11] proved that if ZF without the axiom of choice is consistent, then the system obtained by adding to ZF the axiom of choice and the continuum hypothesis is also consistent. To show this, he constructed a model of ZF satisfying the axiom of choice and the generalized continuum hypothesis as follows. Assume first that M is an arbitrary domain of objects among which the $\in$ relation is defined. By a formula on M, we understand a formula having constants of M as its only constants, having $\in$ as its sole predicate symbol, and further having exclusively variables whose ranges are restricted to M. Let us denote by M' the totality of subsets of M defined by a formula A(x) on M. Now we put $M_0=\{\varnothing\}$, $M_{\alpha+1}=M'_\alpha$, $M_\beta=\bigcup_{\alpha<\beta}M_\alpha$, if β is a †limit ordinal number. We call **x constructible** if $x\in M_\alpha$ for some ordinal number α, assumed to be less than the first inaccessible ordinal number, if any. We denote the totality of constructible sets by L, and the totality of sets of ZF by V. We call the assertion $V=L$, that is, every set is constructible, the **axiom of constructibility.** If we add this axiom to ZF, the axiom of choice and the generalized continuum hypothesis become provable. On the other hand, if we regard elements of L as sets of the model and the original $\in$ relation as the $\in$ relation of the model, we have a model of ZF in which the axiom of constructibility holds.

Independence of the Axiom of Choice and tbe Continuum Hypothesis. Since the result of Gödel, attempts have been made to prove that the axiom of choice is independent of the other axioms. Fraenkel constructed a mode1 of set theory without satisfying the axiom of choice, starting from a countable number of objects that are not sets. A. Mostowski constructed in ZF a mode1 of set theory having objects that are not sets, and he proved that the model satisfies the axiom: Every set can be †linearly ordered, but does not satisfy the axiom of choice [27]. E. Mendelson constructed a mode1 of set theory that does not satisfy the axiom of choice by making use of an infinite descending chain $a_1\ni a_2\ni a_3\ni\dots$ [25]. These models, however, do not satisfy all the axioms of ZF or of Zermelo set theory minus the axiom of choice, even though they satisfy most of the axioms. Consequently, they are not sufficient for proving independence.

P. J. Cohen [3,4] proved the following results in connection with the independence of the axiom of choice, the continuum hypothesis, and $V=L$: If ZF is consistent, each of the following conditions has a model. (1) The axiom of choice as well as the generalized continuum hypothesis holds, but there exists an a satisfying $a\notin L$ and $a\subset\omega$. (2) $\mathfrak{P}(\omega)$ is not well-ordered. (3) The axiom of choice holds, but the continuum hypothesis does not hold. (4) $\mathfrak{P}(\mathfrak{P}(\omega))$ cannot be linearly ordered.

By (1), we see that $V=L$ is independent of the axiom of choice and the generalized continuum hypothesis: (2) shows the independence of the axiom of choice; and (3) shows that the continuum hypothesis is independent of the axiom of choice. In (4), $\mathfrak{P}(\mathfrak{P}(\omega))$ corresponds to the set F of all real-valued functions defined on the interval [0, 1], and (4) shows that the proposition "F can be linearly ordered" is not deducible in ZF without the axiom of choice.

E. Some Recent Results

Boolean-Valued Set Theory. D. Scott and R. M. Solovay defined models of set theory in which the set-theoretic formula has values in a complete Boolean algebra. This viewpoint is motivated by P. J. Cohen's original notions of the forcing relation and the generic filter.

According to the relation between sets and their representing functions, a function f: A+ B corresponds to a subset of A when B is a complete Boolean algebra. So by transfinite induction we put

$$V_\alpha^{(B)}=\bigcup_{\beta<\alpha}P^{(B)}(V_\beta^{(B)}),$$

where $P^{(B)}(A)$ is the Boolean-valued power set of A defined by

$$u\in P^{(B)}(A)\equiv u:A\to B.$$

Let $V^{(B)}$ be the union of all $V_\alpha^{(B)}$'s; then the truth values of the formulas $[\![u\in v]\!]$ and $[\![u=v]\!]$ are defined by transfinite induction as

$$[\![u\in v]\!]=\sum_{x\in\mathrm{dom}(v)}v(x)\cdot[\![x=u]\!],$$

$$[\![u=v]\!]=\prod_{x\in\mathrm{dom}(u)}u(x)\Rightarrow[\![x\in v]\!]\cdot\prod_{x\in\mathrm{dom}(v)}v(x)\Rightarrow[\![x\in u]\!],$$

where $a\Rightarrow b$ is an abbreviation for the element $-a+b$ in B.

In $V^{(B)}$ the following properties are satisfied:
(i) The truth value of any formula provable from ZFC is 1, the largest element of B.
(ii) Any complete homomorphism $h: B_1\to B_2$

can be extended to

$$h: V^{(B_1)} \to V^{(B_2)},$$

so that the commutative relation

$$h([\![A(u_1, \ldots, u_n)]\!]) = [\![A(h(u_1), \ldots, h(u_n))]\!]$$

is satisfied for every restricted formula; further, if h is an epimorphism then it is satisfied for every formula.

(iii) (maximum principle) For any formula $A(x)$ there is an element u of $V^{(B)}$ such that

$$[\![\exists x A(x)]\!] = [\![A(u)]\!].$$

Moreover, if $[\![\exists x R(x)]\!] = 1$ then there an element u of $V^{(B)}$ such that $[\![R(u)]\!] = 1$ and

$$[\![\exists x (R(x) \wedge A(x))]\!] = [\![A(u)]\!].$$

Property (i) means that if $[A] < 1$, then A is not provable from ZFC. Many consistency results in set theory are obtained in this way by constructing special partial order structures, topological spaces, or complete Boolean algebras.

Axiom of Strong Infinity. The axiom of infinity, Axiom 6, of ZF asserts that there exists a set including the set ω of all natural numbers. The set ω is the smallest infinite cardinal number. In general, the axiom of strong intinity is stated in a form such that certain special properties of ω are satistïed for an uncountable cardinal number.

Though such an axiom asserts the existence of large cardinal numbers, their properties sometimes reflect, for example, those of the real numbers, or of the set of real numbers, etc.

We state here some typical examples:

(i) **Weakly inaccessible cardinal numher.** This is a regular limit cardinal number, that is,

$$\mathrm{cf}(\omega_\alpha) = \omega_\alpha, \qquad \forall \beta < \alpha(\omega_{\beta+1} < \omega_\alpha).$$

(ii) **Strongly inaccessible cardinal number.**

$$\mathrm{cf}(\omega_\alpha) = \omega_\alpha, \qquad \forall \beta < \alpha(2^{\omega_\beta} < \omega_\alpha).$$

(iii) **Weakly compact cardinal number.** ω_α is uncountable and the space $X = 2^{\omega_\alpha}$ with $< \omega_\alpha$-topology is ω_α-compact.

(iv) **Measurable cardinal number.** ω_α is uncountable and there is a nonprincipal $< \omega_\alpha$-additive 2-valued measure $\mu: P(\omega_\alpha) \to \{0, 1\}$.

(v) Strongly compact cardinal number. ω_α is uncountable and any product of $2 = \{0, 1\}$ with $< \omega_\alpha$-topology is ω_α-compact.

The strength of these properties increases in the order presented here. Properties (i) to (iii) are compatible with the axiom of constructibility $V = L$. And (iv) implies $V \neq L$ while it is compatible with $V = L(\mu)$. But (v) implies $V \neq L(a)$ for any set a.

There are many other stronger axioms of infinity, and the consistency proofs of these "axioms" within ZF or ZFC would involve essential difficulties. Inconsistency proofs of these properties, if they exist, would be very interesting.

F. Examples of Results

(1) Cardinality and Cofinality (→ 312 Ordinal Numbers). (i) (W. B. Easton [6]) Let $\mathfrak{M}$ be a mode1 of ZFC (the Zermelo-Frankel axioms plus the axiom of choice) in which the GCH (generahzed continuum hypothesis) is valid, i.e., $\forall \alpha(2^{\omega_\alpha} = \omega_{\alpha+1})$, and let g be a function from ordinals to ordinals in $\mathfrak{M}$ such that $\forall \alpha, \beta(\alpha < \beta \Rightarrow \omega_{g(\alpha)} \leqslant \omega_{g(\beta)})$ and $\forall \alpha(\omega_\alpha < \mathrm{cf}(\omega_{g(\alpha)}))$. Then there is a Boolean mode1 $\mathfrak{N}$ of ZFC, $\mathfrak{N} \supset \mathfrak{M}$, having the same cofinality and satisfying $2^{\omega_\alpha} = \omega_{g(\alpha)}$ for every regular cardinal. (This means that König's condition $(\mathrm{cf}(2^{\omega_\alpha}) > \omega_\alpha)$ is the only restriction on the cardinality of powers of regular ω_α.)

(ii) (J. H. Silver [39]) Suppose

$$\omega < \mathrm{cf}(\alpha) = \mathrm{cf}(\omega_\alpha) < \omega_\alpha.$$

Then for any $\lambda < \mathrm{cf}(\alpha)$

$$\forall \nu < \alpha(2^{\omega_\nu} \leqslant \omega_{\nu+\lambda}) \to 2^{\omega_\alpha} \leqslant \omega_{\alpha+\lambda}.$$

However, the validity of the implication

$$\forall n < \omega(2^{\omega_n} = \omega_{n+1}) \to 2^{\omega_\omega} = \omega_{\omega+1}$$

still remains an open question. And it is known that if there is a model for

$$\forall n < \omega(2^{\omega_n} = \omega_{n+1}), \qquad 2^{\omega_\omega} > \omega_{\omega+1},$$

then there is a model in which there are many measurable cardinal numbers.

(iii) (K. Prikry [31]) Let ω_α be a measurable cardinal number in a model M of ZFC. Then there is a Boolean extension N of M in which the notion of cardinality is not changed but the notion of cofinality is changed, that is $\mathrm{cf}(\omega_\alpha) = \omega_\alpha$ in M but $\mathrm{cf}(\omega_\alpha) = \omega$ in N.

(iv) (Solovay) Let ω_α be a strongly compact cardinal number. Then for any strong-limit singular cardinal $\omega_\beta > \omega_\alpha$ the continuum hypothesis holds at ω_β; that is, the implication

$$\mathrm{cf}(\omega_\beta) < \omega_\beta, \qquad \forall \gamma < \beta(2^{\omega_\gamma} < \omega_\beta) \to 2^{\omega_\beta} = \omega_{\beta+1}$$

is provable.

(2) Lebesgue Measurability and the Baire Property. As is well known, every Δ_1^1 (Borel) set (and consequently every Σ_1^1 (analytic) set) of real numbers is Lebesgue measurable and has the Baire property.

(i) (Gödel) $V = L$ implies the existence of a Δ_2^1 set of real numbers that is neither Lebesgue measurable nor has the Baire property.

Let DC denote the **principle of depending choice:**

$$\forall x\in a\exists y\in aP(x,y) \rightarrow\exists f:\omega\rightarrow a\forall n\in\omega P(f(n),f(n+1)).$$

DC is adequate for the development of the classical notions of measure theory, such as Jordan decomposition, the Radon-Nikodym derivate, etc. Let I denote the hypothesis $\exists\alpha(\mathrm{cf}(\omega_\alpha)=\omega_\alpha \wedge \forall\beta<\alpha(2^{\omega_\beta}<\omega_\alpha))$ **(strongly inaccessible).**

(ii) (Solovay [40]) The consistency of ZFC and 1 implies that of ZF and either of the following two axioms:

(a) DC plus the hypothesis that every set of real numbers is Lebesgue measurable and has the Baire property.

(b) The axiom of choice plus the hypothesis that every set of real numbers definable by an ω sequence of ordinals is Lebesgue measurable and has the Baire property.

Axiom (a) means that DC is not strong enough to construct a (Lebesgue) nonmeasurable set, while Axiom (b) implies that every projective set is Lebesgue measurable and has the Baire property, and hence implies that the set of all constructible real numbers, as a Σ^1_2 set, has Lebesgue measure 0, and is of the first category.

(3) Martin's Axiom. Let B be a Boolean algebra. We say that B satisfies ω_α C.C. (chain condition) if the cardinality of every disjoint family of positive elements of B is at most ω_α. Let B^* denote the topological space consisting of all homomorphisms $h: B\rightarrow 2=\{0,1\}$ with the open base $U(a)=\{h\mid h(a)=1\}$ $(a\in B)$; then B^* is a †Baire space. Then **Martin's axiom** (MA) is: Let B be an ω C.C. Boolean algebra and $\alpha<2$". Then the intersection of a dense open sets is dense in B^*. Since $\{h \ \Sigma_\nu h(a_\nu)=$ h(b)} is dense and open in B^* if $\Sigma_\nu a_\nu=b$, MA means the existence of an $h\in B^*$ preserving any given set of α (< 2") equations in B, If 2" = ω_1, MA merely reduces to the Baire property of B^*. However, if 2" > ω_1, then the ω C.C. hypothesis is essential, for there exists a B satisfying ω_1 C.C. such that B^* contains ω_1 dense open sets with empty intersection.

(i) (Solovay and S. Tennenbaum [43]) The consistency of ZF implies that of ZFC, MA, and $2^\omega>\omega_1$.

(ii) (D. A. Martin and Solovay [24]) ZFC, MA, and 2" > ω_1 imply the propositions:
(a) $\forall\omega_\alpha<2^\omega(2^{\omega_\alpha}=2^\omega)$, hence 2" = 2^{ω_1};
(b) the totality of the first category sets of Lebesgue measure zero sets is a-additive for any $\alpha<2$";
(c) every Σ^1_1 set of real numbers is Lebesgue measurable and has the Baire property.

(4) Suslin's Hypothesis (SH) is: Every dense, linear, order complete set without end points, having at most ω disjoint intervals, is order isomorphic to the continuum of real numbers.

(i) (T. J. Jech, Tennenbaum) The consistency of ZF implies that of (a) ZFC, ¬ SH, and GCH, as well as (b) ZFC, ¬ SH, and 2" > ω_1

(ii) (R. Jensen [15]) The consistency of ZF implies that of ZFC, SH, and 2" = ω_1.

(iii) (Solovay and Tennenbaum [43]) ZFC, MA, and 2" > ω_1 imply SH.

(iv) (Jensen [15]) $V=L$ implies ¬ SH.

(5) Measurahle and Real-Valued Measurahle Cardinals. A cardinal $\kappa>\omega$ is said to be **measurahle** if there is a measure $\mu:\mathfrak{P}(\kappa)\rightarrow\{0,1\}$ with (a) $\mu(\kappa)=1$, (b) $\forall\nu<\kappa(\mu(\{\nu\})=0)$ and (c) $\mu(\Sigma_{\nu<\alpha}A_\nu)=\Sigma_{\nu<\alpha}\mu(A_\nu)$; and κ is said to be **real-valued measurahle** if there exists a $\mu:\mathfrak{P}(\kappa)\rightarrow[0,1]$ that satifies (a), (b) and (c), and is not measurable.

Let MC (RMC) denote the existence of a measurable cardinal (real-valued measurable cardinal).

(i) (S. Ulam [45]) The existence of an ω-additive measure $\mu:\mathfrak{P}(A)\rightarrow[0,1]$, with $\mu(A)=1$ and $\forall x\in A(\mu(\{x\})=0)$, implies RMC or MC; RMC implies the existence of an extension of Lebesgue measure defined on $\mathfrak{P}([0,1])$; every real-valued measurable cardinal is $\leqslant$ 2"' and weakly inaccessible.

(ii) Every measurable cardinal is strongly (hyper) inaccessible, and (a) $\exists\alpha F(\alpha)\rightarrow\exists\alpha<\omega_1$ *F(a)* for any Σ^1_1 formula F(a) on the ordinal numbers, (b) $\exists\alpha F(\alpha)\rightarrow\exists\alpha<\mu_1 F(\alpha)$ for any Π^1_1 formula F(a) and the smallest measurable μ_1. (Many results have been obtained concerning the ordinal magnitude of ω_1 and the measurable cardinals.)

(iii) (Solovay [41]) The consistency of ZFC and MC is equivalent to that of ZFC and RMC.

(iv) (Martin and Solovay [24]) ZFC, RMC, and MA are not consistent.

(v) (Lévy and Solovay [21]) The consistency of ZFC and MC implies that of ZFC, MC, MA, and $2^\omega>\omega_2$.

(vi) (Solovay) ZFC and MC imply that every Σ^1_2 set of real numbers is Lebesgue measurable.

(vii) (Martin, Solovay) ZFC, MC, MA, and 2" > ω_2 imply that every Σ^1_3 set of real numbers is Lebesgue measurable and has the Baire property.

(viii) (Silver [38]) The consistency of ZFC and MC implies that of ZFC and MC as well as the existence of (Lebesgue) non-measurable Δ^1_3 sets of real numbers.

(6) Axiom of Determinateness. We consider the following inlïnite game. Let A be a set consist-

ing of functions ω to ω. The players I and II choose alternatively natural numbers n_i and m_i, and, for the resulting infinite sequence, if

$$(n_1, m_1, n_2, \ldots, n_k, m_k, \ldots) \in A,$$

then 1 wins, otherwise II wins.

A function $\sigma: \bigcup_{n\in\omega} \omega^n \to \omega$ is called a **strategy.** If II plays according to the sequence $\underline{m} = (m_1, m_2, \ldots, m_k, \ldots)$ and I plays according to the strategy σ, then the resulting sequence is

$$\sigma * \underline{m} = (\sigma(\emptyset), m_1, \sigma(m_1), \ldots, m_k, \sigma(m_1, \ldots, m_k), \ldots).$$

The strategy σ is called a **winning strategy** for I of the game defined above, if for any play $\underline{m}$ of II, the resulting play $\sigma * \underline{m}$ is in the set A, and similarly for the winning strategy of II.

The axiom of determinateness AD is the assertion that for any $A \subset \omega^\omega$, the player I or II has a winning strategy.

(i) (J. Mycielski [28]) Assume ZF and AD; then we have:
(a) axiom of choice for countable sequence of sets of real numbers;
(b) every set of real numbers is Lebesgue measurable and has the Baire property, so the axiom of choice does not hold;
(c) every uncountable set of real numbers contains a perfect set.

(ii) (Solovay) Assume ZF and AD; then we have:
(a) cardinalities of ω_1 and 2” are not comparable;
(b) ω_1 and ω_2 are measurable cardinals, but ω_3 is a singular cardinal such that $cf(\omega_3) = \omega_2$.

(iii) (Martin) In ZFC, every Borel, namely Δ_1^1, game is determined.

(iv) (Martin, Solovay) In ZFC:
(a) if there exists a measurable cardinal, then every analytic, namely Π_1^1, game is determined;
(b) if every Δ_2^1-game is determined, then there is a model of ZFC with many measurable cardinal numbers.

Since the axiom of determinateness AD is very strong, there is some suspicion that it might contradict ZF. But at present there is no such evidence.

References

[1] J. Bell, Boolean-valued models and independence proofs in set theory, Clarendon Press, 1977.

[2] P. Bernays, A system of axiomatic set theory I, J. Symbolic Logic, 2 (1937), 65577.

[3] P. J. Cohen, The independence of the continuum hypothesis I, II, Proc. Nat. Acad. Sci. US, 50 (1963), 1143-1148; 51 (1964), 105–110.

[4] P. J. Cohen, Set theory and the continuum hypothesis, Benjamin, 1966.

[5] F. R. Drake, Set theory, North-Holland, 1974.

[6] W. B. Easton, Power of regular cardinals, Ann. Math. Logic, 1 (1970), 1399178.

[7] G. Föder, On stationary sets and regressive functions, Acta Sci. Math., 27 (1966), 1055 110.

[S] A. Fraenkel, Zu den Grundlagen der Cantor-Zermeloschen Mengenlehre, Math. Ann., (1922), 230–237.

[9] H. Gaifman, Uniform extension operators for models and their applications, Sets, Models and Recursion Theory, North-Holland, 1967, 1222155.

[10] F. Galvin and A. Haynal, Inequalities for cardinal powers, Ann. Math., (2) 101 (1975), 491-498.

[11] K. Gödel, The consistency of the axiom of choice and of the generalized continuum hypothesis with the axioms of set theory, Ann. Math. Studies, Princeton Univ. Press, 1940.

[12] K. Gödel, What is Cantor’s continuum problem?, Amer. Math. Monthly, 54 (1947), 5155525.

[133 T. Jech, Non-provability of Suslin’s hypothesis, Comm. Math. Univ. Carolinae, 8 (1972), 2299308.

[14] T. Jech, Set theory, Academic Press, 1978.

[15] R. B. Jensen, The fine structure of the constructible hierarchy, Ann. Math. Logic, 4 (1972), 2299308.

[16] A. Kanamori and M. Magidor, The evolution of large cardinal axioms in set theory, Lecture notes in math. 669, Springer, 1978, 99-275.

[17] H. J. Keisler and A. Tarski, From accessible to inaccessible cardinals, Fund. Math., 53 (1965), 2255308; 57 (1965), 119.

[18] K. Kunen, Inaccessibility properties of cardinals, Thesis, Stanford Univ.

[19] K. Kunen and J. B. Paris, Boolean extensions and measurable cardinals, Ann. Math. Logic, 2 (1971), 3599378.

[20] A. Lévy, Basic set theory, Perspectives in mathematical logic, Springer, 1979.

[21] A. Lévy, and R. M. Solovay, Measurable cardinals and the continuum hypothesis, Israel J. Math., 5 (1967), 2344248.

[22] M. Magidor, How large is the first strongly compact cardinal?, Ann. Math. Logic, 10 (1976), 33-57.

[23] P. Mahlo, Zur Theorie und Anwendung der ρ_0-Zahlen, Ber. Verh. Königlish Säch. Ges. Wiss. Leipzig. Math. Phys. Kl., 63 (1912), 108-112; 65 (1913), 2688282.

[24] D. A. Martin and R. M. Solovay, Internal Cohen extensions, Ann. Math. Logic, 2 (1970), 143–178.

[25] E. Mendelson, The independence of weak

axiom of choice, J. Symbolic Logic, 21 (1956), 350-366.
[26] Y. N. Moschovakis, Determinacy and prewellordering of the continuum, Mathematical logic and foundations of set theory, North-Holland, 1970, 24462.
[27] A. Mostowski, Über die Unabhangigkeit des Wohlordnungssatzes vom Ordnungsprinzip, Fund. Math., 32 (1939), 201-252.
[28] J. Mycielski, On the axiom of determinateness, Fund. Math., 53 (1964), 205–224.
[29] K. Namba, An axiom of strong infinity and analytic hierarchy of ordinal numbers, Proc. Symp. Pure Math., 13 (1971), 279-319.
[30] J. v. Neumann, Die Axiomtisierung der Mengenlehre, Math. Z., 27 (1928), 6699752.
[31] K. L. Prikry, Changing measurable into accessible cardinals, Diss. Math., 68 (1970), 5–52.
[32] J. B. Rosser, Boolean-valued models for set theory, Academic Press, 1969.
[33] G. E. Sacks, Forcing with perfect closed sets, Proc. Symp. Pure Math., 13 (1971), 331–355.
[34] D. S. Scott, Measurable cardinals and constructible sets, Bull. Acad. Polon. Sci., Sér. Math. Astr. Phys., 7 (1961), 145-149.
[35] D. S. Scott and R. M. Solovay, Boolean-valued models for set theory, Lecture notes, Summer Institute on Axiomatic Set Theory, UCLA, 1967.
[36] S. Shelah, A compactness theorem for singular cardinals, free algebras, Whitehead problem and transversals, Israel J. Math., 21 (1975), 319-349.
[37] J. H. Silver, Some applications of mode1 theory in set theory, Ann. Math. Logic, 3 (1971), 45-110.
[38] J. H. Silver, The consistency of the GCH with the existence of a measurable cardinal, Proc. Symp. Pure Math., 13 (1971), 391-396.
[39] J. H. Silver, On the singular cardinal problem, Proc. Int. Congr. Math., Vancouver, 1 (1974), 2655286.
[40] R. M. Solovay, A mode1 of set theory in which every set of reals is Lebesgue measurable, Ann. Math., (2) 92 (1970), l-56.
[41] R. M. Solovay, Real-valued measurable cardinals, Proc. Symp. Pure Math., 13 (1971), 4977428.
[42] R. M. Solovay, Strongly compact cardinal and the G.C.H., Proc. Symp. Pure Math., 25 (1974), 3655372.
[43] R. M. Solovay and S. Tennenbaum, Iterated Cohen extensions and Suslin's problem, Ann. Math., (2) 94 (1971), 201–245.
[44] G. Takeuti and W. B. Zaring, Axiomatic set theory, GTM 8, Springer, 1973.
[45] S. Ulam, Zur Masstheorie in der algemainen Mengenlehre, Fund. Math., 16 (1930), 140–150.
[46] P. Vopenka and K. Hrbacek, On strongly measurable cardinals, Bull. Acad. Polon. Sci., Sér. Sci. Math. Astr. Phys., 14 (1966), 587–591.
[47] E. Zermelo, Untersuchungen über die Grundlagen der Mengenlehre I, Math. Ann., 65 (1908), 261–281.

34 (11.5) Axiom of Choice and Equivalents

A. The Axiom of Choice

In set theory, the following axiom is known as the **axiom of choice:** For any nonempty family $\mathfrak{A}$ of nonempty subsets of a set X, there exists a single-valued function *f*, called the **choice function**, whose domain is $\mathfrak{A}$, such that for every element A of $\mathfrak{A}$ the value f(A) is a member of *A*. This axiom is equivalent to each of the following three propositions: (1) If $\{A_\lambda\}_{\lambda\in\Lambda}$ is a family of sets not containing the empty set and with an index set Λ, then the Cartesian product $\prod_\lambda A_\lambda$ is not the empty set. (2) If a set *A* is a disjoint union $\bigcup_\lambda A_\lambda$ of a family of subsets $\{A_\lambda\}_{\lambda\in\Lambda}$ which does not contain the empty set, there exists a subset *B*, called the **choice set,** of *A* such that every intersection of *B* and $A_\lambda (\lambda\in\Lambda)$ contains one and only one element. (3) For every mapping *f* from a set *A* onto a set *B*, there is a mapping from *B* to *A* such that $f\circ g = 1$, (identity mapping). Also equivalent to the axiom of choice are the well-ordering theorem and Zorn's lemma, which are discussed in the following sections.

B. The Well-Ordering Theorem

In 1904, E. Zermelo [2] first stated the axiom of choice and used it for his proof of the **well-ordering theorem,** which says that every set can be †well-ordered by an appropriate †ordering. Conversely, the well-ordering theorem implies the axiom of choice. Many important results in set theory can be obtained by using the axiom of choice, for example, that †cardinal numbers are comparable, or that various definitions of the finiteness or infiniteness of sets are equivalent. Various important theorems outside of set theory, e.g., the existence of bases in a †linear space, †compactness of the direct product of compact †topological spaces (†Tikhonov's theorem), the existence of a subset which is not †Lebesgue measurable in Euclidean space, etc., are proved using the axiom of choice. But for those proofs the well-ordering theorem or Zorn's lemma (stated

below) are used more often than the axiom of choice.

Using the well-ordering theorem, it can be proved in the following manner that for every field k, any linear space X over k has a basis. Let $\{x_\nu\}_{\nu\in\Lambda}$ be an enumeration of X. By †transfinite induction, we can define the function f from X to $B=\{0,1\}$ such that (i) if $x_0=0$, $f(x_0)=0$; if $x_0\neq 0$, $f(x_0)=1$; (ii) for $\nu>0$, if x_ν is expressed as a linear combination of elements of $\{x_\mu \mid \mu<\nu, f(x_\mu)=1\}$, $f(x_\nu)=0$; otherwise $f(x_\nu)=1$. Then $U=\{x\in X \mid f(x)=1\}$ is a basis of X.

C. Zorn's Lemma

An ordered set **X** is called an **inductively ordered set** if every †totally ordered subset of X has an †upper bound. A condition C for sets is called a **condition of finite character** if a set X satisfies C if and only if every finite subset of X satisfies C. A condition C for functions is called a **condition of finite character** if C is a condition of finite character for the graph of the function. **Zorn's lemma** [4] can be stated in any one of the following ways, which are all equivalent to the axiom of choice. It is often more convenient to use than the axiom of choice or the well-ordering theorem.

(1) Every inductively ordered set has at least one maximal element.

(2) If every well-ordered subset of an ordered set M has an upper bound, then there is at least one maximal element in M.

(3) Every ordered set A4 has a well-ordered subset W such that every upper bound of M belongs to W.

(4) For a condition C of finite character for sets, every set X has a maximal (for the relation of the inclusion) subset of X that satisfies C.

(5) Let C be a condition of finite character for functions from X to Y. Then, in the set of functions that satisfy C, there is a function whose domain is maximal (for the relation of the inclusion).

Using Zorn's lemma (1), we can prove again in the following way any linear space X over a field k has a basis. Let $\mathfrak{A}$ be the set of all nonempty subsets A of X such that arbitrary finite subsets of A are linearly independent over k. $\mathfrak{A}$ is not empty. If we order $\mathfrak{A}$ by the relation of inclusion, then $\mathfrak{A}$ is an inductively ordered set. By Zorn's lemma (1), there is a maximal element U of $\mathfrak{A}$. Since U is maximal, U is a basis of X.

The same theorem is proved as follows using Zorn's lemma (4). Condition C for the subset A of X, that arbitrary finite subsets of A are linearly independent over k, is a condition of finite character. Hence there is a maximal subset U that satisfies C and is a basis of X.

Concerning recent developments regarding the axiomatic basis for the axiom of choice → 33 Axiomatic Set Theory D.

References

[1] G. Cantor, Über unendliche, lineare Punktmannigfaltigkeiten, Math. Ann., 21 (1883), 545-591. (Gesammelte Abhandlungen, Springer, 1932.)

[2] E. Zermelo, Beweis, dass jede Menge wohlgeordnet werden kann, Math. Ann., 59 (1904), 5144516.

[3] E. Zermelo, Neuer Beweis für die Möglichkeit einer Wohlordnung, Math. Ann., 65 (1908), 1077128.

[4] M. Zorn, A remark on method in transfinite algebra, Bull. Amer. Math. Soc., 41 (1935), 6677670.

[5] J. W. Tukey, Convergence and uniformity in topology, Ann. Math. Studies, Princeton Univ. Press, 1940.

[6] N. Bourbaki, Eléments de mathématique I, Théorie des ensembles, ch. 3, Actualités Sci. Ind., 1243, Hermann, second edition, 1966; English translation, Theory of sets, Addison-Wesley, 1968.

35 (1.2) Axiom Systems

A. History

A mathematical theory is based on a specific system of axioms, i.e., a system of hypotheses from which the whole theory is deduced without reliance on other assumptions.

One of the first deductive methods of mathematical reasoning was utilized by Thales, who returned to Greece from Egypt with the knowledge of surveying methods, and who deduced additional results from that empirical knowledge. His method gave impetus to the development of Greek geometry, which flowered with the Pythagorean school and research by members of Plato's Academy. In the course of this development, the deductive method led to the idea of constructing the whole theory upon a system of "absolutely obvious" statements from which the whole theory could be deduced. Euclid systematized Greek geometry in his *Elements* utilizing this idea. His work became the basis of geometry after the Renaissance, and Greek geometry came to be called †Euclidean geometry. In the *Elements*

Euclid called the basic obvious statements **common notions** when they were of general nature, and **postulates** when they were specifically geometric. Both were later called **axioms** (or **postulates).**

Among the axioms stated by Euclid, the "fifth postulate" concerning parallels was longer and more complicated than the other axioms. Many efforts were made to deduce this particular axiom from the other axioms. The failure of these attempts suggested the possibility of establishing a †non-Euclidean geometry, which was actually done by N. I. Lobachevskii and J. Bolyai, who replaced the fifth postulate by its negation and showed that the new system of "axioms" was as valid as the classical one. This development naturally led to a new evaluation of the idea of axioms, and eventually the traditional concept of recognizing the axioms as obvious truths was replaced by the understanding that they are hypotheses for a theory. D. Hilbert [1] established the latter idea as **axiomatization** and claimed that the whole science of mathematics should be built upon a system of axioms. His idea became the foundation of present-day mathematics. Hilbert reorganized classical geometry based upon his idea and published his result in *Foundations of Geometry* (1899).

B. Systems of Axioms

The **system of axioms** of a theory, i.e., the system of basic hypotheses from which we hope to deduce the whole theory, is written in **undefined terms** (or in terms of **undefined concepts)** by means of which all other terms are defined. On the other hand, a given theory is **axiomatized** by specifying such a system of axioms upon which the theory may be reorganized. It should also be noted that a system of axioms determines a †structure (→ 409 Structures).

A system of axioms is considered to be mathematically valid if and only if it is **consistent.** It is also desirable that the axioms in such a system be mutually **independent** (i.e., the negation of any one of the axioms is still consistent with the others). When such a system is not independent, it can be simplified by deleting redundant axioms from it.

When any two models of a system of axioms are isomorphic to each other, we call the system **complete** or **categorical** (→ 409 Structures). For example, the system of axioms (I)–(V) postulated by Hilbert as the foundation for Euclidean geometry is complete (→ 155 Foundations of Geometry), whereas the systems of axioms for the theories of †groups, †rings, or †fields are not complete since there are nonisomorphic groups, etc. Although it is desirable that the systems of axioms postulated for a given theory (e.g., the †theory of real numbers, or Euclidean geometry) be complete, the study of partial systems that may not be complete is also important (→ 156 Foundations of Mathematics).

References

[1] D. Hilbert, Axiomatisches Denken, Math. Ann., 78 (1918), 405–415. (Gesammelte Abhandlungen III, Springer, 1935, 146–156.)
[2] N. Bourbaki, Eléments de mathématique, Théorie des ensembles, ch. 1-4 and fascicule de résultats, Actualités Sci. Ind., 1212c, 1243b, 1258a, 1141d, Hermann, 1960, 1967, 1966, 1964.

B

36 (XII.17) Banach Algebras

A. Definition [1–4]

A real or complex †Banach space R is called a **Banach algebra** (or **normed ring)** if a multiplication law between elements of R is introduced that makes R an †associative algebra and if $\|xy\| \leqslant \|x\| \cdot \|y\|$ is satisfied. The complex case is generally easier to handle, and a real Banach algebra can always be embedded in a complex one isomorphically and isometrically. When a Banach algebra contains a unity element e with respect to the multiplication, i.e., it is **unital**, we can suppose $\|e\| = 1$. If it is not unital, we can adjoin a unity element to it.

B. Examples

1. Let $C_0(\mathfrak{M})$ be the set of complex- (real-) valued continuous functions $x(\xi)$ on a locally compact space $\mathfrak{M}$ that vanish at infinity. Endowed with the pointwise operations and the †supremum norm (→ 168 Function Spaces), it forms a complex (real) Banach algebra. In the complex case, it has the involution $x^*(\xi)=x(\xi)$.

2. Let X be a Banach space. The set $\mathscr{B}(X)$ of †bounded linear operators on X forms a Banach algebra if we define addition, multiplication by scalars, multiplication between elements, and the norm in the usual fashion.

3. Let G be a locally compact Hausdorff group (→ 423 Topological Groups) and μ be its †left-invariant Haar measure. The Banach space $L_1(G)$ (with respect to μ) can be made into a Banach algebra by defining $xy(g) = \int x(h)y(h^{-1}g)d\mu(h)$ (→ Section L).

4. Every subalgebra of a Banach algebra R, which is closed in the norm topology, is a Banach algebra with respect to the original algebraic operations and norm and is called a **closed subalgebra** of R.

Throughout this article, all Banach algebras are supposed to be complex Banach algebras unless otherwise specified.

C. Spectrum of an Element

We define a new operation $x \cdot y$ in R by setting $x\cdot y=x+y-xy$. If $x\cdot y=y\cdot x=0$, y is called a **quasi-inverse** of x. When a unity element e exists, y is a quasi-inverse of x if and only if e-y is an inverse of e x. For a complex number A such that $|\lambda|>\|x\|$, we see that $\lambda^{-1}x$ possesses a quasi-inverse y given by the strongly convergent series $y= -\sum_{n=1}^{\infty}\lambda^{-n}x^n$. The set of all complex numbers λ such that $\lambda^{-1}x$ does not have a quasi-inverse is called the **spectrum** of x and is denoted by $\mathrm{Sp}(x)$. (When x itself does not have an inverse, in particular, when R does not contain a unity element, we include 0 in this set.) $\mathrm{Sp}(x)$ is a bounded closed set in the complex plane, and $\sup\{|\lambda| \mid \lambda\in\mathrm{Sp}(x)\}=\lim_{n\to\infty}\|x^n\|^{1/n}$. In Example 2, this spectrum is the spectrum of the operator x, and the problem of determining the behavior of this spectrum constitutes one of the central problems in the theories of Banach and Hilbert spaces.

D. Representations of a General Banach Algebra [1,2]

We understand by a **representation** of a Banach algebra R on a Banach space X any algebraic homomorphism π of R into the algebra $\mathscr{B}(X)$ (see Example 2) satisfying $\|\pi(x)\| \leqslant \|x\|$ for all $x \in R$. We call **X** the **representation space.** A Banach algebra always possesses an isomorphic and isometric representation. Especially important are the **irreducible representations.** A vector subspace (closed or not) Y of X is an invariant subspace if $\pi(x)\,Y\subset Y$ for any $x \in R$. A representation is called algebraically irreducible if the invariant subspaces are trivial, i.e., they are only $\{0\}$ or X. A representation is said to be topologically irreducible if closed invariant subspaces are trivial. The †kernel of an algebraically irreducible representation is called a **primitive ideal**, which can alternatively be defined in the following way: A left ideal J $(\neq \{0\}, R)$ is **regular,** by definition, if R contains an element u, a unity element modulo J, such that $x - xu \in J$ for any $x \in R$. Such an ideal is always contained in a maximal left ideal, which in turn is necessarily regular and closed. A two-sided ideal I is primitive if it is the set of elements a in R for which $aR \subset J$, where J is some fixed regular maximal left ideal. If R is commutative, a primitive ideal is a regular maximal ideal, and conversely. The intersection of all primitive ideals is called the **radical** of R, and when it is $\{0\}$, R is called **semisimple.**

The set of primitive ideals $\mathfrak{J}$ is known as the **structure space** of R, in which the **hull-kernel topology** (or **Jacobson topology)** is introduced. The †closure of a set $\mathfrak{A}$ in $\mathfrak{J}$ is, under this topology, the set of primitive ideals containing the intersection of the ideals in $\mathfrak{A}$. This topology is rather intractable; even in commutative cases, it does not coincide with the Gel'fand topology (→ Section E), in general.

E. The Gel'fand Representation of a Commutative Banach Algebra

A complex Banach algebra is a field if and only if it coincides with the field C of complex numbers **(Gel'fand-Mazur theorem).** This is the most fundamental fact in the study of commutative Banach algebras. Now let R be a commutative Banach algebra. Then every regular maximal ideal J of R is closed and the †quotient algebra R/J is isomorphic to C, so that the quotient mapping $R \to R/J$ is viewed as a complex-valued homomorphism of R. Conversely, if φ is a nonzero complex-valued homomorphism of R, it has norm $\leqslant 1$ as a linear functional on R and its kernel, say R_φ, is a regular maximal ideal of R such that φ is exactly the quotient mapping $R \to R/R_\varphi$. The set %II(R) of nonzero complex-valued homomorphisms (or regular maximal ideals) of R endowed with the tweak* topology of the Banach space †dual R' of R is called the **maximal ideal space** of R. $\mathfrak{M}(R)$ is a locally compact Hausdorff space, and its topology is called the **Gel'fand topology.** For each XE R we define a function $\hat{x}$ on $\mathfrak{M}(R)$ by setting $\hat{x}(\varphi) = \varphi(x)$. Then the mapping $x \to \hat{x}$ is a homomorphism of R into the algebra $C_0(\mathfrak{M}(R))$ of all continuous complex-valued functions on $\mathfrak{M}(R)$ vanishing at infinity. This is the **Gel'fand representation** of R whose image, the **Gel'fand transform** of R, is denoted by $\hat{R}$. Concerning this, we have (1) $\mathfrak{M}(R)$ is compact if R has a unity element; (2) Sp(x) equals the closure of the range $\hat{x}(\mathfrak{M}(R))$ of $\hat{x}$; (3) the representation $x \to \hat{x}$ is norm-decreasing if $C_0(\mathfrak{M}(R))$ is endowed with the †supremum norm; (4) $\|\hat{x}\|_\infty$ $(=\sup\{|\hat{x}(\varphi)| \mid \varphi \in \mathfrak{M}(R)\})$ equals $\lim_{n\to\infty} \|x^n\|^{1/n}$. The †kernel of the Gel'fand representation of R is the **radical** of R, which consists of **generalized nilpotent elements** of R.

F. Banach Star Algebras

An **involution** in a Banach algebra R is an operation $x \to x^*$ that satisfies (1) $(x + y)^* = x^* + y^*$; (2) $(\lambda x)^* = \bar{\lambda} x^*$; (3) $(xy)^* = y^* x^*$; (4) $(x^*)^* = x$. A Banach algebra with an involution is called a **Banach $*$-algebra**. A **$*$-homomorphism** Φ between two Banach $*$-algebras is an algebraic homomorphism which preserves involutions, i.e., $\Phi(x^*) = \Phi(x)^*$. To represent a Banach $*$-algebra we prefer a **$*$-representation**, i.e., a representation $x \to T_x$ on a Hilbert space such that T_{x^*} is equal to the adjoint T_x^* of T_x for any $x \in R$.

G. C*-Algebras [5–10]

A Banach $*$-algebra A satisfying $\|x^*x\| = \|x\|^2$ for all $x \in A$ is called a **C*-algebra.** Every C^*-algebra is $*$-isomorphic and -isometric to a Banach algebra of operators on a Hilbert space (see Example 2) that contains, along with an operator, its adjoint (the **Gel'fand-Vaïmark theorem).** A C*-algebra is semisimple, and a commutative C*-algebra A is $*$-isomorphic to $C_0(\mathfrak{M}(A))$ under the Gel'fand representation. A topologically irreducible $*$-representation of a C*-algebra is also algebraically irreducible (R. V. Kadison), and the set of unitary equivalence classes of these irreducible $*$-representations is called the **dual space.** It becomes a topological space if we introduce the hull-kernel topology inherited from the structure space, but other topologies are also introduced. Moreover, for the study of separable C*-algebras, Bore1 structure in the sense of G. W. Mackey is a very powerful tool. C^*-algebras also have intimate connections with the theory of †unitary representations of a topological group (see below) and with quantum physics. Many works have been published on $*$-representations, dual spaces, etc.

A linear functional ω on a C*-algebra A is said to be **positive** if $\omega(x^*x) \geqslant 0$ for any $x \in A$. For any nonzero positive linear functional ω of A, there exist a $*$-representation (π_ω, H_ω) of A on a Hilbert space If, and a vector $\xi_\omega \in H_\omega$ such that $\omega(a) = (\pi_\omega(a)\xi_\omega, \xi_\omega)$ for all $a \in A$ and that the subspace $\{\pi_\omega(a)\xi_\omega \mid a \in A\}$ is dense in H_ω. We call (π_ω, H_ω) a **cyclic representation** of A induced by ω. A $*$-representation (π, H) of A is called **universal** if for any $*$-representation (ρ, K) of A there exists a †σ-weakly continuous $*$-homomorphism $\tilde{\rho}$ of $n(A)''$ onto $\rho(A)''$ such that $\rho(a) = (\tilde{\rho} \circ \pi)(a)$ for all $a \in A$, where $\pi(A)''$ is the double †commutant of $\pi(A)$ and so is a †von Neumann algebra (→ 308 Operator Algebras). Von Neumann algebras $\pi(A)''$ for universal representations π of A are mutually $*$-isomorphic, so that they determine the **enveloping von Neumann algebra** of A. Especially, the direct sum π of all cyclic representations π_ω is a universal representation and there is a unique isometric isomorphism of $n(A)''$ onto the second dual A^{**} of A that is bicontinuous in the o-weak topology of $n(A)''$ and the $\sigma(A^{**}, A^*)$-topology of A^{**}. So A^{**} is identified with the enveloping von Neumann algebra of A.

H. Some Classes of C*-Algebras

Let A be a C*-algebra. If is a **CCR (liminal C*-) algebra** if it is mapped to the algebra of compact operators under any irreducible $*$-representation. It is a **GCR (postliminal C*-) algehra** if every nonzero quotient C*-algebra of A has a nonzero CCR closed two-sided ideal. These classes of C*-algebras have inter-

esting properties. For example, the dual space of a separable C*-algebra is a †T_0 space (†T_1 space) if and only if the algebra is GCR (CCR) (J. Glimm).

A is called an **AF (approximately finite) algebra** if it is the uniform closure of an increasing sequence of finite-dimensional C^*-algebras (0. Bratteli). AF algebras are classified by so-called dimension groups. To see this, we call an ordered Abelian group G (written additively) a **Riesz group** if (a) for any integer $n>0$ and any $g\in G$, $ng\geqslant 0$ implies $g\geqslant 0$ and (b) for any g_i, h_j ($1\leqslant i\leqslant m$, $1\leqslant j\leqslant n$) in G with $g_i\leqslant h_j$ for all i and j, there exists $k\in$ G with $g_i\leqslant k\leqslant h_j$ for all i, j. Then the isomorphism classes of AF algebras correspond bijectively to the isomorphism classes, as local semigroups, of generating upward-directed hereditary subsets of the positive cones of countable Riesz groups, and a dimension group, defined otherwise, is exactly a countable Riesz group (G. Elliot, E. G. Effros, D. E. Handelman and C. L. Shen) (→ [Il]).

Now consider two C*-algebras A and B. The algebraic †tensor product A $\otimes$ B over C becomes a $*$-algebra in the natural way. A norm β on $A\otimes B$ is a **C*-crossnorm** if $\|xy\|_\beta\leqslant\|x\|_\beta\,\|y\|_\beta$ and $\|x^*x\|_\beta=\|x\|_\beta^2$ for x, $y\in A\otimes B$ and $\|a\otimes b\|_\beta\leqslant\|a\|\,\|b\|$ for $a\in A$, $b\in B$. The completion of A $\otimes$ B under such a norm β is a C*-algebra, which is denoted by $A\otimes_\beta B$. There are two special C*-crossnorms on $A\otimes$ i?: $\|\cdot\|_{\min}$ and $\|\cdot\|_{\max}$. The former is called the spatial (minimal, injective) C*-crossnorm and is defined by $\|x\|_{\min}=\sup_{\varphi,\psi}\|(\varphi\otimes\psi)(x)\|$, where φ and ψ run over all $*$-representations of A and B, respectively. The algebra $A\otimes_{\min}B$ is called the **spatial tensor product** of A and B. The latter is called the greatest C*-crossnorm and is defined by $\|x\|_{\max}=\sup_T\|T(x)\|$, where T runs over all $*$-representations of A $\otimes$ B. And the algebra $A\otimes_{\max}B$ is called the **projective C*-tensor product** of A and B. Any C^*-crossnorm β on $A\otimes B$ lies in between these two norms, i.e., $\|x\|_{\min}\leqslant\|x\|_\beta\leqslant\|x\|_{\max}$ for $x\in A\otimes B$. A C*-algebra A is called **nuclear** if $A\otimes B$ has a unique C*-crossnorm for any C^*-algebra B. It is known that any GCR algebra is nuclear and that an inductive limit of nuclear C*-algebras (e.g., any AF algebra) is nuclear. Given a linear mapping φ between C*-algebras A and B, we define for any integer $n\geqslant 1$ a linear mapping φ_n :A $\otimes$ $M_n\to B\otimes M_n$ by setting $\varphi_n(x\otimes e_{ij})=\varphi(x)\otimes e_{ij}$, where $\{e_{ij}\}$ are the matrix units for the C*-algebra M_n of n x n complex matrices. φ is said to be **completely positive** if φ_n is positive for any $n\geqslant 1$. A C*-algebra A is then called **injective** if, for any C*-algebras B, C with $B\subseteqq$ C and any completely positive contraction φ: $B\to A$, there exists a completely positive contraction $\tilde{\varphi}$: C+A that extends φ. A C*-algebra A is nuclear if and only if its enveloping von Neumann algebra A^{**} is injective. Every injective C*-algebra B is not necessarily a von Neumann algebra but every **AW*-algebra,** i.e., every maximal commutative $*$-subalgebra M of B, is monotone complete (i.e., every increasing net of self-adjoint elements has a least upper bound in M) and the lattice of projections in B is conditionally complete.

I. Crossed Product

Let A be a C*-algebra, G a locally compact Hausdorff group with left-invariant measure μ, and α a continuous action of G on A as $*$-automorphisms α_g, $g\in G$. Let K(G, A) be the linear space of continuous functions from G to A with compact support, which is a $*$-algebra under the multiplication y $*$ z and the involution y* given by $(y*z)(g)=\int_G y(h)\alpha_h(z(h^{-1}g))\,d\mu(h)$ and $y^*(g)=\Delta(g^{-1})\cdot\alpha_g(y(g^{-1})^*)$ for y, $z\in K(G,A)$, where A is the †modular function of G. By completing $K(G,A)$ under the norm $\|y\|_1=\int_G\|y\|\,d\mu(g)$, we get a Banach $*$-algebra $L_1(G,A)$. $L_1(G,A)$ is then made into a C*-algebra by furnishing the norm $\|x\|=\sup_\pi\|\pi(x)\|$, where π runs over all $*$-representations of $L_1(G,A)$. This C*-algebra, denoted by $A\otimes_\alpha$ **G,** is the **crossed product** of A by G relative to the action α. Crossed products are useful in the structure theory of C^*-algebras.

J. Extensions by C*-Algebras (BDF Tbeory) [12]

Let H be a separable infinite-dimensional Hilbert space and $\mathscr{LC}(H)$ the ideal of $\mathscr{B}(H)$ consisting of all compact operators. The quotient C*-algebra $\mathscr{B}(H)/\mathscr{LC}(H)$ is called the **Calkin algebra.** We denote it by $\mathscr{Q}(H)$ and the quotient mapping by $\pi:\mathscr{B}(H)\to\mathscr{Q}(H)$. By an **extension** of $\mathscr{LC}(H)$ by a separable unital C*-algebra A we mean a unital (preserving unity elements) $*$-isomorphism τ of A into $\mathscr{Q}(H)$. We call two extensions τ_1, τ_2 :A $\to\mathscr{Q}(H)$ **equivalent** if, for some unitary operator $u\in\mathscr{B}(H)$, $\pi(u)\tau_1(a)\pi(u)^*=\tau_2(a)$ for any $a\in A$, and denote by Ext(A) the set of all equivalence classes [t] of extensions τ by A. Ext(A) forms a commutative semigroup with respect to the addition $[\tau]=[\tau_1]+[\tau_2]$, where $\tau(a)=\tau_1(a)\oplus\tau_2(a)\in\mathscr{Q}(H)\oplus\mathscr{Q}(H)\subset\mathscr{Q}(H\oplus H)$. Call an extension $\tau:A\to\mathscr{Q}(H)$ trivial if there is a unital *-isomorphism σ of A into $\mathscr{B}(H)$ with $\tau=\pi\sigma$. Then (1) all trivial extensions are equivalent,

and we define the unity element of the semigroup Ext(A) (D. Voiculescu); (2) Ext(A) is a group if A is commutative (L. G. Brown, R. G. Douglas, and P. A. Fillmore) or, more generally, if A is nuclear (M. D. Choi and E. G. Effros); (3) Ext(A) is not always a group (J. Anderson). For further information → 390 Spectral Analysis of Operators J.

K. Derivations in C*-Algebras

A linear operator δ in a C*-algebra A is a **derivation** in A if its domain $D(\delta)$ is a dense subalgebra of A and $\delta(xy) = \delta(x)y + x\delta(y)$ for x, $y \in D(\delta)$. δ is a **$*$-derivation** if, moreover, $x \in D(\delta)$ implies $x^* \in D(\delta)$ and $\delta(x^*)=\delta(x)^*$. δ is called bounded (unbounded, closed) if it is bounded (unbounded, closed) as a linear operator.

Every bounded derivation δ is expressed as $\delta(x)=\tilde{a}x-x\tilde{a}$ with some $\tilde{a}$ in the enveloping von Neumann algebra A^{**} of A (R. V. Kadison and S. Sakai). The element $\tilde{a}$ can be taken from the **multiplier algebra** $M(A) = \{b \in A^{**} \mid bA + Ab \subseteq A\}$ of A if A is simple (Sakai) or if A has continuous trace (C. A. Akemann, Elliott, G. K. Pedersen, and J. Tomiyama). If A is separable, we see that all bounded derivations in A are given by elements in M(A) if and only if A is the C*-direct sum of a family of simple C*-algebras and a C*-algebra with continuous trace (Elliott, Akemann, and Pedersen) [13] and that every bounded derivation in the quotient C*-algebra A/I, I being any closed two-sided ideal, can be lifted to a derivation in A (Pedersen).

Next we consider unbounded derivation. By a **$*$-automorphism group** on A we mean a †one-parameter group ρ_t, $t \in$ **R**, of $*$-automorphisms of A such that, for each $x \in A$, $\rho_t x$ is continuous in $t \in$ **R**. A **C*-dynamical system** is a pair consisting of a C*-algebra A and a $*$-automorphism group ρ_t on A. The fact that the time evolution of a physical system is often represented by such a dynamical system has made the study of unbounded derivations quite active. We have to see if a given derivation δ is †closable and if, in case of a closable δ, the closure $\bar{\delta}$ generates a $*$-automorphism group. Sample results: (1) If a $*$-derivation δ is **well-behaved** in the sense that for every self-adjoint x in $D(\delta)$ there exists a †state φ with $|\varphi(x)| = \|x\|$ and $\varphi(\delta(x))=0$, then δ is closable and its closure is well-behaved (A. Kishimoto and Sakai); (2) if a $*$-derivation δ is closable, its closure $\bar{\delta}$ generates a $*$-automorphism group if and only if δ is well-behaved and $(1 \pm \delta)D(\delta)$ is dense in A (R. T. Powers and S. Sakai; 0. Bratteli and D. W. Robinson). For further results → [14].

L. Applications to the Theory of Topological Groups [2, 3, 6, 15]

Banach algebras have many applications in different branches of mathematics and quantum physics. Here we mention some that concern topological groups. Let G be a locally compact Hausdorff group and μ be its left-invariant measure. We make the complex algebra $L_1(G)$ of Example 3 into a Banach $*$-algebra by defining $x^*(g)=\overline{x(g^{-1})}\Delta(g^{-1})$, where Δ is the †modular function of G. This is called the **L_1-algebra** (or **group algebra).** It is not C* but is semisimple. Considering a unitary representation of G is equivalent to considering a $*$-representation of the L_1-algebra. Replace the norm of $x \in L_1(G)$ by $\sup\|\pi(x)\|$, where the supremum is taken over all the $*$-representations. The new norm satisfies the C* condition, and the completion of $L_1(G)$ with respect to this norm is a C*-algebra, which we call the **C*-group algebra** of G. The dual of the C*-group algebra thus defined is called the dual of the group G, and this notion plays an important role in the study of topological groups. Unitary representations of a group G, $*$-representations of the L_1-algebra of G, and $*$-representations of the C*-group algebra of G are all characterized by positive definite functions on G. A function **p(g)** is **positive definite**, by definition, if it is measurable on G and

$$\iint p(g^{-1}h)x(g)x(h)\,d\mu(g)\,d\mu(h) \geqslant 0$$

for any continuous function x(g) with compact support.

The Abelian case. When G is an Abelian group (→ 422 Topological Abelian Groups), a regular maximal ideal M of the L_1-algebra R of G and a character γ of G are in a one-to-one correspondence by the relation

$$\hat{x}(M) = \int x(g)\gamma(g)\,d\mu(g)$$

(the left-hand side is the value of x at M under the Gel'fand representation). Moreover, the set of regular maximal ideals of R provided with the Gel'fand topology and the set $\hat{G}$ of characters of G provided with the Pontryagin topology (the †character group of G) are homeomorphic by this correspondence. Therefore the Gel'fand transform of an element x of the L_1-algebra R is seen to be a function $\hat{x}(\gamma)$ on $\hat{G}$ defined by the integral on the right-hand side in the above expression, which is properly called the **Fourier transform** of x. Of course, the Fourier transform can be defined for other classes of functions (e.g., for the L_2-space over G. the Fourier transform in the sense of Plan-

cherel), and classical theories of Fourier series, Fourier integrals, and harmonic analysis (→ 192 Harmonic Analysis) are studied from a more extensive point of view. Thus the statement that the Fourier transform of an element x of the L_1-algebra of G is a continuous function vanishing at infinity (→ Section E), for example, is a version of the classical †Riemann-Lebesgue theorem. **Bochner's tbeorem** in classical †Fourier analysis is restated thus: A continuous †positive definite function on an Abelian group can be put in the form

$$p(g)=\int \gamma(g)\,d\rho(\gamma),$$

where ρ is a uniquely determined bounded positive †Radon measure on the character group $\hat{G}$. Developing these theories further, we obtain an alternative proof of the †Pontryagin duality theorem (H. Cartan and R. Godement). A closed ideal I in the L_1-algebra R determines a set Z(I) in $\hat{G}$ as the set of common zeros of the Fourier transforms of elements of I. We ask whether, conversely, I is characterized by Z(1). This question is the problem of **spectral syntbesis,** and many important results have been obtained. The statement that I must coincide with R when Z(1) is empty is a formulation of the **generalized Tauberian tbeorem** of N. Wiener. A considerable simplification of the proof was accomplished by using the theory of Banach algebras, which was the first application of the theory (I. M. Gel'fand).

M. Holomorpbic Functional Calculus [16,171

Let R be a unital commutative Banach algebra with maximal ideal space $\mathfrak{M}=\mathfrak{M}(R)$. We define the **joint spectrum** $\mathrm{Sp}(X)$ of a finite n-tuple $X=\{x_1,\dots,x_n\}$ in R by $\{(\hat{x}_1(\varphi),\,,\ \hat{x}_n(\varphi))$ $\varphi\in\mathfrak{M}\}$, a nonempty compact subset of $\mathbf{C}^n$. Then the **bolomorpbic functional calculus** says the following: For each X = $\{x_1,\,.\,,$ x,$\}\subset R$ there exists a unique algebra homomorphism Φ_X of the algebra $H(Sp(X))$ of †germs of holomorphic functions on $Sp(X)$ into R with the following properties: (1) $\Phi_X(1)=1$, the unity element of R, and $\Phi_X(z_i)=x_i$, $1\leqslant i\leqslant n$; (2) if $n<m$, $X'=\{x_1,\dots.$ x,} and $\tilde{F}(z_1,\dots.\ z_m)=F(z_1,\dots.\ z_n)$ with $F\in H(\mathrm{Sp}(X))$, then $\Phi_{X'}(\tilde{F})=\Phi_X(F)$; (3) if $\{\ F_k\quad k=1,2,\}$ are holomorphic in a fixed neighborhood U of $\mathrm{Sp}(X)$ and $F_k\rightarrow F$ uniformly on U, then $\Phi_X(F_k)\rightarrow\Phi_X(F)$ (L. Waelbroeck). From this follows the **implicit function tbeorem:** Let $x_1,\,.\,,x_n\in R$, $f\in C(\mathfrak{M})$, and let $F(w,z\,,\,.\,.\,.\,,\,z_n)$ be holomorphic in a neighborhood of the set $\mathrm{Sp}(f,x_1,\dots,x_n)=\{(f(\varphi),\hat{x}_1(\varphi),\,,\ \hat{x}_n(\varphi))\mid\varphi\in\mathfrak{M}\}$. Suppose that $F(f,\hat{x}_1,\dots,\hat{x}_n)=0$ but $\partial F/\partial w\neq0$ on $\mathrm{Sp}(f,x_1,$ $\dots,$x,). Then there exists a unique $y\in R$ such that $\hat{y}=f$ and $F(y,x_1,\,.\,,x_n)=0$ (R. Arens and A. Calderon). As an application we obtain Shilov's **idempotent tbeorem:** If K is a closed-open subset of $\mathfrak{M}$, there exists x $\in R$ with $\hat{x}\equiv1$ on K and $\hat{x}\equiv0$ on $\mathfrak{M}-K$. Set $Q(R)=\{x\in R\mid e^{2\pi ix}=1)$. Then $Q(R)$ is an additive subgroup of R and Shilov's theorem says that the Gel'fand representation gives an isomorphism of $Q(R)$ onto Q(C(W)) ($\cong H^0(\mathfrak{M},\mathbf{Z})$). Another theorem of this sort is the **Arens-Royden tbeorem:** $R^{-1}/\exp(R)\cong C(\mathfrak{M})^{-1}/\exp(C(\mathfrak{M}))\cong H^1(\mathfrak{M},\mathbf{Z})$, where R^{-1} is the set of invertible elements of R and $\exp(R)=\{e^x\mid$ XE $R\}$. Further extensions related to K-theory have been given by J. L. Taylor and others. (For functional calculus for one variable → 25 1 Linear Operators)

References

[1] C. E. Rickart, General theory of Banach algebras, Van Nostrand, 1960.
[2] M. A. Naïmark, Normed algebras, Wolters-Noordhoff, 1972. (Original in Russian, 1968.)
[3] L. H. Loomis, An introduction to abstract harmonic analysis, Van Nostrand, 1953.
[4] K. Yosida, Functional analysis, Springer, third edition, 1971.
[5] W. Arveson, An invitation to C*-algebra, Springer, 1976.
[6] J. Dixmier, Les C*-algèbres et leurs représentations, Gauthier-Villars, 1964.
[7] G. K. Pedersen, C*-algebras and their automorphism groups, Academic Press, 1979.
[S] S. Sakai, C*-algebras and W*-algebras, Springer, 1971.
[9] M. Takesaki, Theory of operator algebras 1, Springer, 1979.
[10] 0. Bratteli and D. W. Robinson, Operator algebras and quantum statistical mechanics, Springer, 1979.
[1 1] G. A. Elliott, On the classification of inductive limits of sequences of semisimple finite-dimensional algebras, J. Algebra, 38 (1976), 29-44.
[12] R. G. Douglas, Extensions of C*-algebras and algebraic topology, Proc. Internat. Congress Math. Helsinki 1978, vol. 2, 1980, 6999 704.
[13] C. A. Akemann and G. K. Pedersen, Central sequences and inner derivations of separable C*-algebras, Amer. J. Math., 101 (1979), 1047-1061.
[14] S. Sakai, Recent developments in the theory of unbounded derivations in C*-

algebras, Proc. Internat. Congress Math. Helsinki 1978, vol. 2, 1980, 709–714.
[15] W. Rudin, Fourier analysis on groups, Interscience, 1962.
[16] T. W. Gamelin, Uniform algebras, Prentice-Hall, 1969.
[17] J. L. Taylor, Banach algebras and topology, Algebra in Analysis, J. H. Williamson (ed.), Academic Press, 1975, 118186.

37 (X11.3)
Banach Spaces

A. General Remarks

The notion of **Banach space** was introduced in 1922 by S. Banach and N. Wiener, working independently. The idea was to apply topological and algebraic methods to fundamental problems of analysis, such as mapping problems in infinite-dimensional function spaces (→ 168 Function Spaces, 197 Hilbert Spaces, 251 Linear Operators).

B. Definition of Banach Spaces

We associate to each element x of a †linear space X over the real (or complex) number field a real number $\|x\|$ satisfying the following conditions: (i) $\|x\| \geqslant 0$ for all x, and $\|x\|=0$ is equivalent to $x=0$; (ii) $\|\alpha x\| = |\alpha| \cdot \|x\|$ for any real (complex) number α; (iii) $\|x+y\| \leqslant \|x\| + \|y\|$. Then $\|x\|$ is called the **norm** of the vector x, and X is called a **normed linear space.** The norm is thus an extension of the notion of the length of a vector in a †Euclidean space. A normed linear space X is a †metric space under the distance $\rho(x,y)=\|x-y\|$. We write $\text{s-lim}_{n\to\infty} x_n = x$ or simply $x_n \to x$ when $\lim_{n\to\infty} \|x_n - x\| = 0$ and say that x_n **converges strongly** to x. If this metric space X is †complete, then X is called a **Banach space.** In normed linear spaces the addition and the multiplication by scalars are continuous. A closed linear subspace M of a Banach space X is again a Banach space, and the †quotient space X/M becomes a Banach space if the norm of a †coset is defined by $\|x+M\| = \inf_{m\in M} \|x+m\|$. The subset $\{x \mid \|x\| \leqslant 1\}$ of a normed linear space X is called the (closed) **unit ball** (or unit sphere) of X.

Examples. †Function spaces C, L_p $(1 \leqslant p < \infty)$, M, W_p^l, H_0^l, †sequence spaces c, l, m, and BV are all Banach spaces (→ 168 Function Spaces).

C. Linear Operators and Linear Functionals

Suppose that a linear subspace D(T) of a linear space X is the (definition) **domain** of a mapping T with values in a linear space X_1. T is called a **linear operator** if $T(\alpha x + \beta y) = \alpha Tx + \beta Ty$ for any scalars α, β and x, y ∈ D(T). $R(T) = \{Tx \in X_1 \mid x \in D(T)\}$ is called the **range** of T. In the special case where X_1 is the real or complex number field, T is called a **linear functional**. If X and X_1 are both normed linear spaces, then T is continuous if and only if $\text{s-lim}_{n\to\infty} Tx_n = Tx$ whenever $\text{s-lim}_{n\to\infty} x_n = x$. This is equivalent to the condition $\sup_{x \in D(T),\, \|x\| \leqslant 1} \|Tx\| < \infty$. In particular, if $D(T) = X$, the linear operator T is continuous if and only if the set $\{Tx \mid \|x\| \leqslant 1\}$ is bounded. In this case, T is called a **bounded linear operator,** and $\|T\| = \sup_{\|x\| \leqslant 1} \|Tx\|$ is called the **norm** of the operator T. In particular, a linear operator T satisfying $\|T\| \leqslant 1$ is called a **contraction.** (Sometimes T is called a contraction only when $\|T\| < 1$. In that case, an operator with $\|T\| \leqslant 1$ is called a **nonexpansive operator.**) The **scalar multiple, sum,** and **product** of linear operators are defined by $(\alpha T)x = \alpha(Tx)$, $(T+S)x = Tx + Sx$, and $(ST)x = S(Tx)$, respectively. The **identity operator** I in X is defined by $I \cdot x = x$ for all x ∈ X. If the inverse mapping T^{-1} of $x \to Tx$ exists, then it is called the **inverse operator** of T. A normed linear space X is said to be **isomorphic** to a normed linear space Y if there exists a bounded linear operator T from X onto Y with bounded inverse. If T can be chosen to be isometric (i.e., $\|Tx\| = \|x\|$ for all $x \in X$), then X is said to be **isometrically isomorphic** to Y.

D. The Dual Space and the Dual Operator

The totality of continuous linear functionals f defined on a normed linear space X is a Banach space X' under the previously defined linear operations and the norm $\|f\| = \sup_{\|x\| \leqslant 1} |f(x)|$. This X' is called the **dual** (or **conjugate) space** of X. In view of useful properties of the †inner product in †Hilbert spaces, it is sometimes convenient to write $\langle x, f \rangle$ for f(x). Let X and Y be normed linear spaces, and let T be a linear operator with the dense domain $D(T)$ in X and the range $R(T)$ in Y. If (f, g) is a pair with $f \in Y'$ and $g \in X'$ satisfying the equation $(Tx, f) = \langle x, g \rangle$ for any x ∈ D(T), then g is determined uniquely by f. The operator T' defined by $T'f = g$ is linear and is called the **dual operator** (or **conjugate** or **adjoint operator)** of T. This is an extension of the notion of the †transpose of a matrix in matrix theory. If T is a bounded linear operator then

T' is also a bounded linear operator such that $\|T'\| = \|T\|$.

E. The Weak Topology and the Strong Topology

Let X be a normed linear space and X' its dual space. Take a finite number of elements $x'_1, x'_2, \ldots, x'_n$ from X', and consider the subset of $X: \{x \in X \mid \sup_{1 \leqslant i \leqslant n} |\langle x, x'_i \rangle| \leqslant \varepsilon\}, \varepsilon > 0$. If we take the totality of such subsets of X as a †fundamental system of neighborhoods of 0 of X, then X is a †locally convex topological linear space, denoted sometimes by X_w. This topology is called the **weak topology** of X. If a sequence $\{x_n\} \subset X$ converges to $x \in X$ with respect to the weak topology of X, then it is said to **converge weakly.** This is equivalent to the convergence $\langle x_n, f \rangle \rightarrow \langle x, f \rangle$ for any $f \in X'$. The original topology of X determined by the norm is then called the **strong topology** of X, and to stress the strong topology we sometimes write X_s in place of the original X. Take a finite number of elements $x_1, x_2, \ldots, x_n$ from X, and consider the subset of $X': \{x' \in X' \mid \sup_{1 \leqslant i \leqslant n} |\langle x_i, x' \rangle| \leqslant \varepsilon\}, \varepsilon > 0$. If we take the totality of such subsets of X' as a fundamental system of neighborhoods of 0 of X', then X' is a †locally convex topological linear space. We write this space as X'_{w^*} and call the topology the **weak* topology** of X'. The topology of X' defined by the norm $\|f\|$ is called the **strong topology** of X', and to stress the strong topology we write X'_s. The terms "weak" and "weakly" are used in reference to the weak topology, for instance, weak closedness and weakly closed. Similar conventions are used for the weak* and strong topologies.

The unit ball of the dual space of a normed linear space X is weak*-compact **(Banach-Alaoglu theorem).** Then by the †Krein-Milman theorem (→ 424 Topological Linear Spaces) the unit ball of X' is the weak* closure of the †convex hull of its †extreme points. If X is a Banach space, a convex subset K of X' is weak*-closed if and only if the intersection of K and each weak*-compact subset is weak*-closed **(Krein-Shmul'yan theorem).**

F. The Hahn-Banach Extension Theorem

Let M be a linear subspace of a real linear space X and p(x) a real-valued functional defined on X such that $p(x+y) \leqslant p(x)+p(y)$ and $p(\lambda x) = \lambda p(x)$ for all x, $y \in X$ and $\lambda \geqslant 0$. If a linear functional f_1 defined on M satisfies $f_1(x) \leqslant p(x)$ for all $x \in M$, then there is a linear functional f on X which extends f_1 and satisfies $f(x) \leqslant p(x)$ for all $x \in X$ (the **Hahn-Banach (extension) theorem).** The Hahn-Banach extension theorem has many applications. (1) Let M be a linear subspace of a normed linear space X. Then for any $f_1 \in M'$, we can construct an $f \in X'$ such that $f(x) = f_1(x)$ for all $x \in M$ and $\|f\| = \|f_1\|$. (2) For any $x_0 \neq 0$ of a normed linear space X, we can construct an $f_0 \in X'$ such that $f_0(x_0) = \|x_0\|$ and $\|f_0\| = 1$. (3) For any closed linear subspace M of a normed linear space X and a point $x_0 \notin M$ we can construct $f_0 \in X'$ such that $\|f_0\| = 1$, $f_0(x_0) = \inf_{m \in M} \|x_0 - m\|$ and $f_0(x) = 0$ for all $x \in M$. A proposition more general than **(3)** is **Mazur's theorem,** which is useful in applications: (4) Let a closed subset M of a normed linear space X be †convex. Then for any $x_0 \notin M$ we can construct an $f_0 \in X'$ such that $f_0(x_0) > f_0(x)$ for all $x \in M$. By (4) we can prove, e.g., that a convex set of a normed linear space is weakly closed if it is strongly closed. This proposition has the following corollaries: (5) A convex set A containing 0 is closed if and only if $(A^\circ)^\circ = A$, where A° is the †polar of A (the **bipolar theorem).** (6) For two closed convex sets A and B, containing 0, the polar of $A \cap B$ coincides with the weak* closure of the convex hull of $A^\circ \cup B^\circ$. For complex linear spaces, most of the propositions in this section are valid with Re $f(x)$ instead of $f(x)$.

The Hahn-Banach theorem can also be employed to prove the existence of the **generalized limit** (or **Banach limit)** $\mathrm{Lim}_{n\rightarrow\infty} \zeta_n$ defined for all bounded real sequences $\{\zeta_n\}$ such that $\liminf_{n\rightarrow\infty} \zeta_n \leqslant \mathrm{Lim}_{n\rightarrow\infty} \zeta_n \leqslant \limsup_{n\rightarrow\infty} \zeta_n$ and $\mathrm{Lim}_{n\rightarrow\infty}(\alpha\xi_n + \beta\eta_n) = \alpha \,\mathrm{Lim}_{n\rightarrow\infty} \xi_n + \beta \,\mathrm{Lim}_{n\rightarrow\infty} \eta_n$.

G. Duality in Normed Linear Spaces

An element x_0 of a normed linear space X gives rise to an element x''_0 of $(X'_s)'$ determined by $\langle x_0, x' \rangle = \langle x', x''_0 \rangle$ for all $x' \in X'_s$. If we write $x''_0 = Jx_0$, then J is a linear operator satisfying $\|Jx_0\| = \|x_0\|$ by (2), and so the space X is isometrically isomorphic to a linear subspace of $(X'_s)'_s$. If X_s coincides with $(X'_s)'_s$ under this isomorphism, we call **X** a **reflexive** (or **regular) Banach space.** A necessary and sufficient condition for the normed linear space X to be reflexive is that the unit ball of X be weakly compact. A convenient criterion for the reflexivity of X is that any bounded sequence $\{x_n\}$ of X contains a subsequence weakly convergent to a point of **X (Eberlein-Shmul'yan theorem).** In this connection, a Banach space X is reflexive if and only if each $x' \in X'$ attains its norm, i.e., there is an $x_0 \in X$ such that $\|x_0\| = 1$ and $|\langle x_0, x' \rangle| = \|x'\|$ **(James's theorem).**

A normed linear space is said to be **uniformly convex** if for any $\varepsilon > 0$ there exists a $\delta > 0$ such that $\|x\| \leqslant 1$, $\|y\| \leqslant 1$ and $\|x - y\| \geqslant \varepsilon$

implies $\|x+y\| \leqslant 2-\delta$. A normed linear space is said to be **uniformly smooth** if for any $\varepsilon>0$ there is a $\delta>0$ such that $\|x\|=1$ and $\|y\|\leqslant\delta$ imply $\|x+y\|+\|x-y\|\leqslant 2+\varepsilon\|y\|$. A Banach space X is uniformly convex (resp. uniformly smooth) if and only if the dual X' is uniformly smooth (resp. uniformly convex). If a Banach space X is uniformly convex or smooth, it is isomorphic to a Banach space that is uniformly convex and uniformly smooth (P. Enflo). The space L_p with $1<p<\infty$ is uniformly convex and uniformly smooth. Any uniformly convex or uniformly smooth Banach space is reflexive **(Milman's theorem).**

H. The Resonance Theorem

Let $\{T_n\}$ be a sequence of bounded linear operators from a Banach space X into a normed linear space **Y.** The **uniform boundedness theorem (resonance theorem** or **Banach-Steinhaus theorem)** states that $\sup_{n\geqslant 1}\|T_n\|<\infty$ if $\sup_{n\geqslant 1}\|T_nx\|<\infty$ for every $x\in X$. As a corollary, we have $\sup_{n\geqslant 1}\|x_n\|<\infty$ for any weakly convergent sequence of X. Another corollary states that the set $\{x\in X \mid \limsup_{n\to\infty}\|T_nx\|<\infty\}$ either coincides with X or is a subset of X of the †first category. This implies the so-called **principle of condensation of singularities,** which gives a general existence theorem for functions exhibiting various kinds of singularities, for example, a continuous function whose †Fourier expansion diverges at every point of a †perfect set of points having the cardinal number of the continuum.

I. The Closed Graph Theorem

Let T be a bounded linear operator from a Banach space X into a normed linear space Y. If the image of the unit ball of X under T is dense in the unit ball of Y, then for any $\varepsilon>0$ and any $y_0\in$ Y the equation $Tx=y_0$ has a solution x with $\|x\|\leqslant(1+\varepsilon)\|y_0\|$ **(Banach's theorem).** By using the †Baire category theorem, we can derive from this the **open mapping theorem:** Every bounded linear operator from a Banach space X onto a Banach space Y maps each open set in X onto an open set in Y. As an application of the open mapping theorem, we can prove the **closed graph theorem:** A linear operator T from the whole of a Banach space into a Banach space is continuous if and only if T is a closed operator, i.e., s-$\lim_{n\to\infty}x_n=x$ and s-$\lim_{n\to\infty}Tx_n$ = y imply Tx = y. This theorem plays an important role in modern treatments of linear partial differential equations.

J. The Closed Range Theorem

Let X and Y be Banach spaces and T a linear closed operator with domain D(T) dense in X and with range R(T) in Y. Under these conditions, the following four propositions are mutually equivalent. (1) $R(T)$ is a closed set in Y. (2) $R(T')$ is a closed set in X'. (3) $R(T)=\{y\in Y \mid \langle y,y^*\rangle=0$ for all $y^*\in D(T')$ such that $T'y^*=0\}$. (4) $R(T')=\{x^*\in X' \mid \langle x,x^*\rangle=0$ for all $x\in D(T)$ such that $Tx=0\}$. These four propositions, as a whole, are called the **closed range theorem.** This theorem implies (5) $R(T)$ = Y if and only if T' has a continuous inverse; and (6) $R(T')$ = X' if and only if T has a continuous inverse. The following theorem is of similar nature: the following three propositions on two closed linear subspaces M and N of a Banach space are mutually equivalent. (7) $M+N$ is closed. (8) $M^\circ+N^\circ$ is strongly closed. (9) $M^\circ+N^\circ$ is weak*-closed.

The Hahn-Banach theorem, the resonance theorem, the open mapping theorem, the closed graph theorem, and the closed range theorem can be extended to various classes of †locally convex topological linear spaces. By virtue of this extension, we are able not only to treat various fundamental problems of analysis from a unified viewpoint but also to develop the theory of functional analysis itself in a new direction (→ 424 Topological Linear Spaces; concerning linear operators on a Banach space → 68 Compact and Nuclear Operators, 251 Linear Operators, 390 Spectral Analysis of Operators).

K. Differential and Integral Calculus of Functions with Values in Banach Spaces

Calculus involving functions from a set to a Banach space is also an effective tool in various problems. A function x(t) defined on an interval $[a,b]$ with values in a Banach space X is said to be **strongly (weakly) continuous** if x(t) converges strongly (weakly) to $x(t_0)$ as $t\to t_0$. For a strongly (weakly) continuous function **x(t),** the **Riemann integral** can be defined in a standard way, using strong (weak) convergence of the Riemann sum

$$\int_a^b x(t)\,dt=\lim\sum x(t_i')(t_{i+1}-t_i).$$

The †fundamental theorem of calculus, i.e., strong (weak) differentiability of the indefinite integral, remains true. Various definitions of integrals of a Banach space-valued function on a †measure space are discussed elsewhere (→ 443 Vector-Valued Integrals).

Now let $x(\lambda)$ be defined on a domain Ω in the complex plane with values in a complex

Banach space X. $x(\lambda)$ is said to be **holomorphic** if $f(x(\lambda))$ is †holomorphic in Ω for every $f \in X'$. If $x(\lambda)$ is holomorphic, then there exists an X-valued function $y(\lambda)$ on Ω such that

$$\|\zeta^{-1}(x(\lambda+\zeta)-x(\lambda))-y(\lambda)\| \to 0 \quad \text{as} \quad \zeta \to 0.$$

Therefore there is no difference between "strong" and "weak" in analyticity. †Cauchy's integral theorem remains true for a holomorphic function $x(\lambda)$ with values in X, and the †Laurent expansion

$$x(\lambda)=\sum_{n=-\infty}^{\infty} a_n(\lambda-\lambda_0)^n,$$

$$a_n=\frac{1}{2\pi i}\int_C x(\lambda)(\lambda-\lambda_0)^{-n-1}\,d\lambda,$$

is valid with the integral taken in the Riemannian sense. Banach space-valued holomorphic functions on complex (or real) †analytic manifolds of higher dimension can be defined in a natural way by means of power series expansion. A function $\Theta(\lambda)$ defined on a domain in the complex plane with values in the Banach space $\mathbf{B}(X, Y)$ of bounded linear operators from X to Y becomes holomorphic if $f(\Theta(\lambda)x)$ is holomorphic for every $x \in X$ and $f \in Y'$. An operator-valued holomorphic function is often called an **analytic operator function.**

L. The Approximation Property

A Banach space X is said to have the **approximation property** if there is a family $\{S_\lambda\}$ of bounded linear operators of †finite rank in X such that $\inf_\lambda \|(S_\lambda - I)T\| = 0$ for all †compact linear operators T in X. There is a Banach space that fails to have the approximation property (Enflo [9]). More surprisingly, the space of all bounded linear operators on an infinite-dimensional Hilbert space fails to have the approximation property (A. Szankowski). The approximation property plays a decisive role in the theory of †tensor products of Banach spaces. A Banach space X is said to have the **bounded approximation property** if a family $\{S_\lambda\}$ in the definition of the approximation property can be taken bounded, i.e., $\sup_\lambda \|S_\lambda\| < \infty$. The bounded approximation property does not follow from the approximation property. The bounded approximation property is closely related to the existence of a basis. A sequence $\{e_n\}$ in a Banach space X is called a **Schauder basis** or simply a **basis** (or **base**) for X if to each $x \in X$ there corresponds a unique sequence of numbers $\{\alpha_n\}$ such that $\lim_{n\to\infty} \|x - \sum_{k=1}^n \alpha_k e_k\| = 0$. Most separable Banach spaces appearing in analysis have bases. A separable Banach space has the bounded approximation property if and only if it is isomorphic to a complemented linear subspace (→ Section N) of a Banach space with a basis (A. Pełczyński).

M. Injective and Projective Banach Spaces

Banach spaces of the type $†C(\Omega)$ with compact Ω and of the type $†l_1(\Omega)$ on a set Ω play a special role in the theory of Banach space. This is already seen in the fact that every Banach space is isometrically isomorphic to a subspace of a space $C(\Omega_1)$ as well as to a quotient space of a space $l_1(\Omega_2)$. A Banach space X is said to be **injective** if for any Banach space Y and its linear subspace M, each bounded linear operator T from M to X can be extended to a bounded linear operator $\tilde{T}$ from Y to X, that is, $\tilde{T}x = Tx$ for all $x \in M$. A Banach space is injective if and only if it is isomorphic to a complemented linear subspace of the space $C(\Omega)$ for a compact Hausdorff space Ω with the property that the closure of every open set is open. Such a topological space is called **†extremally disconnected.** The †maximal ideal space of the †Banach algebra $L_\infty(\Omega)$ is extremally disconnected. Hence the Banach space $L_\infty(\Omega)$ is injective. Whether every injective Banach space is isomorphic to a space $C(\Omega)$ is still an open problem. However, a Banach space is isometrically isomorphic to a space $C(\Omega)$ with Ω extremally disconnected if and only if it is injective with the property of norm-preserving extension, i.e., $\|\tilde{T}\| = \|T\|$ is always possible **(Nachbin-Goodner-Kelley theorem).** In this connection, the following propositions on a Banach space X are mutually equivalent. (1) X' is isometrically isomorphic to a space $†L_1(\mu)$. (2) For any Banach space Y, its linear subspace M, and $\varepsilon > 0$, each compact linear operator T from M to X can be extended to a compact linear operator $\tilde{T}$ from Y to X with $\|\tilde{T}\| \leqslant (1+\varepsilon)\|T\|$ (J. Lindenstrauss). A Banach space X is said to be **projective** if for any Banach space Y and its closed linear subspace M, each bounded linear operator S from X to the quotient space Y/M is lifted to a bounded linear operator $\tilde{S}$ from X to Y, i.e., $\varphi(\tilde{S}x) = Sx$ for all $x \in X$, where φ is the quotient mapping from Y to Y/M. Projectivity of X is characterized by its being isomorphic to the space $l_1(\Omega)$ on a set Ω. A Banach space is isometrically isomorphic to a space $l_1(\Omega)$ if and only if it is projective with the property of norm-preserving lifting, i.e., $\|\tilde{S}\| = \|S\|$ is always possible (→ [8]).

N. Complemented Subspaces Problems

A linear subspace M of a Banach space X is **complemented,** i.e., there is a closed linear

subspace N such that $M \cap N = \{0\}$ and $M + N = X$ if and only if M is the range of a bounded projection P, i.e., $P^2 = P$ and $R(P) = M$. Each nonzero closed hnear subspace of a Hilbert space is complemented, or more precisely, it is the range of a projection of norm one (- 197 Hilbert Spaces). This property distinguishes Hilbert spaces from general Banach spaces: (1) A Banach space of more than 3 dimensions is isometrically isomorphic to a Hilbert space if each nonzero closed linear subspace is the range of a projection of norm one **(Kakutani's theorem).** (2) A Banach space is isomorphic to a Hilbert space if each closed linear subspace is complemented (J. Lindenstrauss and L. Tzafriri [10]).

0. Quasi-Banach Spaces and Fréchet Spaces

Let X be a linear space over the real (or complex) number field. Suppose that a real-valued function $\|x\|$ on X satisfies (i) and (ii) of Section B and (iii') $\|x+y\| \leqslant k(\|x\|+\|y\|)$ with a constant $k \geqslant 1$ independent of x and y. Then $\|x\|$ is called the **quasinorm** of x, and X equipped with a quasinorm is called a **quasinormed linear space.** Let $0 < p \leqslant 1$ be the root of the equation $k = 2^{(1/p)-1}$. Then there is a distance $d(x, y) = d(x-y)$ depending only on x -y such that $d(x-y) \leqslant \|x-y\|^p \leqslant 2d(x-y)$. Hence a quasinormed linear space is a metric space in which a sequence x_n converges to x if and only if $\|x_n - x\| \to 0$. If a quasinormed linear space X is complete under this metric, then X is called a **quasi-Banach space.** The †function space L_p is a quasi-Banach space for $0 < p < 1$.

If we denote by $\|x\|$ the distance $d(x-0)$ of a quasinormed hnear space, then it satisfies (i) and (iii) of Section B and (ii') $\|-x\| = \|x\|$ and $\lim_{n\to\infty} \|\alpha_n x_n - \alpha x\| = 0$ whenever $\alpha_n \to \alpha$ and $\lim_{n\to\infty} \|x_n - x\| = 0$. A functional $\|x\|$ satisfying (i), (ii'), and (iii) is called a **pseudonorm.** If a hnear space X equipped with a pseudonorm is complete, then X is called a **Fréchet space (in the sense of Banach).** The †function space $S(\Omega)$ is a Fréchet space. Quasinormed linear spaces and Fréchet spaces are ttopological linear spaces, but they need not be †locally convex. Hence it is possible that there is no continuous linear functional except for zero. However, the open mapping theorem and the closed graph theorem hold for linear operators from a Fréchet space into a Fréchet space.

References

[1] S. Banach, Theorie des operations lineaires, Warsaw, 1932 (Chelsea, 1963).
[2] N. Dunford and J. T. Schwartz, Linear operators I, Interscience, 1958.
[3] K. Yosida, Functional analysis, Springer, 1978.
[4] M. M. Day, Normed linear spaces, Springer, 19.58.
[5] H. E. Lacey, The isometric theory of classical Banach spaces, Springer, 1974.
[6] J. Lindenstrauss and L. Tzafriri, Classical Banach spaces I and II, Springer, 1977, 1979.
[7] J. Diestel, Geometry of Banach space, Springer, 1975.
[8] A. Pełczyński, Some aspects of the present theory of Banach spaces, Oeuvres de Stefan Banach II, Warsaw, 1979.
[9] P. Enflo, A counterexample to the approximation problem in Banach spaces, Acta Math., 130 (1973), 3099317.
[10] J. Lindenstrauss and L. Tzafriri, On complemented subspaces problem, Israel J. Math., 9 (1971), 2633269.

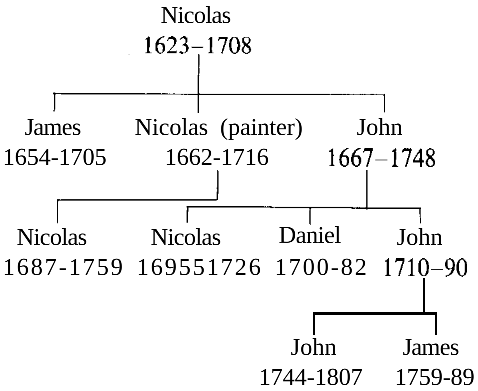
38 (XXI.1 4) Bernoulli Family

The Bernoullis, Protestants who came originally from Holland and settled in Switzerland, were a signiťicant family to the mathematics of the 17th Century. In a single Century, the family produced eight brilliant mathematicians, all of whom played important roles in the development of calculus.

The brothers James (1654-1705) and John (1667– 1748) and Daniel (1700– 1782), John's second son, were especially outstanding. James and John were close friends of G. W. †Leibniz, with whom they exchanged the correspondence through which it might be said that calculus developed. James studied problems related to the †tautochrone and †brachistochrone, as well as problems in geometry, dynamics, and other fields, including the †isoperimetric problem. He was the first to change the name *calculus summatoris* to *calculus integralis* (1690). His ***Ars conjectandi*** was published after his death in 1713; in it is found the †law of large numbers, which made his name prominent in the theory of tprobability. James had little guidance, learning mathematics on his own. He was a professor of experimental physics at the University of Base1 and later became a professor of mathematics. He taught mathematics to his brother John, who succeeded him as professor at the University. John's many achievements appeared in such publications of the time as ***Acta eruditorum*** and ***Journal des savants.*** In 1701, the beginnings of the †calculus of variations were seen in his solution to the isoperimetric problem. He

was the first to use the term *functio*, the root of the present term *function* (1714).

Despite discord between the brothers and also between fathers and sons, the Bernoullis were ardent teachers and brilliant researchers, who instructed not only their sons but also such mathematicians as †Euler. Their achievements were numerous in consolidating the content and form of calculus and also in expanding its application. Daniel was especially outstanding in the theory of probability; he also made contributions to the field of †hydrodynamics and to the †kinetic theory of gases. The eldest John's eldest son Nicolas (1695-1726) achieved distinction as a professor of mathematics in St. Petersburg. Daniel's youngest brother, John (1710–1790), succeeded his father, John Sr., as a professor at the University of Basel. The son of John, Jr., also named John (1744–1807), was the chairman of mathematics at the Academy of Berlin. His brother, another James (1759–1789), was a professor of experimental physics at the University of Basel. Nicolas (1687–1759), a grandson of the founder Nicolas (162331708) and son of Nicolas the painter (1662–1716), held Galileo's old chair of mathematics at Padua from 1716 to 1719.



In this article, the first names have been given in English. The German names corresponding to James, John, and Nicolas are Jakob, Johann, and Nikolaus, respectively.

References

[1] M. B. Cantor, Vorlesungen über Geschichte der Mathematik III, Teubner, 1898.
[2] Jacobi Bernoulli, Opera 1, II, Cramer, 1744.
[3] Jakob Bernoulli, Die Werke von Jakob Bernoulli, Birkhäuser, 1969.
[4] Jacobi Bernoulli, Ars conjectandi, Bâle, 1713.
[5] Johannis Bernoulli, Opera omnia I-IV, M. M. Bousquet, 1742.

39 (XIV.8) Bessel Functions

A. General Remarks

Bessel functions were first introduced in order to solve †Kepler's equation concerning planetary motions and were systematically investigated by F. W. Bessel in 1824. Since then they have appeared in various problems and have become important.

B. Bessel Functions

Separating variables for the Helmholtz equation $\Delta\Psi + k^2\Psi = 0$ in terms of cylindrical coordinates, we obtain **Bessel's differential equation**

$$\frac{d^2w}{dz^2}+\frac{1}{z}\frac{dw}{dz}+\left(1-\frac{\nu^2}{z^2}\right)w=0 \tag{1}$$

for the component of the radius vector. The following two linearly independent solutions of(l):

$$H_\nu^{(1)}(z)=\frac{1}{\pi}\int_{L_1} e^{-iz\sin\zeta+i\nu\zeta}\,d\zeta,$$

$$H_\nu^{(2)}(z)=\frac{1}{\pi}\int_{L_2} e^{-iz\sin\zeta+i\nu\zeta}\,d\zeta, \tag{2}$$

are called the **Hankel functions of the first and second kind,** respectively, where the contour L_1 of the first integration is a curve from $(-\pi+0)+i\infty$ to $-0-i\infty$, and L_2 is a curve from $+0-i\infty$ to $(\pi-0)+i\infty$. If both z and v are real, we have

$$\overline{H_\nu^{(1)}(z)}=H_\nu^{(2)}(z),\qquad \overline{H_\nu^{(2)}(z)}=H_\nu^{(1)}(z), \tag{3}$$

where $\bar z$ is the complex conjugate of z. Hence

$$J_\nu(z)=(H_\nu^{(1)}(z)+H_\nu^{(2)}(z))/2,$$

$$N_\nu(z)\equiv Y_\nu(z)=(H_\nu^{(1)}(z)-H_\nu^{(2)}(z))/2i \tag{4}$$

are real functions. If both z and v are complex, the functions J,(z) and $N_\nu(z)$ defined in (4) are also called **Bessel functions** and **Neumann functions**, respectively. The other names for J,(z), $N_\nu(z)$, and **H,(z)** are **Bessel functions of the first, second, and third kind,** respectively. Each of them satisfies the following recurrence formulas:

$$2\frac{dC_\nu(z)}{dz}=C_{\nu-1}(z)-C_{\nu+1}(z),$$

$$(2\nu/z)C_\nu(z)=C_{\nu-1}(z)+C_{\nu+1}(z). \tag{5}$$

In general, functions satisfying the simultaneous †differential-difference equations (5) are called **cylindrical functions**. Every cylindrical

function C,(z) is represented in the form C,(z) $=a_1(v)H_v^{(1)}(z)+a_2(v)H_v^{(2)}(z)$, where a,(v) and $a_2(v)$ are arbitrary periodic functions of period 1 with respect to v.

If v = n (an integer) we have

$$J_{-n}(z)=(-1)^n J_n(z),$$

$$N_{-n}(z)=(-1)^n N_n(z), \qquad (6)$$

which show the linear dependency of J_{-n} and J_n, and N_{-n} and N_n, respectively. If v $\neq n$ (an integer), as the fundamental solutions of (1) we can take a pair J_v and J_{-v}, or N_v and N_{-v}. In (2) if we take as a contour of integration a curve from $(-\pi+0)+i\infty$ to $(\pi-0)+i\infty$, *we* obtain an integral representation for ***J,(z),*** which yields the relations

$$J_v(ze^{im\pi})=e^{imv\pi}J_v(z),$$

$$J_{-v}(ze^{im\pi})=e^{-imv\pi}J_{-v}(z). \qquad (7)$$

If v = n (an integer) and Re z > 0, we obtain

$$J_n(z)=\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{iz\sin\zeta+in\zeta}d\zeta \qquad (8)$$

$$=\frac{1}{\pi}\int_0^{\pi}\cos(z\sin\zeta-\eta\zeta)d\zeta, \qquad (9)$$

which is called **Bessel's integral.** These representations imply the following expansions by means of †generating functions:

$$e^{iz\sin\zeta}=\sum_{n=-\infty}^{\infty}J_n(z)e^{in\zeta}, \qquad (10)$$

$$\cos(z\sin\zeta)=J_0(z)+2\sum_{n=1}^{\infty}J_{2n}(z)\cos 2n\zeta,$$

$$\sin(z\sin\zeta)=2\sum_{n=0}^{\infty}J_{2n+1}(z)\sin(2n+1)\zeta. \qquad (11)$$

Making a change of variable u = exp($i\zeta$) in (2), we obtain

$$J_v(z)=\frac{1}{2\pi i}\int_L \exp\left(\frac{z}{2}\left(u-\frac{1}{u}\right)\right)u^{-v-1}du, \qquad (12)$$

where L is a contour starting at the point at infinity with the argument $-\pi$, encircling the origin in the positive direction, and tending to the point at infinity with the argument π. From (12) we obtain a power series expansion

$$J_v(z)=\left(\frac{z}{2}\right)^v\sum_{m=0}^{\infty}\frac{(-1)^m}{m!\Gamma(v+m+1)}\left(\frac{z}{2}\right)^{2m}, \qquad (13)$$

obtained also from (1) by a power series expansion at z = 0, which is a †regular singular point of (1). Substituting (13) into

$$N_v(z) = (\cos vzJ_v(z) - J_{-v}(z))/\sin v\pi, \qquad (14)$$

we obtain a power series expansion for $N_v(z)$. A power series expansion for N,(z) for an integer n is obtained by taking the limit $v\to n$ (→ Appendix A, Table 19). In particular, if v = $n+1/2$ ***(n*** is an integer), we have

$$J_{n+1/2}(z)=(-1)^n\frac{(2z)^{n+1/2}}{\sqrt{\pi}}\frac{d^n}{d(z^2)^n}\left(\frac{\sin z}{z}\right),$$

$$n=0,1,2,\ldots,$$

which is represented by elementary functions and is sometimes called simply the **half Bessel function.** Bessel functions for half-integers have appeared also as radius vector components when the variables in the Helmholtz equation are separated by spherical coordinates. The function

$$j_n(z)=\sqrt{\pi/2z}\,J_{n+1/2}(z)$$

is called the **spherical Bessel function.**

We have the following **addition theorem:**

$$H_n^{(\mu)}(k\rho)e^{in\psi}=\sum_{m=-\infty}^{\infty}J_n(kr_2)H_{n+m}^{(\mu)}(kr_1)e^{im\varphi},$$

$$\mu=1,2,$$

where

$$\rho=\sqrt{r_1^2+r_2^2-2r_1r_2\cos\varphi},$$

$$\rho\cos\psi=r_1-r_2\cos\varphi, \qquad \rho\sin\psi=r_2\sin\varphi.$$

C. Zero Points of the Function *J,(z)*

From the differential equations satisfied by $J_v(\alpha z)$, *we* have

$$(\alpha^2-\beta^2)\int_0^1 zJ_v(\alpha z)J_v(\beta z)dz$$

$$=\beta J_v(\alpha)J_v'(\beta) \quad \alpha J_v'(\alpha)J_v(\beta). \qquad (15)$$

By letting $\beta\to\alpha$ in (15), we have

$$\int_0^1 z(J_v(\alpha z))^2dz$$

$$=\frac{1}{2}\left(\left(1-\frac{v^2}{\alpha^2}\right)(J_v(\alpha))^2+(J_v'(\alpha))^2\right) \qquad (16)$$

If α and β are distinct roots of ***J,(z) = 0, we*** have from (15)

$$\int_0^1 zJ_v(\alpha z)J_v(\beta z)dz=0, \qquad \mathrm{Re}\,v> -1. \qquad (17)$$

The integral formulas (15), (16), and (17) are called **Lommel's integrals.**

As for the zero points of ***J,(z),*** the following facts are well known: ***J,(O) = 0*** if v > 0. ***J,(z)*** has no multiple zero points other than z = 0. $J_v(-\alpha)$ = 0 if $J_v(\alpha)$ = 0. Every zero point of ***J,(z)*** is real if v > -1. Between two adjacent zero points that are positive, there exists one and only one zero point of $J_{v-1}(z)$ and $J_{v+1}(z)$, respectively. ***J,(z)*** has a countably infinite set of zero points on the real axis. When v $\geqslant$ 0 is rational, every zero point of ***J,(z)*** except z = 0 is a †transcendental number. The transcendentality of π is a special case of this result for v = 1/2.

D. Expansion by means of Bessel Functions

Let $f(r,\varphi)$ be defined for $0<r<1$, $-\pi<\varphi<\pi$, and $\alpha_{n,1},\alpha_{n,2},\dots,\alpha_{n,s},\dots(0<\alpha_{n,s}<\alpha_{n,s+1}, s=$ 1, 2, for every n) be zero points of $J_n(x)$ $(n=0,1,2,\dots)$. Then we have an expansion

$$f(r,\varphi)=\sum_{n=0}^{\infty}\sum_{s=1}^{\infty}(a_{n,s}\cos n\varphi+b_{n,s}\sin n\varphi)J_n(\alpha_{n,s}r), \tag{18}$$

which is called the **Fourier-Bessel series**. The coefficients $a_{n,s}$ and $b_{n,s}$ are determined by the properties of †Fourier series and (16) and (17) as follows:

$$\left.\begin{matrix}a_{n,s}\\ b_{n,s}\end{matrix}\right\}=\frac{\varepsilon_n}{\pi(J_{n+1}(\alpha_{n,s}))^2}$$

$$\times\int_0^1\int_{-\pi}^{\pi}f(r,\varphi)J_n(\alpha_{n,s}r)\begin{matrix}\cos n\varphi\\ \sin n\varphi\end{matrix}\, r\,d\varphi\,dr;$$

$$\varepsilon_0=1,\qquad \varepsilon_1=\varepsilon_2=\dots=2.$$

The integral transformation

$$g(y)=\int_0^{\infty}xf(x)J_n(xy)\,dx \tag{19}$$

is called the **Fourier-Bessel transform.** If f(x) is sufficiently smooth and tends rapidly to zero as $x\to\infty$, the following inversion formula holds:

$$f(x)=\int_0^{\infty}yg(y)J_n(xy)\,dy. \tag{20}$$

There are other types of series expansions in terms of Bessel functions as follows: **Dini's series**

$$\sum_{m=1}^{\infty}a_mJ_\nu(\lambda_m x)$$

(λ_m is the mth positive root of $xJ'_\nu(x)+HJ_\nu(x)=0$, where H is a real constant); **Kapteyn's series**

$$\sum_{m=1}^{\infty}a_mJ_{\nu+m}((\nu+m)x);$$

Schlömilch's series

$$\frac{a_0}{2}+\sum_{m=0}^{\infty}a_mJ_0(mx);$$

and the **generalized Schlömilch series**

$$\frac{1}{2}\frac{a_0}{\Gamma(\nu+1)}+\sum_{m=1}^{\infty}\frac{a_mJ_\nu(mx)+b_m\mathbf{H}_\nu(mx)}{(mx/2)^\nu},$$

where $\mathbf{H}_\nu(mx)$ is the Struve function (- Section F).

E. Asymptotic Expansion

If $|z|$ or ν is sufftciently large, the asymptotic representation for Bessel functions is obtained by applying the †method of the steepest descent for (2). If $|z|>|\nu|$, we have

$$H_\nu^{(1)}(z)\sim\sqrt{\frac{2}{\pi z}}\exp i\left(z-\frac{\pi}{2}\nu-\frac{\pi}{4}\right),$$

$$-\pi<\arg z<2\pi,$$

$$H_\nu^{(2)}(z)\sim\sqrt{\frac{2}{\pi z}}\exp\left(-i\left(z-\frac{\pi}{2}\nu-\frac{\pi}{4}\right)\right),$$

$$-2\pi<\arg z<\pi,$$

$$J_\nu(z)\sim\sqrt{\frac{2}{\pi z}}\cos\left(z-\frac{\pi}{2}\nu-\frac{\pi}{4}\right),$$

$$-\pi<\arg z<\pi,$$

$$N_\nu(z)\sim\sqrt{\frac{2}{\pi z}}\sin\left(z-\frac{\pi}{2}\nu-\frac{\pi}{4}\right).$$

Hence $H_\nu^{(1)}(z)$ tends to zero as $|z|\to\infty$ in the upper half-plane, and becomes large exponentially as $|z|\to\infty$ in the lower half-plane. The results for $H_\nu^{(2)}$ are obtained by interchanging "upper half-plane" with "lower half-plane" in this statement.

If both $|z|$ and $|\nu|$ are sufficiently large, we have the **Debye asymptotic representations**. For example, if $z=\nu\sec\beta$ ($\nu>0$, $\beta>0$), we have

$$H_\nu^{(1,2)}(\nu\sec\beta)\sim(\pi\nu\tan\beta/2)^{-1/2}$$

$$\times\exp(\pm i(\nu(\tan\beta-\beta)-\pi/4)).$$

If $z=\nu\,\mathrm{sech}\,\alpha$ $(\nu>0,\alpha>0)$, we have

$$J_\nu(\nu\,\mathrm{sech}\,\alpha)$$

$$\sim(2\pi\nu\tanh\alpha)^{-1/2}\exp\nu(\tanh\alpha-\alpha),$$

$$N_\nu(\nu\,\mathrm{sech}\,\alpha)$$

$$\sim(\pi\nu\tanh\alpha/2)^{-1/2}\exp\nu(\alpha-\tanh\alpha),$$

If $|\nu|\sim|z|$, we have

$$H_\nu^{(1,2)}(\nu\sec\beta)\sim\frac{\tan\beta}{\sqrt{3}}$$

$$\times\exp\left(\pm i\left(\frac{\pi}{6}+\nu\left(\tan\beta-\frac{1}{3}\tan^3\beta-\beta\right)\right)\right)$$

$$\times H_{1/2}^{(1,2)}((\nu/3)\tan^3\beta)+O(\nu^{-1}),$$

which is called **Watson's formula.**

F. The Wagner Function

As an application of Bessel functions to the theory of nonstationary aircraft wings, T. Theodorsen introduced the function

$$C(z)=H_1^{(2)}(z)/(H_0^{(2)}(z)+H_1^{(2)}(z))$$

[6], and H. Wagner considered the function

$$k_1(s)=\frac{1}{2\pi i}\int_{\mathrm{Br}}e^{ws}\frac{2C(-iw)}{w}dw$$

[5], where $H_0^{(2)}(z)$, $H_1^{(2)}(z)$ are Hankel functions, and $\int_{Br}$ means a Bromwich integral giving the inverse Laplace transform. Then C(z) and $k_1(s)$ are called the **Theodorsen function** and **Wagner function,** respectively. The function $k_1(s)$ is equal to the lift coefficient when a 2-dimensional flat wing suddenly proceeds forward a distance s at an angle of incidence $1/\pi$.

G. Functions Related to Bessel Functions

The following functions are closely related to Bessel functions (- Appendix A, Table 19.IV).

(1) **Modified Bessel functions.**

$$I_\nu(z) \equiv e^{-i\nu\pi/2} J_\nu(e^{i\pi/2} z),$$

$$K_\nu(z) = \frac{\pi}{2} \frac{I_{-\nu}(z) - I_\nu(z)}{\sin \nu\pi}$$

(2) **Kelvin functions.**

$$\mathrm{ber}_\nu(z) \pm i\,\mathrm{bei}_\nu(z) \equiv J_\nu(e^{\pm 3\pi i/4} z),$$

$$\mathrm{her}_\nu(z) \pm i\,\mathrm{hei}_\nu(z) \equiv H_\nu^{(1)}(e^{\pm 3\pi i/4} z),$$

$$\mathrm{ker}_\nu(z) \equiv -(\pi/2)\mathrm{hei}_\nu(z),$$

$$\mathrm{kei}_\nu(z) \equiv (\pi/2)\mathrm{her}_\nu(z).$$

(3) **Struve function.**

$$\mathbf{H}_\nu(z) = \sum_{m=0}^{\infty} \frac{(-1)^m (z/2)^{\nu+2m+1}}{\Gamma[m+(3/2)]\Gamma[\nu+m+(3/2)]}.$$

(4) **Anger function.**

$$\mathbf{J}_\nu(z) \equiv \frac{1}{\pi}\int_0^\pi \cos(\nu\theta - z\sin\theta)\,d\theta.$$

When ν is an integer *n, we* have $\mathbf{J}_n(z) = J_n(z)$.

(5) **H. F. Weber function.**

$$\mathbf{E}_\nu(z) \equiv \frac{1}{\pi}\int_0^\pi \sin(\nu\theta - z\sin\theta)\,d\theta.$$

The last three functions satisfy certain inhomogeneous Bessel differential equations. Many other functions, such as †Airy's integral, can be represented by Bessel functions.

References

[1] G. N. Watson, A treatise on the theory of Bessel functions, Cambridge Univ. Press, 1922.
[2] A. Gray and G. B. Mathews, A treatise on Bessel functions and their application to physics, Macmillan, second edition, 1922.
[3] R. Weyrich, Die Zyhnderfunktionen und ihre Anwendungen, Teubner, 1937.
[4] F. Bowman, Introduction to Bessel functions, Longmans-Green, 1938.
[5] H. Wagner, Über die Entstehung des dynamische Auftriebes von Tragflügeln, Z. Angew. Math. Mech., 5 (1925), 17-35.
[6] T. Theodorsen, General theory of aerodynamic instability and the mechanism of flutter, NASA Tech. Rep. 496 (1935).
Also → references to 167 Functions of Confluent Type, 389 Special Functions.

40 (XVIII.1 6) Biometrics

A. General Remarks

Biometrics is the branch of science that applies mathematical and statistical methods to biological problems, and deals with all the phenomena that affect the physical, social, and psychological well-being of human beings. These phenomena involve the relationships of groups of human beings to other human beings, to animals, microbes, and plants, and to physical and chemical elements in the environment. In dealing with these problems the biometrician encounters such theoretical tasks as analyzing autocorrelated data in time series, and such practical undertakings as cost-vs-benefit evaluations of health programs.

Biometrics began in the middle of the 17th century when Sir William Petty and John Graunt developed a new method of analyzing the London Bills of Mortahty. Petty and Graunt essentially invented the field of *vital statistics* by studying the reports of christenings and causes of death and proposed a method called "political arithmetic."

Some fields to which biometries is relevant are given below.

B. Statistical Genetics

After early development in vital statistics, *statistical genetics* was founded on the new ideas emerging in statistics. Major contributions were made by Charles Darwin (1809–1882), Francis Galton (1822–1911), Karl Pearson (1857– 1936), and Ronald A. Fisher (1890-1962).

Galton was the first to use the term "regression" in statistics (→ 403 Statistical Models D), when he observed that sons regressed linearly on their fathers with respect to stature. He called the phenomenon a "regression to mediocrity" because the deviations of the stature of sons were less than those of fathers. This gave rise to the measurement of correlation in the bivariate normal distribution by means of the coefficient of correlation (Pearson, 1897). Pearson is credited with the

creation of the discipline of biometry (biometrics), and he established the journal *Biometrika* to promote studies in the field. Fisher's major contributions were to genetics and statistical theories. His *General theory of natural selection* appeared in 1930. This landmark book, along with earlier and later publications, represented Fisher's attempts to give quantitative form to Darwin's views and to frame a statistical theory of evolution.

C. Bioassay

Bioassay is a set of techniques for evaluating the effectiveness of dosages of drugs by monitoring biological responses. It entails the use of special transformations, such as probits and logits, as well as the application of regression to the estimation of dosages that are *p* percent effective within stated confidence limits. Problems to be solved include measuring relative potency, slope-ratio assays, and quanta1 responses vis-à-vis tolerance distributions.

D. Demography

Demography, which includes traditional vital statistics, rates and ratios, life tables, competing risks, actuarial statistics, and census enumeration techniques, is a part of biometrics. In this category, many tabulations of data consist of time series of events or rates classified by age. For the analysis of such data, the cohort analysis techniques described in Hastings and Berry [3] are employed.

E. Epidemiology

The quantitative description of an epidemic should state the sensitivity and specificity of any diagnostic tests, as well as the true incidence or prevalence of the epidemic from survey results. Within an epidemiological theory, a disease is studied by the use of deterministic and stochastic models, wherein the theory of Markov chains can be applied (- 260 Markov Chains). Differential equations involving probability-generating functions or tmoment-generating functions can be solved to yield the tprobability distribution, †mean, and †variance of the number of infected individuals as functions of time [4, 5]. Fundamental to this whole field of application is a clear understanding of causality and association.

When clinical trials are possible, two groups of persons, "treated" and "untreated," are monitored over a period of time with regard to the incidence or recovery from the disease under study. The techniques in this procedure include compiling reports on persons to be observed, using double-blind techniques, and combining multiple-response variables by use of multivariate analysis (- 280 Multivariate Analysis).

If clinical intervention is forbidden on ethical grounds, e.g., studying congenital malformation by infecting pregnant women with German measles, the relative risk of exposure is estimated from retrospective studies. In practice, almost all statistical studies in epidemiology are retrospective with the sole exception of clinical trials. The research is ex post facto because investigators are mostly confined to describing and analyzing sudden and/or obvious events in the etiology of the disease.

F. Clinical Trials

In clinical trials, many problems have arisen for which biometricians have had to develop special techniques. One such technique takes account of unexpected adverse effects of drugs, and the consequent early termination of a trial. Moreover, when data demonstrate a trend earlier than expected, investigators will desire to end the accession of patients and to stop further treatment with what may be an inferior regimen. This means that the biometrician must be familiar with the problems of multiple examinations of data, multiple comparisons, and other adjustment procedures required by the ex post facto dredging of data.

G. Future Trends

There are two areas in which the biometrician has played a leading role recently. These are pertinent in many different applications and problems, and considerable methodological research has been devoted to the two areas. The areas are "mathematical modeling" and "effects of hazardous substances."

Mathematical Modeling. The relationship between a set of independent variables and the dependent or response variable(s) is usually referred to as a mathematical model. The mode1 may take the form of a standard multiple regression analysis with a single response variable or with multiple response variables as in multivariate analysis (- 280 Multivariate Analysis).

It is generally assumed that specialists with substantive knowledge of the specific field of applications (epidemiology, toxicology, pharmacology, radiology, genetics, etc.) play a

crucial role in determining a model or relationship between a set of independent variables and response variable(s). However, it is mainly the biometrician who finally selects the specific model establishing the functional relationship and who attempts to measure the strength or influence of the independent variables therein. The biometrician is also expected to contribute substantially to the decision as to whether the relationship is causal, or merely one of association or correlation. For example, in the measurement of carcinogenicity of a food additive or drug, questions arise as to whether a substance can be judged harmful if it "accelerates" the appearance of a tumor even though it does not increase the incidence of the abnormal growth. In general, the answer to this question is in the affirmative when an unequivocal dosage-response relationship is indicated between the substance and the tumor.

Effects of Hazardous Substances. With the successful conquest of most of the infectious diseases that have plagued mankind throughout history, health authorities have recently been concentrating on two chronic diseases whose etiology is yet to be determined: cardiovascular disease and cancer. In both cases there is no disagreement with the thesis that heredity exercises a determining influence, but the role of the environment in causing many cases is also unquestioned. Measurement of the risk due to potentially toxic substances in the environment, principally with respect to these two diseases, represents the greatest challenge to the biometrician today. The social benefits of success make this a tantalizing area of research, though exceptional complexities are involved.

For related topics → 263 Mathematical Models in Biology.

References

[1] D. J. Finney, Statistical method in biological assay, Griffin, second edition, 1964.
[2] D. R. Cox, The analysis of binary data, Methuen, 1970.
[3] D. W. Hastings and L. G. Berry (eds.), Cohort analysis: A collection of interdisciplinary readings, Scripps Foundation for Research in Population Problems, Oxford, Ohio, 1979.
[4] N. T. J. Bailey, The mathematical theory of epidemics, Griffin, 1957.
[5] M. S. Bartlett, Stochastic population models in ecology and epidemiology, Methuen, 1960.

41 (XX.20) Boltzmann Equation

A. Introduction

The **Boltzmann equation** is an **equation of motion of a rarefied gas** given by L. Boltzmann in 1872 [1]. He used it successfully in his pioneering work on the kinetic theory of gases and on the more general statistical mechanics (→ 402 Statistical Mechanics), but the equation was held to be inconsistent with the †classical mechanics used in its derivation; objections were raised by, e.g., J. Loschmidt (1895) and E. Zermelo (1896). Through the controversies it became widely recognized that the Boltzman equation should be justified by means of the †theory of probability, and many such justifications have been proposed. Also, this equation has been studied extensively as a significant nonlinear partial differential equation.

B. Boltzmann Equation

Let $f=f(t, x, \xi)$ be the density of gas molecules having position $x \in \mathbf{R}^3$ and velocity $\xi \in \mathbf{R}^3$ at time t. The Boltzmann equation is a conservation law for f of the form [1–3]

$$f_t=-\xi\cdot\nabla_x f-a(t,x)\cdot\nabla_\xi f+Q[f]. \qquad (1)$$

Here $a(t, x)$ denotes the external force and Q is a quadratic nonlinear integral operator in ξ-space describing binary collisions of the molecules. The integral kernel of Q, called the collision cross section, depends on the intermolecular forces. The two classical examples are the hard ball model (a gas of rigid spheres) and the inverse power law potential (one proportional to r^{-s}, where r is the intermolecular distance and $s>1$). The latter gives rise to singularities in the kernel, and a cutoff is customarily employed to avoid this difficulty. Grad's hard ($s \geqslant 5$) and soft ($s<5$) cutoff potentials may be effectively used in the study of (1) [3].

If the gas is contained in a vessel (domain), then f must also satisfy boundary conditions determined by the assumed law of reflection at the walls (boundaries) (specular reflection, random reflection, etc.). For spatially homogeneous gases with no external forces, (1) reduces to

$$f_t=Q[f], \qquad (2)$$

with $f=f(t, \xi)$. The Maxwellians, those solutions satisfying the †Maxwell-Boltzmann distribution law, are the only stationary solutions of (2); they describe equilibrium states.

C. Justification by the Theory of Probability

Consider (2). The collision process is a †Markov process, and the †master equation obtained from the †Chapman-Kolmogorov equality is equivalent to (2) under the assumption of "propagation of molecular chaos" (G. Uhlenbeck (1942)). This assumption becomes valid in the limit as N (total number of molecules) $\rightarrow\infty$ (M. Kac (1955), H. McKean (1967), F. Grünbaum (1971)) [4]. Also, nonlinear Markov processes defined by (2) (with probability density f/N) have been studied (McKean (1967), H. Tanaka [5]). No such results have yet been established for (1).

D. Existence of Solutions

T. Carleman [2] gave the first solutions of the Boltzmann equation, solving globally in time the †Cauchy problem for (2) for the hard ball model. His result has been extended to a wide class of potentials.

The spatially inhomogeneous case (1) is also known to have global solutions if $a(t, x) = 0$ and if the initial data for f are nearly Maxwellian. This was first proved by S. Ukai [6] assuming periodicity in x, and then proved for the Cauchy problem by Ukai [7] and T. Nishida and K. Imai [8], and for the initial boundary problem by J. P. Guiraud [9] (random reflection) and Y. Shizuta and K. Asano [10] (specular reflection) for bounded domains, and by Ukai and Asano [11] for exterior domains, all assuming Grad's hard cutoff potentials. The case of soft cutoff potentials was also solved (Ukai and Asano, R. Caflisch (1980)). All the solutions are unique and tend to Maxwellians as $t\rightarrow\infty$, with certain decay rates. There are some results also on 1-dimensional shock wave solutions (B. Nicolaenko (1975)) and stationary solutions (J. P. Guiraud (1972) (bounded domains), Ukai and Asano (1980) (exterior domains)).

For initial data far from the Maxwellian, (1) remains unsolved even locally in time for hard cutoff potentials. As for noncutoff potentials, the existence theorem in [5] is the only result known so far for either (1) or (2).

E. Hilbert Expansion

Put $f^{\varepsilon} = f/N$ with $\varepsilon = 1/N$. If f^{ε} has a power series expansion in ε, then (1) gives an infinite system of equations for the coefficients. D. Hilbert (1912) proposed a method of solving this system as an application of his theory of integral equations, initiating the attempt to solve the Boltzmann equation. No proofs of convergence exist, but his expansion and the improved one by S. Chapman (1917) and D. Enskog (1917) made possible the applications of (1) to hydrodynamics. The zeroth-order approximation to the expansion gives rise to the compressible Euler equation, and the first-order approximation yields the compressible Navier-Stokes equation. The solutions f^{ε} of the Cauchy problem for (1) with nearly Maxwellian initial data converge to those of the compressible Euler equation locally in t as $\varepsilon\rightarrow0$ (Nishida [12], Caflisch (1980), and to those of the compressible Navier-Stokes equation asymptotically as $t\rightarrow\infty$ when $\varepsilon>0$ is fixed [13].

References

[1] L. Boltzmann, Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen, Wien. Ber., 66 (1872), 275–370.
[2] T. Carlemann, Problèmes mathématiques dans la théorie cinétique des gaz, Almqvist and Wiksell, 1957.
[3] H. Grad, Asymptotic theory of the Boltzmann equation, Rarefied Gas Dynamics, J. A. Laurmann (ed.), Academic Press, 1963, vol. 1.
[4] M. Kac, Some probabilistic aspects of the Boltzmann equation, Acta Physica Austriaca, suppl. x (1973), 379–400.
[5] H. Tanaka, On Markov process corresponding to Boltzmann's equation of Maxwellian gas, Japan-USSR Symp. on Probability Theory, 2nd, Kyoto, 1972, G. Maruyama and Yu. V. Prokhorov (eds.): Lecture notes in math. 330, Springer, 1973.
[6] S. Ukai, On the existence of global solutions of mixed problem for the Boltzmann equation, Proc. Japan Acad., 50 (1974), 179–184.
[7] S. Ukai, Les solutions globales de l'équation de Boltzmann dans l'espace tout entier, C. R. Acad. Sci. Paris, 282A (1976), 317–320.
[8] T. Nishida and K. Imai, Global solutions to the initial value problem for the nonlinear Boltzmann equation, Publ. Res. Inst. Math. Sci., 12 (1976), 229–239.
[9] J. P. Guiraud, An H theorem for a gas of rigid spheres in a bounded domain, Théories Cinétiques Classiques et Relativistes, Colloq. CNRS, 1975.
[10] Y. Shizuta and K. Asano, Global solutions of the Boltzmann equation in a bounded domain, Proc. Japan Acad., 53A (1977), 3-5.
[11] S. Ukai and K. Asano, On the initial boundary value problem of the linearized Boltzmann equation in an exterior domain, Proc. Japan Acad., 56A (1980), 12–17.

[12] T. Nishida, Fluid dynamical limit of the Boltzmann equation to the level of the compressible Euler equation, Comm. Math. Phys., 61 (1978), 119-148.
[13] S. Kawashima, A. Matsumura, and T. Nishida, On the fluid dynamical approximation to the Boltzmann equation at the level of the Navier-Stokes equation, Comm. Math. Phys., 70 (1979), 97-124.

42 (11.15) Boolean Algebras

A. Boolean Algebras

Boolean algebra was introduced by G. Boole to study logical operations (→ 411 Symbolic Logic). It is now included within the more general concept of †lattice or lattice-ordered set (→ 243 Lattices) and appears not only in logic but also very often in analysis in the form of a particular lattice of sets, e.g., the lattice of †measurable sets.

Let L be a given set and suppose that to any pair of its elements x, y there correspond two elements $x \cup y$, $x \cap y$ of L (called **join** and **meet** of x and y, respectively) such that the following laws are valid: (1) $x \cup y = y \cup x$, $x \cap y = y \cap x$ (commutative law); (2) $x \cup (y \cup z) = (x \cup y) \cup z$, $x \cap (y \cap z) = (x \cap y) \cap z$ (associative law); (3) $x \cup (y \cap x) = (x \cup y) \cap x = x$ (absorption law); (4) $x \cup (y \cap z) = (x \cup y) \cap (x \cup z)$, $x \cap (y \cup z) = (x \cap y) \cup (x \cap z)$ (distributive law); $(x, y, z \in L)$. From (1), (2), and (3) it follows further that $x \cup x = x \cap x = x$ (idempotent law). If $x \leqslant y$ is defined to mean $x \cup y = y$, L becomes an †ordered set with respect to the ordering $\leqslant$. Now suppose, moreover, that the following law holds: (5) there exist a least element 0 and a greatest element I, and for any element x there exists an element x' satisfying $x \cup x' = I$, $x \cap x' = 0$ **(law of complementation)**. Then L is called a **Boolean algebra** (or †**Boolean lattice**). In this case x' is uniquely determined by x and is called the **complement** of x. The binary operations $(x, y) \to x \cup y$, $x \cap y$ together with the operation $x \to x'$ are called **Boolean operations.** These operations obey **de Morgan's law** $(x \cup y)' = x' \cap y'$, $(x \cap y)' = x' \cup y'$.

B. Generalized Boolean Algebras

Suppose that $a \leqslant b$ holds for two given elements a, b of an ordered set. Then the set of all elements x satisfying $a \leqslant x \leqslant b$ is denoted by $[a, b]$ and is called an **interval.** An interval of a Boolean algebra is also a Boolean algebra with respect to the induced operations $\cup$ and $\cap$, where the least and greatest elements are a and b, respectively, and the complement of x in $[a, b]$ is equal to $a \cup (x' \cap b) = (a \cup x') \cap b$. More generally, if a set L with two operations $\cup$, $\cap$ satisfying (1)–(4) above has a least element 0 and if each interval of L satisfies (5) (i.e., is a Boolean algebra), then L is called a **generalized Boolean algebra.**

C. Boolean Rings

A ring L satisfying the condition $xx = x$ for all $x \in L$ (i.e., all of its elements are †idempotent) is called a **generalized Boolean ring,** and if it has a unity element then it is called a **Boolean ring.** A generalized Boolean ring L satisfies $x + x = 0$ for all $x \in L$ and is necessarily a commutative ring. A (generalized) Boolean algebra L becomes a (generalized) Boolean ring if for any elements x, y of L the sum $x + y$ is defined to be the complement of $x \cap y$ in the interval $[0, x \cup y]$, and the product xy is defined to be $x \cap y$. A nonempty subset J of a (generalized) Boolean algebra L is an ideal with respect to the corresponding structure of the ring if and only if $x \cup y \in J$ for x, $y \in J$ and $x \cap y \in J$ for $x \in J$, $y \in L$. More generally, in any lattice, a nonempty subset that satisfies these conditions is sometimes called an **ideal** of the lattice.

D. Representation of a Boolean Algebra

Any Boolean algebra L is isomorphic to a Boolean lattice of subsets in a set X. If L is of finite †height, then L is isomorphic to the Boolean lattice $\mathfrak{P}(X)$ of all subsets of X. In general X can be taken to be the set of all maximal ideals of L. Let $a \in L$ and let $O(a) = \{\mathfrak{m} \mid \mathfrak{m} \in X, a \notin \mathfrak{m}\}$. The isomorphism is obtained by the mapping $a \to O(a)$. If we define a topology in X such that $\{O(a) \mid a \in L\}$ is the †open base, then X is a compact, totally disconnected T_1 space and $O(a)$ is characterized as a compact open set in X. Such a space X is called a **Boolean space** (M. H. Stone [3, 4]).

In any complete Boolean algebra L, the complete distributive laws hold: $(\sup_I x_i) \cap y = \sup_I (x_i \cap y)$ and its dual. These are equivalent to the stronger relations: $(\sup_I x_i) \cap (\sup_J y_j) = \sup_{I,J}(x_i \cap y_j)$ and its dual. In order for a Boolean algebra L to be isomorphic to the Boolean algebra $\mathfrak{P}(X)$ of all subsets of X, it is necessary and sufficient that the following strongest complete distributive laws hold: $\inf_I(\sup_{J(i)} x_{ij}) = \sup_F(\inf_I x_{i,\varphi(i)})$ (where F is the set of all functions φ assigning to each $i \in I$ a value $\varphi(i) \in J(i)$) and its dual.

References

[1] P. R. Halmos, Lectures on Boolean algebras, Van Nostrand, 1963.
[2] R. Sikorski, Boolean algebras, Erg. Math., Springer, second edition, 1964.
[3] M. H. Stone, The theory of representations for Boolean algebras, Trans. Amer. Math. Soc., 40 (1936), 37–111.
[4] M. H. Stone, Applications of the theory of Boolean rings to general topology, Trans. Amer. Math. Soc., 41 (1937), 375-481.
[5] G. Boole, Collected logical works 1, II, Open Court, 1916.

43 (X1.4) Bounded Functions

A. General Remarks

A complex-valued function defined on a subset E of the complex z-plane is called a **bounded function** defined on E if its range f(E) is bounded, that is, if there exists a positive constant M such that $|f(z)| \leqslant M$ on E. However, when studying the theory of bounded functions, we usually restrict ourselves to the consideration of †analytic or †harmonic functions. On the other hand, the classes of functions f(z) satisfying conditions such as $\operatorname{Re} f(z) > 0$ or $\alpha < \arg f(z) < \beta$ rather than $|f(z)| \leqslant M$ are studied by a method similar to that applied to the study of bounded functions.

†Schwarz's lemma, †Liouville's theorem, and †Riemann's theorem on the removability of singularities (which will be explained later) are among the classical theorems in the theory of bounded functions.

B. Maximum Principle

When a function f(z) is holomorphic and not constant in a domain D of the complex plane, $|f(z)|$ never attains its maximum in the interior of D. In particular, when f(z) is continuous on the bounded closed domain $\bar{D} = D \cup \partial D$, the maximum of $|f(z)|$ on D is taken on its boundary ∂D. This fact is called the **maximum (modulus) principle**.

As a direct application of the maximum principle, we can deduce **Schwarz's lemma:** If a holomorphic function f(z) in $|z| < R$ satisfies $|f(z)| \leqslant M$ and $f(0) = 0$, we have $|f(z)| \leqslant M \cdot |z|/R$ $(|z| < R)$. The equality at z_0, $0 < |z_0| < R$, occurs only for the functions f(z) = $e^{i\lambda} Mz/R$ (where A is a real constant).

C. Lindelöf's Principle

E. Lindelof extended Schwarz's lemma and obtained various extensions of the maximum principle, from which he, together with E. Phragmén, deduced several useful theorems on the behavior of a function that is single-valued and holomorphic in a neighborhood of the boundary. We mention some representative theorems:

Let z = $\varphi(\zeta)$ and w = $\psi(\zeta)$ both be †meromorphic and †univalent functions in $|\zeta| < 1$ that map $|\zeta| < 1$ onto D_z and D_w, respectively. Set $\varphi(0) = z_0$ and $\psi(0) = w_0$. Let $D,(p)$ and $D,(p)$ denote the images of $|\zeta| \leqslant \rho$ $(0 < \rho < 1)$ under the mappings φ and ψ, respectively. Under these circumstances, if a function f(z) that is holomorphic in D_z satisfies $f(D_z) \subset D_w$ and $f(z_0) = w_0$, then $f(D_z(\rho)) \subset D,(p)$. Furthermore, unless f(z) maps D_z onto D_w univalently, $f(D_z(\rho))$ is contained in the interior of $D,(p)$ (an extension of Schwarz's lemma).

Let f(z) be analytic in a bounded domain D but not necessarily single-valued. Suppose that $|f(z)|$ is single-valued. Suppose, furthermore, that there is a positive constant M such that, for each boundary point ζ of D, except for a finite number of boundary points and for each $\varepsilon > 0$, the inequality $|f(z)| < M+E$ holds on the intersection of D with a suitable neighborhood of ζ, and suppose also that each of the exceptional points has a neighborhood such that $f(z)$ is bounded on the intersection of D with this neighborhood. Under these assumptions we have $|f(z)| \leqslant M$. Moreover, if $|f(z_0)| = M$ at a point z_0 of D, then f(z) is a constant (an extension of the maximum principle).

Let f(z) be holomorphic in an angular domain W: $\arg z| < \alpha\pi/2$. Suppose that there is a constant M such that, for each $\varepsilon > 0$, each finite boundary point has a neighborhood such that $|f(z)| < M + \varepsilon$ on the intersection of D with this neighborhood, and that for some positive number $\beta > \alpha$ and for sufftciently large $|z|$ the inequality $|f(z)| < \exp|z|^{1/\beta}$ holds. Under these assumptions we have $|f(z) \leqslant M$ in D **(Phragmén-Lindelöf theorem).**

Let f(z) be a function that is holomorphic and bounded in a closed angular domain W: $\alpha \leqslant \arg z \leqslant \beta$ except for the point at infinity. Suppose that $f(z) \to a$ as $z \to \infty$ along a side of the angle and that $f(z) \to b$ as $z \to \infty$ along the other side of it. Then we have $a = b$ and $f(z) \to a$ uniformly as $z \to \infty$ in W **(Lindelöf's asymptotic value theorem).**

D. Bounded Functions in a Disk

If f(z) is a bounded holomorphic function in the unit disk $|z| < 1$, it has a limit at every

point z_0 on the circle C: $|z|=1$, except for a set of 1-dimensional measure zero, as z tends to z_0 from within an angle with vertex at z_0 and contained in $|z|<1$ (or along a †Stolz's path at z_0) **(Fatou's theorem).** Under the same assumption, if the boundary value function $f(e^{i\theta}) = \lim_{r\to 1-0} f(re^{i\theta})$, which exists by Fatou's theorem, is equal to a constant a for a set of positive measure on the circle C, then f(z) ≡ a in $|z|<1$ **(F. and M. Riesz theorem).** These theorems are valid for some kinds of meromorphic functions in $|z|<1$ (- 272 Meromorphic Functions D).

E. Three-Circle Theorem and Related Theorems

Let f(z) be a function that is single-valued, holomorphic, and not identically zero in an annulus $\rho<|z|<R$. Set $M(r)\equiv\max_{|z|=r}|f(z)|$ ($\rho<r<R$) for f(z). Then log M(r) is a †convex function of log r in log $\rho<$ log $r<$ log R **(Hadamard's three-circle theorem).** The same assertion holds for a function $f(z)$ that is not necessarily single-valued, as long as $|f(z)|$ is single-valued. When f(z) is single-valued, a stronger assertion can be obtained (0. Teichmüller). The following theorems are regarded as the analog of Hadamard's three-circle theorem for the respective basic regions:

Set $L(\sigma)\equiv\sup_{-\infty<t<\infty}|f(\sigma+it)|$ $(\alpha<\sigma<\beta)$ for a function that is bounded and regular in a strip $\alpha<\operatorname{Re} z<\beta$. Then log $L(\sigma)$ is a convex function of σ in $a<\sigma<\beta$ **(Doetsch's three-line theorem).**

Set $l(\sigma)=\limsup_{t\to\infty}|f(\sigma+it)|$ $(\alpha<\sigma<\beta)$ for a function that is holomorphic and bounded in a half-strip $\alpha<\operatorname{Re} z<\beta$, $\operatorname{Im} z>0$. Then log I(a) is a convex function of σ in $\alpha<\sigma<\beta$ **(Hardy-Littlewood theorem).**

Set

$$I_p(r)=\frac{1}{2\pi}\int_0^{2\pi}|f(re^{i\theta})|^p\,d\theta$$

for a holomorphic function in a disk $|z|<R$. Then for every $p>0$, log $I_p(r)$ is an increasing convex function of logr for $-\infty<\log r<\log R$ **(Hardy's theorem).**

F. Hardy Class

Hardy's theorem motivates us to introduce a class of functions. An analytic function f in the unit disk is said to belong to the **Hardy class** H^p $(0<p<\infty)$ if $I_p(r,f)=(2\pi)^{-1}\int_0^{2\pi}|f(re^{i\theta})|^p d\theta$ remains bounded as $r\to 1$. For the case $p=\infty$, $f\in H^\infty$ if $I_\infty(r,f)=\max_{|z|=r}|f(z)|$ is bounded.

An analytic function fis said to be of the class N if $\int_0^{2\pi}\log^+|f(re^{i\theta})|\,d\theta$ is bounded for $r<1$. $f\in$ N if and only if $f=\varphi/\psi$, where $\varphi,\psi\in H^\infty$. $f\in H^p$ if and only if $|f|^p$ has a †harmonic majorant. If $1\leqslant p\leqslant\infty$, H^p is a †Banach space with the norm $\|f\|_p=(\sup_{r<1}I_p(r,f))^{1/p}$ for $p<\infty$ and $\|f\|_\infty=\sup_{r<1}I_\infty(r,f)$ for $p=\infty$, and if $0<p<1$, H^p is a complete metric space with the metric $\sup_{r<1}I_p(r,f-g)$. For $0<p<q<\infty$ the inclusion relation $N\supset H^p\supset H^q\supset H^\infty$ holds, so that Fatou's theorem and that of F. and M. Riesz are valid for functions of H^p $(p>0)$.

For a nonnegative integer p and a sequence of nonzero complex numbers a_n, $|\alpha_n|<1$, the infinite product $B(z)=z^p\prod_{n=1}^\infty(\bar\alpha_n(\alpha_n-z)/|\alpha_n|(1-\bar\alpha_n z))$ converges locally uniformly in the unit disk if and only if $\sum_{n=1}^\infty(1-|\alpha_n|)<\infty$. Then B(z) is called a **Blaschke product** and $\{a_n\}$ a **Blaschke sequence.** If the sequence $\{a_n\}$ **U {0}** is the set of zeros of an analytic function f (an m-tuple zero appears m times in $\{\alpha_n\}$, and z = 0 is a p-tuple zero), B(z) is said to be generated by the zeros of f. If $f\in N$, then $\sum_{n=1}^\infty(1-|\alpha_n|)<\infty$ for the set $\{a_n\}$ of zeros of f. An analytic function f in the unit disk is called an **inner function** if its modulus is less than 1 and its nontangential limit on the unit circle is of modulus 1 almost everywhere. A Blaschke product is an inner function. An inner function f without zeros is said to be singular if f(0) > 0. A **singular inner function** f can be represented by $f(z)=\exp\{-\int_0^{2\pi}(e^{i\theta}+z)/(e^{i\theta}-z)\,d\mu(\theta)\}$ with a positive measure μ †singular with respect to Lebesgue measure. An **outer function F(z)** for H^p $(p>0)$ is an analytic function that can be represented by

$$F(z)=e^{ir}\exp\left\{\frac{1}{2\pi}\int_0^{2\pi}\frac{e^{i\theta}+z}{e^{i\theta}-z}\log\psi(\theta)\,d\theta\right\},$$

where r is a real number and $\psi(\theta)$ is a nonnegative function such that $\psi(\theta)\in L^p$ and $\log\psi(\theta)\in L^1$. An H^p-function f $(p>0)$ is factorized uniquely as $f(z)=B(z)S(z)F(z)$, where B(z) is a Blaschke product (generated by the zeros of f), S(z) a singular inner function, and $F(z)$ is an outer function for H^p (with $\psi(\theta)=|f(e^{i\theta})|$) **(factorization theorem).** Conversely, such a product belongs to H^p. The **interpolation problem** raised by R. C. Buck asks whether or not there exists a sequence of distinct points $\{z_n\mid |z_n|<1\}$ such that, given any bounded sequence of complex numbers $\{w_n\}$, there is a function $f\in H^\infty$ for which $f(z_n)=w_n$. The sequence $\{z_n\}$ in the problem is called an **interpolating sequence.** L. Carleson proved that a necessary and sufficient condition for $\{z_n\}$ to be an interpolating sequence is that there be a $\delta>0$ such that $\prod_{n\neq k}|(z_n-z_k)/(1-\bar z_k z_n)|\geqslant\delta$ $(k=1,2,\ldots)$. An interpolating sequence is a Blaschke sequence, and any sequence $\{z_n \mid z_n\to 1\}$ contains an interpolating subsequence

tending to the unit circle. An analogous theorem is obtained for a function of H^p $(p \geqslant 1)$.

G. Corona Problem

The Hardy space H^∞ on the unit disk D is furthermore a commutative †Banach algebra with identity. Let M denote the †maximal ideal space of H^∞. It is a compact Hausdorff space with respect to the †Gel'fand topology. To each point $z \in D$, there correspond a homomorphism φ_z: $\varphi_z(f)$ =f(z) for $f \in H^\infty$, and a maximal ideal $M_z = \{ f \in H^\infty \mid f(z) = 0\}$ as its kernel. This correspondence gives a continuous injection of D into the maximal ideal space M of H^∞. By identifying z and M_z, D is regarded as an open subset of M, and the **corona problem** then asks: Are there points of M outside the closure of D? The negative answer was given by L. Carleson, that is, D is dense in **A4 (corona theorem).** The corona theorem is obtained as a direct consequence of the following theorem and is equivalent to it. For functions $f_i \in H^\infty$, $i=1,2,\ldots,n$, satisfying $\sum|f_i(z)|>\delta$ in D with some $\delta>0$, there exist $g_i \in H^\infty$, $i=1,\ldots,n$, such that $\sum f_i(z)g_i(z) = 1$ in D. As a trivial consequence of the corona theorem, we have the following **cluster value theorem:** Let $C(f,\zeta)$ denote the †cluster set of $f \in H^\infty$ at ζ with $|\zeta| = 1$; then $C(f,\zeta)=\hat{f}(M_\zeta)$, where $\hat{f}$ is the Gel'fand transform off and M_ζ is the †fiber of M over ζ. M is decomposed into pairwise disjoint †Gleason parts. A Gleason part is a connected open set or a singleton according as each point of it is captured in the closure of an interpolating sequence or not. Each point of the †Shilov boundary Γ of M forms by itself a trivial Gleason part, and Γ is †adherent to a sequence S in D with the property that the set of its nontangential limit points covers the unit circle **(Brown-Shield-Zeller theorem).** Let D be, in turn, a general bounded domain. The corona theorem is proved for some classes of such domains. A Shilov boundary Γ lies only over the set of points on ∂D which are not †removable for bounded analytic functions. It is adherent to a sequence S in D on which $\sup_S |f(z)| = \|f\|_\infty$. For any $f \in H^\infty$ and $\zeta \in \partial D$, the boundary of $\hat{f}(M_\zeta)$ is contained in $\hat{f}(\Gamma_\zeta)$ if $\Gamma_\zeta = \Gamma \cap M_\zeta$ is not empty. This, together with the cluster value theorem, has an important implication in the theory of cluster sets. The extremal problem of maximizing $|f'(z_0)|$, $z_0 \in D$, among all $f \in H^\infty$ satisfying $\|f\|_\infty \leqslant 1$ has the unique normalized solution G such that $G'(z_0) = \max|f'(z_0)|$. G(z) is called the **Ahlfors function.** The Ahlfors function has unit modulus on the Shilov boundary **(Fisher's theorem).** The theory of H^∞ is also applied to the problem of †rational approximation of bounded analytic functions.

H. Applications of the Maximum Principle

Theorems of the following type are useful for some problems of holomorphic functions:

Let D be a domain. Suppose that there exists an arc of angular measure α that is on a circle of radius R centered at a point z_0 of D and not contained in D. Let C denote the intersection of the boundary of D with the disk $|z-z_0|<R$. If f(z) is a single-valued holomorphic function that satisfies $|f(z)| \leqslant M$, and if $\limsup_{z\to\zeta} |f(z)| \leqslant M$ for every $\zeta \in C$, then the inequality $|f(z_0)| \leqslant M^{1-1/n} m^{1/n}$ holds for every positive integer n satisfying $2\pi/n \leqslant \alpha$ **(Lindelöf's theorem).**

Let D be a domain bounded by two segments OA, OB both starting from 0 and making an angle $\pi\alpha$, and a Jordan arc $\widehat{AB}$, and let R be the maximal distance between 0 and the points on $\widehat{AB}$. Suppose that $f(z)$ is holomorphic in D and that $\limsup |f(z)|$ as $z \to \zeta \in \partial D$ with $z \in D$ is not greater than M for $\zeta \in OA \cup OB$ and m for $\zeta \in \widehat{AB}$. Then we have $|f(z)| \leqslant M^{1-\lambda} m^\lambda$ (where $\lambda = (|z|/R)^{1/\alpha}$) at every point on the bisector of the angle $\angle AOB$ in D **(Carlemann's theorem).**

I. Holomorphic Functions with Positive Real Parts

Holomorphic functions with positive real parts are intimately connected with bounded functions. Concerning these functions we have the following classical result, which is equivalent to Schwarz's lemma: If $f(z)$ is holomorphic in $|z|<R$, $\operatorname{Re} f(z) \geqslant 0$ in the same domain, and $f(0)=1$, then $(R-|z|)/(R+|z|) \leqslant \operatorname{Re} f(z) \leqslant (R+|z|)/(R-|z|)$ $(|z|<R)$. The right or left inequality becomes equality for some z_0, $0<|z_0|<R$, only if $f(z)=(Rz_0 \mp z_0 z)/(R \pm z_0 z)$, respectively.

In order to prove various results for the class of functions with positive real part, **Herglotz's integral representation**, which is based on Poisson's integral representation and unique to this class, can be used effectively. It is given by

$$f(z) = \int_0^{2\pi} \frac{e^{i\varphi}+z}{e^{i\varphi}-z} d\rho(\varphi), \quad |z|<R,$$

where $\rho(\varphi)$ is monotone increasing (real-valued) with total variation 1 and is determined uniquely up to an additive constant by $f(z)$. An analogous integral representation is introduced for a holomorphic function in an annulus.

Similar to the case of holomorphic functions is the notion of Hardy classes to harmonic

functions, that is, a harmonic function u in the unit disk is said to be of the class h^p $(0<p\leqslant\infty)$ if $I_p(r,u)$ is bounded for $r<1$. $u\in h^p$ if and only if $u+iv\in H^p$, where v is conjugate to u. h^1 is equal to the vector space of functions which are representable as a difference of two nonnegative harmonic functions. The Herglotz theorem stated above is then a corollary of the theorem of integral representation for h^1-functions.

J. Coefficient Problems

There are many classical results for partial sums and coefficients of the Taylor expansion of bounded functions in a disk. Let $f(z)=\sum_{n=0}^{\infty}c_nz^n$ be the Taylor expansion of a bounded function in $|z|<1$. Set its partial sum $s_n(z)=\sum_{\nu=0}^{n}c_\nu z^\nu$ $(n=0,1,\ldots)$. and let $t_n(z)=(1/(n+1))\sum_{\nu=0}^{n}s_\nu(z)$ $(n=0,1,\ldots)$, which is the sequence of the arithmetic means of the partial sums (the Fejér sums). Then $|f(z)|\leqslant1$ in $|z|<1$ if and only if $|t_n(z)|\leqslant1$ for $|z|=1$ $(n=0,1,\ldots)$ (L. Fejér). Thus the sequences $\{t_n(z)\}$ for bounded functions f(z) are uniformly bounded, whereas the sequences $\{s_n(z)\}$ are not uniformly bounded. Indeed the maximum value of $|s_n(1)|$ over the set of function f satisfying $|f(z)|\leqslant1$ for $|z|<1$ is

$$G_n=1+\sum_{j=1}^{n}\binom{-1/2}{j}^2$$

$$=\sum_{j=0}^{n}\left(\frac{1\cdot3\cdot\ldots\cdot(2j-1)}{2\cdot4\cdot\ldots\cdot2j}\right)^2$$

$(G_n\sim\pi^{-1}\log n$ as $n\to\infty)$.

The following result is decisive for coefficient problems: Set $h_{\mu\nu}=\sum_{j=0}^{\mu}\bar c_{\mu-j}c_{\nu-j}$ $(\mu\leqslant\nu)$, $h_{\nu\mu}=h_{\mu\nu}$ for $f(z)=\sum_{n=0}^{\infty}c_nz^n$, let m_n^2 be the maximal eigenvalue (a nonnegative real number) of the Hermitian matrix $(-h_{\mu\nu})_{\mu,\nu=0}^{n}$, and let $m=\lim m_n(\geqslant0)$.

Let H be a Hermitian form in infinitely many variables given by

$$H=m^2\sum_{\nu=0}^{\infty}\bar x_\nu x_\nu-\sum_{\mu,\nu=0}^{\infty}h_{\mu\nu}\bar x_\mu x_\nu$$

$$=m^2\sum_{\nu=0}^{\infty}|x_\nu|^2-\sum_{\mu=0}^{\infty}\left|\sum_{\nu=0}^{\infty}c_\nu x_{\mu+\nu}\right|^2.$$

Then a necessary and sufficient condition for $|f(z)|\leqslant1$ when $|z|<1$ is that H be positive semidefinite, i.e., the sequence of the †principal minor determinants of H,

$$\Delta\begin{pmatrix}1&2&\ldots&m\\1&2&\ldots&m\end{pmatrix},\quad m=1,2,\ldots,$$

is all positive or positive for an initial finite number of them and zero for the remainder (I. Schur).

A corresponding result for functions with positive real part in a disk can be stated in a simpler form: A holomorphic function $f(z)=1/2+\sum_{n=1}^{\infty}c_nz^n$ in $|z|<1$ satisfies $\mathrm{Re}\,f(z)\geqslant0$ if and only if

$$\begin{vmatrix}1&c_1&\ldots&c_n\\\bar c_1&1&\ldots&c_{n-1}\\\vdots&&&\vdots\\\bar c_n&\bar c_{n-1}&\ldots&1\end{vmatrix}\geqslant0,\quad n=1,2,\ldots$$

(C. Carathéodory). Furthermore, for $n=1,2,\ldots$, when we regard $(c_1,\ldots,c_n)$ as a point of complex n-dimensional Euclidean space, we can determine the domain of existence of points satisfying this criterion by Carathéodory. This result is generalized for coefficients of the Laurent expansion of a function that is holomorphic and single-valued in an annulus.

Next, there are some results for a function that omits two values: If $f(z)=a_0+a_1z+\ldots$ $(a_1\neq0)$ is holomorphic in $|z|<R$ and $f(z)\neq0,1$ in the same domain $|z|<R$, then there exists a constant $L(a_0,a_1)$, depending only on a_0 and a_1, such that $R\leqslant L(a_0,a_1)$ **(Laudau's tbeorem).** Under these circumstances, $|f(z)|\leqslant S(a_0,\theta)$ in $|z|\leqslant\theta R$ for $0<\theta<1$, where $S(a_0,\theta)$ is a constant depending only on a_0 and θ **(Scbottoky's tbeorem).** These theorems have applications in value distribution theory. On the other hand, coefficient problems have been investigated as extremal problems in H^p-spaces under general conditions.

K. Angular Derivative

Let f(z) be holomorphic in $|z|<1$. If $f(z)\to w_0$ uniformly as $z\to z_0$ along Stolz paths with end point at z_0 and if the limit $\lim_{z\to z_0}((f(z)-w_0)/(z-z_0))\equiv D$ exists, we call D the **angular derivative** of $f(z)$ at z_0. In the case of the half-plane Re z > 0, the angular derivative at the point z_0 on the imaginary axis is similarly defined. It should be noted that $f(z)-w_0$ is replaced by $1/f(z)$ for $w_0=\infty$ and $1/(z-z_0)$ by z for $z_0=\infty$. In the latter case, a Stolz path is a path contained in an angular domain $|\arg z|\leqslant\alpha$ $(<\pi/2)$ and tending to ∞. The study of angular derivatives was initiated by G. Julia (1920) and J. Wolff (1926) and was further advanced by Carathéodory (1929) and E. Landau and G. Varilon (1929).

A fundamental theorem for angular derivatives can be stated as follows: If a holomorphic function f(z) in Re z > 0 satisfies $\mathrm{Re}\,f(z)\geqslant0$, there exists a constant c $(0\leqslant c<+\infty)$ such that $f(z)/z\to c$ and $f'(z)\to c$ uniformly as $z\to\infty$ along every Stolz path. Moreover, the pth derivative of f(z) for an arbitrary positive p, denoted by $D^pf(z)$, has the property that $z^{p-1}D^pf(z)\to c/\Gamma(2-p)$ uniformly. Further-

more, the inequality Re f(z) $\geqslant$ c Re z holds everywhere in Re $z > 0$. An analogous theorem is valid for the unit disk.

An important problem in the theory of conformal mapping is to find some condition for a mapping $w = f(z)$ of the unit disk (or half-plane) G onto a simply connected domain B to have a nonzero and finite angular derivative at a boundary point z_0, that is, the condition for conformality at the boundary point. Carathéodory showed that a sufficient condition is the existence of two circles that are mutually inscribed and circumscribed at the boundary point $w_0 = f(z_0) \neq \infty$ of B and that lie inside and outside of B, respectively. Ahlfors later established a necessary and sufficient condition for the existence of the angular derivative by making use of his †distortion theorem for a strip domain. The angular derivative was used by Wolff in his research on the iteration of conformal mappings.

References

[1] L. Bieberbach, Lehrbuch der Funktionentheorie II, Teubner, 1931 (Johnson Reprint Co., 1969).
[2] C. Carathéodory, Funktionentheorie 1, II, Birkhäuser, Basel, 1950; English translation, Theory of functions, Chelsea, I, 1958; II, 1960.
[3] E. G. H. Landau, Darstellung und Begründung einiger neuerer Ergebnisse der Funktionentheorie, Springer, 1929.
[4] J. E. Littlewood, Lectures on the theory of functions I, Clarendon Press, 1944.
[5] P. L. Duren, Theory of H^p-spaces, Academic Press, 1970.
[6] R. Nevanlinna, Eindeutige Analytische Funktionen, Springer, second edition, 1953; English translation, Analytic functions, Springer, 1970.
[7] L. Carleson, Interpolations by bounded analytic functions and the corona problem, Ann. Math., (2) 76 (1962).
[S] H. S. Shapiro and A. L. Shield, On some interpolation problem for analytic functions, Amer. J. Math., 83 (1961).
[9] K. Hoffman, Banach spaces of analytic functions, Prentice-Hall, 1962.
[10] K. Hoffman, Bounded analytic functions and Gleason parts, Ann. Math., (2) 86 (1967).
[11] M. L. Weiss, Cluster sets of bounded analytic functions from a Banach algebraic viewpoint, Ann. Acad. Sci. Fenn., 367 (1965).
[12] T. W. Gamelin, Lectures on H"(D), Univ. Nacional de la Plata, 1972.
[13] T. W. Gamelin, The algebra of bounded analytic functions, Bull. Amer. Math. Soc., 79 (1973).

44 (XVII.1 0) Branching Processes

A. General Remarks

A **brancbing process** is a mathematical model for random motion of a family of particles each of which is in an isolated process of multiplication and death. Examples of such random motions are population growth, miosis of genes, growth of the numbers of neutrons in an atomic chain reaction, and cascade showers of cosmic rays. The simplest and most fundamental branching process is discussed in Section B. For a historical introduction to the study of branching processes → Kendall [2] and Harris [3].

B. Galton-Watson Processes

Although there is a similar process with continuous parameter $t \in \mathbf{R}$, we consider here only the case of a discrete time parameter ($i = 0, 1, 2, .$). Suppose that we are given a family of particles of the same kind. Each member of the family splits into several particles according to a given probability law independently of the other members and its own past history. Let Z_n be the number of particles of the family at a moment (or generation) n; then $\{Z_n\}$ gives rise to a †Markov chain. This is called the **Galton-Watson process**. A precise mathematical description of the process is as follows: A Galton-Watson process is a Markov chain $\{Z_n\ n = 0, 1, \ldots\}$ on the nonnegative integers with †transition probability defined in terms of a given †probability distribution $\{p_k\ k = 0, 1, \}$, $p_k \geqslant 0$, $\sum_{k=0}^{\infty} p_k = 1$, by

$$p_{ij} = P[Z_{n+1} = j \mid Z_n = i] = \begin{cases} p_j^{*i} & \text{if } i \geqslant 1,\ j \geqslant 0, \\ \delta_{ij} & \text{if } i = 0,\ j \geqslant 0, \end{cases}$$

where the probability distribution $\{p_k^{*i}\ k = 0, 1, .\}$ denotes the i-fold †convolution of the probability distribution $\{p_k\ k = 0, 1, \ldots\}$ and δ_{ij} is the †Kronecker delta. Then p_k is interpreted as the probability that an object existing in the nth generation has k children in the (n + 1)th generation. We assume that $Z_0 = 1$ a.s. for the rest of this section. An important tool in the analysis of the process is the †generating function f(s) = $\sum_{j=0}^{\infty} p_j s^j$, $|s| \leqslant 1$. Denote the generating function $\sum_{k=0}^{\infty} P(Z_n = k) s^k$ ($s \leqslant 1$) of Z_n by $f_n(s)$. Then we have $f_0(s) = s$, $f_1(s) = f(s)$, $f_{i+j}(s) = f_i(f_j(s))$ ($i, j = 0, 1, 2, .$), and the †expectation of Z_n is given by $E(Z_n) = m^n$, where $m = f'(1) = E(Z_1)$ is the expectation of Z,. A Galton-Watson process is said to be **subcritical, critical,** or **supercritical** if $m < 1$, $= 1$, or > 1, respectively. If we have $Z_n =$

0 for some n, then $Z_{n+1}=Z_{n+2}=\ldots=0$. The probability $q=P$ (lim, $Z_n=0$) is called the **extinction probability.** The case $f_1(s)\equiv s$ is excluded in the following. The extinction probability q of the process $\{Z_n\}$ is the smallest nonnegative solution of the equation $s=f(s)$. It is 1 if $m\leqslant 1$ and <1 if $m>1$. Moreover, $P(\lim_n Z_n=\infty)=1-q$ when $m>1$, and hence the process $\{Z_n\}$ is †transient.

If we put $W_n=Z_n/m^n$ when $m<\infty$, $\{W_n\}$ gives rise to a †martingale, and the limit $W=\lim_{n\to\infty} W_n$ exists with probability 1. If $m>1$, then the †moment-generating function $\varphi(s)=E(\exp(-sW))$ of W satisfies the **Königs-Schröder equation** $\varphi(ms)=f(\varphi(s))$, Re $s\geqslant 0$. If $m\leqslant 1$, then $P(Z_n=0$ for some $n)=1$. The following theorem on the conditional distribution of Z_n, given that $Z_n\neq 0$, was first proved by Yaglom [6] under moment restrictions. If $m<1$, there exist $b_k=\lim$, $P(Z_n=k \mid Z_n\neq 0)$, $k=1,2,\ldots$. Then $\{b_k\mid k=1,2,\ldots\}$ is a probability distribution and its generating function $g(s)=\sum_{k=1}^{\infty} b_k s^k$, $|s|\leqslant 1$, is the unique solution of the equation $g(f(s))=mg(s)+1-m$, $|s|\leqslant 1$, among generating functions vanishing at 0. Furthermore, when $m<1$, $\sum_{j=1}^{\infty} jb_j<\infty$ if and only if $\sum_{j=1}^{\infty} p_j(j\log j)<\infty$ (i.e., $E[Z_1\log Z_1]<\infty$).

If $m=1$ and $\sigma^2=$ variance of $Z_1<\infty$, then $\lim_n P[Z_n/n>u\mid Z_n\neq 0]=\exp[-2u/\sigma^2]$, $u\geqslant 0$. If $\sigma^2=\infty$, then this result is still correct with the interpretation that the limit in the left-hand side equals 1.

For further details regarding the theory of Galton-Watson processes → [3, S].

C. Multitype Galton-Watson Processes

These are generalizations of the Galton-Watson process, involving k types of particles ($k\geqslant 2$), say, $T_1, T_2, \ldots, T_k$. Let $p^l(r_1, r_2, \ldots, r_k)$ be the probability that a particle of type T_l splits into r_m particles of type T_m ($m=1,2,\ldots,k$), and set $f^l(s_1, s_2, \ldots, s_k)=\sum_r p^l(r_1, r_2, \ldots, r_k)s_1^{r_1}s_2^{r_2}\ldots s_k^{r_k}$. The number of particles $Z_n=(Z_n^1, Z_n^2, \ldots, Z_n^k)$ at a moment n gives rise to a †Markov chain over Z_+^k, and its transition probability is determined by $p^l(r_1, r_2, \ldots, r_k)$ as for Galton-Watson processes. Here $\mathbf{Z}_+^k$ denotes the space of all k-dimensional †lattice points whose components are nonnegative integers. We now define a mathematical model of the process mentioned above. A Markov chain $\{Z_n\mid n=0,1,\ldots\}$ on $\mathbf{Z}_+^k$ is called a **multi (k)-type Galton-Watson process** if its transition probability is given by

$$P(\mathbf{i},\mathbf{j})=P[Z_{n+1}=\mathbf{j}\mid Z_n=\mathbf{i}]=\text{coefficient of } \prod_{l=1}^{k} s_l^{j_l} \text{ in } \prod_{l=1}^{k}(f^l(s_1,s_2,\ldots,s_k))^{i_l}$$

for $\mathbf{i}=(i_1,i_2,\ldots,i_k)$, $\mathbf{j}=(j_1,j_2,\ldots,j_k)\in\mathbf{Z}_+^k$, $n=0,1,\ldots$ and $s=(s_1,s_2,\ldots,s_k)\in D_k$, where $D_k=\{s=(s_1,s_2,\ldots,s_k)\in\mathbf{R}^k \mid |s_l|\leqslant 1, l=1,2,\ldots,k\}$.

When $m_{ij}=\partial f^i(1,\ldots,1)/\partial s_j<\infty$ ($i,j=1,2,\ldots,k$), the †conditional expectation of Z_{n+m}, given Z_m, is given as $E[Z_{m+n}\mid Z_m]=Z_m M^n$, where M is the matrix (m_{ij}). From now on we assume that $m_{ij}<\infty$ ($i,j=1,2,\ldots,k$). A multi-type Galton-Watson process $\{Z_n \mid n=0,1,\ldots\}$ is said to be singular if the generating functions $f^l(s_1,s_2,\ldots,s_k)$, $l=1,2,\ldots,k$, are all linear in $s_1, s_2, \ldots, s_k$ with no constant terms. We assume nonsingularity throughout. If there is a positive integer N such that every component of M^N is positive, the process Z_n is said to be **positively regular.** In this case, M has a positive †eigenvalue λ that is simple, and $\lambda>|\mu|$ for all other eigenvalues μ. The eigenvalue λ plays the role of m in the case of the Galton-Watson process. For every i, $1\leqslant i\leqslant k$, set $q^i=P[Z_n=0$ for some $n\mid Z_0=e_i]$, where $e_i=(e_i^1,e_i^2,\ldots,e_i^k)$, $e_i^j=\delta_{ij}$. Then q^i is called the extinction probability of the process $\{Z_n \mid n=0,1,\ldots\}$ starting with a single particle of type T_i; let $\mathbf{q}=(q^1,q^2,\ldots,q^k)$. Suppose that the process $\{Z_n \mid n=0,1,\ldots\}$ is positively regular and not singular. Then $\mathbf{q}=(1,1,\ldots,1)$ if $\lambda\leqslant 1$, and $0\leqslant q^i<1$ ($i=1,2,\ldots,k$) if $\lambda>1$, and $\mathbf{q}$ satisfies the equation $\mathbf{q}=(f^1(\mathbf{q}),f^2(\mathbf{q}),\ldots,f^k(\mathbf{q}))$. Furthermore, the process $\{Z_n \mid n=0,1,\ldots\}$ is †transient, i.e., $P[Z_n=\mathbf{j}$ infinitely often$]=0$ for any $\mathbf{j}\neq 0$, $\mathbf{j}\in\mathbf{Z}_+^k$. Moreover, if $\lambda>1$, then $W_n=Z_n/\lambda^n$ converges with probability 1, and

$$\lim_{n\to\infty} Z_n/\lambda^n=vW \quad \text{a.s.},$$

where v denotes the positive left †eigenvector $v=(v_1,\ldots,v_k)$ for λ of the matrix M, and W is a nonnegative †random variable. $P[W>0]>0$ holds if and only if

$$E[Z_{1j}^{(i)}\log Z_{1j}^{(i)}]<\infty \quad \text{for all} \quad 1\leqslant i,j\leqslant k. \tag{1}$$

Here $Z_{1j}^{(i)}$ denotes the number of type T_j particles in the first generation for a process with $Z_0=e_i$, $1\leqslant i\leqslant k$. If condition (1) holds, the †moment-generating functions of W, $\varphi_i(\alpha)=E[\exp[-\alpha W]\mid Z_0=e_i]$, $i=1,2,\ldots,k$, $\alpha>0$, satisfy $\varphi_i(\alpha)=f^i(\varphi_1(\alpha/\lambda), \varphi_2(\alpha/\lambda),\ldots,\varphi_k(\alpha/\lambda))$, $1\leqslant i\leqslant k$.

When $\lambda\leqslant 1$, a theorem analogous to Yaglom's holds. For further information → Joffe and Spitzer (*J. Math. Anal. Appl.* 19 (1967)) and Athreya and Ney [S].

D. Markov Branching Processes

The branching processes dealt with in Sections B and C are limited in the sense that generation times are fixed. We now formulate a continuous-time version of branching processes. The treatment in this section is limited

to Markov processes that are extensions to continuous time of the Galton-Watson process. A Markov process $\{X_t, P_i\}$ on $\mathbf{Z}_+^1$ is called a **Markov branching process** if the transition probability $P(t,i,j)=P_i[X_t=j]$ $(i,j\in\mathbf{Z}_+^1, t\geqslant 0)$ satisfies

$$\sum_{j=0}^{\infty} P(t,i,j)s^j=\left(\sum_{j=0}^{\infty} P(t,1,j)s^j\right)^i$$

for all $i\in\mathbf{Z}_+^1$ and $|s|\leqslant 1$. Then there exist a positive constant a and a probability distribution $\{p_k;\ k=0,2,3,\ldots\}$ such that $\lim_{t\downarrow 0}(1-P(t,i,i))/t=ia$ if $i\in\mathbf{Z}_+^1$ and $\lim_{t\downarrow 0}P(t,i,j)/t=iap_{j-i+1}$ if $j\geqslant i-1,\ j\neq i$. The †Kolmogorov equations are

$$\frac{d}{dt}P(t,i,j)=-jaP(t,i,j)+a\sum_{l=1,l\neq j}^{j+1} lp_{j-l+1}P(t,i,l)$$

(†forward equation)

and

$$\frac{d}{dt}P(t,i,j)=-iaP(t,i,j)+ia\sum_{l=i-1,l\neq i}^{\infty} p_{l-i+1}P(t,l,j)$$

(†backward equation).

Then a and $\{p_k\ \ k=0,2,3,\ldots\}$ are interpreted as follows. An object existing at t has a probability adt of dying in the time interval $[t,t+dt)$ of length dt. If it dies at any time t, the probabilities are $p_0,p_2,p_3,\ldots$ that it is replaced by $0,2,3,\ldots$ objects. When $\sum_{k=0}^{\infty}P(t,i,k)<1$, the number of particles attains $+\infty$ in a finite time interval with positive probability. In order that $\sum_{k=0}^{\infty}P(t,i,k)=1$ for all $i\in\mathbf{Z}_+^1$, it is necessary and sufficient that for each $1>\varepsilon>0$, $\int_{1-\varepsilon}^{1}(f(u)-u)^{-1}du=\infty$, where $f(s)=\sum_{j=0,j\neq 1}^{\infty}p_js^j$ for $|s|\leqslant 1$.

Most of the theory of the Galton-Watson process carries over to the continuous case. In particular, concerning the †limit distribution of X_t, more precise results have been obtained. (For detailed discussion of Markov branching processes → [3, 8].)

Some work has also been done on the asymptotic behavior for temporally inhomogeneous cases of multitype Markov branching processes.

E. Branching Markov Processes

We now give a systematic treatment of general branching processes in which an object is characterized by a parameter x in a †compact †Hausdorff space S with a countable open †base. The results are formulated in terms of strong Markov processes. Set $S^0=\{\partial\}$, where ∂ is an extra point. For every positive integer n, let $S^n=\underbrace{S\times S\times\ldots\times S}_{n}/\sim$, where $\sim$ is the †equivalence relation given by the †permutations of coordinates. S^n is compact with respect to the †quotient topology, and its †topological direct sum $\mathbf{S}=\sum_{n=0}^{\infty}S^n$ is †locally compact. Let $\hat{\mathbf{S}}=\mathbf{S}\cup\{\Delta\}$ be the one-point †compactification of $\mathbf{S}$. A point $\mathbf{x}\in\hat{\mathbf{S}}$ is denoted by $\mathbf{x}=[x_1,x_2,\ldots,x_n]$ if $\mathbf{x}$ is the †equivalence class containing $(x_1,x_2,\ldots,x_n\in S^n)$. $\mathbf{x}=[x]$ is denoted simply by x. Set $\mathbf{B}(S)=\{f\,|\,f$ a bounded Borel measurable function on $S\}$, $\mathbf{B}^*(S)=\{f\in\mathbf{B}(S)\,|\,\|f\|<1\}$, $\mathbf{C}(S)=\{f\,|\,f$ a bounded continuous function on $S\}$, and $\mathbf{C}^*(S)=\mathbf{C}(S)\cap\mathbf{B}^*(S)$. $\mathbf{B}(\hat{\mathbf{S}})$, $\mathbf{C}(\mathbf{S}),\ldots$ are defined similarly. For $f\in\mathbf{B}^*(S)$, define $\hat{f}\in\mathbf{B}(\hat{\mathbf{S}})$ by $\hat{f}(\mathbf{x})=1$ if $\mathbf{x}=\partial$, $=\prod_{j=1}^{n}f(x_j)$ if $\mathbf{x}=[x_1,x_2,\ldots,x_n]\in S^n$, and $=0$ if $\mathbf{x}=\Delta$. A †strong Markov process $X=\{\mathbf{X}_t,\mathbf{P}_{\mathbf{x}}\}$ on $\hat{\mathbf{S}}$ is called a **branching Markov process** if its semigroup $\{\mathbf{T}_t\}$ (→ 261 Markov Processes) satisfies

$$\mathbf{T}_t\hat{f}(\mathbf{x})=\widehat{(\mathbf{T}_t\hat{f})_{|S}}(\mathbf{x})\quad\text{for every}\quad \mathbf{x}\in\hat{\mathbf{S}}\quad\text{and}\quad f\in\mathbf{B}^*(S),$$

where for $g\in\mathbf{B}(\hat{\mathbf{S}})$, $g_{|S}$ is the restriction of g on S [10–12]. We set (i) $Z_t=n$ if $\mathbf{X}_t\in S^n$, $n=0,1,\ldots,\infty$, where $S^\infty=\{\Delta\}$, and (ii) $\tau=\inf\{t\ \ Z_t\neq Z_0\}$. Z_t and τ are called the number of particles and the **first splitting time,** respectively. Let $X=\{x_t,P_x\}$ be a strong Markov process on S. We assume that the semigroup $\{H_t\}$ of X is a strongly continuous semigroup on C(S). Let φ_t be a nonnegative continuous †additive functional of X. Let $p_n(x)$, $n=0,2,3,\ldots$ be a sequence of nonnegative functions in $\mathbf{B}(S)$ such that $\sum_{n=0,n\neq 1}^{\infty}p_n(x)=1$. Consider a sequence of stochastic kernels $\pi_n(x,dy)$, $n=0,2,3,\ldots$ on $S\times S^n$, i.e., for fixed $A\in\mathfrak{B}(S^n)$, $\pi_n(x,A)$ is a measurable function of x, and for fixed x it is a probability on $(S^n,\mathfrak{B}(S^n))$, where $\mathfrak{B}(S^n)$ is the †topological σ-field on S^n. Set $\pi(x,D)=\sum_{n=0,n\neq 1}^{\infty}p_n(x)\pi_n(x,D\cap S^n)$ for $D\in\mathfrak{B}(\hat{\mathbf{S}})$, $x\in S$, where $\mathfrak{B}(\hat{\mathbf{S}})$ is the topological σ-field on $\hat{\mathbf{S}}$. Then there exists a unique branching Markov process $X=\{\mathbf{X}_t,\mathbf{P}_{\mathbf{x}}\}$ on $\hat{\mathbf{S}}$ satisfying the following: (i) The Markov process $\{\mathbf{X}_t, t<\tau, \mathbf{P}_x\}$ on S obtained from X by shortening its †lifetime is equivalent to the Markov process obtained by †killing X at a rate $d\varphi_t$ (→ 261 Markov Processes). (ii) For any $D\in\mathfrak{B}(\hat{\mathbf{S}})$ and $\lambda>0$, $E_x[\exp[-\lambda\tau];\ \mathbf{X}_\tau\in D\,|\,\mathbf{X}_{\tau-}]=E_x[\exp[-\lambda\tau]\,|\,\mathbf{X}_{\tau-}]\pi(\mathbf{X}_{\tau-},D)$ a.s. on $\{\tau<\infty\}$ for every $x\in S$. (iii) With probability one, $\mathbf{X}_t=\Delta$ for all $t\geqslant\lim_{n\uparrow\infty}\tau_n$, where τ_n denotes the nth splitting time [10]. This process $X=\{\mathbf{X}_t,\mathbf{P}_{\mathbf{x}}\}$ is referred to as the $\{x_t,\varphi_t,\pi\}$-branching Markov process.

Example 1. Let $S=\{a_1,a_2,\ldots,a_k\}$; then $\mathbf{S}$ can

be identified with $\mathbf{Z}_+^k$ and $\hat{S}$ with $\widehat{\mathbf{Z}_+^k} = \mathbf{Z}_+^k \cup \{\infty\}$. Therefore a branching Markov process X on $\hat{S}$ is a Markov process on $\widehat{\mathbf{Z}_+^k}$ such that its transition probability $P(t, \mathbf{i}, \mathbf{j})$, $\mathbf{i}, \mathbf{j} \in \mathbf{Z}_+^k$ satisfies

$$\sum_{j_1, j_2, \ldots, j_k \in \mathbf{Z}_+^k} P(t, \mathbf{i}, (j_1, j_2, \ldots, j_k)) s_1^{j_1} s_2^{j_2} \ldots s_k^{j_k}$$

$$= \prod_{l=1}^{k} \left(\sum_{(j_1, j_2, \ldots, j_k) \in \mathbf{Z}_+^k} P(t, e_l, (j_1, j_2, \ldots, j_k)) \times s_1^{j_1} s_2^{j_2} \ldots s_k^{j_k} \right)^{i_l},$$

$P(t, \infty, \infty) = 1$ for $\mathbf{i} = (i_1, i_2, \ldots, i_k) \in \mathbf{Z}_+^k$,

$s = (s_1, s_2, \ldots, s_k) \in D_k$.

Then the process X is called a **multi (Q-type Markov branching process.**

Example 2. Let $S = [0, \infty]$, $k(x)$ be a nonnegative locally integrable function on $[0, \infty)$, and $p_n(x)$, $n = 0, 1, \ldots$, be a sequence of nonnegative measurable functions on $[0, \infty)$ such that $\sum_{n=0}^{\infty} p_n(x) = 1$ and $p_1(x) = 0$. We extend $k(x)$ and $p_n(x)$ ($n \neq 2$) as functions on $[0, \infty]$ by setting them to 0 at ∞, and set $p_2(\infty) = 1$. Define a stochastic kernel $\pi(x, dy)$ on $S \times \hat{S}$ by

$$\pi(x, D) = \begin{cases} \sum_{n=0}^{\infty} p_n(x) \delta_{\{[\underbrace{0, 0, \ldots, 0}_{n}]\}} (D \cap S^n) & \text{if } x \in [0, \infty), \\ \delta_{\{[\infty, \infty]\}} (D \cap S^2) & \text{if } x = \infty. \end{cases}$$

Let $X = \{x_t, P_x\}$ be the uniform motion on S with the semigroup $\{H_t\}$ such that for $f \in C(S)$, $H_t f(x) = f(x+t)$ if $x \in [0, \infty)$ and $= f(\infty)$ if $x = \infty$. Set $\varphi_t = \int_0^t k(x_s) ds$. Then we have a $\{x_t, \varphi_t, \pi\}$-branching Markov process, and we call it an **age-dependent branching process.**

We now discuss some fundamental properties of $\{x_t, \varphi_t, \pi\}$-branching Markov processes. For the sake of simplicity we assume the following: (i) $\varphi_t = \int_0^t k(x_s) ds$, $k \in \mathbf{C}^+(S) = \{f \in \mathbf{C}(S) \mid f \geqslant 0\}$, (ii) $p_n(x) \in \mathbf{C}^+(S)$ and

$$\pi_n(x, D) = \delta_{\{[\underbrace{x, x, \ldots, x}]\}}(D)$$

for $D \in \mathfrak{B}(S^n)$. Then $\{\mathbf{T}_t\}$ is a strongly continuous semigroup on $C_\Delta(S) = \{f \in \mathbf{C}(\hat{\mathbf{S}}) \mid \lim_{\mathbf{x} \to \Delta} f(\mathbf{x}) = 0\}$. Let A be the †infinitesimal generator of $\{H_t\}$ (→ 261 Markov Processes) and $\mathscr{D}(A)$ be the domain of A. If $f \in \mathscr{D}(A) \cap C^*(S)$, then $u(t, x) = (\mathbf{T}_t \hat{f})_{|S}(x) \in \mathscr{D}(A)$, and it satisfies $\partial u / \partial t = Au + k(F(\cdot; u) - u)$, $u(0+, x) = f(x)$, where $F(x; g) = \int_{\hat{S}} \hat{g}(\mathbf{y}) \pi(x, d\mathbf{y})$ [10].

Let $\mathscr{M}$ be the space of all nonnegative †Radon measures on S and $\mathscr{M}_0$ be the subspace of $\mathscr{M}$ consisting of all probability Radon measures on S endowed with the topology of †weak convergence. Set $\bar{\mathscr{M}} = \mathscr{M} \cup \{\delta\}$, where δ is an extra point and $\tilde{\mathscr{M}} = [0, \infty] \times \mathscr{M}_0$. Define a mapping $\rho: \tilde{\mathscr{M}} \ni (\bar{\lambda}, \lambda_0) \mapsto \rho(\bar{\lambda}, \lambda_0) \in \bar{\mathscr{M}}$ by $\rho(\bar{\lambda}, \lambda_0) = \bar{\lambda}\lambda_0$ if $\lambda < \infty$ and $= \delta$ if $\lambda = \infty$, and define the topology of $\bar{\mathscr{M}}$ as the strongest of all the topologies rendering ρ continuous. Set $\mathbf{C}^{++}(S) = \{f \in \mathbf{C}(S) \mid f > 0\}$, and for $f \in \mathbf{C}^{++}(S)$, define a function $\Phi_f(\lambda)$ on $\bar{\mathscr{M}}$ by $\Phi_f(\lambda) = \exp[-(\lambda, f)]$ if $\lambda \in \mathscr{M}$ and $= 0$ if $\lambda = \delta$, where $(\lambda, f) = \int_S f(x) \lambda(dx)$. Consider a mapping $\psi: \hat{\mathbf{S}} \to \bar{\mathscr{M}}$ defined by $\psi(\mathbf{x}) = 0$ if $\mathbf{x} = \partial$, $= \sum_{k=1}^{n} \delta_{\{x_k\}}$ if $\mathbf{x} = [x_1, x_2, \ldots, x_n] \in S^n$, and $= \delta$ if $\mathbf{x} = \Delta$. Let $\mathscr{M}_p$ be the subspace of $\bar{\mathscr{M}}$ consisting of all nonnegative integer-valued Radon measures on S, i.e., $\mathscr{M}_p = \{\psi(\mathbf{x}) \mid \mathbf{x} \in \hat{\mathbf{S}}\}$. We denote by $\mathscr{D}$ the space of all right-continuous mappings v: $[0, \infty) \to \mathscr{M}_p$ whose discontinuities are at most of the first kind and such that $v_t = \delta$ for $t \geqslant s$ if $v_s = \delta$. Consider a branching Markov process $\{\mathbf{X}_t, \mathbf{P}_{\mathbf{x}}\}$, and let $P_{\mu_{\mathbf{x}}}$ be the †probability law on $\mathscr{D}$ of the stochastic process $\{v_t = \psi(\mathbf{X}_t); 0 \leqslant t < \infty\}$ with $\mathbf{X}_0 = \mathbf{x}$, where $\mu_{\mathbf{x}} = \psi(\mathbf{x})$. Then the branching property can be rewritten as follows: For every $t \geqslant 0$, $f \in \mathbf{C}^{++}(S)$ and $\mu_1, \mu_2 \in \mathscr{M}_p$, $E_{\mu_1 + \mu_2}[\exp[-(v_t, f)]] = E_{\mu_1}[\exp[-(v_t, f)]] E_{\mu_2}[\exp[-(v_t, f)]]$ (→ Silverstein, Z. *Wahrscheinlichkeitstheorie und Verw. Gebiete,* 9 (1968); for a treatment of branching processes in the framework of †martingale problems → Holley and Stroock, *Publ. Res. Inst. Math. Sci.*, 14 (1978)).

There are analogous processes which preserve the basic features of the branching property. A Markov process $\{\mu_t, P_\mu\}$ on $\bar{\mathscr{M}}$ with δ as a †trap is called a **continuous-state branching process** if it satisfies the property that for every $t \geqslant 0$, $f \in \mathbf{C}^{++}(S)$, and $\mu_1, \mu_2 \in M$, $E_{\mu_1 + \mu_2}[\exp[-(\mu_t, f)]] = E_{\mu_1}[\exp[-(\mu_t, f)]] E_{\mu_2}[\exp[-(\mu_t, f)]]$. This concept was introduced by Jiřina [15] for some special cases. For further information → Jiřina *(Third Prague Conf.*, 1964), Lamperti (Bull. *Amer. Math.* Soc., 73 (1967)), Watanabe (*J. Math. Kyoto Univ.*, 8 (1968)), and Fujimagari and Motoo *(Kôdai Math. Sem. Rep., 23* (1971)).

References

[1] F. Galton and H. W. Watson, On the probability of the extinction of families, J. Roy. Anthropol. Inst., 4 (1874), 138-144.

[2] D. Kendall, Branching processes since 1873, J. London Math. Soc., 41 (1966), 385–406.

[3] T. E. Harris, The theory of branching processes, Springer, 1963.

[4] M. S. Bartlett, An introduction to stochastic processes, Cambridge Univ. Press, 1966.

[5] A. N. Kolmogorov and N. A. Dmitriev, Branching random processes (in Russian), Dokl. Akad. Nauk SSSR, (N.S.) 56 (1947), 5–8.

[6] A. M. Yaglom, Some limiting theorems of branching random processes (in Russian), Dokl. Akad. Nauk SSSR, (N.S.) 56 (1947), 795–798.
[7] H. Kesten and B. P. Stigum, A limit theorem for multidimensional Galton-Watson processes, Ann. Math. Statist., 37 (1966), 1211–1223.
[8] K. B. Athreya and P. E. Ney, Branching processes, Springer, 1972.
[9] B. A. Sevast'yanov, Branching processes (in Russian), Nauka, 1971.
[10] N. Ikeda, M. Nagasawa, and S. Watanabe, Branching Markov processes I, II, III, J. Math. Kyoto Univ., 8 (1968), 233-278, 365–410; 9 (1969), 95–160; 11 (1971), 195–196.
[11] J. E. Moyal, The general theory of stochastic population processes, Acta Math., 108 (1962), 1–31.
[12] A. V. Skorokhod, Branching diffusion processes, Theory of Prob. Appl., 9 (1964), 445–449. (Original in Russian, 1964.)
[13] T. W. Mullikin, Limiting distributions for critical multitype branching processes with discrete time, Trans. Amer. Math. Soc., 106 (1963), 469-494.
[14] S. Watanabe, Limit theorem for a class of branching processes, Markov Processes and Potential Theory, J. Chover (ed.), Wiley, 1967, 205–232.
[15] M. Jiřina, Stochastic branching processes with continuous state space, Czech. Math. J., 8 (1958), 292–313.

45 (XVII.7) Brownian Motion

A. General Remarks

R. Brown, an English botanist, observed in 1827 that the minute particles comprising the pollen of plants, when suspended in water, exhibit peculiarly erratic movements [1]. The physical explanation of this phenomenon is that haphazard impulses are given to the suspended particles by collisions with molecules of the fluid. Let X(t) be the x-coordinate of a particle at time t. Then X(t) is treated as a †random variable, and the distribution of **X(t)** -X(s) is a †normal distribution $N(0, D|t-s|)$, with †mean 0 and †variance $D|t-s|$, where D is a positive constant. To be more exact, such a family of random variables {X(t)} is now considered as the family of random variables determining a †stochastic process. Various aspects of the theory were analyzed by A. Einstein [2], L. Bachelier, N. Wiener, P. Lévy, and others.

B. Wiener Processes

Let T be the real line $\mathbf{R}^1$ or a subinterval. A †stochastic process $\{X(t)\}_{t\in T}$ defined on a †probability space $(\Omega, \mathfrak{B}, P)$ is called a **Wiener process** on $\mathbf{R}^d$ if it satisfies the following three conditions: (1) $X(t,\omega)\in\mathbf{R}^d$ $(t\in T, \omega\in\Omega)$. (2) The $\mathbf{R}^d$-valued random variables $X(t_1)$, $X(t_j)-X(t_{j-1})$, $j=2,3,\ldots,n$, are †independent for any $t_1<t_2<\ldots<t_n$, $t_j\in T$ $(j=1,2,\ldots,n)$, where n is an arbitrary positive integer. (3) If $X_i(t)$ is the ith component of the vector X(t), then the $\{X_i(t)\}_{t\in T}$ $(1\leqslant i\leqslant d)$ are independent as stochastic processes and every increment $X_i(t)-X_i(s)$ is normally distributed with mean 0 and variance $|t-s|$. A Wiener process is also called a **Brownian motion.** A Wiener process is a †temporary homogeneous additive process. A †separable Wiener process has continuous paths with probability 1. Conversely, if $\{X(t)\}_{t\in T}$ is a temporary homogeneous additive process on $\mathbf{R}^1$ whose †sample function is continuous with probability 1 and the increment $X(t)-X(s)$ has mean 0 and variance $|t-s|$, then it is a Wiener process (→ 407 Stochastic Processes, 406 Stochastic Differential Equations). Let $\{\mathscr{F}_t\}_{t\geqslant 0}$ be an †increasing family of σ-subfields of $\mathfrak{B}$. We assume that $\{\mathscr{F}_t\}$ is right continuous. A d-dimensional continuous process $X=(X(t))_{t\geqslant 0}$ is called a d-dimensional $\{\mathscr{F}_t\}$**-Brownian motion** if it is †$\{\mathscr{F}_t\}$-adapted and satisfies

$$E[\exp[i\langle\xi, X(t)-X(s)\rangle]\,|\,\mathscr{F}_s]$$
$$=\exp[-(t-s)|\xi|^2/2]$$
$$\text{a.s. for every } \xi\in\mathbf{R}^d,\quad 0\leqslant s<t,$$

where $|\xi|$ denotes the †norm of $\xi\in\mathbf{R}^d$ [9]. Then X satisfies conditions (1)–(3) mentioned above, and hence it is a Brownian motion on $\mathbf{R}^d$ (→ 406 Stochastic Differential Equations).

Let $\{X_k(\omega)\}$ $(k=0,1,\ldots)$ be a sequence of independent random variables defined on a probability space $(\Omega, \mathfrak{B}, P)$ such that each X_k has the normal distribution $N(0,1)$ with mean 0 and variance 1. Then the series

$$\frac{t}{\sqrt{\pi}}X_0(\omega)+\sqrt{\frac{2}{\pi}}\sum_{n=1}^{\infty}X_n(\omega)\frac{\sin nt}{n},\qquad t\in[0,\pi]\equiv T,$$

converges uniformly in $t\in T$ with probability 1, and its limit, denoted by $X(t,\omega)$, is a Wiener process [3, 10].

Let $W^d=C([0,\infty)\to\mathbf{R}^d)$ be the space of all continuous functions w: $[0,\infty)\to\mathbf{R}^d$ endowed with the topology of uniform convergence on finite intervals and $\mathfrak{B}(W^d)$ be the †topological σ-field. Let $X=(X(t))_{t\geqslant 0}$ be a Brownian motion on $\mathbf{R}^d$ and μ be the probability law of X(0). The probability law P^X of the Brownian motion X on $(W^d, \mathfrak{B}(W^d))$ is called the d-

dimensional **Wiener measure with the initial distribution** μ [9, 11, 12]. For a probability μ on $(\mathbf{R}^d, \mathfrak{B}(\mathbf{R}^d))$ the d-dimensional Wiener measure P_μ with the initial distribution μ is a probability on $(W^d, \mathfrak{B}(W^d))$ characterized by the property that for every $0 = t_0 < t_1 < \ldots < t_n$ and $A_j \in \mathfrak{B}(\mathbf{R}^d), j = 1, 2, \ldots, n,$

$$P_\mu[w(t_1)\in A_1, w(t_2)\in A_2, \ldots, w(t_n)\in A_n]$$

$$= \int_{\mathbf{R}^d}\int_{A_1}\cdots\int_{A_n}\prod_{j=1}^{n}\left(\frac{1}{\sqrt{2\pi(t_j-t_{j-1})}}\right)^d \times \exp\left[-\frac{|x_j-x_{j-1}|^2}{2(t_j-t_{j-1})}\right]\mu(dx_0)\,dx_1\ldots dx_n,$$

where $\mathfrak{B}(\mathbf{R}^d)$ denotes the topological σ-field on Rd.

C. Brownian Motion as a Diffusion Process

In terms of the general framework of †Markov processes, Brownian motion is a typical example of a diffusion process (→ 115 Diffusion Processes). Let $X = \{X_t(w), \mathbf{R}^d, P_x\}$ be a continuous Markov process on $\mathbf{R}^d$ (→ 261 Markov Processes) with the †transition probability

$$P(t,x,B) = \int_B (2\pi t)^{-d/2}\exp\left(-\frac{1}{2t}|x-y|^2\right)dy,$$

$$t > 0, \quad x\in\mathbf{R}^d, \quad B\in\mathfrak{B}(\mathbf{R}^d).$$

For each $x \in \mathbf{R}^d$, the process $\{X_t(w), P_x\}$ is a Wiener process in the sense mentioned above. The Markov process X, which is a collection of Wiener processes $\{X_t, P_x\}$ starting at x, is said to be a **d-dimensional Brownian motion.** A d-dimensional Brownian motion possesses the †strong Markov property. Let $\mathfrak{G}$ be the †generator of the †semigroup T_t corresponding to X. A bounded uniformly continuous function f defined on $\mathbf{R}^d$ belongs to the domain of $\mathfrak{G}$ if its partial derivatives $\partial f/\partial x_i$ and $\partial^2 f/\partial x_i \partial x_j$, $i,j = 1, 2, \ldots d$, are bounded and uniformly continuous. For such a function f we have $\mathfrak{G}f(x) = (1/2)\Delta f(x)$, where Δ is the †Laplacian [8, 12].

D. Brownian Motions and Potentials

For $\alpha > 0$, the function $G_\alpha(x)$ defined by

$$G_\alpha(x) = \frac{1}{2}\int_0^\infty e^{-\alpha t}(2\pi t)^{-d/2}\exp\left(-\frac{1}{2t}|x|^2\right)dt$$

is said to be the **α-order Green's function.** Since Brownian motion is †nonrecurrent for $d \geqslant 3$, the limit $G_{0+}(x) = \lim_{\alpha\downarrow 0} G_\alpha(x)$, $x\in\mathbf{R}^d$, exists for $d \geqslant 3$, and $G_{0+}(x)$ is equal to $K_0(x) = (\Gamma(d/2-1)/4\pi^{d/2})|x|^{-d+2}$, which is the kernel for the †Newtonian potentials. Brownian motion is †recurrent when $d \leqslant 2$ and $G_{0+}(x) = +\infty$, $x\in\mathbf{R}^d$. In this case, $K_0(x)$ is defined by $K_0(x) = \lim_{\alpha\downarrow 0}(G_\alpha(x) - G_\alpha(x_0))$, and $K_0(x) = (1/2\pi)\log(1/|x|)$, when $d=2$ and $|x_0|=1$. This is the kernel for the †logarithmic potentials. When $d = 1$ and $x_0 = 0$, $K_0(x) = -|x|/2$. Using this relationship, we can express many concepts of classical potential theory in an elegant form in probability language. Let $X = \{X(t), P_x\}$ be the d-dimensional Brownian motion. Given a set A, set $\sigma_A = \inf\{t \mid t > 0, X(t)\in A\}$, where the infimum over the empty set is understood to be $+\infty$. Then σ_A is called the †hitting time of X for the set A (→ 261 Markov Processes). For a Green domain (i.e., a domain which is a †Green space) D in $\mathbf{R}^d$ $(d \geqslant 2)$, set $2g^D(t,x,y)dy = P_x(X(t)\in dy, \sigma_{\partial D} > t)$, $x, y\in D$. Then the right-hand side of this equation is the †transition probability of the Brownian motion on D with the †absorbing barrier ∂D. Then $G^D(x,y) = \int_0^\infty g^D(t,x,y)dt$ is †Green's function of D. If B is a compact subset of a Green domain D or an open subset with compact closure $\bar{B}\subset D$, then the hitting probability $p(x) = P_x(\sigma_B < \sigma_{\partial D})$ is the equilibrium potential of B relative to D.

Suppose A is an †analytic subset of $\mathbf{R}^d$. †Blumenthal's 0-1 law implies that $P_x(\sigma_A = 0) = 1$ or 0. The point x is said to be **regular** for A if this probability is 1 and **irregular** for A otherwise. Let B be a compact subset in $\mathbf{R}^d$ $(d \geqslant 2)$. Then †Wiener's test (→ 120 Dirichlet Problem) states that x is regular or irregular for B according as the following series diverge or converge:

$$\sum_{k=1}^{\infty} 2^{k(d-2)}C(B_k), \qquad d\geqslant 3,$$

$$\sum_{k=1}^{\infty} kC(B_k), \qquad d=2,$$

where $C(B_k)$ is the †Newtonian capacity (the logarithmic capacity relative to a bounded domain when $d = 2$) of the set $B_k = \{y \mid 2^{-(k+1)} \leqslant |y-x| < 2^{-k}\}\cap B$ [12]. Suppose D is a bounded domain in $\mathbf{R}^d$ $(d \geqslant 2)$. A point $x \in \partial D$ is regular or irregular for $R^d - D$ according as x is regular or irregular in the sense of the †Dirichlet problem for D. Given a continuous boundary function f on ∂D, $u(x) = E_x(f(X(\sigma_{\partial D})))$ is the solution of the generalized Dirichlet problem for D. Given $x\in D$, the distribution $h(x,B) = P_x(X(\sigma_{\partial D})\in B)$ $(B\in\mathfrak{B}(\partial D))$ used in the solution $u(x) = \int_{\partial D} f(y)h(x,dy)$ of the generalized Dirichlet problem is the †harmonic measure of ∂D as viewed from x.

E. 1-Dimensional Brownian Motion

In his monograph [6], P. Lévy gave a profound description of the fine structure of the individual 1-dimensional Brownian path

$\{X(t,\omega)\}_{t\geqslant 0}$. Let us set

σ_a = the hitting time to the point a in $\mathbf{R}'$,

$$m(t,\omega)=\min_{\sigma_0\leqslant s\leqslant t} X(s,\omega),$$

$$M(t,\omega)=\max_{\sigma_0\leqslant s\leqslant t} X(s,\omega),$$

$$Y_0(t,\omega)=|X(t,\omega)|,$$

$$Y_1(t,\omega)=\begin{cases}X(t,\omega), & t<\sigma_0(\omega),\\ M(t,\omega)-X(t,\omega), & t\geqslant\sigma_0(\omega),\end{cases}$$

and

$$Y_2(t,\omega)=\begin{cases}X(t,\omega), & t<\sigma_0(\omega),\\ X(t,\omega)-m(t,\omega), & t\geqslant\sigma_0(\omega).\end{cases}$$

Then we have (1) $P_0(M(t)>a)=2P_0(M(t)>a, X(t)>a)=2P_0(M(t)>a, X(t)<a)=P_0(|X(t)|>a)$ (**reflection principle** of D.André). (2) The stochastic process $\{\sigma_a, 0\leqslant a<\infty, P_0\}$ is a †one-sided stable process with exponent $1/2$, that is, it is additive and homogeneous with the law

$$P_0(\sigma_b-\sigma_a\leqslant t)=P_0(\sigma_{b-a}\leqslant t)$$

$$=\int_0^t \frac{b-a}{\sqrt{2\pi s^3}}e^{-(b-a)^2/2s}ds,$$

$$0\leqslant a<b,\qquad t\geqslant 0.$$

(3) Let $\varphi^{-1}(t,\omega)$ be the right continuous inverse function of

$$\varphi(t,\omega)=\int_0^t \chi_{[0,\infty)}(X(s,\omega))ds,$$

where $\chi_{[0,\infty)}(\cdot)$ is the †indicator function of the interval $[0,\infty)$. Set $Y_3(t,\omega)=X(\varphi^{-1}(t,\omega),\omega)$. Then the four processes $\{Y_i(t,\text{w}), 0\leqslant t<\infty, P_x\}$ $(x\in[0,\infty), 0\leqslant i\leqslant 3)$ on $[0,\infty)$ have the same probability law. Each of them is a diffusion process with transition probability

$$P(t,x,B)$$

$$=\int_B\frac{1}{\sqrt{2\pi t}}(e^{-|x-y|^2/2t}+e^{-|x+y|^2/2t})dy,$$

$$B\in\mathfrak{B}([0,\infty)),\qquad t>0,\qquad x\in[0,\infty),$$

and is said to be a Brownian motion on $[0,\infty)$ with a †reflecting barrier at the origin. (4) As a consequence of (3), if $X(0,\omega)=0$ a.s., then, for fxed t, M(t), -m(t), and $Y_i(t)$ $(0\leqslant i\leqslant 3)$ have a common distribution. For example,

$$P_0(M(t)\geqslant a)=P_0(\sigma_a\leqslant t)=2P_0(X(t)>a)$$

$$=\sqrt{\frac{2}{\pi t}}\int_a^\infty e^{-x^2/2t}dx\qquad a>0,$$

$$P_0(X(t)\in da, M(t)\in db)$$

$$=\sqrt{\frac{2}{\pi t^3}}(2b-a)e^{-(2b-a)^2/2t}da\,db,\qquad 0\leqslant a<b.$$

(5) The diffusion process $\{X(t,\omega), 0\leqslant t<\sigma_0, P_x\}$ $(x\in(0,\infty))$ obtained from a 1-dimensional Brownian motion by shortening its †lifetime is called a Brownian motion on $(0,\infty)$ with an †absorbing barrier at the origin, and its transition probability is given by

$$P(t,x,B)$$

$$=\int_B\frac{1}{\sqrt{2\pi t}}(e^{-|x-y|^2/2t}-e^{-|x+y|^2/2t})dy,$$

$$B\in\mathfrak{B}([0,\infty)),\qquad t>0,\qquad x\in(0,\infty).$$

The **arcsin law** is valid for many functionals of 1-dimensional Brownian motion. For example,

$$P_0\left(\int_0^t\chi_{[0,\infty)}(X(s,\omega))ds\leqslant\theta\right)=\frac{2}{\pi}\arcsin\sqrt{\frac{\theta}{t}},$$

$$0<\theta\leqslant t,$$

$$P_0(\tau_t(\omega)\leqslant s)=(2/\pi)\arcsin\sqrt{s/t},\qquad 0<s\leqslant t,$$

where $\tau_t(\omega)=\sup\{s\,|\,X(s,\omega)=0, 0\leqslant s\leqslant t\}$.

The visiting set $\mathscr{Z}(\omega)=\{t\,|\,X(t,\omega)=0\}$ of a Brownian path is a †totally disconnected set. Its †Lebesgue measure is 0, and the †Hausdorff-Besikovich dimension number of $\mathscr{Z}(\omega)$ is $1/2$.

Consider a 1-dimensional Wiener process $\{X(t)\}_{0\leqslant t<\infty}$ starting at the origin. Let us set

$$B_1(t)=\begin{cases}0 & t=0,\\ tX(1/t) & t>0,\end{cases}$$

and

$$B_2(t)=cX(t/c^2),\qquad t\geqslant 0,\quad (c>0).$$

Then the stochastic processes $\{B_1(t)\}_{0\leqslant t<\infty}$ and $\{B_2(t)\}_{0\leqslant t<\infty}$ are likewise 1-dimensional Wiener processes starting at the origin. Hence the properties of the Wiener process starting at the origin in a neighborhood of $t=0$ $(t=\infty)$ can be obtained from those in a neighborhood of $t=\infty$ $(t=0)$.

F. d-Dimensional Brownian Motion

Almost all paths of d-dimensional Brownian motion are continuous but are not of †bounded variation on any finite interval. Accordingly, they cannot have lengths. A positive, continuous, increasing function φ defined on $[t_0,\infty)$ with $t_0>0$ is said to **belong to the upper (lower) class with respect to local continuity** if $P_0((\inf\{t\,|\,|X(t,\omega)|>\sqrt{t}\,\varphi(1/t), t>0\})>0)=1$ (0). **Kolmogorov's test** states that φ belongs to the upper class or to the lower class with respect to local continuity according as

$$\int_{t_0}^\infty t^{-1}(\varphi(t))^d e^{-\varphi^2(t)/2}dt$$

converges or diverges. For example,

$$\varphi(t)=(2\log_{(2)} t+(d+2)\log_{(3)} t+2\log_{(4)} t+\ldots+2\log_{(n-1)} t+(2+\delta)\log_{(n)} t)^{1/2}$$

belongs to the upper or lower class with respect to local continuity according as $\delta>0$ or $\delta\leqslant 0$, where $\log_{(n)} t=\log\log_{(n-1)} t$ and $\log_{(1)} t=\log t$. The **law of the iterated logarithm** for 1-dimensional Brownian motion,

$$P_0\left(\limsup_{t\to s}\frac{|X(t,\omega)-X(s,\omega)|}{\sqrt{2|t-s|\log\log 1/|t-s|}}=1\right)=1,$$

is a special case of this example. For the case $d=1$, consider the **space-time Brownian motion** $\{(-t, X(t)), P_0\}$. Set $D_\varphi=\{(s,x)\mid -1/t_0\leqslant s\leqslant 0, \sqrt{-s}\,\varphi(-1/s)\leqslant x<\infty\}$. Then O=(O,O) is a regular point of D_φ or an irregular point of D_φ for the space-time Brownian motion according as φ belongs to the lower or upper class with respect to local continuity. (→ 261 Markov Processes). Thus Kolmogorov's test is the Wiener test for space-time Brownian motion. Let ψ be a positive, continuous decreasing function defined on $[t_0,\infty)$ with $t_0>0$. Then

$$P_0\left(\left(\inf\{t\mid |X(t,\omega)|<\sqrt{t}\,\psi(1/t), t>0\}\right)>0\right)$$

=0 or 1.

We have

$$\int_{t_0}^{\infty}\frac{1}{t}(\psi(t))^{d-2}\,dt=\infty \text{ or } <\infty \qquad \text{when } d\geqslant 3,$$

and

$$\int_{t_0}^{\infty}\frac{dt}{t|\log\psi(t)|}=\infty \text{ or } <\infty \qquad \text{when } d=2.$$

To describe uniform continuity of a path on the interval [0, 1], take a positive, continuous increasing function ψ defined on $[t_0,\infty)$ with $t_0>0$ and set $\varphi(t)=\sqrt{t}\,\psi(1/t)$. Then ψ is said to **belong to the upper class with respect to uniform continuity** if almost all paths $X(t,\omega)$ $(0\leqslant t\leqslant 1)$ satisfy the †Lipschitz condition relative to φ, that is, for almost all ω there exists an $\varepsilon(\omega)>0$ such that $0<|t-s|<\varepsilon(\omega)$ implies $|X(t,\omega)-X(s,\omega)|<\varphi(|t-s|)$. And ψ is said to **belong to the lower class with respect to uniform continuity** if almost all paths $X(t,\omega)$ $(0\leqslant t\leqslant 1)$ do not satisfy the Lipschitz condition relative to φ. Then ψ belongs to the upper or lower class with respect to uniform continuity according as

$$\int_{t_0}^{\infty}\psi(t)^{d+2}e^{-\psi^2(t)/2}\,dt$$

converges or diverges [16]. For example,

$$\psi(t)=(2\log t+(d+4)\log_{(2)} t+2\log_{(3)} t+\ldots+2\log_{(n-1)} t+(2+\delta)\log_{(n)} t)^{1/2}$$

belongs to the upper or lower class with respect to uniform continuity according as $\delta>0$ or $\delta\leqslant 0$. The following theorem on the uniform continuity of 1-dimensional Brownian motion is a special case of this criterion:

$$P_0\left(\limsup_{\substack{|t-s|\to 0\\ 0\leqslant t,s\leqslant 1}}\frac{|X(t,\omega)-X(s,\omega)|}{\sqrt{2|t-s|\log 1/|t-s|}}=1\right)=1.$$

Now we state some other properties of Brownian paths. Let A be a set of zero †outer capacity in $\mathbf{R}^d$ $(d\geqslant 2)$. Then $P_x(X(t)\in A$ for some $t>0)=0$ for any $x\in\mathbf{R}^d$. Let A be a plane set with positive †inner capacity. Then $P_x(X(t)\in A$ for infinitely many t larger than any given $s>0)=1$ for any $x\in\mathbf{R}^2$. With probability 1, the Lebesgue measure of the set $\{X(t,\omega)\mid 0\leqslant t<\infty\}$ in $\mathbf{R}^d$ is zero for $d\geqslant 2$. With probability 1 this set is everywhere dense in $\mathbf{R}^2$ when $d=2$ and nowhere dense in $\mathbf{R}^d$ when $d\geqslant 3$ [13, 15, 17]. Almost all 2-dimensional Brownian paths have k-fold multiple points for any integer $k\geqslant 2$. In the 3-dimensional case, almost all paths have †double points but cannot have any triple point. In the $d(\geqslant 4)$-dimensional case, almost all paths have no double point [18–20].

G. Itô's Formula and Brownian Local Time

The following formula of Itô [21] is of fundamental importance in the theory of stochastic processes.

Itô's formula. Let $X=\{X(t)=(X_1(t),X_2(t),\ldots,X_d(t)), P_x\}$ be a d-dimensional Brownian motion and $f(x)=f(x_1,x_2,\ldots,x_d)$ be a C^2-function defined on $\mathbf{R}^d$. Let us set $\varphi(t)=f(X(t))-f(X(0))$. Then $\{\varphi(t)\}_{t\geqslant 0}$ is a continuous †additive functional of the Brownian motion X (→ 261 Markov Processes), and we have

$$f(X(t))-f(X(0))=\sum_{i=1}^{d}\int_0^t\left(\frac{\partial}{\partial x_i}f\right)(X(s))\,dX_i(s)+\frac{1}{2}\int_0^t(\Delta f)(X(s))\,ds,\qquad t\geqslant 0,$$

where the first term in the right-hand side is a †stochastic integral with respect to the Brownian motion X [22] (→ 406 Stochastic Differential Equations). In particular, if f is a C^3-function defined on $\mathbf{R}^d$, Itô's formula can be rewritten as

$$f(X(t))-f(X(0))=\sum_{i=1}^{d}\int_0^t\left(\frac{\partial}{\partial x_i}f\right)(X(s))\circ dX_i(s),\qquad t\geqslant 0,$$

where the right-hand side is a stochastic integral of the †Stratonovich type with respect to

the Brownian motion X [9] (- 406 Stochastic Differential Equations).

Consider a 1-dimensional Brownian motion $X = (X(t), P_x)$. By the **local time** or the **sojourn time density** of X we mean a family of nonnegative random variables $\{\varphi(t,x)\}$ ($t \in [0, \infty)$, $x \in \mathbf{R}^1$) such that, with probability 1, the following holds:
(a) the mapping $[0, \infty) \times \mathbf{R}^1 \ni (t,x) \mapsto \varphi(t,x) \in \mathbf{R}^1$ is continuous,
(b) for every Bore1 subset A of $\mathbf{R}^1$ and $t \geqslant 0$,

$$\int_0^t \chi_A(X(s))\,ds = 2\int_A \varphi(t,x)\,dx.$$

The notion of the local time of Brownian motion was first introduced by Lévy [6]. The family of random variables $\{\varphi(t, x)\}$ ($t \geqslant 0$, $x \in \mathbf{R}^1$) defmed by

$$\varphi(t,x) = (X(t)-x)^+ - (X(0)-x)^+ - \int_0^t \chi_{(x,\infty)}(X(s))\,dX(s)$$

satistïes properties (a) and (b) mentioned above and hence it is a local time [9, 22]. Here a+ is the bigger of a and 0. It is clear that the local time $\varphi(t, x)$ of X is given by

$$\varphi(t,x) = \lim_{\varepsilon\downarrow 0}\frac{1}{4\varepsilon}\int_0^t \chi_{(x-\varepsilon,x+\varepsilon)}(X(s))\,ds \qquad \text{for every } x\in\mathbf{R}^1, t\geqslant 0.$$

Furthermore, we have

$$P_0\left(\limsup_{b-a=\delta\downarrow 0,\, a<b}\frac{|\varphi(t,b)-\varphi(t,a)|}{\sqrt{\delta\log 1/\delta}} \leqslant 2\sqrt{\max_{a\in\mathbf{R}^1}\varphi(t,a)}\right)=1,$$

and

$$P_0\left(\limsup_{\delta\downarrow 0}\frac{|\varphi(t,\delta)-\varphi(t,0)|}{\sqrt{\delta\log\log 1/\delta}} \leqslant 2\sqrt{\varphi(t,0)}\right)=1.$$

Lévy [6] studied the fine structure of the local time. Define the visiting sets $\mathscr{Z}^+ = \{t \mid Y_0(t,\omega) = 0\}$ and $\mathscr{Z}^- = \{t \mid Y_1(t,\omega)=0\}$, where $Y_0(t,\omega)$ and Y, (t, ω) are the stochastic processes defined in Section E above. Then we have

$$P_0\left(\lim_{\varepsilon\downarrow 0}\sqrt{\frac{\pi\varepsilon}{2}} \times [\text{the number of flat stretches of } M(s,\omega)\ (0\leqslant s\leqslant t) \text{ of length} \geqslant\varepsilon] = M(t,\omega), t\geqslant 0\right)=1.$$

Since the two diffusion processes $X^+ = \{Y_0(t), 0\leqslant t<\infty, P_x\}$ and $X^- = \{Y_1(t), 0\leqslant t<\infty, P_x\}$ define the same probability law on $W([0,\infty))$, there exists a functional $\varphi^+(t,\omega)$ of X^+ corresponding to $M(t,\omega)$ of X^-. Here, $W([0,\infty))$ denotes the space of all continuous functions w: $[0,\infty)\to[0,\infty)$. The flat stretches of the graph of $\varphi^+(t,\omega)$ are the open intervals $\mathscr{Z}_n$ $(n\geqslant 1)$ such that $\sum_{n\geqslant 1}\mathscr{Z}_n = [0,\infty) - \mathscr{Z}^+$. Then we have

$$P_0\left(\lim_{\varepsilon\downarrow 0}\sqrt{\frac{\pi\varepsilon}{2}} \times [\text{the number of intervals } \mathscr{Z}_n\subset[0,t) \text{ of length} \geqslant\varepsilon] = \varphi^+(t,\omega), t\geqslant 0\right)=1.$$

Furthermore, we have

$$P_0\left(\lim_{\varepsilon\downarrow 0}\sqrt{\frac{\pi}{2\varepsilon}} \times [\text{the total length of the intervals } \mathscr{Z}_n\subset[0,t) \text{ of length} <\varepsilon] = \varphi^+(t,\omega), t\geqslant 0\right)=1$$

and

$$P_0\left(\lim_{\varepsilon\downarrow 0}(2\varepsilon)^{-1}\int_0^t \chi_{[0,\varepsilon)}(Y_0(s,\omega))\,ds = \varphi^+(t,\omega), t\geqslant 0\right)=1.$$

Let $d_\varepsilon(t,\omega)$ be the number of times that the reflecting Brownian path $Y_0(s,\omega)$ crosses down from $\varepsilon>0$ to 0 before time t. Then

$$P_0\left(\lim_{\varepsilon\downarrow 0}\varepsilon d_\varepsilon(t,\omega) = \varphi^+(t,\omega), t\geqslant 0\right)=1.$$

H. Flows and Random Distributions

The flow derived from the 1-dimensional Wiener process $\{X_t\}_{-\infty<t<\infty}$ is †Kolmogorov's flow. It has †mixing properties of all orders and is †ergodic (- 136 Ergodic Theory, 39.5 Stationary Processes). The †stationary process with independent values at every point corresponding to the †characteristic functional

$$\exp\left(-\frac{1}{2}\int_{-\infty}^{\infty}\varphi(t)^2\,dt\right)$$

on the †Schwartz space Y defines the same probability law with the stationary process obtained by differentiation of the Wiener process in the †distribution sense [23] (- 395 Stationary Processes, 407 Stochastic Processes).

I. Generalizations of Brownian Motion

In addition to the Brownian motion described above, there are several stochastic processes that are also called Brownian motion. A Gaussian system $\{X(a)\}_{a\in\mathbf{R}^N}$ defined on a probability space $(\Omega, \mathfrak{B}, P)$ is said to be a **Brownian motion witb an N-dimensional time parameter**

if(i) $E(X(a))=0$, (ii) $E(X(a)X(b))=\frac{1}{2}(|a|+|b|-|a-b|)$, (iii) $P(X(0)=0)=1$. Let a^* be the †spherical inversion of $a \in \mathbf{R}^N$ with respect to the unit sphere. Set $X^*(a)=|a|X(a^*)$ $(a \neq 0)$ and $X^*(0) = 0$. Then $\{X^*(a)\}_{a \in \mathbf{R}^N}$ defines the same probability law with a Brownian motion with an N-dimensional time parameter. Almost all paths of a Brownian motion with an N-dimensional time parameter are continuous. A positive, continuous, increasing function φ defined on $[t_0, \infty)$ with $t_0>0$ is said to belong to the upper (lower) class with respect to local continuity if the probability that the closure of the set $\{a \mid |X(a, \omega)|>\sqrt{|a|}\varphi(1/|a|), |a|>0\}$ contains the origin 0 is equal to 0 (1). Then φ belongs to the upper or lower class with respect to local continuity according as the integral

$$\int_{t_0}^{\infty} \frac{1}{t}(\varphi(t))^{2N-1} e^{-\varphi^2(t)/2}\, dt$$

converges or diverges. For example,

$$\varphi(t)=(2\log_{(2)} t+(2N+1)\log_{(3)} t+2\log_{(4)} t+ \ldots +2\log_{(n-1)} t+(2+\delta)\log_{(n)} t)^{1/2}$$

belongs to the upper or lower class with respect to local continuity according as $\delta > 0$ or $\delta \leqslant 0$. As a special case, we have

$$P\left(\limsup_{a \to 0} \frac{|X(a, \omega)|}{\sqrt{2|a|\log\log 1/|a|}}=1\right)=1.$$

Take a positive, continuous increasing function φ defined on $[t_0, \infty)$ with $t_0 > 0$, and set $\psi(t)=\sqrt{t}\,\varphi(1/t)$. If almost all paths $X(a, \omega)$ $(|a| \leqslant 1)$ satisfy the Lipschitz condition relative to $\psi(t)$, then φ is said to belong to the upper class with respect to uniform continuity. It is said to belong to the lower class with respect to uniform continuity if, with probability 1, these paths do not satisfy the Lipschitz condition relative to $\psi(t)$. Then φ belongs to the upper or lower class with respect to uniform continuity according as

$$\int_{t_0}^{\infty} t^{N-1}(\varphi(t))^{4N-1} e^{-\varphi^2(t)/2}\, dt$$

converges or diverges (T. Sirao, *Nagoya Math. J.*, 17 (1960)). For example,

$$\varphi(t) = (2N\log t+(4N+1)\log_{(2)} t+2\log_{(3)} t + \ldots +2\log_{(n-1)} t+(2+\delta)\log_{(n)} t)^{1/2}$$

belongs to the upper or lower class with respect to uniform continuity according as $\delta > 0$ or $\delta \leqslant 0$. As a special case,

$$P\left(\limsup_{|a-b| \to 0, |a|, |b| \leqslant 1} \frac{|X(a, \omega)-X(b, \omega)|}{\sqrt{2N|a-b|\log 1/|a-b|}}=1\right)=1$$

(For general information about Brownian motion with a multidimensional time parameter → P. Lévy [6] and H. P. McKean, *Theory of Prob. Appl.* 8 (1963).)

Let $X(t)$ be a Wiener process. L. S. Ornstein and G. E. Uhlenbeck based their investigation of the irregular movements of small particles immersed in a liquid on **Langevin's equation**

$$dU(t)=-\alpha U(t)\,dt+\beta\, dX(t),$$

where $U(t)$ is the velocity of a particle. The first term on the right-hand side of this †stochastic differential equation (→ 406 Stochastic Differential Equations) is due to frictional resistance or its analog, which is thought to be proportional to the velocity. The second term represents random external force. The stationary solution of this equation is given by

$$U(t) = \int_{-\infty}^{t} \beta e^{-\alpha(t-u)}\, dX(u).$$

The stochastic process $\{U(t)\}_{-\infty<t<\infty}$ is a stationary †Gaussian Markov process with covariance function $\gamma(t) = (\beta^2/2\alpha)e^{-\alpha|t|}$. This process is called **Ornstein-Uhlenbeck Brownian motion** [24] (→ 176 Gaussian Processes).

Brownian motion has been defined on state spaces that are †Riemannian spaces or †Lie groups, and its properties are being investigated (→ 5 Additive Processes, 115 Diffusion Processes).

References

[1] R. Brown, A brief account of microscopical observations made in the months of June, July, and August, 1827, on the particles contained in the pollen of plants; On the general existence of active molecules in organic and inorganic bodies, Philos. Mag., Ann. Philos. New Series, 4 (1828).

[2] A. Einstein, Investigations on the theory of the Brownian movement, Methuen, 1926.

[3] R. E. A. C. Paley and N. Wiener, Fourier transforms in the complex domain, Amer. Math. Soc. Colloq. Publ., 1934.

[4] N. Wiener, Differential space, J. Math. and Phys., 2 (1923), 131–174.

[5] P. Lévy, Le mouvement brownien plan, Amer. J. Math., 62 (1940), 487–550.

[6] P. Lévy, Processus stochastiques et mouvement brownien, Gauthier-Villars, 1948.

[7] J. L. Doob, Stochastic processes, Wiley, 1953.

[8] K. Itô, Lectures on stochastic processes, Tata Inst, 1960.

[9] N. Ikeda and S. Watanabe, Stochastic differential equations and diffusion processes, North-Holland and Kodansha, 198 1.

[10] K. Itô and M. Nisio, On the convergence of sums of independent Banach space-valued random variables, Osaka J. Math., 5 (1968), 35-48.
[11] P. Billingsley, Convergence of probability measure, Wiley, 1963.
[12] K. Itô and H. P. McKean, Jr., Diffusion processes and their sample paths, Springer, 196.5.
[13] J. L. Doob, Semi-martingales and subharmonic functions, Trans. Amer. Math. Soc., 77 (1954), 86-121.
[14] A. Ya. Khinchin (Khintchin) Asymptotische Gesetze der Wahrscheinlichkeitsrechnung, Erg. Math., Springer, 1933 (Chelsea, 1948).
[15] S. Kakutani, Two-dimensional Brownian motion and harmonic functions, Proc. Imp. Acad. Tokyo, 20 (1944), 7066714.
[16] K. L. Chung, P. Erdos, and T. Sirao, On the Lipschitz's condition for Brownian motion, J. Math. Soc. Japan, 11 (1959), 263-274.
[17] S. Kakutani, On Brownian motions in n-space, Proc. Imp. Acad. Tokyo, 20 (1944), 648-652.
[18] A. Dvoretsky, P. Erdos, and S. Kakutani, Double points of Brownian motion in n-space, Acta Sci. Math. Szeged., 12 (1950), 75-81.
[19] A. Dvoretsky, P. Erdos, and S. Kakutani, Multiple points of paths of Brownian motion in the plane, Bull. Res. Council Israel, 3 (1954), 3644371.
[20] A. Dvoretsky, P. Erdos, S. Kakutani, and S. J. Taylor, Triple points of Brownian paths in 3-space, Proc. Cambridge Philos. Soc., 53 (1957), 8566862.
[21] K. Itô, On a formula concerning stochastic differentials, Nagoya Math. J., 3 (1951), 55–65.
[22] H. P. McKean, Stochastic integrals, Academic Press, 1972.
[23] T. Hida, Brownian motion, Springer, 1980.
[24] G. E. Uhlenbeck and L. S. Ornstein, On the theory of Brownian motion, Phys. Rev., 36 (1930), 823-841.

C

46 (X.32) Calculus of Variations

A. General Remarks

One of the first objects of differential calculus was to systematize the theory of the extrema of functions of a finite number of independent variables. In the calculus of variations we consider **functionals** (i.e., real-valued or complex-valued functions defined on a function space {u} consisting of functions defined on a certain domain B), originally so named by J. Hadamard. Such a functional is denoted by $J[u\ B]$ (or simply $J[u]$). A function u, which is considered an independent variable of the functional, is called an **argument function** (or **admissible function).**

Concrete examples of functionals are the length of a curve y $=f(x)$ and the area of a surface z = $z(x$, y), which are expressed by

$$L[y]=\int_{x_0}^{x_1}\sqrt{1+(y'(x))^2}\,dx$$

and

$$S[z]=\iint_B\sqrt{1+z_x^2+z_y^2}\,dx\,dy,$$

respectively. Furthermore, consider a curve y = y(x) connecting two given points (x,, y_0) and (x_1, y,) with y, >y,. The time in which a particle slides down without friction from (x,, y,,) to (x_1, y_1) along this curve under constant gravity acting in the direction of the positive y-axis is expressed by the functional

$$J[y]=k\int_{x_0}^{x_1}\sqrt{(1+y'^2)/(y-y_0)}\,dx,$$

k constant. (1)

Let $F(\ldots)$ be a known real-valued function that depends on a certain number of independent variables and on argument functions of these variables, together with derivatives of these functions up to a certain order. Then a typical problem in the calculus of variations is formulated as an extremal problem of a functional that is expressed by the integral with $F(\ldots)$ as the integrand, for example,

$$J[u]=\int_{x_0}^{x_1}F(x, u(x), u'(x), \ldots, u^{(m)}(x))\,dx,$$

$$J[u, v]= \iint_B F(x, y, u(x, y), u_x, u_y, \ldots, v(x, y), v_x, v_y, \ldots)\,dx\,dy.$$

For instance, a curve minimizing (1) is the **curve of steepest descent.**

In extremal problems, suitable boundary conditions can be assigned to argument functions. On the other hand, there are so-called **conditional problems in the calculus of variations.** A typical example of this sort is the tisoperimetric problem, i.e., the determination of the curve that bounds a domain with maximal area among all curves on a plane with given length. In general, an extremal problem of a functional under a subsidiary condition that the value of another given functional remain fixed is called a **generalized isoperimetric problem.** In addition to these, there are **Lagrange's problem,** in which a finiteness condition is imposed, and **Hilbert's problem,** in which a condition consisting of differential equations is imposed.

The birth of the calculus of variations was almost simultaneous with that of differential and integral calculus. Johann Bernoulli, Jakob Bernoulli, L. Euler, and others had dealt with several concrete problems of the calculus of variations when in 1760, J. L. Lagrange introduced a general method of dealing with variational problems connected with mechanics. Then an equation bearing the name of Euler or Lagrange was introduced.

B. Euler's Equation

As an example, consider the simplest variational problem

$$J[y]=\int_{x_0}^{x_1}F(x, y(x), y'(x))dx = \min. \tag{2}$$

Let the boundary condition $y(x_0)$ = y,, $y(x_1)$ =y, be assigned to the argument function y(x). Consider a family of admissible functions $Y(x;\varepsilon)=y(x)+\varepsilon\eta(x)$, where $\eta(x)$ is any fixed function vanishing at both endpoints and ε is a parameter. If y(x) gives the minimum of $J[y]$, then the function of ε, $J[Y]$, must attain a minimum for ε = 0. The condition $(\partial J[Y]/\partial\varepsilon)_{\varepsilon=0}$ = 0 is written in the form

$$0=\int_{x_0}^{x_1}(F_y\eta+F_{y'}\eta')\,dx$$

$$=\int_{x_0}^{x_1}\eta\left(F_y-\frac{d}{dx}F_{y'}\right)dx$$

by taking into account the boundary condition. By making use of the arbitrariness of $\eta(x)$, we conclude that

$$0=F_y-\frac{d}{dx}F_{y'}=F_y-F_{y'x}-y'F_{y'y}-y''F_{y'y'} \tag{3}$$

holds, in view of the following lemma: Let $\varphi(x)$ be continuous in $[x_0, x_1]$ and $\eta(x)$ be a function of class C^p that satisfies $\eta(x_0)=\eta'(x_0)=\ldots=\eta^{(q)}(x_0)=0$, $\eta(x_1)=\eta'(x_1)=\ldots=\eta^{(q)}(x_1)=0$ $(0\leqslant q\leqslant p)$. If $\int_{x_0}^{x_1}\eta(x)\varphi(x)dx=0$ holds for any such $\eta(x)$, then $\varphi(x)=0$. (Here class C^p can be

replaced by class C^ω when q is supposed finite. If $\eta \in C^\omega$, then $q = \infty$ is admitted.) This is called the **fundamental lemma in the calculus of variations.** We call (3) the **Euler-Lagrange differential equation** (or **Euler's equation** for the extremal problem). Since this equation is of the second order, y(x) can be determined by means of the boundary condition.

The quantity $[F]_y = F_y - \frac{d}{dx} F_{y'}$ contained in equation (3) is called the **variational derivative** of F with respect to y. Furthermore, $\delta y = \eta d\varepsilon$ and $\delta J = (\partial J$ [Y]/&),,, $d\varepsilon$ are called the **first variations** of the argument function y and of the functional $J[y]$, respectively. If $\eta(x)$ is not subject to the condition that it must vanish at the endpoints, then the first variation of J becomes

$$\delta J = \int_{x_0}^{x_1} [F]_y \delta y \, dx + [F_{y'} \delta y]_{x_0}^{x_1}.$$

In comparison with an ordinary extremal problem $f(x_1, , x_n)$ = min in differential calculus, $[F]_y$ and δJ correspond to gradfand df, respectively. In general, a solution of the Euler-Lagrange differential equation $[F]_y$ =0 is called a **stationary function** for the variational problem, and its graph is called a **stationary curve**.

For a variational problem involving several argument functions, we have only to write the system of Euler-Lagrange differential equations corresponding to them. For a problem

$$J[y] = \int_{x_0}^{x_1} F(x, y, y', \ldots, y^{(m)}) \, dx = \min$$

whose integrand involves derivatives of higher orders of an argument function, the Euler-Lagrange differential equation is

$$[F]_y = \sum_{\mu=0}^{m} (-1)^\mu \frac{d^\mu}{dx^\mu} F_{y^{(\mu)}} = 0.$$

For a problem involving a double integral

$$J[u] = \iint_B F(x, y, u, u_x, y_y) \, dx \, dy = \min,$$

the equation is

$$0 = [F]_u = F_u - \frac{\partial}{\partial x} F_{u_x} - \frac{\partial}{\partial y} F_{u_y}.$$

For a generalized isoperimetric problem, for example $J[y]$ = min, K [y] = c, the Euler-Lagrange equation becomes

$$[F + \lambda G]_y = 0$$

with the so-called **Lagrange multiplier** λ, where F and G denote the integrands of J and K, respectively. Two integration constants contained in a general solution of this differential equation of the second order and an undetermined constant λ can be determined, for instance, by the boundary condition $y(x_0) = y_0$, $y(x_1)$ = y, and a subsidiary condition $K[y] = c$. As an example, for the classical proper isoperimetric problem F = y, G = $\sqrt{1+y'^2}$, the equation is $1 - \lambda(y'/\sqrt{1+y'^2})' = 0$, which after integration leads to $(x - \alpha)^2 + (y - \beta)^2 = \lambda^2$ (→ 228 Isoperimetric Problems).

Besides the case of fixed endpoints, there is a boundary condition, for instance, that an endpoint (x_1, y,) of the argument function y = y(x) must lie on a given curve $T(x, y) = 0$. For the case of such a movable endpoint, the extremal function is subject to the **condition of transversality,**

$$(F - y'F_{y'})T_y - F_{y'} T_x = 0, \qquad x = x_1.$$

C. Sufficient Conditions

A. M. Legendre introduced the notion of the second variation, corresponding to differential quotients of the second order in differential calculus, in order to discuss sufficient conditions. Concerning the simplest problem (2), the inequality $F_{y'y'}(x, y_0(x), y_0'(x)) \geqslant 0$ is necessary in order for $y_0(x)$ to give the minimum. Conversely, the inequality $F_{y'y'} > 0$ and Jacobi's condition (which is stated below) imply that y = $y_0(x)$ gives a **weak minimum.** Here "weak minimum" means the minimum when a family of admissible functions $\{ \, |y - y_0| < \varepsilon, \ |y' - y_0'| < \varepsilon\}$ is considered a neighborhood of y,.

Jacobi's condition: Let u be a solution of a linear ordinary differential equation of the second order,

$$\frac{d}{dx}\left(F_{y'y'} \frac{du}{dx}\right) - \left(F_{yy} - \frac{d}{dx} F_{yy'}\right) u = 0,$$

$$u(x_0) = 0;$$

then the smallest zero of u that is greater than x_0 (i.e., the **conjugate point** of x_0) is greater than the right endpoint x_i.

K. Weierstrass derived sufficient conditions for a strong minimum by extending the range of admissible functions to $\{ \, |y - y_0| < \varepsilon\}$. Results that were obtained until about that time constitute the content of what is usually called the **classical theory of the calculus of variations.**

If for the variations problem (2) there exists a unique curve through every point in a domain on the xy-plane that belongs to a one-parameter family of stationary curves of the functional

$$J[y] = \int_{x_0}^{x_1} F(x, y, y') \, dx,$$

then the domain is called a **field** of stationary curves. Let the parameter value of the curve through a point (x, y) in such a family of

stationary curves $y = \varphi(x; a)$ be denoted by $\alpha = \alpha(x, y)$. The slope $p(x, y) = [\varphi'(x; \alpha)]_{\alpha=\alpha(x,y)}$ is called the slope of the field at the point (x, y) or, by regarding x, y as variables, the **slope function** of the field. The value of a curvilinear integral

$$I_C = I[y] = \int_C (F(x,y,p) - (y'-p)F_{y'}(x,y,p))\,dx$$

is then determined and depends only on the two endpoints of the curve C. We call I_C **Hilbert's invariant integral.** In view of the property mentioned above, we can denote the value of the functional J for a function $y = y(x)$ representing a curve C by J_C. Then any admissible curve C that passes through a field embedding a stationary curve C_0 satisfies

$$0 \leqslant \Delta J = J_C - J_{C_0} = \int_C \mathscr{E}(x, y; p, y')\,dx.$$

Here

$$\mathscr{E}(x,y;p,y') = F(x,y,y') - F(x,y,p) - (y'-p)F_{y'}(x,y,p)$$

is the **$\mathscr{E}$-function** introduced by Weierstrass. For C_0 to give the minimum of $J[y]$, it suffices that $\mathscr{E} \geqslant 0$ hold for every point (x, y) in. the field and every value y' (→ 279 Morse Theory).

D. Optimal Control

Let a system of differential equations

$$dx_i/dt = f_i(x_1, \ldots, x_n; u_1, \ldots, u_k),$$

$$(u_1, \ldots, u_k) \in \Omega; \quad x_i(t_0) = x_i, \quad i = 1, \ldots, n, \qquad (4)$$

be given, where $u_1, \ldots, u_k$ are parameters. In general, a problem of **optimal control** is to determine $u_j = u_j(t)$ $(t_0 \leqslant t \leqslant t_1)$ such that the value of a functional

$$J[u] = \int_{t_0}^{t_1} F(x_1, \ldots, x_n; u_1, \ldots, u_k)\,dt$$

assumes a minimum, where $x_i(t)$ are the solutions of (4) and are considered functions in $u_1, \ldots, u_k$, and t. Such a problem is a kind of conditional variational problem. But since the existence region of u is restricted, certain conditions in the form of inequalities are imposed, and furthermore u is not necessarily continuous, and in many cases the problem cannot be treated within the classical theory of the calculus of variations (→ 86 Control Theory).

E. The Direct Method in the Calculus of Variations

In mathematical physics †variational principles are derived from discussions of formal correspondence between a functional $J[u]$ to be minimized and Euler's equation for $J[u]$. This is certainly one of the important methods in the calculus of variations, but it is also possible to investigate a stationary function u_0 on the basis of its stationary character and independently of Euler's equation. This *is* called the **direct method** in the calculus of variations. It plays an important role in the theoretical treatment of the existence and uniqueness of solutions, and it is also significant as a technique for approximate or numerical solutions. When a differential equation is given independently of the calculus of variations, it is possible to apply the direct method if a functional whose Euler's equation is the given differential equation can be constructed.

Let D be a bounded domain in m-dimensional space and $f \in L_2(D)$ be a real-valued function. Consider the variational problem of minimizing the functional

$$J[u] = \int_D |\operatorname{grad} u|^2\,dx - 2\int_D fu\,dx.$$

Here we suppose the set of admissible functions, denoted by A_J, to be the Hilbert space obtained by completing the function space $C_0^\infty(D)$ with respect to the norm

$$N(u) = \left(\int_D |\operatorname{grad} u|^2\,dx + \int_D u^2\,dx\right)^{1/2}$$

Utilizing F. Riesz's representation theorem in Hilbert spaces, it can be shown that there exists a minimum value l in A, which is uniquely realized by certain $u_0 \in \bar{A}_J$. Since the function u_0 belongs to A,, it can be shown that the boundary condition

$$u|_{\partial D} = 0 \qquad (5)$$

is satisfied in a generalized sense. Furthermore, in view of $J[u_0] \leqslant J[u_0 + \varphi]$ being valid for any $\varphi \in C_0^\infty(D)$, it can be verified that the equation

$$-\Delta u = f \qquad (6)$$

is satisfied in D in the sense of differentiation of †distributions. In other words, the stationary function u_0 is a solution in the wide sense (a tweak solution) of the classical boundary value problem for †Poisson's equation formulated by (6) and (5). If a function space A, with $\bar{A}_J \supset A_J \supset C_0^\infty(D)$ is taken as the set of admissible functions, the value $l = J[u_0]$ becomes the greatest lower bound of J in A_J. In this case, if $\{u_n\}_{n=1}^\infty$ is any **minimizing sequence** from A_J, that is, if $u_n \in A$,, $n = 1, 2, \ldots$; $J[u_n] \to l$ $(n \to \infty)$, then it converges to u_0 in the sense that

$$N(u_n - u_0) \to 0 \quad (n \to \infty). \qquad (7)$$

In other words, the solution in the wide sense

u_0 of the boundary value problem can be constructed as the limit of a minimizing sequence that consists of sufficiently smooth functions vanishing on the boundary. In the proof of the fact that a solution in the wide sense u_0 coincides with the classical solution of the boundary value problem under an assumption of suitable smoothness for f and ∂D, a standard argument has been established for proving the regularity of a solution in the wide sense.

The technique of obtaining the solution of the boundary value problem as the limit of a minimizing sequence was proposed by B. Riemann concerning the classical †Dirichlet problem and was completed by D. Hilbert. This pioneering work led to the recent treatment of boundary value problems by utilizing Hilbert spaces. For †Self-adjoint boundary value problems, the method stated for the above example has been generalized almost directly to the cases of differential operators of higher order and with variable coefficients (→ 323 Partial Differential Equations of Elliptic Type). By making use of some auxiliary arguments, this technique can be applied extensively to the construction of several kinds of mapping functions in the theory of functions of a complex variable, to the solution of †integral equations of the second kind, and also in other fields [3, 4].

The eigenvalue problem, which is formulated by

$$Hu = \lambda u, \qquad u \neq 0, \tag{8}$$

with a †Self-adjoint operator H in a Hilbert space, can also be transformed into a variational problem for the †Rayleigh quotient

$$R[u] = (Hu, u)/\|u\|^2 \tag{9}$$

(→ 298 Numerical Computation of Eigenvalues).

F. Solution of Differential Equations by the Direct Method

In view of the convergence shown in (7), a minimizing sequence can be regarded as an approximating sequence for a solution of the boundary value problem or a stationary function u_0. Let a function

$$u_n = u_n(x; c) = u_n(x; c_1, \ldots, c_n) \tag{10}$$

involving an n-vector $c = (c_1, \ldots, c_n)$ as a parameter be admissible for any c. If $J[u_n(\cdot; c)] = F(c)$, obtained by substituting u_n into J, is minimized at $c = c^0$, then $u_n(\cdot; c^0)$ is considered as a function that approximates u_0 most precisely within the family $u_n(\cdot; c)$. This vector c^0 is obtained, in general, by solving the simultaneous equations

$$\partial J[u_n(\cdot; c)]/\partial c_j = 0, \qquad j = 1, \ldots, n. \tag{11}$$

The function u_n appearing in (10) is often taken to be a so-called linear admissible function. For instance, in the example above, let $\{\varphi_k\}_{k=1}^{\infty}$ be a system of independent functions complete in $\tilde{A}_J$, and set

$$u_n = c_1\varphi_1 + \ldots + c_n\varphi_n. \tag{12}$$

The method that constructs a minimizing sequence $u_1, u_2, \ldots$ by determining the value of the c_k in (12) by (11) is called **Ritz's method,** and φ_k is called a coordinate function in this method. As for the rate of convergence in approximation by Ritz's method, as well as for the estimation of errors, there are several results by the Soviet school in addition to those of E. Trefftz [3, 5]. (Other methods of constructing minimizing sequences are stated in detail in [3]; concerning a connection with †Galerkin's method → 304 Numerical Solution of Partial Differential Equations.)

Ritz's method applied to eigenvalue problems is called the **Rayleigh-Ritz method.** Since a stationary value of the Rayleigh quotient is itself an eigenvalue in this case, the precision of approximation is far better for eigenvalues than for eigenfunctions, so that it is a convenient method for the approximate computation of eigenvalues.

References

[1] R. Courant and D. Hilbert, Methods of mathematical physics, Interscience, 1, 1953; II, 1962.

[2] P. Funk, Variationsrechnung und ihre Anwendungen in Physik und Technik, Springer, 1962.

[3] L. V. Kantrovich and V. I. Krylov, Approximation methods of higher analysis (in Russian), Gostekhizdat, 1952.

[4] S. G. Mikhlin, Variational methods in mathematical physics, Macmillan, 1964. (Original in Russian, 1957.)

[5] S. G. Mikhlin, The problem of the minimum of a quadratic functional, Holden-Day, 1965. (Original in Russian, 1952.)

[6] S. L. Sobolev, Applications of functional analysis in mathematical physics, Amer. Math. Soc. Transl. of Math. Monographs, 1963. (Original in Russian, 1950.)

[7] E. Goursat, Cours d'analyses mathématique, III, Gauthier-Villars, 1927.

[8] C. Carathéodory, Calculus of variations and partial differential equations of the first order, Holden-Day, 1967.

[9] O. Bolza, Vorlesungen über Variationsrechnung, Teubner, 1909.

[10] I. M. Gel'fand and S. V. Fomin, Calculus of variations, Prentice-Hall, 1963. (Original in Russian, 1961.)
[11] C. B. Morrey, Multiple integrals in the calculus of variation, Springer, 1966.

47 (XXI.15) Cantor, Georg

George Cantor (March 3, 1845–January 6, 1918), the founder of set theory, was born in St. Petersburg into a Jewish merchant family that settled in Germany in 1856. He studied mathematics, physics, and philosophy in Zurich and at the University of Berlin. After receiving his degree in 1867 in Berlin, he became a lecturer at the University of Halle and served as a professor at that university from 1879 to 1905. In 1884, under the strain of opposition to his ideas and his efforts to prove the †continuum hypothesis, he suffered the first of many attacks of depression which continued to hospitalize him from time to time until his death.

The thesis he wrote for his degree concerned the theory of numbers; however, he arrived at set theory from his research concerning the uniqueness of †trigonometric series. In 1874, he introduced for the first time the concept of †cardinal numbers, with which he proved that there were "more" †transcendental numbers than †algebraic numbers. This result caused a sensation in the mathematical world and became the subject of a great deal of controversy. Cantor was troubled by the opposition of L. †Kronecker, but he was supported by J. W. R. †Dedekind and G. Mittag-Leffler. In his note on point-set theory, he wrote, in connection regard with his concept of infinity, "The essence of mathematics lies in its freedom!" In addition to his work on cardinal numbers, he laid the basis for the concepts of †order types, †transfinite ordinals, and the theory of real numbers by means of †fundamental sequences. He also studied general point sets in Euclidean space and defined the concepts of †accumulation point, †closed set, and †open set. He was a pioneer of the point-set theory that led to the development of general †topology.

References

[1] G. Cantor, Gesammelte Abhandlungen, E. Zermelo and A. Fraenkel (eds.), Springer, 1932 (Olms, 1962).
[2] A. Schoenflies, Die Krisis in Cantors mathematischem Schaffen, Acta Math., 50 (1927), 1-23.

48 (X.31) Capacity

A. General Remarks

The electric capacity of a conductor in the 3-dimensional Euclidean space $\mathbf{R}^3$ is defined as the ratio of a given positive charge on the conductor to the value of the potential on the surface. This definition of capacity is independent of the given charge. The capacity of a set as a mathematical notion was defined first by N. Wiener (1924) and was developed by O. Frostman, C. J. de La Vallée Poussin, and several other French mathematicians in connection with †potential theory.

B. Energy

Let Ω be a †locally compact Hausdorff space and $\Phi(x, y)$ be a †lower semicontinuous function on $\Omega \times \Omega$ such that $-\infty < \Phi \leqslant \infty$. A measure μ will mean a nonnegative †Radon measure with compact †support S_μ. Denote by $\Phi(x, \mu)$ the †potential $\int \Phi(x, y) d\mu(y)$ of a measure μ with kernel Φ and by (μ, μ) the †energy $\iint \Phi \, d\mu \, d\mu$ of μ. Let X be a set in Ω, and denote by $\mathscr{U}_X$ the class of normalized measures μ (i.e., of measures μ satisfying $\mu(\Omega) = 1$) with $S_\mu \subset X$. Let K be a nonempty compact set in Ω. Set $W(K) = \inf(\mu, \mu)$ for $\mu \in \mathscr{U}_K$, and $W(\varnothing) = \infty$ for the empty set $\varnothing$. For $\Phi(x, y) = 1/|x-y|$ in $\Omega = \mathbf{R}^3$, the general solution $u(x)$ of the †Dirichlet problem (†exterior problem) for the boundary function 1 in the unbounded component of $\mathbf{R}^3 - K$ is equal to the potential of an †equilibrium mass-distribution. Therefore, if S is a smooth surface surrounding K and normals are drawn outward to S, the integral $-(1/4\pi)\cdot\int_S (\partial u/\partial n) d\sigma$ of the normal derivative is equal to $1/W(K)$. This is the **capacity** of K defined by N. Wiener (*J. Math. Phys.*, MIT, 3 (1924)) when K is a closed region. Vallée-Poussin (*Ann. Inst. H. Poincaré*, 2 (1932)) called the supremum of $\mu(\mathbf{R}^3)$ the Newtonian capacity of a bounded †Borel set E, where μ runs through the class of measures μ with $S_\mu \subset E$ whose †Newtonian potentials are not greater than 1 in $\mathbf{R}^3$. If E is compact, the Newtonian capacity coincides with Wiener's capacity. For the †logarithmic potential in $\mathbf{R}^2$, $e^{-W(K)}$ is called the **logarithmic capacity**. When †Green's function $g(z, \infty)$ with the pole at the point at infinity exists in the unbounded component of $\mathbf{R}^2 - K$, $\lim_{z\to\infty}(g(z,\infty) - \log|z|)$ is called **Rohin's constant** and can be shown to be equal to $W(K)$. (For the relation between Robin's constant and †reduced extremal distance → 143 Extremal Length.) In the case of a general

kernel it is difficult to define capacity as above by means of W(K), and hence the value of W(K) itself instead of the capacity of K is often used. When $W(K)=\infty$, we can say that K is of capacity zero. The minimum value of the †Gauss integral $(\mu,\mu)-2\int f\,d\mu$ is a generalization of $W(K)$, where $\mu\in\mathscr{U}_K$ and f is an upper semicontinuous function bounded above on K.

C. Minimax Value

Suppose that we are given the kernel Φ as above. For a set $X\subset\Omega$ and a measure μ, set $U(\mu;X)=\sup_{x\in X}\Phi(x,\mu)$ and $V(\mu;X)=\inf_{x\in X}\Phi(x,\mu)$. Next, for $Y\subset\Omega$, set $U(Y)=\inf U(\mu;S_\mu)$, $V(Y)=\sup V(\mu;S_\mu)$, $U(X,Y)=\inf U(\mu;X)$, and $V(X,Y)=\sup V(\mu;X)$, where $\mu\in\mathscr{U}_Y$. If $\check\Phi(x,y)=\Phi(y,x)$ is taken as a kernel instead of $\Phi(x,y)$, then the notations $\check U(K)$, $\check U(\mu;X)$, $\check V(\mu;X)$, are used correspondingly. For any compact set K the following relations hold:

$$W(K)=\check W(K)\leqslant U(K)=\check U(K)$$

$$\leqslant\begin{cases}U(K,K)=\check V(K,K)\\ \check U(K,K)=V(K,K)\end{cases}\leqslant\begin{cases}U(\Omega,K)=\check V(K,\Omega)\\ \check U(\Omega,K)=V(K,\Omega)\\ V(K)=\check V(K).\end{cases}$$

Examples show that all the inequalities can be strict. The †minimax theorem in the †theory of games plays an important role in the proof of these inequalities [7]. Even if the kernel is symmetric, the inequalities can be strict except for the equality $W(K)=U(K)$. When the kernel is positive, we can define the quantities which correspond to $U(Y)$, $V(Y)$, $U(X,Y)$, $V(X,Y)$ by considering the class of μ with $S_\mu\subset Y$ and $(\mu,\mu)=1$ instead of $\mathscr{U}_Y$.

D. Transfinite Diameter

As $k\to\infty$,

$$D_k(K)=k^{-1}(k-1)^{-1}\inf_{x_1,\ldots,x_k\in K}\sum_{i\neq j}\Phi(x_i,x_j)$$

decreases and the limit $D(K)$ is equal to $W(K)$. For the logarithmic kernel in $\mathbf{R}^2$, M. Fekete defined $D(K)$ and called $e^{-D(K)}$ the **transfinite diameter** of K (1923). F. Leja and his school in Poland studied relations between transfinite diameter and †conformal mapping. Next, set

$$kR_k(X,Y)=\sup_{x_1,\ldots,x_k\in Y}\inf_{x\in X}\sum_{i=1}^{k}\Phi(x,x_i).$$

Then $R(X,Y)=\lim R_k(X,Y)$ exists as $k\to\infty$, and we have $R(K,Y)=V(K,Y)$.

Fekete introduced $R(K)=R(K,K)$ in $\mathbf{R}^2$ (1923). G. Polya and G. Szegö computed $D(K)$ and $R(K)$ for special K and α-kernel $r^{-\alpha}$ ($\alpha\gtrless 0$) in $\mathbf{R}^2$ and $\mathbf{R}^3$ [12]. The equality $D(K)=R(K)$ holds for the logarithmic kernel in $\mathbf{R}^2$ and the Newtonian kernel in $\mathbf{R}^3$. The maximum of the absolute value on K of †Chebyshev's polynomial (→ 336 Polynomial Approximation) of order k with respect to K in $\mathbf{R}^2$ is equal to $\exp(-kR_k(K))$.

E. Evans's Theorem

In order that K be of Newtonian capacity zero, it is necessary and sufficient that there exist a measure μ on K such that the Newtonian potential of μ is equal to ∞ at every point of K. This result was proved by G. C. Evans and H. Selberg independently (1935) and is called **Evans's theorem** (or the **Evans-Selberg theorem**). The corresponding theorem in $\mathbf{R}^2$ is often applied in the theory of functions [15, 16]. A similar potential exists in case of a general kernel if and only if $R(K,K)=\infty$.

F. Nonadditivity of Capacity

Many kinds of capacity satisfy the inequality $\operatorname{cap}(\bigcup_n X_n)\leqslant\sum_n\operatorname{cap}X_n$, where a capacity is denoted by cap. Even the Newtonian capacity C is not necessarily additive, but it satisfies $C(K_1\cup K_2)+C(K_1\cap K_2)\leqslant C(K_1)+C(K_2)$ (G. Choquet [2]). Choquet [3] proved that X can be divided into mutually disjoint sets X_1 and X_2 such that $C_i(X)=C_i(X_1)=C_i(X_2)$, where $C_i(X)$ is the **Newtonian inner** (or **interior**) **capacity** defined to be $\sup_{K\subset X}C(K)$ if $X\neq\varnothing$ and 0 if $X=\varnothing$.

G. Relation to Hausdorff Measure

There are many studies of relations between capacity and †Hausdorff measure [1]. Frostman [6] introduced the notion of **capacitary dimension** and observed that it coincides with the Hausdorff dimension. The capacity of product sets has been evaluated from above and below [11]. For compact sets $K\subset\mathbf{R}^n$, $K'\subset\mathbf{R}^m$, their dimensions α, β, and the dimension γ of $K\times K'$, we have the relation $\alpha+\beta\leqslant\gamma\leqslant\min(m+\alpha,n+\beta)$, where the equalities are attained by general †Cantor sets. There are also works on the evaluation of capacities of general Cantor sets [10, 15]. If K is a continuum of logarithmic capacity 1 in a plane, then its diameter d satisfies $2\leqslant d\leqslant 4$, and its area A satisfies $A<\pi$ [S]. Consider the sum $K=\{z_1+\ldots+z_n\mid z_k\in K_k, 1\leqslant k\leqslant n\}$ of continua $K_1,\ldots,K_n$ in a plane. The logarithmic capacity of K is strictly greater than the sum of the logarithmic capacities of $K_1,\ldots,K_n$ except when all K_k are convex and mutually similar [14]. By various †symmetrizations the logarithmic capacity decreases in general (→ 228 Isoperimetric Problems; also [133).

H. Capacitability

The **Newtonian outer** (or **exterior) capacity** $C_e(X)$ is defined by inf $C_i(G)$, where G ranges over an open set containing X. The inequality $C_i(X) \leqslant C_e(X)$ holds, in general. When the equality holds, X is called **capacitable.** Choquet (1955) [2] proved that all †analytic sets and hence †Borel sets are capacitable but there exists an analytic set whose complement is not capacitable. He himself generalized his result on capacitability in the following way [4]: Let Ω be an abstract space, φ a nondecreasing function defined on the family of all subsets of Ω, and $\mathscr{K}$ some family of subsets of Ω that is closed under the formation of finite unions and countable intersections. Assume that $\varphi(H_n)\downarrow\varphi(H)$ as H_n in $\mathscr{K}$ decreases to H and that $\varphi(X_n)\uparrow\varphi(X)$ as $X_n\uparrow X$. When $\varphi(X)$ is equal to $\sup\{\varphi(H) \mid H\in\mathscr{K}, H\subset X\}$, X is called $(\varphi, \mathscr{K})$-capacitable. Choquet defined $\mathscr{K}$-Suslin sets and showed that they are $(\varphi, \mathscr{K})$-capacitable. M. Kishi [9], Choquet [5], and B. Fuglede [7] investigated capacitability with respect to several kinds of capacity more general than Newtonian capacity. We can discuss capacitability with respect to quantities defined in connection with the †Gauss variational problem.

References

[1] L. Carleson, Selected problems on exceptional sets, Van Nostrand, 1967.
[2] G. Choquet, Theory of capacities, Ann. Inst. Fourier, 5 (1955), 131-295.
[3] G. Choquet, Potentiels sur un ensemble de capacité nulle, Suites de potentiels, C. R. Acad. Sci. Paris, 244 (1957), 1707–1710.
[4] G. Choquet, Forme abstraite du théorème de capacitabilité, Ann. Inst. Fourier, 9 (1959), 83-89.
[5] G. Choquet, Diamètre transfini et comparaison de diverses capacités, Sém. Théorie du Potentiel, 3 (1958–1959), no. 4.
[6] 0. Frostman, Potentiel d'équilibre et capacité des ensembles avec quelques applications à la théorie des fonctions, Medd. Lunds Univ. Mat. Sem., 3 (1935).
[7] B. Fuglede, Une application du théorème du minimax à la théorie du potentiel, Colloque Internat. sur la Théorie du Potentiel, Paris, Orsay, 1964, C. N. R. S. exposé no. 8.
[S] G. M. Goluzin, Geometrische Funktionentheorie, Deutscher Verlag der Wiss., 1957.
[9] M. Kishi, Capacitability of analytic sets, Nagoya Math. J., 16 (1960), 91–109.
[10] M. Ohtsuka, Capacité d'ensembles de Cantor généralisés, Nagoya Math. J., 11 (1957), 151–160.
[1 1] M. Ohtsuka, Capacité des ensembles produits, Nagoya Math. J., 12 (1957), 95–130.
[12] G. Pólya and G. Szegö, Über den transfiniten Durchmesser (Kapazitatskonstante) von ebenen und räumlichen Punktmengen, J. Reine Angew. Math., 165 (1931), 4-49.
[13] G. Pólya and G. Szegö, Isoperimetric inequalities in mathematical physics, Princeton Univ. Press, 1951.
[143 C. Pomerenke, Zwei Bemerkungen zur Kapazitat ebener Kontinuen, Arch. Math., 12 (1961), 122–128.
[15] M. Tsuji, Potential theory in modern function theory, Maruzen, 1959.
[16] L. Sario and K. Oikawa, Capacity functions, Springer, 1969.

49 (11.6) Cardinal Numbers

A. Definition

The general concept of cardinal number is an extension of that of natural number (Cantor [1]). When there exists a †one-to-one correspondence whose †domain is a set A and whose †range is a set *B*, this set *B* is said to be **equipotent** (or **equipollent)** to *A*, and this relation is denoted by *A-B*. The relation $\sim$ is an †equivalence relation, and each equivalence class under this relation is said to be a **cardinal number.** The class of all sets equipotent to a set *A* is denoted by *A* (or $|A|$) and is said to be the **cardinal number (power, cardinality,** or **potency) of tbe set** *A*. When *A* is a finite set, *A* is said to be **finite**, and when *A* is an infinite set, *A* is said to be **infinite** (or **transfinite).** When the cardinal number of a set *A* is m, *A* is also said to consist of m members (or m elements). In this sense, 0 and the natural numbers are considered to express finite cardinal numbers. For example, 0= $\varnothing$, 1 = {0}, 2 = {0,1}, etc. Examples of infinite cardinal numbers: A set *A* which is equipotent to the set N of all natural numbers is said to be **countably infinite,** and the cardinal number of the set N is denoted by a. A set *A* which is finite or countably infinite is said to be **countable.** The cardinal number of the set of all real numbers is denoted by c and is called the **cardinal number of tbe continuum.** Moreover, the cardinal number of the set of all real-valued functions whose domain is the interval [0, 1] is denoted by f. These three cardinal numbers are known to be distinct. Henceforth in this article, lower-case German letters denote cardinal numbers. For a definition of cardinal numbers using the concept of ordinal numbers → 312 Ordinal Numbers.

B. Ordering of Cardinal Numbers

$\mathfrak{m} \geqslant \mathfrak{n}$ or $\mathfrak{n} \leqslant \mathfrak{m}$ will mean that there exist sets A and B such that $\mathfrak{m}=\overline{\overline{A}}$, $\mathfrak{n}=\overline{\overline{B}}$, and $A \supset B$. $A \neq B$ does not necessarily imply $\mathfrak{m} \neq \mathfrak{n}$. For example, the cardinal number of the set of all positive even numbers is also $\mathfrak{a}$. $\mathfrak{m} \geqslant \mathfrak{n}$ and $\mathfrak{n} \geqslant \mathfrak{m}$ imply $\mathfrak{m}=\mathfrak{n}$ (**Bernshteĭn's theorem**). Since the †reflexive and †transitive laws for the relation $\leqslant$ between cardinal numbers are obvious, the relation is an †ordering relation. The †well-ordering theorem imples that $\geqslant$ is a †total ordering (**comparability theorem for cardinal numbers**). $\mathfrak{m} > \mathfrak{n}$ means that $\mathfrak{m} \geqslant \mathfrak{n}$ and $\mathfrak{m} \neq \mathfrak{n}$. When $B \leqslant \mathfrak{m}$, B is said to be **at most** $\mathfrak{m}$.

C. Sum, Product, and Power of Cardinal Numbers

For cardinal numbers $\mathfrak{m}$ and $\mathfrak{n}$, choose sets A and B so that $\mathfrak{m}=\overline{\overline{A}}$, $\mathfrak{n}=\overline{\overline{B}}$, and $A \cap B=\varnothing$, and put $\mathfrak{s}=\overline{\overline{A \cup B}}$. Then $\mathfrak{s}$ is uniquely determined by $\mathfrak{m}$ and $\mathfrak{n}$. The $\mathfrak{s}$ is said to be the **sum** of the cardinal numbers $\mathfrak{m}$ and $\mathfrak{n}$ and is denoted by $\mathfrak{m}+\mathfrak{n}$. If the sets A, B are chosen as described above, the cardinal numbers of the †Cartesian product $A \times B$ and of the set of functions A^B are called the **product** of $\mathfrak{m}$ and $\mathfrak{n}$ and the $\mathfrak{n}$th **power** of $\mathfrak{m}$, denoted by $\mathfrak{m}\mathfrak{n}$ and $\mathfrak{m}^{\mathfrak{n}}$, respectively. These operations are also determined by $\mathfrak{m}$ and $\mathfrak{n}$. For these three operations, the following laws are valid: **commutative laws** $\mathfrak{m}+\mathfrak{n}=\mathfrak{n}+\mathfrak{m}$, $\mathfrak{m}\mathfrak{n}=\mathfrak{n}\mathfrak{m}$; **associative laws** $(\mathfrak{m}+\mathfrak{n})+\mathfrak{p}=\mathfrak{m}+(\mathfrak{n}+\mathfrak{p})$, $(\mathfrak{m}\mathfrak{n})\mathfrak{p}=\mathfrak{m}(\mathfrak{n}\mathfrak{p})$; **distributive law** $\mathfrak{p}(\mathfrak{m}+\mathfrak{n})=\mathfrak{p}\mathfrak{m}+\mathfrak{p}\mathfrak{n}$; **exponential laws** $\mathfrak{m}^{\mathfrak{n}+\mathfrak{p}}=\mathfrak{m}^{\mathfrak{n}}\mathfrak{m}^{\mathfrak{p}}$, $\mathfrak{m}^{\mathfrak{n}\mathfrak{p}}=(\mathfrak{m}^{\mathfrak{n}})^{\mathfrak{p}}$, $(\mathfrak{m}\mathfrak{n})^{\mathfrak{p}}=\mathfrak{m}^{\mathfrak{p}}\mathfrak{n}^{\mathfrak{p}}$. In particular, if $\overline{\overline{A}}=\mathfrak{m}$, then $2^{\mathfrak{m}}$ is the cardinal number of the †power set $\mathfrak{P}(A)$ of A.

Addition and multiplication of more than two cardinal numbers can be defined as follows. Let Λ be any set, and suppose that to any element λ of Λ there corresponds a unique cardinal number $\mathfrak{m}_\lambda$. Let M_λ be a set such that $\overline{\overline{M_\lambda}}=\mathfrak{m}_\lambda$, and $M_\lambda \cap M'_\lambda=\varnothing$ for $\lambda \neq \lambda'$. Then the cardinal number of the †disjoint union $\sum_\lambda M_\lambda$ is said to be the **sum** of all $\mathfrak{m}_\lambda$ and is denoted by $\sum_\lambda \mathfrak{m}_\lambda$. The cardinal number of the Cartesian product $\prod_\lambda M_\lambda$ is said to be the **product** of all $\mathfrak{m}_\lambda (\lambda \in \Lambda)$ and is denoted by $\prod_\lambda \mathfrak{m}_\lambda$. The axiom of choice can be stated as follows: If $\mathfrak{m}_\lambda \neq 0$ for all $\lambda \in \Lambda$, then $\prod_\lambda \mathfrak{m}_\lambda \neq 0$.

D. The Continuum Hypothesis

For $\mathfrak{a}$, $\mathfrak{c}$, and $\mathfrak{f}$ defined as before, $\mathfrak{f}=2^{\mathfrak{c}}>\mathfrak{c}=2^{\mathfrak{a}}>\mathfrak{a}$. In general, $2^{\mathfrak{m}}>\mathfrak{m}$ holds for any cardinal number $\mathfrak{m}$ (Cantor). The hypothesis which asserts that for any $\mathfrak{m}$, there does not exist an $\mathfrak{n}$ such that $2^{\mathfrak{m}}>\mathfrak{n}>\mathfrak{m}$ is called the **generalized continuum hypothesis**. In particular, this hypothesis restricted to the case where $\mathfrak{m}=\mathfrak{a}$ is called the **continuum hypothesis**. After Cantor stated this hypothesis (*J. Reine Angew. Math.*, 84 (1878)), it remained an open question for many years. In particular, Cantor himself repeatedly tried to prove it, and W. Sierpiński pursued various related hypotheses. Finally, the continuum hypothesis and the generalized continuum hypothesis were proved to be independent of the axioms of set theory by K. Gödel (1940) [3] and P. J. Cohen (1963) [4] (→ 33 Axiomatic Set Theory).

E. Cardinality of Ordinal Numbers

Lower-case Greek letters will stand here for †ordinal numbers. The cardinal number of $\{\xi \mid \xi<\alpha\}$ will be denoted by $\overline{\alpha}$, which is called the **cardinality of the ordinal number** α or the **cardinal number corresponding to** α. When a cardinal number $\mathfrak{m}$ corresponds to some ordinal number, the minimum among ordinal numbers α with $\overline{\alpha}=\mathfrak{m}$ is called the **initial ordinal number** corresponding to $\mathfrak{m}$. An initial ordinal number corresponding to an infinite cardinal number is called a **transfinite initial ordinal number**. There exists a unique correspondence $\beta \rightarrow \omega_\beta$ from the class of ordinal numbers onto the class of all transfinite initial ordinal numbers such that $\beta>\gamma$ implies $\omega_\beta>\omega_\gamma$. In particular, $\omega_0=\omega$, and an ordinal number ξ such that $\xi<\omega_1$ is called a **countable ordinal number**. ω_β is called the βth transfinite initial ordinal number. The cardinality of ω_β is denoted by $\aleph_\beta$ ($\aleph$ is the Hebrew letter **aleph**). In particular, $\mathfrak{a}$ is denoted by $\aleph_0$ (**aleph zero**). We have $\aleph_\beta \geqslant \aleph_\gamma$ if and only if $\beta \geqslant \gamma$, and in this case $\aleph_\beta+\aleph_\gamma=\aleph_\beta$, $\aleph_\beta\aleph_\gamma=\aleph_\beta$. The axiom of choice implies that every infinite cardinal number is an $\aleph_\beta$. Hence, in this case, the continuum hypothesis can be formulated as $2^{\aleph_0}=\aleph_1$, and the generalized continuum hypothesis can be formulated as $2^{\aleph_\beta}=\aleph_{\beta+1}$ for every ordinal number β.

F. Finiteness and Infiniteness

Dedekind [5] defined a set A to be **infinite** if A is equipotent to a proper subset of itself, and to be **finite** otherwise. It is also possible to define finiteness and infiniteness of sets as follows: A set A is finite if there exists a †well-ordering of A such that its †dual ordering is also a well-ordering, and A is infinite other-

wise. If a set is finite in the latter sense, then it is also finite in the sense of Dedekind. Under the axiom of choice, these two definitions can be shown to be equivalent.

References

[1] G. Cantor, Beiträge zur Begründung der transfinite Mengenlehre I, Math. Ann., 46 (1895), 471–512. Gesammelte Abhandlungen, Springer, 1932; English translation, Contributions to the founding of the theory of transfinite numbers, Open Court, 1915.
[2] W. Sierpiński, Hypothèse du continu, Warsaw, 1934.
[3] K. Gödel, The consistency of the continuum hypothesis, Ann. Math. Studies, Princeton Univ. Press, 1940.
[4] P. J. Cohen, Set theory and the continuum hypothesis, Benjamin, 1966.
[5] R. Dedekind, Was sind und was sollen die Zahlen? Vieweg, 1888. Gesammelte mathematische Werke 3, Vieweg, 1932; English translation, Essays on the theory of numbers, Open Court, 1901.
Also → references to 381 Sets.

50 (XXI.16) Cartan, Elie

Elie Cartan (April 9, 1869–May 6, 1951) was born at Dolomieu in the French province of Isère. He entered the Ecole Normale Supérieure in Paris in 1888 and graduated in 1891, having at the same time qualified in the agrégé examination. Beginning his research immediately, he completed his thesis on the structure of continuous transformation groups [2] in 1894 at the age of 25.

Cartan was a professor first at the University of Montpellier, later at the University of Lyon, then the University of Nancy, and finally in 1912 at the University of Paris. He freely used the †moving coordinate system introduced by J. G. Darboux, and contributed to many areas, such as the theory of †Lie groups, the theory of †Pfaffian forms, the theory of †invariant integrals, †topology, †differential geometry (especially the geometry of †connections), and theoretical physics. His doctoral thesis is still an object of interest among young researchers today, and the concept of connection introduced by him is fundamental in the field of differential geometry. Henri Cartan (1904–) is his eldest son.

References

[1] E. Cartan, Oeuvres complètes I–III, Gauthier-Villars, 1952–1955.
[2] E. Cartan, Sur la structure des groupes de transformations finis et continus, Thesis, 1894. (Oeuvres complètes, pt. I, vol. 1, 137–287.)
[3] S. S. Chern and C. Chevalley, Elie Cartan and his mathematical work, Bull. Amer. Math. Soc., 58 (1952), 217–250.

51 (IX.24) Catastrophe Theory

A. General Remarks

Catastrophe theory was originally proposed by R. Thom [1, 2] in late 1960. This theory provides certain mathematical models for the evolution of forms in nature, in particular in biology and the natural languages. In recent years, E. C. Zeeman and others have developed and applied the theory to various fields, including the physical sciences, medicine, economics, and sociology; a detailed bibliography may be found in [3, 9, 10].

B. Static Models

The basic concept of the theory is the **static model**. This is a family of potential functions $f_u: X \to \mathbf{R}$, where X is a subset of $\mathbf{R}^n$, containing a neighborhood of the origin, and the parameter u lies in a neighborhood U of the origin in $\mathbf{R}^r$.

We regard $\mathbf{R}^n$ as the **internal space** or **state space**, which is parametrized by the various variables that are relevant to the process under study, and $\mathbf{R}^r$ is the **external space** or **control space**, which can be either the physical space-time continuum in which the process under consideration takes place or a space of control parameters that govern the process. As a rule, we assume that the dimension $r \leqslant 4$, although the dimension n can be arbitrary large.

The static model is local in nature, and any one of the local minima of f_u, called a **local regime** at $u \in U$, is a candidate for the state of the model corresponding to the control point u.

Mathematically, a static model is a germ of C^∞-functions $f: \mathbf{R}^n \times \mathbf{R}^r \to \mathbf{R}$ at 0 that is an unfolding (r-dimensional extension family) of a germ of C^∞-functions $\eta = f \mid \mathbf{R}^n \times \{0\}: \mathbf{R}^n \to \mathbf{R}$ at 0; further details are given in Section D.

C. Classification of Singularities

Let $\mathscr{E}(n,m)$ be the vector space of †germs of C^∞-functions $f:\mathbf{R}^n\to\mathbf{R}^m$ at 0, and let $\mathscr{B}(n)\subset\mathscr{E}(n,n)$ be the subset of invertible germs $\mathbf{R}^n\to\mathbf{R}^n$ mapping 0 to 0.

The germs η, $\xi\in\mathscr{E}(n,m)$ are called **right equivalent** if there exists an $h\in\mathscr{B}(n)$ such that $\eta\circ h=\xi$. Let $\mathscr{E}(n)=\mathscr{E}(n,1)$. Then it follows in the usual manner that $\mathscr{E}(n)$ is a local algebra with the unique maximal ideal $\mathscr{M}=\mathscr{M}(n)=\{\eta\in\mathscr{E}(n)\,|\,\eta(0)=0\}$. A germ $\eta\in\mathscr{E}(n)$ such that $\eta(0)=D\eta(0)=0$ is called a **singularity**. For any singularity η, we define the **codimension** of η by

$$\operatorname{codim}\eta=\dim_{\mathbf{R}}(\mathscr{M}/\langle\partial\eta/\partial x_i\rangle_{\mathscr{E}(n)}),$$

where $\langle\partial\eta/\partial x_i\rangle_{\mathscr{E}(n)}$ is the ideal of $\mathscr{E}(n)$ generated by $\partial\eta/\partial x_1,\ldots,\partial\eta/\partial x_n$ and $x=(x_1,\ldots,x_n)$ denotes the coordinate system of $\mathbf{R}^n$.

The following result was proved by J. N. Mather [4]: Up to the addition of a nondegenerate quadratic form in other variables, and up to multiplication by ±1, a singularity of codimension $\leqslant 4$ and $\geqslant 1$ is right equivalent to one of the η appearing in Thom's list of the seven elementary catastrophes below.

D. Unfoldings

Let η be a singularity. An **r-unfolding** of η is a germ $f\in\mathscr{M}(n+r)$ such that $f\,|\,\mathbf{R}^n\times\{0\}=\eta$; this unfolding is denoted by (r,f). Let (r,f) and (s,g) be unfoldings of η. A **morphism** $(\varphi,\Phi,\varepsilon)$: $(r,f)\to(s,g)$ consists of (i) a germ $\varphi\in\mathscr{E}(n+r,n+s)$ such that $\varphi\,|\,\mathbf{R}^n\times\{0\}=$identity, (ii) a germ $\Phi\in\mathscr{E}(r,s)$ such that $\pi_s\circ\varphi=\Phi\circ\pi_r$, (iii) a germ $\varepsilon\in\mathscr{M}(r)$ such that $f=g\circ\varphi+\varepsilon\circ\pi_r$, where $\pi_r:\mathbf{R}^n\times\mathbf{R}^r\to\mathbf{R}^r$ is the projection. In this case we say that the unfolding (r,f) is **induced** by $(\varphi,\Phi,\varepsilon)$ from (s,g). A morphism is an **isomorphism** if φ and Φ are diffeomorphic germs.

The **addition** of unfoldings (r,f) and (s,g) of η is defined by $(r,f)+(s,g)=(r+s,f+g-\eta)$, where the last term on the right is given by $(f+g-\eta)(x,u,v)=f(x,u)+g(x,v)-\eta(x)$.

Thus if the **constant unfolding** (s,η) of η is defined by $\eta(x,u)=\eta(x)$, we have $(r,f)+(s,\eta)=(r+s,f)$.

An unfolding (r,f) of η is said to be **versal** if any unfolding of η is induced by (r,f) and a suitable morphism. A versal unfolding (r,f) with minimal r is said to be **universal**. The following facts have also been proved by Mather [4]:

A singularity $\eta\in\mathscr{M}(n)$ has a versal unfolding if and only if $\operatorname{codim}\eta$ is finite. Any two r-versal unfoldings of η are isomorphic. Every versal unfolding is isomorphic to $(r,f)+$constant, where $r=\operatorname{codim}\eta$ and (r,f) is a universal unfolding defined as follows: If $\{b_1,\ldots,b_r\}\subset\mathscr{M}(n)$ is a system of representatives for a basis of $\mathscr{M}(n)/\langle\partial\eta/\partial x_i\rangle_{\mathscr{E}(n)}$, then the unfolding f of η is defined by $f(x,u)=\eta(x)+b_1(x)u_1+\ldots+b_r(x)u_r$.

E. The Seven Elementary Catastrophes

Let f and g be germs in $\mathscr{E}(n+r)$. We say that f and g are **equivalent** as r-unfoldings if there exist $h\in\mathscr{B}(r)$, a family of $H_u\in\mathscr{B}(n)$, where $u\in U\subset\mathbf{R}^r$, and an $\varepsilon\in\mathscr{M}(r)$ such that $f(x,y)=g(H_u(x),h(u))+\varepsilon(u)$.

We say that a static model (r,f) is **stable** if any small perturbation (r,g) of (r,f) in $\mathscr{E}(n+r)$ (with the Whitney C^∞-topology) is equivalent to (r,f). The following is the main result observed by Thom and proved by Mather and others (for references see the bibliography in [3]). Suppose that $r\leqslant 4$. Then the set of stable static models (r,f) is an open dense subset of $\mathscr{E}(n+r)$, and up to the addition of a nondegenerate quadratic form and multiplication by ±1, any stable static model (r,f) is equivalent to one of the models with the standard potentials F, which are the universal unfoldings of singularities η in Table 1 (x,y denote internal variables, and u, v, w, t denote external variables).

The static models with these standard potentials are known as the **elementary catastrophes** and can serve as qualitative models of various natural processes.

Table 1. Thom's List of the Seven Elementary Catastrophes

r	Singularity η	Standard Potential F	Name
1	x^3	x^3+ux	Fold
2	x^4	x^4+ux^2+vx	Cusp (Riemann-Hugo-Niot)
3	x^5	$x^5+ux^3+vx^2+wx$	Swallowtail (Dovetail)
3	x^3+y^3	$x^3+y^3+uxy+vx+wy$	Hyperbolic umbilic
3	x^3-xy^2	$x^3-xy^2+u(x^2+y^2)+vx+wy$	Elliptic umbilic
4	x^6	$x^6+tx^4+ux^3+vx^2+wx$	Butterfly
4	x^2y+y^4	$x^2y+y^4+ux^2+vy^2+wx+ty$	Parabolic umbilic

F. Catastrophe Sets

A **process** for the static model $f \in \mathscr{E}(n+r)$ is a subset s of $X \times U$, where X and U are neighborhoods of the origin in $\mathbf{R}^n$ and $\mathbf{R}^r$, respectively. If s is a process and $u \in U$, we define $s_u = s \cap (X \times \{u\})$. We say that $u \in U$ is a **regular point** for s if there exists a neighborhood V of u in U and a homeomorphism $h: X \times V \rightarrow X \times V$ such that $\pi_r \circ h = \pi_r$ on $X \times V$ and $h(s \cap (X \times V)) = s_u \times V$. A **catastrophe point** is a nonregular point in U. The set of all catastrophe points is called the **catastrophe set**.

There are several conventions with regard to the definition of a process for the unfolding (r, f). One of these is the **Maxwell convention**, which requires that s_u be a least local regime at $u \in U$. The catastrophe set of the model (r, f) according to the Maxwell convention consists of those points u of the control space $\mathbf{R}^r$ where the least minimum of f_u is attained for at least two points or where this minimum is attained at a unique point but is not stable.

Another is the **perfect delay convention**, which assigns to each path τ in U a mapping (possibly discontinuous) $m_\tau: \tau \rightarrow X \times \tau$ such that $\pi_r(m_\tau(u)) = u$, $(m_\tau(u), u)$ is a local regime, and m_τ remains continuous for a maximum interval on the path τ. Consider the set

$$\Delta = \{(x, u) \in \mathbf{R}^n \times \mathbf{R}^r \mid d_x^2 f(x, u) \text{ is degenerate}\}$$

and its image $B = \pi_r(\Delta)$, called the **bifurcation set**, under the projection $\pi_r: \mathbf{R}^n \times \mathbf{R}^r \rightarrow \mathbf{R}^r$. Then the points of the bifurcation set B are important candidates for catastrophe points of the static model with respect to the perfect delay convention.

For geometrical studies of the elementary catastrophes → [6–8]. The static models have been generalized to **metabolic models**, the structure of which largely remains an open question [2]. In this connection, the bibliographies of [3, 9, 10] appear to be fairly complete.

References

[1] R. Thom, Topological models in biology, Topology, 8 (1969), 313–335.
[2] R. Thom, Stabilité structuelle et morphogenèse, Benjamin, 1972; English translation, structural stability and morphogenesis, Benjamin, 1975.
[3] A. Manning (ed.), Dynamical systems—Warwick 1974, Lecture notes in math. 468, Springer, 1975.
[4] J. N. Mather, Right equivalence, Warwick preprint, 1969.
[5] G. Wassermann, Stability of unfoldings, Lecture notes in math. 393, Springer, 1974.
[6] T. Bröcker, Differentiable germs and catastrophes, London Math. Soc. Lecture Notes Series 17, Cambridge Univ. Press, 1975.
[7] A. N. Godwin, Three-dimensional pictures for Thom's parabolic umbilic, Publ. Math Inst. HES, 40 (1971), 117–138.
[8] A. E. R. Woodcock and T. Poston, A geometrical study of the elementary catastrophes, Lecture notes in math. 373, Springer, 1974.
[9] T. Poston and I. Stewart, Catastrophe theory and its applications, Pitman, 1978.
[10] E. C. Zeeman, Catastrophe theory, Selected papers 1972–1977, Addison-Wesley, 1977.

52 (II.24) Categories and Functors

A. Categories

Consider the family of all †groups. Given two groups X and Y, denote the set of all †homomorphisms of X to Y by $\mathrm{Hom}(X, Y)$. If X, Y, and Z are groups and if $f: X \rightarrow Y$ and $g: Y \rightarrow Z$ are homomorphisms, we can compose them to get a homomorphism $g \circ f: X \rightarrow Z$.

In general, suppose that we are given, as in this example, (1) a family $\mathfrak{M}$ of **mathematical objects**, and (2) for every pair (X, Y) of objects in $\mathfrak{M}$, a set $\mathrm{Hom}(X, Y)$ whose elements are called **morphisms**; and let $f \in \mathrm{Hom}(X, Y)$ and $g \in \mathrm{Hom}(Y, Z)$ determine a morphism $g \circ f \in \mathrm{Hom}(X, Z)$, which is called their **composite**. A morphism $f \in \mathrm{Hom}(X, Y)$ is also written $f: X \rightarrow Y$. Suppose further that these morphisms satisfy the following axioms: (1) if $f: X \rightarrow Y$, $g: Y \rightarrow Z$, and $h: Z \rightarrow W$ are morphisms, then $(h \circ g) \circ f = h \circ (g \circ f)$; (2) for each object $X \in \mathfrak{M}$ there exists a morphism $1_X: X \rightarrow X$ such that for any $f: X \rightarrow Y$ and $g: Z \rightarrow X$ we have $f \circ 1_X = f$ and $1_X \circ g = g$; (3) $\mathrm{Hom}(X, Y)$ and $\mathrm{Hom}(X', Y')$ are disjoint unless $X = X'$ and $Y = Y'$. Then we call the whole system (i.e., the family of objects $\mathfrak{M}$, the morphisms, and the composition of morphisms) a **category**. The elements in $\mathfrak{M}$ are called the **objects** of the category.

By axioms (1) and (2), the set $\mathrm{Hom}(X, X)$ is a semigroup (with respect to the composition of morphisms) which has 1_X as the identity element. Hence 1_X is determined uniquely by X. On the other hand, axiom (3) implies that a morphism f determines the objects X and Y such that $f \in \mathrm{Hom}(X, Y)$. From these facts we can give an alternative definition of category using only the morphisms and their composition.

The totality of the objects (or morphisms) in a category $\mathscr{C}$ is denoted by $\mathbf{Ob}(\mathscr{C})$ (or $\mathbf{Fl}(\mathscr{C})$;

the notation Fl comes from the French word *flèche*). The relation $x \in \mathrm{Ob}(\mathscr{C})$ is often abbreviated to $x \in \mathscr{C}$, while $\mathrm{Hom}(X, Y)$ is written $\mathrm{Hom}_{\mathscr{C}}(X, Y)$ if necessary. A **subcategory** of a category $\mathscr{C}$ is a category $\mathscr{C}'$ with $\mathrm{Ob}(\mathscr{C}') \subset \mathrm{Ob}(\mathscr{C})$, such that for $X, Y \in \mathscr{C}'$ we have $\mathrm{Hom}_{\mathscr{C}'}(X, Y) \subset \mathrm{Hom}_{\mathscr{C}}(X, Y)$ and the composition in $\mathscr{C}'$ is the restriction of $\mathscr{C}$ to $\mathscr{C}'$. If $\mathrm{Hom}_{\mathscr{C}'}(X, Y) = \mathrm{Hom}_{\mathscr{C}}(X, Y)$ for all $X, Y \in \mathscr{C}'$, we say that $\mathscr{C}'$ is a **full subcategory** of $\mathscr{C}$.

We define the **product category** $\mathscr{C}_1 \times \mathscr{C}_2$ of two categories in the canonical way, using the pairs of objects and the pairs of morphisms.

B. Examples of Categories

(1) Taking all sets as the objects, all mappings as the morphisms, and the composition of mappings as the composition, we obtain a category called the **category of sets**, denoted by (Sets) (or (Ens) from the French *ensemble*). For the empty set $\varnothing$ we make the convention that $\mathrm{Hom}(\varnothing, Y)$ contains just one element for any Y and that $\mathrm{Hom}(Y, \varnothing)$ is empty if $Y \neq \varnothing$.

(2) As we have seen, taking all groups as the objects and the homomorphisms as the morphisms, we get the **category of groups**, written (Gr). If we limit the objects to †Abelian groups, we get the **category of Abelian groups** (Ab) as a full subcategory of (Gr).

(3) Fix a ring R. The left R-modules and their R-linear mappings define the **category of left R-modules**, which we denote by ${}_R\mathscr{M}$. The category of right R-modules, $\mathscr{M}_R$, is defined similarly. When R is †unitary, we usually limit the objects of ${}_R\mathscr{M}$ and $\mathscr{M}_R$ to †unitary modules. If R is commutative we can identify ${}_R\mathscr{M}$ with $\mathscr{M}_R$. When $R = \mathbf{Z}$ (the ring of rational integers), ${}_R\mathscr{M}$ can be identified with (Ab). When R is a field, ${}_R\mathscr{M}$ is also called the **category of linear spaces over R**.

(4) Taking rings as objects and homomorphisms of rings as morphisms, we obtain the **category of rings**. The subcategory consisting of unitary commutative rings and unitary homomorphisms is called the **category of commutative rings** and is denoted by (Rings).

(5) If we take †differentiable manifolds as objects and differentiable mappings as morphisms, we obtain the **category of differentiable manifolds**. Similarly, for †analytic manifolds and analytic mappings we obtain the **category of analytic manifolds**.

(6) Taking topological spaces as objects and continuous mappings as morphisms, we get a category called the **category of topological spaces** and denoted by (Top). On the other hand, if we take the †homotopy classes of continuous mappings as morphisms and define their composition in the natural way, we obtain another category, which is called the **homotopy category of topological spaces**.

(7) Fix a †preordered set I. Taking the elements of I as the objects and the pairs (x, y) of elements of I with $x \leqslant y$ as the (unique) morphism from x to y, we get a category, in which we define the composite of the morphisms (x, y) and (y, z) to be (x, z).

In examples (1) through (6), the totality of the objects $\mathrm{Ob}(\mathscr{C})$ is not a set, but a †class (→ 381 Sets G; for the logical foundation of category theory → [3,9]).

C. Diagrams

If a set of arrows $\{A_\alpha\}$ and a set of points $\{B_\beta\}$ are given in such a way that each arrow A_α has a unique initial point and a unique endpoint, then we say that $\{A_\alpha, B_\beta\}$ is a **diagram**. (Usually, we consider the case where each point B_β is the initial point or the endpoint of at least one A_α (Fig. 1).) Let $\mathscr{C}$ be a category and $\{A_\alpha, B_\beta\}$ a diagram. If we associate a morphism f_α in $\mathscr{C}$ with each arrow A_α and an object $Z_\beta \in \mathscr{C}$ with each point B_β so that $f_\alpha \in \mathrm{Hom}(Z_\beta, Z_\gamma)$ whenever A_α has the initial point B_β and the endpoint B_γ, then we say that $\{f_\alpha, Z_\beta\}$ is a **diagram in the category** $\mathscr{C}$ (Fig. 2). Suppose, furthermore, that the following condition is satisfied: For any pair of points B_β and B_γ, and for any sequence of adjacent arrows $A_{\alpha_1}, \ldots, A_{\alpha_m}$ starting at B_β and ending at B_γ (i.e., the initial point of A_{α_1} is B_β, the endpoint of A_{α_i} is the initial point of $A_{\alpha_{i+1}}$ and the endpoint of A_{α_m} is B_γ), the composite $f_{\alpha_m} \circ f_{\alpha_{m-1}} \circ \ldots \circ f_{\alpha_1}$ ($\in \mathrm{Hom}(B_\beta, B_\gamma)$) depends only on B_β and B_γ. Then the diagram in $\mathscr{C}$ is said to be a **commutative diagram**. For example, commutativity of Fig. 2 is equivalent to $f_3 \circ f_1 = f_4 \circ f_2 = f_5$.

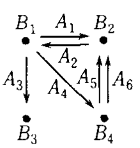


Fig. 1

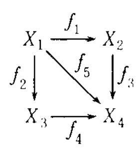


Fig. 2

D. Miscellaneous Definitions

A morphism $f: X \rightarrow Y$ in a category $\mathscr{C}$ is called an **isomorphism** (or **equivalence**) if there exists a morphism $g: Y \rightarrow X$ such that $f \circ g = 1_Y$, $g \circ f = 1_X$. In this case, g is determined uniquely by f and is itself an isomorphism. We call g the **inverse morphism** of f. Then the inverse of g is f. An isomorphism is sometimes written $f: X \simeq Y$. Two objects X and Y are said to be **isomorphic** if there is an isomorphism $X \rightarrow Y$, and then we write $X \cong Y$. The composite of

isomorphisms is again an isomorphism. In particular, an isomorphism $X \to X$ is an invertible element of the semigroup $\mathrm{Hom}(X, X)$, and is called an **automorphism** of X. Examples of isomorphisms are bijections in (Sets), group isomorphisms in (Gr), the R-isomorphisms in ${}_R\mathscr{M}$, ring isomorphisms in the category of rings, †diffeomorphisms in the category of differentiable manifolds, and homeomorphisms in (Top).

A morphism $f: X \to Y$ is called a **monomorphism** (or **injection**) if for any object Z and for any morphisms $u, v: Z \to X (u \neq v)$ we have $f \circ u \neq f \circ v$. Dually, $f: X \to Y$ is called an **epimorphism** (or **surjection**) if for any $u, v: Y \to Z (u \neq v)$ we have $u \circ f \neq v \circ f$. In the category of sets the monomorphisms and the epimorphisms coincide, respectively, with the injections and the surjections as mappings (→ 381 Sets). A monomorphism which is at the same time an epimorphism is called a **bijection**. An isomorphism is always a bijection, but the converse is false in some categories.

Two monomorphisms $f_1: X_1 \to X$ and $f_2: X_2 \to X$ (with the same X) are said to be equivalent if there exist $g_1: X_1 \to X_2$ and $g_2: X_2 \to X_1$ such that $f_1 = f_2 \circ g_1$ and $f_2 = f_1 \circ g_2$ (Fig. 3). An equivalence class with respect to this equivalence relation is called a **subobject** of X. Similarly, we define a **quotient object** of X as an equivalence class of epimorphisms $X \to X_i$.

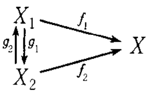


Fig. 3

An **object** e of a category $\mathscr{C}$ is called a **final object** of $\mathscr{C}$ if for every object Y of $\mathscr{C}$, $\mathrm{Hom}(Y, e)$ contains one and only one element. Dually, an object e' is called an **initial object** (or **cofinal object**) if $\mathrm{Hom}(e', Y)$ contains one and only one element for every $Y \in \mathscr{C}$. If e_1 and e_2 are final objects, then there is a unique isomorphism $e_1 \cong e_2$, and similarly for initial objects. A set with only one element is the final object in (Sets), and a space with only one point is the final object in (Top). In the category (Gr) (resp. (Ab)), the trivial group $\{1\}$ (or $\{0\}$) is the final object and the initial object at the same time. In the category of commutative rings, the zero ring $\{0\}$ is the final object and the ring of rational integers $\mathbf{Z}$ is the initial object.

E. Product and Coproduct

Let X_1 and X_2 be objects of a category $\mathscr{C}$. We say that a triple (P, p_1, p_2) consisting of an object P and morphisms $p_i: P \to X_i$ $(i = 1, 2)$ is the **product** (or **direct product**) of X_1 and X_2 if for any pair of morphisms $f_i: X \to X_i$ $(i = 1, 2)$, there exists a unique morphism $f: X \to P$ with $p_i \circ f = f_i$ $(i = 1, 2)$ (Fig. 4). If (P', p'_1, p'_2) is another product of X_1 and X_2, then by virtue of this definition there is a unique morphism $f: P \to P'$ such that $p'_i \circ f = p_i$ $(i = 1, 2)$, and f is an isomorphism. The product is unique in this sense. The product (or any one of the products) of X_1 and X_2 is denoted by $X_1 \times X_2$ or by $X_1 \prod X_2$.

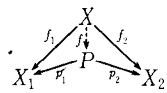


Fig. 4

The product in the categories of sets, of groups, of rings, and of topological spaces coincides with the notion of †direct product in the respective systems. In a general category, the product does not always exist. Suppose the product $X \times X$ exists for an object X; then there is a unique morphism $\Delta_X: X \to X \times X$ such that $1_X = p_1 \circ \Delta_X = p_2 \circ \Delta_x$, which is called the **diagonal morphism** of X. Let $f_i: X \to X'_i$ $(i = 1, 2)$ be morphisms and assume that the products $(X_1 \times X_2, p_1, p_2)$, $(X'_1 \times X'_2, p'_1, p'_2)$ exist. Then there is a unique morphism $f: X_1 \times X_2 \to X'_1 \times X'_2$ satisfying $p'_i \circ f = f_i \circ p_i$ $(i = 1, 2)$. This f is denoted by $f_1 \times f_2$. On the other hand, if $g_i: X \to X_i (i = 1, 2)$ are given, the unique morphism $g: X \to X_1 \times X_2$ with $p_i \circ g = g_i$ $(i = 1, 2)$ is denoted by (g_1, g_2). We have $(g_1, g_2) = (g_1 \times g_2) \circ \Delta_X$ if $X \times X$ exists.

The dual notion of product is coproduct. We say that a triple (S, j_1, j_2) of an object S and morphisms $j_i: X_i \to S$ $(i = 1, 2)$ is the **coproduct** (or **direct sum**) of X_1 and X_2 if for any morphisms $f_i: X_i \to X$ $(i = 1, 2)$ there exists a unique morphism $f: S \to X$ with $f \circ j_i = f_i$ $(i = 1, 2)$ (Fig. 5). The coproduct, like the product, is uniquely determined up to canonical isomorphisms. It is denoted by $X_1 + X_2$ or by $X_1 \coprod X_2$. The coproduct in (Gr) is the †free product. In (Ab), or more generally in ${}_R\mathscr{M}$, the product of two objects can be identified with the coproduct (= direct sum) (→ 277 Modules F). The coproduct in the category of commutative rings is the †tensor product over $\mathbf{Z}$.

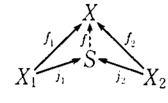


Fig. 5

Product and coproduct can also be defined for a family $\{X_i\}_{i \in I}$ of objects. Namely, the

product of $\{X_i\}_{i\in I}$ is an object P together with a family of morphisms $p_i: P \to X_i$ $(i \in I)$ having the property that for any family of morphisms $f_i: X \to X_i$ $(i \in I)$, there exists a unique morphism $f: X \to P$ such that $p_i \circ f = f_i$ $(i \in I)$. The product is unique up to canonical isomorphisms, and similarly for the coproduct (→ Sections F, L).

F. Dual Category

In the theory of categories we often encounter the dual treatment of notions and propositions. To be precise, we may define the notion of the **dual category** $\mathscr{C}^\circ$ of a category $\mathscr{C}$ as follows: The objects of $\mathscr{C}^\circ$ are those of $\mathscr{C}$, i.e., $\mathrm{Ob}(\mathscr{C}^\circ) = \mathrm{Ob}(\mathscr{C})$; for any objects X and Y we put $\mathrm{Hom}_{\mathscr{C}^\circ}(X, Y) = \mathrm{Hom}_{\mathscr{C}}(Y, X)$; if $f: X \to Y$ and $g: Y \to Z$ in $\mathscr{C}^\circ$ (i.e., $f: Y \to X$ and $g: Z \to Y$ in $\mathscr{C}$), then the composite $g \circ f$ in $\mathscr{C}^\circ$ is defined to be $f \circ g$ in $\mathscr{C}$. It is clear that $\mathscr{C}^\circ$ then satisfies the axioms of a category. Quite generally, given a proposition concerning objects and morphisms we can construct another proposition by reversing the directions of the morphisms, and we call the latter the dual proposition of the former. The dual proposition of a proposition in $\mathscr{C}$ coincides with a proposition in $\mathscr{C}^\circ$. For instance, a monomorphism (epimorphism) in $\mathscr{C}$ is an epimorphism (monomorphism) in $\mathscr{C}^\circ$, and the final (initial) object in $\mathscr{C}$ is the initial (final) object in $\mathscr{C}^\circ$. The product (coproduct) in $\mathscr{C}$ is the coproduct (product) in $\mathscr{C}^\circ$. Although the notion of the dual category is defined quite formally, it is useful in describing relations between specific categories. The dual category of (Ab), for instance, is equivalent to the category of commutative compact topological groups (†Pontryagin's duality theorem).

G. Categories over an Object

Fix a category $\mathscr{C}$ and an object $S \in \mathscr{C}$. A pair (X, f) of an object $X \in \mathscr{C}$ and a morphism $f: X \to S$ is called an object over S or an **S-object**, and f is called its **structure morphism**. We often omit f and simply say "an S-object X" if there is no danger of misunderstanding. If (X, f) and (Y, g) are S-objects, a morphism $h: X \to Y$ such that $f = g \circ h$ is called an **S-morphism** from (X, f) to (Y, g). The category whose objects are the S-objects and whose morphisms are the S-morphisms is called the **category of S-objects** in $\mathscr{C}$, and is denoted by $\mathscr{C}/S$. It has $(S, 1_S)$ as the final object. The product of two S-objects X and Y, taken in $\mathscr{C}/S$, is called the **fiber product** of X and Y **over S** (in $\mathscr{C}$), and is denoted by $X \times_S Y$ or $X \prod_S Y$. The dual notion of the fiber product is called **fiber sum** (or **amalgamated sum**). Thus for two morphisms $f: S \to X$ and $g: S \to Y$, the fiber product of X and Y over S in $\mathscr{C}^\circ/S$ is the fiber sum of X and Y (with respect to S); it is denoted by $X \coprod_s Y$.

Let $\mathscr{C}$ be the category of commutative rings and $K \in \mathscr{C}$. Then the family of K-objects in $\mathscr{C}^\circ$ is precisely the family of commutative K-algebras. The fiber product $A \times_K B$ in $\mathscr{C}^\circ$, i.e., $A \coprod_K B$ in $\mathscr{C}$, is the tensor product $A \otimes_K B$ of algebras.

H. Functors

Let $\mathscr{C}$ and $\mathscr{C}'$ be categories. A **covariant functor** F from $\mathscr{C}$ to $\mathscr{C}'$ is a rule which associates (1) with each object X in $\mathscr{C}$, an object $F(X)$ in $\mathscr{C}'$, and (2) with each morphism $f: X \to Y$ in $\mathscr{C}$, a morphism $F(f): F(X) \to F(Y)$ such that $F(g \circ f) = F(g) \circ F(f)$, $F(1_X) = 1_{F(X)}$. A **contravariant functor** is defined dually, by modifying this definition to $F(f): F(Y) \to F(X)$, $F(g \circ f) = F(f) \circ F(g)$. Thus a contravariant functor from $\mathscr{C}$ to $\mathscr{C}'$ is the same as a covariant functor from the dual category $\mathscr{C}^\circ$ to $\mathscr{C}'$ (or from $\mathscr{C}$ to $\mathscr{C}'^\circ$). **Functor** is a general term for both covariant functors and contravariant functors, but some authors use the word exclusively in the sense of a covariant functor. A functor in several variables is defined to be a functor from the product category of the categories in which the variables take their values.

A covariant functor $F: \mathscr{C} \to \mathscr{C}'$ is said to be **faithful (fully faithful)** if for any X, $Y \in \mathscr{C}$, the mapping $\mathrm{Hom}(X, Y) \to \mathrm{Hom}(F(X), F(Y))$ induced by F is injective (bijective), and similarly for contravariant functors. A faithful covariant functor $F: \mathscr{C} \to \mathscr{C}'$ which maps distinct objects of $\mathscr{C}$ to distinct objects of $\mathscr{C}'$ is called an **embedding**, and in this case $\mathscr{C}$ can be identified with a subcategory of $\mathscr{C}'$ by F. A fully faithful covariant functor $F: \mathscr{C} \to \mathscr{C}'$ is called an **equivalence** (between the categories) if it satisfies the condition that for any object X' of $\mathscr{C}'$, there exists an object X of $\mathscr{C}$ such that $F(X) \cong X'$. In this case we can consider the two categories essentially the same. A contravariant functor from $\mathscr{C}$ to $\mathscr{C}'$ which defines an equivalence from $\mathscr{C}^\circ$ to $\mathscr{C}'$ is called an **antiequivalence**.

I. Examples of Functors

(1) Let $\mathscr{C}$ be the category of groups (or rings). For any $X \in \mathscr{C}$ let $F(X)$ be the underlying set of X (i.e., the set obtained from X by "forgetting" its structure as a group or ring), and for any homomorphism f put $F(f) = f$. Then we get a faithful covariant functor (often called the **forgetful functor**) $F: \mathscr{C} \to (\mathrm{Sets})$.

(2) Let $\mathscr{C}$ be any category and fix an object X of $\mathscr{C}$. Then we get a covariant functor $h_X:\mathscr{C}\rightarrow$(Sets) as follows: With each $Y\in\mathscr{C}$ we associate the set $\mathrm{Hom}(X, Y)$, and with each morphism $f: Y\rightarrow Y'$ in $\mathscr{C}$ the mapping $f\circ$(where $f\circ:\mathrm{Hom}(X, Y)\rightarrow\mathrm{Hom}(X, Y')$ is defined by $(f\circ)(g)=f\circ g$). Similarly we define a contravariant functor $h^X:\mathscr{C}\rightarrow$(Sets) by $h^X(Y)=\mathrm{Hom}(Y,X)$ and $h^X(f)=\circ f$.

(3) Let $\rho: A\rightarrow B$ be a homomorphism of rings. With each left A-module M associate the †scalar extension $\rho^*(M)=B\otimes_A M$, and with each A-homomorphism f associate the B-homomorphism $\rho^*(f)=1_B\otimes f$. Then we get a covariant functor $\rho^*: {}_A\mathscr{M}\rightarrow{}_B\mathscr{M}$.

(4) Let R be a ring. With each left R-module M associate its dual module $M^*=\mathrm{Hom}_R(M, R)$, and to each R-linear mapping f associate its †dual mapping ${}^tf=\circ f$. Then we get a contravariant functor ${}_R\mathscr{M}\rightarrow\mathscr{M}_R$, and similarly for $\mathscr{M}_R\rightarrow{}_R\mathscr{M}$.

(5) For each differentiable manifold X let $F(X)$ denote the commutative ring of the differentiable functions on X, and for each differentiable mapping $f:X\rightarrow Y$ let $F(f)$ be the ring homomorphism $\circ f: F(Y)\rightarrow F(X)$. Then F is a faithful contravariant functor.

(6) Fix an Abelian group A. By associating with each topological space X the cohomology group $H(X,A)$ and with each continuous mapping $f:X\rightarrow Y$ the homomorphism $H(Y,A)\rightarrow H(X,A)$ induced by f, we obtain a contravariant functor from (Top) to (Ab).

(7) Fix a topological space X, and let $T(X)$ be the set of the open sets in X. Then $T(X)$ is ordered by inclusion, so it is a category (→ Section B, no. 7). The contravariant functors from $T(X)$ to (Ab) are precisely the †presheaves of Abelian groups over X. We can use any category instead of (Ab) to define a presheaf over X (→ 383 Sheaves).

J. Natural Transformations

Let $\mathscr{C}$ and $\mathscr{C}'$ be categories, and denote by $\mathrm{Hom}(\mathscr{C},\mathscr{C}')$ the collection of all covariant functors $\mathscr{C}\rightarrow\mathscr{C}'$. Let $F, G\in\mathrm{Hom}(\mathscr{C},\mathscr{C}')$. A **natural transformation** (or **functorial morphism**) from F to G is a function which assigns to each object X of $\mathscr{C}$ a morphism $\varphi(X): F(X)\rightarrow G(X)$ in $\mathscr{C}'$ such that for any morphism $f:X\rightarrow Y$ in $\mathscr{C}$, the equation $G(f)\circ\varphi(X)=\varphi(Y)\circ F(f)$ holds; in other words, the accompanying diagram is commutative:

$$\begin{array}{cccc} X & F(X) & \xrightarrow{\varphi(X)} & G(X) \\ {\scriptstyle f}\downarrow & {\scriptstyle F(f)}\downarrow & & \downarrow{\scriptstyle G(f)} \\ Y & F(Y) & \xrightarrow{\varphi(Y)} & G(Y) \end{array}$$

A natural transformation between contravariant functors is defined similarly. For instance, let A and B be Abelian groups and let $H^i(\cdot, A)$ and $H^j(\cdot, B)$ be the contravariant functors of †cohomology viewed as functors (Top)→(Sets). Then the natural transformations between them are the †cohomology operations.

Let $\varphi: F\rightarrow G$ be a natural transformation, and suppose that $\varphi(X): F(X)\rightarrow G(X)$ is an isomorphism for every $X\in\mathscr{C}$. Then the inverse transformation $G\rightarrow F$ of φ exists, and φ is called a **natural equivalence (functorial isomorphism** or **isomorphism)** and is written $\varphi: F\cong G$.

Suppose that $\mathrm{Ob}(\mathscr{C})$ is a set. Then the collection $\mathrm{Hom}(F, G)$ of all natural transformations $F\rightarrow G$ is also a set, and hence we can consider $\mathrm{Hom}(\mathscr{C},\mathscr{C}')$ a category in which the objects are the covariant functors $\mathscr{C}\rightarrow\mathscr{C}'$, the morphisms are the natural transformations, and the composition of morphisms is the natural one. Then $\mathrm{Hom}(\mathscr{C}^\circ,\mathscr{C}')$ is the category of contravariant functors from $\mathscr{C}$ to $\mathscr{C}'$. In particular, the category $\mathrm{Hom}(\mathscr{C}^\circ,\text{(Sets)})$ is sometimes denoted by $\hat{\mathscr{C}}$.

Given a category $\mathscr{C}$, a covariant (resp. contravariant) functor $F:\mathscr{C}\rightarrow$(Sets), and an object $X\in\mathscr{C}$, we can define a canonical bijection $\Phi_X:\mathrm{Hom}(h_X, F)\cong F(X)$ (resp. $\mathrm{Hom}(h^X, F)\cong F(X)$) by $\Phi_X(\varphi)=\varphi(X)1_X$. (The functors h_X and h^X were defined in Section I.) The inverse mapping of Φ_X assigns to $\xi\in F(X)$ the natural transformation $\varphi: h_X\rightarrow F$ defined by $\varphi(Y)u=F(u)\xi(Y\in\mathscr{C})$. In particular, if we take $F=h_Y(h^Y)$, we obtain a canonical bijection $\mathrm{Hom}(h_X, h_Y)\cong\mathrm{Hom}(Y,X)$ $(\mathrm{Hom}(h^X, h^Y)\cong\mathrm{Hom}(X, Y))$. It follows that there is a fully faithful contravariant (covariant) functor $\mathscr{C}\rightarrow\mathrm{Hom}(\mathscr{C},\text{(Sets)})$ $(\mathscr{C}\rightarrow\mathrm{Hom}(\mathscr{C}^\circ,\text{(Sets)})=\hat{\mathscr{C}})$ which associates $h_X(h^X)$ with $X\in\mathscr{C}$.

K. Adjoint Functors

Let $F:\mathscr{C}\rightarrow\mathscr{C}'$ and $F':\mathscr{C}'\rightarrow\mathscr{C}$ be covariant functors. Suppose that there is a rule which assigns to each pair of objects $M\in\mathscr{C}$ and $M'\in\mathscr{C}'$ a bijective mapping $\theta_{M,M'}:\mathrm{Hom}_{\mathscr{C}}(M, F'(M'))\cong\mathrm{Hom}_{\mathscr{C}'}(F(M), M')$ such that for any pair of morphisms $N\rightarrow M$ in $\mathscr{C}$ and $M'\rightarrow N'$ in $\mathscr{C}'$, the following diagram induced by the morphisms is commutative:

$$\begin{array}{ccc} \mathrm{Hom}_{\mathscr{C}}(M,F'(M')) & \xrightarrow{\theta_{M,M'}} & \mathrm{Hom}_{\mathscr{C}'}(F(M),M') \\ \downarrow & & \downarrow \\ \mathrm{Hom}_{\mathscr{C}}(N,F'(N')) & \xrightarrow{\theta_{N,N'}} & \mathrm{Hom}_{\mathscr{C}'}(F(N),N') \end{array}$$

Then we say that F is a **left adjoint functor** of F' and that F' is a **right adjoint functor** of F. We can regard $\mathrm{Hom}_{\mathscr{C}'}(F(M), M')$ as a functor from $\mathscr{C}\times\mathscr{C}'$ (contravariant in the variable

$M \in \mathscr{C}$, covariant in the variable $M' \in \mathscr{C}'$) to the category of sets, and similarly for $\mathrm{Hom}_{\mathscr{C}}(M, F'(M'))$. This commutativity of the diagram means that these two functors are isomorphic (→ Section J).

For instance, let A and B be rings and L a fixed B-A-†bimodule. Let $F: {}_A\mathscr{M} \to {}_B\mathscr{M}$ and $F': {}_B\mathscr{M} \to {}_A\mathscr{M}$ be the functors defined by

$$F(M) = L \otimes_A M, \quad F'(M') = \mathrm{Hom}_B(L, M'),$$

where the assignment for morphisms is defined in the natural way. Then F is the left adjoint of F' and F' is the right adjoint of F. In particular, let $\rho: A \to B$ be a homomorphism and consider the case $L = B$. Then F is the functor $\rho^*: {}_A\mathscr{M} \to {}_B\mathscr{M}$ and F' is the functor $\rho_*: {}_B\mathscr{M} \to {}_A\mathscr{M}$, so that ρ^* is the left adjoint of ρ_* (and ρ_* is the right adjoint of ρ^*) (→ 277 Modules, K, L; for more examples of adjoint functors → [11]).

L. Representation of Functors

We begin by discussing an example. Let T be a set, and consider the following problem: Is it possible to find a group X and a mapping $\xi: T \to X$ such that, for any group Y and for any mapping $\eta: T \to Y$, there exists a unique homomorphism $u: X \to Y$ with $u \circ \xi = \eta$? The answer is yes; it is enough to take the †free group X generated by T and the canonical injection $\xi: T \to X$ (Fig. 6). On the other hand, let $F(Y)$ be the set of all mappings $T \to Y$, and for each group homomorphism $f: Y \to Y'$ define the mapping $F(f): F(Y) \to F(Y')$ by $F(f)\eta = f \circ \eta$ ($\eta \in F(Y)$). Then we get a covariant functor from the category $\mathscr{C}$ of groups to the category of sets, $F: \mathscr{C} \to (\mathrm{Sets})$. We can now reformulate the condition on $X \in \mathscr{C}$ and $\xi \in F(X)$ as follows:

For any $Y \in \mathscr{C}$ and for any $\eta \in F(Y)$, there exists a unique morphism $u: X \to Y$ such that $F(u)\xi = \eta$.

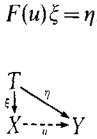


Fig. 6

Proceeding to the general case, let $\mathscr{C}$ be an arbitrary category and let $F: \mathscr{C} \to (\mathrm{Sets})$ be a functor. If there exist an object X of $\mathscr{C}$ and an element ξ of $F(X)$ satisfying the condition just stated (with the modification $u: Y \to X$ in the contravariant case), then we say that the pair (X, ξ) **represents** the functor F, or less specifically, that X represents F, and we call ξ the **canonical element** of $F(X)$. We also say that F is **representable**. The condition stated above is a formulation of the so-called **universal mapping property**. If (X', ξ') also represents F, the unique morphism $u: X \to X'$ (or $X' \to X$) with $F(u)\xi = \xi'$ is necessarily an isomorphism.

When (X, ξ) represents F, the natural transformation $\varphi: h_X \to F$ ($h^X \to F$ in the contravariant case) which corresponds to ξ by the canonical bijection $\Phi_X: \mathrm{Hom}(h_X, F) \cong F(X)$ is an isomorphism. Conversely, if there is a functorial isomorphism $\varphi: h_X \to F$ (or $h^X \to F$) for some $X \in \mathscr{C}$, then the object X represents F, with the canonical element of $F(X)$ the element which corresponds to φ by the canonical bijection Φ_X, i.e., $\xi = \varphi(X)1_X$.

We have already seen the example of a free group; here we list a few more examples. (1) Let $\{X_i\}_{i \in I}$ be a family of objects in a category $\mathscr{C}$. For each $Y \in \mathscr{C}$ we put $F(Y) = \prod_{i \in I} \mathrm{Hom}(Y, X_i)$, and for each morphism $f: Y \to Y'$ we define the mapping $F(f): F(Y') \to F(Y)$ by $F(f)(f_i) = (f_i \circ f)$. Then we get a contravariant functor $F: \mathscr{C} \to (\mathrm{Sets})$. A pair (X, ξ) which represents F (where $\xi \in F(X) = \prod_{i \in I} \mathrm{Hom}(X, X_i)$) is the product of $\{X_i\}$. Thus representability of F is equivalent to the existence of the product of $\{X_i\}$, and similarly for the coproduct.

(2) Let R be a ring, M a right R-module, and N a left R-module. For each Abelian group Y let $F(Y)$ denote the set of the R-balanced mappings $M \times N \to Y$ (→ 277 Modules J). Since a homomorphism $f: Y \to Y'$ induces a natural mapping $F(f): F(Y) \to F(Y')$ by composition, we obtain a covariant functor $F: (\mathrm{Ab}) \to (\mathrm{Sets})$. This functor is representable: the pair consisting of the tensor product $M \otimes_R N$ and the canonical mapping $M \times N \to M \otimes_R N$ represent it.

(3) Let R be a commutative ring and S a subset of R. For each commutative ring Y, let $F(Y)$ denote the set of homomorphisms $R \to Y$ that map the elements of S to invertible elements of Y. As in the preceding example, we obtain a covariant functor $F: (\mathrm{Rings}) \to (\mathrm{Sets})$. This functor is represented by the †ring of quotients $S^{-1}R$ and the canonical homomorphism $R \to S^{-1}R$.

M. Groups in a Category

Let $\mathscr{C}$ be a category with a final object e, and assume that a finite product always exists in $\mathscr{C}$. If an object $G \in \mathscr{C}$ and morphisms $\alpha: G \times G \to G$, $\beta: G \to G$, $\varepsilon: e \to G$ are given such that the the diagrams of Fig. 7 are commutative, then $(G, \alpha, \beta, \varepsilon)$ is called a group in $\mathscr{C}$ (**group object** in $\mathscr{C}$ or $\mathscr{C}$**-group**).

If $\mathscr{C}$ is the category of sets, then α defines a law of composition in the set G, and the image of e by ε is the identity element and $\beta(x)$ is the inverse of x, so that G is an ordinary group. If $\mathscr{C}$ is the category of topological spaces (analy-

tic manifolds, algebraic varieties, †schemes) then G is a †topological group (†Lie group, †algebraic group, †group scheme).

We can also define the $\mathscr{C}$-group by lifting the group concept in (Sets) to the category $\mathscr{C}$ by means of the functor h^X. Namely, let G be an object of $\mathscr{C}$, and suppose that for each $Y \in \mathscr{C}$ the set $h^G(Y) = \mathrm{Hom}(Y, G)$ is equipped with a group structure and that for each morphism $f: Y \rightarrow Y'$ the induced mapping $h^G(Y') \rightarrow h^G(Y)$ is a group homomorphism. In other words, suppose that h^G is a contravariant functor from $\mathscr{C}$ to the category of groups. Then the object G with the additional structure on h^G is called a $\mathscr{C}$-group. This definition is equivalent to the one given above.

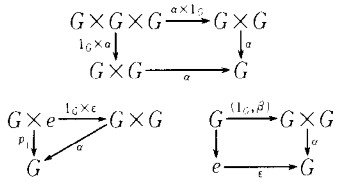


Fig. 7

N. Additive Categories

A category $\mathscr{C}$ is called an **additive category** if for each pair X, $Y \in \mathscr{C}$, the set of morphisms $\mathrm{Hom}(X, Y)$ has the structure of an additive group such that (1) the composition of morphisms is distributive in both ways: $h \circ (f+g) = h \circ f + h \circ g$, $(f+g) \circ h = f \circ h + g \circ h$; (2) there exists an object $0'$ with $\mathrm{Hom}(0', 0') = \{0\}$; (3) the product (or the coproduct) of any two objects exists. Then the object $0'$ in (2) is a final and initial object, and is called the **zero object**. Both the product and the coproduct of any two objects exist and can be identified. The dual category of an additive category is also an additive category. A functor F from an additive category to another is called an **additive functor** if $F(f+g) = F(f) + F(g)$ holds for morphisms. In an additive category $\mathscr{C}$, $\mathrm{Hom}(X, Y)$ is an additive functor from $\mathscr{C}$ to (Ab) in each variable.

For any ring R, the category of left (or right) R-modules is an additive category. The following definitions are generalizations of the corresponding concepts in the theory of modules. The **kernel** of a morphism $f: A \rightarrow B$ is a pair consisting of an object A' and a monomorphism $i: A' \rightarrow A$ with $f \circ i = 0$, such that any morphism $u: X \rightarrow A$ with $f \circ u = 0$ is divisible by i (that is, $u = i \circ v$ for some $v: X \rightarrow A'$). Dually, the **cokernel** of f is a pair consisting of an object B' and an epimorphism $j: B \rightarrow B'$ with $j \circ f = 0$ which divides any morphism $u: B \rightarrow X$ with $u \circ f = 0$. We write $A' = \mathrm{Ker}\, f$, $B' = \mathrm{Coker}\, f$. The kernel of $j: B \rightarrow \mathrm{Coker}\, f$ is called the **image** of f and is denoted by $\mathrm{Im}\, f$; the cokernel of $i: \mathrm{Ker}\, f \rightarrow A$ is called the **coimage** of f and is denoted by $\mathrm{Coim}\, f$. If all these exist, it follows from the definitions that there is a unique morphism $\mathrm{Coim}\, f \rightarrow \mathrm{Im}\, f$ such that the composite of $A \rightarrow \mathrm{Coim}\, f \rightarrow \mathrm{Im}\, f \rightarrow B$ is equala to f.

An additive category $\mathscr{C}$ is called an **Abelian category** if it satisfies the following conditions: (1) every morphism has a kernel and a cokernel, (2) for every morphism f, the morphism $\mathrm{Coim}\, f \rightarrow \mathrm{Im}\, f$ just mentioned is an isomorphism. The dual category of an Abelian category is also Abelian. The categories of Abelian groups, of R-modules, and of sheaves of $\mathcal{O}$-modules on a †ringed space $(X, \mathcal{O})$ are important examples of Abelian categories. Many propositions which are valaid in (Ab) remain valid in any Abelian category. In particular, the notion of an †exact sequence is defined in an Abelian category in the same way as in (Ab), and the fiber product and fiber sum of a finite number of objects always exist in an Abelian category. A functor between Abelian categories which carries exact sequences into exact sequences is called an **exact functor**; (such a functor is automatically additive). If $\mathscr{C}$ is a category of which $\mathrm{Ob}(\mathscr{C})$ is a set and if $\mathscr{C}'$ is an Abelian category, then $\mathrm{Hom}(\mathscr{C}, \mathscr{C}')$ is an Abelian category. Given an Abelian category $\mathscr{C}$ and a subcategory $\mathscr{C}'$ which satisfies certain conditions, one can construct an Abelian category $\mathscr{C}/\mathscr{C}'$ which is called the **quotient category** (Serre's theory of classes of Abelian groups; → [8]).

If $\mathscr{C}$ is an Abelian category of which $\mathrm{Ob}(\mathscr{C})$ is a set, there is an embedding of $\mathscr{C}$ into the category ${}_R\mathscr{M}$ of modules over some ring R by a fully faithful flat exact covariant functor (**full embedding theorem**, B. Mitchell, *Amer. J. Math.*, 86 (1964)). This remarkable theorem enables us to extend results obtained for modules to the case of Abelian categories.

The notions of category and functor were introduced in [7] and were applied first in topology and then in homological algebra and algebraic geometry (→ 200 Homological Algebra).

References

[1] S. Eilenberg and N. Steenrod, Foundations of algebraic topology, Princeton Univ. Press, 1952.
[2] H. Cartan and S. Eilenberg, Homological algebra, Princeton Univ. Press, 1956.
[3] S. MacLane, Homology, Springer, 1963.

[4] P. Freyd, Abelian categories, Harper & Row, 1964.
[5] B. Mitchell, The theory of categories, Academic Press, 1965.
[6] I. Bucur and A. Deleanu, Introduction to the theory of categories and functors, Wiley-Interscience, 1968.
[7] S. Eilenberg and S. MacLane, General theory of natural equivalences, Trans. Amer. Math. Soc., 58 (1945), 231–294.
[8] A. Grothendieck, Sur quelques points d'algèbre homologique, Tôhoku Math. J., (2) 9 (1957), 119–221.
[9] P. Gabriel, Des catégories abéliennes, Thesis, Paris, 1962.
[10] N. Yoneda, On universality I, II (in Japanese), Sûgaku, 13 (1961–1962), 109–112; 14 (1962–1963), 39–43.
[11] S. MacLane, Categorical algebra, Bull. Amer. Math. Soc., 71 (1965), 40–106.

53 (XXI.17) Cauchy, Augustin Louis

The French mathematician Augustin Louis Cauchy (August 21, 1789–May 25, 1857) graduated from the Ecole Polytechnique in 1807 and from the Ecole des Ponts et Chaussées in 1810, to become a civil engineer. In 1816, his mathematical works were recognized, and he was appointed a member of the Académie des Sciences while a professor at the Ecole Polytechnique. After the July revolution in 1830, he refused to pledge loyalty to Louis-Philippe and fled to Turin; he later moved to Prague. He returned to France after the revolution of 1848 and became a professor at the University of Paris, where he remained until his death. He was a Catholic and a Royalist all his life.

His scientific contributions were numerous and covered many fields. In algebra, he did pioneering work in †determinants and in the theory of †groups. He also made notable achievements in theoretical physics, optics, and the theory of elasticity. His main field was analysis. He was interested in making analysis rigorous by giving calculus a solid foundation in such works as *Cours d'analyse de l'Ecole Polytechnique* (1821). In his paper "Memoire sur les intégrales définies prises entre les limites imaginaires" (1825), he proved the main theorem of the theory of functions of a complex variable. Another important work is his proof of the existence theorem for the solutions of †differential equations in the cases of real and complex variables

References

[1] A. Cauchy, Oeuvres complètes I.1–12; II.1–14, Gauthier-Villars, 1882–1958.
[2] F. Klein, Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert I, Springer, 1926 (Chelsea, 1956).

54 (III.21) Cayley Algebras

Let Q be a †quaternion algebra over a field K of characteristic zero. A **general Cayley algebra** $\mathscr{C}$ is a 2-dimensional Q-†module $Q+Qe$ with the multiplication $(q+re)(s+te)=(qs+\gamma\bar{t}r)+(tq+r\bar{s})e$, where $q, r, s, t\in Q$, γ is a given element in K, and $\bar{t}, \bar{s}$ are the †conjugate quaternions of t, s, respectively. The elements of $\mathscr{C}$ are called **Cayley numbers**; $\mathscr{C}$ is a nonassociative, †alternative algebra of dimension 8 over K (→ 231 Jordan Algebras). The map $a=q+re\rightarrow\bar{a}=\bar{q}-re$ is an †antiautomorphism of $\mathscr{C}$. Define two maps $\mathscr{C}\rightarrow K$ by $N(a)=a\bar{a}=\bar{a}a$ (**norm** of a) and $T(a)=a+\bar{a}$ (**trace** of a). Then every a in $\mathscr{C}$ satisfies the equation $x^2-T(a)x+N(a)=0$. Furthermore, $N(ab)=N(a)N(b)$ for a, b in $\mathscr{C}$. The †quadratic form $N(x)=T(x\bar{x})/2$ characterizes $\mathscr{C}$. In particular, any two (nonassociative) general Cayley algebras over the same field K which are not †alternative fields are isomorphic.

In order for $\mathscr{C}$ to be an alternative field, either of the following two conditions is necessary and sufficient: (i) $N(a)=0$ implies $a=0$; (ii) Q is a noncommutative division algebra and γ cannot be expressed in the form $\sigma^2-\lambda\xi^2-\mu\eta^2+\lambda\mu\zeta^2$ $(\sigma, \xi, \eta, \zeta\in K)$. (For the meaning of λ, μ with respect to Q → 29 Associative Algebras D.) Every alternative field of finite dimension is a general Cayley algebra.

In particular, when Q is the †quaternion field over the real number field with $\lambda=\mu=-1$, the general Cayley algebra over Q with $\gamma=-1$ is called the **Cayley algebra**. When K is an †algebraic number field of finite degree, there are only a finite number of nonisomorphic general Cayley algebras over K.

The Lie algebra $\mathfrak{D}(\mathscr{C})$ of all †derivations of a general Cayley algebra $\mathscr{C}$ is a †simple Lie algebra of type G_2. If K is the real number field, the identity component of the group of all †automorphisms of the Cayley algebra $\mathscr{C}$ is a compact simply connected †simple Lie group of type G_2. The Cayley algebra $\mathscr{C}$ is the unique alternative field over the real number field K. This last fact is important because of the following proposition: In the theory of †non-

Desarguesian projective planes, the field which gives rise to the coordinates is an alternative field. Let $\mathfrak{P}P_2$ be the set of all 3×3 †Hermitian matrices A over the Cayley algebra $\mathscr{C}$ such that $\mathrm{tr}\, A = 1$, $A^2 = A$. Then we can define a structure of a projective plane on $\mathfrak{P}P_2$, which with this structure is called the **Cayley projective plane**. Furthermore, let $\mathfrak{J}$ be the set of all 3×3 Hermitian matrices over $\mathscr{C}$, with a multiplication in $\mathfrak{J}$ defined by $A \cdot B = (1/2)(BA + AB)$. The identity component G of the group of all automorphisms of $\mathfrak{J}$ is a compact simply connected simple Lie group of type F_4. This group G acts on $\mathfrak{P}P_2$ transitively, and $\mathfrak{P}P_2 = G/Spin(9)$ (→ 249 Lie Groups; Appendix A, Table 5.III).

References

[1] L. E. Dickson, Algebren und ihre Zahlentheorie, Orell Füssli, 1927.
[2] M. Zorn, Alternativkörper und quadratische Systeme, Abh. Math. Sem. Univ. Hamburg, 9 (1933), 395–402.
[3] N. Jacobson, Cayley numbers and normal simple Lie algebras of type G, Duke Math. J., 5 (1939), 775–783.

55 (XX.6) Celestial Mechanics

A. General Remarks

The motions of planets, comets, the moon, and satellites in our solar system are the main topics in **celestial mechanics** (→ 392 Spherical Astronomy). However, studies in this subject can also include motions of fixed and binary stars in our galaxy, equilibrium figures of celestial bodies, and rotational motions of the earth and the moon.

Celestial mechanics is usually based on †Newtonian mechanics, with the effects of †general relativity sometimes taken into account to determine corrections to the orbits of celestial bodies. Therefore the main task of celestial mechanics is to solve differential equations of motion based on Newtonian mechanics. However, since the equations for the problem of $n > 2$ bodies cannot be solved rigorously (→ 420 Three-Body Problem), appropriate methods have been devised whereby we may obtain approximate solutions of the equations with accuracy comparable to that of observations.

The **two-body problem**, which concerns the behavior of two celestial bodies regarded as points exerting mutual interactions, can be reduced to a one-body problem with reference to a central force, since integrals of motion of the center of gravity for the system exist. The †Hamilton-Jacobi equation for the one-body problem is of †separable type and can be solved completely. The orbit for the two-body problem is a †conic with one of its †foci at the center of gravity. The majority of celestial bodies in the solar system actually perform **elliptic motions**. †Kepler's orbital elements for elliptic motion are functions of the integration constants in the solution of the Hamilton-Jacobi equation and are determined by the initial conditions.

B. Perturbations

In studying the †n-body problem, we first solve certain two-body problems and then apply the method of †perturbations, i.e., the **method of variation of constants**, in order to obtain solutions developed as †power series of small parameters. The parameters are ratios of the masses of planets to that of the sun for planetary motions and the ratio of the geocentric lunar distance to the solar distance for lunar motion.

Electronic computers have made it possible to compute planetary coordinates for long intervals of time by solving numerically differential equations of motion including all possible interactions. However, in discussing the stability of the solar system, analytic methods are more effective, particularly the method of obtaining **secular perturbations** by eliminating short-periodic terms by canonical transformations. This is one of the averaging methods of solving differential equations. However, as the solutions obtained by means of perturbation methods are not always convergent, most important problems related to the stability of motion have not yet been solved rigorously. Secular perturbations for planetary motions can be derived by solving differential equations that are linearized by neglecting cubic powers of orbital eccentricities and inclinations to the ecliptic (→ 309 Orbit Determination), which are small quantities. The †eigenvalues of these linear differential equations correspond to mean angular velocities of the perihelion and the ascending node. The equations for the eigenvalues are called **secular equations**.

C. Artificial Satellites

To discuss motions of artificial satellites close to the earth, the latter cannot be regarded as a

point or as a sphere but must be assumed to be an oblate spheroid, i.e., an †ellipsoid of revolution. The effects of oblateness on the motion of satellites can be derived as perturbations of the theoretical elliptic motions obtained as the solutions of this two-body problem under the assumption that the earth is spherical. Also, by utilizing a special potential very close to the geopotential, we can find a Hamilton-Jacobi equation of separable type which is solvable. This special potential appears in the problem of two fixed centers with equal masses situated on an imaginary axis. When the geopotential is assumed to be axially symmetric, the equations of motion for the satellites have two degrees of freedom; therefore there appear two fundamental frequencies related to the special potential. When these two frequencies are equal, the problem is called a **critical inclination problem** and is important from the mathematical point of view. Theories for satellites can be applied to the motions of fixed stars in the galaxy.

D. Equilibrium Figures

There is a large literature concerning **equilibrium figures** and the stabilities of celestial bodies assumed to consist of spinning fluids. The two-body problem with tidal interactions is particularly important; problems concerning the evolution of the earth-moon system are special cases of such a problem.

The theory of rotation of the earth as it is affected by †precession, †nutation, and latitude variations is also a part of celestial mechanics; for this theory, elastic theory and geophysics are applied.

For the n-body problem → 420 Three-Body Problem.

References

[1] F. F. Tisserand, Traité de méchanique céleste 1–4, Gauthier-Villars, 1889–1896.
[2] H. Poincaré, Les méthodes nouvelles de la méchanique céleste I, II, III, Gauthier-Villars, 1892–1899.
[3] F. L. Moulton, An introduction to celestial mechanics, Macmillan, 1914.
[4] A. Wintner, The analytical foundation of celestial mechanics, Princeton Univ. Press, 1941.
[5] W. M. Smart, Celestial mechanics, Longmans-Green, 1953.
[6] D. Brouwer and G. M. Clemence, Methods of celestial mechanics, Academic Press, 1961.
[7] R. Deutsch, Orbital dynamics of space vehicles, Prentice-Hall, 1963.
[8] Y. Hagihara, Celestial mechanics. I, Dynamical principles and transformation theory, MIT Press, 1970; II, Perturbation theory, pts. 1, 2, MIT, 1972; III, Differential equations in celestial mechanics, pts. 1, 2, Japan Society for the promotion of Science, 1974; IV, Periodic and quadi-periodic solutions, pts. 1, 2, Japan Society, 1975; V, Topology of the three-body problem, pts. 1, 2, Japan Society, 1976.
[9] C. L. Siegel and J. K. Moser, Lectures on celestial mechanics, Springer, 1971.
[10] E. L. Stiefel and G. Scheifele, Linear and regular celestial mechanics, Springer, 1971.

56 (IX.14) Characteristic Classes

A. General Remarks

The theory of characteristic classes arose from the problem of whether or not there exists a †tangent r-frame field on a †differentiable manifold (E. Stiefel [5]) (→ 105 Differentiable Manifolds). The importance of the characteristic class as a fundamental invariant of †vector bundle structure is now fully recognized (→ 114 Differential Topology, 147 Fiber Bundles, 427 Topology of Lie Groups and Homogeneous Spaces).

B. Stiefel-Whitney Classes

Let $\xi=(E,B,\mathbf{R}^n)$ be an n-dimensional real †vector bundle (called an $\mathbf{R}^n$-bundle) with †paracompact Hausdorff †base space B, fiber $\mathbf{R}^n$, and †orthogonal group $O(n)$ as the †structure group. Then the element $w_i(\xi)$ of the i-dimensional †cohomology group $H^i(B;\mathbf{Z}_2)$ $(i=1,2,\ldots,n)$ of the base space B with coefficient in $\mathbf{Z}_2=\mathbf{Z}/2\mathbf{Z}$, called the i-dimensional (or ith) **Stiefel-Whitney class**, and the element $w(\xi)=1+w_1(\xi)+\ldots+w_n(\xi)$ of the †cohomology ring $H^*(B;\mathbf{Z}_2)$, called the **total Stiefel-Whitney class**, are defined as follows. First, we deal with the case $n=1$. We call the **infinite-dimensional real projective space** the †inductive limit $\mathbf{P}^\infty(\mathbf{R})=\varinjlim \mathbf{P}^n(\mathbf{R})$ of the finite-dimensional real †projective space $\mathbf{P}^n(\mathbf{R})$. The nontrivial †line bundle γ_1 over the infinite-dimensional real projective space $\mathbf{P}^\infty(\mathbf{R})$ is a universal $\mathbf{R}^1$-bundle (†universal bundle for the orthogonal group $O(1)$); therefore any $\mathbf{R}^1$-bundle $\xi=(E,B,\mathbf{R}^1)$ is equivalent to an †induced bundle from the universal bundle γ_1 by a †classifying mapping $f_\xi: B\rightarrow \mathbf{P}^\infty(\mathbf{R})$: $\xi\equiv f_\xi^*\gamma_1$. We define the 1-dimensional universal Stiefel-Whitney class $w_1(\gamma_1)$ to be the generator of $H^1(\mathbf{P}^\infty(\mathbf{R});\mathbf{Z}_2)$ and set $w_1(\xi)=$

$f_\xi^* w_1(\gamma_1)$. For general n, we consider the principal $O(n)$-bundle $(P, B, O(n))$ †associated with the given $\mathbf{R}^n$-bundle $\xi=(E, B, \mathbf{R}^n)$. Let Q_n be the subgroup of $O(n)$ consisting of all diagonal matrices. Then the orbit space $Y=P/Q_n$ is the base space of the †principal Q_n-bundle $\eta=(P, P/Q_n, Q_n)$. Let $\rho: Y \to B=P/O(n)$ be the natural projection. Then the $\mathbf{R}^n$-bundle $\rho^*\xi$ over Y induced by ρ is associated with η and is equivalent to the †Whitney sum of n line bundles (F. Hirzebruch [4]), $\rho^*\xi \equiv \xi_1 \oplus \ldots \oplus \xi_n$. Moreover, the homomorphism $\rho^*: H^*(B; \mathbf{Z}_2) \to H^*(Y, \mathbf{Z}_2)$ is injective (A. Borel [1]). Therefore we can uniquely define the total Stiefel-Whitney class $w(\xi)$ of the $\mathbf{R}^n$-bundle by the relation $\rho^* w(\xi)=w(\xi_1) \ldots w(\xi_n)$. The Stiefel-Whitney classes defined above are compatible with bundle mappings $f: w(f^*\xi)=f^*w(\xi)$ (→ 147 Fiber Bundles M). The Stiefel-Whitney class $w_i(\gamma_n) \in H^i(B_{O(n)}; \mathbf{Z}_2)$ $(1 \leqslant i \leqslant n)$ of the universal $\mathbf{R}^n$-bundle γ_n over the †classifying space $B_{O(n)}$ is called the i-dimensional **universal Stiefel-Whitney class**. For the Whitney sum of vector bundles, we have $w(\xi \oplus \eta)=w(\xi)w(\eta)$.

In order for an $\mathbf{R}^n$-bundle ξ to be †orientable, namely, for ξ to have an $SO(n)$-structure, it is necessary and sufficient that $w_1(\xi)=0$. For an oriented $\mathbf{R}^n$-bundle ξ, the **Euler-Poincaré class** $X_n(\xi)$ is defined to be the †primary obstruction $X_n(\xi) \in H^n(B; \mathbf{Z})$ for constructing a †cross section of the †associated $(n-1)$-sphere bundle. In particular, the Euler-Poincaré class of the universal bundle for $SO(n)$ is called the **universal Euler-Poincaré class**. $X_n(\xi) \bmod 2$ is equal to $w_n(\xi)$. If n is odd, we have $2X_n(\xi)=0$.

C. Chern Classes

We consider an n-dimensional complex vector bundle $\omega=(E, B, \mathbf{C}^n)$ (called $\mathbf{C}^n$-bundle in the following) with a paracompact Hausdorff base space B, fiber $\mathbf{C}^n$, and unitary group $U(n)$ as the structure group. The cohomology class $c_i(\omega) \in H^{2i}(B; \mathbf{Z})$ $(i=1,2,\ldots,n)$, called the $2i$-dimensional (or ith) **Chern class**, and the **total Chern class** $c(\omega)=1+c_1(\omega)+\ldots+c_n(\omega) \in H^*(B; \mathbf{Z})$ are defined as follows. In the case $n=1$, let $\mathbf{C}^\infty=\varinjlim \mathbf{C}^n$ (the inductive limit of the †complex Euclidean spaces $\mathbf{C}^n$), S^∞ be its unit sphere, and $\mathbf{P}^\infty(\mathbf{C})$ be the **infinite-dimensional complex projective space** consisting of all complex lines through the orgin O of $\mathbf{C}^\infty$. Then the natural mapping $S^\infty \to \mathbf{P}^\infty(\mathbf{C})$ defines a universal principal $U(1)$-bundle $(S^\infty, \mathbf{P}^\infty(\mathbf{C}), U(1))$. Let γ_1 be its associated universal $\mathbf{C}^1$-bundle. Then we define the 1-dimensional universal Chern class $c_1(\gamma_1) \in H^2(\mathbf{P}^\infty(\mathbf{C}); \mathbf{Z})$ to be the cohomology class that takes the value -1 on the cycle $S^2(\approx \mathbf{P}^1(\mathbf{C}) \subset \mathbf{P}^\infty(\mathbf{C}))$ with the natural orientation. Since a general $\mathbf{C}^1$-bundle $\xi=(E, B, \mathbf{C}^1)$ is induced from γ_1 by a classifying mapping $f_\xi: B \to \mathbf{P}^\infty(\mathbf{C})$, we set $c_1(\xi)=f_\xi^* c_1(\gamma_1)$. When $n>1$, let $(P, B, U(n))$ be the principal $U(n)$-bundle associated with the given $\mathbf{C}^n$-bundle $\xi=(E, B, \mathbf{C}^n)$. Let T_n be the subgroup of $U(n)$ consisting of all diagonal matrices (which is a †maximal torus of $U(n)$). Then the quotient space $Y=P/T_n$ is the base space of the principal T_n-bundle $\eta=(P, P/T_n, T_n)$. Let $\rho: Y \to B=P/U(n)$ be the natural projection. Then the $\mathbf{C}^n$-bundle $\rho^*\xi$ over Y is associated with η and is equivalent to the †Whitney sum of n complex line bundles: $\rho^*\xi \equiv \xi_1 \oplus \ldots \oplus \xi_n$. Moreover, $\rho^*: H^*(B; \mathbf{Z}) \to H^*(Y; \mathbf{Z})$ is a monomorphism (Borel [1], Hirzebruch [4]). Therefore we can uniquely define the total Chern class $c(\xi)$ of the $\mathbf{C}^n$-bundle ξ by the relation $\rho^* c(\xi)=c(\xi_1) \ldots c(\xi_n)$. The Chern classes, as defined above, are compatible with bundle mappings f (i.e., $c(f^*\xi)=f^*c(\xi)$) (→ 147 Fiber Bundles N). The Chern class $c_i(\gamma_n) \in H^{2i}(B_{U(n)}; \mathbf{Z})$ $(1 \leqslant i \leqslant n)$ of the universal $\mathbf{C}^n$-bundle γ_n over $B_{U(n)}$ is called the $2i$-dimensional **universal Chern class**. Let ξ, η be complex vector bundles over B. Then we have $c(\xi \oplus \eta)=c(\xi)c(\eta)$. By the natural inclusion $U(n) \subset SO(2n)$, we can identify a $\mathbf{C}^n$-bundle ω with an oriented $\mathbf{R}^{2n}$-bundle $\omega_{\mathbf{R}}$. Then we have $c_i(\omega) \bmod 2=w_{2i}(\omega_{\mathbf{R}})$, $w_{2i+1}(\omega_{\mathbf{R}})=0$ $(i=0,1,\ldots,n)$, $c_n(\omega)=X_{2n}(\omega_{\mathbf{R}})$.

Examples. Let $(S^{2n+1}, \mathbf{P}^n(\mathbf{C}), U(1))$ be the †Hopf bundle, and let γ_1^n be the associated complex line bundle. Then the classifying mapping of γ_1^n is the natural inclusion $\mathbf{P}^n(\mathbf{C}) \to BU(1)=\mathbf{P}^\infty(\mathbf{C})$, and $c(\gamma_1^n)=1-g_n$, where $g_n \in H^2(\mathbf{P}^n(\mathbf{C}); \mathbf{Z})$ is the cohomology class dual to the homology class represented by the hyperplane $\mathbf{P}^{n-1}(\mathbf{C})$. On the other hand, for the complex line bundle $\xi_1^n=\{\mathbf{P}^{n-1}(\mathbf{C})\}$ determined by the †divisor (→ 72 Complex Manifolds) $\mathbf{P}^{n-1}(\mathbf{C}) \subset \mathbf{P}^n(\mathbf{C})$, we have $c(\xi_1^n)=1+g_n$, and ξ_1^n, γ_1^n are dual to each other. ξ_1^n is called the **canonical line bundle** over the complex projective space. Moreover, the Whitney sum $\tau \oplus \varepsilon_1$ of the tangent bundle $\tau(\mathbf{P}^n(\mathbf{C}))$ of the n-dimensional complex projective space $\mathbf{P}^n(\mathbf{C})$ and the trivial $\mathbf{C}^1$-bundle ε_1 is equivalent to the Whitney sum of $(n+1)$ copies of ξ_1^n. Therefore we have $c(\tau(\mathbf{P}^n(\mathbf{C})))=c(\xi_1^n \oplus \xi_1^n \oplus \ldots \oplus \xi_1^n)=(1+g_n)^{n+1}$.

D. Pontryagin Classes

Utilizing the inclusion mapping $O(n) \subset U(n)$ of structure groups, we can associate a $\mathbf{C}^n$-bundle $\xi_{\mathbf{C}}=\xi \oplus \sqrt{-1}\,\xi$ with an $\mathbf{R}^n$-bundle ξ. We define the $4i$-dimensional (or ith) **Pontryagin class** of the $\mathbf{R}^n$-bundle ξ by $p_i(\xi)=(-1)^i c_{2i}(\xi_{\mathbf{C}})$ $(\in H^{4i}(B; \mathbf{Z}))$ $(i=1,2,\ldots,[n/2])$ (Hirzebruch

[4]). In particular, the Pontryagin classes of the universal bundle for $O(n)$ are called the **universal Pontryagin classes**. We also define the **total Pontryagin classes** $p(\xi)=1+p_1(\xi)+\ldots+p_{[n/2]}(\xi)$. We have $2c_{2i+1}(\xi_{\mathbf{C}})=0$. For $\mathbf{R}^n$-bundles ξ and η, $p(\xi\oplus\eta)\equiv p(\xi)p(\eta)$ modulo 2-torsion elements. We have $p_i(\xi)\pmod 2=(w_{2i}(\xi))^2$ and $p_n(\xi)=(X_{2n}(\xi))^2$ for the oriented $\mathbf{R}^{2n}$-bundle ξ. Moreover, for a complex vector bundle ω, we have (Wu [8])

$$(-1)^k p_k(\omega_{\mathbf{R}})=\sum_{i=0}^{2k}(-1)^i c_i(\omega)c_{2k-i}(\omega),$$

where we put $c_0(\omega)=1$.

All such classes as defined in Sections B–D are called **characteristic classes**.

E. Other Definitions of Characteristic Classes

Axiomatic Definition. (1) For a $\mathbf{C}^n$-bundle ξ over a paracompact Hausdorff base space B, Chern classes $c_i(\xi)\in H^{2i}(B;\mathbf{Z})$ $(i\geqslant 0)$ are defined, and we have $c_0(\xi)=1$, $c_i(\xi)=0$ $(i>n)$. (2) For the total Chern class $c(\xi)=\sum_{i=0}^{\infty}c_i(\xi)$, we have $c(f^*\xi)=f^*c(\xi)$ for each bundle mapping f. (3) For the Whitney sum, we have $c(\xi\oplus\eta)=c(\xi)\cdot c(\eta)$. (4) Normalization condition: For the canonical line bundle ξ_1^n, we have $c(\xi_1^n)=1+g_n$ (→ Section C). We can verify the existence and uniqueness of $c_i(\xi)$ satisfying these four conditions, so Chern classes can be defined axiomatically by these conditions (Hirzebruch [4]). We can similarly defined Stiefel-Whitney classes axiomatically.

Definition in Terms of Obstruction Classes. When the base space is a †CW complex, we can define the Chern class of a $\mathbf{C}^n$-bundle ξ over B as follows: Let $V_{n,n-q+1}(\mathbf{C})=U(n)/I_{n-q+1}\times U(q-1)$ be the †complex Stiefel manifold of all orthonormal $(n-q+1)$-frames in $\mathbf{C}^n$ with Hermitian metric. Then $V_{n,n-q+1}(\mathbf{C})$ is †$(2q-2)$-connected, and its $(2q-1)$-dimensional †homotopy group $\pi_{2q-1}(V_{n,n-q+1}(\mathbf{C}))=\mathbf{Z}$. Let ξ' be the †associated bundle of ξ with fiber $V_{n,n-q+1}(\mathbf{C})$. Then the †primary obstruction $(\in H^{2q}(B;\mathbf{Z}))$ to constructing a †cross section of ξ' coincides with the Chern class $c_q(\xi)$. Analogously, we can interpret $w_q(\xi)$ for an $\mathbf{R}^n$-bundle ξ as an obstruction class (→ 147 Fiber Bundles).

Definition in Terms of Schubert Varieties. We denote by $\mathbf{C}^k$ the subspace defined by $z_{k+1}=z_{k+2}=\ldots=z_{n+N}=0$ of the space $\mathbf{C}^{n+N}=\{(z_1,\ldots,z_{n+N})\mid z_i\in\mathbf{C}, i=1,\ldots,n+N\}$, and fix the sequence of subspaces $\mathbf{C}^1\subset\mathbf{C}^2\subset\ldots\subset\mathbf{C}^{n+N}$. The set of all complex n-planes X through the origin O in $\mathbf{C}^{n+N}$ forms the †complex Grassmann manifold $M_{n+N,n}(\mathbf{C})$. We denote by $E(\gamma_n^N)$ the set of all pairs (X,v), where $X\in M_{n+N,n}(\mathbf{C})$ and v is a vector in X. Then we can define a $2N$-universal complex n-dimensional vector bundle γ_n^N with base space $M_{n+N,n}(\mathbf{C})$, with total space $E(\gamma_n^N)$ and projection $(X,v)\to X$. Let $\omega=(\omega(1),\ldots,\omega(n))$ be a sequence of integers satisfying the condition $0\leqslant\omega(1)\leqslant\ldots\leqslant\omega(n)\leqslant N$. Then the set e_ω of all n-planes $X\subset\mathbf{C}^{n+N}$ through the origin O satisfying $\dim(X\cap\mathbf{C}^{i+\omega(i)})=i$, $\dim(X\cap\mathbf{C}^{i+\omega(i)-1})=i-1$, $i=1,2,\ldots,n$, forms a real $(2\sum_{i=1}^n\omega(i))$-dimensional open †cell. The set of all these open cells e_ω gives a †cellular subdivision of $M_{n+N,n}(\mathbf{C})$ as a CW complex. The closure $\bar{e}_\omega$ of e_ω is a cellular subcomplex of $M_{n+N,n}(\mathbf{C})$, called a **Schubert variety**. This is a †pseudomanifold with canonical orientation and represents a $(2\sum_{i=1}^n\omega(i))$-dimensional integral cycle, called a **Schubert cycle**. We denote $\bar{e}_\omega$ by $(\omega(1),\ldots,\omega(n))$. All these homology classes form the basis of the homology group $H_*(M_{n+N,n}(\mathbf{C});\mathbf{Z})$. The cocycle dual to the cycle $(\underbrace{0,\ldots,0}_{n-q}\underbrace{1,\ldots,1}_{q})$ represents the Chern class $c_q(\gamma_n^N)\in H^{2q}(M_{n+N,n}(\mathbf{C});\mathbf{Z})$. For the real Grassmann manifold $M_{n+N,n}(\mathbf{R})$, we can analogously define the universal Stiefel-Whitney classes.

Thom's Definition. Let ξ be an $\mathbf{R}^n$-bundle over B, B_ξ be its †Thom space, and $U\in H^n(B_\xi;\mathbf{Z}_2)$ be the †fundamental class of B_ξ. Let $j:B\to B_\xi$ be the inclusion induced from the zero cross section and $\varphi:H^k(B;\mathbf{Z}_2)\cong H^{k+n}(B_\xi;\mathbf{Z}_2)$ be the †Thom-Gysin isomorphism. Then we have $j^*U=w_n(\xi)$, $\varphi^{-1}(Sq^iU)=w_i(\xi)$ $(0\leqslant i\leqslant n)$, where Sq^i is the †Steenrod square (R. Thom [6]).

Definition in Terms of Differential Forms. Let B be a †differentiable manifold and $\xi=(P_\xi,B,U(n))$ be a differentiable principal $U(n)$-bundle over B. Let $\Omega=(\Omega_{ij})$ be the †curvature form corresponding to the †connection form $\omega=(\omega_{ij})$, $i,j=1,\ldots,n$, on P_ξ. Then Ω_{ij} is a complex-valued 2-form, and $\bar{\Omega}_{ij}=-\Omega_{ji}$. For the matrix Ω, we consider the differential form

$$\psi=\sum_q\psi_q=\det|I+(2\pi\sqrt{-1})^{-1}\Omega|,$$

where I is the unit matrix, the multiplication in the determinant is the †exterior product, and ψ_q is the part of degree $2q$ in ψ. Then ψ is defined as a real form independent of the connection ω. We have $d\psi_q=0$, and the cohomology class of $(-1)^q\psi_q$ in $H^{2q}(B;\mathbf{R})$ is the Chern class $c_q(\xi)$ with real coefficients (Borel and Hirzebruch [2], S. S. Chern [3]).

Definition in Terms of Symmetric Polynomials. (→ 427 Topology of Lie Groups and Homogeneous Spaces.)

F. Characteristic Classes of Manifolds

For a differentiable (complex or almost complex) manifold M, the characteristic classes of its tangent bundle are called **characteristic classes of the manifold** M. We shall denote **Stiefel-Whitney classes, Pontryagin classes, Euler-Poincaré classes, and Chern classes of** M by $w_i(M)$, $p_i(M)$, $X_n(M)$, and $c_i(M)$, respectively. These are invariants of differentiable structures, orientations, or (almost) complex structures of a manifold M if M is a differentiable, oriented differential, or (almost) complex manifold. By the **Stiefel-Whitney numbers** of an n-dimensional manifold M, we mean the values of n-dimensional monomials of Stiefel-Whitney classes of M on the fundamental homology class $((w_1(M)^{r_1} w_2(M)^{r_2} \ldots w_n(M)^{r_n})[M] \in \mathbf{Z}_2$, where $r_1 + 2r_2 + \ldots + nr_n = n$, $r_i \geqslant 0$. We can define integer-valued **Pontryagin numbers** and **Chern numbers** similarly. These numbers are called generally **characteristic numbers** of the given manifold. In particular, $X_n(M)[M] = \chi(M)$ is the †Euler-Poincaré characteristic.

In the case of topological manifolds, we can define characteristic classes in the following sense. Let M be a closed n-dimensional topological manifold and X^n the generator of $H^n(M; \mathbf{Z}_2)$. By defining $X^i(Y^{n-i}) = (X^i Y^{n-i})[M] \in \mathbf{Z}_2$ for $X^i \in H^i(M; \mathbf{Z}_2)$, $Y^{n-i} \in H^{n-i}(M; \mathbf{Z}_2)$, we have an isomorphism $H^i(M; \mathbf{Z}_2) \cong \mathrm{Hom}(H^{n-i}(M, \mathbf{Z}_2); \mathbf{Z}_2)$. The element $u_i \in H^i(M; \mathbf{Z}_2)$, corresponding to the homomorphism $Y^{n-i} \rightarrow Sq^i Y^{n-i}[M]$ under this isomorphism is called the **Wu class** of M, where Sq^i is the †Steenrod square. Moreover, we call $w_j = \sum_{i=0}^{j} Sq^{j-i} u_i \in H^j(M; \mathbf{Z}_2)$ the **Stiefel-Whitney class** of the topological manifold M. Then for any †differentiable structure $\mathscr{D}$, we have $w_j(M, \mathscr{D}) = w_j$. Therefore Stiefel-Whitney classes of differentiable manifolds are topological invariants (more precisely homotopy type invariants) (Thom [6], W. T. Wu [8]). J. W. Milnor [9] proved that Pontryagin classes of differentiable manifolds are not topological invariants. The image of $p_i(M)$ by the homomorphism $H^{4i}(M; \mathbf{Z}) \rightarrow H^{4i}(M; \mathbf{Q})$ induced by the inclusion $\mathbf{Z} \subset \mathbf{Q}$ (the rational number field) is called the **rational Pontryagin class**. In 1966, S. P. Novikov [10] proved the topological invariance of the rational Pontryagin class (→ Section H).

G. Index Theorem for Differentiable Manifolds

Let M be an oriented closed manifold of dimension $4k$. Putting $f(x, y) = x \cdot y[M]$ for elements x, y of the $2k$-dimensional real cohomology group $H^{2k}(M; \mathbf{R})$, we obtain a bilinear form on $H^{2k}(M; \mathbf{R})$. The †signature of the quadratic form $f(x, x)$ (namely, (number of positive terms) − (number of negative terms) in its canonical form) is a topological invariant (homotopy type invariant) of the manifold M. We call it the **index** or the **signature** of the manifold M and denote it by $\tau(M)$. If the dimension of M is not divisible by 4, we define $\tau(M) = 0$. For the product of manifolds we have $\tau(M \times N) = \tau(M) \cdot \tau(N)$. Also $\tau(M)$ is an invariant of the †cobordism class of M (Thom [6]).

The index τ of a differentiable manifold gives a homomorphism of the †cobordism ring Ω into the ring $\mathbf{Z}$ of integers. Hirzebruch investigated the multiplicative property of τ and gave its expression by means of Pontryagin numbers. Let P_i be the ith †elementary symmetric function of indeterminates $\beta_1, \ldots, \beta_n$. Then a homogeneous part of the formal power series

$$\prod_{i=1}^{n} \frac{\sqrt{\beta_i}}{\tanh \sqrt{\beta_i}}$$

of $\beta_1, \ldots, \beta_n$ is a symmetric polynomial of $\beta_1, \ldots, \beta_n$, and therefore a polynomial of P_i with rational coefficients. For $k \leqslant n$, we denote the homogeneous part of degree k by $L_k(P_1, \ldots, P_k)$. Specifically, if P_i are the Pontryagin classes $p_i(M^{4k})$ of a $4k$-dimensional closed differentiable manifold M^{4k}, then $L_k(P_1, \ldots, P_k)$ is a $4k$-dimensional cohomology class of M^{4k}. Then we have the formula

$$\tau(M^{4k}) = L_k(P_1, \ldots, P_k)[M^{4k}],$$

called the **index theorem** of differentiable manifolds (or **Hirzebruch index theorem**). For example, $L_1 = (1/3)P_1$, $L_2 = (1/45)(7P_2 - P_1^2)$, and $L_3 = (1/945)(62P_3 - 13P_2P_1 + 2P_1^3), \ldots$ (→ 114 Differential Topology). Later this index theorem was generalized to the †Atiyah-Singer index theorem (→ 237 K-Theory).

H. Combinatorial Pontryagin Classes

Let K be an oriented n-dimensional compact †homology manifold, and let Σ^r be the boundary of an oriented $(r+1)$-simplex, namely, the combinatorial r-sphere. Let $f: K \rightarrow \Sigma^{n-4i}$ be a †piecewise linear mapping. Then for almost all points y of Σ^{n-4i}, $f^{-1}(y)$ is an oriented $4i$-dimensional compact homology manifold, and its index $\tau(f^{-1}(y))$ is independent of y. We denote this by $\tau(f)$. The $\tau(f)$ is an invariant of the homotopy class of f. Let σ be the fundamental class of $H^{n-4i}(\Sigma^{n-4i}; \mathbf{Z})$. Then for $n \geqslant 8i + 2$, there exists a unique cohomology class $l_i = l_i(K) \in H^{4i}(K; \mathbf{Q})$ such that for any piecewise linear mapping $f: K \rightarrow \Sigma^{n-4i}$, we have

$(l_i \cdot f^*\sigma)[K] = \tau(f)$. We can remove the restriction $n \geqslant 8i+2$ if we take $K \times \Sigma^m$ for K and define $l_i(K)$ to be $l_i(K \times \Sigma^m)$ for sufficiently large m. If K is a ${}^{\dagger}C^1$-triangulation of a differentiable manifold M, $l_i(K)$ coincides with the class $L_i(P_1, \ldots, P_i)$ defined by Hirzebruch (R. Thom [7]; V. Rokhlin and A. Shvarts), where $P_j = p_j(M)$ is the Pontryagin class of M. Since the variable P_i can be expressed as a polynomial with rational coefficients of $L_j(P_1, \ldots, P_j)$, $j \leqslant i$, we define the **combinatorial Pontryagin class** $p_i(K)$ of a homology manifold K as the polynomial of $l_j(K)$ with rational coefficients. Therefore, if K is a C^1-triangulation of a differentiable manifold M, we have $p_i(K) = p_i(M)$ (modulo torsion elements). The class $l_i(K)$ and consequently $p_i(K)$ are important combinatorial invariants of K. These classes are topological invariants [10].

References

[1] A. Borel, Sur la cohomologie des espaces fibrés principaux et des espaces homogènes de groupes de Lie compacts, Ann. Math., (2) 57 (1953), 115–207.
[2] A. Borel and F. Hirzebruch, Characteristic classes and homogeneous spaces I, II, III, Amer. J. Math., 80 (1958), 459–538; 81 (1959), 315–382; 82 (1960), 491–504.
[3] S. S. Chern, Characteristic classes of Hermitian manifolds, Ann. Math., (2) 47 (1946), 58–121.
[4] F. Hirzebruch, Topological methods in algebraic geometry, Springer, third edition, 1966.
[5] E. Stiefel, Richtungsfelder und Fernparallelismus in n-dimensionalen Mannigfaltigkeiten, Comment. Math. Helv., 8 (1936), 3–51.
[6] R. Thom, Espaces fibrés en sphères et carrés de Steenrod, Ann. Sci. Ecole Norm. Sup., (3) 69 (1952), 109–182.
[7] R. Thom, Les classes caractéristiques de Pontrjagin des variétés triangulées, Symposium Internacional de Topologia Algebraica, Mexico, 1958.
[8] W. T. Wu, Sur les classes caractéristiques des structures fibrés sphériques, Actualités Sci. Ind., Hermann, 1952.
[9] J. W. Milnor, Microboundles I, Topology, 3 (1964), Suppl. I., 53–80.
[10] S. P. Novikov, Manifolds with free Abelian fundamental groups and their applications (Pontrjagin classes, smoothness, multidimensional knots), Amer. Math. Soc. Transl., 71 (1968), 1–42. (Original in Russian, 1966.)
[11] N. Steenrod, The topology of fibre bundles, Princeton Univ. Press, 1951.
[12] D. Husemoller, Fibre bundles, McGraw-Hill, 1966.
[13] J. Milnor and J. Stasheff, Characteristic classes, Princeton Univ. Press, 1974.

57 (XXI.6) Chinese Mathematics

A. Mathematics in the Chao, Han, and Tang Dynasties (3rd Century B.C.–10th Century A.D.)

In ancient China, the art of divination, called *yi*, was used in government administration. This was a kind of calculation that used pieces called *tse*. The book embodying it, called the *I-Ching*, is still popularly used. It shows that "numbers" or mathematics was seriously utilized in China at that time. The multiplication table for numbers up to nine (called the Pythagorean table in the West) was known in China from the legendary period. However, mathematics in the Greek sense, that is, mathematics as a logically systematized science, was unknown in ancient China.

Suanching-Shihshu, or the Ten Books on Arithmetic—namely, *Choupi-Suanching*, *Chiuchang-Suanshu*, *Haitao-Suanching* (edited by Liu Hui), *Suntzu-Suanching*, *Wutsao-Suanching*, *Hsiahouyang-Suanching*, *Changchiu-Suanching*, *Wuching-Suanshu*, *Chiku-Suanching* (edited by Wang Hsiao-Tong), and *Shushu-Chiyi* (edited by Hsu Yue) —came into being between the 2nd century B.C. and the 6th century A.D., from the Chao to the Han eras, with the exception of the *Chiku-Suanching* compiled in the Tang era. These are the only mathematical texts from this early period whose authors and times of publication are known. They were used in the civil service examination for selecting administrators up to the beginning of the Sung era (960 A.D.). The most important among them is *Chinchang-Suanshu*, or the Book of Arithmetic, which contains nine chapters. It treats positive and negative fractions with laws of operations on signed numbers, equations, and the elementary mathematical knowledge of daily life. The *Chiku-Suanching* contains a number of problems reducible to equations of the 3rd and 4th degrees.

There were also two works called *San Tung Shu* (edited by Tong Chuan) and *Chui Shu* (edited by Tsu Chung-Chih), but no copies of them are extant. Later works, one from the Sui era (published in 636) and another from the Tang era, tell us that the latter contained the result $3.1415927 > \pi > 3.1415926$ and the approximate values $355/113$ and $22/7$ for π.

In the 1st century A.D. Buddhism was introduced from India, and paper was invented.

However, despite the communication with India, neither the Indian numeration system, written calculation, nor the abacus was at that time widely used in China. The extraction of square or cube roots was done with calculating rods.

B. Mathematics in the Sung and Yuan Dynasties (10th–14th Centuries)

In the Sung and Yuan periods contact was made with the Arab world. In the 13th century, a mechanical algebra utilizing calculating rods made remarkable progress; this can be attributed to Arab influence. Toward the end of the Sung era appeared the *Shushu-Chiuchang* by Ch'in Chiu-Shao and the *Yiku-Yentan* by Li Chih. The former gives a method like Horner's for approximate solution of equations, and the latter gives the principle of *tienyuan-shu*, i.e. the mechanical algebra of this period. The principle of *tienyuan-shu* was further expounded in the *Suanhsueh Chimeng* (1295) and the *Suyuan Yuchien* (1303) by Shih Shih-Chich, the *Yanghui Suanfa* by Yang Hui, and other works. These were introduced into Japan and they influenced the *wasan* (Japanese mathematics) of early times. Until recently, no further original mathematical ideas appeared in China.

C. Mathematics after the Ming Era (15th Century)

In this epoch, European renaissance civilization began to influence the Orient. In 1607, Matteo Ricci (1552–1610) translated Books I–IV of Euclid's *Elements* into Chinese with the aid of Hsu Kuang-Chi. In 1592 *Suanfa Tang-tsung* by Ch'êng Ta-Wei appeared, which dealt with the use of the abacus. This book had great influence upon *wasan*.

No development was seen in the indigenous mathematics of the Ching era, that is, after the 17th century, but science and technology were imported by Christian missionaries. This brought about calendar reform from the lunar to the solar method. On the other hand, new editions of classical works such as the Ten Books on Arithmetic began to appear in this period. Emperor Kang Hsi-Ti (1655–1722), who was in correspondence with Leibniz, asked Ferdinand Verbiest (renamed Nan Huai Jen in Chinese) to compile *Shuli-Chingwen* (Advanced mathematics), whose 53 chapters were completed in 1723. This book dealt with European-style algebra and trigonometry. In the latter half of the 19th century, Alexander Wylie translated a number of Western mathematical books into Chinese, including Books VII–XIII of Euclid's *Elements* and some works on calculus. Many current Chinese and Japanese mathematical terms originated with this translation.

References

[1] Y. Li, History of mathematics in China (Japanese translation), Seikatusya, 1940.
[2] Y. Mikami, The development of mathematics in China and Japan, Teubner, 1913 (Chelsea, 1961).

58 (X.9) C^∞-Functions and Quasi-Analytic Functions

A. General Remarks

An example of a †C^∞-function is a †real analytic function, which is defined to be a function that can be expressed as a power series that converges in a neighborhood of each point of the domain where the function is defined. Many examples, however, show that real analytic functions form a rather small subset of the C^∞-functions. Sometimes C^∞-functions (not real analytic functions) play essential roles in the development of theories of analysis (→ 105 Differentiable Manifolds S). On the other hand, there is a subfamily of C^∞-functions having some remarkable properties in common with the family of real analytic functions. This family is called the **family of quasi-analytic functions**. It has been an important object of study since the beginning of the 20th century. On the other hand, the Gevrey class functions, which are no longer quasi-analytic in general, also constitute an important subset of C^∞-functions. The first part of this article deals with C^∞-functions, the second part with quasi-analytic functions, and the third part with Gevrey class functions.

B. C^∞-Functions

Let Ω be an open set of the n-dimensional real Euclidean space $\mathbf{R}^n$. A real-valued function $f(x_1, \ldots, x_n)$ defined on Ω is called a **function of class** C^∞ on Ω (or **C^∞-function** on Ω) if $f(x_1, \ldots, x_n)$ is continuously differentiable up to any order. The totality of C^∞-functions defined on Ω is denoted by $C^\infty(\Omega)$. It is an †associative algebra over the real number field $\mathbf{R}$. A continuous function f defined on some closed set F

of $\mathbf{R}^n$ is called a C^∞-function on F if there exist an open neighborhood U of F and a function $g \in C^\infty(U)$ such that $f = g | F$. This definition is equivalent to the following (H. Whitney [6]): For any multi-index $\alpha = (\alpha_1, \ldots, \alpha_n)$, we can find a continuous function $f^\alpha(x_1, \ldots, x_n)$ on F such that (i) $f^0(x_1, \ldots, x_n) = f(x_1, \ldots, x_n)$ and (ii) for any positive integer r and for every multi-index α with $|\alpha| \leqslant r$,

$$\lim_{y \to x} \left[\frac{1}{\| y - x \|^{r - |\alpha|}} \left| f^\alpha(x) - \sum_{|\alpha + \beta| < r} f^{\alpha+\beta}(x) \frac{(x-y)^\beta}{\beta!} \right| \right] = 0,$$

where $\| \quad \|$ denotes the Euclidean norm of $\mathbf{R}^n$.

C. Local Theory of C^∞-Functions

We shall now introduce an equivalence relation $\sim$ in $C^\infty(\mathbf{R}^n)$, defined as follows: $f \sim g \Leftrightarrow f | U = g | U$ for some open neighborhood U of the origin. Let $\mathscr{E}_n$ denote the quotient set $C^\infty(\mathbf{R}^n)/\sim$, which naturally inherits the structure of an associative algebra from $C^\infty(\mathbf{R}^n)$. An element of $\mathscr{E}_n$ is called a **germ of a C^∞-function at the origin**. We denote the germ of $f \in C^\infty(\mathbf{R}^n)$ by $\tilde{f}$. To $\tilde{f} \in \mathscr{E}_n$, we assign the **formal Taylor expansion** $\sum (D^\alpha f(0)/\alpha!) x^\alpha$ around 0, where $D^\alpha f$ means $(\partial^{r_1 + \cdots + r_n} f)/(\partial^{r_1} x_1 \ldots \partial^{r_n} x_n)$ for $\alpha = (r_1, \ldots, r_n)$. This assignment induces a homomorphism τ from $\mathscr{E}_n$ to the †ring of formal power series $\mathbf{R}[[x_1, \ldots, x_n]]$ of n variables. The homomorphism τ is surjective but not injective. Put $\Lambda_n = \tau^{-1}(0) \subset \mathscr{E}_n$. A function f whose germ $\tilde{f}$ belongs to Λ_n is called a **flat function**. The function $\varphi(x)$ defined by $\varphi(x) = \exp(-1/x^2)$ when $x \neq 0$ and $\varphi(0) = 0$ is an example of a flat function on $\mathbf{R}^1$. A close study of the relationship between $\mathscr{E}_n$ and $\mathbf{R}[[x_1, \ldots, x_n]]$ leads to the **preparation theorem for C^∞-functions**, which can be stated as follows: Let $\tilde{F}(x_1, \ldots, x_n) \in \mathscr{E}_n$ satisfy $\tilde{F}(0, \ldots, 0, x_n) = x_n^p \tilde{g}(x_n)$ $(\tilde{g} \in \mathscr{E}_1, g(0) \neq 0)$. Then any $\tilde{f} \in \mathscr{E}_n$ can be expressed as $\tilde{f} = \tilde{F}\tilde{Q} + \tilde{R}$, where $\tilde{Q} \in \mathscr{E}_n$ and $R = \sum_{i=0}^{p-1} r_i(x_1, \ldots, x_{n-1}) x_n^i$ with $r_i \in \mathscr{E}_{n-1}$ (B. Malgrange [3]).

Let $f(x_1, \ldots, x_n)$ be a symmetric function in $(x_1, \ldots, x_n)$ of class C^∞. Then there exists a germ $\tilde{g} \in \mathscr{E}_n$ such that $\tilde{f}(x_1, \ldots, x_n) = \tilde{g}(\sigma_1, \ldots, \sigma_n)$, where $\sigma_1, \ldots, \sigma_n$ denote elementary symmetric functions with respect to $x_1, \ldots, x_n$ (G. Glaeser, Malgrange). Let $\tilde{f} \in \mathscr{E}_1$ satisfy $f(x) = f(-x)$. Then there exists a germ $\tilde{g} \in \mathscr{E}_1$ such that $\tilde{f}(x) = \tilde{g}(x^2)$ (H. Whitney).

D. Global Results

Case of n Variables. $C^\infty(\Omega)$ becomes a †Fréchet space when it is endowed with the topology of uniform convergence on compact sets for all partial derivatives. Let J and J_1 be two closed †ideals of $C^\infty(\Omega)$. Then we have $J = J_1$ if and only if $\tau_x(J) = \tau_x(J_1)$ for each $x \in \Omega$, where τ_x is the mapping from $C^\infty(\Omega)$ to the ring of the formal power series that assigns the formal Taylor series of f around x (Whitney).

Case of One Variable. In the case of one variable, further information can be obtained from various points of view. In the following, f denotes a C^∞-function defined on the unit interval $I = [0, 1]$. If f satisfies $f(1) = 1, f'(0) = \ldots = f^{(r-1)}(0) = 0$, then we have

$$\frac{1}{m_1} + \frac{1}{\sqrt{m_2}} + \ldots + \frac{1}{\sqrt[r]{m_r}} < k,$$

where $m_i = \sup \{ |f^{(i)}(x)| \mid x \in I \}$ and where k is some constant independent of the choice of f and r (E. Borel). Similar kinds of inequalities were obtained by A. N. Kolmogorov, A. Gorny, and H. Cartan. Let A be an arbitrary countable set of real numbers. If for any $x \in I$ we can find an integer $r(x)$ such that $f^{(r(x))}(x) \in A$, then such a function f is necessarily a polynomial. The interval I can be divided into three disjoint subsets: $S_1^{(f)}$, $S_2^{(f)}$, and $S_3^{(f)}$. These are characterized as follows: For $x \in S_1^{(f)}$ the formal Taylor series $\tau_x(f)$ of f around x converges to f in some neighborhood of x. For $x \in S_2^{(f)}$, $\tau_x(f)$ diverges. And for $x \in S_3^{(f)}$, $\tau_x(f)$ converges in some neighborhood of x but does not tend to f. Then $S_1^{(f)}$ is an open set and $S_2^{(f)}$ is a G_δ-set, while $S_3^{(f)}$ is an F_σ-set of the †first category. Conversely, let $I = S_1 + S_2 + S_3$ be any partition of I into an open set S_1, a G_δ-set S_2, and an F_σ-set S_3 of the first category. Then there is some $f \in C^\infty(I)$ with $S_i = S_i^{(f)}$ $(i = 1, 2, 3)$ [5].

E. Relations between C^∞-Functions and Real Analytic Functions

Let $C^\omega(I)$ be the set of real analytic functions on I. Then $C^\omega(I)$ is a subalgebra of $C^\infty(I)$. Applying the above result in the case of $S_1 = \varnothing$, we find a function $f \in C^\infty(I)$ that admits no real analytic function coinciding with f in a subinterval of I. Actually, functions with such a property are distributed densely in $C^\infty(I)$. A necessary and sufficient condition for a function $f \in C^\infty(I)$ to belong to $C^\omega(I)$ is that for suitable constants A and k, $|f^{(n)}(x)| \leqslant A k^n n!$, $x \in I$, $n = 0, 1, 2, \ldots$, be valid (**Pringsheim's theorem**). If $f^{(n)}(x) \geqslant 0$ for all $x \in I$ and $n = 0, 1, 2, \ldots$, then $f \in C^\omega(I)$ (S. N. Bernshteĭn). For any open set Ω $(\subset \mathbf{R}^n)$, the set $C^\omega(\Omega)$ of real analytic functions on Ω is dense in $C^\infty(\Omega)$ (**polynomial approximation theorem**). This result is true even when the topology of $C^\infty(\Omega)$

is replaced by a stronger one (Whitney [6]). Let $f\in C^\infty(\Omega)$ and $\varphi\in C^\infty(\Omega)$. Then we can find $g\in C^\infty(\Omega)$ satisfying $f=g\varphi$ if and only if for any $x\in\Omega$, $\tau_x(f)$ is divisible by $\tau_x(\varphi)$ in the ring of formal power series (S. Łojasiewicz, Malgrange [3]).

F. Quasi-Analytic Functions

The investigation of quasi-analytic functions began with the attempt to obtain an intrinsic characterization of analytic functions. Borel defined monogenic functions as functions differentiable on their domains of definition, which can be any subset of the complex plane, not necessarily assumed to be open (→ 198 Holomorphic Functions Q). Similar to complex analytic functions, monogenic functions are uniquely determined by their values on any curve. While quasi-analyticity can be defined by such properties, it is customary to approach quasi-analytic functions from another aspect, that is, the behavior of higher derivatives of C$^\infty$-functions.

Generally, a subset B of $C^\infty(I)$ is called a **set of quasi-analytic functions** if the mapping τ_x: $B\to\mathbf{R}[[x]]$ defined in Section D is injective at each point $x\in I$. The functions belonging to B are called **quasi-analytic** (with respect to B). Here, an important problem is to characterize a set of quasi-analytic functions B by specializing the image of $\tau_x(B)$.

Now let $\{M_n\}$ be a sequence of positive numbers. Let $C(M_n)$ be the subset of $C^\infty(I)$ consisting of f such that

$$|f^{(n)}(x)|\leqslant Ak^nM_n,\qquad x\in I,\quad n=0,1,2,\ldots,$$

where $A=A(f)$ and $k=k(f)$ are constant. Then Pringsheim's theorem simply asserts that $C(n!)=C^\omega(I)$.

In 1912, J. Hadamard raised the problem of determining the condition that the sequence $\{M_n\}$ should satisfy so that $C(M_n)$ becomes a set of quasi-analytic functions [9]. A. Denjoy showed that if

$$M_n=(n\log^1 n\log^2 n\ldots\log^p n)^n,$$

where

$$\log^1 n=\log n,\ \log^p n=\log(\log^{p-1} n),\qquad p=2,3,\ldots,$$

then $C(M_n)$ is a set of quasi-analytic functions [10]. Later he derived an improved result that the condition $\sum M_n^{-1/n}=\infty$ is sufficient. T. Carleman first gave a necessary and sufficient condition for $C(M_n)$ to be a set of quasi-analytic functions, and later A. Ostrowski and T. Bang gave another version of the same condition [11–13]. The condition states essentially the following: A necessary and sufficient condition for $C(M_n)$ to be a family of quasi-analytic functions in the interval (a,b) is given by either (i) $\sum\beta_n^{-1}=+\infty$, where $\beta_n=\inf_{k\geqslant n}M_k^{1/k}$ (Carleman), or (ii) $\int^\infty(\log T(r))/(r^2)dr=\infty$, where $T(r)=\sup_{n\geqslant1}(r^n/M_n)$ (Ostrowski, Bang). S. Mandelbrojt and T. Bang also gave another condition [13, 14]. (The simplest proof of these results is found in [13] or [15], where the proof follows Bang's idea.)

Related to the above theorem, we also have the following: Let $\{M_n\}$ be a sequence of positive numbers with $\sum(M_n/M_{n+1})<\infty$. For $\alpha>0$ we can find $f\in C(M_n)$ defined on $(-\infty,\infty)$ such that $f(0)>0$, $f^{(n)}(\pm\alpha)=0$. Moreover, for $0<\alpha<\beta$ there exists $f\in C(M_n)$ such that $f(0)>0$, $f^{(n)}=0$ $(\alpha\leqslant x\leqslant\beta,\ n=0,1,2,\ldots)$ [16].

Suppose that we are given an interval I and increasing sequences $\{\nu_n\}$ and $\{M_n\}$ of positive numbers. Then we have the problem of finding suitable conditions on $\{\nu_n\}$ and $\{M_n\}$ under which the mapping $f\to\{f^{(\nu_n)}(x_0)\}$ gives an injective mapping from $C(M_n)$ to the sequences above. When $\{\nu_n\}$ and $\{M_n\}$ satisfy the above conditions, then a function belonging to $C(M_n)$ is called **quasi-analytic (ν_n) in the generalized sense**. The study of the inclusion relation between two families $C(M_n)$ and $C(M_n')$ also deserves attention. In [15] the relation between $C(M_n)$ and $C(n!)=C^\omega(I)$ is discussed in detail. There are many open problems concerning the relationship between $C(M_n)$ and $C(M_n')$ in general.

Quasi-analytic functions are closely related to problems in various branches of analysis, in particular the theories of complex analytic functions, Fourier series, Fourier integrals, Dirichlet series, and asymptotic expansions [8, 15, 17].

G. Gevrey Class Functions

This class of functions has its origin in the study of †parabolic equations. A C$^\infty$-function $f(x)$ defined in a domain of $\mathbf{R}^n$ is called of **Gevrey class** s $(1\leqslant s<\infty)$ if for every compact set K in that domain there exist positive constants A_K and C_K such that

$$\max_K|D^\alpha f(x)|\leqslant A_KC_K^{|\alpha|}\alpha!^s$$

for all multi-indices α $(|\alpha|=\alpha_1+\ldots+\alpha_n,\ \alpha!=\alpha_1!\alpha_2!\ldots\alpha_n!)$. A typical example is

$$f(x)=\begin{cases}\exp(-1/x) & \text{for } x>0,\\ 0 & \text{for } x\leqslant0,\end{cases}$$

which is of Gevrey class 2. Let us denote the function space of a Gevrey class s by $\gamma^{(s)}$. Evidently $\gamma^{(s)}\subset\gamma^{(s')}$ if $s<s'$. Let $f(x)\in\gamma^{(s)}$ and $F(y)\in\gamma^{(s)}$; then $x\mapsto F(f(x))\in\gamma^{(s)}$. In particular,

the sum and the product of two functions in $\gamma^{(s)}$ belong also to $\gamma^{(s)}$. Furthermore, the †implicit function theorem holds in this class. Unlike the class of analytic functions, we can use this class of functions of class s (>1) for a †partition of unity, and several problems in partial differential equations can be treated within this function space. Gevrey class functions are discussed in a complete form in [20] (→ also [19]).

References

For functions of class C^{∞},
[1] R. P. Boas, A primer of real functions, Carus Math. Monographs, Wiley, 1960.
[2] G. Glaeser, Fonctions composées différentiables, Ann. Math., (2) 77 (1963), 193–209.
[3] B. Malgrange, Le théorème de préparation en géométrie différentiables, Sém. H. Cartan, 1962–1963, Exp. 11, Inst. H. Poincaré, Univ. Paris, 1964.
[4] B. Malgrange, Ideals of differentiable functions, Oxford Univ. Press, 1966.
[5] H. Salzmann and K. Zeller, Singularitäten unendlich oft differenzierbarer Funktionen, Math. Z., 62 (1955), 354–367.
[6] H. Whitney, Analytic extensions of differentiable functions defined in closed sets, Trans. Amer. Math. Soc., 36 (1934), 63–89.
[7] H. Whitney, Differentiable even functions, Duke Math. J., 10 (1943), 159–160.
[8] H. Whitney, On ideals of differentiable functions, Amer. J. Math., 70 (1948), 635–658.
For quasi-analytic functions,
[9] J. Hadamard, Sur la généralisation de la notion de fonction analytique, Bull. Soc. Math. France, 40 (1912), 28–29.
[10] A. Denjoy, Sur les fonctions quasi-analytiques de variable réelle, C. R. Acad. Sci. Paris, 173 (1921), 1329–1331.
[11] T. Carleman, Les fonctions quasi-analytiques, Gauthier-Villars, 1926.
[12] A. Ostrowski, Über quasianalytische Funktionen und Bestimmtheit asymptotischer Entwickelungen, Acta Math., 53 (1929), 181–266.
[13] T. S. V. Bang, Om quasi-analytiske Funktioner, Thesis, Univ. of Copenhagen, 1946.
[14] S. Mandelbrojt, Analytic functions and classes of infinitely differentiable functions, Rice Inst. Pamphlet 29, no. 1, 1942.
[15] S. Mandelbrojt, Séries adhérentes, régularisation des suites, applications, Gauthier-Villars, 1952.
[16] S. Mandelbrojt, Some theorems connected with the theory of infinitely differentiable functions, Duke Math. J., 11 (1949), 341–349.
[17] S. Mandelbrojt, Séries de Fourier et classes quasi-analytiques de fonctions, Gauthier-Villars, 1935.
[18] R. E. A. C. Paley and N. Wiener, Fourier transformations in the complex domain, Amer. Math. Soc. Colloq. Publ., 1934.
For Gevrey class functions,
[19] J. Hadamard, Les fonctions de classe supérieure dans l'équation de Volterra, J. Analyse Math., 1 (1951), 1–10.
[20] M. Gevrey, Sur la nature des solutions des équations aux dérivées partielles, Ann. Sci. École Norm. Sup., ser. 3, 35 (1918), 243–304.

59 (V.14) Class Field Theory

A. History

The notion of a class field was first introduced by D. Hilbert (1898). Let k be an †algebraic number field and K a †Galois extension of k. Hilbert called such a field K a class field over k (or K/k was called a class field) when the following property was satisfied: A †prime ideal $\mathfrak{p}$ of k of absolute degree 1 (i.e., a prime ideal whose †absolute norm is a prime number) is decomposed in K as the product of prime ideals of K of absolute degree 1 if and only if $\mathfrak{p}$ is a †principal ideal. (Such a field K is now said to be an **absolute class field** over k in order to distinguish it from a class field defined later more generally by T. Takagi as explained below.) Hilbert conjectured the following theorems (1)–(4) together with the principal ideal theorem (→ Section D), and proved them in some special cases. (1) For any algebraic number field k there exists one and only one class field K/k. (2) A class field K/k is an †Abelian extension whose †Galois group is isomorphic to the †ideal class group of k. Hence the degree $n=[K:k]$ is equal to the †class number h of k. (3) The †relative different of a class field K/k is the principal order; thus K/k is an †unramified extension. (4) Let $\mathfrak{p}$ be a prime ideal of k, and let f be the smallest positive integer such that $\mathfrak{p}^f$ is a principal ideal. Then $\mathfrak{p}$ is decomposed in the class field K/k as $\mathfrak{p}=\mathfrak{P}_1\mathfrak{P}_2\dots\mathfrak{P}_g$, $N_{K/k}(\mathfrak{P}_i)=\mathfrak{p}^f$, $fg=n$.

Hilbert was led to these conjectures by the analog to the theory of †algebraic functions in one variable. Theorems (1)–(4) were proved by P. Furtwängler (*Math. Ann.*, 63 (1907)), but these results were subsumed under the class field theory of Takagi, who generalized the notion of class field and proved that every Abelian extension of k is a class field over k (*J. Coll. Sci. Imp. Univ. Tokyo*, (9) 41 (1920)). Since

then, the arithmetic of Abelian extensions of k has developed through this theory. In Takagi's paper, L. Kronecker's problem concerning Abelian extensions of an imaginary quadratic field (→ 73 Complex Multiplication) was solved simultaneously; this had been an open problem since the 19th century. Later, E. Artin proved the general law of reciprocity (*Abh. Math. Sem. Univ. Hamburg*, 5 (1927)) which put class field theory into its complete form. The original proof by Takagi was rather complicated, and H. Hasse, Artin, J. Herbrand, C. Chevalley, and others tried to simplify it. In particular, Chevalley introduced the notion of †ideles and gave a purely arithmetic proof. On the other hand, attempts are also being made to generalize this theory to non-Abelian extensions. We mention here the results of G. Shimura [15] and Y. Ihara [16]. Today, class field theory is considered one of the most beautiful theories in mathematics.

B. Definition of a Class Field

Let k be an algebraic number field. For the definition of a general class field over k, we need a generalization of the ideal class group of k (→ 14 Algebraic Number Fields H). Let $\mathfrak{m}$ be an †integral divisor of k, and let $\mathfrak{J}(\mathfrak{m})$ be the multiplicative group of all †fractional ideals of k which are †relatively prime to $\mathfrak{m}$. For the rest of this article, we mean by an ideal of k a fractional ideal of k. Denote by $S(\mathfrak{m})$ the †ray modulo $\mathfrak{m}$. Let $H(\mathfrak{m})$ be an †ideal group modulo $\mathfrak{m}$, that is, a subgroup of $\mathfrak{J}(\mathfrak{m})$ containing $S(\mathfrak{m})$. A Galois extension K of k is said to be a **class field** over k for the ideal group $H(\mathfrak{m})$ if the following property is satisfied: A prime ideal $\mathfrak{p}$ of k of absolute degree 1 which is relatively prime to $\mathfrak{m}$ is decomposed in K as the product of prime ideals of K of absolute degree 1 if and only if $\mathfrak{p}$ belongs to $H(\mathfrak{m})$. The absolute class field of Hilbert is the case where $\mathfrak{m}=(1)$ and $H(\mathfrak{m})$ is the group of all principal ideals of k.

A class field K/k for an ideal group H is uniquely determined by H (**uniqueness theorem**). The †conductor $\mathfrak{f}$ of H is said to be the **conductor** of the class field for H. The ideal group H corresponding to the class field K/k is determined by K as follows: $H(\mathfrak{f})/S(\mathfrak{f})$ is the union of all cosets C of $\mathfrak{J}(\mathfrak{f})$ modulo $S(\mathfrak{f})$ such that C contains a †relative norm $N_{K/k}(\mathfrak{A})$ of some ideal $\mathfrak{A}$ of K which is relatively prime to $\mathfrak{f}$. In general, let K/k be a Galois extension and $\mathfrak{m}$ be an integral divisor of k. Let $H(\mathfrak{m})$ be the union of all cosets C of $\mathfrak{J}(\mathfrak{m})$ modulo $S(\mathfrak{m})$ such that C contains a relative norm $N_{K/k}(\mathfrak{A})$ of some ideal $\mathfrak{A}$ of K which is relatively prime to $\mathfrak{m}$. Then $H(\mathfrak{m})$ is a multiplicative subgroup of $\mathfrak{J}(\mathfrak{m})$, and the index $h=(\mathfrak{J}(\mathfrak{m}):H(\mathfrak{m}))$ is not greater than the degree $n=[K:k]$. We have $h=n$ if and only if K/k is the class field for H. Hence a class field K/k can be defined as a Galois extension of k such that $h=n$ for a suitable integral divisor $\mathfrak{m}$ of k.

C. Fundamental Theorems in Class Field Theory

(1) **Main theorem**: Any Abelian extension K/k is a class field over k for a suitable ideal group H.

(2) **Existence theorem**: For any ideal group $H(\mathfrak{m})$ there exists one and only one class field for $H(\mathfrak{m})$.

(3) **Composition theorem**: Let K_1 and K_2 be class fields for H_1 and H_2, respectively. Then the composite field K_1K_2 is the class field over k for $H_1 \cap H_2$. Consequently, $K_1 \supset K_2$ if and only if $H_1 \subset H_2$.

(4) **Isomorphism theorem**: The Galois group of a class field K/k for $H(\mathfrak{m})$ is isomorphic to $\mathfrak{J}(\mathfrak{m})/H(\mathfrak{m})$. In particular, every class field K/k is an Abelian extension of k.

(5) **Decomposition theorem**: Let $\mathfrak{f}$ be the conductor of the class field for H. If $\mathfrak{p}$ is a prime ideal of k relatively prime to $\mathfrak{f}$ and f is the smallest positive integer with $\mathfrak{p}^f \in H$, then $\mathfrak{p}$ is decomposed in K/k as $\mathfrak{p}=\mathfrak{P}_1\mathfrak{P}_2\ldots\mathfrak{P}_g$, $N_{K/k}(\mathfrak{P}_i)=\mathfrak{p}^f$, $fg=n$.

(6) **Conductor-ramification theorem**: Let $\mathfrak{f}$ be the conductor of a class field K/k. Then $\mathfrak{f}$ is not divisible by any prime divisor that is unramified for K/k, and $\mathfrak{f}$ is divisible by every prime divisor that ramifies for K/k. Let $\mathfrak{f}=\prod\mathfrak{f}_\mathfrak{p}$, $\mathfrak{f}_\mathfrak{p}=\mathfrak{p}^c$. Then $\mathfrak{f}_\mathfrak{p}$ coincides with the †$\mathfrak{p}$-conductor of K/k, and the exponent c can be explicitly expressed by the order of the †ramification groups and the †ramification numbers of $\mathfrak{p}$ for K/k (→ 14 Algebraic Number Fields P).

(7) Let $\mathfrak{p}$ be a prime ideal of k that ramifies for K/k. Let $H_\mathfrak{p}$ be the ideal group of k such that (i) the conductor of $H_\mathfrak{p}$ is relatively prime to $\mathfrak{p}$ and (ii) $H_\mathfrak{p}$ is the minimal ideal group of k containing H with property (i). Let $n=[K:k]$, $e=(H_\mathfrak{p}:H)$, and $\mathfrak{p}^f \in H_\mathfrak{p}$, where $\mathfrak{p}^d (d<f) \notin H_\mathfrak{p}$. Then $\mathfrak{p}$ is decomposed in K/k as $\mathfrak{p}=(\mathfrak{P}_1\mathfrak{P}_2\ldots\mathfrak{P}_g)^e$, $N_{K/k}(\mathfrak{P}_i)=\mathfrak{p}^f$, $n=efg$.

(8) **Translation theorem**: Let K/k be the class field for an ideal group $H(\mathfrak{m})$, and let Ω be an arbitrary finite extension of k. Then $K\Omega/\Omega$ is the class field for H^*, where H^* is the ideal group of Ω consisting of all ideals $\mathfrak{b}$ of Ω with $N_{\Omega/k}(\mathfrak{b}) \in H(\mathfrak{m})$. In particular, the conductor of $K\Omega/\Omega$ is a divisor of the conductor of K/k.

(9) **Artin's general law of reciprocity**: Let K/k be the class field for an idea group H with the

conductor $\mathfrak{f}$. We denote the †Artin symbol of an ideal $\mathfrak{a}$ of k that is relatively prime to $\mathfrak{f}$ by

$$(K/\mathfrak{a})=\left(\frac{K/k}{\mathfrak{a}}\right).$$

Let a mapping Φ from $\mathfrak{J}(\mathfrak{f})$ to the Galois group G of K/k be defined by $\Phi(\mathfrak{a})=(K/\mathfrak{a})$ for $\mathfrak{a}\in\mathfrak{J}(\mathfrak{f})$. Then Φ induces the isomorphism $\mathfrak{J}(\mathfrak{f})/H(\mathfrak{f})\cong G$. Namely, the isomorphism mentioned in (4) is explicitly given by the Artin symbol. Also, the ideal group $H(\mathfrak{f})$ is characterized as the set of all ideals $\mathfrak{a}$ such that $\mathfrak{a}\in\mathfrak{J}(\mathfrak{f})$ and $(K/\mathfrak{a})=1$. From this theorem we can prove all the known laws of reciprocity for power-residue and norm-residue symbols (→ 14 Algebraic Number Fields O, Q, R).

From the general results of class field theory we can systematically derive all the known theorems concerning the arithmetic of quadratic fields, cyclotomic fields, and Kummer extensions.

D. Principal Ideal Theorem

Let K/k be an absolute class field. Then the extension of any ideal of k to K is a principal ideal of K. This theorem is called the **principal ideal theorem**. It was conjectured by Hilbert, formulated by Artin as a theorem of group theory, and proved by Furtwängler (*Abh. Math. Sem. Univ. Hamburg*, 7 (1930)). Later a simple proof was given by S. Iyanaga (*Abh. Math. Sem. Univ. Hamburg*, 10 (1934)). This theorem was also generalized to the following general principal ideal theorem (Iyanaga, *Japan. J. Math.*, 7 (1930)): Let K/k be the class field for the ray $S(\mathfrak{f})$, and let $\mathfrak{f}=\mathfrak{F}\mathfrak{D}$, where $\mathfrak{D}$ is the relative different of K/k. Then the extension to K of any ideal of k that is relatively prime to $\mathfrak{f}$ belongs to $S(\mathfrak{F})$. Put $\mathfrak{F}=\prod\mathfrak{P}^{v}$. Then v is equal to the ramification number v_r+1 (→ 14 Algebraic Number Fields K). For an absolute class field K/k, let the extension of an ideal $\mathfrak{a}$ to K be $(\Theta(\mathfrak{a}))$. Then we can choose $\Theta(\mathfrak{a})\in K$ such that $\Theta(\mathfrak{a})\Theta(\mathfrak{b})^{\sigma(\mathfrak{a})}\Theta(\mathfrak{a}\mathfrak{b})^{-1}\in k$, where $\sigma(\mathfrak{a})=(K/\mathfrak{a})$ is the Artin symbol for $\mathfrak{a}$ (T. Tannaka, *Ann. Math.*, 67 (1958)). This result can also be generalized for the class field for $S(\mathfrak{f})$.

E. Theory of Genera

Let K/k be a Galois extension and let $H(\mathfrak{m})$ be an ideal group of k. The set of all ideals $\mathfrak{A}$ of K relatively prime to $\mathfrak{m}$ such that $N_{K/k}(\mathfrak{A})$ belongs to $H(\mathfrak{m})$ forms an ideal group of K. This ideal group is said to be the **principal genus** for H. Each coset of $\mathfrak{J}(\mathfrak{m})$ modulo the principal genus for H is said to be a **genus** for H. In particular, let K/k be a cyclic extension with the conductor $\mathfrak{f}$, and let $H(\mathfrak{f})$ be the ideal group of k generated by $N_{K/k}(A)$ $(A\in K)$ and $S(\mathfrak{f})$. Then the principal genus for $H(\mathfrak{f})$ is the ideal group formed by the ideal classes of K of the form $C^{1-\sigma}$, where σ is a generator of the Galois group of K/k (→ 347 Quadratic Fields F). In general, let K/k be an Abelian extension, and let $\mathfrak{f}$ be the conductor of K/k. Then for the ideal $\mathfrak{F}=\prod\mathfrak{P}^{v}$ of K defined in Section D, $N_{K/k}(S(\mathfrak{m}\mathfrak{F}))=S(\mathfrak{m}\mathfrak{f})$ for an arbitrary integral ideal $\mathfrak{m}$ of k. In particular, let K/k be a cyclic extension, and let $H=S(\mathfrak{m}\mathfrak{f})$. Then the principal genus for H is the ideal group consisting of all cosets of $\mathfrak{J}(\mathfrak{m}\mathfrak{F})$ modulo $S(\mathfrak{m}\mathfrak{F})$ of the form $B^{1-\sigma}$ (Herbrand; Iyanaga, *J. Reine Angew. Math.*, 171 (1934)).

F. Class Field Tower Problem and Construction Problem

Furtwängler considered the following problem: Let k be a given algebraic number field, $k=k_0\subset k_1\subset k_2\ldots$ be the sequence of fields such that k_i is the absolute class field over k_{i-1}, and K_∞ be the union of all the k_i. Is K_∞ a finite extension of k? The answer is yes if and only if k_n is of class number 1 for some n. This problem is called the **class field tower problem**. Artin remarked that if for every algebraic number field F of degree n we have the inequality $|D_F|>(\pi/4)^{2r_2}(n^n/n!)^2>(\pi e^2/4)^n/(2\pi n e^{1/6n})$ for the †discriminant D_F, then K_∞/k is always finite [1, p. 46]. E. S. Golod and I. R. Shafarevich (1964) solved the class field tower problem negatively; they proved that K_∞/k is infinite if k_i $(i=1,2,\ldots)$ is the maximal unramified Abelian p-extension of k_{i-1} for a fixed prime number p and if the inequality $\gamma\geqslant 3+2\sqrt{\rho+2}$ holds, where γ is the minimal number of generators of the p-component of the ideal class group of k and ρ is the rank of the unit group of k. (We call an extension K/k a **p-extension** if the degree $[K:k]$ is a power of a prime number p.) For example, the class field tower K_∞/k is actually infinite if k is an imaginary quadratic field ($\rho=1$) and $\gamma\geqslant 7$ for $p=2$, for example, $k=Q(\sqrt{-3\cdot5\cdot7\cdot11\cdot13\cdot17\cdot19})$.

Construction problem. Let k be a given algebraic number field and G a finite group. The construction problem asks us whether there exists a Galois extension K/k such that its Galois group $\mathrm{Gal}(K/k)$ is isomorphic to G. If G is Abelian the problem can be solved affirmatively by using class field theory. This problem was also solved affirmatively for p-groups by A. Scholtz and H. Reinhardt in 1937, and for general solvable groups by Shafarevich in 1954 (*Izv. Akad. Nauk SSSR*, ser. mat. 18).

G. Class Field Theory for Algebraic Function Fields and Local Class Field Theory

F. K. Schmidt developed an analog of class field theory for Abelian extensions over an algebraic function field in one variable with finite coefficient field (1930; → [8]). An arithmetic proof was given by M. Moriya (1938). An analog of class field theory for local fields with finite residue-class fields, called **local class field theory** (→ 257 Local Fields) was first developed by Hasse, and later Chevalley gave an algebraic derivation (1933).

H. Cohomology of Groups and Class Field Theory

For the purpose of simplifying the proof of the main theorems in class field theory, the theory of †Galois cohomology was developed by T. Nakayama, G. Hochschild, A. Weil, Artin, J. Tate, and others. In particular, Artin and Tate [9] constructed class field theory on the basis of the cohomology theory of finite groups as follows: Let G be a finite group, A a †G-module (or a multiplicative commutative group with the operator domain G), and $\hat{H}^n(G,A)$ the nth †cohomology group ($n=0,\pm 1,\pm 2,\dots$) of G with coefficients in A (→ 200 Homological Algebra N). Then we have $\hat{H}^0(G,A)\cong A^G/N_G(A)$, where A^G is the set of all G-invariant elements in A and $N_G(A)$ is the set of all elements of the form $N_G(a)=\sum_{\sigma\in G}\sigma a$ $(a\in A)$. We can consider $\mathbf{Z}$ a G-module by defining $\sigma n=n$ $(n\in\mathbf{Z},\sigma\in G)$. Let A, B, C be G-modules such that a G-bilinear mapping $(A,B)\to C$ is defined. Then we can define the †cup product $(\alpha,\beta)\to\alpha\smile\beta$ $(\alpha\in\hat{H}^r(G,A),\beta\in\hat{H}^s(G,B),\alpha\smile\beta\in\hat{H}^{r+s}(G,C))$ for $r,s\in\mathbf{Z}$ with the usual properties. Let A be a G-module and H a subgroup of G. Then the †restriction homomorphism $\mathrm{R}_{G/H}:\hat{H}^n(G,A)\to\hat{H}^n(H,A)$ and the †injection homomorphism $\mathrm{Inj}_{H/G}:\hat{H}^n(H,A)\to\hat{H}^n(G,A)$ are defined for $n\in\mathbf{Z}$. If H is a normal subgroup of G, then the †inflation homomorphism $\mathrm{Inf}_{(G/H)/G}:\hat{H}^n(G/H,A^H)\to\hat{H}^n(G,A)$ can be defined for $n\geqslant 1$ (→ 200 Homological Algebra M).

Let k be an algebraic number field, and let K be a Galois extension of k of degree n with the Galois group $G=G(K/k)$. The multiplicative group $K^\times=K-\{0\}$, the †idele group J_K of K, and the idele class group C_K of K are multiplicative commutative groups with G as their operator domain. The fundamental formulas in Galois cohomology for class field theory are

$$\hat{H}^1(G,C_K)=0, \tag{1}$$

$$\hat{H}^2(G,C_K))\cong\mathbf{Z}/n\mathbf{Z}. \tag{2}$$

It is possible to realize the isomorphism of (2) by the **invariant** $\mathrm{inv}_{K/k}:\hat{H}^2(G,C_K)\overset{\pi}{\to}\{(r/n)\ (\mathrm{mod}\,\mathbf{Z})\mid r=0,1,\dots,n-1\}$ in such a way that the following properties hold, where the **canonical cohomology class** for K/k is the element $\xi_{K/k}$ of $\hat{H}^2(G,C_K)$ such that $\mathrm{inv}_{K/k}\xi_{K/k}=(1/n)\ (\mathrm{mod}\,\mathbf{Z})$: (i) for $k\subset l\subset K$, $G=G(K/k)$, $H=G(K/l)$, the relation $\mathrm{Res}_{G/H}\xi_{K/k}=\xi_{K/l}$ holds; (ii) if l/k is also a Galois extension with $F=G(l/k)$, then we have $\mathrm{Inf}_{F/G}\xi_{l/k}=\xi_{K/k}^m$ $(m=[K:l])$; (iii) for a cyclic extension K/k we have $\mathrm{inv}_{K/k}\xi_{K/k}=\sum_{\mathfrak{p}}\mathrm{inv}_{\mathfrak{p}}(\xi_{K/k})_{\mathfrak{p}}\ (\mathrm{mod}\,\mathbf{Z})$, where $\mathfrak{p}$ runs over all prime divisors of k and $\mathrm{inv}_{\mathfrak{p}}$ is the invariant in the local theory (→ 257 Local Fields E). By these properties, the canonical cohomology class $\xi_{K/k}$ is uniquely determined. After these preliminaries we can state Tate's theorem, from which the fundamental theorems in class field theory follow.

Tate's theorem. Let K/k be a Galois extension with the Galois group G. Then we have the isomorphism $\Phi_n:\hat{H}^{n-2}(G,\mathbf{Z})\cong\hat{H}^n(G,C_K)$ $(n=0,\pm 1,\pm 2,\dots)$ that is given explicitly by $\Phi_n(\alpha)=\xi_{K/k}\smile\alpha$, where $\xi_{K/k}\in\hat{H}^2(G,C_K)$ is the canonical cohomology class for K/k (*Ann. Math.*, (2) 56 (1952)).

Corollary 1. Since $\hat{H}^{-2}(G,\mathbf{Z})\cong G/[G,G]$ and $\hat{H}^0(G,C_K)\cong C_k/N_{K/k}(C_K)$, we have the isomorphism $\Phi_0:G/[G,G]\cong C_k/N_{K/k}(C_K)$. Let $f(\tau,\sigma)$ $(\tau,\sigma\in G)$ be a 2-†cocycle belonging to $\xi_{K/k}$. Then by the explicit expression for the cup product we obtain the isomorphism

$$\Phi_0:\sigma\left(\mathrm{mod}[G,G]\right)\to\left(\prod_{\tau\in G}f(\tau,\sigma)^{-1}(\mathrm{mod}\ N_{K/k}(C_K)\right).$$

This is an analog of the result in local theory that was proved earlier by T. Nakayama and Y. Akizuki (*Math. Ann.*, 112 (1936)).

Corollary 2. For an Abelian extension K/k we have the isomorphism $\Phi_0:G\cong C_k/N_{K/k}(C_K)$. Φ_0^{-1} has the property of being the norm-residue symbol for C_k, and from this isomorphism we can prove immediately Artin's law of reciprocity. Thus we can prove the main theorems in class field theory by cohomology-theoretic methods [9].

We can also see, by generalizing this isomorphism to infinite Abelian extensions, that the Galois group of the maximal Abelian extension of k over the ground field k with †Krull topology is algebraically and topologically isomorphic to C_k/D_k, where D_k is the connected component of the unit element in C_k. The structures of D_k and C_k/D_k were explicitly determined by Artin and T. Kubota, respectively (→ 6 Adeles and Ideles D).

If we assume the fundamental formulas (1) and (2) stated above and several other simple assumptions as axioms for an infinite ex-

tension of a fixed ground field, we can develop the results stated in this section purely cohomology-theoretically. Such a system is called a **class formation** (Artin; → [9]). In addition to the cases of algebraic number fields, algebraic function fields in one variable with finite coefficient fields, and local fields with finite residue-class fields, which we have mentioned already, we also know several other cases for which analogies of class field theory are valid. These analogies can be explained systematically by using class formation theory (Y. Kawada, *Duke Math. J.*, 22 (1955)). Examples are (1) the theory of unramified Abelian extensions of an algebraic function field in one variable with algebraically closed constant field of characteristic 0 (Tate and Kawada, *Amer. J. Math.*, 77 (1955)); (2) the theory of Kummer extensions over a field k such that (i) the characteristic of k is 0, (ii) k contains all the roots of unity, and (iii) for any Galois extension K/k, $N_{K/k}(K)=k$; (3) the theory of Abelian p-extensions of a field of characteristic p (E. Witt, *J. Reine Angew. Math.*, 176 (1963); I. Satake and Kawada, *J. Fac. Sci. Univ. Tokyo*, 7 (1955)); (4) the theory of unramified Abelian p-extensions of an algebraic function field in one variable with algebraically closed constant field of characteristic p (Hasse and Witt, *Monatsh. Math.*, 43 (1936), H. L. Schmid, I. R. Shafarevich, Kawada, T. Tamagawa); and (5) the theory of Abelian extensions of a local field with algebraically closed residue-class fields (J.-P. Serre, *Bull. Soc. Math. France*, 89 (1961)).

An analogy of class field theory for infinite Abelian extensions was considered by Herbrand, Moriya, M. Mori, and Kawada.

References

[1] H. Hasse, Bericht über neuere Untersuchungen und Probleme aus der Theorie der algebraischen Zahlkörper, Jber. Deutsch. Math. Verein. I, 35 (1926), 1–55; Ia, 36 (1927), 231–311; II, 1930 (Physica Verlag, 1965).
[2] H. Hasse, Vorlesungen über Klassenkörpertheorie, Lecture notes, Univ. Marburg, 1932 (Physica Verlag, 1967).
[3] C. Chevalley, Sur la théorie du corps des classes dans les corps finis et les corps locaux, J. Fac. Sci. Univ. Tokyo, 2 (1933), 365–476.
[4] J. Herbrand, Le dévelopement moderne de la théorie des corps algébriques, Gauthier-Villars, 1936.
[5] E. Artin, Algebraic numbers and algebraic functions, Lecture notes, Princeton Univ., 1950–1951 (Gordon & Breach, 1967).
[6] C. Chevalley, La théorie du corps de classes, Ann. Math., (2) 41 (1940), 394–418.
[7] A. Weil, Sur la théorie du corps de classes, J. Math. Soc. Japan, 3 (1951), 1–35.
[8] F. K. Schmidt, Die Theorie des Klassenkörper über einem Körper algebraischer Funktionen in einer Unbestimmten und mit endlichem Konstantenbereich, S.-B. Phy.-Med. Soz. Erlangen, 62 (1931), 267–284.
[9] E. Artin and J. Tate, Class field theory, Princeton Univ. Press, 1951 (Benjamin, 1967).
[10] S. Iyanaga, Class field theory notes, Univ. of Chicago, 1961.
[11] J.-P. Serre, Groupes algébriques et corps de classes, Actualités Sci. Ind., Hermann, 1959.
[12] J.-P. Serre, Corps locaux, Actualités Sci. Ind., Hermann, 1962; English translation, Local fields, Springer, 1979.
[13] A. Weil, Basic number theory, Springer, 1967.
[14] J. W. S. Cassels and A. Fröhlich (eds.), Algebraic number theory, Academic Press, 1967.
[15] G. Shimura, A reciprocity law in nonsolvable extensions, J. Reine Angew. Math., 221 (1966), 209–220.
[16] Y. Ihara, Non-Abelian class fields over function fields in special cases, Actes Congr. Intern. Math., 1970, Nice, Gauthier-Villars, 381–389.
[17] Y. Ihara, Some fundamental groups in the arithmetic of algebraic curves over finite fields, Proc. Nat. Acad. Sci. US, 72 (1975), 3281–3284.
[18] Y. Kawada, Class formations, 1969 Number Theory Institute, Amer. Math. Soc. Proc. Symposia in Pure Math., 20 (1971), 96–114.
[19] S. Iyanaga (ed.), Theory of numbers, North-Holland, 1975. (Original in Japanese, 1969.)

60 (IV.5)
Classical Groups

A. Introduction

The general linear groups, unitary groups, orthogonal groups, symplectic groups, etc., that are described below are all called **classical groups** (→ 13 Algebraic Groups, 151 Finite Groups, 248 Lie Algebras, 249 Lie Groups).

B. General Linear Groups

Let V be a †linear space of dimension n over a †field K, and let $GL(V)$ denote the set of all †linear mappings of V onto V (hence they are

all bijections). Then $GL(V)$ is a group under the composition of mappings. This group is called the **general linear group** (or **full linear group**) on V. Let $e_1, \ldots, e_n$ be a basis of V over K, and let (α_j^i) be the matrix associated with an element A of $GL(V)$: $Ae_i = \sum_j \alpha_i^j e_j$. Then the mapping $A \rightarrow (\alpha_j^i)$ is an isomorphism of $GL(V)$ onto the multiplicative group $GL(n, K)$ of all $n \times n$ †invertible matrices over K. We can thus identify the group $GL(V)$ with $GL(n, K)$. $GL(n, K)$ is called the **general linear group of degree n over K**. Consider the homomorphism $A \rightarrow |A|$ ($|A|$ is the determinant of A) of $GL(V)$ onto the multiplicative group $K^* = K - \{0\}$. Its kernel $SL(V)$ is a normal subgroup of $GL(V)$ and is called the **special linear group** (or **unimodular group**) on V. The subgroup $SL(n, K) = \{A \mid A \in GL(n, K), |A| = 1\}$ of $GL(n, K)$ corresponds to $SL(V)$ under the above isomorphism $GL(V) \cong GL(n, K)$. $SL(n, K)$ is called the **special linear group of degree n over K**. Unless $n = 2$ and K is the †finite field $\mathbf{F}_2 = GF(2)$, $SL(n, K)$ is the †commutator subgroup of $GL(n, K)$. The †center $\mathfrak{z}$ of $GL(n, K)$ coincides with the set of all scalar matrices αI ($\alpha \in K^*$), and the center $\mathfrak{z}_0$ of $SL(n, K)$ is a finite group given by $\mathfrak{z} \cap SL(n, K) = \{\alpha I \mid \alpha \in K, \alpha^n = 1\}$.

Now let $P(V)$ be the †projective space of dimension $n-1$ obtained from a linear space V of dimension n. Namely, $P(V)$ is the set of all linear subspaces of dimension 1. Then there exists a natural homomorphism φ of $GL(V)$ into the group of all projective transformations of $P(V)$, and the †kernel of φ coincides with the center $\mathfrak{z}$ of $GL(V)$. Hence $\varphi(GL(V)) \cong GL(V)/\mathfrak{z}$. This group is written as $PGL(V)$ and is called the **projective general linear group** on $P(V)$. Similarly, $PGL(n, K) = GL(n, K)/\mathfrak{z}$ is called the **projective general linear group of degree n over K**. The quotient group $SL(n, K)/\mathfrak{z}_0$ of $SL(n, K)$ by the center $\mathfrak{z}_0$ is called the **projective special linear group** and is written as $PSL(n, K)$ or $LF(n, K)$ (**linear fractional group**).

The groups $GL(n, K)$, $SL(n, K)$, etc., are also written as $GL_n(K)$, $SL_n(K)$, etc. In particular, when K is the †finite field $\mathbf{F}_q$, these groups are denoted by $GL(n, q)$, $SL(n; q)$, $PGL(n, q)$, $PSL(n, q)$, $LF(n, q)$.

Simplicity of $PSL(n, k)$. When $n = 2$ and $K = \mathbf{F}_2$, $PSL(2, 2) \cong \mathfrak{S}_3$ (the †symmetric group of degree 3). When $n = 2$ and $K = \mathbf{F}_3$, $PSL(2, 3) \cong \mathfrak{A}_4$ (the alternating group of degree 4). Except for these cases, the group $PSL(n, K)$ ($n \geqslant 2$) is a noncommutative †simple group (→ 151 Finite Groups I).

Suppose that K is the finite field $\mathbf{F}_q$, and let $\alpha(n, q)$, $\beta(n, q)$, $\gamma(n, q)$, $\delta(n, q)$ denote the orders of $GL(n, q)$, $SL(n, q)$, $PGL(n, q)$, $PSL(n, q)$, respectively. Then we have

$$\alpha(n, q) = (q^n - 1)(q^n - q)\ldots(q^n - q^{n-1}),$$

$$\beta(n, q) = \gamma(n, q) = \alpha(n, q)/(q-1),$$

$$\delta(n, q) = \gamma(n, q)/d,$$

where $d = (n, q-1)$ (the greatest common divisor of n and $q-1$).

C. Properties as Lie Groups

If the ground field K is the field $\mathbf{R}$ of real numbers (the field $\mathbf{C}$ of complex numbers), the above groups are all †Lie groups (†complex Lie groups). In particular, $SL(n, \mathbf{C})$ is a †simply connected, †simple, and †semisimple complex Lie group of type A_{n-1}, and $PSL(n, \mathbf{C})$ is the †adjoint group of the complex simple Lie algebra of type A_{n-1}.

D. Determination of the Rational Representations of $GL(V)$

In Sections D and E, the field K is assumed to be of characteristic 0. Let ρ be a homomorphism of $GL(V) = GL(n, K)$ into $GL(m, K)$ (ρ: $A = (\alpha_j^i) \rightarrow B = (\beta_q^p)$). Then if each β_q^p is a rational function (or polynomial or analytic function) in $(\alpha_1^1, \alpha_2^1, \ldots, \alpha_n^n)$ over K, ρ is called a **rational representation** (or **polynomial** or **analytic representation**) of degree m of $GL(V)$. (We suppose that K is $\mathbf{R}$ or $\mathbf{C}$ when we consider analytic functions.) For example, every rational representation of degree 1 can be expressed as $A \rightarrow |A|^e$ (e is an integer). In particular, if K is the field $\mathbf{C}$ of complex numbers, every analytic representation of $GL(n, \mathbf{C})$ is a rational representation. Since $GL(n, \mathbf{C})$ is the †complexification of the †unitary group $U(n)$, there exists a one-to-one correspondence between the complex analytic representations of $GL(n, \mathbf{C})$ and the continuous representations of $U(n)$; this correspondence preserves equivalence, irreducibility, †tensor product, and direct sum of the representations (→ 249 Lie Groups). Hence, determining the rational representations of $GL(n, \mathbf{C})$ is equivalent to determining the continuous representations of $U(n)$. In the general case, the rational representations of $GL(V) = GL(n, K)$ are all completely reducible. For any rational representation ρ of $GL(V)$, there exists a natural number e such that the representation $\rho': A \rightarrow |A|^e \rho(A)$ is a polynomial representation. Hence in order to determine the rational representations of $GL(V)$, it is sufficient to determine the irreducible polynomial representations of $GL(V)$, which, as described below, can be obtained by decomposing the representations on the †tensor space

$V^m = V \otimes \ldots \otimes V$ of degree m (m copies of V) ($m = 1, 2, \ldots$). For $A \in GL(V)$, define $D_m(A) \in GL(V^m)$ as the tensor product

$$D_m(A) = A \otimes \ldots \otimes A \ (m \text{ copies of } A).$$

Namely, for $v_1, \ldots, v_m \in V$, we have

$$D_m(A)(v_1 \otimes \ldots \otimes v_m) = Av_1 \otimes \ldots \otimes Av_m.$$

The mapping $GL(V) \ni A \to D_m(A) \in GL(V^m)$ is a polynomial representation of degree n^m of $GL(V)$. Now let $\mathfrak{L}(V^m)$ be the †associative algebra of all linear mappings of V^m into V^m (†total matrix algebra), and let $\mathfrak{A}$ be the subalgebra of $\mathfrak{L}(V^m)$ generated by $\{D_m(A) \mid A \in GL(V)\}$. Next, for an element σ of the symmetric group $\mathfrak{S}_m$ of degree m, define $B_\sigma \in GL(V^m)$ by $B_\sigma(v_1 \otimes \ldots \otimes v_m) = v_{\sigma^{-1}(1)} \otimes \ldots \otimes v_{\sigma^{-1}(m)}$. Then the mapping $\sigma \to B_\sigma$ is a representation of $\mathfrak{S}_m$ on V^m. Thus we obtain a representation τ of the †group ring $K[\mathfrak{S}_m]$ of $\mathfrak{S}_m$ over K on V^m: $K[\mathfrak{S}_m] \to \mathfrak{L}(V^m)$. Set $\tau(K[\mathfrak{S}_m]) = \mathfrak{B}$. Then $\mathfrak{A}$ and $\mathfrak{B}$ are †commutors of each other in $\mathfrak{L}(V^m)$, i.e., $\mathfrak{A} = \{X \in \mathfrak{L}(V^m) \mid XB = BX$ (for all $B \in \mathfrak{B}$)$\}$, $\mathfrak{B} = \{X \in \mathfrak{L}(V^m) \mid AX = XA$ (for all $A \in \mathfrak{A}$)$\}$.

Now for a right ideal $\mathfrak{r}$ of $\mathfrak{B}$, let $\mathfrak{r}(V^m)$ be the subspace of V^m composed of all the finite sums of the form $\sum Bx$ ($B \in \mathfrak{r}, x \in V^m$). Then the following statements hold:

(1) $\mathfrak{r}(V^m)$ is invariant under $\mathfrak{A}$; hence it is a subspace of V^m invariant under $GL(V)$. Conversely, for any subspace U of V^m invariant under $GL(V)$, there exists a unique right ideal $\mathfrak{r}$ of $\mathfrak{B}$ such that $U = \mathfrak{r}(V^m)$.

(2) Let $\mathfrak{r}_1, \mathfrak{r}_2$ be right ideals of $\mathfrak{B}$, and put $U_1 = \mathfrak{r}_1(V^m)$, $U_2 = \mathfrak{r}_2(V^m)$. Then $\mathfrak{r}_1 \cong \mathfrak{r}_2$ (as right $\mathfrak{S}_m$-modules) if and only if $U_1 \cong U_2$ (as representation spaces of $GL(V)$).

(3) The mapping $\mathfrak{r} \to \mathfrak{r}(V^m)$ is a lattice isomorphism of the †lattice of right ideals of $\mathfrak{B}$ onto the lattice of $GL(V)$-invariant subspaces of V^m. Hence if $\mathfrak{r} = \mathfrak{r}_1 + \mathfrak{r}_2$ (direct sum), then $U = U_1 + U_2$ (direct sum). Also, $\mathfrak{r}(V^m)$ gives an irreducible representation of $GL(V)$ if and only if $\mathfrak{r}$ is a minimal right ideal of $\mathfrak{B}$.

Since the algebra $K[\mathfrak{S}_m]$ is a †semisimple algebra, $\mathfrak{B}$ can be considered as a two-sided ideal of $K[\mathfrak{S}_m]$. Hence a minimal right ideal $\mathfrak{r}$ of $\mathfrak{B}$ is also a minimal right ideal of $K[\mathfrak{S}_m]$, and the †idempotent element ε which generates $\mathfrak{r}$ is a †primitive idempotent of $K[\mathfrak{S}_m]$. From the theory of symmetric groups (→ 362 Representations H) the primitive idempotents of $K[\mathfrak{S}_m]$ are all given (up to isomorphism) by †Young's diagrams $T(f_1, \ldots, f_k)$ ($f_1 \geqslant f_2 \geqslant \ldots \geqslant f_k > 0$, $m = f_1 + \ldots + f_k$). In this setting, we have

(4) Let $\varepsilon = \varepsilon(f_1, \ldots, f_k)$ be the primitive idempotent determined by Young's diagram $T(f_1, \ldots, f_k)$. Then $\varepsilon K[\mathfrak{S}_m] \subset \mathfrak{B}$ if and only if $k \leqslant n$. In this case, put $\varepsilon K[\mathfrak{S}_m] = \mathfrak{r}$, $\mathfrak{r}(V^m) = \varepsilon(V^m) = V^m(T(f_1, \ldots, f_k))$ and denote the irreducible representation of $GL(V)$ on $V^m(T(f_1, \ldots, f_k))$ by $A \to D(A; f_1, \ldots, f_k)$. We call $(f_1, \ldots, f_k)$ the **signature** of this irreducible representation.

(5) The representation $D(A; f_1, \ldots, f_k)$ is an irreducible polynomial representation of $GL(V)$. Furthermore, for any irreducible polynomial representation ρ of $GL(V)$, there exists a unique $D(A; f_1, \ldots, f_k)$ equivalent to ρ. For example, if $k = 1$, then $f_1 = m$, $\varepsilon = (m!)^{-1} \sum_{\sigma \in \mathfrak{S}_m} \sigma$, and $V^m(T(m))$ is the space of †symmetric tensors of degree m. If $f_1 = \ldots = f_k = 1$, then $k = m$, $\varepsilon = (m!)^{-1} \cdot \sum_{\sigma \in \mathfrak{S}_m} (\operatorname{sgn} \sigma) \sigma$, and $V^m(T(1, \ldots, 1))$ is the space of †alternating tensors of degree m.

(6) Let $\chi(A; f_1, \ldots, f_k)$ be the †character of the irreducible representation $D(A; f_1, \ldots, f_k)$. Then

$$\chi(A; f_1, \ldots, f_k) = \begin{vmatrix} \varepsilon_1^{l_1} & \varepsilon_1^{l_2} & \cdots & \varepsilon_1^{l_n} \\ \cdots & \cdots & \cdots & \cdots \\ \varepsilon_n^{l_1} & \varepsilon_n^{l_2} & \cdots & \varepsilon_n^{l_n} \end{vmatrix} \div \begin{vmatrix} \varepsilon_1^{n-1} & \varepsilon_1^{n-2} & \cdots & \varepsilon_1 & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \varepsilon_n^{n-1} & \varepsilon_n^{n-2} & \cdots & \varepsilon_n & 1 \end{vmatrix},$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are the †eigenvalues of A and $l_1 = f_1 + (n-1)$, $l_2 = f_2 + (n-2), \ldots, l_n = f_n$ (set $f_{k+1} = \ldots = f_n = 0$). Hence the degree d of $D(A; f_1, \ldots, f_k)$ is expressed as

$$d = D(l_1, \ldots, l_n) / D(n-1, \ldots, 1, 0),$$

where $D(x_1, \ldots, x_n) = \prod_{i<j} (x_i - x_j)$.

(7) In particular, denote the character of $D(A; m)$ by $p_m = p_m(A)$. Then they satisfy $|I - zA|^{-1} = p_0 + p_1 z + p_2 z^2 + \ldots$ and

$$\chi(A; f_1, \ldots, f_k) = \begin{vmatrix} p_{f_1} & p_{f_1+1} & \cdots & p_{f_1+(n-1)} \\ p_{f_2-1} & p_{f_2} & \cdots & p_{f_2+(n-2)} \\ \cdots & \cdots & \cdots & \cdots \\ p_{f_n-(n-1)} & p_{f_n-(n-2)} & \cdots & p_{f_n} \end{vmatrix},$$

where we put $f_{k+1} = f_{k+2} = \ldots = f_n = 0$, $p_{-1} = p_{-2} = \ldots = 0$. This matrix is simply written as $|p_{l-(n-1)}, \ldots, p_l|$, with the convention that in each row, we set $l_1 = f_1 + (n-1), \ldots, l_{n-1} = f_{n-1} + 1$, $l_n = f_n$.

E. Determination of the Rational Representations of $SL(V)$

The rational representations of $SL(V)$ are completely reducible. By restricting any irreducible representation $D(A; f_1, \ldots, f_n)$ ($f_1 \geqslant f_2 \geqslant \ldots \geqslant f_n \geqslant 0$) of $GL(V)$ to $SL(V)$, we get an irreducible rational representation $\tilde{D}(A; f_1, \ldots, f_n)$ of $SL(V)$. Furthermore, any irreducible rational representation of $SL(V)$

can be obtained in this way. $\tilde{D}(A; f_1, \ldots, f_n)$ and $\tilde{D}(A; f_1', \ldots, f_n')$ are equivalent representations of $SL(V)$ if and only if $f_i - f_{i+1} = f_i' - f_{i+1}'$ $(i = 1, \ldots, n-1)$.

F. Unitary Groups

The set $U(n)$ of all $n \times n$ †unitary matrices with complex elements is a group under multiplication (→ 269 Matrices). This group $U(n)$ is called the **unitary group** (or **unitary transformation group**) of degree n. The subset of $U(n)$ consisting of all matrices of determinant 1 is a normal subgroup of $U(n)$. This group is called the **special unitary group** and is denoted by $SU(n)$.

$U(n)$ and $SU(n)$ are subgroups of $GL(n, \mathbf{C})$ and $SL(n, \mathbf{C})$, respectively, and can be obtained from these groups through the †unitary restriction. Hence they are both compact, connected Lie groups; in particular, $SU(1)$ is composed only of the identity and $U(1)$ is the multiplicative group of all complex numbers of absolute value 1. The center $\mathfrak{z}$ of $U(n)$ is the set of all diagonal matrices $\lambda I (\lambda \in \mathbf{C}, |\lambda| = 1)$, and we have

$$\mathfrak{z} \cong U(1), \quad \mathfrak{z} \cdot SU(n) = U(n),$$

$$U(n)/SU(n) \cong U(1).$$

Moreover, for $n \geqslant 2$, $SU(n)$ is a simple, semisimple, and simply connected Lie group, which gives one of four infinite series of simple compact Lie groups.

$U(n)/\mathfrak{z}$ is denoted by $PU(n)$ and is called the **projective unitary group**. We have the relations $PU(n) \cong SU(n)/\mathfrak{z} \cap SU(n)$, $\mathfrak{z} \cap SU(n) \cong \mathbf{Z}/n\mathbf{Z}$. Hence $PU(n)$ is locally isomorphic to $SU(n)$.

G. Irreducible Representations of $U(n)$

Restricting the irreducible representation $D(A; f_1, \ldots, f_k)$ of $GL(n, \mathbf{C})$ on $SU(n)$, we obtain a continuous irreducible representation of $SU(n)$, and conversely, all continuous irreducible representations of $SU(n)$ are obtained in this manner. Similarly, any continuous irreducible representations of $U(n)$ are given by $A \rightarrow |A|^e D(A; f_1, \ldots, f_k)$, where e is an integer. Since both $U(n)$ and $SU(n)$ are compact, any continuous representation of these groups can be decomposed into a direct sum of the irreducible representations mentioned above (→ 69 Compact Groups).

The representation theory of $U(n)$ and $SU(n)$ is important as the most typical and concrete example of the representation theory of general compact Lie groups (→ 69 Compact Groups, 248 Lie Algebras, 249 Lie Groups).

H. Unitary Groups over General Fields

A unitary matrix and the unitary group can also be defined over some fields other than the field $\mathbf{C}$ of complex numbers. Namely, let P be a field and K a quadratic extension field of P; for an element ξ of K, let $\bar{\xi}$ be the †conjugate of ξ over P. Then a matrix of degree n with entries in K is called a unitary matrix of K (relative to P) if it leaves invariant the †Hermitian form $\xi_1\bar{\xi}_1 + \xi_2\bar{\xi}_2 + \ldots + \xi_n\bar{\xi}_n$. The multiplicative group consisting of all unitary matrices is called the **unitary group over** K (relative to P) and is denoted by $U(n, K, P)$; its subgroup consisting of all unitary matrices of determinant 1 is called the **special unitary group over** K and is denoted by $SU(n, K, P)$. The quotient group of $SU(n, K, P)$ by its subgroup consisting of all λI $(\lambda^n = 1, |\lambda| = 1)$ is called the **projective special unitary group over** K and is denoted by $PSU(n, K, P)$. In particular, when K and P are the finite fields $\mathbf{F}_{q^2}$ and $\mathbf{F}_q$ $(q = p^m)$, $U(n, K, P)$, $SU(n, K, P)$, $PSU(n, K, P)$ are written simply as $U(n, q)$, $SU(n, q)$, $PSU(n, q)$. Then for $n \geqslant 3$, each $PSU(n, q)$ is a noncommutative simple group, except for $PSU(3, 2)$ (→ 151 Finite Groups I).

I. Orthogonal Groups

The set of all †orthogonal matrices of degree n (with real entries) forms a group under multiplication. This group $O(n)$ is called the **orthogonal group** (or **orthogonal transformation group**) of degree n. The subset of $O(n)$ consisting of all orthogonal matrices of determinant 1 forms a normal subgroup of $O(n)$ of index 2. This group $SO(n)$ (also denoted by O_n^+) is called the **rotation group (special orthogonal group** or **proper orthogonal group**) of degree n. Geometrically, $O(n)$ is the set of all orthogonal transformations leaving a point in Euclidean space of dimension n fixed, and $SO(n)$ is composed of all rotations around the point.

Both $O(n)$ and $SO(n)$ are compact Lie groups, and $SO(n)$ coincides with the connected component of $O(n)$ which contains the identity. For $n = 3$ or $n \geqslant 5$, each $SO(n)$ is a simple and semisimple Lie group. Following the theory of Lie algebras, we divide the set of all $SO(n)$ $(n \geqslant 3$ but $n \neq 4)$ into two classes according as n is even or odd, and we thus get two of the four infinite series of simple and semisimple compact Lie groups (for $SO(4)$, for example, see [1]).

Although $SO(n)$ $(n \geqslant 3)$ is a connected Lie group, it is not simply connected. The simply connected Lie group which is locally isomorphic to $SO(n)$ is called the **spinor group** and is denoted by $Spin(n)$. $SO(n)$ is isomorphic to the

quotient group of $Spin(n)$ by a normal subgroup of order 2. Let $\mathfrak{z}$ be the center of $Spin(n)$. Then $\mathfrak{z} \cong \mathbf{Z}/2\mathbf{Z}$ for odd n, $\mathfrak{z} \cong \mathbf{Z}/4\mathbf{Z}$ for $n \equiv 2 \pmod 4$, and $\mathfrak{z} \cong (\mathbf{Z}/2\mathbf{Z}) \oplus (\mathbf{Z}/2\mathbf{Z})$ for $n \equiv 0 \pmod 4$ (→ 61 Clifford Algebras).

The group $O(n, \mathbf{C})$ of all †complex orthogonal matrices is called the **complex orthogonal group**, and the group $SO(n, \mathbf{C})$ of all matrices in $O(n, \mathbf{C})$ of determinant 1 is called the **complex special orthogonal group**. $SO(n, \mathbf{C})$ $(n \geqslant 3, n \neq 4)$ is a simple and semisimple complex Lie group.

J. Irreducible Representations of Orthogonal Groups

In the same way as for $GL(n, K)$, the irreducible representations of $O(n)$ can be obtained by decomposing the tensor product $D_m(A) = A \otimes \ldots \otimes A$ of m copies of an orthogonal matrix A using Young's diagram. Namely, consider the Young's diagram $T(f_1, f_2, \ldots, f_k)$ such that the sum of the lengths of the first column and of the second column is not greater than n, and call it an $O(n)$ diagram. Then to any $O(n)$ diagram $T = T(f_1, \ldots, f_k)$, there corresponds an absolutely irreducible representation $D^0(A; f_1, f_2, \ldots, f_k)$, and the representations $D^0(A; f_1, f_2, \ldots, f_k)$ corresponding to two distinct $O(n)$ diagrams are mutually inequivalent. $D_m(A)$ can be decomposed into the direct sum of those representations $D^0(A; f_1, f_2, \ldots, f_k)$ such that $f = f_1 + \ldots + f_k$ takes the values $m, m-2, m-4, \ldots$. Furthermore, any continuous irreducible representation of $O(n)$ is equivalent to a $D^0(A; f_1, f_2, \ldots, f_k)$ obtained from some $O(n)$ diagram $T = T(f_1, f_2, \ldots, f_k)$.

In general, two $O(n)$ diagrams T and T' are called **mutually associated diagrams** if the sum of the lengths of their first columns is equal to n and if the lengths of each column other than the first one coincide. In particular, if $T = T(f_1, f_2, \ldots, f_k)$ and $2k = n$, then T is said to be self-associated. The set of all $O(n)$ diagrams can be divided into pairs of mutually associated T, T' (and self-associated $T = T'$). Suppose that we are given mutually associated diagrams T and T' and that the length k of the first column of $T = T(f_1, f_2, \ldots, f_k)$ is not greater than $n/2$. Then the character $\chi_T(A)$ of $D^0(A; f_1, f_2, \ldots, f_k)$ corresponding to T and the character $\chi_{T'}(A)$ of the irreducible representation corresponding to T' are given by

$$\chi_T(A) = |p_{l-(v-1)} - p_{l-(v+1)}, \quad p_{l-(v-2)} - p_{l-(v+2)}, \ldots, p_l - p_{l-2v}|,$$

$$\chi_{T'}(A) = |A| \chi_T(A), \quad v = [n/2],$$

where p_i and $|p_{l-(v-1)} - p_{l-(v+1)}, \ldots|$ have the same meaning as in the formula for the characters of irreducible representations of $GL(n, K)$.

The irreducible representations of $SO(n)$ can be obtained immediately from those of $O(n)$. Namely, if $T = T(f_1, f_2, \ldots, f_k)$ is not self-associated, $D^0(A; f_1, f_2, \ldots, f_k)$ is irreducible as a representation of $SO(n)$, and the representations of $SO(n)$ derived from T and the associated T' coincide. If T is self-associated, $D^0(A; f_1, f_2, \ldots, f_k)$ can be decomposed into two irreducible representations of $SO(n)$ of the same degree over the field of complex numbers. Furthermore, the irreducible representations of $SO(n)$ obtained in this way from different pairs of associated diagrams are mutually inequivalent, while any continuous irreducible representation of $SO(n)$ is equivalent to one of these representations. For the representations of $SO(3)$ (the rotation group of degree 3) → 353 Racah Algebra.

Since $SO(n)$ is isomorphic to the quotient group of $Spin(n)$ by a normal subgroup N of order 2, a continuous representation of $Spin(n)$ which is not the identity representation on N can be considered as a double-valued representation of $SO(n)$. This representation is called the **spin representation** and is important in the field of applied mathematics.

The orthogonal group $O(n)$ consists of all $n \times n$ real matrices which leave invariant the quadratic form $\xi_1^2 + \ldots + \xi_n^2$, while the group of all $n \times n$ real matrices which leave invariant the quadratic form $\xi_1^2 + \ldots + \xi_r^2 - \xi_{r+1}^2 - \ldots - \xi_n^2$ of †signature $(r, n-r)$ is called the **Lorentz group** of signature $(r, n-r)$. The case for $n = 4$ and $r = 3$ is used in special relativity (→ 359 Relativity). Let G_0 be the connected component of the identity of the Lorentz group of signature $(3, 1)$. Then G_0 is called the **proper Lorentz group**. For $\sigma = (g_{ij}) \in G$, we have $|\sigma| = \pm 1$ and $g_{44} \geqslant 1$ or $g_{44} \leqslant -1$. Moreover, we have $G_0 = \{\sigma \mid |\sigma| = 1, g_{44} \geqslant 1\}$, $G/G_0 \cong (\mathbf{Z}/2\mathbf{Z}) \oplus (\mathbf{Z}/2\mathbf{Z})$ (†four group), and $G_0 \cong SL(2, \mathbf{C})/\{\pm I\}$.

K. Orthogonal Groups over General Fields

Orthogonal groups can also be defined over other general fields than the field of real numbers as follows: Fix a †quadratic form $Q(\xi, \xi) = \sum_{i,j=1}^n \alpha_{ij} \xi_i \xi_j$ $(|\alpha_{ij}| \neq 0)$ over a field K. Then a linear transformation of $\xi_i (i = 1, 2, \ldots, n)$ over K which leaves Q invariant is called an **orthogonal transformation** with respect to Q. The set of all orthogonal transformations forms a group. This group is denoted by $O(n, K, Q)$ or simply $O(Q)$ and is called the **orthogonal (transformation) group over K with respect to Q**. In particular, the normal subgroup of all transformations in $O(n, K, Q)$ of determinant 1 is called the **special orthogonal group over K with**

respect to Q and is denoted by $SO(n, K, Q)$ (or simply $SO(Q)$). $O(n)$ and $SO(n)$ are special cases of $O(n, K, Q)$ and $SO(n, K, Q)$, where K is the field $\mathbf{R}$ of real numbers and $Q(\xi, \xi)$ is the unit quadratic form $\xi_1^2+\xi_2^2+\ldots+\xi_n^2$.

Let $\Omega(n, K, Q)$ be the †commutator subgroup of $O(n, K, Q)$. Then this subgroup coincides with the commutator subgroup of $SO(n, K, Q)$. If K is of characteristic $\neq 2$, and if $n \geqslant 5$ and the †index ν of $Q \geqslant 1$, then $\Omega(n, K, Q)/\mathfrak{z}$ ($\mathfrak{z}$ is the center of $\Omega(n, K, Q)$) is a simple group, where $\mathfrak{z}=\{I\}$ or $\mathfrak{z}=\{\pm I\}$ (L. Dickson, J. Dieudonné). Suppose that K is a finite field $\mathbf{F}_q$ (of characteristic $\neq 2$). Then we have $\nu=m$ if $n=2m+1$, and $\nu=m$ or $m-1$ if $n=2m$. Hence $\nu \geqslant 2$ if $n \geqslant 5$. If $\nu=0$ and $K=\mathbf{R}$, we have $\Omega(n, \mathbf{R}, Q)=SO(n)$ and, as mentioned before, $SO(n)/\mathfrak{z}$ is simple for $n \geqslant 5$. The same proposition also holds when K is an †algebraic number field (M. Kneser, 1956). If K is of characteristic 2, then $O(n, K, Q)=SO(n, K, Q)$, $\mathfrak{z}=\{I\}$, and $\Omega(n, K, Q)$ is a simple group in many cases (Dieudonné [5]). For the case where K is a finite field (Dickson) → 151 Finite Groups I.

L. Symplectic Groups

Let $\xi_1, \xi_2, \ldots, \xi_{2n}$ and $\eta_1, \eta_2, \ldots, \eta_{2n}$ be two sets of variables, and suppose that the same linear transformation A over a field K acts on them (from the left). If A leaves the †bilinear form $\sum_{i=1}^{n}(\xi_{2i-1}\eta_{2i}-\xi_{2i}\eta_{2i-1})$ invariant, this linear transformation (or the corresponding matrix) A is called a **symplectic transformation (symplectic matrix)** of degree $2n$. The set of all symplectic transformations (or matrices) of degree $2n$ over K forms a group denoted by $Sp(n, K)$ and called the **symplectic group (symplectic transformation group, complex group**, or **Abelian linear group) over** K.

Any matrix in $Sp(n, K)$ is always of determinant 1, and the center $\mathfrak{z}$ of $Sp(n, K)$ consists of I and $-I$. The quotient group $PSp(n, K)$ of $Sp(n, K)$ by $\mathfrak{z}$ is called the **projective symplectic group over** K. Except for the three cases $n=1$, $K=\mathbf{F}_2$; $n=1$, $K=\mathbf{F}_3$; and $n=2$, $K=\mathbf{F}_2$, the group $PSp(n, K)$ ($n \geqslant 1$) is always simple.

Properties of Symplectic Groups as Lie Groups. When K is the field $\mathbf{C}$ of complex numbers or the field $\mathbf{R}$ of real numbers, $Sp(n, K)$ is a Lie group. The intersection of the **complex symplectic group** $Sp(n, \mathbf{C})$ and the unitary group $U(2n)$, namely, the unitary restriction of $Sp(n, \mathbf{C})$, is denoted by $Sp(n)$ and is called the **unitary symplectic group** (or simply **symplectic group**). $Sp(n, \mathbf{C})$ is a simple and semisimple complex Lie group, and both $Sp(n, \mathbf{R})$ and $Sp(n)$ are simple and semisimple Lie groups. Moreover, $Sp(n)$ is compact and simply connected and gives one of the four series of simple, semisimple and compact Lie groups (→ 249 Lie Groups).

Let $\mathbf{H}^n$ be the linear space of dimension n over the †quaternion field $\mathbf{H}$. Define the inner product of two elements $x=(x_1, \ldots, x_n)$ and $y=(y_1, \ldots, y_n)$ in $\mathbf{H}^n$ by $(x, y)=x_1\bar{y}_1+\ldots+x_n\bar{y}_n$ ($\bar{y}_i$ is the †conjugate quaternion of y_i), and consider the group of all linear transformations which leave this inner product invariant. Then this group is isomorphic to $Sp(n)$. $Sp(n)$ is thus compared with the orthogonal group $O(n)$, which leaves invariant the inner product of a linear space over the field $\mathbf{R}$ of real numbers and with the unitary group $U(n)$, which has the same property over the field $\mathbf{C}$ of complex numbers (C. Chevalley [4, ch. 1]).

M. Irreducible Representations of Symplectic Groups

In the same way as for $GL(n, K)$, the representation $D_m(A)=A \otimes \ldots \otimes A$ (tensor product of m copies of A) of $Sp(n, \mathbf{C})$ can be decomposed into irreducible components using Young's diagram. Namely, for any Young's diagram $T=T(f_1, f_2, \ldots, f_k)$ ($k \leqslant n$) such that the number k of rows is not greater than n, an irreducible representation $D^s(A; f_1, \ldots, f_k)$ of $Sp(n, \mathbf{C})$ is determined. These $D^s(A; f_1, \ldots, f_k)$ are mutually inequivalent, and $D_m(A)$ can be decomposed into the direct sum of representations $D^s(A; f_1, f_2, \ldots, f_k)$ such that $f=f_1+\ldots+f_k$ is equal to any of the values $m, m-2, m-4, \ldots$. The character of $D^s(A; f_1, f_2, \ldots, f_k)$ is given by

$$\chi_T(A)=|p_{l-n+1}, p_{l-n+2}+p_{l-n}, \ldots, p_l+p_{l-2n+2}|,$$

where p_i and $|p_{l-n+1}, p_{l-n+2}+p_{l-n}, \ldots|$ have the same meaning as in the formula for the characters of the irreducible representations of $GL(n, K)$.

For the matrices A in $Sp(n)$, $D^s(A; f_1, f_2, \ldots, f_k)$ gives rise to a continuous irreducible representation of $Sp(n)$. Furthermore, any continuous irreducible representation of $Sp(n)$ is equivalent to a representation $D^s(A; f_1, f_2, \ldots, f_k)$ corresponding to some diagram T.

N. Relations among Various Classical Groups

There are some isomorphisms (homomorphisms) among the classical groups mentioned above. For general fields K → [1, 5]. For finite fields K → 151 Finite Groups I. For $K=\mathbf{R}$ or $\mathbf{C}$, the following isomorphisms hold: $SO(3) \cong SU(2)/\{\pm I\}$, $SU(2) \cong Sp(1)$, $SO(5) \cong Sp(2)/\{\pm I\}$, $SO(6) \cong SO(4)/\{\pm I\}$ (→ 248 Lie Algebras, 249 Lie Groups).

O. Classical Groups over Noncommutative Fields

Let V be a right linear space over a noncommutative field K. Then the set of all linear transformations of V forms a group under the multiplication defined by the composition of mappings. This group $GL(V)$ is called the **general linear group** on V. It is isomorphic to the multiplicative group of all $n \times n$ invertible matrices with entries in K. The commutator subgroups $SL(V)$ and $SL(n, K)$ of $GL(V)$ and $GL(n, K)$, respectively, are called the **special linear group** of degree n on V and over K, respectively. Now, suppose that an element A of $GL(V)$ leaves each element of a subspace U of dimension $n-1$ of V fixed. Choose an element x of V which does not belong to U, and set $Ax \equiv x\alpha \pmod U$; $\alpha \in K$ depends not only on A but also on the choice of x. However, the conjugate class $\dot{\alpha} = \{\lambda\alpha\lambda^{-1} \mid \lambda \in K^*\}$ of α in the multiplicative group K^* of K is determined only by A. In particular, if $\dot{\alpha} = \{1\}$ and $A \neq I$, then A is called a **transvection**. For a †matrix unit E_{ij}, $B_{ij}(\alpha) = I + \alpha E_{ij}$ is a transvection if $i \neq j$ and $\alpha \neq 0$. $SL(V)$ coincides with the subgroup of $GL(V)$ generated by all transvections. This fact also holds when K is a commutative field, except when $n=2$ and $K = \mathbf{F}_2$. In this case, transvections generate the whole $GL(2,2)$, which is isomorphic to the symmetric group $\mathfrak{S}_3$ of degree 3 and does not coincide with the commutator subgroup. The center $\mathfrak{z}$ of $GL(n, K)$ consists of all scalar matrices corresponding to nonzero elements in the center of K. Let C be the commutator subgroup of the multiplicative group K^* of K. Then for $n \geqslant 2$, $GL(n; K)/SL(n, K)$ is isomorphic to K^*/C. This isomorphism can be obtained by appropriately defining, for $A \in GL(n, K)$, an element $\det A$ of K^*/C which is called the **determinant** of A [6, 9]. The center $\mathfrak{z}_0$ of $SL(n, K)$ is $\{\alpha I \mid \alpha^n \in C\}$. The quotient group $PSL(n, K) = SL(n, K)/\mathfrak{z}_0$ is called the **projective special linear group** of degree n over K. If K is a noncommutative field, then $PSL(n, K)$ $(n \geqslant 2)$ is always a simple group [5, 8].

Next, let K be any field (commutative or noncommutative), and let V be a right linear space of dimension n over K. Consider a Hermitian form $f(x, y)$ (→ 256 Linear Spaces) on V relative to an †involution J of K. If for a fixed element ε in the center of K we have $f(x, y) = \varepsilon f(y, x)$, then f is called an **ε-Hermitian form**. For the rest of this article, f is assumed to be an ε-Hermitian form on V. Let W be a subspace of V. If $f(x, y) = 0$ for any $x, y \in W$, then W is called a **totally isotropic subspace**. The largest dimension m of the totally isotropic subspaces of V is called the **index** of f. We always have $2m \leqslant n$. If $f(Ax, Ay) = f(x, y)$ for any $x, y \in V$, then A is called a **unitary transformation** relative to f.

The set $U(n, K, f)$ of all unitary transformations relative to f forms a subgroup of $GL(V)$. This group is called the **unitary group** relative to f. Also, the group $SU(n, K, f) = U(n, K, f) \cap SL(n, K)$ is called the **special unitary group**. When $J = 1$ and $\varepsilon = 1$, a unitary transformation (unitary group) is called an **orthogonal transformation (orthogonal group)**, and $U(n, K, f)$ is written as $O(n, K, f)$. Also, when $J = 1$ and $\varepsilon = -1$, a unitary transformation (unitary group) is called a **symplectic transformation (symplectic group)**, and $U(n, K, f)$ is written as $Sp(n, K)$. In fact, in these cases, for arbitrary choice of f, the corresponding groups are mutually isomorphic.

An ε-Hermitian form f is called an **ε-trace form** if for any $x \in V$, there exists an $\alpha \in K$ which satisfies $f(x, x) = \alpha + \varepsilon\alpha^J$. If $J = 1$, $\varepsilon = -1$ (hence K is commutative) or $\varepsilon = 1$ and K is of characteristic $\neq 2$, then any ε-Hermitian form is an ε-trace form. If f is an ε-trace form, a linear mapping B of any subspace W of V into V such that for any $x, y \in W$, $f(Bx, By) = f(x, y)$ can be extended to an element A of the unitary group $U(n, K, f)$ relative to f (**Witt's theorem**). In particular, $U(n, K, f)$ acts transitively on the maximal totally isotropic subspaces, and their dimensions are equal to the index m of f. Now, let P be a **Pythagorean ordered field** (an ordered field which contains square roots of any positive element). If $K = P$ and $J = 1$, or if $K = P(\sqrt{-1})$, or if the noncommutative field K is a †quaternion algebra over P and J is the operation of †conjugation of K, then for two Hermitian forms f, f', their unitary groups $U(n, K, f)$ and $U(n', K, f')$ are isomorphic if and only if $n = n'$, and the indices of f and f' are equal. In this case, $U(n, K, f)$ can be written as $U(n, m, K)$, where m is the index of f. If the field K is a †quaternion algebra over P and f is an †anti-Hermitian form, there exists an orthogonal basis (e_i) of V such that $f(e_i, e_i) = j$ (quaternion unit), $1 \leqslant i \leqslant n$. Hence, in this case, the unitary group $U(n, K, f)$ relative to f is determined only by n and K.

Suppose that we are given an ε-trace form f over a general field K whose index m is not equal to 0. We exclude the case where $J = 1$ and $\varepsilon = 1$. Then the unitary group $U(n, K, f)$ contains transvections. Let $T(n, K, f)$ denote the subgroup of $U(n, K, f)$ generated by transvections which are unitary transformations. If $m \geqslant 2$, then $T(n, K, f)$ is the commutator subgroup of $U(n, K, f)$. The center W_n of $T(n, K, f)$ coincides with the intersection of $T(n, K, f)$, and the center $\mathfrak{z}$ of $GL(n, K)$. If $n \geqslant 3$ and K contains more than 25 elements, then the quotient group $T(n, K, f)/W_n$ is a simple group [6]. Also, if K is commutative and $n \geqslant 2$, $m \geqslant 1$,

$J \neq 1$, then $T(n,K,f)=SU(n,K,f)$, except for the case where $n=3$, $K=\mathbf{F}_4$.

If K is the field $\mathbf{R}$ of real numbers, the field $\mathbf{C}$ of complex numbers, or the quaternion field $\mathbf{H}$, then $GL(n,K)$, $SL(n,K)$, and $U(n,K,f)$ are all Lie groups. In particular, $SL(n,K)$ and $U(n,K,f)$ are simple Lie groups except in the following three cases: (1) $n=1$, $K=\mathbf{R}$ or $\mathbf{C}$; (2) $n=2$, $K=\mathbf{R}$, $J=1$, $\varepsilon=1$; (3) $n=4$, $K=\mathbf{R}$ or $\mathbf{C}$, $J=1$, $\varepsilon=1$, $m=2$. In cases (1) and (2) they are commutative groups, and in case (3) they are locally direct sums of two noncommutative simple groups.

Suppose that $K=\mathbf{H}$. Since $\mathbf{H}$ contains $\mathbf{C}$ as a subfield, a vector space V of dimension n over $\mathbf{H}$ has the structure of a vector space of dimension $2n$ over $\mathbf{C}$. From this fact, $GL(n,\mathbf{H})$ can be considered as a subgroup of $GL(2n,\mathbf{C})$ in a natural way.

Each of the complex classical simple groups $G=SL(n,\mathbf{C})$, $SO(n,\mathbf{C})$, $Sp(n,\mathbf{C})$ has the structure of an †algebraic group defined over $\mathbf{R}$ (→ 13 Algebraic Groups). The **real forms** of G, i.e., the algebraic subgroups of G whose scalar extension to $\mathbf{C}$ is G, can be realized as $SL(n,K)$, $U(n,K,f)$ corresponding to $K=\mathbf{R}$, $\mathbf{C}$, $\mathbf{H}$. Namely, a real form of a complex classical group G is conjugate in G to one of the following groups: (i) The real forms of $SL(n,\mathbf{C})$: $SL(n,\mathbf{R})$ (type AI); $SL(k,\mathbf{H})$ only for $n=2k$ (type AII); and the special unitary group $SU(n,m,\mathbf{C})$, $0 \leqslant m \leqslant [n/2]$, relative to a Hermitian form f of index m (type AIII). (ii) The real forms of $SO(2n+1,\mathbf{C})$: the proper orthogonal group $SO(2n+1,m,\mathbf{R})$, $0 \leqslant m \leqslant n$, relative to a quadratic form of index m on a space of dimension $2n+1$ (type BI and BII). (iii) The real forms of $SO(2n,\mathbf{C})$: $SO(2n,m,R)$, $0 \leqslant m \leqslant n$ (type DI and DII); and $U(n,H,f)$ relative to an anti-Hermitian form f on $\mathbf{H}$ (type DIII). (iv) The real forms of $Sp(n,\mathbf{C})$: $Sp(n,\mathbf{R})$ (type CI); the unitary group $U(2n,m,\mathbf{H})$, $0 \leqslant m \leqslant n$, relative to a Hermitian form f of index m on $\mathbf{H}$ (type CII); and $Sp(n)$ corresponds to the special case $m=0$. The quotient groups of these real forms by their centers can all be realized as the groups of automorphisms of semisimple algebras with involutions J which commute with J (A. Weil [10]).

References

[1] B. L. van der Waerden, Gruppen von linearen Transformationen, Erg. Math., Springer, 1935 (Chelsea, 1948).
[2] H. Weyl, The classical groups, Princeton Univ. Press, 1939, revised edition, 1946.
[3] N. Bourbaki, Eléments de mathématique, Algèbre, ch. 9, Actualités Sci. Ind., 1272a, Hermann, 1959.
[4] C. Chevalley, Theory of Lie groups I, Princeton Univ. Press, 1946.
[5] J. Dieudonné, Sur les groupes classiques, Actualités Sci. Ind., Hermann, 1948.
[6] J. Dieudonné, La géométrie des groupes classiques, Springer, 1955.
[7] C. Chevalley, The algebraic theory of spinors, Columbia Univ. Press, 1954.
[8] M. Eichler, Quadratische Formen und orthogonale Gruppen, Springer, 1952.
[9] E. Artin, Geometric algebra, Interscience, 1957.
[10] A. Weil, Algebras with involutions and the classical groups, J. Indian Math. Soc., 24 (1960), 589–623.

61 (III.16) Clifford Algebras

A. Definitions and Basic Properties

Let V be an n-dimensional †linear space over a field K, and let Q be a †quadratic form on V. Denote the †tensor algebra over V by $T(V)$, the tensor multiplication by $\otimes$. Let $I(Q)$ be the two-sided ideal of $T(V)$ generated by the elements $x \otimes x - Q(x) \cdot 1$ $(x \in V)$. The resulting †quotient associative algebra $T(V)/I(Q)$ is then denoted by $C(Q)$ and is called the **Clifford algebra** of the quadratic form Q. The elements of $C(Q)$ are called **Clifford numbers**.

The composite of two canonical mappings $\tau: V \to T(V)$, $\sigma: T(V) \to C(Q)$ is a linear injection $\sigma \circ \tau: V \to C(Q)$. Hence we can regard V as a linear subspace of $C(Q)$ via $\sigma \circ \tau$. Then $C(Q)$ is an associative algebra over K generated by 1 and V. Furthermore, $x^2 = Q(x) \cdot 1$ for every x in V.

Indeed, $C(Q)$ is the universal associative algebra with these properties. That is, let A be any associative algebra with a unity element, and let $f: V \to A$ be a linear mapping such that $f(x)^2 = Q(X) \cdot 1$ for every x in V. Then f can be extended uniquely to an algebra homomorphism $\tilde{f}: C(Q) \to A$ with $\tilde{f}(1)=1$. Furthermore, let Φ be the †symmetric bilinear form associated with $Q: \Phi(x,y) = Q(x+y) - Q(x) - Q(y)$, $x, y \in V$. Then $xy + yx = \Phi(x,y) \cdot 1$ for every x, y in V. $C(Q)$ is of dimension 2^n over K. If $e_1, \dots, e_n$ is a basis of V, then

$$1,\ e_i,\ e_ie_j\ (i<j), \dots,\ e_1e_2 \dots e_n$$

form a basis of $C(Q)$. In particular, if $\{e_i\}$ is an orthogonal basis relative to Q, we have

$$e_ie_j = -e_je_i, \quad e_i^2 = Q(e_i) \cdot 1; \qquad i,j = 1, \dots, n, \quad i \neq j. \quad (1)$$

In this case, $C(Q)$ can be defined as an associative algebra (with a unity element) generated by the $\{e_i\}$ together with the defining relations (1). In particular, for $Q=0$, $C(Q)$ is the †exterior algebra (†Grassmann algebra) over V.

B. The Principal Automorphism and the Principal Antiautomorphism of $C(Q)$

There exists a unique automorphism α of the algebra $C(Q)$ such that $\alpha(x)=-x$ for every x in V. This automorphism α is called the **principal automorphism** of $C(Q)$, and we have $\alpha^2=1$. Also, there exists a unique antiautomorphism β of the algebra $C(Q)$ such that $\beta(x)=x$ for every x in V. This antiautomorphism β is called the **principal antiautomorphism** of $C(Q)$, and we have $\beta^2=1$.

For the rest of this article we assume that the †discriminant of Q is $\neq 0$. We also assume for the sake of simplicity that the characteristic of K is $\neq 2$. Let $C^+=C^+(Q)=K\cdot 1+V^2+V^4+\ldots$, and $C^-=C^-(Q)=V+V^3+V^5+\ldots$. Then $C(Q)$ is the direct sum of the linear subspaces $C^+(Q)$ and $C^-(Q)$. Furthermore, $C^+C^+\subset C^+$, $C^+C^-\subset C^-$, $C^-C^+\subset C^-$, and $C^-C^-\subset C^+$. Thus $C(Q)=C^++C^-$ has the structure of a †graded algebra with the index group $\{\pm 1\}$, and C^+ is a subalgebra of $C(Q)$. The elements of $C^+(Q)$, $C^-(Q)$ are called **even elements** and **odd elements**, respectively. We have $\dim C^+(Q)=\dim C^-(Q)=2^{n-1}$.

C. The Structure of $C(Q)$ and $C^+(Q)$

$C(Q)$ and $C^+(Q)$ are both †separable, †semisimple associative algebras over K. Suppose n is even: $n=2r$. Then $C(Q)$ is a †simple algebra with K as its center; the center Z of $C^+(Q)$ is 2-dimensional over K. Let $e_1,\ldots,e_n$ be an orthogonal basis of V. Then 1 and $z=2^re_1\ldots e_n$ form a basis of Z, and we have

$$z^2=2^{2r}(-1)^rQ(e_1)\ldots Q(e_n)=(-1)^rD,$$

where D is the †discriminant of Φ relative to the basis $\{e_i\}$. Thus if $(-1)^rD$ has a square root in K, $Z\cong K\oplus K$ (direct sum), and so $C^+(Q)$ is decomposed into the direct sum of two simple algebras. If $(-1)^rD$ does not have a square root in K, then Z is a field and $C^+(Q)$ is a simple algebra. In particular, if the †index of Q (i.e., the dimension of a maximal †totally singular subspace of V (→ 348 Quadratic Forms)) is r, $C(Q)$ is isomorphic to the †total matrix algebra of degree 2^r over K, and $C^+(Q)$ is isomorphic to the direct sum of two copies of the total matrix algebra of degree 2^{r-1} over K.

Now suppose that n is odd: $n=2r+1$. Then $C^+(Q)$ is a simple algebra with K as its center. (In particular, if Q is of index r, then $C^+(Q)$ is isomorphic to the total matrix algebra of degree 2^r over K.) The center Z of $C(Q)$ is 2-dimensional over K, and we have $C(Q)\cong Z\otimes_K C^+(Q)$. If $e_1,\ldots,e_n$ is an orthogonal basis of V, then 1 and $z=e_1\ldots e_n$ form a basis of Z. Putting $z'=2^{r+1}z$, we have $z'^2=2(-1)^rD$, where D is the discriminant of Φ relative to $\{e_i\}$. Thus if $2(-1)^rD$ has a square root in K, $C(Q)$ is the direct sum of two 2^{2r}-dimensional simple algebras. If $2(-1)^rD$ has no square root in K, then $C(Q)$ is a simple algebra.

D. The Clifford Group

Let G be the set of all invertible elements s in $C(Q)$ such that $sVs^{-1}=V$. Then G forms a group relative to the multiplication of $C(Q)$. This group G is called the **Clifford group** of the quadratic form Q. The subgroup $G^+=G\cap C^+(Q)$ is called the **special Clifford group**. The linear transformation $\varphi(s):x\to sxs^{-1}$ of V induced by $s\in G$ belongs to the †orthogonal group $O(Q)$ of V relative to Q. Moreover, the mapping $s\to\varphi(s)$ is a homomorphism from G into $O(Q)$. Thus φ is a †representation of G on V. This representation φ is called the **vector representation** of G. The †kernel of φ consists of invertible elements in the center Z of $C(Q)$. If $x\in G\cap V$, then $Q(x)\neq 0$ and $-\varphi(x)$ is the reflection mapping of V relative to the hyperplane orthogonal to x. If $n=\dim V$ is odd, $\varphi(G)=\varphi(G^+)=SO(Q)$. If n is even, $\varphi(G)=O(Q)$, $\varphi(G^+)=SO(Q)$.

Exploiting the principal antiautomorphism β of $C(Q)$, we obtain a homomorphism $N:G^+\to K^*$ (the multiplicative group of K) defined by $N(s)=\beta(s)s$ ($s\in G^+$), and $N(s)$ is called the **spinorial norm** of $s\in G^+$. The normal subgroup of G^+ defined as the kernel of N is denoted by G_0^+ and is called the **reduced Clifford group** (of Q). The subgroup $\varphi(G_0^+)$ of $SO(Q)$ is denoted by $O_0^+(Q)$ and is called the **reduced orthogonal group**.

In particular, when the ground field K is the real number field $\mathbf{R}$, $O_0^+(Q)$ coincides with the †identity component of the †Lorentz group $O(Q)$. Furthermore, if Q is definite, $O_0^+(Q)\cong SO(n)$, so that the identity component $Spin(n)$ of G_0^+ is a †simply connected †covering group of $SO(n)$ via the covering homomorphism φ (with each point in $SO(n)$ covered twice). The group $Spin(n)$ is called the **spinor group** of degree n.

E. Spin Representations

In this section we assume that the ground field K is the complex number field $\mathbf{C}$ and that n

$=\dim V \geqslant 3$. Then we have $O_0^+(Q) \cong SO(n, \mathbf{C})$, so G_0^+ is a simply connected covering group of $SO(n, \mathbf{C})$ via the covering homomorphism φ. In this section we denote G_0^+ by $Spin(n, \mathbf{C})$ and call it the **complex spinor group** of degree n. $Spin(n, \mathbf{C})$ is the †complexification (→ 249 Lie Groups) of the compact Lie group $Spin(n)$ and is a complex analytic subgroup of the †complex Lie group $C(Q)^*$ consisting of all invertible elements of $C(Q)$. With the bracket operation $[x, y] = xy - yx$, $C(Q)$ becomes the †Lie algebra of $C(Q)^*$. Furthermore, the Lie subalgebra of $C(Q)$ associated with the complex analytic subgroup $Spin(n, \mathbf{C})$ is given by $\sum_{i<j} \mathbf{C} e_i e_j$, where $e_1, \ldots, e_n$ is an orthogonal basis of V. The spin representations of the group $Spin(n, \mathbf{C})$ are defined as follows:

(1) When n is odd: $n = 2r + 1$. Since $C^+(Q)$ is isomorphic to a total matrix algebra of degree 2^r over $\mathbf{C}$, $C^+(Q)$ has a unique (up to equivalence) †irreducible representation $\tilde{\rho}$, which is of degree 2^r. The restriction of $\tilde{\rho}$ on $Spin(n, \mathbf{C})$ (on $Spin(n)$) defines an irreducible representation ρ of degree 2^r of $Spin(n, \mathbf{C})$ (of $Spin(n)$); ρ is called the **spin representation** of the group $Spin(n, \mathbf{C})$ (of $Spin(n)$). The elements in the representation space of ρ are called **spinors**. Thus we can say that a spinor is a quantity with 2^r components which obey the transformation law according to the spin representation (→ 258 Lorentz Group). This representation ρ defines a representation of the Lie algebra $\mathfrak{so}(n, \mathbf{C})$ of $Spin(n, \mathbf{C})$ (note that $\mathfrak{so}(n, \mathbf{C})$ is a †complex simple Lie algebra of type B_r). This representation of $\mathfrak{so}(n, \mathbf{C})$ is also called the spin representation of $\mathfrak{so}(n, \mathbf{C})$. Note that ρ is not well defined on $SO(n, \mathbf{C})$ or on $SO(n)$; ρ is of valence 2 on $SO(n, \mathbf{C})$ or on $SO(n)$.

(2) When n is even: $n = 2r$. Since $C(Q)$ is isomorphic to a total matrix algebra of degree 2^r over $\mathbf{C}$, $C(Q)$ has a unique (up to equivalence) irreducible representation $\tilde{\rho}$, which is of degree 2^r. The restriction of $\tilde{\rho}$ on $Spin(n, \mathbf{C})$ (on $Spin(n)$) defines a representation ρ of degree 2^r of $Spin(n, \mathbf{C})$ (of $Spin(n)$); ρ is called the spin representation of the group $Spin(n, \mathbf{C})$ (of $Spin(n)$). This representation ρ is, however, not irreducible; ρ is decomposed into the direct sum of two irreducible representations ρ_+ and ρ_-. They are not equivalent to each other, and both are of degree 2^{r-1}. By taking a suitable minimal left ideal L of $C(Q)$ as the representation space of the representation $\tilde{\rho}$, we obtain the representation spaces L^+, L^- of ρ_+, ρ_-, respectively, by putting $L^+ = L \cap C^+(Q)$ and $L^- = L \cap C^-(Q)$. The representation ρ^+ (or ρ^-) is called the **half-spin representation** of the group $Spin(n, \mathbf{C})$ or of the group $Spin(n)$. The elements in the representation space of ρ^+ (or ρ^-) are called **half-spinors**. Again, ρ^+ and ρ^- are not well defined on $SO(n, \mathbf{C})$ or on $SO(n)$. They are of valence 2 on these groups. The representations of the Lie algebra of the Lie group $SO(n, \mathbf{C})$ (note that this Lie algebra is a †complex simple Lie algebra of type D_r) associated with ρ^+, ρ^- are also called **half-spin representations** of this Lie algebra.

References

[1] N. Bourbaki, Eléments de mathématique, Algèbre, ch. 9, Actualités Sci. Ind., 1272a, Hermann, 1959.
[2] C. Chevalley, The algebraic theory of spinors, Columbia Univ. Press, 1954.
[3] M. Eichler, Quadratische Formen und orthogonale Gruppen, Springer, 1952.

62 (XI.9) Cluster Sets

A. Cluster Sets of Functions Meromorphic in an Arbitrary Domain

Let D be an arbitrary †domain in the complex z-plane, Γ its boundary, and E a †totally disconnected closed set contained in Γ. Let $w = f(z)$ be a single-valued †meromorphic function defined in D. Then for each point z_0 in Γ, we can define the following sets related to the mapping $w = f(z)$ in the complex w-plane (or on the complex w-sphere $\hat{\mathbf{C}}$).

The Cluster Set. A value α is called a **cluster value** of $f(z)$ at z_0 if there exists a sequence of points $\{z_n\}$ such that

$$z_n \in D, \quad z_n \to z_0, \quad f(z_n) \to \alpha.$$

The totality $C_D(f, z_0)$ of all the cluster values of $f(z)$ at z_0 is called the **cluster set** of $f(z)$ at z_0 or, more precisely, the **interior cluster set**. It is a nonempty, closed, but not necessarily connected set.

The Boundary Cluster Set. The set of all values α such that there exist a sequence of points $\{\zeta_n\}$ of $\Gamma - \{z_0\}$ (resp. $\Gamma - \{z_0\} - E$) and a sequence of points $\{w_n\}$ in the complex w-plane satisfying

$$\zeta_n \to z_0, \quad w_n \in C_D(f, \zeta_n), \quad w_n \to \alpha$$

is called the **boundary cluster set** of $f(z)$ at z_0 and is denoted by $C_\Gamma(f, z_0)$ (resp. $C_{\Gamma-E}(f, z_0)$). These are closed sets, and

$$C_{\Gamma-E}(f, z_0) \subset C_\Gamma(f, z_0) \subset C_D(f, z_0).$$

If $z_0 \in \Gamma - E$ or z_0 is an isolated point of E, then $C_{\Gamma-E}(f, z_0) = C_\Gamma(f, z_0)$. Furthermore, $C_\Gamma(f, z_0)$

(resp. $C_{\Gamma-E}(f, z_0)$) is empty if and only if z_0 is an isolated boundary point (resp. z_0 is an exterior point of $\Gamma - E$).

Range of Values. The set of values α such that

$$z_n \in D, \quad z_n \to z_0, \quad f(z_n) = \alpha$$

is called the **range of values** of $f(z)$ at z_0 and is denoted by $R_D(f, z_0)$. In other words, $R_D(f, z_0)$ is the set of values α assumed by $f(z)$ infinitely often in any neighborhood of z_0 and is a †G_δ-set.

The Asymptotic Set. Let z_0 be an †accessible boundary point of D. If $f(z)$ converges to a value α as z tends to z_0 along a simple arc in D terminating at z_0, then α is called an **asymptotic value** of f at z_0. The totality $A_D(f, z_0)$ of asymptotic values of f at z_0 is called the **asymptotic set** of f at z_0. If z_0 is an inaccessible boundary point, we let $A_D(f, z_0)$ be the empty set.

B. Iversen-Beurling-Kunugui Theorems

Suppose now that E is empty, and put

$$\Omega = C_D(f, z_0) - C_\Gamma(f, z_0).$$

†Iversen's theorems in the case where D is the unit disk $|z| < 1$ and z_0 is a point on $|z| = 1$ are generalized as follows.
(1) $\Omega - R_D(f, z_0) \subset A_D(f, z_0)$ (K. Noshiro, 1936).
(2) First Beurling-Kunugui theorem: If z_0 is not an isolated boundary point, then Ω is an open set.
(3) Second Beurling-Kunugui theorem: Suppose that the open set Ω is not empty. Then $f(z)$ assumes every value belonging to each component Ω_n of Ω infinitely often, with two possible exceptions belonging to Ω_n, that is, $\Omega_n \cap (\hat{\mathbf{C}} - R_D(f, z_0))$ consists of at most two values (an extension of †Picard's theorem on an isolated essential singularity).

Next suppose that E is not empty and of †logarithmic capacity zero, and put $\Omega = C_D(f, z_0) - C_{\Gamma-E}(f, z_0)$. If $\alpha \in \Omega - R_D(f, z_0)$, then either α is an asymptotic value of $f(z)$ at z_0 or there exists a sequence of points $\zeta_n \in E$ ($n = 1, 2, \ldots$) converging to z_0 such that α is an asymptotic value of $f(z)$ at each ζ_n (Noshiro, 1937). Furthermore, if z_0 is contained in the closure of $\Gamma - E$, then Ω is an open set (which may be empty), and $\Omega - R_D(f, z_0)$ is at most of logarithmic capacity zero (M. Tsuji, 1943). If E is contained in a single component Γ_0 of the boundary Γ, z_0 is contained in the closure of $\Gamma - E$, and Ω is nonempty, then $w = f(z)$ assumes every value belonging to each component Ω_n of Ω infinitely often, with two possible exceptions belonging to Ω_n (Noshiro, 1950). In particular, if f is bounded in a neighborhood of z_0, the number of such exceptional values is at most 1. This is still true if each point of E is contained in a boundary component that is a †continuum containing at least two points and Ω is not empty (M. Hervé, 1955). However, this conclusion does not hold if we remove the hypothesis on the set E (K. Matsumoto, 1960).

C. Cluster Sets of Functions Meromorphic in the Unit Disk

Let D be the unit disk $\{|z| < 1\}$, $z_0 = e^{i\theta_0}$ be a fixed point on the unit circumference Γ, A be an open arc of Γ containing z_0, and E be a set of †linear measure zero such that $z_0 \in E \subset A$. With every $e^{i\theta} \in A - E$ we associate an arbitrary simple arc Λ_θ in D terminating at $e^{i\theta}$ and the **curvilinear cluster set** $C_{\Lambda_\theta}(f, e^{i\theta})$, defined as the set of all values α such that $z_n \in \Lambda_\theta$, $z_n \to e^{i\theta}$, $f(z_n) \to \alpha$. We put

$$C^*_{\Gamma-E}(f, z_0) = \bigcap_{r>0} M_r,$$

where M_r denotes the closure of the union $C_{\Lambda_\theta}(f, e^{i\theta})$ for all $e^{i\theta}$ in the intersection of $A - E$ with $|z - z_0| < r$.

By using the above cluster set $C^*_{\Gamma-E}(f, z_0)$ instead of the boundary cluster set $C_{\Gamma-E}(f, z_0)$, we obtain results similar to those in the preceding section (M. Ohtsuka, 1950; Noshiro, 1955). Many interesting results have been obtained by F. Bagemihl and W. Seidel, E. F. Collingwood, O. Lehto and K. I. Virtanen, and others concerning the cluster sets of functions meromorphic in the unit disk. They studied functions of Seidel's class U, **normal** meromorphic functions, and other functions where the class U is the totality of regular bounded functions in the unit disk possessing almost everywhere on $|z| = 1$ radial limits of the constant modulus 1; a nonconstant meromorphic function $f(z)$ in $|z| < 1$ is normal if the family $\{f(T(z))\}$ is †normal in the sense of Montel, where $T(z)$ is an arbitrary conformal mapping of $|z| < 1$ onto itself.

D. The More General Case

The definitions of cluster sets are also available for arbitrary functions for which neither analyticity nor continuity is assumed. If there exist two simple arcs Λ_1 and Λ_2 in the unit disk D terminating at a point $z = e^{i\theta}$ such that

$$C_{\Lambda_1}(f, e^{i\theta}) \cap C_{\Lambda_2}(f, e^{i\theta}) = \varnothing,$$

then $z = e^{i\theta}$ is called an **ambiguous point** of f. Bagemihl proved the following: The set of ambiguous points of an arbitrary complex-valued function defined in the unit disk D is at

most countable [5]. Under the same hypothesis, the set of points $e^{i\theta}$ such that $C_D(f, e^{i\theta}) \neq C_\Gamma(f, e^{i\theta})$ is at most countable (Collingwood, 1960). This result shows the importance of introducing the cluster set $C^*_{\Gamma-E}(f, z_0)$ that was previously mentioned.

E. History

The theory of cluster sets originated from the †value distribution theory of analytic functions in the neighborhood of their essential singularities. The first systematic results were those of F. Iversen and W. Gross, obtained about 1920. Subsequent significant contributions were made by Seidel, J. L. Doob, M. L. Cartwright, A. Beurling, and others. Since 1940, some important results have been obtained by K. Kunugui, S. Irie, Y. Tôki, Y. Tumura, S. Kametani, Tsuji, Noshiro, and other Japanese mathematicians. Many results have been extended to †pseudoanalytic functions. As can be seen from the Bagemihl ambiguous point theorem, some properties of cluster sets are not intrinsic to analytic mappings [2, 5]. On the other hand, it seems to be an interesting problem to extend the theory of cluster sets to the case of analytic mappings between open Riemann surfaces.

References

[1] K. Noshiro, Cluster sets, Erg. Math., Springer, 1960.
[2] E. F. Collingwood and A. J. Lohwater, The theory of cluster sets, Cambridge Univ. Press, 1966.
[3] K. Kunugui, Sur un théorème de MM. Seidel-Beurling, Proc. Japan Acad., 15 (1939), 27–32; Sur un problème de M. A. Beurling, Proc. Japan Acad., 16 (1940), 361–366.
[4] M. Ohtsuka, On the cluster sets of analytic functions in a Jordan domain, J. Math. Soc. Japan, 2 (1950), 1–15.
[5] F. Bagemihl, Curvilinear cluster sets of arbitrary functions, Proc. Nat. Acad. Sci. US, 41 (1955), 379–382.

63 (XVI.3) Coding Theory

A. General Notions of Coding Theory

When we wish to store, to search for, or to send information in the presence of noise efficiently and with the least error, we can apply various bounds to the efficiency (→ 213 Information theory). Sophisticated coding operations are required in order to achieve efficiencies as close as possible to the bounds. In the abstract sense, the information is generally understood to be a choice of an element from a finite set X. For implementation, we take a set K of q elements called **alphabets**. Each element of K is called a **letter**. We consider the direct product K^n, i.e., the set of all sequences of n letters. An injection ψ from X into K^n is called **encoding**. The sequence $\psi(x)$ for $x \in X$ is called a **code word**, and the image $\psi(X)$ (a collection of code words) is called a **code**. Such a code is also called a **block code**. The noise is represented by a mapping $\omega: K^n \to K^n$, but usually it is restricted to a certain subset Ω of the set of such mappings with special properties. For example, we usually assume that the sequences $x \in K^n$ and $\omega(x)$ are different only at less than d letters, where d is a preassigned constant. The inverse mapping φ, i.e., a mapping $\varphi: K^n \to X$ satisfying $\varphi \circ \psi(x) = x$ for all $x \in X$, is called the **decoding** of ψ. A code ψ satisfying the property $\omega \circ \psi(x) \notin \psi(X)$ for all $x \in X$ and for all $\omega \in \Omega$ is called **error-detecting** with respect to the noise Ω. If ψ has the decoding φ satisfying $\varphi \circ \omega \circ \psi(x) = x$ for all $x \in X$ and for all $\omega \in \Omega$, ψ is called **error-correcting** with respect to the noise Ω. To discuss such properties, we can assume $X = \psi(X)$ $(\subset K^n)$ without restricting the generality, so hereafter we assume this condition. Also, we take $q = 2$ unless explicitly stated otherwise. Because in many communication systems $q = 2$ is commonly adopted, and the generalization for other prime powers q can be obtained naturally.

B. Bounds for the Size of Codes

Let $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n) \in K^n$. The **Hamming distance** $d(x, y) = d_H(x, y)$ between the elements x and y is the total number of unequal bits $(x_i \neq y_i)$. We put $d_{\min} = \min\{d_H(x, y) \mid x, y \in X, x \neq y\}$. When $e = \max\{d_H(x, \omega(x)) \mid \omega \in \Omega, x \in X\}$ is less than $d_{\min}$, the errors due to Ω can be detected, and if $d_{\min} \geqslant 2e + 1$, the errors can be corrected. The maximal integer t satisfying $d_{\min} \geqslant 2t + 1$ is often called the **error-correcting capability**. Several important relations are known among $d_{\dim}$, t, n (the length of the code), and $|X|$ (the size of the code word) as follows. **Hamming bound**: $|X| \leqslant 2^n / \sum_{i=1}^{t} \binom{n}{i}$. A code satisfying the equality here is called a **perfect code** [6]. **Plotkin bound**: $d_{\min} \leqslant n|X|/(2(|X|-1))$. On the other hand, if the **Varsharmov-Gilbert-Sacks bound** $2^{n-k} > \sum_{i=0}^{d_{\min}-2} \binom{n-1}{i}$ is satisfied, there exists an (n, k)-linear code (Section C).

C. Linear Codes

Let $K=GF(2)=\{0,1\}$, K^n be an n-dimensional vector space over K, and X be a k-dimensional linear subspace. Then X is called a **group code** or **(n, k)-linear code**. In the present case, we can take a suitable basis of K for which there exists a $k \times (n-k)$ matrix P over K such that every vector $\mathbf{x}=(x_1, \ldots, x_n)$ representing a code word is expressed as $\mathbf{x}=(\mathbf{z}P, \mathbf{z})$ by a suitable vector $\mathbf{z}$, and conversely, every vector of this form is a code word, i.e., a vector $\mathbf{x} \in X$ is a code word if and only if

$$(x_1, \ldots, x_{n-k})=(x_{n-k+1}, \ldots, x_n)P=\mathbf{0}. \qquad (*)$$

Introducing the matrix $H=[I_{n-k}, -P^T]$, where I_{n-k} is the unit matrix of order $n-k$ and ^{T} indicates the transpose, the condition (*) is equivalent to $\mathbf{x}H^T=\mathbf{0}$. Therefore, H is called the **parity check matrix**, $x_{n-k+1}, \ldots, x_n$ are called the **information bits**, and $x_1, \ldots, x_{n-k}$ are called the **check bits**. If x is deformed to $\mathbf{y}=\omega\mathbf{x}$ by a noise ω, i.e., the original signal $\mathbf{x}$ is transmitted as signal $\mathbf{y}$, we call $\mathbf{s}=\mathbf{y}H^T=(\mathbf{y}-\mathbf{x})H^T$ the **syndrome** of the transmitted signal $\mathbf{y}$. If we have an algorithm to determine the error vector $\mathbf{e}=\mathbf{y}-\mathbf{x}$ from the syndrome $\mathbf{s}$, we have a decoding with error-correcting property. For the linear code, it is evident that $d_{\min}=\min\{$the number of nonzero elements of $\mathbf{x} \mid \mathbf{x}\ (\neq 0) \in X\}$.

The **Hamming code** is given by $n=2^m-1$ (m being an integer $\geqslant 2$), $k=n-m$, where the (i,j)-component h_{ij} ($i=1, \ldots, m; j=1, \ldots, n$) of the parity check matrix H is given by the ith bit of the number $j-1$ expressed in the binary (2-adic) number system. The check bits are x_j ($j=2^0, 2^1, \ldots, 2^{m-1}$), and all other bits are the information bits. This code has the error-correcting property with respect to the noise ω for which the nonzero component of $\mathbf{e}=\omega\mathbf{x}-\mathbf{x}$ is at most 1. In fact, from the syndrome $\mathbf{s}=(s_1, \ldots, s_m)$, we compute $\hat{j}=s_1+2s_2+\ldots+2^{m-1}s_m$ and put $e_{\hat{j}}=1$, $e_j=0$ for all $j \neq \hat{j}$; if $\hat{j}=0$, we can put $\mathbf{e}=\mathbf{0}$. The Hamming code is a perfect code.

D. Cyclic Codes

The cyclic code is a special case of a linear code. This is the object of one of the most important applications of the theory of finite fields. By using this theory, we can actually construct the codes with high for error-correcting capability in which the encoding and the decoding operations are performed algebraically.

An (n, k)-linear code X over $GF(2)$ is called a **cyclic code** if $\mathbf{x}=(x_1, x_2, \ldots, x_n) \in X$ implies $(x_n, x_1, x_2, \ldots, x_{n-1}), (x_{n-1}, x_n, x_1, \ldots, x_{n-2}), \ldots, (x_2, \ldots, x_n, x_1) \in X$. To each $\mathbf{x} \in X$, we can associate a polynomial of one variable $x(\xi)=x_1+x_2\xi+\ldots+x_n\xi^{n-1}$ over $GF(2)$. Let $g_X(\xi)$ be the polynomial of the least degree $n-k$ among the polynomials corresponding to $\mathbf{x} \in X$. $g_X(\xi)$ is called the **generator** of X, because the property $x \in X$ and the property that $x(\xi)$ is divisible by $g_X(\xi)$ are equivalent. $g_X(\xi)$ also divides ξ^n-1. Let the quotient be $h_X(\xi)=(\xi^n-1)/g_X(\xi)$. Then $x \in X$ is equivalent to $x(\xi)h_X(\xi) \equiv 0 \pmod{(\xi^n-1)}$. $h_X(\xi)$ is called the **parity check polynomial**. The **BCH code (Bose-Chaudhuri-Hocquenghem code)** defined below is a typical example of a cyclic code.

Let α be an element of order n in $GF(2^m)$ (the finite field with 2^m elements), and let $g_i(\xi)$ be the polynomial of the lowest degree in $GF(2)[\xi]$ for which α^i is a root ($i=1, 3, \ldots, 2t-1$). The BCH code is given by the least common multiple $g(\xi)$ of the polynomials $g_1(\xi), g_3(\xi), \ldots, g_{2t-1}(\xi)$. The BCH code has at least $n-mt$ information bits and satisfies $d_{\min} \geqslant 2t+1$, and algebraic decoding methods are known [3, 5]. The Hamming code is a BCH code when $t=1$, $n=2^m-1$.

E. Other Codes

Other important types of codes include **convolutional codes** (not block codes) for correcting errors in consecutive digits (**burst errors**) and **Goppa codes** (an extension of the BCH class). For correcting burst errors, special cyclic codes or their mixtures are also used. This field of research is closely connected to information theory, algebra, and various applications of combinatorial analysis, such as †experimental design [3].

References

[1] J. Wolfowitz, Coding theorems of information theory, Springer, 1961.
[2] R. G. Gallager, Low-denity parity-check codes, MIT Press, 1963.
[3] E. R. Berlekamp, Algebraic coding theory, McGraw-Hill, 1968.
[4] S. Lin, An introduction to error-correcting codes, Prentice-Hall, 1970.
[5] W. W. Peterson and E. J. Weldon, Jr., Error-correcting codes, MIT Press, second edition, 1972.
[6] A. Tietäväinen, On the non-existence of perfect codes over finite fields, SIAM J. Appl. Math., 24 (1973), 88–96.
[7] E. R. Berlekamp, Goppa codes, IEEE Trans. on Information Theory, IT-19 (1973), 590–592.

64 (IX.8) Cohomology Operations

A. General Remarks

The notion of cohomology operations was introduced by L. S. Pontryagin and N. E. Steenrod in order to solve homotopy classification problems (→ 202 Homotopy Theory). Since then, numerous works have proved the importance of cohomology operations as applied to †homotopy theory, †differential topology, and other branches of topology. In fact, the use of cohomology operations is indispensable in studying problems related to †homotopy groups, †characteristic classes of manifolds, etc.

We denote by the symbol $H^*(X;A)=\sum H^n(X;A)$ the †singular cohomology ring of a topological space X with coefficients in an †Abelian group A.

B. Primary Cohomology Operations

A **(primary) cohomology operation** (or simply an operation) φ is a †natural transformation

$$\varphi:\prod_\lambda H^{l_\lambda}(\ ;A_\lambda)\rightarrow\prod_\mu H^{m_\mu}(\ ;B_\mu)$$

between the cohomology functors defined on the †category of topological spaces and continuous mappings. That is, φ is a family of mappings satisfying the following conditions:

(1) For each space X, φ defines a mapping

$$\varphi:\prod_\lambda H^{l_\lambda}(X;A_\lambda)\rightarrow\prod_\mu H^{m_\mu}(X;B_\mu)$$

that is not necessarily additive.

(2) For each mapping $f:X\rightarrow Y$, the commutativity $f^*\circ\varphi=\varphi\circ f^*$ holds in the diagram

$$\begin{array}{ccc}\prod_\lambda H^{l_\lambda}(X;A) & \xrightarrow{\varphi} & \prod_\mu H^{m_\mu}(X;B)\\ \uparrow f^* & & \uparrow f^*\\ \prod_\lambda H^{l_\lambda}(Y;A) & \xrightarrow{\varphi} & \prod_\mu H^{m_\mu}(Y;B).\end{array}$$

We list here two trivial examples.

(I) Addition of cohomology groups determines an operation $\varphi:H^l(X;A)\times H^l(X;A)\rightarrow H^l(X;A)$.

(II) The †cup product determines an operation

$$\varphi:H^{l_1}(X;A_1)\times H^{l_2}(X;A_2)\rightarrow H^{l_1+l_2}(X;A_1\otimes A_2)$$

denoted by

$$\varphi(\alpha,\beta)=\alpha\smile\beta.$$

The **composite** of two cohomology operations is defined in the obvious way. Among cohomology operations the most important ones are operations of one variable. A cohomology operation of type $(A,l;B,m)$ is a natural transformation

$$\varphi:H^l(\ ;A)\rightarrow H^m(\ ;B).$$

$\mathfrak{O}(A,l;B,m)$ denotes the Abelian group consisting of all such operations.

Denote by $H^m(A,l;B)$ the mth †cohomology group of an †Eilenberg-MacLane space $K(A,l)$, and let $u\in H^l(A,l;A)$ be the †fundamental class of $K(A,l)$. If X is a CW complex, by assigning f^*u to each $f:X\rightarrow K(A,l)$, we obtain a one-to-one correspondence between the set of the homotopy classes $\pi(X,K(A,l))$ and the cohomology group $H^l(X;A)$ (→ 70 Complexes F). Hence by utilizing condition (2), it can be shown that the value of φ on $H^l(X;A)$ is uniquely determined by its operation on $H^l(A,l;A)$. Thus the assignment $\varphi\rightarrow\varphi u$ defines the isomorphism $\mathfrak{O}(A,l;B,m)\cong H^m(A,l;B)$. Here we should remark that the isomorphism $\pi(K(A,l),K(B,m))\cong H^m(A,l;B)$ defines a one-to-one correspondence $\mathfrak{O}(A,l;B,m)\rightarrow\pi(K(A,l),K(B,m))$. In some cases $H^m(A,l;B)$ vanishes, for example, when $0<m<l$; $l=1<m$, $A=\mathbf{Z}$; $l=2$, $m=2m'+1$, $A=\mathbf{Z}$; $l=2l'<m=2m'+1$, $A=\mathbf{Z}\cdot B=\mathbf{Q}$; $l=2l'+1<m$, $A=\mathbf{Z}$, $B=\mathbf{Q}$; A is finite, $B=\mathbf{Q}$; etc.

The following four types of operations together with the two above are called elementary operations:

(III) Homomorphisms induced by a coefficient homomorphism: There are homomorphisms $\eta_*:H^l(X;A)\rightarrow H^l(X;B)$ induced by a homomorphism $\eta:A\rightarrow B$.

(IV) Bockstein (cohomology) operation: This operation is given by the †connecting homomorphisms $\delta^*:H^l(X;C)\rightarrow H^{l+1}(X;A)$ associated with a short exact sequence $0\rightarrow A\rightarrow B\rightarrow C\rightarrow 0$ of coefficient groups (→ 200 Homological Algebra). For example, the coefficient sequence $0\rightarrow\mathbf{Z}\xrightarrow{n}\mathbf{Z}\rightarrow\mathbf{Z}_n\rightarrow 0$ $(\mathbf{Z}_n=\mathbf{Z}/n\mathbf{Z})$ defines a **Bockstein operation** (or **Bockstein homomorphism**), which is usually denoted by $(1/n)\delta$ or Δ_n.

(V) **Steenrod** (or **reduced**) **square operations** Sq^i $(i\geqslant 0)$: Sq^i are sequences of operations defined by the following five axioms [2, 5]:

(V1) For each pair of integers $i\geqslant 0$ and $l\geqslant 0$,

$$Sq^i:H^l(X;\mathbf{Z}_2)\rightarrow H^{l+i}(X;\mathbf{Z}_2)$$

is a natural transformation of functors that is a homomorphism.

(V2) $Sq^0=1$.

(V3) If $\deg x=i$, then $Sq^i x=x\smile x$ (cup product).

(V4) If $\deg x<i$, then $Sq^i x=0$.

(V5) **(Cartan formula)**

$$Sq^i(x\smile y)=\sum_{j=0}^{i}Sq^j x\smile Sq^{i-j}y.$$

These five axioms imply the following two formulas:

(V6) Sq^1 is the Bockstein operation β_2 of the coefficient sequence $0\to\mathbf{Z}_2\to\mathbf{Z}_4\to\mathbf{Z}_2\to 0$.

(V7) **(Adem relations)** If $0<i<2j$, then

$$Sq^i\circ Sq^j=\sum_{k=0}^{[i/2]}\binom{j-1-k}{i-2k}Sq^{i+j-k}\circ Sq^k.$$

The binomial coefficient is taken mod 2.

We can extend the definition of the Sq^i so that they operate on the relative cohomology groups. Now (V1), (V2), and (V5) imply:

(V8) If $\delta: H^l(Y;\mathbf{Z}_2)\to H^{l+1}(X,Y;\mathbf{Z}_2)$ is the †coboundary homomorphism, then $\delta\circ Sq^i= Sq^i\circ\delta$.

(V′) **Steenrod** pth **power operations** $\mathscr{P}^i$ $(i\geqslant 0)$: Let p be an odd prime. Then $\mathscr{P}^i$ is a sequence of operations defined by the following five axioms [5].

(V′1) For each pair of integers $i\geqslant 0$ and $l\geqslant 0$,

$$\mathscr{P}^i: H^l(X;\mathbf{Z}_p)\to H^{l+2i(p-1)}(X;\mathbf{Z}_p)$$

is a natural transformation that is a homomorphism.

(V′2) $\mathscr{P}^0=1$.

(V′3) If $\deg x=2k$, then $\mathscr{P}^k x=x^p$.

(V′4) If $\deg x<2k$, then $\mathscr{P}^k x=0$.

(V′5) **(Cartan's formula)**

$$\mathscr{P}^i(x\smile y)=\sum_{j=0}^{i}\mathscr{P}^j x\smile\mathscr{P}^{i-j}y.$$

These axioms imply the **Adem relations** for $\mathscr{P}^i$ and the Bockstein homomorphism β_p associated with the coefficient sequence $0\to\mathbf{Z}_p\to\mathbf{Z}_{p^2}\to\mathbf{Z}_p\to 0$ (→ Appendix A, Table 6.II).

We can extend the definition of $\mathscr{P}^i$ to the relative cohomology groups too, and we obtain $\delta\circ\mathscr{P}^i=\mathscr{P}^i\circ\delta$.

(VI) **Pontryagin** pth **power operations** $\mathfrak{P}_p$. Let p be a prime. $\mathfrak{P}_p$ is a system of operations satisfying the following five conditions [3]:

(VI1) For each pair of integers $l\geqslant 0$ and $h\geqslant 1$,

$$\mathfrak{P}_p: H^l(X;\mathbf{Z}_{p^h})\to H^{pl}(X;\mathbf{Z}_{p^{h+1}})$$

is a natural transformation.

(VI2) If $\eta:\mathbf{Z}_{p^{h+1}}\to\mathbf{Z}_{p^h}$ is a homomorphism defined by $\eta(1)=1$, then we have $\eta_*\circ\mathfrak{P}_p x=x^p$.

(VI3) If $p:\mathbf{Z}_{p^h}\to\mathbf{Z}_{p^{h+1}}$ is a homomorphism defined by $p(1)=p$, then we have

$$\mathfrak{P}_p(x+y)$$
$$=\mathfrak{P}_p x+\mathfrak{P}_p y+\sum_{i=1}^{p-1}\left(\binom{p}{i}\Big/p\right)p_*(x^i\smile y^{p-i}).$$

(VI4) $\mathfrak{P}_p(x\smile y)=\mathfrak{P}_p x\smile\mathfrak{P}_p y$.

(VI5) If $p>2$ and $\deg x=2k+1$ (odd), then $\mathfrak{P}_p x=0$.

Let A, B be finitely generated Abelian groups. Then the computation of $H^*(A,l)$ shows that each element of $\mathfrak{O}(A,l;B,m)$ can be written as the composite of a finite number of the operations of the types (I)–(VI) (H. Cartan).

Here we show some examples. Let u be the fundamental class of $K(\mathbf{Z}_2,2)$. Then the elements of $H^2(\mathbf{Z}_2,2;\mathbf{Z}_4)\cong\mathbf{Z}_2$, $H^3(\mathbf{Z}_2,2;\mathbf{Z}_4)\cong\mathbf{Z}_2$, $H^4(\mathbf{Z}_2,2;\mathbf{Z}_4)\cong\mathbf{Z}_4$, and $H^5(\mathbf{Z}_2,2;\mathbf{Z}_4)\cong\mathbf{Z}_4+\mathbf{Z}_2$ correspond to the cohomology operations $\lambda 2_*$, $\lambda\eta_*\circ(\delta/2)$, $\mu\mathfrak{P}_2$, and $\mu\eta_*\circ(\delta/4)\circ\mathfrak{P}_2+\lambda 2_*\circ Sq^2\circ Sq^1$, respectively, where λ, μ denote integers satisfying $0\leqslant\lambda\leqslant 1$, $0\leqslant\mu\leqslant 3$ and $\eta:\mathbf{Z}\to\mathbf{Z}_4$, $2:\mathbf{Z}_2\to\mathbf{Z}_4$ stand for the homomorphisms defined by $\eta(1)=1$, $2(1)=2$, respectively.

Suppose that we are given sets of integers l_i and m_j. We define the **stable (primary) cohomology operation** φ with respect to these sets of integers l_i and m_j as a system of natural transformations

$$\varphi:\prod_{\lambda}H^{n+l_\lambda}(\ ;A_\lambda)\to\prod_{\mu}H^{n+m_\mu}(\ ;B_\mu)$$

satisfying the following condition, for all integers $n\geqslant 0$;

(3) Let $S: H^{l+1}(SX;A)\to H^l(X;A)$ denote the †suspension. Then the commutativity $S\circ\varphi=\varphi\circ S$ holds in the diagram

$$\begin{array}{ccc}\prod_{\lambda}H^{n+l_\lambda}(X;A) & \xrightarrow{\varphi} & \prod_{\mu}H^{n+m_\mu}(X;B)\\ \uparrow S & & \uparrow S\\ \prod_{\lambda}H^{n+l_\lambda+1}(SX;A) & \xrightarrow{\varphi} & \prod_{\mu}H^{n+m_\mu+1}(SX;B).\end{array}$$

This condition is equivalent to the commutativity with the coboundary homomorphisms.

For example, the cohomology operations

$$(-1)^n\beta_p: H^n(X;\mathbf{Z}_p)\to H^{n+1}(X;\mathbf{Z}_p)$$

define a stable cohomology operation β. Sq^i, $\mathfrak{P}^i$ are also examples of stable cohomology operations.

A stable cohomology operation φ of type (A,B) and of degree q is a sequence of cohomology operations of type $(A,n;B,n+q)$ defined for all integers $n\geqslant 0$. $\mathfrak{A}(A,B)_q$ denotes the Abelian group consisting of all stable cohomology operations of type (A,B) and of degree q. When $A=B$, $\mathfrak{A}(A)=\sum_{q=0}^{\infty}\mathfrak{A}(A,A)_q$ is a †graded algebra, where multiplication of two operations is given by their composition. Let p be a prime. Then $\mathfrak{A}(\mathbf{Z}_p)$ is called the **Steenrod algebra** mod p and is denoted by $\mathfrak{A}(p)$. $\mathfrak{A}(2)$ is the augmented graded algebra over $\mathbf{Z}_2$ generated by Sq^i subject to the Adem relations. Suppose that we are given a sequence of nonnegative integers $I=(i_1,i_2,\ldots,i_k)$. We call I an **admissible sequence** if $i_{s-1}\geqslant 2i_s$ holds for $2\leqslant s\leqslant k$. We write $Sq^I=Sq^{i_1}\circ Sq^{i_2}\circ\ldots\circ Sq^{i_k}$. If I is an admissible sequence, we say that Sq^I is an **admissible monomial**. The admissible monomials form an additive basis for $\mathfrak{A}(2)$, which

has the structure of a †Hopf algebra whose †comultiplication $\psi:\mathfrak{A}(2)\to\mathfrak{A}(2)\otimes\mathfrak{A}(2)$ is given by $\psi Sq^i=\sum_{j=0}^{i}Sq^j\otimes Sq^{i-j}$. The dual space $\mathfrak{A}(2)^*=\mathrm{Hom}_{\mathbf{Z}_2}(\mathfrak{A}(2),\mathbf{Z}_2)$ gives a Hopf algebra that is the polynomial algebra generated by ξ_i of degree 2^i-1, where ξ_i is the dual element of $Sq^{2^{i-1}}Sq^{2^{i-2}}\ldots Sq^1$ with respect to the additive basis. The comultiplication $\varphi^*:\mathfrak{A}(2)^*\to\mathfrak{A}(2)^*\otimes\mathfrak{A}(2)^*$ is given by $\varphi^*\xi_i=\sum_{j=0}^{i}\xi_{i-j}^{2j}\otimes\xi_j$.

$\mathfrak{A}(p)$ has properties similar to $\mathfrak{A}(2)$ (p is an odd prime). In particular, $\mathfrak{A}(p)$ is a †Hopf algebra generated by $\mathscr{P}^i$ and β subject to the Adem relations with comultiplication ψ given by $\psi\beta=\beta\otimes 1+1\otimes\beta$ and $\psi\mathscr{P}^i=\sum_{j=0}^{i}\mathscr{P}^j\otimes\mathscr{P}^{i-j}$. An additive basis of $\mathfrak{A}(p)$ is given by $\{\beta^{\varepsilon_0}\mathscr{P}^{i_1}\beta^{\varepsilon_1}\ldots\mathscr{P}^{i_k}\beta^{\varepsilon_k}\}$, where $\varepsilon_l=0$ or 1, $i_m\geqslant pi_{m+1}+\varepsilon_m$ $(m\geqslant 1)$ and $i_k\geqslant 1$. Such a sequence $I=(\varepsilon_0,i_1,\varepsilon_1,\ldots,i_k,\varepsilon_k)$ is also called an **admissible sequence**. Denote the dual element of $\mathscr{P}^{p^{i-1}}\mathscr{P}^{p^{i-2}}\ldots\mathscr{P}^1$ and $\mathscr{P}^{p^{i-1}}\mathscr{P}^{p^{i-2}}\ldots\mathscr{P}^1\beta$ by ξ_i and τ_i, respectively; then the dual algebra $\mathfrak{A}(p)^*$ is isomorphic to the tensor product of the polynomial ring over $\mathbf{Z}_p$ generated by ξ_1, $\xi_2,\ldots$ and the exterior algebra over $\mathbf{Z}_p$ generated by $\tau_1,\tau_2,\ldots$. The comultiplication φ^* is given by $\varphi^*(\xi_i)=\sum_{j=0}^{i}\xi_{i-j}^{pj}\otimes\xi_j$ and $\varphi^*(\tau_i)=\tau_i\otimes 1+\sum_{j=0}^{i}\xi_{i-j}^{pj}\otimes\tau_j$ [5] (→ Appendix A, Table 6.III).

C. Secondary Cohomology Operations

Here we restrict our attention to a special type of operation treated by J. F. Adams [4] that has been proved to be powerful in applications.

For a specified prime p, we write $H^+(X)=\sum_{i=1}^{\infty}H^i(X;\mathbf{Z}_p)$. We can regard $H^+(X)$ as a †graded left module over $\mathfrak{A}(p)$. Now let C_s $(s=0,1)$ be a pair of †left free modules over $\mathfrak{A}(p)$, and let $d:C_1\to C_0$ be a graded homomorphism over $\mathfrak{A}(p)$. Suppose that we are given an element $z\in C_1$ such that $dz=0$. For future purposes, C_0 and C_1 are assumed to have bases $\{c_{0,\lambda}\}$ and $\{c_{1,\mu}\}$ with $\deg c_{0,\lambda}=l_\lambda$ and $\deg c_{1,\mu}=m_\mu$, respectively, in terms of which d has the representation $dc_{1,\mu}=\sum_\lambda a_{\mu,\lambda}c_{0,\lambda}$. Then z is expressed in the form $z=\sum_\mu b_\mu c_{1,\mu}$.

We say that Φ is a **stable secondary cohomology operation** associated with the pair (d,z) if it satisfies the following four axioms:

(1) Let $D^n(d,X)$ be the module consisting of homomorphisms $\varepsilon:C_0\to H^+(X)$ over $\mathfrak{A}(p)$ of degree n such that $\varepsilon d=0$. Putting $\varepsilon(c_{0,\lambda})=x_\lambda$, we can assume that $D^n(d,X)=\{\varepsilon=\prod_\lambda x_\lambda\in\prod_\lambda H^{n+l_\lambda}(X)\mid\sum_\lambda a_{\mu,\lambda}x_\lambda=0\}$. Let $Q^n(z,X)$ be the submodule of $H^+(X)$ consisting of elements of the form $\xi(z)$, where ξ is an $\mathfrak{A}(p)$-homomorphism: $C_1\to H^+(X)$ of degree $n-1$. In other words, $Q^n(z,X)=\sum_\mu b_\mu H^{n+m_\mu-1}(X)$. Then for each $n\geqslant 0$ and each X, Φ is a mapping

$$\Phi:D^n(d,X)\to H^+(X)/Q^n(z,X).$$

(2) For each mapping $f:X\to Y$, the commutativity $f^*\circ\Phi=\Phi\circ f^*$ holds in the diagram

$$\begin{array}{ccc}D^n(d,X)&\xrightarrow{\Phi}&H^+(X)/Q^n(z,X)\\ \uparrow f^*&&\uparrow f^*\\ D^n(d,Y)&\xrightarrow{\Phi}&H^+(Y)/Q^n(z,Y).\end{array}$$

(3) Let $S:H^{n+1}(SX)\to H^n(X)$ denote the suspension. Then the commutativity $S\circ\Phi=\Phi\circ S$ holds in the diagram

$$\begin{array}{ccc}D^n(d,X)&\xrightarrow{\Phi}&H^+(X)/Q^n(z,X)\\ \uparrow S&&\uparrow S\\ D^{n+1}(d,SX)&\xrightarrow{\Phi}&H^+(SX)/Q^{n+1}(z,SX).\end{array}$$

(4) Let $i:Y\to X$ be an injection such that $i^*\circ\varepsilon=0$. That is, $i^*x_\lambda=0$. We can then find homomorphisms (over $\mathfrak{A}(p)$) $\eta:C_0\to H^+(X,Y)$ of degree n and $\zeta:C_1\to H^+(Y)$ of degree $n-1$ such that the following diagram commutes:

$$\begin{array}{ccccccccc}H^+(Y)&\xleftarrow{i^*}&H^+(X)&\xleftarrow{j^*}&H^+(X,Y)&\xleftarrow{\delta^*}&H^+(Y)&\xleftarrow{i^*}&H^+(X)\\ &&\uparrow&&\uparrow\eta&&\uparrow\zeta&&\\ &&\llcorner&\xrightarrow{\varepsilon}&C_0&\xleftarrow{d}&C_1&&\end{array}$$

Then for any such pair (η,ζ) we have $i^*\circ\Phi\varepsilon\equiv\zeta z\bmod i^*Q^n(z,X)$, where δ^* denotes the homomorphism defined by $\delta^*y=(-1)^{\dim y}\delta y$.

For each pair (d,z), there is at least one associated operation Φ. Existence is proved by means of a †Postnikov system. Let Φ and Φ' be two operations associated with the same (d,z). Then they differ by a stable primary operation in the sense that there is an element $\varphi\in C_0/dC_1$ such that $\Phi'\varepsilon\equiv\Phi\varepsilon+\varphi\varepsilon\bmod Q(z,X)$, where $\varphi\equiv\sum_\lambda a_\lambda c_{0,\lambda}\bmod dC_1$ means $\varphi\varepsilon=\sum_\lambda a_\lambda x_\lambda$.

For example, define the action of $\mathfrak{A}(2)$ on $\mathbf{Z}_2$ by the rule $a\cdot v=0$ if $\deg a>0$ and $1\cdot v=v$ for each $v\in\mathbf{Z}_2$. We consider a minimal projective resolution

$$0\leftarrow\mathbf{Z}_2\xleftarrow{\varepsilon}C_0\xleftarrow{d_1}C_1\xleftarrow{d_2}C_2\leftarrow\ldots$$

(→ 200 Homological Algebra). First we take $C_0=\mathfrak{A}(2)$ and define $\varepsilon(1)=1$. Next we take C_1 to be free over $\mathfrak{A}(2)$ with generators c_i, where $0\leqslant i$ and $d_1c_i=Sq^{2^i}$. Furthermore, we take C_2 to be free over $\mathfrak{A}(2)$ with generators $c_{i,j}$ with $0\leqslant i\leqslant j$, $j\neq i+1$ and

$$d_2(c_{i,j})=Sq^{2^i}c_j+\sum_{0\leqslant k<j}b_{i,j,k}c_k$$

with $b_{i,j,k}\in\mathfrak{A}(2)$. Here the $d_1c_i=Sq^{2^i}$ form a minimal set of generators of $\mathfrak{A}(2)$, and the equations $0=d_1d_2c_{i,j}=Sq^{2^i}\circ Sq^{2^j}+\sum_{0\leqslant k<j}b_{i,j,k}Sq^{2^k}$ form a minimal set of generators over $\mathfrak{A}(2)$ of the Adem relations. Let $C_1(j)$ be the submodule over $\mathfrak{A}(2)$ generated by c_k with $0\leqslant k\leqslant j$ in C_1, and let $d_1(j)$ be the restriction of d_1 on $C_1(j)$. We write $z_{i,j}=d_2c_{i,j}$.

Let $\Phi_{i,j}$ denote an operation associated with $(d_1(j), z_{i,j})$. Then $\Phi_{i,j}$ is defined on the submodule $D^n_{i,j}(X)$ consisting of elements $x \in H^n(X)$ such that $Sq^{2^k}x=0$ for $0 \leqslant k \leqslant j$ and takes values in $H^{n+2^i+2^j-1}(X)$ modulo the submodule

$$Q^n_{i,j}(X)=Sq^{2^i}H^{n+2^j-1}(X) + \sum_{0\leqslant k<j} b_{i,j,k}H^{n+2^k-1}(X).$$

For example, $\Phi_{0,0}$: $\operatorname{Ker}\beta_2 \to \operatorname{Cok}\beta_2$ is the generalized Bockstein homomorphism defined by $\delta/2^2$ and $\Phi_{1,1}$ is the operation discovered by J. Adem that appears as the †tertiary obstruction of S^n for $n \geqslant 2$. If $k \geqslant 3$, we have the relation

$$Sq^{2^{k+1}} \equiv \sum_{\substack{0\leqslant i\leqslant j\leqslant k\\ j\neq i+1}} a_{i,j,k}\Phi_{i,j} \quad \left(\mathrm{mod} \sum_{\substack{0\leqslant i\leqslant j\leqslant k\\ j\neq i+1}} a_{i,j,k}Q^n_{i,j}(X)\right).$$

These formulas can be applied to prove the nonexistence of an element with †Hopf-invariant 1 in $\pi_{2n-1}(S^n)$ unless $n=1,2,4,8$ (Adams [4]). Analogous formulas also hold for $\mathfrak{A}(p)$ with $p>2$. We have no satisfactory theories concerning cohomology operations of orders higher than 2. For cohomology operations in generalized cohomology theory → 237 K-Theory.

References

[1] N. E. Steenrod, Products of cocycles and extension of mappings, Ann. Math., (2) 48 (1947), 290–320.
[2] J.-P. Serre, Cohomologie modulo 2 des complexes d'Eilenberg-MacLane, Comment. Math. Helv., 27 (1953), 198–232.
[3] N. E. Steenrod and P. E. Thomas, Cohomology operations derived from cyclic groups, Comment. Math. Helv., 32 (1957), 129–152.
[4] J. F. Adams, On the non-existence of elements of Hopf invariant one, Ann. Math., (2) 72 (1960), 20–104.
[5] N. E. Steenrod, Cohomology operations, Ann. Math. Studies, Princeton Univ. Press, 1962.
[6] J. F. Adams, On the structure and applications of the Steenrod algebra, Comment. Math. Helv., 32 (1958), 180–214.
[7] R. E. Mosher and M. C. Tangora, Cohomology operations and applications in homotopy theory, Harper, 1968.
[8] E. H. Spanier, Algebraic topology, McGraw-Hill, 1966.
[9] G. W. Whitehead, Elements of homotopy theory, Springer, 1978.

65 (IX.17) Combinatorial Manifolds

A. Introduction

First of all, we explain some notions in geometric combinatorial topology, also called **PL** (piecewise linear) **topology**, in order to distinguish between the usual topological viewpoint and the combinatorial (or PL) one. By an n-dimensional **simplicial complex** K we mean an n-dimensional, †locally finite and †rectilinear simplicial complex in the Euclidean N-space $\mathbf{R}^N$, i.e., an n-dimensional †Euclidean simplicial complex (→ 70 Complexes). The subspace $P=|K|$ of $\mathbf{R}^N$ covered by all simplexes of K is called an n-dimensional **polyhedron**, and K is referred to as a **simplicial division** or simply a **division** of P. A polyhedron covered by a †subcomplex of a division of P is called a subpolyhedron of P. A polyhedron contained in a polyhedron P is a subpolyhedron of P if and only if it is a closed subset of P. An open subset of a polyhedron P is always a polyhedron but not necessarily a subpolyhedron of P [13]. For arbitrary divisions K_1 and K_2 of a polyhedron there is a common †subdivision of K_1 and K_2. Thus a property of a simplicial complex K which is invariant under subdivision is called a **combinatorial property** of K or of a polyhedron $|K|$. Two polyhedra P and Q are said to be **combinatorially equivalent** if they have †isomorphic divisions K and L, respectively. Then a homeomorphism $f: P \to Q$ which is induced from an isomorphism $f: K \to L$ is called a **combinatorial equivalence**. Since a combinatorial equivalence $f: P \to Q$ is an isomorphism in the category of polyhedra and **PL** (†piecewise linear) **mappings**, (**PL maps**) which is called the **PL category**, it is also called a **PL homeomorphism** or a **PL isomorphism**, and P and Q are said to be PL homeomorphic or PL isomorphic. A †topological space X which is homeomorphic with an n-dimensional polyhedron P is said to be **triangulable** and is referred to as an n-dimensional topological polyhedron. Then a division K of P, or a homeomorphism $t: P \to X$, is called a **triangulation** of X. A division K of P is a triangulation of P. However, there is a compact polyhedron of dimension n ($\geqslant 5$) which has a triangulation K not combinatorially equivalent to its division (→ 70 Complexes; for triangulation problems on manifolds → Section C below). The notion of a simplicial complex or cell complex had been introduced as a tool to describe topological invariants of triangulable manifolds (H. Poincaré [1]). The notion of an n-dimensional combinatorial manifold (i.e., a

polyhedron locally combinatorially equivalent to an n-dimensional simplex) was established as an important geometric object in combinatorial topology by means of J. H. C. Whitehead's theory of regular neighborhoods [2, 1939, 1940]. The 1960s saw remarkable results in the study of combinatorial manifolds together with the topological study of differentiable manifolds, which gave rise to contemporary combinatorial or PL topology [3, 4]. The fundamental conjecture (Hauptvermutung) for combinatorial manifolds and the combinatorial triangulation problem of manifolds were negatively solved by R. C. Kirby and L. C. Siebenmann [5, 1969]. In the 1970s, important results about general triangulations were obtained by T. A. Chapman [6] R. D. Edwards, and J. W. Cannon [7].

B. Pseudomanifolds and Homology Manifolds

In this section we explain some topological properties of triangulable manifolds which can be defined by means of incidence relations among their triangulations. A †polyhedron $M = |K|$ is called an n-dimensional **pseudomanifold** if K satisfies the following three conditions: (i) Every simplex of K is either an n-simplex or a face of an n-simplex; (ii) each $(n-1)$-simplex is a face of at most two n-simplexes; (iii) for any two n-simplexes σ, τ of K, there exists a finite sequence of n-simplexes $\sigma=\sigma_0, \sigma_1, \dots, \sigma_s=\tau$ such that σ_i and σ_{i+1} have an $(n-1)$-face in common.

Consider the set S of $(n-1)$-simplexes of K, each of which is a face of only one n-simplex. Then the set of all $\sigma \in S$ together with their faces forms a subcomplex L of K. The polyhedron $|L|$ of L is called the **boundary** of the pseudomanifold $M=|K|$ and is denoted by ∂M as well as by $\partial K = L$. The boundary of a pseudomanifold is not necessarily a pseudomanifold.

Let $|K|$ be an n-dimensional polyhedron (or †cell complex). A point p in $|K|$ is called a **regular point** of $|K|$ if it has a neighborhood in $|K|$ homeomorphic to an n-dimensional simplex; otherwise, p is called a **singular point**. A pseudomanifold without singular points is a topological manifold, called a **triangulated manifold**. An n-dimensional polyhedron $|K|$ is a pseudomanifold if and only if the set of all regular points in $|K|$ is dense and connected and the set of all the singular points is of dimension less than $n-1$. Thus a connected triangulated manifold is a pseudomanifold whose boundary coincides with the boundary as a topological manifold.

Let τ_1, τ_2 be oriented n-simplexes of an n-dimensional pseudomanifold $M=|K|$. Suppose that they have an $(n-1)$-face σ in common. If the †incidence numbers of these simplexes with σ satisfy the relation $[\tau_1,\sigma] = -[\tau_2,\sigma]$, then τ_1 and τ_2 are said to be **coherently oriented**. We call $M=|K|$ with all its n-simplexes oriented an **oriented pseudomanifold** if any two n-simplexes τ_1 and τ_2 are coherently oriented whenever they have an $(n-1)$-face in common. For an oriented pseudomanifold, the formal sum of all its oriented n-simplexes forms an integral cycle of the pair $(M, \partial M)$, which is called the **fundamental cycle** of $(M, \partial M)$. Its homology class, or fundamental class, generates the homology group $H_n(M, \partial M; \mathbf{Z}) = \mathbf{Z}$. When $M=|K|$ is nonorientable, the **fundamental class with coefficient $\mathbf{Z}_2$** ($\mathbf{Z}_2 = \mathbf{Z}/2\mathbf{Z}$) is similarly defined. A pseudomanifold is **orientable** if and only if the topological manifold consisting of all its regular points is orientable. If for each point p of a polyhedron $M=|K|$ the local homology groups are $H_i(p)=0$ $(i \neq n)$ and $H_n(p)=\mathbf{Z}$ or 0, we call $M=|K|$ an n-dimensional **homology manifold**, where $H_i(p) = H_i(M, M-p; \mathbf{Z})$. If M is connected, it is also a pseudomanifold. The set ∂M of points p of $M=|K|$ with $H_n(p)=0$ is called the **boundary** of M and is an $(n-1)$-dimensional homology manifold without boundary.

Let $M=|K|$ be an n-dimensional homology manifold. Given simplexes σ, τ of K, $\sigma < \tau$ means that σ is a face of τ and $\sigma \neq \tau$. By definition, a vertex of the †barycentric subdivision $\operatorname{Sd} K$ of K is the barycenter of a simplex of K, and a k-simplex τ of $\operatorname{Sd} K$ is defined to be $|\hat{\sigma}_0 \hat{\sigma}_1 \dots \hat{\sigma}_k|$, where $\sigma_0, \sigma_1, \dots, \sigma_k$ are simplexes of K such that $\sigma_0 < \sigma_1 < \dots < \sigma_k$ and $\hat{\sigma}_i$ is the barycenter of σ_i. In the following we denote $\operatorname{Sd} K$ by K'. For a q-simplex σ of K, we denote by $K(\sigma)$ the union of all $(n-q)$-simplexes $|\hat{\sigma}\hat{\sigma}_1 \dots \hat{\sigma}_{n-q}|$ of K' such that $\sigma < \sigma_1 < \dots < \sigma_{n-q}$. Then $K(\sigma)$ is an $(n-q)$-dimensional homology manifold and pseudomanifold contained in M. $K(\sigma)$ is called the $(n-q)$-**dual cell** or simply the **dual** of σ. Let K_σ be the subcomplex of K' consisting of simplexes of K' contained in $K(\sigma)$. Thus we have $|K_\sigma| = |K(\sigma)|$. We denote by $(K(\sigma))^{\cdot}$ the union of simplexes of K_σ which do not intersect with σ. The set of dual cells $K^* = \{K(\sigma) \mid \sigma \in K\}$ of simplexes of K is called the **dual complex** of K; and K^* is also called the **dual subdivision** of $|K|$. K^* satisfies the following conditions:

(i) $\bigcup_{\sigma \in K} K(\sigma) = |K|$. Each $K(\sigma)$ is the †join $\hat{\sigma} * (K(\sigma))^{\cdot}$ of the barycenter $\hat{\sigma}$ and $(K(\sigma))^{\cdot}$.

(ii) $\sigma < \tau \Leftrightarrow (K(\sigma))^{\cdot} \supset K(\tau)$ $(\sigma, \tau \in K)$, $(K(\sigma))^{\cdot} = \bigcup_{\sigma < \tau \in K} K(\tau)$.

(iii) If $K(\sigma) \cap K(\tau) \neq \emptyset$, then there exist simplexes of K with σ and τ as faces, and for the least simplex ρ of K with σ and τ as faces, it holds that $K(\sigma) \cap K(\tau) = K(\rho)$.

The boundary $\partial(K(\sigma))$ of the $(n-q)$-

dimensional homology manifold $K(\sigma)$ satisfies $\partial(K(\sigma))=(K(\sigma))^{\cdot}\cup((\partial K)(\sigma))$, where $(\partial K)(\sigma)$ denotes the dual cell of σ in ∂K, and we understand that $(\partial K)(\sigma)=\varnothing$ if $\sigma\notin\partial K$. $K^*\cup(\partial K)^*$ is called the **dual complex** of $(K,\partial K)$ and is denoted by $(K,\partial K)^*$. For an r-simplex σ $(0\leqslant r\leqslant n-1)$, it holds that

$$H_i((K(\sigma))^{\cdot},(\partial K(\sigma))^{\cdot};\mathbf{Z})=0,\qquad i\neq n-r-1,$$

$$H_{n-r-1}((K(\sigma))^{\cdot},(\partial K(\sigma))^{\cdot};\mathbf{Z})\cong\mathbf{Z}.$$

Let $|K^*|^q$ be the union of i-dual cells of $(K,\partial K)^*=K^*\cup(\partial K)^*$ such that $i\leqslant q$. Let $C_q(K^*)=\tilde{H}_q(|K^*|^q/|K^*|^{q-1})$, and let $\partial_q:C_q(K^*)\rightarrow C_{q-1}(K^*)$ be the †connecting homomorphism $\partial_*:\tilde{H}_q(|K^*|^q/|K^*|^{q-1})\rightarrow\tilde{H}_{q-1}(|K^*|^{q-1}/|K^*|^{q-2})$ in the †reduced homology exact sequence of $(|K^*|^q,|K^*|^{q-1},|K^*|^{q-2})$ (→ 201 Homology Theory F). Then $C_q(K^*)$ is a free Abelian group generated by the q-dual cells of $(K,\partial K)^*$, and $C(K^*)=\{C_q(K^*),\partial_q\}$ is a †chain complex. It follows from the property of $H_*((K(\sigma))^{\cdot},(\partial K(\sigma))^{\cdot};\mathbf{Z})$ as above that $H_q(C(K^*))=H_q(|K|)$ as in the case of a †CW complex.

Now suppose that K is oriented. For a q-simplex $\sigma=|a_0a_1\dots a_q|$ we give an †orientation $[a_0,a_1,\dots,a_q]$. Furthermore, we choose a_{q+1}, $a_{q+2},\dots,a_n$ so that $[a_0,a_1,\dots,a_n]$ is the orientation of $|a_0a_1\dots a_n|$ induced from the orientation of K, and we give $K(\sigma)$ the orientation determined by $[\hat{\sigma},\hat{\sigma}_1,\dots,\hat{\sigma}_{n-q}]$, where $\sigma_i=|a_0a_1\dots a_{q+i}|$. We define $\sigma\cdot K(\sigma)=1$ for σ and $K(\sigma)$ thus oriented, and $\tau\cdot K(\sigma)=0$ if $\tau\neq\sigma$. Let $c=\sum\lambda_i\sigma_i\in C_q(K)$ and $c^*=\sum\mu_jK(\sigma_j)$. Then we define $c\cdot c^*\in\mathbf{Z}$ by $c\cdot c^*=\sum_{i,j}\lambda_i\mu_j(\sigma_i\cdot K(\sigma_j))$. The integer $c\cdot c^*$ is called the **intersection number** of c and c^*. For homology classes $[z]\in H_q(K)$ and $[z^*]\in H_{n-q}(K^*)$, the intersection number $z\cdot z^*$ is independent of the choice of representing cycles, and thus we define the **intersection number** $[z]\cdot[z^*]=z\cdot z^*$.

When X is compact, the isomorphism of the Poincaré-Lefschetz duality theorem $H_q(X,\partial X;\mathbf{R})\cong H^{n-q}(X;\mathbf{R})$ is given in terms of intersection numbers.

C. Combinatorial Manifolds and PL Manifolds

A polyhedron which is PL homeomorphic with a k-dimensional simplex or its boundary is called a **PL k-ball** or a **PL (k-1)-sphere** and is denoted by B^k, S^{k-1}, respectively. A simplicial complex K or a polyhedron $M=|K|$ is called an n-dimensional **combinatorial manifold** if the †star of each vertex in K is a PL k-ball. If M is an n-dimensional combinatorial manifold, then so is an open subset of M. An n-dimensional PL manifold is a topological manifold X with a distinguished maximal coordinate system $\pi=\{(U_i,h_i)\}$, called a **PL structure** on X, which consists of homeomorphisms $h_i:U_i\rightarrow V_i$ from open subsets U_i onto open subsets V_i of an n-dimensional simplex such that $h_j\circ h_i^{-1}:h_i(U_i\cap U_j)\rightarrow h_j(U_i\cap U_j)$ are PL homeomorphisms. For a PL manifold (X,π) there is a triangulation $t:M\rightarrow X$ from a combinatorial manifold M such that for each $(U_i,h_i)\in\pi$, $h_i\circ t|t^{-1}(U_i)$ is a PL homeomorphism from an open subset $t^{-1}(U_i)$ of M onto $h_i(U_i)$. Thus the notions of combinatorial manifold and PL manifold are essentially the same. For a †smooth manifold (X,σ), there is a unique PL homeomorphism class of a combinatorial manifold by means of a †smooth triangulation. In this sense, smooth manifolds are combinatorial manifolds; but the converse of this is not true (→ 114 Differential Topology C). A triangulation of a topological manifold which is a combinatorial manifold is called a **combinatorial triangulation**. The triangulation problem for n-dimensional topological manifolds, especially the existence and uniqueness problem of their combinatorial triangulations, a long-standing problem, is stated as follows. CT_n (**Combinatorial Triangulation Problem**): Is every n-dimensional topological manifold homeomorphic with a combinatorial manifold? CH_n (**Hauptvermutung** for combinatorial manifolds): Homeomorphic n-dimensional combinatorial manifolds should be PL homeomorphic. When $n\leqslant 3$, CT_n and CH_n hold; this was shown for $n=2$ by T. Radó [15] and for $n=3$ by E. E. Moise [16]. Moreover, in this case, topological manifolds admit unique smooth structures, and homology manifolds are combinatorial manifolds. For $n\geqslant 5$, †surgery theory was used and developed as an obstruction theory for "homotoping" a homeomorphism $h:M\rightarrow M'$ between n-dimensional combinatorial manifolds to a PL homeomorphism by D. Sullivan (1967). By the discovery of the torus-unfurling method for "isotoping" $h:M\rightarrow M'$ to a PL homeomorphism, Kirby reduced CT_n and CH_n for $n\geqslant 5$ to CH_n for some specific manifolds, such as thickened tori modulo the boundary, for which surgery methods had been sufficiently developed (for isotopy → Section D). The solutions (1)–(9) can be stated as follows.

(1) Classification of combinatorial triangulations. Assume $n\geqslant 5$. Existence: An n-dimensional closed topological manifold X admits a combinatorial triangulation if and only if an obstruction $\Delta(X)\in H^4(X;\mathbf{Z}_2)$ vanishes. Uniqueness: A homeomorphism $h:M\rightarrow M'$ between n-dimensional closed combinatorial manifolds is isotopic (→ Section D) to a PL homeomorphism if and only if an obstruction $\Delta(h)\in H^3(X;\mathbf{Z}_2)$ vanishes (Kirby and Siebenmann [5]).

In fact, for each $n\geqslant 5$, there is an n-

dimensional combinatorial manifold which is homeomorphic to a torus $S^1 \times \ldots \times S^1$ but is not PL homeomorphic, and there is an n-dimensional closed manifold which is not homeomorphic to a combinatorial manifold (Siebenmann [18]).

As a special case, the following conjecture is proved for $n \geqslant 5$.

(2) The n-dimensional **annulus conjecture**. The closed region bounded by two mutually disjoint †locally flat $(n-1)$-dimensional topological spheres S_1 and S_2 in an n-dimensional sphere S^n is homeomorphic with an n-dimensional annulus $S^{n-1} \times [0,1]$ (Kirby [19]).

Regarding the uniqueness problem for (not necessarily combinatorial) triangulations of topological manifolds, we mention the following deep result.

(3) **Double suspension theorem** for homology spheres. The double suspensions of homology n-spheres are homeomorphic with S^{n+2}, where a homology n-sphere is a closed topological manifold with the same homology groups as those of S^n (Cannon [7]).

For each $n \geqslant 3$, there is a homology n-sphere P^n which is a combinatorial manifold with the nontrivial †fundamental group (for $n=3$, see Poincaré [1, 1904] and for $n \geqslant 4$, see M. Newman, 1948). The suspension of P^n is an example of an $(n+1)$-dimensional homology manifold which is not a topological manifold. The double suspension of P^n is not a combinatorial manifold but is homeomorphic with S^{n+2}. The problem of whether every n-dimensional topological manifold is homeomorphic to a polyhedron is still open for $n \geqslant 4$. This problem is reduced to a problem in homology 3-spheres (T. Matumoto; D. E. Galewski and R. J. Stern).

In the following we list some results on combinatorial manifolds focusing on Whitehead's theory of regular neighborhoods [2]. Let $M=|K|$ be an n-dimensional combinatorial manifold without boundary. The dual complex K^* consists of PL balls. Thus the division K of M gives rise to a decomposition into PL †handles $H(K)$ of M in such a way that for each k-dimensional simplex $\sigma \in K$, the star $\mathrm{St}_{K''}(\hat{\sigma})$ of its barycenter $\hat{\sigma}$ in the second barycentric subdivision K'' of K is a PL handle of index k. If L is a subcomplex of K or K^*, then the star $\mathrm{St}_{K''}(|L|)$ of $|L|$ in K'' is a subhandlebody of $H(K)$ or of its dual handlebody $H^*(K)=H(K^*)$, denoted by $H_K(L)$ or $H_{K^*}(L)$, respectively. In particular, if M is closed, orientable, and of dimension 3, then $U=H_K(K^{(1)})$ and $V=H_{K^*}(K^*_{(1)})$ are PL homeomorphic with a PL boundary-connected sum of some copies of a solid torus $S^1 \times B^2$, and $M=U \cup V$, $U \cap V = \partial U = \partial V$, where $K^{(k)}$ and $K^*_{(k)}$ stand for the k-†skeletons of K and K^*, respectively. This decomposition $(U, V; \partial U = \partial V)$ of M is called a **Heegaard decomposition** (or **splitting**) of M [12]. We say that a simplicial complex L collapses elementarily to a subcomplex L_1 of L, referred to as an **elementary collapse** $L \searrow L_1$, if $L - L_1 = \{\sigma, \tau\}$, where σ is a k-dimensional simplex of L and τ is a $(k-1)$-dimensional face of σ which is not a face of any simplex in $L - \{\sigma, \tau\}$. We say that a polyhedron P collapses to a subpolyhedron Q of P or Q expands to P, referred to as a **collapse** $P \searrow Q$ or **expansion** $Q \nearrow P$, if there is a division L of P containing a subcomplex L_r dividing Q so that there is a sequence of elementary collapsings $L \searrow L_1 \searrow \ldots \searrow L_r$. For a subpolyhedron P of a combinatorial manifold M, a subpolyhedron U of M is called a **regular neighborhood** of P in M if (1) U is a combinatorial manifold which is a closed neighborhood of P in M and (2) $U \searrow P$. In general, for a subcomplex L of a simplicial complex K, the †star $\mathrm{St}_{K''}(|L|)$ of $|L|$ in the second barycentric subdivision K'' of K is called a **second barycentric derived neighborhood** or simply a **derived neighborhood** of $|L|$ in $|K|$ (E. C. Zeeman [4]).

(4) **Regular neighborhood theorem**. Let P be a compact subpolyhedron of a combinatorial manifold M. Existence: Every derived neighborhood of P in M is a regular neighborhood. Uniqueness: Any two regular neighborhoods of P in M are PL homeomorphic by a mapping keeping P pointwise fixed [2, 1939]. On one hand, this theorem is regarded as a combinatorial counterpart to the †tubular neighborhood theorem of differential topology. For a topological manifold X with boundary ∂X, there are a neighborhood U of ∂X in X and a homeomorphism $h: \partial X \times [0,1] \rightarrow U$, called a collar of ∂X in X, such that $h(x,0)=x$ $(x \in \partial X)$. In particular, for a combinatorial manifold $M=X$ we can take U as a derived neighborhood of ∂M in M and h as a PL homeomorphism, called a PL collar of ∂M in M. (For more about normal bundle theory → 147 Fiber Bundles Q). On the other hand, it is also regarded as a combinatorial deformation theorem of a handlebody decomposition $H_{K''}(L)$ of a derived neighborhood $\mathrm{St}_{K''}(|L|)$.

Two polyhedra P and Q are said to be **simple homotopy equivalent** if Q is obtained from P by a finite sequence of collapsings and expansions. A composite mapping $f: P \rightarrow Q$ of the inclusion mappings and deformation retractions associated naturally to the expansions and the collapsings from P to Q is called a **simple homotopy equivalence**. Moreover, if P and Q are subpolyhedra of a combinatorial manifold M and those collapsings and expansions take place in M, then their regular neighborhoods are PL homeomorphic [2, 1940].

(5) A simply homotopy equivalence is clear-

ly a homotopy equivalence. Conversely, we have the **simple homotopy theorem**: For the homotopy class of a homotopy equivalence $f: P \rightarrow Q$ between compact connected polyhedra P and Q, there is a well-defined element $W(f)$, called the **Whitehead torsion** of f, in the †Whitehead group $Wh(G)$ of the fundamental group $G = \pi_1(P)$ such that f is homotopic to a simple homotopy equivalence if and only if $W(f) = 0$, ([2, 1950]; → also J. W. Milnor [20]).

Let W be a compact combinatorial manifold of dimension $n+1$. If ∂W consists of two connected components M and N such that the inclusion mappings $M \subset W$ and $N \subset W$ are simple homotopy equivalences, then $(W; M, N)$ is called an **s-cobordism** of dimension $n+1$. The following theorem of D. Barden, B. Mazur, and J. Stallings is the nonsimply connected PL version of Smale's †h-cobordism theorem.

(6) **s-Cobordism theorem.** For $n \geqslant 5$, an s-cobordism $(W; M, N)$ of dimension $n+1$ is PL homeomorphic with $(M \times [0,1]; M \times 0, M \times 1)$ (B. Mazur [21]; M. Kervaire [22]; C. Weber [23]; J. F. P. Hudson [8]).

(7) Topological invariance of simple homotopy types. Two homeomorphic compact polyhedra are simple homotopy equivalent (T. A. Chapman [6]).

The famous **Poincaré conjecture** that a simply connected closed manifold of dimension 3 is homeomorphic to a 3-sphere is still unsolved despite much effort by many mathematicians. This conjecture is generalized for dimension n as follows.

(8) **Generalized Poincaré conjecture** in dimension n. Any homotopy n-sphere Σ^n is homeomorphic to an n-sphere, where a homotopy n-sphere is defined to be an n-dimensional topological manifold homotopy equivalent to an n-sphere.

This has been affirmatively answered for $n \geqslant 5$. The first proof was given by S. Smale (→ 114 Differential Topology F, K). From the viewpoints of collapsing and general position (→ Section D), J. Stallings [3, 1960] and E. C. Zeeman [11, 198–204] proved the following.

(9) **Engulfing lemma**. For a compact polyhedron P of dimension m in the interior $\mathring{M} = M - \partial M$ of an n-dimensional combinatorial manifold M, there is a PL n-ball in $\mathring{M}$ containing P in its interior (i.e., engulfing P), provided that M is m-connected (i.e., $\pi_i(M) = 0$ $(i \leqslant m)$) and $n - m \geqslant 3$. (For the topological version → M. H. A. Newman [24].) If a combinatorial manifold $M = |K|$ is a homotopy n-sphere $(n \geqslant 5)$, then by the engulfing lemma, there are PL n-balls B_1 and B_2 engulfing $H_{K''}(K^{(n-3)})$ and $H_{K''}(K^*_{(2)})$ so that $\mathring{B}_1 \cup \mathring{B}_2 = M$. By the generalized Schoenflies theorem (→ Section G), $B_2 - \mathring{B}_1 = M - \mathring{B}_1$ is homeomorphic to an n-ball. It follows that by the Alexander trick (→ Section D) M is homeomorphic to an n-sphere. Since CH_n holds for S^n $(n \geqslant 5)$, the homotopy PL n-sphere M is actually PL homeomorphic with S^n.

D. Embeddings of Combinatorial Manifolds

An injective PL mapping $f: P \rightarrow Q$ between polyhedra is called a **PL embedding** if $f(P)$ is a subpolyhedron of Q (→ Section A). Namely, a PL embedding $f: P \rightarrow Q$ is a PL mapping which is induced from a simplicial injection $f: K \rightarrow L$, called a division of f, for some divisions K and L of P and Q, respectively. In the following we explain PL embeddings from an m-dimensional combinatorial manifold M into an n-dimensional combinatorial manifold N. For a PL embedding $f: M \rightarrow N$, $c = n - m$ is called the **codimension** of f. If $f^{-1}(\partial N) = \partial M$, f is said to be **proper**. Since an injective PL mapping $f: M \rightarrow N$ is a PL embedding if and only if it is proper as a mapping between topological spaces M and N (i.e., the preimages of compact sets are compact), this term "proper" for PL embeddings should not be confused with the term "proper" for mappings. Two PL embeddings $f: M \rightarrow N$ and $f': M' \rightarrow N'$ are said to be (PL) **equivalent** if there is a PL homeomorphism $h: N \rightarrow N'$, called an equivalence from f to f', such that $h \circ f(M) = f'(M')$. When $M = M'$, f and f' are said to be (PL) **isomorphic** if there is an equivalence $h: N \rightarrow N'$ from f to f' such that $h \circ f = f'$. A **submanifold** of N is defined as a subpolyhedron of N which is a combinatorial manifold. Thus the equivalence class of a PL embedding $f: M \rightarrow N$ is nothing but the PL homeomorphism class of a pair $(N, f(M))$ of N and its submanifold $f(M)$. For a submanifold of N, the same terms as those for the inclusion mapping are used. For every division $f: K \rightarrow L$ of a (proper) PL embedding $f: M \rightarrow N$ and for $x \in M$ we have a (proper) PL embedding $f | \mathrm{St}_K(x)$ from a PL m-ball $\mathrm{St}_K(x)$ into a PL n-ball $\mathrm{St}_L(f(x))$ which represents the †germ of f at $x \in M$. A PL embedding of a PL m-sphere into a PL n-sphere is called a **PL (n, m)-knot** or (n, m)-sphere pair and a proper PL embedding of a PL m-ball into a PL n-ball is called a **PL (n, m)-ball knot** or (n, m)-ball pair. A PL (n, m)-knot (a PL (n, m)-ball knot) is said to be **trivial** or **unknotted** if it is equivalent to the inclusion mapping $\partial I^{m+1} \times 0 \subset \partial I^{n+1}$ $(I^m \subset I^n)$, where $I = [-1, 1]$. Then we have the following statements (1)–(5).

(1) **Zeeman's unknotting theorem**. Every PL (n, m)-knot and every PL (n, m)-ball knot are unknotted, provided that $n - m \geqslant 3$.

In general, an injection $g: X \rightarrow Y$ between

topological manifolds is said to be **locally flat** if for each point $x \in X$, there is a neighborhood U of x in X so that the local germ of f at x is equivalent to the inclusion mapping $U \times 0 \subset U \times I^c$. Then f is necessarily proper, i.e., $f^{-1}(\partial Y) = \partial X$. A PL embedding $f: M \to N$ is said to be **locally flat** if it is locally flat in the PL category. Namely, $f|\mathrm{St}_K(x)$ is equivalent to $I^m \times 0 \subset I^n$. Conversely, Zeeman's unknotting theorem implies the following.

(2) Every proper PL embedding of codimension $\geqslant 3$ is locally flat.

Since every locally flat PL embedding $f: M \to N$ admits a normal †block bundle, the classification of $F: M \to U$ up to isomorphism is reduced to the classification of block bundles (→ 147 Fiber Bundles Q), where U is a derived neighborhood of $f(M)$ in N. (For the general classification of locally flat PL embeddings by surgical methods → C. T. C. Wall [25, Corollary 11.3.1]).

In the codimension 2 case, there exist nontrivial knots (the classical knots) (→ 235 Knot Theory) and there is a proper PL submanifold which is not locally flat. For example, a complex projective curve V defined by $x^n + y^{n-1}z = 0$ can be regarded as a PL 2-sphere in the complex projective plane $\mathbf{C}P^2$, and for the inclusion mapping $i: V \subset \mathbf{C}P^2$, at $0 = (0:0:1)$, $i|\mathrm{St}_V(0)$ is PL equivalent to a cone of a torus knot $k: S^1 \subset S^3$ of type $(n, n-1)$. For further results in higher-dimensional knot theory (→ 235 Knot Theory) we mention the works of M. Kato and Y. Matsumoto [26] and S. Capell and J. Shaneson [27].

Here we consider only proper PL embeddings of a fixed m-dimensional combinatorial manifold M into a fixed connected n-dimensional combinatorial manifold N, so that a codimension 0 PL embedding is a PL homeomorphism. Two PL embeddings f_0, $f_1: M \to N$ are said to be pseudoisotopic if there is a proper PL embedding $F: M \times [0,1] \to N \times [0,1]$, called a pseudoisotopy from f_0 to f_1, such that $F(x,0) = (f_0(x), 0)$ and $F(x,1) = (f_1(x), 1)$ $(x \in M)$. Moreover, if F satisfies the level-preserving condition $F(M \times \{t\}) \subset N \times \{t\}$ $(0 \leqslant t \leqslant 1)$, then f_0 and f_1 are said to be **isotopic**, and F or a homotopy $\{f_t\}$ defined by $F(x,t) = (f_t(x), t)$ $((x,t) \in M \times [0,1])$ is called an **isotopy** from f_0 to f_1. When $M = N$, a PL homeomorphism which is isotopic to the identity, or the isotopy of the identity itself, is called an **ambient isotopy** of N. Two PL embeddings from M to N are said to be **ambient isotopic** if they are isomorphic by an ambient isotopy of N.

Since I^k $(I = [-1,1])$ is a k-dimensional †convex cell and is a cone $0 * \partial I^k$ of ∂I^k from the center 0, each point x of $I^k - \{0\}$ can be written uniquely as $x = t \cdot u$ for $0 < t \leqslant 1$ and $u \in \partial I^k$. Thus a PL embedding $f: \partial I^m \to \partial I^n$ can be extended to a PL embedding $F: I^m \to I^n$, called a **cone extension** or simply a **cone** of f, by setting $F(0) = 0$ and $F(t \cdot u) = t \cdot f(u)$ for $t \cdot u \in I^m - \{0\}$. This method of cone extension is often referred to as **Alexander's trick**. By using Alexander's trick and the uniqueness of regular neighborhoods, one can show without much difficulty that (i) the isotopy classes of PL homeomorphisms of PL spheres or PL balls onto themselves are classified by their †degrees ($= \pm 1$) (V. K. A. M. Gugenheim, 1953), and (ii) every PL (n,m)-knot or PL (n,m)-ball knot is equivalent to the standard one $\partial I^{m+1} \times 0 \subset \partial I^{n+1}$ or $I^m \times 0 \subset I^n$ up to isotopy, provided that $n - m \geqslant 1$. This is why we define the notion of ambient isotopy in a different way from the differentiable category, where the smooth isotopy is always covered by the ambient isotopy. However, for codimension $\geqslant 3$, because of Zeeman's unknotting theorem, the following holds.

(3) The pseudoisotopy class, the isotopy class, and the ambient isotopy class of a PL embedding of codimension $\geqslant 3$ are the same (J. F. P. Hudson and E. C. Zeeman, 1964).

In order to construct a PL embedding from a continuous mapping within its homotopy class, the notion of general position is useful. A PL mapping $\varphi: M \to N$ is said to be **in general position** if $\dim \varphi^{-1}(y) \leqslant 0$ $(y \in N)$ and $\dim S(\varphi) \leqslant 2m - n$, where $S(\varphi)$ is the closure of $\{x \in M \mid \varphi^{-1}(\varphi(x)) \neq \{x\}\}$ in M (J. F. P. Hudson [8]).

(4) **General position theorem.** Every continuous mapping $\psi: M \to N$ can be approximated by a PL mapping $\varphi: M \to N$ which is in general position.

The following theorem can be proved by sharpening the general position theorem and the engulfing lemma.

(5) **Irwin's embedding theorem.** Let $\psi: M \to N$ be a continuous mapping from a compact combinatorial manifold of dimension m into a combinatorial manifold of dimension n such that $\psi|\partial M$ is a PL embedding from ∂M into ∂N. Assume that (i) M is d-connected, (ii) N is $(d+1)$-connected, and (iii) $n - m \geqslant 3$, where $d = 2m - n$. Then $\psi: M \to N$ is homotopic to a proper PL embedding $f: M \to N$ relative to ∂M. Moreover, if M and N are closed and if (i*) M is $(d+1)$-connected and (ii*) N is $(d+2)$-connected, then two homotopic PL embeddings from M into N are ambient isotopic (M. C. Irwin, 1965).

E. 3-Manifolds

There has been a great deal of research done on 3-manifolds, and we have, among others, the following results (1)–(4).

(1) **Sphere theorem**. Let M be an orientable 3-manifold with $\pi_2(M) \neq 0$. Then there exists a PL embedding f of a 2-sphere into M such that f is not homotopic to 0 in M (C. Papakyriakopoulos [28]).

(2) **Dehn's lemma**. Let D be a 2-cell with self-intersections in a 3-manifold M, with its boundary a simple polygon C. If some neighborhood U of C in D has no singularities, then there exists a 2-cell in M without singularities whose boundary is C. In 1910, M. Dehn asserted this lemma, but his proof was incomplete. In 1957, Papakyriakopoulos and T. Homma proved the lemma independently [28, 29].

(3) **Loop theorem**. Let N be the boundary of a 3-manifold M and U be an open set in a component of N. If there is a closed curve in U which is homotopic to 0 in M but not in N, then there exists a simple closed curve in U homotopic to 0 in M but not in N (Papakyriakopoulos [30], J. Stallings [31]).

(4) **Unique decomposition theorem** for a 3-manifold (J. W. Milnor [32], H. Kneser). We assume that all manifolds are connected, oriented, and triangulated 3-manifolds without boundaries and that all homeomorphisms are piecewise linear. Two manifolds M, M' are said to be isomorphic ($M \approx M'$) if there exists an orientation-preserving homeomorphism of M onto M'. Removing an open 3-cell from each of two 3-manifolds M, M' and identifying the boundaries of these removed cells, we obtain a 3-manifold $M \# M'$, called the **connected sum** of M and M'. A manifold that is not isomorphic to the 3-sphere S^3 is called nontrivial. A nontrivial manifold P is called prime if P cannot be decomposed as $P = M_1 \# M_2$ with M_1 and M_2 both nontrivial. A manifold M is called **irreducible** if every 2-sphere in M bounds a 3-cell. Then from the sphere theorem the following results can be deduced: If a compact 3-manifold M is nontrivial, then M is isomorphic to a connected sum $P_1 \# P_2 \# \ldots \# P_k$ of prime manifolds P_i $(i = 1, 2, \ldots, k)$, where $P_1, P_2, \ldots, P_k$ are determined uniquely up to order. Every irreducible 3-manifold M has $\pi_2(M) = 0$. Conversely, if a 3-manifold M has $\pi_2(M) = 0$, then M is irreducible, provided that the Poincaré conjecture is correct. (For further results on 3-manifolds → F. Waldhausen [10] and J. Hempel [12].)

Irreducible sufficiently large 3-manifolds [10] can further be decomposed into Seifert fibered spaces and "simple manifolds" which contain no essential tori (W. Jaco and P. B. Shalen, K. Johannson). Existence of hyperbolic structures on "simple manifolds" has been extensively studied by W. Thurston.

F. Wild Spaces

Nonclosed 3-manifolds behave very differently from closed 3-manifolds. For example, there exists an open 3-manifold U that has the same †homotopy type as an open 3-cell but is not homeomorphic to it (M. H. A. Newman and J. H. C. Whitehead, 1937). The construction of this and similar examples generally involves infinite processes. Such examples, which do not exhibit the properties we expect, are commonly termed **pathological** (or **wild**) spaces. For the manifold U above, the product $U \times \mathbf{R}^1$ is found to be homeomorphic to the Euclidean 4-space $\mathbf{R}^4$. R. Moore proved that the quotient space of the Euclidean 2-space $\mathbf{R}^2$ by any upper semicontinuous decomposition consisting of continua not separating $\mathbf{R}^2$ is homeomorphic to $\mathbf{R}^2$. On the other hand, R. H. Bing (1959) gave an example of a similar upper semicontinuous decomposition of $\mathbf{R}^3$ for which the quotient space B is not homeomorphic to $\mathbf{R}^3$ and is not even a manifold. This quotient space has the property $B \times \mathbf{R}^1 \approx \mathbf{R}^4$. Bing (1957) also proved that at most countably many copies of a wild closed surface can simultaneously be embedded in $\mathbf{R}^3$. On the other hand, Stallings (1960) proved that there exists a wild disk of which a set of copies with the power of the continuum can simultaneously be embedded in $\mathbf{R}^3$ [11].

The study of the quotient space of n-dimensional manifolds by cell-like decompositions has been fully generalized by Edwards and Cannon [7] for $n \geqslant 5$ to prove the double suspension theorem of homology spheres (→ Section C (3)).

G. The Schoenflies Theorem

In 1906, A. M. Schoenflies proved that every †simple closed curve in the plane is the boundary of a 2-cell. This result is sharper than the †Jordan curve theorem and is called the **Schoenflies theorem**. However there was a gap in the proof, and the complete proof was given by L. E. J. Brouwer in 1910.

The following higher-dimensional analog of the above theorem is called the **Schoenflies problem**: Is every $(n-1)$-sphere embedded in the n-dimensional Euclidean space $\mathbf{R}^n$ $(n \geqslant 3)$ the boundary of an embedded n-disk?

For the case where the embedding is topological, L. Antoine constructed a counterexample, called **Antoine's necklace**, for $n = 3$ in 1921 and solved this problem negatively. Also, J. W. Alexander constructed another counterexample, called **Alexander's horned sphere**, in 1924 (Fig. 1). A similar example was given by R. H. Fox and E. Artin in 1948 by making use

of knot theory (Fig. 2). The study of wild embeddings began with these examples, and many interesting results were obtained (Bing).

For the case where the embedding is PL or smooth, the Schoenflies problem has been solved affirmatively for $n=3$ (Alexander).

M. Brown (1960) and B. Mazur (1959) proved the following: Let $h: S^{n-1} \times [-1, 1] \rightarrow S^n$ be a topological embedding; then $S^n - h(S^{n-1} \times \{0\})$ consists of connected open sets D_+ and D_- of S^n, and the closures of D_+ and D_- are both homeomorphic to the n-disk. This result solves the Schoenflies problem affirmatively for $n \geqslant 5$ when the embedding is PL and locally flat or smooth.

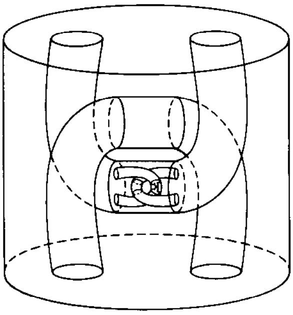


Fig. 1

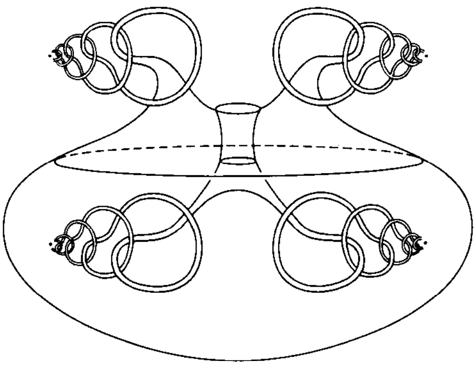


Fig. 2

References

[1] H. Poincaré, Analysis situs, J. Ecole Polytech., 1 (1895), 1–121; Cinquième complément à l'analysis situs, Rend. Circ. Mat. Palermo, 18 (1904), 45–110.

[2] J. H. C. Whitehead, Simplicial spaces, nuclei and m-groups, Proc. London Math. Soc., 45 (1939), 243–327; On the homotopy type of manifolds, Ann. Math., (2) 41 (1940), 825–832; Simple homotopy types, Amer. J. Math., 72 (1950), 1–57.

[3] J. Stallings, Polyhedral homotopy spheres, Bull. Amer. Math. Soc., 66 (1960), 485–488; Notes on polyhedral topology, Tata Inst., 1967.

[4] E. C. Zeeman, Unknotting combinatorial balls, Ann. Math., (2) 78 (1963), 501–526; Seminars on combinatorial topology, Publ. Math. Inst. HES and Univ. of Warwick (Conventry), 1963–1966.

[5] R. C. Kirby and L. C. Siebenmann, On the triangulation of manifolds and the Hauptvermutung, Bull. Amer. Math. Soc., 75 (1969), 742–749; Foundational essays on topological manifolds, smoothings and triangulations, Ann. Math. Studies 88, Princeton Univ. Press, 1977.

[6] T. A. Chapman, Topological invariance of Whitehead torsion, Amer. J. Math., 96 (1974), 488–497.

[7] J. W. Cannon, Shrinking cell-like decompositions of manifolds. Codimension three, Ann. Math., (2) 110 (1979), 83–112.

[8] J. F. P. Hudson, Piecewise linear topology, Benjamin, 1969.

[9] C. P. Rourke and B. J. Sanderson, Introduction to piecewise linear topology, Erg. Math., 69 (1972).

[10] F. Waldhausen, On irreducible 3-manifolds which are sufficiently large, Ann. Math., (2) 87 (1968), 56–88, Amer. Math. Soc. Proc. Symposia in Pure Math., vol. 32, part 2, 1978.

[11] M. K. Fort, Jr. (ed.), Topology of 3-manifolds and related topics, Prentice-Hall, 1962.

[12] J. Hempel, 3-manifolds, Ann. Math. Study 86, Princeton Univ. Press, 1976.

[13] P. Alexandroff and H. Hopf, Topologie I, Springer, 1935.

[14] H. Seifert and W. Threlfall, Lehrbuch der Topologie, Teubner, 1934.

[15] T. Rado, Ueber den Begriff der Riemannschen Fläche, Acta Litt. Sci. Szeged, 2 (1925), 101–121.

[16] E. E. Moise, Affine structures on 3-manifolds, Ann. Math., (2) 56 (1952), 96–114.

[17] D. Sullivan, On the Hauptvermutung for manifolds, Bull. Amer. Math. Soc., 73 (1967), 698–700.

[18] L. C. Siebenmann, Topological manifolds, Proc. Int. Congress Math. (Nice, 1970), vol. 2, 133–163.

[19] R. C. Kirby, Stable homeomorphisms and the annulus conjecture, Ann. Math., (2) 89 (1969), 575–582.

[20] J. W. Milnor, Whitehead torsion, Bull. Amer. Math. Soc., 72 (1966), 358–426.

[21] B. Mazur, Differential topology from the viewpoint of simple homotopy theory, Publ. Math. Inst. HES, 15 (1963), 5–93.

[22] M. Kervaire, Le Théorèm de Barden-Mazur-Stallings, Comment. Math. Helv., 40 (1965), 31–42.

[23] C. Weber, L'élimination des points doubles dans le cas combinatoire, Comment. Math. Helv., 41 (1966), 179–182.

[24] M. H. A. Newman, The engulfing theorem for topological manifolds, Ann. Math., (2) 84 (1966), 555–571.
[25] C. T. C. Wall, Surgery on compact manifolds, Academic Press, 1970.
[26] M. Kato and Y. Matsumoto, Simply connected surgery of submanifolds in codimension two, J. Math. Soc. Japan, 24 (1972), 586–608.
[27] S. Cappell and J. Shaneson, The codimension two placement problem and homology equivalent manifolds, Ann. Math., (2) 99 (1974), 277–348.
[28] C. Papakyriakopoulos, On Dehn's lemma and the asphericity of knots, Ann. Math., (2) 66 (1957), 1–26.
[29] T. Homma, On Dehn's lemma for S^3, Yokohama Math. J., 5 (1957), 223–244.
[30] C. Papakyriakopoulos, On solid tori, Proc. London Math. Soc., 7 (1957), 281–299.
[31] J. Stallings, On the loop theorem, Ann. Math., (2) 72 (1960), 12–19.
[32] J. W. Milnor, A unique factorization theorem for 3-manifolds, Amer. J. Math., 84 (1962), 1–7.
[33] W. Thurston, The geometry and topology of three-manifolds, Princeton Univ. Press, 1982.

66 (XVI.10) Combinatorics

A. General Review of Combinatorics

Combinatorics is concerned mainly with problems of discrete sets, such as the enumeration of subsets satisfying certain conditions, or their actual construction and the selection of an optimal subset with respect to a suitable criterion. Generally speaking, both the theoretical analysis and the construction of discrete sets are much more difficult than those problems in analysis concerning infinite sets with suitable topologies.

The main emphasis and the name of this field have changed from time to time and from person to person. Other names such as **combinatorial analysis, combinatorial theory**, or **combinatorial mathematics** are commonly used and mean almost the same thing. Recently, the expression "**discrete mathematics**" has also been used to describe the same field. In many cases a problem originally conceived as number-theoretic, algebraic, analytic, or geometric has eventually turned out to be combinatorial (→ 102 Design of Experiments, 241 Latin Squares).

†Graph theory, originally a combinatorial study of 1-dimensional topological †complexes, has developed into an important branch of mathematics.

In this article we discuss various topics not yet considered to be established branches or organized fields of mathematics. The topics treated under the title "combinatorics" are at present not unified, and consist of a mixture of isolated techniques. However, recent progress with electronic computers is playing an important role in the solution of various combinatorial problems arising from large systems, and the study of combinatorics is now quite active both in theoretical development and application. We now have special journals in this field, such as the *Journal of Combinatorial Theory* (Academic Press), *Combinatorica* (Elsevier), the *European Journal of Combinatorics* (Academic Press), *Discrete Mathematics* (North-Holland), and *Discrete Applied Mathematics* (North-Holland).

B. Three Fundamental Principles

The most fundamental among the various principles in combinatorics are the following three principles for enumeration problems. Let Ω be a finite set and let $|A|$ denote the number of elements in a subset A of Ω.

(1) **Rule of sums**: If $A \cap B = \varnothing$, then $|A+B| = |A|+|B|$.

(2) **Rule of products**: $|A \times B| = |A| \times |B|$, where $A \times B$ is the †direct product of the two sets A and B.

(3) **Principle of inclusion and exclusion**:

$$|A_1 \cup A_2 \cup \ldots \cup A_n| = \sum_i |A_i| - \sum_{i<j} |A_i \cap A_j| + \sum_{i<j<k} |A_i \cap A_j \cap A_k| - \ldots + (-1)^{n-1} |A_1 \cap A_2 \cap \ldots \cap A_n|.$$

These principles can be generalized as statements concerning various measures.

C. Möbius Inversion Formula

Let P be an ordered set whose arbitrary interval is a finite set. Let $[P \times P] = \{(x, y) \in P \times P \mid x \leqslant y\}$ be the set of intervals of P. A function satisfying the following two conditions is determined uniquely and is called the **Möbius function** over P:

(i) $\mu(x, x) = 1$ for all $x \in P$;

(ii) $\sum_{z: x \leqslant z \leqslant y} \mu(x, z) = 0$ for all $x, y \in P$ with $x < y$ and $x \neq y$.

When K is a field, for a mapping $f: [P \times P] \rightarrow K$, we can define another mapping $g: [P \times$

$P] \to K$ by

$$g(x,y)=\sum_{z\in[x,y]} f(x,z).$$

Then we have the **Möbius inversion formula**

$$f(x,y)=\sum_{z\in[x,y]} g(x,z)\mu(z,y).$$

This formula implies, as its special cases, many important formulas, such as the †Möbius inversion formula in number theory, the principle of inclusion and exclusion, and the inverse relation of difference and summation.

D. Some Special Sequences

In the enumeration problem, many important sequences $\{a_n\}$ are given by the †generating function $f(x)=\sum a_n x^n$. Typical examples are as follows: (1) †Binomial coefficients: $(1+x)^m = \sum_n \binom{m}{n} x^n$. (2) Multinomial coefficients: $(x_1+\ldots+x_r)^m=\sum \binom{m}{n_1 \ldots n_r} x_1^{n_1}\ldots x_r^{n_r}$. (3) Multiple combination $a_n={}_mH_n=\binom{m+n-1}{n}:(1-x)^{-m}=\sum_n ({}_mH_n)x^n$. (4) Bell numbers: $\exp(e^x-1)=\sum B_n x^n/n!$ (5) The **Stirling number of the second kind** S_m^n is the total number of bijections from $\mathbf{N}_n$ to $\mathbf{N}_m$ divided by $m!$, where $\mathbf{N}_n=\{1,2,\ldots,n\}$. This is equivalent to the total number of partitions of $\mathbf{N}_n$ into m nonempty blocks: $(e^x-1)^m/m! = \sum_{n=m}^{\infty} S_m^n x^n/n!$.

In general, when the generating function $f(x)$ is an analytic function of one or several variables as above, we can obtain recurrence formulas for the sequence from the functional equations or the differential equations which $f(x)$ satisfies; or we can get asymptotic estimates by using the †saddle point method.

E. Pólya's Enumeration Theorem

Let U be a finite set, V be a (finite or infinite) set, G be a transformation group operating on U, and w be a function assigning to each element of V a "weight." A weight is an element of a ring and is usually defined by a suitable generating function. The weight of a function $f: U \to V$ is defined by $w(f)=\prod_{u\in U} w(f(u))$. Two functions f_1 and f_2 are considered identical ($f_1 \equiv f_2$) if there exists a $\pi \in G$ such that $f_1 = f_2 \circ \pi$. Put $F=V^U/(\Xi)$, the set of all equivalence classes of the relation Ξ. $w(\tilde{f})$ is well-defined for $\tilde{f}\in F$, because $f_1\equiv f_2$ implies $w(f_1) = w(f_2)$. Then we have the equality $\sum_{\tilde{f}\in F} w(\tilde{f}) = P_G(\sum_{v\in V} w(v), \sum_{v\in V} w(v)^2, \ldots, \sum_{v\in V} w(v)^m)$. Here, $P_G(y_1, y_2, \ldots, y_m)$ is the **cycle index** of G, which is given by $n^{-1}\sum_{\pi\in G} y_1^{j_1(\pi)} y_2^{j_2(\pi)} \ldots y_m^{j_m(\pi)}$, where $n=|G|$, $m=|U|$, and $j_k(\pi)$ is the number of cyclic permutations of length k contained in π. This is called **Pólya's enumeration theorem**. When $G=S_m$ (the †symmetric group), $P_{S_m}(y_1, \ldots, y_m)=\sum_{\lambda_k\geqslant 0, \Sigma k\lambda_k=m}(\lambda_1!2^{\lambda_2}\lambda_2!\ldots m^{\lambda_m}\lambda_m!)^{-1} y_1^{\lambda_1} \ldots y_m^{\lambda_m}$. When $G=A_m$ (the †alternating group), $P_{A_m}(y_1, \ldots, y_m)=\sum_{\lambda_k\geqslant 0, k\lambda_k=m, \Sigma\lambda_k=\text{even}} 2(\lambda_1!2^{\lambda_2}\lambda_2! \ldots m^{\lambda_m}\lambda_m!)^{-1} y_1^{\lambda_1}\ldots y_m^{\lambda_m}$. When $G=C_m$ (the †cyclic group), $P_{C_m}(y_1, \ldots, y_m)=m^{-1}\sum_{k|m}\varphi(k) y_k^{m/k}$, where $\varphi(k)$ is the †Euler function, i.e., the number of integers less than k and coprime to k. This theorem is widely applicable to various problems with suitable choices of weights. Several generalizations have also been studied (see, e.g., [3]).

F. The Notion of Polymatroids

The theory of matroids was originally introduced as an abstraction of the notion of †linear dependence in a vector space or of some properties of †graphs or of the notion of †algebraic dependence in the theory of fields (see, e.g., [4, 5]). But recently, as many important applications were discovered, it has become an important branch of combinatorics (see [6–8]).

Let μ be a function from a †lattice L to a totally ordered module R. μ is called **submodular** if $\mu(x)+\mu(y) \geqslant \mu(x\vee y)+\mu(x\wedge y)$ for every pair of elements $x, y\in L$. The set of $a\in L$ such that $\mu(a)=\min_{x\in L}\mu(x)$ constitutes a sublattice in L. Let L be the family of all subsets of a finite set E. The sublattices K in $L=2^E$ and the pseudo-orders over E correspond in a one-to-one manner as follows (here pseudo-order means a binary relation taken as an †ordering without assuming the †antisymmetric law): (i) The class of a partition P of E corresponds to the difference of two subsets which are the immediate predecessor and immediate successor in K, or the complement with respect to E of the maximum element in K, or the complement of the minimum element. (ii) The complement of the maximum element in K is maximum in the order of P, and the minimum element in K is minimum in the order of P. (iii) Two classes ξ_1, ξ_2 in P, given as the difference of two elements in K, satisfy the relation $\xi_1 \geqslant \xi_2$ with respect to the order of P if and only if all elements of K such that $x \supset \xi_1$ in E satisfy $x\supset\xi_2$.

Now let E be a finite set, $L=2^E$, and let $\mathbf{R}$ be the real number field with the usual addition and order relations. If a submodular function $\mu: 2^E \to \mathbf{R}$ satisfies the following conditions, then $\mathbf{P}=(E,\mu)$ is called a **polymatroid** over E, and μ is called the **rank function** over $\mathbf{P}$. (i) $\mu(\varnothing)=0$, where $\varnothing$ is the empty set. (ii) $x\subset y$ $(\subset E)$ implies $\mu(x)\leqslant\mu(y)$. A function $u: E\to\mathbf{R}$ is automatically extended to $u: 2^E\to\mathbf{R}$ by $u(x)=\sum_{e\in x} u(e)$ $(x\subset E)$, which is also denoted by the same symbol u. If a function u satisfies $0\leqslant u(x)\leqslant\mu(x)$ for every $x\in E$, then u is called an **independent vector** over $\mathbf{P}$. An independent

vector u with $u(E)=\mu(E)$ is called a **base** of $\mathbf{P}$. A function $u:E\to\mathbf{R}$ is identified with an element in $\mathbf{R}^n$, where $n=|E|$. Then a set of independent vectors in $\mathbf{P}$ forms a †convex polyhedron in $\mathbf{R}^n$. If we denote the polyhedron by $\mathbf{P}$, then such a $\mathbf{P}$ is characterized by the following two conditions: (P1) If $u\in\mathbf{P}$ then each v satisfying $0\leqslant v(e)\leqslant u(e)$ for every $e\in E$ also belongs to $\mathbf{P}$. (P2) If two $u, v\in\mathbf{P}$ satisfy $u(E)< v(E)$, then there exists $w\in P$, $w\neq u$, satisfying $u(e)\leqslant w(e)\leqslant\max[u(e), v(e)]$ for every $e\in E$. If there are two polymatroids $\mathbf{P}_1=(E,\mu_1)$ and $\mathbf{P}_2=(E,\mu_2)$ over the same finite set E, then we have $\max\{u(E)\mid u\in\mathbf{P}_1\cap\mathbf{P}_2\}=\min\{\mu_1(x)+\mu_2(E-x)\mid x\subset E\}$. According to †information theory, an example of a polymatroid (E,μ) is given by the following: $\mu(x)$, $x\in E$, is the simultaneous †entropy of a subset x of the information source E with correlation.

G. Matroids

A polymatroid $\mathbf{P}=(E,\mu)$ is called a **matroid** if the values of μ are always integers and $\mu(x)\leqslant|x|=$ the cardinality of the set x. A set $x\subset E$ is called an **independent set** if $\mu(x)=|x|$, and a **dependent set** otherwise. An independent set with maximal cardinality is called a **base**. A minimal dependent set is called a **circuit**. The set y with $\max\{|y|\mid y\supset x, \mu(x)=\mu(y)\}$ is uniquely determined by x, which is called the **closure** of x and denoted by $\mathrm{cl}\,x$. Several axioms stated in terms of these notions exist, and can be used to characterize matroids. Some of them are:

(1) The axiom for the family of independent sets $\mathscr{I}$: (i) $\varnothing\in\mathscr{I}$; (ii) $x\in\mathscr{I}$ and $y\subset x$ implies $y\in I$; (iii) If $x, y\in\mathscr{I}$ and $|x|<|y|$, then there exists an $e\in E$ in $y-x$ such that $x\cup\{e\}\in\mathscr{I}$.

(2) The axiom for the family of bases $\mathscr{B}$: (i) $\mathscr{B}$ is not empty; (ii) If $b, b'\in\mathscr{B}$ and $b\neq b'$, then for every $e\in b-b'$, there exists an e' such that $e'\in b'-b$ and $(b-\{e\})\cup\{e'\}\in\mathscr{B}$.

(3) The axiom for the family of circuits $\mathscr{C}$: (i) $\varnothing\in\mathscr{C}$; (ii) For every pair $x, y\in\mathscr{C}$, y is not a proper subset of x; (iii) If $x, y\in\mathscr{C}$ and $x\neq y$, then for every $e\in x\cap y$ there exists a $z\in\mathscr{C}$ such that $z\subset x\cup y-\{e\}$.

(4) The axiom for the closure function cl: $2^E\to 2^E$: (i) $x\subset\mathrm{cl}\,x=\mathrm{cl}(\mathrm{cl}\,x)$ for every $x\in E$; (ii) $x\subset y$ implies $\mathrm{cl}\,x\subset\mathrm{cl}\,y$; (iii) If $e\in\mathrm{cl}(x\cup\{e'\})-\mathrm{cl}\,x$, then $e'\in\mathrm{cl}(x\cup\{e\})-\mathrm{cl}\,x$.

The following are typical examples of matroids: (1) Let E be the family of finite set of vectors in a vector space V and $\mathscr{I}$ be the family of sets of linearly independent vectors. (2) Let E be the set of edges of a †graph and $\mathscr{C}$ be the set of edges constituting a fundamental circuit. (3) Let E be a finite set and S be a family of subsets of E. $x=\{e_1,\ldots,e_r\}\subset E$ is defined to be independent if there exist different elements $s_1,\ldots,s_r$ in S such that $e_i\in s_i$ for every $i=1,\ldots,r$.

Let a weight function $w:E\to R$ be defined for every element $e\in E$ in a matroid (E,μ). To obtain a base b of the matroid with minimal weight $w(b)=\sum_{e\in b}w(e)$, we can apply the following procedure, called the **greedy algorithm**: (1) Arrange the elements $e_1,\ldots,e_n$ of E in such a manner that $w(e_1)\leqslant\ldots\leqslant w(e_n)$. (2) Put $b^{(0)}:=\varnothing$, and $i:=0$. (3) If $x\equiv b^{(i)}\cup\{e_{i+1}\}\in\mathscr{I}$, then $b^{(i+1)}:=x$; otherwise, $b^{(i+1)}:=b^{(i)}$. (4) Repeat process (3) for $i=0,1,\ldots,n-1$. Then the final $b^{(n)}$ is the base of minimal weight.

H. Operations on Matroids

For a matroid $M=(E,\mu)$ with the family $\mathscr{B}$ of its base, $\mathscr{B}^*=\{E-x\mid x\in\mathscr{B}\}$ also satisfies the axioms of the base. The matroid $M^*=(E,\mu^*)$ with $\mathscr{B}^*$ as its base is called the **dual** of M. Here we define $\mu^*(x)=|x|+\mu(E-x)-\mu(E)$. When $F\subset E$, $M\mid F=(F,\mu_F)$ is called the **reduction** of $M=(E,\mu)$ to F, where μ_F is the restriction of μ on 2^F. A matroid $M\times F=(M^*\mid F)^*$ is called the **contraction** of M to F. A matroid given by the form $(M\mid F)\times F'$ is called a **minor** of M. From two matroids $M_1=(E,\mu_1)$ and $M_2=(E,\mu_2)$ over the same set E, we can construct a matroid $M=M_1\vee M_2=(E,\mu)$ whose rank function $\mu(x)$ is defined by the minimum of $\{\mu_1(y)+\mu_2(y)+|x-y|\mid y\subset x\}$. This is called the **union** of M_1 and M_2.

Let $M=(E,\mu)$ be a matroid and α be a positive constant. The function $\tilde{\mu}(x)=\mu(x)-\alpha|x|$ is submodular. The set L of x which minimizes $\tilde{\mu}(x)$ is a sublattice of 2^E. The partition of E and the order relation of its classes defined by L are called the **principal partition** of M with respect to the parameter α. When the value of the parameter α is not specified, α is taken to be 2.

As was seen in example (1) at the end of the preceding section, the family of vectors in a vector space V over a field K is a matroid M. A matroid isomorphic to such an M is called **linearly representable** over K. If $K=GF(p)$, the finite field with p elements, then it is called a **p-ary matroid**. A matroid linearly representable over any field is called **regular**. If $[G]$ is a matroid isomorphic to the matroid defined by the family of edges of a graph as shown in example (2), it is called **graphic**. If the dual M^* is graphic, M is called **cographic**. Graphic and cographic matroids are always regular. There exist matroids not linearly representable over a field K, or linearly representable over a particular field K but not regular, or regular but neither graphic nor cographic. If $[G]$ is a matroid isomorphic to

the matroid of the type in example (3), it is called **transversal**. Such a matroid is linearly representable when the cardinality of the field K is sufficiently large.

References

[1] G. C. Rota, On the foundations of combinatorial theory I. Theory of Möbius functions, Z. Wahrscheinlichkeitstheorie, 2 (1964), 340–368.
[2] M. Hall, Jr. Combinatorial theory, Blaisdell, 1967.
[3] C. L. Liu, Introduction to combinatorial mathematics, McGraw-Hill, 1968.
[4] H. Whitney, On the abstract properties of linear dependence, Amer. J. Math., 57 (1935), 509–533.
[5] B. L. van der Waerden, Algebra, Ungar, 1970. (Original in German, 1937.)
[6] D. J. A. Welsh, Matroid theory, Academic Press, 1976.
[7] E. L. Lawler, Combinatorial optimization —Networks and matroids, Holt, Rinehart and Winston, 1976.
[8] M. Iri and S. Fujishige, Use of matroid theory in operations research, circuits and system theory, Int. J. System Sci., 12 (1981), 27–54.
[9] J. Riordan, An introduction to combinatorial analysis, Wiley, 1958.
[10] H. J. Ryser, Combinatorial mathematics, Wiley, 1963.

67 (III.11) Commutative Rings

A. General Remarks

A ring R (→ 368 Rings) whose multiplication is commutative is called a **commutative ring**. Throughout this article, we mean by a ring a commutative ring with unity element.

B. Ideals

Since our rings are commutative, we need not distinguish right or left ideals from †ideals. A subset $\mathfrak{a}$ of a ring R is an ideal of R if and only if $\mathfrak{a}$ is an R-submodule of R (→ 277 Modules), if and only if $\mathfrak{a}$ is the †kernel of a ring homomorphism from R into some ring (except for the case $\mathfrak{a}=R$). Given an ideal $\mathfrak{a}$ of a ring R, the set of elements which are †nilpotent modulo $\mathfrak{a}$, i.e., $\{x\in R \mid x^n\in\mathfrak{a}\ (\exists n)\}$, is called the **radical** of $\mathfrak{a}$ and is often denoted by $\sqrt{\mathfrak{a}}$. The radical of the zero ideal is called the **radical** of R, or more precisely, the **nilradical** of R (→ 368 Rings H).

For a subset S of a ring R, the smallest ideal $\mathfrak{a}$ containing S is called the ideal **generated** by S, and S is called a **basis** for $\mathfrak{a}$; if S is a finite set, then S is called a **finite basis**. When $\mathfrak{a}_\lambda$ $(\lambda\in\Lambda)$ are ideals of R, the **sum** of these ideals, denoted by $\sum\mathfrak{a}_\lambda$, is defined to be the ideal generated by the union of the ideals $\mathfrak{a}_\lambda$ $(\lambda\in\Lambda)$. If Λ is a finite set, say $\{1,2,\ldots,n\}$, then the sum is denoted also by $\mathfrak{a}_1+\ldots+\mathfrak{a}_n$, and in this case, the sum is also called a finite sum. Note that $\mathfrak{a}_1+\ldots+\mathfrak{a}_n=\{a_1+\ldots+a_n \mid a_i\in\mathfrak{a}_i\}$. The **product** $\mathfrak{a}_1\ldots\mathfrak{a}_n$ of a finite number of ideals $\mathfrak{a}_1,\ldots,\mathfrak{a}_n$ is defined to be the ideal generated by the set $\{a_1\ldots a_n \mid a_i\in\mathfrak{a}_i\}$. The intersection of an arbitrary number of ideals is an ideal. When $\mathfrak{a}$ is an ideal of a ring R and S is a subset of R, the **quotient** $\mathfrak{a}:S$ is defined to be the ideal $\{x\in R \mid xS\subset\mathfrak{a}\}$. If $\mathfrak{a},\mathfrak{b},\mathfrak{c},\mathfrak{d}_\lambda$ $(\lambda\in\Lambda)$ are ideals, we have $(\mathfrak{a}:\mathfrak{b}):\mathfrak{c}=\mathfrak{a}:\mathfrak{b}\mathfrak{c}$, $\mathfrak{a}:\sum\mathfrak{d}_\lambda=\bigcap_\lambda(\mathfrak{a}:\mathfrak{b}_\lambda)$.

C. Prime Ideals

An ideal $\mathfrak{p}$ of a ring R is called a **prime ideal** if $R/\mathfrak{p}$ is an †integral domain; $\mathfrak{p}$ is a prime ideal if and only if $\mathfrak{p}\neq R$ and also $ab\in\mathfrak{p}$ $(a,b\in R)$ implies $a\in\mathfrak{p}$ or $b\in\mathfrak{p}$ (some literature includes the ring R itself in the set of prime ideals). Let S be a **multiplicatively closed subset** of a ring R, i.e., a nonempty †subsemigroup of R with respect to multiplication. A maximal member among the set of ideals which do not meet S is called a **maximal ideal with respect to** S. Such a member is necessarily a prime ideal; when $S=\{1\}$, it is called a **maximal ideal** of R. An ideal $\mathfrak{m}$ is a maximal ideal of R if and only if $R/\mathfrak{m}$ is a field.

D. Jacobson Radical

The intersection J of all maximal ideals of a ring R is called the **Jacobson radical** of R; in some cases, this intersection J is called the radical of R (→ 368 Rings H). Let N be an R-submodule of a finite R-module M. If $MJ+N=M$, then $M=N$ (**Krull-Azumaya lemma** or **Nakayama lemma**).

E. Krull Dimension

For a prime ideal $\mathfrak{p}$ of a ring R, the maximum of the lengths n of †descending chains of prime ideals $\mathfrak{p}=\mathfrak{p}_0\supsetneqq\mathfrak{p}_1\supsetneqq\ldots\supsetneqq\mathfrak{p}_n$ which begin with $\mathfrak{p}$ (or ∞ if the maximum does not exist) is called the **height** or **rank** of the prime ideal $\mathfrak{p}$. For an ideal $\mathfrak{a}$, the minimum of the heights of prime ideals containing $\mathfrak{a}$ is called the **height** of the ideal $\mathfrak{a}$. The maximum of heights of prime

ideals of R is called the **Krull dimension** (or **altitude**) of the ring R. For an ideal $\mathfrak{a}$ of R, the Krull dimension of $R/\mathfrak{a}$ is called the **Krull dimension** (or **depth**) of the ideal $\mathfrak{a}$. (The meanings of the terms *rank*, *dimension*, and *depth* now depend on the writings in which they are found; standardization of definition of these terms is becoming more and more of a necessity.)

F. Primary Ideals

For an ideal $\mathfrak{q}$ of R, if $\mathfrak{q} \neq R$ and every zero divisor of $R/\mathfrak{q}$ is nilpotent, then $\mathfrak{q}$ is called a **primary ideal**. (If R is included in the set of prime ideals, then R is regarded as a primary ideal.) In this case, $\mathfrak{p} = \sqrt{\mathfrak{q}}$ is a prime ideal, and $\mathfrak{q}$ is then said to **belong** to $\mathfrak{p}$ or to be a $\mathfrak{p}$**-primary ideal**. The intersection of a finite number of primary ideals belonging to the same prime ideal $\mathfrak{p}$ is $\mathfrak{p}$-primary. Assume that an ideal $\mathfrak{a}$ is expressed as the intersection $\mathfrak{a} = \mathfrak{q}_1 \cap \ldots \cap \mathfrak{q}_n$ of a finite number of primary ideals $\mathfrak{q}_1, \ldots, \mathfrak{q}_n$. If this intersection is **irredundant**, that is, if none of the $\mathfrak{q}_i$ is superfluous in the expression of $\mathfrak{a}$, then the set of $\sqrt{\mathfrak{q}_i}$ $(i = 1, \ldots, n)$ is uniquely determined by $\mathfrak{a}$. The $\sqrt{\mathfrak{q}_i}$ are called **prime divisors** (or **associated prime ideals**) of the ideal $\mathfrak{a}$; a minimal one among these is called a **minimal** (or **isolated**) **prime divisor** of the ideal $\mathfrak{a}$, and those which are not minimal are called **embedded prime divisors** of the ideal $\mathfrak{a}$. A maximal one among the prime divisors of $\mathfrak{a}$ is called a **maximal prime divisor** of the ideal $\mathfrak{a}$. (For definitions of these concepts in the case where $\mathfrak{a}$ is an arbitrary ideal, see [4].) If, in the expression of $\mathfrak{a}$ as above, n is the smallest occurring in similar expressions, then the expression is called the **shortest representation** of the ideal $\mathfrak{a}$ by primary ideals. In this case, each $\mathfrak{q}_i$ is called a **primary component** of the ideal $\mathfrak{a}$; if $\sqrt{\mathfrak{q}_i}$ is isolated, then $\mathfrak{q}_i$ is called an **isolated primary component** of the ideal $\mathfrak{a}$, and otherwise $\mathfrak{q}_i$ is called an **embedded primary component** of the ideal $\mathfrak{a}$. Isolated primary components are uniquely determined by $\mathfrak{a}$, but embedded primary components are not.

G. Rings of Quotients

Let R be a ring. Then the set U of elements of R which are not zero divisors is multiplicatively closed. In the set $R \times U = \{(r, u) \mid r \in R, u \in U\}$, we define a relation $\equiv$ by $(r, u) \equiv (r', u') \Leftrightarrow ru' = r'u$. Then $\equiv$ is an equivalence relation, and the equivalence class of (r, u) is denoted by r/u. In the set Q of these r/u, addition and multiplication are defined by $r/u + r'/u' = (ru' + r'u)/uu'$, $(r/u)(r'/u') = rr'/uu'$. Then Q becomes a ring, and $r/1$ can be identified with r. Thus Q is a ring containing R, generated by R and inverses of elements of U. This property characterizes Q, which is called the **ring of total quotients** of the ring R. If R is an integral domain, then Q is a field, called the **field of quotients** of the integral domain R. Let S be a multiplicatively closed subset of R such that $0 \notin S$, and let $\mathfrak{n}$ be $\{x \in R \mid xs = 0\ (\exists s \in S)\}$ and φ be the natural homomorphism $R \rightarrow R/\mathfrak{n}$. Then none of the elements of $\varphi(S)$ is a zero divisor. The subring of the ring of total quotients of $R/\mathfrak{n}$, generated by $R/\mathfrak{n}$ and inverses of elements of $\varphi(S)$, is called the **ring of quotients of the ring** R **with respect to** S, and is denoted by R_S ($R[S^{-1}]$ or RS^{-1}). When M is an R-module, $M \otimes_R R_S$ is called the **module of quotients of the** R**-module** M **with respect to** S. It is significant that R_S is †R-flat. There is a one-to-one correspondence between the set of primary ideals $\mathfrak{q}$ of R which do not meet S and the set of primary ideals $\mathfrak{Q}$ of R_S such that $\mathfrak{q}$ corresponds to $\mathfrak{Q}$ if and only if $\mathfrak{Q} = \mathfrak{q}R_S$ ($\mathfrak{q} = \mathfrak{Q} \cap R$). When $\mathfrak{p}$ is a prime ideal of R, the complement $R - \mathfrak{p}$ is multiplicatively closed, and $R_{R-\mathfrak{p}}$ is called the **local ring of** $\mathfrak{p}$ or the **ring of quotients of the ring** R **with respect to the prime ideal** $\mathfrak{p}$, and is denoted by $R_\mathfrak{p}$ (→ 284 Noetherian Rings C, D). A ring of quotients is also called a **ring of fractions**.

H. Divisibility

In a ring R, if $a = bc$ $(a, b, c \in R)$, then we say that b is a **divisor** (or **factor**) of a, and that a is a **multiple** of b, or a is **divisible** by b. We denote this by $b \mid a$. This relation between $a, b \in R$ is called **divisibility relation** in R. If, in this situation, c has its inverse in R, we say that a is an **associate** of b. A factor b of a is called a **proper factor** if b is neither an associate of a nor †invertible. An element which has no proper factor is called an **irreducible element**. A nonzero element which generates a prime ideal is called a **prime element**.

If in an integral domain R every nonzero element is a product of prime elements (up to invertible factors), then we say that the **unique factorization theorem** holds in R, and that R is a **unique factorization domain** (or simply u.f.d.).

Let $A = \{a_1, \ldots, a_n\}$ be a set of nonzero elements of a ring R. A **common divisor** of A is an element which is a factor of a_i. A **common multiple** of A is defined similarly. The **greatest common divisor** (G.C.D.) of A is a common divisor which is a multiple of any common divisor; the **least common multiple** (L.C.M.) of A is a common multiple m which is a factor of any common multiple. Thus, the G.C.D. and L.C.M. exist if R is a u.f.d.

I. Integral Dependence

Let R be a subring of a ring R'' sharing a unity element with R''. An element $a \in R''$ is said to be **integral** (or **integrally dependent**) over R if there are a natural number n and elements c_i of R such that $a^n + c_1 a^{n-1} + \ldots + c_n = 0$. If every element of a subset S of R'' is integral over R, we say that S is **integral** over R. (When R has no unity element, a similar definition is given under an additional condition that $c_i \in R^i$. An important special case is where R is an ideal. See D. G. Northcott and D. Rees, *Proc. Cambridge Philos. Soc.*, 50 (1954); M. Nagata, *Mem. Coll. Sci. Univ. Kyôto*, 30 (1956).) The set $\tilde{R}$ of elements of R'' which are integral over R is a ring and is called the **integral closure** of R in R''. If $\tilde{R} = R$, then R is said to be **integrally closed in** R''. If R is integrally closed in its ring of total quotients, we say that R is **integrally closed**. An integrally closed integral domain is called a **normal ring**. (In some literature, an integrally closed ring is called a normal ring.) An element $a \in R''$ is called **almost integral** over R if there is an element b of R such that b is not a zero divisor and $a^n b \in R$ for every natural number n. If an element a of the ring of total quotients of R is integral over R, then a is almost integral over R. R is said to be **completely integrally closed** if its ring of total quotients contains no elements which are almost integral over R except the elements of R itself.

J. Group Theorem

Let Q be the ring of total quotients of a ring R. An R-submodule $\mathfrak{a}$ of Q is called a **fractional ideal** of R if there is a non-zero-divisor c of R such that $c\mathfrak{a} \subset R$. The product of fractional ideals is defined similarly as in the case of products of ideals. The inverse $\mathfrak{a}^{-1}$ of the fractional ideal $\mathfrak{a}$ is defined to be $\{x \in Q \mid x\mathfrak{a} \subset R\}$. If $\mathfrak{a}$ contains an element which is not a zero divisor, then $\mathfrak{a}^{-1}$ is also a fractional ideal. When R is completely integrally closed, we define fractional ideals $\mathfrak{a}$ and $\mathfrak{b}$ to be equivalent if $\mathfrak{a}^{-1} = \mathfrak{b}^{-1}$. This gives rise to an equivalence relation between fractional ideals. The set of equivalence classes of fractional ideals which contain non-zero-divisors forms a group. This result is called the **group theorem**.

An integral domain R is called a **Krull ring** if (i) for every prime ideal $\mathfrak{p}$ of height 1, the ring $R_\mathfrak{p}$ is a †discrete valuation ring; (ii) R is the intersection of all the valuation rings $R_\mathfrak{p}$ and (iii) every nonzero element a of R is contained in only a finite number of prime ideals of height 1. In a Krull ring R, for an arbitrary nonzero fractional ideal $\mathfrak{a}$, there is a uniquely determined product of powers of prime ideals of height 1 which is equivalent to $\mathfrak{a}$ in the sense stated above (→ 439 Valuations).

K. Dedekind Domains and Principal Ideal Domains

A ring R is called a **Dedekind domain** if (i) R is a †Noetherian integral domain, (ii) R is a normal ring, and (iii) the Krull dimension of R is 1. For an integral domain R which is not a field, R is a Dedekind domain if and only if the set of all nonzero fractional ideals is a group, if and only if every nonzero ideal of R is expressed as the product of a finite number of prime ideals and such expression is unique up to the order of prime factors. An important example of a Dedekind domain is the ring of all †algebraic integers, i.e., the †principal order of an †algebraic number field of finite degree. In general, if R is a Dedekind domain with field of quotients K and L is a finite algebraic extension of K, then the integral closure of R in L and any ring R' such that $R \subset R' \subsetneq K$ are Dedekind domains.

An ideal generated by an element is called a **principal ideal**; a fractional ideal generated by an element is called a **principal fractional ideal** (or simply **principal ideal**). In a Dedekind domain, the set P of nonzero principal fractional ideals is a subgroup of the group I of nonzero fractional ideals; I/P is called the **ideal class group** of R, and a member of it is called an **ideal class**. The †order of I/P is called the **class number** of R (→ 14 Algebraic Number Fields). There are many Dedekind domains whose class numbers are infinite.

A ring R is called a **principal ideal ring** if every ideal is principal; furthermore, if R is an integral domain, then R is called a **principal ideal domain**. A principal ideal ring is the direct sum of a finite number of rings of which each direct summand is either a principal ideal domain or a †local ring whose maximal ideal is a principal nilpotent ideal. A principal ideal domain which is not a field is a Dedekind domain and a u.f.d. We consider, for an arbitrary natural number n and a principal ideal ring R, the set $M(n, R)$ of all $n \times n$ matrices over R. Given an element A of $M(n, R)$, there exist elements X, Y in $M(n, R)$ such that (i) X^{-1}, Y^{-1} are in $M(n, R)$, and (ii) denoting by b_{ij} the (i,j)-entry of XAY, we have $b_{11}R \supset b_{22}R \supset \ldots \supset b_{nn}R$ and $b_{ij} = 0$ if $i \neq j$. The nonzero members of the set $\{b_{11}, b_{22}, \ldots, b_{nn}\}$ are called the **elementary divisors** of the matrix A. Applying this to a finite module M over the principal ideal ring R, we see that M is the direct sum of $m_1 R, \ldots, m_r R$ $(m_i \in M)$ such that, with $\mathfrak{a}_i = \{x \in R \mid m_i x = 0\}$, we have $\mathfrak{a}_1 \subset \mathfrak{a}_2 \subset \ldots \subset \mathfrak{a}_r$ (→ 2 Abelian Groups B; 269 Matrices E).

L. Euclid Rings

A ring R is called a **Euclid ring** if there is a map φ of $R-\{0\}$ into a †well-ordered set W (the set of natural numbers is a special case) satisfying the condition that if $a, b \in R$, $a \neq 0$, then there are $r, q \in R$ such that $b = aq + r$ and either $r = 0$ or $\varphi(r) < \varphi(a)$.

Every Euclid ring is a principal ideal ring. Besides the ring $\mathbf{Z}$ of rational integers, there are several familiar examples of Euclid rings, such as $\mathbf{Z}[\sqrt{-1}]$, $\mathbf{Z}[\sqrt{-2}]$, $\mathbf{Z}[\omega]$ ($\omega^3 = 1, \omega \neq 1$), and the polynomial ring of one variable over a field.

References

[1] W. Krull, Idealtheorie, Erg. Math., Springer, second edition, 1968.
[2] B. L. van der Waerden, Algebra I, II, Springer, 1966, 1967.
[3] O. Zariski and P. Samuel, Commutative algebra I, II, Van Nostrand, 1958, 1960.
[4] M. Nagata, Local rings, Interscience, 1962.
[5] N. Bourbaki, Eléments de mathématique, Algèbre, ch. 1, 8, Actualités Sci. Ind., Hermann, 1144b, 1964; 1261a, 1958.
[6] N. Bourbaki, Eléments de mathématique, Algèbre commutative, ch. 1–7, Actualités Sci. Ind., Hermann, ch. 1, 2, 1290a, 1961; ch. 3, 4, 1293a, 1967; ch. 5, 6, 1308, 1964; ch. 7, 1314, 1965.
[7] A. Grothendieck, Eléments de géometrie algébrique, Publ. Math. Inst. HES, 4 (1960).
[8] D. G. Northcott, Lessons on rings, modules and multiplicities, Cambridge Univ. Press, 1968.

68 (XII.10) Compact and Nuclear Operators

A. General Remarks

Let X be a finite-dimensional linear space. Then a linear operator in X is surjective if and only if it is injective. If X is infinite-dimensional, this is no longer the case in general. For an operator of the form $1+K$ with an integral operator K of continuous kernel, I. Fredholm [1] developed the theory of determinants and retrieved the above equivalence (→ 217 Integral Equations). Later, F. Riesz [3] simplified the proof and showed that the equivalence holds if K is a compact operator in a Banach space. I. Ts. Gokhberg and M. G. Kreĭn [4] reformulated the result as the stability of indices of operators. The concept of compact operators itself was introduced by D. Hilbert (who used the terminology, "completely continuous operators"). The Hilbert-Schmidt class is also due to him. These classes of operators are employed in the spectral theory of integral operators [2]. The trace class of operators in Hilbert spaces is defined as the class of operators for which the trace has meaning. A. Grothendieck [5] introduced nuclear operators, extending the definition of trace class operators to general Banach spaces. Nuclear operators and related classes of operators, such as integral operators and absolutely summing operators, play important roles in the theory of topological tensor products [5], determinants [6], vector measures, and measures in linear spaces [7].

B. Compact Operators

A †linear operator T from a †Banach space X to a Banach space Y is said to be **compact** (or **completely continuous**) if T maps any †bounded set of X to a †relatively compact set of Y. In other words, T is compact if, for any bounded sequence $\{x_n\}$ in X, the sequence $\{Tx_n\}$ in Y contains a †strongly convergent subsequence. A compact operator is necessarily †bounded and hence continuous. A compact operator from X to Y maps any †weakly convergent sequence in X to a strongly convergent sequence in Y. If X is †reflexive, the converse is also true.

In this article the set of all bounded (resp. compact) linear operators from X to Y are denoted by $\mathbf{B}(X, Y)$ (resp. $\mathbf{B}^{(c)}(X, Y)$).

C. Examples of Compact Operators

(1) **Degenerate operators**. An operator $T \in \mathbf{B}(X, Y)$ is said to be degenerate or of **finite rank** if the †range $R(T)$ of T is finite-dimensional. A degenerate operator is necessarily compact. The †identity operator in X is compact ($Y = X$) if and only if X is finite-dimensional. (2) †**Integral operators with continuous kernel**. Let E and F be bounded closed regions in the Euclidean spaces $\mathbf{R}^m$ and $\mathbf{R}^n$, respectively, and let $k = k(t,s)$, $t \in F$, $s \in E$, be a continuous function defined on $F \times E$. Then the integral operator T with †kernel k,

$$(Tx)(t) = \int_E k(t,s)x(s)\,ds, \qquad t \in F, \tag{1}$$

is a compact operator from the Banach space $C(E)$ to $C(F)$. (For the notation for various function spaces → 168 Function Spaces.) (3) **Integral operators of Hilbert-Schmidt type**. In example (2) let E and F be †Lebesgue mea-

surable sets, and let $k \in L_2(F \times E)$. Then the integral operator T determines a compact operator from the †Hilbert space $L_2(E)$ to the Hilbert space $L_2(F)$. (4) Let Ω be a bounded open set in $\mathbf{R}^n$, $s<t$ two real numbers, and $H^s(\Omega)$ the †Sobolev space of order s (→ 168 Function Spaces). Then the natural injection from $H^t(\Omega)$ into $H^s(\Omega)$ is compact (**F. Rellich's lemma**).

D. Properties of Compact Operators

Any linear combination of compact operators is again compact. If a sequence $\{T_n\}$ of compact operators converges in the norm (i.e., in the †uniform operator topology), then the limit T is compact. Thus $\mathbf{B}^{(c)}(X, Y)$ is a closed subspace of the Banach space $\mathbf{B}(X, Y)$ with the †operator norm. Any product of a compact operator and a bounded operator is compact. Namely, $A \in \mathbf{B}^{(c)}(X, Y)$, $B \in \mathbf{B}(Y, Z)$, and $C \in \mathbf{B}(Z, X)$ imply $BA \in \mathbf{B}^{(c)}(X, Z)$ and $AC \in \mathbf{B}^{(c)}(Z, Y)$. In particular, $\mathbf{B}^{(c)}(X) = \mathbf{B}^{(c)}(X, X)$ forms a closed †two-sided ideal of $\mathbf{B}(X) = \mathbf{B}(X, X)$. An operator $T \in \mathbf{B}(X, Y)$ is compact if and only if its †dual operator $T' \in \mathbf{B}(Y', X')$ is compact. The range of a compact operator is always †separable. Let X and Y be Hilbert spaces. Then for any $T \in \mathbf{B}^{(c)}(X, Y)$ there exist †orthonormal sets $\{\varphi_n\}$ in X and $\{\psi_n\}$ in Y and a sequence $\{c_n\}$ of nonnegative numbers with $\lim c_n = 0$ such that

$$Tu = \sum_n c_n(u, \varphi_n)\psi_n, \qquad u \in X.$$

Consequently, any compact operator between Hilbert spaces X, Y can be approximated by a sequence of degenerate operators in the operator norm. However, there are Banach spaces X and Y for which the statement is no longer true. In fact, a Banach space X (resp. the dual Y' of a Banach space Y) has the †approximation property (→ 37 Banach Spaces L) if and only if every $T \in \mathbf{B}^{(c)}(Y, X)$ is the limit in norm of a sequence of degenerate operators for any Banach space Y (resp. X).

E. The Riesz-Schauder Theorem

Let $T \in \mathbf{B}^{(c)}(X)$ and consider a pair of linear equations

$$u - Tu = f, \qquad f \in X, \tag{2}$$

$$\varphi - T'\varphi = g, \qquad g \in X', \tag{3}$$

where $T' \in \mathbf{B}(X')$ is the dual operator of T in the †dual space X' of X. Put $\mathcal{M} = \{u \in X \mid u = Tu\}$ and $\mathcal{M}' = \{\varphi \in X' \mid \varphi = T'\varphi\}$. Then one and only one of the following two cases (i) and (ii) occurs. (i) $\mathcal{M} = \{0\}$, $\mathcal{M}' = \{0\}$; for any $f \in X$ equation (2) has a unique solution; and for any $g \in X'$ equation (3) has a unique solution. (ii) $\dim \mathcal{M} = \dim \mathcal{M}' = m$, $1 \leqslant m < \infty$; (2) has a solution if and only if f is orthogonal to $\mathcal{M}'$ (i.e., $\varphi(f) = 0$ for any $\varphi \in \mathcal{M}'$); and (3) has a solution if and only if g is orthogonal to $\mathcal{M}$ (i.e., $g(u) = 0$ for any $u \in \mathcal{M}$), This is called the **Riesz-Schauder theorem**. In particular, when T is an integral operator in a suitable function space, this theorem is also called **Fredholm's alternative theorem** for integral equation (2).

F. Fredholm Operators and Their Indices

Let T be a †closed linear operator from a Banach space X to a Banach space Y. T is called a **Fredholm operator** or an **operator with index** if both $\operatorname{Ker} T$ (= null space $N(T)$) and $\operatorname{Coker} T$ ($= Y/R(T)$) are finite-dimensional. Then the integer $\operatorname{ind} T = \dim \operatorname{Ker} T - \dim \operatorname{Coker} T$ is called the **index** of T. If a closed linear operator T has $\operatorname{Coker} T$ of finite dimension, then the range $R(T)$ is closed. Moreover, if $R(T)$ is closed and the domain $D(T)$ is dense, then $\dim \operatorname{Coker} T = \dim \operatorname{Ker} T'$, where T' is the dual of T, and hence $\operatorname{ind} T = \dim \operatorname{Ker} T - \dim \operatorname{Ker} T'$. A linear operator K from X to Y is said to be **T-compact** if the domain $D(K)$ contains $D(T)$ and if K from $D(T)$ with the †graph norm into Y is compact, i.e., for any bounded sequence $\{x_n\}$ in $D(T)$ with $\sup \|Tx_n\| < \infty$, the sequence $\{Kx_n\}$ contains a strongly convergent subsequence.

The following are basic properties of Fredholm operators [4, 11]: (1) If T is Fredholm and K is T-compact, then $T+K$ is Fredholm and $\operatorname{ind}(T+K) = \operatorname{ind} T$. (2) If T is Fredholm and another linear operator S is sufficiently close to T in (graph) norm, then S is also Fredholm and $\operatorname{ind} S = \operatorname{ind} T$. (3) If T is a Fredholm operator from X into Y and S is a Fredholm operator with dense domain in Y into Z, then ST is Fredholm and $\operatorname{ind} ST = \operatorname{ind} S + \operatorname{ind} T$. If $D(T)$ is dense, then $D(ST)$ is also dense. (4) Let T be a closed linear operator with dense domain; then T is Fredholm if and only if the dual T' is Fredholm, and then $\operatorname{ind} T' = -\operatorname{ind} T$.

Fredholm operators and their indices were first studied by Russian mathematicians in connection with boundary value problems of differential equations and singular integral equations. M. F. Atiyah and I. M. Singer have proved that an †elliptic linear differential operator P of order m on a compact differentiable manifold M is a Fredholm operator from the Sobolev space $H^s(M)$ into $H^{s-m}(M)$ for any s and that its index is computed from the †symbol of P and the †characteristic classes of M (→ 237 K-Theory). Similarly, the indices of linear ordinary differential operators (and

more generally of maximally overdetermined systems of linear partial differential operators) are computed in various spaces of functions and generalized functions (H. Komatsu, B. Malgrange, J.-P. Ramis, M. Kashiwara).

G. Spectra of Compact Operators

The following structure of the †spectrum of compact operators $T \in \mathbf{B}^{(c)}(X)$ is derived from the Riesz-Schauder theorem. The spectrum $\sigma(T)$ of T consists of at most countably many points and has no accumulation points except possibly for 0. Any nonzero point of $\sigma(T)$ is an †eigenvalue of T. When X is infinite-dimensional, 0 always belongs to $\sigma(T)$ but is not necessarily an eigenvalue of T. Each nonzero eigenvalue of T has finite (algebraic) †multiplicity, and hence the †eigenspace $\mathscr{M}_\lambda(T) = \{u \mid Tu = \lambda y\}$, $\lambda \neq 0$, is finite-dimensional. If λ is an eigenvalue of T, then $\bar{\lambda}$ is an eigenvalue of the †adjoint operator T^* of T with the same (either algebraic or geometric) multiplicity as that of λ.

H. Spectral Representations of Compact Normal Operators

Let T be a compact †normal operator in a Hilbert space H. Then we can find a †complete orthonormal set consisting solely of eigenvectors of T. Namely, for each nonzero eigenvalue λ_j of T, take an orthonormal basis $\{\varphi_k^{(j)}\}$ of the eigenspace associated with λ_j. Rearrange all the $\varphi_k^{(j)}$ into a sequence $\{\varphi_n\}$ and add to it, if 0 is an eigenvalue, a complete orthonormal set of the eigenspace associated with 0. Then we obtain a desired complete orthonormal set of H. Let μ_n be the eigenvalue associated with φ_n. Then the sequence $\{\mu_n\}$ is precisely an enumeration of nonzero eigenvalues of T with repetitions according to multiplicity. In terms of $\{\varphi_n\}$ and $\{\mu_n\}$, a †spectral representation of T is given as

$$Tu = \sum_{n=1}^{\infty} \mu_n(u, \varphi_n)\varphi_n, \qquad u \in H.$$

The eigenvalue problem of a compact nonnegative †self-adjoint operator T can be solved by means of the following **Rayleigh principle**. Consider **Rayleigh's quotient** $R(x) = (Tx, x)/\|x\|^2$, $x \neq 0$. The largest eigenvalue μ_1 is obtained as $\mu_1 = \max_{x \in X} R(x)$ and any vector x_1 which attains this maximum is an eigenvector. When the largest $n-1$ eigenvalues $\mu_1, \ldots, \mu_{n-1}$ and eigenvectors $x_1, \ldots, x_{n-1}$ are determined, the nth eigenvalue μ_n is the maximum of $R(x)$ on the subspace orthogonal to $x_1, \ldots, x_{n-1}$ and any vector which attains the maximum is an eigenvector. A more direct characterization of μ_n, involving no previous eigenvectors, is given by

$$\mu_n = \min_{\substack{f_1, \ldots, f_{n-1} \in X \\ f_1, \ldots, f_{n-1} \neq 0}} \left(\max_{\substack{(x, f_j) = 0 \\ j = 1, \ldots, n-1}} R(x) \right).$$

This formula is referred to as the **minimax principle**.

I. Classification of Compact Operators in Hilbert Spaces

Let H and K be Hilbert spaces. Then a $T \in \mathbf{B}(H, K)$ is compact if and only if $(Te_n, f_n) \to 0$ for any orthonormal sequences e_n in H and f_n in K. J. von Neumann and R. Schatten (*Ann. Math.*, 49 (1948)) classified operators $T \in \mathbf{B}^{(c)}(H, K)$ in the following way. The operator $A = (T^*T)^{1/2}$ is a compact nonnegative self-adjoint operator. Let $\alpha_1 \geqslant \alpha_2 \geqslant \ldots$ be the enumeration in decreasing order of the positive eigenvalues of A, with each repeated according to its multiplicity. The $\alpha_n = \alpha_n(T)$ are sometimes called the **characteristic numbers** of T. For any $p > 0$ the set of all $T \in \mathbf{B}^{(c)}(H)$ such that

$$\|T\|_p = \left(\sum_{n=1}^{\infty} \alpha_n^p \right)^{1/p} < +\infty$$

is denoted by $\mathbf{B}_p(H)$ (or simply $\mathbf{B}_p$). Among these classes, $\mathbf{B}_1$ and $\mathbf{B}_2$ are most important. $\mathbf{B}_1$ is called the **trace class** and $\mathbf{B}_2$ the **Hilbert-Schmidt class**. Correspondingly, $\|T\|_1$ and $\|T\|_2$ are called the **trace norm** and **Hilbert-Schmidt norm** of T, respectively. $\mathbf{B}_1$ is also called the **nuclear class** and any operator $T \in \mathbf{B}_1$ a †**nuclear operator**. The norm $\|T\|_p$ can also be defined more directly as follows. (i) If $1 < p < \infty$, $\|T\|_p = \sup \|(Te_n, f_n)\|_{l_p}$; (ii) If $0 < p \leqslant 2$, $\|T\|_p = \inf \|(\|Te_\nu\|)\|_{l_p}$; (iii) If $2 \leqslant p < \infty$, $\|T\|_p = \sup \|(\|Te_\nu\|)\|_{l_p}$, where e_n and f_n range over the orthonormal sequences in H and K, and e_ν over the orthonormal bases in H. The class $\mathbf{B}_p$ is a two-sided * ideal in the Banach algebra $\mathbf{B}$. Moreover precisely, $T \in \mathbf{B}_p$ and $R \in \mathbf{B}$ imply $\|RT\|_p \leqslant \|R\| \, \|T\|_p$, $\|TR\|_p \leqslant \|R\| \, \|T\|_p$ and $\|T^*\|_p = \|T\|_p$. When $1 \leqslant p < \infty$, the class $\mathbf{B}_p$ becomes a Banach space with the norm $\|T\|_p$. It is always a †quasi-Banach space, in which the set of all degenerate operators is dense.

The norm $\|T\|_p$ $(T \neq 0)$ is a decreasing function of p. Hence $\mathbf{B}_p \subset \mathbf{B}_q$ if $p \leqslant q$. Also, $T \in \mathbf{B}_p$ and $S \in \mathbf{B}_q$ imply $TS \in \mathbf{B}_r$, where $1/r = 1/p + 1/q$. Let $T \in \mathbf{B}_p$, and let $\{\mu_j\}$ be an enumeration of eigenvalues of T with repetitions according to (either geometric or algebraic) †multiplicity. Then

$$\sum_j |\mu_j|^p \leqslant \|T\|_p^p.$$

Hilbert-Schmidt class. For an arbitrary complete orthonormal set $\{u_k\}$, we have

$$\|T\|_2^2 = \sum \|Tu_k\|^2.$$

Thus $\mathbf{B}_2$ can be defined as the set of all $T \in \mathbf{B}$ such that the sum on the right-hand side is finite for a certain complete orthonormal set $\{u_k\}$. A linear operator $T: L_2(E) \rightarrow L_2(F)$ is of Hilbert-Schmidt class if and only if it is an integral operator of Hilbert-Schmidt type (example (3) in Section C). The space $\mathbf{B}_2$ with the norm $\|T\|_2$ becomes a Hilbert space with the inner product defined by

$$(T, S)_2 = \sum (Tu_k, Su_k),$$

where $\{u_k\}$ is as above.

Trace class. For an operator $T \in \mathbf{B}_1$, the trace $\mathrm{tr}(T)$ of T is defined as

$$\mathrm{tr}(T) = \sum (Tu_k, u_k).$$

Here the right-hand side converges absolutely and does not depend on the complete orthonormal set $\{u_k\}$. The trace is a bounded linear functional on the Banach space $\mathbf{B}_1$. The product of two operators of Hilbert-Schmidt type belongs to the trace class, and the converse is also true. If $1 < p, q < \infty$ satisfy $1/p + 1/q = 1$, then the inner product $\langle T, S \rangle$ of $T \in \mathbf{B}_p(H, K)$ and $S \in \mathbf{B}_q(K, H)$ is defined by $\langle T, S \rangle = \mathrm{tr}(ST)$. Under this inner product the dual of the Banach space $\mathbf{B}_p(H, K)$ is identified with $\mathbf{B}_q(K, H)$. Similarly, the dual of $\mathbf{B}^{(c)}(H, K)$ is isomorphic to $\mathbf{B}_1(K, H)$, and the dual of $\mathbf{B}_1(K, H)$ to $\mathbf{B}(H, K)$ (J. Dixmier, Schatten).

J. Volterra Operators

A compact operator T in a Hilbert space is said to be a **Volterra operator** if it is †quasi-nilpotent, i.e., its †spectral radius is 0. The integral operator

$$(Tx)(t) = \int_t^b k(t, s) x(s)\, ds, \qquad x \in L_2(a, b),$$

appearing in the integral equation of Volterra type is a Volterra operator. Conversely, a Volterra operator satisfying a suitable condition is unitarily equivalent to such an integral operator (M. S. Livshits, Gokhberg and Kreĭn [13]). A Volterra operator admits an abstract triangular representation in a manner similar to Jordan's canonical form (→ 390 Spectral Analysis of Operators H).

K. Nuclear Operators

Extending the definition of trace class to operators in Banach spaces, A. Grothendieck [5] defined a **nuclear operator** (or **Fredholm operator** according to [6]) from a Banach space X into a Banach space Y to be a linear operator $T: X \rightarrow Y$ that is represented as $Tx = \sum \lambda_j \langle x, a_j \rangle b_j$ with a sequence $\lambda_j \geqslant 0$ in l_1, a bounded sequence a_j in X', and a bounded sequence b_j in Y. In other words, a linear operator $T: X \rightarrow Y$ is nuclear if and only if it is decomposed as the product

$$X \xrightarrow{A} l_\infty \xrightarrow{\Lambda} l_1 \xrightarrow{B} Y, \qquad (4)$$

where A and B are bounded linear operators and Λ is multiplication by λ_j in l_1. The infimum of $\|A\| \, \|\lambda_j\|_{l_1} \|B\|$ is called the **nuclear norm** of T and is denoted by $\|T\|_1$. When X and Y are Hilbert spaces, this coincides with the trace norm. The integral operator T defined by (1) is nuclear with $\|T\|_1 = \int \sup_t |k(t,s)|\, ds$. The totality $\mathbf{B}_1(X, Y)$ of nuclear operators $T: X \rightarrow Y$ forms a Banach space under the nuclear norm. If T is nuclear and A is bounded, then $\|TA\|_1 \leqslant \|T\|_1 \|A\|$ and $\|AT\|_1 \leqslant \|A\| \, \|T\|_1$. If T is nuclear, then the dual T' is nuclear, and $\|T'\|_1 \leqslant \|T\|_1$. Suppose that X and Y are Banach spaces satisfying one of the following conditions: (i) Y is the dual of a Banach space; (ii) X' has the approximation property; (iii) Y'' has the approximation property. Then conversely a $T \in \mathbf{B}(X, Y)$ is nuclear if the dual T' is nuclear, and $\|T\|_1 = \|T'\|_1$. However, this is not necessarily the case in general (T. Figiel and W. B. Johnson).

Replacing l_1 by l_p, we obtain the definition of **operators** $T: X \rightarrow Y$ **of summable** p**th power**. Grothendieck [5] considered the case $0 < p \leqslant 1$ and showed that if $T_1, \ldots, T_n$ are of summable $p_1, \ldots, p_n$th power, then the product $T_1 \ldots T_n$ is of summable rth power, where r is given by $1/r = (\sum 1/p_i) - (n+1)/2$.

L. Traces and Determinants of Operators

Let T be a linear operator of †finite rank in a linear space X. There are a finite number of elements $a_i \in X'$ and $b_i \in X$ such that

$$Tx = \sum_{i=1}^{n} \langle x, a_i \rangle b_i. \qquad (5)$$

Then $\mathrm{tr}(T) = \sum \langle b_i, a_i \rangle$ and $\det(1 - T) = \det(\delta_{i,j} - \langle b_i, a_j \rangle)$ are independent of the representation (5). Let X be a Banach space. If $T \in \mathbf{B}_1(X)$ is represented as $Tx = \sum \lambda_j \langle x, a_j \rangle b_j$, then it seems reasonable that the **trace** $\mathrm{tr}(T)$ be defined by $\sum \lambda_j \langle b_j, a_j \rangle$, but the sum may depend on the representation. It is known that a Banach space X has the approximation property if and only if the sum does not depend on the representation for any nuclear operator T. However, if T is of summable (2/3)rd power, the trace $\mathrm{tr}(T)$ is always defined uniquely. If X or T satisfies the afore-

mentioned condition, then the **determinant** $\det(1-T)$ is also defined uniquely as the limit of $\det(1-T_n)$, where T_n is the nth partial sum. $\det(1-zT)$ is an entire function of z and its zeros are exactly the reciprocals of the nonzero eigenvalues of T. If $\det(1-zT)\neq 0$, then the resolvent $(1-zT)^{-1}$ is given by **Fredholm's formula** extending †Cramer's rule. Eigenvectors are also computed explicitly [6].

Let H be a Hilbert space, and let $T\in\mathbf{B}_1(H)$. Then $\det(1-zT)=\prod(1-z\lambda_j)$, where λ_j are the nonzero eigenvalues of T, and is an entire function of genus 0. Moreover, let $p\geqslant 2$ be an integer. Then the modified determinant

$$\det_p(1-zT)=\prod_j\left[(1-z\lambda_j)\exp\left(\sum_{k=1}^{p-1}k^{-1}z^k\lambda_j^k\right)\right]$$

is defined for $T\in\mathbf{B}_p(H)$ and is an entire function of genus $p-1$. This type of determinant was introduced by T. Carleman (*Math. Z.*, 9 (1921)) for the Hilbert-Schmidt class of operators (→ also Hilbert [2]). It is utilized to prove the completeness of the †root vectors of T and many other facts [9, 12, 13].

M. Weakly Compact Operators

A linear operator from a Banach space X to a Banach space Y is said to be **weakly compact** if T maps any bounded set of X to a relatively weakly compact set. Weakly compact operators are bounded. They have properties similar to those of compact operators. The following are equivalent conditions for a bounded linear operator $T:X\rightarrow Y$. (1) T is weakly compact; (2) $T':Y'\rightarrow X'$ is weakly compact; (3) $T'':X''\rightarrow Y''$ maps X'' into Y; (4) There is a reflexive Banach space Z and bounded linear operators $S:X\rightarrow Z$ and R: $Z\rightarrow Y$ such that $T=RS$. The last characterization is due to W. J. Davis, Figiel, Johnson, and A. Pełczyński. Any linear combination of weakly compact operators is weakly compact. The uniform limit of a sequence of weakly compact operators is weakly compact. The product of a weakly compact operator and a bounded operator is weakly compact.

Dunford-Pettis theorem ([9]; R. S. Phillips, *Trans. Amer. Math. Soc.*, 48 (1940)). Let (Ω,μ) be a σ-finite measure space and X a Banach space. If $T:L_1(\Omega)\rightarrow X$ is a weakly compact linear operator, then there exists a bounded strongly measurable function g on Ω with values in X such that

$$Tf=\int f(t)g(t)\,d\mu(t).$$

Then T maps weakly convergent sequences in $L_1(\Omega)$ into strongly convergent sequences in X. In particular, the product of two weakly compact linear operators in $L_1(\Omega)$ is compact. Similarly, Grothendieck and R. G. Bartle, N. Dunford, and J. Schwartz (→ [9]) have proved the following: Let Ω be a compact Hausdorff space and X a Banach space. If $T:C(\Omega)\rightarrow X$ is a weakly compact linear mapping, then there is a †vector measure μ defined on the †Borel sets in Ω with values in X such that $\langle\mu,x'\rangle$ is a †Radon measure for any $x'\in X'$ and

$$Tf=\int f(t)\,d\mu(t).$$

T maps weakly convergent sequences in $C(\Omega)$ into strongly convergent sequences in X. The product of two weakly compact linear operators in $C(\Omega)$ is also compact.

N. Absolutely Summing Operators

A linear operator T from a Banach space X to a Banach space Y is called an **integral operator** if there is a positive †Radon measure μ on the product $A\times B$ of the unit ball A in X' and the unit ball B in Y'' both equipped with the weak* topology such that

$$\langle Tx,y'\rangle=\int\langle x,x'\rangle\langle y',y''\rangle\,d\mu(x',y''),$$

or equivalently if $T:X\rightarrow Y''$ is decomposed as (4) with l_∞ and l_1 replaced by $L_\infty(M)$ and $L_1(M)$ for a suitable compact space M with a Radon measure. Nuclear operators are clearly integral. Integral operators are weakly compact, but they are not necessarily compact or nuclear. However, the product of an integral operator and a weakly compact operator is nuclear, and hence every integral operator in a reflexive Banach space is nuclear.

If a compact (resp. weakly compact) operator T from a Banach space X to a Banach space Y maps a closed linear subspace X_1 of X into a closed linear subspace Y_1 of Y, then the restriction $T_1:X_1\rightarrow Y_1$ and the induced operator $T^1:X/X_1\rightarrow Y/Y_1$ are also compact (resp. weakly compact). The corresponding result does not hold for nuclear operators and integral operators. The following conditions are equivalent for a bounded linear operator $T:X\rightarrow Y$: (i) There is a Banach space $\tilde{Y}\supset Y$ such that $T:X\rightarrow\tilde{Y}$ is integral; (ii) there is a positive Radon measure μ on the unit ball A of X' such that

$$\|Tx\|\leqslant\int|\langle x,x'\rangle|\,d\mu(x');$$

(iii) there is a constant C such that

$$\sum\|Tx_i\|\leqslant C\sup\{\sum|\langle x_i,x'\rangle|\mid\|x'\|\leqslant 1\}$$

holds for any finite set $\{x_1,\ldots,x_n\}$ in X; (iv) if $\sum x_i$ is an †unconditionally convergent series

in X, then $\sum \|Tx_i\|$ converges. Grothendieck called such an operator a **right semi-integral** in view of property (i), and A. Pietsch called it **absolutely summing** in view of (iv). Similarly, a bounded linear operator $T: X \to Y$ is called a **left semi-integral** if there is a Banach space $\tilde{X}$ such that $X = \tilde{X}/\tilde{X}_1$ for a closed subspace $\tilde{X}_1$, and the product $\tilde{T}: \tilde{X} \to Y$ is integral. The product of two right (resp. left) semi-integral operators is nuclear. A bounded linear operator in a Hilbert space is right (or left) semi-integral if and only if it is of Hilbert-Schmidt class. A bounded linear operator T from a Banach space X into a Banach space Y is integral if and only if the dual T' is integral. It is a right (resp. left) semi-integral if and only if T' is a left (resp. right) semi-integral.

For other related classes of operators → Grothendieck's paper in *Boletin Soc. Mat. São Paulo*, 8 (1953).

References

[1] I. Fredholm, Sur une classe d'équations fonctionnelles, Acta Math., 27 (1903), 365–390.
[2] D. Hilbert, Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen, Teubner, 1912.
[3] F. Riesz, Über lineare Funktionalgleichungen, Acta Math., 41 (1918), 71–98.
[4] I. Ts. Gokhberg (Gohberg) and M. G. Kreĭn, The basic propositions on defect numbers, root numbers and indices of linear operators, Amer. Math. Soc. Transl., ser. 2, 13 (1960), 185–264. (Original in Russian, 1957.)
[5] A. Grothendieck, Produits tensoriels topologiques et espaces nucléaires, Mem. Amer. Math. Soc., 16 (1955).
[6] A. Grothendieck, Théorie de Fredholm, Bull. Soc. Math. France, 84 (1956), 319–384.
[7] L. Schwartz, Geometry and probability in Banach spaces, Lecture notes in math. 852, Springer, 1980.
[8] F. Riesz and B. Sz.-Nagy, Leçon d'analyse fonctionnelle, Akad. Kiadó, Budapest, 1952.
[9] N. Dunford and J. T. Schwartz, Linear operators I, II, Interscience, 1958, 1963.
[10] K. Yosida, Functional analysis, Springer, 1965.
[11] T. Kato, Perturbation theory for linear operators, Springer, 1966.
[12] I. Ts. Gokhberg (Gohberg) and M. G. Kreĭn, Introduction to the theory of linear nonselfadjoint operators, Amer. Math. Soc., 1969. (Original in Russian, 1965.)
[13] I. Ts. Gokhberg (Gohberg) and M. G. Kreĭn, Theory and applications of Volterra operators in Hilbert space, Amer. Math. Soc., 1970. (Original in Russian, 1967.)
[14] A. Pietsch, Nuclear locally convex spaces, Springer, 1972.

69 (IV.8) Compact Groups

A. Compact Groups

A †topological group G is called a **compact group** if the underlying topological space of G is a †compact Hausdorff space. The †torus group $\mathbf{T}^n = \mathbf{R}^n/\mathbf{Z}^n$ ($n = 1, 2, \ldots$) (commutative group), the †orthogonal group $O(n)$, the †unitary group $U(n)$, the †symplectic group $Sp(n)$, and the additive group $\mathbf{Z}_p$ of †p-adic integers are compact groups (→ 60 Classical Groups; for other compact Lie groups → 249 Lie Groups, 248 Lie Algebras). Let $C(G)$ be the †linear space formed by all the complex-valued continuous functions $f, g, h, \ldots$ defined on a compact group G; $C(G)$ is a †Banach space with the norm

$$\|f\| = \sup_{x \in G} |f(x)|.$$

Since a compact group G is †locally compact, there exists a right-invariant †Haar measure on G. Because of the compactness of G, the total measure of G is finite, and the measure is also left-invariant. By the condition that the total measure is 1, such a measure is uniquely determined. The integral of f in $C(G)$ relative to this measure is called the **mean value** of f. Since for $f, g \in C(G)$, $f(xy^{-1})g(y)$ is continuous in two variables x, y, the †convolution $f \times g(x) = \int f(xy^{-1})g(y)dy$ also belongs to $C(G)$. $C(G)$ constitutes a ring under the multiplication defined by the convolution. This ring can be considered as an extension of the notion of a †group ring for finite groups; it is called the **group ring** of the compact group G. The function $x \to \overline{f(x^{-1})}$ will be denoted by f^*, and the inner product in the †function space $L_2(G)$ will be written as (f, g).

B. Representations of Compact Groups

Let $G(E)$ be the group of units of †bounded linear operators on a Banach space E, and suppose that we have a homomorphism U of a topological group G into $G(E)$. The homomorphism U is called a **strongly (weakly) continuous representation** on E of G if, for any $a \in E$, the map $x \to U(x)a$ of G into E is continuous with respect to the †strong (†weak) topology on E. When E is a †Hilbert space, a strongly continuous representation U such that every

$U(a)$ is a unitary operator is called a †unitary representation. Let U be any strongly continuous representation of a compact group G on a Hilbert space E; for a, b in E, let $\langle a,b\rangle$ be the mean value of the †inner product $(U(x)a, U(x)b)$ on E. Then U is a unitary representation on the Hilbert space E with the new inner product $\langle a,b\rangle$.

A representation U of G on a Banach space E is said to be **irreducible** if E contains no closed subspace other than $\{0\}$ and E, which is invariant under every $U(x)(x \in G)$. If a weakly continuous representation of a compact group on a Banach space E is irreducible, then E is finite-dimensional. Moreover, any unitary representation U of a compact group on a Hilbert space E can be decomposed into a discrete †direct sum of irreducible representations. Namely, there exists a family $\{E_\alpha\}_{\alpha\in A}$ of irreducible (hence finite-dimensional) invariant subspaces E_α of E which are orthogonal to each other such that $E=\sum_{\alpha\in A}E_\alpha$. In particular, any continuous representation of a compact group on a finite-dimensional space is †completely reducible. In $L_2(G)$ with respect to the †Haar measure on a †locally compact group G, the representation U defined by $(U(x)f)(y)=f(yx)$ $(x,y\in G, f\in L_2(G))$ is a unitary representation of G. This representation U is called the (right) **regular representation** of G. Decomposition of the regular representation of a compact group G into irreducible representations is given by the **Peter-Weyl theory**, described later.

In the rest of this article, the representations under consideration will be (continuous) representations by matrices of finite degree. Let $D_1(x)=(d_{ij}^{(1)}(x))$ and $D_2(x)=(d_{ij}^{(2)}(x))$ be irreducible unitary representations which are not mutually †equivalent. Then from †Schur's lemma follow the orthogonality relations $(d_{ij}^{(1)}, d_{kl}^{(2)})=0$ and $(\sqrt{n_1}\,d_{ij}^{(1)}, \sqrt{n_1}\,d_{mn}^{(1)})=\delta_{im}\delta_{jn}$ (where n_1 is the degree of the representation D_1). From each †class D_α of irreducible representations of G, choose a unitary representative $D_\alpha(x)=(d_{ij}^\alpha(x))$, and let n_α denote its degree. Then from the orthogonality relations it follows that the $\sqrt{n_\alpha}d_{ij}^\alpha(x)$ form an †orthonormal system of $L_2(G)$.

Let $h\in C(G)$ and consider the map $H: f\rightarrow h\times f$ of $C(G)$ to $C(G)$. Then H is a †compact operator in $C(G)$. Since $C(G)$ is contained in $L_2(G)$, we can define the inner product in $C(G)$. By $(h\times f, g)=(f, h^*\times g)$, $h=h^*$ implies $(Hf, g)=(f, Hg)$; that is, H is a †Hermitian operator. For a given f in $C(G)$, there exists $h(=h^*)$ in $C(G)$ such that $h\times f$ is uniformly arbitrarily near f. From the theory of compact Hermitian operators, Hf can be uniformly approximated by linear combinations of the †eigenfunctions of H. Since the †eigenspace of H is finite-dimensional and invariant under $U(a)$, the eigenfunctions of H are linear combinations of a finite number of $\sqrt{n_\alpha}\,d_{ij}^\alpha(x)$. Hence any function f in $C(G)$ can be uniformly approximated by linear combinations of a finite number of $\sqrt{n_\alpha}\,d_{ij}^\alpha(x)$. This fact, like the similar result in †Fourier series, is called the **approximation theorem**. From this, it follows that the orthogonal system $\{\sqrt{n_\alpha}\,d_{ij}^\alpha(x)\}$ is †complete; i.e., if an element of $C(G)$ is orthogonal to each element in this system, then it is 0.

Since $C(G)$ is dense in $L_2(G)$, $\{\sqrt{n_\alpha}\,d_{ij}^\alpha(x)\}$ is a †complete orthonormal system of the Hilbert space $L_2(G)$. Hence for any $f\in L_2(G)$, its "Fourier series" $\sum_\alpha\sum_{ij}c_{ij}^\alpha\sqrt{n_\alpha}\,d_{ij}^\alpha(x)$ (where $c_{ij}=(f,\sqrt{n_\alpha}\,d_{ij}^\alpha)$) converges to f in the mean of order 2 (i.e., with respect to the †metric of $L_2(G)$). In particular, if G is a compact †Lie group and f is sufficiently many times differentiable, then this series converges uniformly to f.

The space V_i^α of dimension n_α spanned by the elements $d_{ij}^\alpha(x)$ $(1\leqslant j\leqslant n_\alpha)$ in the ith row of the matrix $D_\alpha(x)$ is invariant under the right regular representation U. The representation on V_i^α given by U can only be D_α. Then the fact that $\{\sqrt{n_\alpha}\,d_{ij}^\alpha(x)\}$ is an orthonormal system of $L_2(G)$ means that the regular representation U of a compact group G is decomposable into a discrete direct sum of finite-dimensional irreducible representations. Each irreducible representation D_α is contained in U with multiplicity equal to its degree n_α.

If a function $\varphi(x)$ in $C(G)$ satisfies $\varphi(y^{-1}xy)=\varphi(x)$ for any x, y, then it is called a **class function**. The set $K(G)$ of all the continuous class functions coincides with the †center of the group ring $C(G)$. The †character of an irreducible representation of G is a class function, and the set $\{\chi_\alpha(x)\}$ of all characters plays the same role as the orthogonal system $\{\sqrt{n_\alpha}\,d_{ij}^\alpha(x)\}$ in $C(G)$. Namely, $\{\chi_\alpha(x)\}$ is a complete orthonormal system in (the †completion of) $K(G)$, and any class function can be uniformly approximated by linear combinations of a finite number of these characters.

The preceding paragraphs give a brief description of the Peter-Weyl theory. If G is the 1-dimensional †torus group $\mathbf{T}^1=\mathbf{R}/\mathbf{Z}$, namely, the compact group of real numbers mod 1, then this is actually the theory of Fourier series concerning periodic functions on the line. (For concrete irreducible representations of $O(n)$, $U(n)$, $Sp(n)$, and formulas for characters → 60 Classical Groups. For representations of compact Lie groups → 249 Lie Groups, 248 Lie Algebras.) The theory of compact groups was completed by F. Peter and H. Weyl (*Math. Ann.*, 97 (1927)), and J. von Neumann's theory concerning almost periodic functions in a group (1934) united the

theory of compact groups with H. Bohr's theory of almost periodic functions (→ 18 Almost Periodic Functions).

C. Structure of Compact Groups

Let a be an element of a compact group G different from e. Since the underlying space of a topological group is a †completely regular space, there exists a function in $C(G)$ such that $f(a) \neq f(e)$. Hence there exists a representation $D(x)$ of G such that $D(a)$ is not equal to the unit matrix. This means that any compact group G can be expressed as a †projective limit group of compact Lie groups. Beginning with this fact, von Neumann (1933) showed that a †locally Euclidean compact group is a Lie group (→ 423 Topological Groups N).

D. Set of Representations

The set $G' = \{D\}$ of representations of G by matrices admits the following operations: (i) $D_1 \otimes D_2$ (†tensor product representation); (ii) $D_1 \oplus D_2 = \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix}$ (†direct sum representation); (iii) $P^{-1}DP$ (equivalent representation); and (iv) $\bar{D}$ (complex conjugate representation). Let M be a subset of G' such that $\bar{D} \in M$ whenever $D \in M$ and the irreducible components of $D_1 \otimes D_2$ are in M whenever $D_1, D_2 \in M$. Then M is called a **module of representations** of G. There is a one-to-one correspondence between closed normal subgroups H of G and modules M formed by all the representations of G/H.

A representation of G' is a correspondence which assigns to each D a matrix $A(D)$ of the same degree as that of D and preserves the operations of G': $A(D_1 \otimes D_2) = A(D_1) \otimes A(D_2)$, $A(D_1 \oplus D_2) = A(D_1) \oplus A(D_2)$, $A(P^{-1}DP) = P^{-1}A(D)P$, and $A(\bar{D}) = \overline{A(D)}$. Let G'' be the set of all the representations of G'. Define the product of $A_1, A_2 \in G''$ by $A_1A_2(D) = A_1(D)A_2(D)$ and a topology on G'' by the †weak topology of the functions $A(D)$ of D. Namely, a typical neighborhood of A_0 is of the form $U(A_0; D_1, \ldots, D_s; \varepsilon) = \{A \mid \|A(D_i) - A_0(D_i)\| < \varepsilon, i = 1, \ldots, s\}$. G'' is a topological group under this multiplication and topology. Then the **Tannaka duality theorem** states that $G'' \cong G$ holds (T. Tannaka, *Tôhoku Math. J.*, 45 (1939)). Let $R(G)$ be the †algebra over the complex number field $\mathbf{C}$ formed by the set of all the linear combinations of a finite number of $d_{ij}^{\alpha}(x)$, and let $\operatorname{Aut} R(G)$ be the automorphism group of this algebra $R(G)$. Let G^* be the set of all the elements σ in $\operatorname{Aut} R(G)$ which commute with every left translation $L(x)$ $((L(x)f)(y) = f(xy))$ and which satisfy $\sigma(\bar{f}) = \overline{\sigma(f)}$. Then G^* is a topological group with respect to the weak topology, and the Tannaka duality theorem implies that the correspondence which assigns to each $x \in G$ the right translation $R(x)$ (the restriction of $U(x)$ to $R(G)$) is an isomorphism of G onto G^* as topological groups. For the case where G is a compact Lie group, C. Chevalley restated the Tannaka duality theorem as one giving a relation between compact Lie groups and complex algebraic groups (→ 249 Lie Groups U).

References

See references to 18 Almost Periodic Functions and 423 Topological Groups.

For compact Lie groups → references to 249 Lie Groups.

70 (IX.5) Complexes

A. General Remarks

The notion of complexes was introduced by H. Poincaré to study the topology of †manifolds by combinatorial methods (→ 65 Combinatorial Manifolds). Various kinds of complexes have been introduced in the course of the development of topology. These will be dealt with individually in the sections that follow.

B. Euclidean Simplicial Complexes

Let $\mathbf{R}^N$ be the N-dimensional †Euclidean space, and let $a_0, a_1, \ldots, a_m$ be points of $\mathbf{R}^N$ with coordinates $a_i = (a_i^{(1)}, a_i^{(2)}, \ldots, a_i^{(N)})$ for $i = 0, 1, \ldots, m$. For real numbers $\lambda_0, \lambda_1, \ldots, \lambda_m$, we denote by $\lambda_0 a_0 + \lambda_1 a_1 + \ldots + \lambda_m a_m$ the point $(\sum_{i=0}^m \lambda_i a_i^{(1)}, \sum_{i=0}^m \lambda_i a_i^{(2)}, \ldots, \sum_{i=0}^m \lambda_i a_i^{(N)})$ of $\mathbf{R}^N$.

A set of $n+1$ points $a_0, a_1, \ldots, a_n$ of $\mathbf{R}^N$ is said to be **independent** or in **general position** if vectors $\overrightarrow{a_0a_1}, \overrightarrow{a_0a_2}, \ldots, \overrightarrow{a_0a_n}$ are linearly independent. For given independent points $a_0, a_1, \ldots, a_n$, let $|a_0a_1 \ldots a_n|$ denote the subset of $\mathbf{R}^N$ given by

$$|a_0a_1 \ldots a_n| = \{\lambda_0 a_0 + \lambda_1 a_1 + \ldots + \lambda_n a_n \mid \lambda_0 + \lambda_1 + \ldots + \lambda_n = 1, \lambda_i \geqslant 0\}.$$

$|a_0a_1 \ldots a_n|$ is called an n-**simplex** with **vertices** $a_0, a_1, \ldots, a_n$. Simplexes are denoted by Δ, Δ', σ, τ, etc. For example, a 0-simplex $|a_0|$ is a point a_0, a 1-simplex $|a_0a_1|$ is a segment $\overline{a_0a_1}$, and a 2-simplex $|a_0a_1a_2|$ is a triangle $\Delta a_0a_1a_2$.

We attach to $\lambda_0 a_0 + \lambda_1 a_1 + \ldots + \lambda_n a_n$ the coordinates $(\lambda_0, \lambda_1, \ldots, \lambda_n)$, called the **barycentric coordinates**. If $|a_0 a_1 \ldots a_n|$ is an n-simplex, every subset $\{a_{i_0}, a_{i_1}, \ldots, a_{i_q}\}$ of $\{a_0, a_1, \ldots, a_n\}$ is independent. A q-simplex $|a_{i_0} a_{i_1} \ldots a_{i_q}|$ is called a **q-face** of $|a_0 a_1 \ldots a_n|$.

A set $\mathfrak{K}$ of simplexes in a Euclidean space $\mathbf{R}^N$ is called a **Euclidean (simplicial) complex** in $\mathbf{R}^N$ if $\mathfrak{K}$ satisfies the following three conditions: (i) Every face of a simplex belonging to $\mathfrak{K}$ is also an element of $\mathfrak{K}$. (ii) The intersection of two simplexes belonging to $\mathfrak{K}$ is either empty or a face of each of them. (iii) Each point of a simplex belonging to $\mathfrak{K}$ has a neighborhood in $\mathbf{R}^N$ that intersects only a finite number of simplexes belonging to $\mathfrak{K}$. A Euclidean complex is also called a **geometric complex** (or **rectilinear complex**). Each 0-simplex in $\mathfrak{K}$ is called a **vertex** in $\mathfrak{K}$. We define the **dimension of** $\mathfrak{K}$ to be n if $\mathfrak{K}$ contains an n-simplex but no $(n+1)$-simplex, and ∞ if $\mathfrak{K}$ contains n-simplexes for all $n \geqslant 0$.

By a **subcomplex** of a Euclidean complex $\mathfrak{K}$ we mean a Euclidean complex that is a subset of $\mathfrak{K}$. For a Euclidean complex $\mathfrak{K}$, its **r-section** (or **r-skeleton**) is defined to be the subcomplex of $\mathfrak{K}$ consisting of all n-simplexes $(n \leqslant r)$ in $\mathfrak{K}$.

If $\mathfrak{K}$ is a Euclidean complex in $\mathbf{R}^N$, we denote by $|\mathfrak{K}|$ the set of points in $\mathbf{R}^N$ belonging to simplexes in $\mathfrak{K}$. This set $|\mathfrak{K}|$ is called the **Euclidean polyhedron** of $\mathfrak{K}$.

By **subdivision** $\mathfrak{K}'$ of a Euclidean complex $\mathfrak{K}$ we mean a Euclidean complex such that $|\mathfrak{K}'| = |\mathfrak{K}|$ and each simplex in $\mathfrak{K}'$ is contained in a simplex in $\mathfrak{K}$.

Specifically, we can construct a subdivision $\mathfrak{K}'$ of $\mathfrak{K}$ utilizing †barycenters of simplexes in $\mathfrak{K}$; namely, we let $\mathfrak{K}'$ be the set of all r-simplexes whose vertices consist of barycenters of the series $\Delta_0 \subset \Delta_1 \subset \ldots \subset \Delta_r$ of simplexes in $\mathfrak{K}$. Then $\mathfrak{K}'$ is a subdivision of $\mathfrak{K}$, called the **barycentric subdivision** of $\mathfrak{K}$ and denoted by $\mathrm{Sd}\,\mathfrak{K}$.

Given a Euclidean complex $\mathfrak{K}$ and a subset A of $|\mathfrak{K}|$, we define the **star** of A in $\mathfrak{K}$ to be the subcomplex of $\mathfrak{K}$ that consists of simplexes $\{\Delta\}$ and their faces such that $\Delta \cap A \neq \varnothing$. Furthermore, we define the **open star** of A in $\mathfrak{K}$ as the union of †open simplexes (the interiors of simplexes) of $\mathfrak{K}$ whose closures intersect A. We denote by $St_{\mathfrak{K}}(A)$ the star of A in $\mathfrak{K}$ and by $O_{\mathfrak{K}}(A)$ the open star of A in $\mathfrak{K}$; then $O_{\mathfrak{K}}(A)$ is an open set whose closure is $|St_{\mathfrak{K}}(A)|$.

The notion of Euclidean simplicial complexes can be generalized to that of **Euclidean cell complexes**; this is done by replacing the term *simplex* by †*convex cell* in the definition of Euclidean simplicial complex. For a Euclidean cell complex $\mathfrak{K}$, the notions of vertex, dimension, subcomplex, r-section, and subdivision are defined similarly as in the case of Euclidean simplicial complexes.

C. Simplicial Complexes

Given a Euclidean simplicial complex $\mathfrak{K}$, let K denote the set of all the vertices in $\mathfrak{K}$, and let Σ denote the set consisting of those subsets $\{v_0, \ldots, v_r\}$ of K for which there exist simplexes Δ in $\mathfrak{K}$ such that $\{v_0, \ldots, v_r\}$ coincide with the set of vertices of Δ. Then we have (1) if $s \in \Sigma$ and $s \supset s'$, $s' \neq \varnothing$, then $s' \in \Sigma$; (2) every set consisting of a single element in K is in Σ, and the empty set is not in Σ.

In general, if a pair (K, Σ) of a set K and a set Σ consisting of finite subsets of K satisfy (1) and (2), then the pair (or the set K) is called an **abstract simplicial complex** (or simply **simplicial complex**). If K is a simplicial complex, each element of the set K is called a **vertex** in K, and each set of Σ is called a **simplex** in K. A simplex consisting of $n+1$ vertices is called an **n-simplex**. We say that a simplicial complex K is **finite** if it consists of a finite number of vertices; it is **locally finite** if every vertex of K belongs to only finitely many simplexes in K. We define similarly **countable simplicial complexes** and **locally countable simplicial complexes**. The **dimension** and **r-section** of a simplicial complex are defined as in the case of Euclidean complexes. By a **subcomplex** of a simplicial complex K we mean a simplicial complex K_0 such that each simplex of K_0 is a simplex in K.

Let K and L be simplicial complexes. A mapping $\varphi: K \rightarrow L$ is called a **simplicial mapping (simplicial map)** of K to L if the following condition is satisfied: If $v_0, v_1, \ldots, v_n$ are vertices of a simplex of K, then $\varphi(v_0)$, $\varphi(v_1), \ldots,$ $\varphi(v_n)$ are vertices of some simplex of L. Two simplicial complexes K and L are said to be **isomorphic** if there exist simplicial mappings $\varphi: K \rightarrow L$, $\psi: L \rightarrow K$ such that $\psi \circ \varphi$ and $\varphi \circ \psi$ are the identity mappings.

Given a simplicial complex K, let $|K|$ denote the set of all functions x from the set of vertices of K to the closed interval $I = [0, 1]$ satisfying the following conditions: (i) The set $\{v \in K \mid x(v) \neq 0\}$ is a simplex of K. (ii) $\Sigma_{v \in K} x(v) = 1$. The value $x(v)$ is called the **barycentric coordinate** of the point $x \in |K|$ with respect to the vertex v. Each vertex v of K is identified with the point of $|K|$ whose barycentric coordinate with respect to the vertex v is 1 and is called a **vertex** in $|K|$. For a simplex $s = \{v_0, v_1, \ldots, v_n\}$ in K, we define $|s| = \{x \in |K| \mid x(v) = 0$ $(v \notin s)\}$, which is called a **simplex** in $|K|$. We call $\{x \in |s| \mid x(v_i) > 0\ (i = 0, 1, \ldots, n)\}$ an **open simplex** of $|K|$ or the **interior** of $|s|$. We remark here that for an arbitrary simplex s in K a point $x \in |s|$ can be written in the form $x = \Sigma_{v \in s} x(v) \cdot v$. We can define a metric d on $|K|$ by $d(x, y) = (\Sigma_{v \in K}(x(v) - y(v))^2)^{1/2}$. However, $|K|$ is usually supplied with a †stronger topology

defined as follows: (1) Each simplex $|s|$ in $|K|$ has the topology given by the metric d. (2) A subset U of $|K|$ is open if and only if $U \cap |s|$ is an open subset of $|s|$ for each simplex s of K. Henceforward, by the topology of $|K|$ we mean the topology just defined unless otherwise stated. The set $|K|$ with such a topology is called the **polyhedron** of K. The topology of $|K|$ coincides with the above metric topology if and only if K is locally finite. If simplicial complexes K and L are isomorphic, then $|K|$ and $|L|$ are homeomorphic. If K is the simplicial complex defined by a Euclidean simplicial complex $\mathfrak{K}$, then $|K|$ and $|\mathfrak{K}|$ are homeomorphic. When K is finite, there exists an Euclidean simplicial complex $\mathfrak{K}$ whose simplicial complex is isomorphic to K. Accordingly $|K|$ is homeomorphic to the Euclidean polyhedron $|\mathfrak{K}|$.

If K and L are simplicial complexes, a mapping $f:|K| \to |L|$ satisfying the following condition is said to be **linear**: If $s = \{v_0, v_1, \ldots, v_n\}$ is a simplex in K and $x = \lambda_0 v_0 + \ldots + \lambda_n v_n$ ($\lambda_0 + \ldots + \lambda_n = 1, \lambda_i \geqslant 0$), then $f(v_0), \ldots, f(v_n)$ belong to a simplex in L and $f(x) = \lambda_0 f(v_0) + \ldots + \lambda_n f(v_n)$. The linear mapping determined by a simplicial mapping $\varphi: K \to L$ is denoted by $|\varphi|:|K| \to |L|$ and is also called a **simplicial mapping** and denoted by the same letter φ. Let K and K' be simplicial complexes. If there exists a linear mapping $l:|K'| \to |K|$, which is a homeomorphism, then we identify $|K|$ and $|K'|$ by l and call K' a **subdivision** of K. The **barycentric subdivision** $\mathrm{Sd}\, K$ of a simplicial complex K is defined as in the case of Euclidean complexes. We also have notions of **star** $St_K(A)$ and **open star** $O_K(A)$ for a simplicial complex K and a subset A of $|K|$. If K and L are simplicial complexes, a mapping $f:|K| \to |L|$ is called a **piecewise linear mapping** if there exist subdivisions K' and L' of K and L, respectively, such that $f:|K'| \to |L'|$ is linear.

Given an open †covering $\mathfrak{M} = \{M_v\}_{v \in K}$ of a topological space X, the index set K becomes a simplicial complex if we consider each finite nonempty subset s of K such that $\bigcap_{v \in S} M_v \neq \varnothing$ (empty) to be a simplex. The resulting simplicial complex K is called the **nerve** of the open covering $\mathfrak{M}$. Furthermore, if $\mathfrak{M} = \{M_v\}_{v \in K}$ and $\mathfrak{N} = \{N_w\}_{w \in L}$ are open coverings of a set X, $\mathfrak{N}$ is a †refinement of the covering $\mathfrak{M}$, and L is the nerve of $\mathfrak{N}$, then a simplicial mapping $\varphi: L \to K$ is defined by sending each vertex w in L to a vertex v in K such that $N_w \subset M_v$.

Given two disjoint simplicial complexes K and L, a simplicial complex $K * L$, called the **join** of K and L, is defined by the following: (1) the vertices of $K * L$ are the vertices of K and the vertices of L. (2) A nonempty subset of vertices is a simplex of $K * L$ if and only if its subsets in K and L are empty or simplexes there. In particular, the join of a simplicial complex K and a single point is called the **cone** of K.

A simplicial complex K is said to be **ordered** if a †partial ordering is given in the set of vertices in K such that the set of vertices of each simplex is †totally ordered. Given ordered simplicial complexes K and L, an ordered simplicial complex $K \times L$, called the **Cartesian product** of K and L, is defined by the following: (1) The vertices in $K \times L$ are pairs (v, w), where v and w are any vertices in K and L, respectively. (2) A set of vertices $(v_0, w_0), \ldots, (v_n, w_n)$ such that $v_0 \leqslant \ldots \leqslant v_n$ and $w_0 \leqslant \ldots \leqslant w_n$ is a simplex in $K \times L$ if $(v_0, \ldots, v_n)$ and $(w_0, \ldots, w_n)$ are simplexes in K and L, respectively; all simplexes in $K \times L$ are obtained in this manner. (3) $(v_1, w_1) \leqslant (v_2, w_2)$ if and only if $v_1 \leqslant v_2$ and $w_1 \leqslant w_2$.

Assume that either K or L is locally finite or that both K and L are locally countable. Let X and Y be topological spaces and I be the closed interval $[0, 1]$. We define an equivalence realtion $\sim$ in the topological space $X \cup (X \times Y \times I) \cup Y$ by $x \sim (x, y, 0)$ and $y \sim (x, y, 1)$, where $x \in X$, $y \in Y$. The quotient space of $X \cup (X \times Y \times I) \cup Y$ by this relation is called the **join** of X and Y and is denoted by $X * Y$. Then the polyhedron $|K \times L|$ is homeomorphic to the product space $|K| \times |L|$ of the topological spaces $|K|$ and $|L|$. And the polyhedron $|K * L|$ is homeomorphic to $|K| * |L|$.

By a **triangulation** T of a topological space X we mean a pair (K, t) consisting of a simplicial complex K and a homeomorphism $t:|K| \to X$. A triangulation is also called a **simplicial decomposition**. If $T = (K, t)$ is a triangulation of X, the various concepts defined for K can be transferred to X by means of the mapping t. For example, by a **simplex** of the triangulation T we mean the image of a simplex of $|K|$ under t. We say that a triangulation $T = (K, t)$ is **finite** if K is a finite simplicial complex. If $T = (K, t)$ is a triangulation and (K', l) is a subdivision of K, then $T' = (K', t \circ l)$ is called a **subdivision** of T. If $T_1 = (K_1, t_1)$, $T_2 = (K_2, t_2)$ are triangulations of topological spaces X_1, X_2, respectively, a mapping $f: X_1 \to X_2$ is called a **simplicial mapping** relative to T_1 and T_2 if $t_2^{-1} \circ f \circ t_1 : |K_1| \to |K_2|$ is a simplicial mapping. The following two problems on triangulations are famous: (1) Under what topological conditions is it possible for a given topological space to be supplied with a triangulation? (2) Given two triangulations T_1, T_2 of a space X, are there subdivisions $T_1' = (K_1', t_1')$, $T_2' = (K_2', t_2')$ of T_1 and T_2, respectively, such that K_1' and K_2' are isomorphic? Concerning the second problem, the conjecture asserting the existence of subdivisions T_1'

and T_2' as above is known as the **fundamental conjecture (Hauptvermutung)** in topology. Every 3-dimensional manifold is triangulable, and any two of its triangulations admit subdivisions satisfying the condition in (2) (E. E. Moise [7]).

In 1961, J. Milnor showed that the fundamental conjecture is not true for polyhedra [8]. The triangulation problem and the fundamental conjecture for topological manifolds were negatively solved by R. Kirby and L. Siebenmann [9] (→ 65 Combinatorial Manifolds). Any †differentiable manifold is triangulable, and the fundamental conjecture holds for its †C^r-triangulations ($r \geqslant 1$) [6] (→ 114 Differential Topology).

Let T_1, T_2 be triangulations of topological spaces X_1, X_2, respectively, and let $f: X_1 \rightarrow X_2$ be a continuous mapping. Then a simplicial mapping $\varphi: X_1 \rightarrow X_2$ relative to T_1 and T_2 is called a **simplicial approximation** to f if, for each $x \in X$, the image $\varphi(x)$ lies on the simplex of T_2 whose interior contains $f(x)$. The following existence theorem is called the **simplicial approximation theorem**: For every continuous mapping $f: X_1 \rightarrow X_2$, there exist a subdivision T_1' of T_1 and a simplicial mapping $\varphi: X_1 \rightarrow X_2$ relative to T_1' and T_2 that is a simplicial approximation of f. If the triangulation T_1 is finite, then for a sufficiently large n we can choose $\mathrm{Sd}^n T_1$ as the T_1' above (where $\mathrm{Sd}^0 T = T$ and $\mathrm{Sd}^n T = \mathrm{Sd}(\mathrm{Sd}^{n-1} T)$ $(n \geqslant 1)$). If $\varphi: X_1 \rightarrow X_2$ is a simplicial approximation to a continuous mapping f, then f and φ are †homotopic.

D. Cell Complexes

Let V^n be the †unit n-disk, S^{n-1} be the †unit $(n-1)$-sphere, and X be a Hausdorff space. For a subset e of X, let $\bar{e}$ be the closure of e in X, and let $\dot{e} = \bar{e} - e$. A subset e of the space X is called an ***n*-cell**, or **open *n*-cell** in X if there is a relative homeomorphism $\varphi: (V^n, S^{n-1}) \rightarrow (\bar{e}, \dot{e})$, i.e., a continuous mapping $\varphi: V^n \rightarrow \bar{e}$ such that $\varphi(S^{n-1}) \subset \dot{e}$ and $\varphi: V^n - S^{n-1} \rightarrow \bar{e} - \dot{e}$ is a homeomorphism. For example, $S^n - \{p\}$ $(p \in S^n)$ is an n-cell. A set $\{e_\lambda \mid \lambda \in \Lambda\}$ of cells in the Hausdorff space X is called a **cellular decomposition** of X if the following three conditions are satisfied: (i) $e_\lambda \cap e_\mu$ is empty if $\lambda \neq \mu$; (ii) $X = \bigcup_{\lambda \in \Lambda} e_\lambda$; (iii) If the dimension of e_λ is $n+1$, then $\dot{e}_\lambda \subset X^n$, where X^n is the union of all the cells e_μ $(\mu \in \Lambda)$ whose dimensions are not greater than n. For example, the n-sphere S^n has a cellular decomposition consisting of a single 0-cell and a single n-cell.

A Hausdorff space X together with its cellular decomposition $\{e_\lambda\}$ is called a **cell complex**, and each e_λ is called a **cell** in the cell complex X. For a cell complex, the notions of **vertex**, ***n*-section**, and **dimension** are defined in the same way as the corresponding notions in Euclidean complexes. Let X be a cell complex and A a topological subspace of X such that the closure of each cell of X intersecting A is contained in A. Then the set of cells e such that $e \cap A \neq \varnothing$ forms a cellular decomposition of A. The set A together with this cellular decomposition is called a **subcomplex** of the cell complex X. A cell complex X with its cells $\{e_\lambda\}$ is said to be **finite** if the number of e_λ is finite. If each point in a cell complex X is an interior point of some finite subcomplex of X, then X is said to be **locally finite**. We define similarly a **countable cell complex** and a **locally countable cell complex**. If the closure of each n-cell of a cell complex X is homeomorphic to V^n $(n = 0, 1, \ldots)$, X is said to be **regular**. If X and Y are cell complexes, a continuous mapping $f: X \rightarrow Y$ such that $f(X^n) \subset Y^n$ $(n = 0, 1, \ldots)$ is called a **cellular mapping**. If X and Y are cell complexes, the set of †topological products $e_1 \times e_2$, where e_1, e_2 run over all cells of X, Y, respectively, is a cellular decomposition of the product space $X \times Y$. The resulting cell complex $X \times Y$ is called the **product complex** of the cell complexes X and Y.

A cell complex X is said to be **closure finite** if each cell in X is contained in a finite subcomplex of X; and X is said to have the **weak topology** if a subset $U \subset X$ is open if and only if $U \cap \bar{e}$ is relatively open in $\bar{e}$ for each cell e of X. We call a cell complex a **CW complex** if it is closure finite and has the weak topology. The cellular decomposition of a CW complex is called a **CW decomposition**. A locally finite cell complex is a CW complex.

Fundamental properties of CW complexes are as follows: (i) A CW complex is a †paracompact (hence †normal) space and is †locally contractible. (ii) A subcomplex A of a CW complex X is a closed subspace of X, and A itself is a CW complex. (iii) A mapping $f: X \rightarrow Y$ of a CW complex X to a topological space Y is continuous if and only if the restriction $f \mid \bar{e}$ is continuous for each cell e of X. (iv) If X and Y are CW complexes and $f: X \rightarrow Y$ is any continuous mapping then there exists a cellular mapping of X to Y that is homotopic to f (**cellular approximation theorem**). (v) A pair (X, A) consisting of a CW complex X and its subcomplex A has the †homotopy extension property for any topological space. (vi) A CW complex has the †covering homotopy property for any †fiber space. (vii) The product complex $X \times Y$ of two CW complexes X and Y is not necessarily a CW complex, but it is †homotopy equivalent to a CW complex. (viii) If either X or Y is locally finite, or if both X and Y are locally countable, then the product

complex $X \times Y$ is a CW complex. (ix) For CW complexes X and Y, the †mapping space Y^X is homotopy equivalent to a CW complex. (x) The †covering space of a CW complex has a CW decomposition.

If K is a simplicial complex, the polyhedron $|K|$ is a regular CW complex whose cells are all open simplexes in $|K|$. A simplicial complex K is (locally) finite if and only if the CW complex $|K|$ is (locally) finite. In particular, the Euclidean polyhedron of a Euclidean simplicial (or cell) complex is a locally finite CW complex. A polyhedron $|K|$ generally admits a CW decomposition whose cells are far smaller in number than the simplexes constituting a simplicial decomposition of K. For any CW complex X, there exists a polyhedron $|K|$ that is homotopy equivalent to X. In particular, if X is an n-dimensional finite (countable) CW complex, we can choose as K an n-dimensional finite (locally finite and countable) simplicial complex.

E. Semisimplicial Complexes

By an **ordered simplex** in a simplicial complex K we mean a finite sequence $(v_0, v_1, \ldots, v_n)$ $(n \geqslant 0)$ of vertices in K, contained in the set of vertices of a simplex in K. Let $O(K)_n$ be the set of all ordered simplexes of K of length $n+1$, and define mappings $\partial_i: O(K)_n \rightarrow O(K)_{n-1}$ and $s_i: O(K)_n \rightarrow O(K)_{n+1}$ for $i=0, 1, \ldots, n$ by $\partial_i(v_0, \ldots, v_n) = (v_0, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n)$ and $s_i(v_0, \ldots, v_n) = (v_0, \ldots, v_{i-1}, v_i, v_i, v_{i+1}, \ldots, v_n)$. Then the following relations hold:

$$\begin{aligned} &\partial_i \circ \partial_j = \partial_{j-1} \circ \partial_i && (i<j), \\ &s_i \circ s_j = s_{j+1} \circ s_i && (i \leqslant j), \\ &\partial_i \circ s_j = s_{j-1} \circ \partial_i && (i<j), \\ &\partial_i \circ s_j = s_j \circ \partial_{i-1} && (i>j+1), \\ &\partial_i \circ s_i = \partial_{i+1} \circ s_i = \text{identity}. && \qquad (1) \end{aligned}$$

Let Δ^n be the n-dimensional simplex in $\mathbf{R}^n$ with vertices $e_0=(0,0,\ldots,0)$, $e_1=(1,0,\ldots,0)$, $\ldots, e_n=(0,\ldots,1)$. By a **singular n-simplex** in a topological space X we mean a continuous mapping $T: \Delta^n \rightarrow X$. Let $S(X)_n$ be the set of all singular n-simplexes in X, and define mappings $\partial_i: S(X)_n \rightarrow S(X)_{n-1}$ and $s_i: S(X)_n \rightarrow S(X)_{n+1}$ for $i=0, 1, \ldots, n$ by $\partial_i T(\lambda_0, \ldots, \lambda_n) = T(\lambda_0, \ldots, \lambda_{i-1}, 0, \lambda_i, \ldots, \lambda_n)$ and $s_i T(\lambda_0, \ldots, \lambda_{n+1}) = T(\lambda_0, \ldots, \lambda_{i-1}, \lambda_i + \lambda_{i+1}, \lambda_{i+1}, \ldots, \lambda_{n+1})$, where $(\lambda_0, \ldots, \lambda_n)$ is the point $\sum_{i=0}^n \lambda_i e_i$, $\lambda_i \geqslant 0$, $\sum_{i=0}^n \lambda_i = 1$. Then relation (1) holds between ∂_i and s_i.

Because of the importance of relation (1), which is basic in defining †homology of simplicial complexes and topological spaces (→ 201 Homology Theory), S. Eilenberg and J. A. Zilber gave the following definition: A **semisimplicial complex** K consists of a sequence of sets K_n $(n=0, 1, \ldots)$ together with mappings $\partial_i: K_n \rightarrow K_{n-1}$, $s_i: K_n \rightarrow K_{n+1}$ $(i=0, 1, \ldots, n)$ satisfying relations (1). An element of K_n is called an **n-simplex** in K, and ∂_i, s_i are called the ith **face operator** and the ith **degeneracy operator**, respectively. A simplex is said to be **degenerate** if it is the image of a simplex under some s_i. A semisimplicial complex is abbreviated as **s.s. complex**. The s.s. complexes $O(K) = \{O(K)_n, \partial_i, s_i\}$ and $S(X) = \{S(X)_n, \partial_i, s_i\}$ are called the **ordered complex** of K and the **singular complex** of X, respectively.

Let K be an s.s. complex, and let L_n be a subset of K_n for $n=0, 1, \ldots$. If $\partial_i(L_n) \subset L_{n-1}$ and $s_i(L_n) \subset L_{n+1}$ for each i, then $L = \{L_n, \partial_i | L_n, s_i | L_n\}$ is an s.s. complex, and L is called a **subcomplex** of the s.s. complex K. If A is a subspace of a topological space X, $S(A)$ is a subcomplex of $S(X)$. If K is an ordered simplicial complex, a subcomplex $O'(K)$ of $O(K)$ is obtained by considering the set of all ordered simplexes $(v_0, v_1, \ldots, v_n)$ such that $v_0 \leqslant v_1 \leqslant \ldots \leqslant v_n$ $(n=0, 1, \ldots)$. If K and L are s.s. complexes, a sequence $f = \{f_n\}$ of mappings $f_n: K_n \rightarrow L_n$ defined for each n is called an **s.s. mapping** if $\partial_i \circ f_n = f_{n-1} \circ \partial_i$ and $s_i \circ f_n = f_{n+1} \circ s_i$ $(0 \leqslant i \leqslant n)$. If $f: X \rightarrow Y$ is a continuous mapping of topological spaces, then f determines an s.s. mapping $S(f): S(X) \rightarrow S(Y)$ by $S(f)(T) = f \circ T$. Two s.s. complexes K and L are said to be **isomorphic** if there is a bijective s.s. mapping of K to L. For two s.s. complexes K and L, we define the **Cartesian product** $K \times L$ to be the s.s. complex given by $(K \times L)_n = K_n \times L_n$, $\partial_i(\sigma, \tau) = (\partial_i \sigma, \partial_i \tau)$, $s_i(\sigma, \tau) = (s_i \sigma, s_i \tau)$ $(\sigma \in K_n, \tau \in L_n)$. If K and L are ordered simplicial complexes, the s.s. complexes $O'(K) \times O'(L)$ and $O'(K \times L)$ are isomorphic. If X and Y are topological spaces, the s.s. complexes $S(X) \times S(Y)$ and $S(X \times Y)$ are isomorphic.

Given an s.s. complex K, we construct a topological space $|K|$ as follows: First, we provide K_n with the †discrete topology and consider the topological space $\bar{K} = \bigcup_{n \geqslant 0} K_n \times \Delta^n$. Next we consider simplicial mappings $\varepsilon^i: \Delta^{n-1} \rightarrow \Delta^n$ and $\eta^i: \Delta^{n+1} \rightarrow \Delta^n$ defined by $\varepsilon^i(p_j) = p_j$ $(j<i)$, $\varepsilon^i(p_j) = p_{j+1}$ $(j \geqslant i)$ and $\eta^i(p_j) = p_j$ $(j \leqslant i)$, $\eta^i(p_j) = p_{j-1}$ $(j>i)$, where $p_0, \ldots, p_n$ are the vertices of Δ^n. The topological space $|K|$ is defined to be the †quotient space of $\bar{K}$ with respect to an equivalence relation $\sim$ that is defined by the following: $(\partial_i \sigma, y) \sim (\sigma, \varepsilon^i(y))$ $(\sigma \in K_n, y \in \Delta^{n-1})$, $(s_i \sigma, y) \sim (\sigma, \eta^i(y))$ $(\sigma \in K_n, y \in \Delta^{n+1})$, where $i=0, 1, \ldots, n$. The space $|K|$ is called the **(geometric) realization of the s.s. complex** K. Given an s.s. mapping $f: K \rightarrow L$, we obtain a continuous mapping $|f|: |K| \rightarrow |L|$ defined by $|f|(|\sigma, y|) = |f(\sigma), y|$, where $|\sigma, y|$ is the point in $|K|$ represented by $(\sigma, y) \in \bar{K}$. We call $|f|$ the **realization of the s.s. mapping** f.

The realization $|K|$ of an s.s. complex K is a CW complex whose cells are in one-to-one correspondence with the nondegenerate simplexes in K. For a topological space X, a †weak homotopy equivalence $\rho:|S(X)|\rightarrow X$ is defined by $\rho(|T,y|)=T(y)$ $(T\in S(X)_n, y\in\Delta^n)$. This mapping gives rise to a homotopy equivalence when X is a CW complex.

The singular complex $S(X)$ of a topological space X has the following property: Given simplexes $\sigma_0,\ldots,\sigma_{k-1},\sigma_{k+1},\ldots,\sigma_{n+1}\in K_n$ with $\partial_i\sigma_j=\partial_{j-1}\sigma_i$ $(i<j, i,j\neq k)$, there exists a simplex $\sigma\in K_{n+1}$ with $\partial_i\sigma=\sigma_i$ $(i\neq k)$. An s.s. complex K with this property is called a **Kan complex**. If, in addition, $\partial_i\sigma=\partial_i\sigma'$ $(\sigma,\sigma'\in K_n, i\neq k)$ imply $\partial_k\sigma=\partial_k\sigma'$, we call K a **minimal complex**. For every Kan complex K, there are minimal subcomplexes M of K that are isomorphic to each other. Moreover, $|M|$ is a †deformation retract of $|K|$. For a Kan complex K, the †homotopy group can be defined combinatorially.

Two CW complexes X and Y are homotopy equivalent if and only if the minimal subcomplexes of $S(X)$ and $S(Y)$ are isomorphic.

F. Eilenberg-MacLane Complex

Given an integer $n\geqslant 1$ and a group π (Abelian if $n\geqslant 2$), there exists an †arcwise connected topological space X for which the †homotopy groups $\pi_i(X)$ are trivial for $i\neq n$ and $\pi_n(X)\cong\pi$. Such a space is called an **Eilenberg-MacLane space** of type (π,n). Let $\Omega(X;X,*)$ be the †path space over X, and let $p_0:\Omega(X;X,*)\rightarrow X$ be the natural projection. If X is an Eilenberg-MacLane space of type (π,n), then $(\Omega(X;X,*), p_0, X)$ gives rise to a standard †contractible †fiber space whose fiber is an Eilenberg-MacLane space of type $(\pi,n-1)$. Assume that X is an Eilenberg MacLane space of type (π,n) with Abelian group π. Then X is $(n-1)$-connected; hence the †Hurewicz theorem can be utilized to show the existence of an isomorphism $h:\pi_n(X)\cong H_n(X)$, while the universal coefficient theorem can be utilized to show that $H^n(X;\pi)\cong\mathrm{Hom}(H_n(X),\pi)$. Since in this case we have $\pi_n(X)\cong\pi$, the element $h^{-1}\in\mathrm{Hom}(H_n(X),\pi_n(X))$ can be regarded as an element of $\mathrm{Hom}(H_n(X),\pi)$. Now the **fundamental class** of X is defined to be the cohomology class $u\in H^n(X;\pi)$ corresponding to h^{-1}. Let Y be a †CW complex and $\pi(Y;X)$ be the set of †homotopy classes of continuous mappings from Y to X. Then there exists a one-to-one correspondence $\pi(Y;X)\rightarrow H^n(Y;\pi)$ given by the assignment $[f]\rightarrow f^*u$ (→ 305 Obstructions). Let $S(X)$ be the singular complex of X, and let $M(X)$ be a minimal complex of $S(X)$. If X is an Eilenberg-MacLane space of type (π,n), $M(X)$ is isomorphic to a certain complex determined uniquely by π and n. This complex is called the **Eilenberg-MacLane complex** of type (π,n) and is denoted by $K(\pi,n)$. The notation $K(\pi,n)$ is also used to mean the space X itself.

Let $\Delta(q)$ be a simplicial complex whose simplexes are all nonvacuous subsets of $\{0,1,\ldots,q\}$. Let $\varepsilon_i:\Delta(q-1)\rightarrow\Delta(q)$ be the simplicial mapping defined by $\varepsilon_i(j)=j$ $(0\leqslant j\leqslant i-1)$, $\varepsilon_i(j)=j+1$ $(i\leqslant j\leqslant q-1)$, and let $\eta_i:\Delta(q+1)\rightarrow\Delta(q)$ be the mapping defined by $\eta_i(j)=j$ $(0\leqslant j\leqslant i)$, $\eta_i(j)=j-1$ $(i+1\leqslant j\leqslant q+1)$. Now $K(\pi,n)$ is a Kan complex defined by $K(\pi,n)_q=Z^n(\Delta(q);\pi)$, $\partial_i\sigma=\sigma\circ\varepsilon_i$, $s_i\sigma=\sigma\circ\eta_i$, where $Z^0(\Delta(q);\pi)=\pi$, $Z^1(\Delta(q);\pi)$ is the set of π-valued functions defined on the set of pairs (i,j) such that $0\leqslant i<j\leqslant q$ and satisfying the equality $\sigma(j,k)\cdot\sigma(i,k)^{-1}\cdot\sigma(i,j)=1$ for $0\leqslant i<j<k\leqslant q$, and $Z^n(\Delta(q);\pi)$ $(n\geqslant 2)$ is the group of †oriented cocycles of the simplicial complex $\Delta(q)$. If π is Abelian, the structure of the Abelian group $Z^n(\Delta(q);\pi)$ gives $K(\pi,n)$ the structure of an Abelian group in the s.s. category. This structure yields a one-to-one correspondence $K(\pi,n)_q\rightarrow K(\pi,n-1)_{q-1}\times\ldots\times K(\pi,n-1)_0$ for $n\geqslant 1$ and leads to the expression of $\tau\in K(\pi,n)_q$ in the form $\langle\sigma_{q-1},\ldots,\sigma_0\rangle$ with $\sigma_i\in K(\pi,n-1)_i$. The ***W*-construction** of $K(\pi,n-1)$ for $n\geqslant 1$ is a Kan complex $W(\pi,n-1)$ defined by $W(\pi,n-1)_q=K(\pi,n-1)_q\times K(\pi,n)_q$ and $\partial_0(\sigma_q\times\tau_q)=(\partial_0\sigma_q)\cdot\sigma_{q-1}\times\partial_0\tau_q$, $\partial_i(\sigma_q\times\tau_q)=\partial_i\sigma_q\times\partial_i\tau_q$ for $1\leqslant i\leqslant q$, $s_i(\sigma_q\times\tau_q)=s_i\sigma_q\times s_i\tau_q$, where $\sigma_q\in K(\pi,n-1)_q$ and $\tau_q=\langle\sigma_{q-1},\ldots,\sigma_0\rangle\in K(\pi,n)_q$. Let $p: W(\pi,n-1)\rightarrow K(\pi,n)$ be a natural projection. Then $(W(\pi,n-1),p,K(\pi,n))$ plays the role of the †universal bundle for $K(\pi,n-1)$ in the s.s. category in the following sense: Let L be an s.s. complex, and let $f:L\rightarrow K(\pi,n)$ be an s.s. mapping. We define $f^{\#}W(\pi,n-1)$ to be the subcomplex of $W(\pi,n-1)\times L$ generated by simplexes $(\sigma_q\times\tau_q)\times\rho_q$ such that $\tau_q=f(\rho_q)$, where $\sigma_q\times\tau_q\in W(\pi,n-1)_q$ and $\rho_q\in L_q$. Let $p:f^{\#}W(\pi,n-1)\rightarrow L$ be the natural projection. Then $(f^{\#}W(\pi,n-1),p,L)$ is called the principal fiber bundle induced from $W(\pi,n-1)$ by f. Any principal bundle over L with group $K(\pi,n-1)$ can be expressed as an induced bundle. This property means that $(W(\pi,n-1), p,K(\pi,n))$ is universal. On the other hand, we have an algebraic analog of the universal bundle for the chain group of $K(\pi,n-1)$, called the **bar construction** [12]. Both these concepts were defined by Eilenberg and MacLane in order to determine the structure of the (co)homology of $K(\pi,n)$, which is denoted by $(H^*(\pi,n))H_*(\pi,n)$. This object was later achieved by H. Cartan, who introduced an improved notion called Cartan construction (→ Appendix A, Table 6.III). Let $\pi(L,K(\pi,n))$ be the set of s.s. homotopy classes of s.s. map-

pings from L to $K(\pi, n)$. If π is Abelian, there exists a one-to-one correspondence $\pi(L, K(\pi, n)) \to H^n(L; \pi)$ given by the assignment $[f] \to f^*u$, where $u \in H^n(\pi, n; \pi)$ is the fundamental class of $K(\pi, n)$. By virtue of this correspondence, $k = f^*u \in H^n(L; \pi)$ determines an induced bundle $f^\# W(\pi, n-1)$ uniquely up to equivalence, which is denoted by $K(\pi, n-1) \times_k L$.

Let X be an arcwise connected topological space. Let X^n be the Cartesian product of n copies of X. Clearly, the symmetric group $\mathfrak{S}_n$ of degree n operates on X^n. The n-fold **symmetric product** $SP^n X$ of X is defined to be the quotient space of X^n under the action of $\mathfrak{S}_n$. If we specify a reference point of X, we have a natural inclusion $SP^{n-1} X \subset SP^n X$ and can consider the inductive limit space $\bigcup_{1 \leqslant n} SP^n X$, denoted by $SP^\infty X$. Then the Dold-Thom theorem [15] shows that $M(SP^\infty X) \cong \prod_{i=1}^\infty K(H_i(X), i)$. In particular, we have $M(SP^\infty S^n) \cong K(\mathbf{Z}, n)$ for $n \geqslant 1$. This result can be applied to obtain a direct relationship between the axiomatic definition of †cohomology operations using $K(\pi, n)$ due to Eilenberg and Serre and the constructive definition using the symmetric groups due to Steenrod (A. Dold [16], T. Nakamura [17]). For a detailed study of the (co)homology of $SP^n X$ → [18].

G. Postnikov Complex

Let X be an arcwise connected topological space. For the sake of simplicity, throughout this section we assume that X is †simple (→ 202 Homotopy Theory). Then the †Postnikov system of X can be defined as an inverse system (X_n, p_n) $(n = 0, 1, 2, \dots)$ consisting of topological spaces X_n, continuous mappings p_n: $X_n \to X_{n-1}$, and a system (X, q_n) $(n = 0, 1, 2, \dots)$ consisting of continuous mappings $q_n : X \to X_n$ such that $p_n \circ q_n = q_{n-1}$ and satisfying the following three properties: (1) X_0 is one point. (2) (X_n, p_n, X_{n-1}) is a †fiber space induced from a standard contractible fiber space over an Eilenberg-MacLane space of type $(\pi_n(X), n+1)$ by a mapping corresponding to a cohomology class $k^{n+1} \in H^{n+1}(X_{n-1}; \pi_n(X))$. (3) $q_{n*} : \pi_i(X) \to \pi_i(X_n)$ gives an isomorphism for $0 \leqslant i \leqslant n$. These cohomology classes k^{n+1} are called **Eilenberg-Postnikov invariants** (or simply **k-invariants**). Corresponding to the above facts, the minimal complex $M(X)$ can be obtained as the inverse limit of a certain inverse system $(K(n), p(n))$ consisting of Kan complexes $K(n)$ and s.s. mappings $p(n): K(n) \to K(n-1)$ defined by $K(0) = K(0, 0)$ and $K(n) = K(\pi_n, n) \times_{k^{n+1}} K(n-1)$ for $n \geqslant 1$ with natural projections $p(n)$, where $\pi_n = \pi_n(X)$ and $k^{n+1} \in H^{n+1}(K(n-1); \pi_n)$. This system is determined uniquely up to s.s. homotopy equivalence by its limit, called the **Postnikov complex** and denoted by $K(\pi_1, k^3, \pi_2, \dots, k^{n+1}, \pi_n, \dots)$. As yet we are ignorant of an effective method of computing the cohomology of a Postnikov complex from π_n and k^{n+1}.

References

[1] H. Cartan, Sur la théorie de Kan, Sém. H. Cartan, Ecole Norm. Sup., 1956–1957.
[2] S. Eilenberg and N. E. Steenrod, Foundations of algebraic topology, Princeton Univ. Press, 1952.
[3] P. J. Hilton and S. Wylie, Homology theory, Cambridge Univ. Press, 1960.
[4] E. H. Spanier, Algebraic topology, McGraw-Hill, 1966.
[5] R. E. Mosher and M. C. Tangora, Cohomology operations and applications in homotopy theory, Harper & Row, 1968.
[6] J. R. Munkres, Elementary differential topology, Princeton Univ. Press, second edition, 1966.
[7] E. E. Moise, Geometric topology in dimensions 2 and 3, Springer, 1977.
[8] J. Milnor, Two complexes that are homeomorphic but combinatorially distinct, Ann. Math., (2) 74 (1961), 575–590.
[9] R. C. Kirby and L. C. Siebenmann, Foundational essays on topological manifolds, smoothings and triangulations, Princeton Univ. Press, 1977.
[10] J. P. May, Simplicial objects in algebraic topology, Van Nostrand, 1967.
[11] E. Curtis, Simplicial homotopy theory, Lecture notes, Aarhus Univ., 1967.
[12] S. Eilenberg and S. MacLane, On the groups $H(\pi, n)$ I, II, III, Ann. Math., (2) 58 (1953), 55–106; 60 (1954), 49–139; 60 (1954), 513–557.
[13] J.-P. Serre, Cohomologie modulo 2 des complexes d'Eilenberg-MacLane, Comment. Math. Helv., 27 (1953), 198–232.
[14] Sém. H. Cartan, Ecole Norm. Sup., 1954–1955.
[15] A. Dold and R. Thom, Quasifaserungen und unendliche symmetrische Produkte, Ann. Math., (2) 67 (1958), 239–281.
[16] A. Dold, Über die Steenrodschen Kohomologieoperationen, Ann. Math., (2) 73 (1961), 258–294.
[17] T. Nakamura, On cohomology operations, Japan. J. Math., 33 (1963), 93–145.
[18] M. Nakaoka, Decomposition theorem for homology groups of symmetric groups, Ann. Math., (2) 71 (1960), 16–42.

[19] G. W. Whitehead, Elements of homotopy theory, Springer, 1978.

71 (XVI.9) Complexity of Computations

A. Measures for Complexity of Computation

Intuitively, **complexity of computation** means the amount of computing efforts measured on some suitable scale. When a problem can be solved by means of any one of several algorithms, it is highly desirable to compare the complexities of those algorithms. For example, the power x^{10} can be evaluated with a pocket calculator by using either of the algorithms

$$x^{10}=x\times x\times x\times x\times x\times x\times x\times x\times x\times x, \qquad \text{(A)}$$

$$u=x\times x, \quad v=u\times u, \quad w=v\times v,$$

$$x^{10}=u\times w. \qquad \text{(B)}$$

The complexities of these algorithms can be measured by various quantities: ($T1$) time in seconds, ($T2$) the number of times keys and buttons are touched, ($T3$) the number of basic operations (here, multiplications), and (S) the number of values to be stored in the calculator. The quantities ($T1$), ($T2$), and ($T3$) are called **time complexities**, and (S) is called the **space complexity**. Obviously, algorithm (B) is preferable with respect to time complexity, although (A) is better with respect to space complexity. (If the calculator has only one memory register, then only (A) can be executed without recording numbers by hand.) For many calculators, the following algorithm is best with respect to both time and space complexities:

$$x^{10}=((x^2)^2\times x)^2. \qquad \text{(C)}$$

In the general theory of complexity, the number of basic operations ($T3$) is often taken as the basis of time complexity, since it represents the intrinsic complexity of the algorithm, rather than of extrinsic factors such as human skill or mechanical performance.

For the evaluation of a power x^n in general, the complexity depends on the value of n: such a parameter dominating the complexity is called the **size** of the problem. Let $T_X(n)$ be the number of multiplications required for evaluating x^n by an algorithm X. Let $\mathscr{F}$ be the set of all algorithms for evaluating x^n. The time complexity $T(n)$ for evaluation of the power is then defined by

$$T(n)=\mathrm{Min}_{X\in\mathscr{F}}\, T_X(n).$$

This is satisfied when [6]

$$\log_2 n\leqslant T(n)\leqslant 2\log_2 n$$

and

$$\lim_{n\to\infty}\frac{T(n)}{\log_2 n}=1.$$

In a problem concerning a †graph, the number m of nodes in the graph, the number n of its edges, or their sum $m+n$ are chosen as the size of the graph. However, the complexity is not determined by such a size, because there are many different instances (particular graphs) of the same size. So the complexity of this problem is defined in the following manner. Let $T(I)$ be the complexity of solving an instance I of the problem and $L(n)$ be the set of all possible instances of size n. Then

$$W(n)=\mathrm{Max}_{I\in L(n)}\, T(I)$$

and

$$A(n)=\sum p(I)T(I),$$

where $p(I)$ denotes the relative frequency of the instance I. The value $W(n)$ is called the **worst-case complexity**, and $A(n)$ is called the **average complexity** of the problem.

In some cases the complexity of a circuit is measured by the number of its building blocks [1], and the complexity of a program is measured by its length [2, 3].

B. Complexity of a Decision Problem

A decision problem and its complexity are rigorously formulated in terms of †Turing machines. Let Σ be a finite set of symbols and Σ^+ be the whole set of nonempty strings of symbols in Σ. Every instance of the problem is assumed to be represented by a string in Σ^+. A **decision problem** is a triple (Σ, L, P) of the set Σ of symbols, a subset L of Σ^+, and a mapping P from L to $\{0$ (false), 1 (true)$\}$. The set L represents the set of all possible instances of the problem. The problem is said to be **solvable** if and only if there exists a Turing machine M satisfying the following condition for every string α in the set L: When a tape containing the string α is given to the machine M whose head is initially put on the leftmost symbol of α, the machine writes the value of $P(\alpha)$ on the tape and stops after a finite number of steps. Such a Turing machine is said to **compute** the function P. The number $T_M(\alpha)$ of steps is the time complexity of solving the instance α by the machine M, and the length $S_M(\alpha)$ of the used area of the tape is the space complexity. The worst-case complexity $T_M(n)$ is defined by

$$T_M(n)=\mathrm{Max}_{\alpha\in L,|\alpha|=n}\, T_M(\alpha).$$

The space complexity $S_M(n)$ for the size n is defined in a similar way. The complexities $T(n)$ and $S(n)$ of the problem are not defined, because improvement by a constant factor is always possible for any Turing machine [5].

A problem is said to be solved **in linear** or **polynomial time** if its time complexity $T_M(n)$ is bounded by a linear or polynomial function of the size n.

C. Objectives of Complexity Theory

The objectives of complexity theory are as follows:
(i) Analyze an algorithm X for a problem P and evaluate its complexity. When exact evaluation is hard, the order of magnitude $O(T_X(n))$ is investigated.
(ii) Construct better algorithms for the problem P. This gives a better upper bound to the complexity $T(n)$ of the problem.
(iii) Make clear the limitation of improving algorithms for the problem P. This is done by establishing a lower bound to the complexity $T(n)$ of the problem.

A common approach to constructing better algorithms is to partition the problem into smaller parts, apply algorithms to the parts, and then combine the solutions for the parts into a solution for the whole (divide and conquer, [5]). This approach often yields an efficient algorithm for the problem, especially when the partitioning can be repeated recursively. For demonstrating the limitations of algorithms of a certain class, diagonalization [5], determination of information-theoretic lower bounds [6], or the oracle method [6] are often utilized.

D. Elementary Results on Time Complexity

(1) **Number of arithmetic operations**. A polynomial of degree n with one variable can be evaluated in about $3n/2$ arithmetic operations if preconditioning on coefficients is allowed. Otherwise, $2n-1$ operations are necessary and sufficient for evaluating a polynomial of degree n. For computing the product of two square matrices of degree n, $O(n^2)$ operations are necessary and $O(n^{2.5})$ operations are sufficient. The inverse matrix and the value of the determinant of a square matrix of degree n are computed with the same order of operations as the product. Discrete Fourier transformation of n points is executed in $O(n\log n)$ arithmetic operations (fast Fourier transformation).

(2) **Number of comparisons and data transfers**. Rearrangement of n items in increasing order is realized in $O(n\log n)$ comparisons and data transfers. Selection of the kth largest item among n items is done in $O(n)$ operations. The position of an item in a list containing n items is found in $O(n)$ comparisons by linear search, in $O(\log n)$ comparisons by binary chopping, and in $O(1)$ comparisons on average by the †hashing method.

Many other results can be found in [5]–[7].

E. NP-Completeness

A problem is easily solved if its complexity is $O(n)$ or less. The problem is very hard if its complexity is $O(2^n)$, unless the size n is small. Between these types of problems there is a class of problems, each of which can be solved in polynomial time by a Turing machine.

A †nondeterministic Turing machine is said to **solve** a decision problem (Σ, L, P) if it can detect the case when $P(\alpha)=1$ by exercising good choices; more precisely, if it can stop after a finite number of steps and write the value 1 of $P(\alpha)$ on the tape, starting from an initial state with the head on the leftmost symbol of a string α on the tape, provided that $P(\alpha)=1$. If $P(\alpha)=0$, then the machine may never stop. For a string α such that $P(\alpha)=1$, $\mathrm{N}T_M(\alpha)$ and $\mathrm{N}S_M(\alpha)$ represent the minimum number of steps and the minimum length of tape used in computing $P(\alpha)$. Nondeterministic complexities $\mathrm{N}T_M(n)$ and $\mathrm{N}S_M(n)$ for the size n are defined in a similar way as before. A problem is said to be solvable in **NP-time** or **NP-space** if it can be solved by a nondeterministic Turing machine M whose time complexity $\mathrm{N}T_M(n)$ or space complexity $\mathrm{N}S_M(n)$ is bounded by a polynomial function of size n. The class of all problems solvable in NP-time is denoted by **NP**, while that of the problems solvable in polynomial time by ordinary (deterministic) Turing machines is denoted by **P**. Obviously, the class **NP** contains the class **P**. Nevertheless, whether **NP** = **P** or not remains one of the biggest unsolved problems.

A decision problem (Σ, L, P) is **polynomially transformable** to a problem (Σ', L', P') if there exists a mapping h from Σ^+ to Σ'^+ satisfying the following conditions: (1) $h(L)\subseteq L'$; (2) for every string α in L, $P(\alpha)=P'(h(\alpha))$; (3) the mapping h is computed by a Turing machine in polynomial time. If a problem C is polynomially transformable to another problem C' in **P** or **NP**, then the problem C is also in **P** or **NP**. A problem C is said to be **NP-hard** if every problem in **NP** is polynomially transformable to C. It is called **NP-complete** if it is in **NP** and NP-hard. Many problems that have been known empirically to be very hard have recently been shown to be NP-complete [8]. For instance, integer linear programming is

NP-complete. Linear programming has recently been shown to be in **P**. The satisfiability problem of a conjunctive normal form was proven to be NP-complete in 1971 [9].

The **complement** of a decision problem (Σ, L, P) is the problem (Σ, L, P'), where $P'(\alpha) = 1 - P(\alpha)$. If a problem C is in **P**, then its complement C' is also in **P**. When C is in **NP**, it is not guaranteed that C' is in **NP**. The class of all problems whose complements are in **NP** is denoted by **co-NP**. The intersection of **NP** and co-**NP** contains the class **P**, but it is an open problem whether this containment is proper or not. The problem PRIME, which asks whether a number N is prime or not, the size n being the number of digits for representing the number N, belongs to both **NP** and co-**NP** [10]. However, it is not known whether it is in **P**.

F. Other Topics

The notion of completeness is introduced in many other classes of problems solvable in linear space, polynomial space, exponential time, etc. Comparisons among these classes have been done, although many problems still remain open. The fundamental conjecture in this field is the nonequality $\mathbf{NP} \neq \mathbf{P}$, the proving of which seems to be extremely difficult. This conjecture can be relativized in ways such that it can be either true or false, so that it is not provable in some formal system.

Some tricks may be used for hard problems. For decision problems, some algorithms have been proposed for guessing the correct answer with high probability [11]. Some algorithms give nearly optimal solutions efficiently, instead of constructing optimal ones over too long a time [7].

A general theory of computational complexity would include investigations on relations among various complexity measures, the complexity hierarchy, and an axiomatic approach independent of any machine models [12].

References

[1] J. E. Savage, The complexity of computing, Wiley, 1971.
[2] A. N. Kolmogorov, Three approaches to the quantitative definition of information, Problems of Information Transmission, 1 (1965), 1–7.
[3] P. Martin-Löf, The definition of random sequences, Information and Control, 9 (1966), 602–619.
[4] M. J. Fischer and M. O. Rabin, Super-exponential complexity of Presburger arithmetic, Complexity of Computation, R. M. Karp (ed.), Proc. SIAM-AMS Symposium in Appl. Math., 1974.
[5] A. V. Aho, J. E. Hopcroft, and J. D. Ullman, The design and analysis of computer algorithms, Addison-Wesley, 1974.
[6] D. E. Knuth, The art of computer programming, Addison-Wesley, 1, 1968; 2, 1969; 3, 1973.
[7] E. Horowitz and S. Sahni, Fundamentals of computer algorithms, Computer Science Press, 1978.
[8] M. R. Garey and D. S. Johnson, Computers and intractability, a guide to the theory of NP-completeness, Freeman, 1979.
[9] S. A. Cook, The complexity of theorem-proving procedures, Proc. 3rd Ann. ACM Symposium on Theory of Computing, 1971.
[10] G. L. Miller, Riemann's hypothesis and test for primarity, Proc. 7th Ann. ACM Symposium on Theory of Computing, 1975.
[11] J. F. Traub (ed.), Algorithms and complexity: New directions and recent results, Academic Press, 1976.
[12] M. Blum, A machine-independent theory of the complexity of recursive functions, J. Assoc. Comput. Mach., 14 (1967), 322–336.

72 (VII.9) Complex Manifolds

A. Definitions

A Hausdorff topological space X is called a **complex manifold** (or **complex analytic manifold**) of **complex dimension** n if there are given an open covering $\{U_i\}_{i\in I}$ and a family $\{\varphi_i\}_{i\in I}$ of homeomorphisms of U_i onto open sets in the n-dimensional complex affine space $\mathbf{C}^n$ such that in case $U_i \cap U_j \neq \varnothing$, the mapping $\varphi_i \circ \varphi_j^{-1}$: $\varphi_j(U_i \cap U_j) \rightarrow \varphi_i(U_i \cap U_j)$ is biholomorphic (i.e., $\varphi_i \circ \varphi_j^{-1}$ and its inverse are both †holomorphic functions when expressed in terms of coordinate functions in $\mathbf{C}^n$). We call X the **underlying topological space** of this complex manifold, and we say that an open covering $\{U_i\}_{i\in I}$ and a family $\{\varphi_i\}_{i\in I}$ define a **complex analytic structure** (or simply **complex structure**) on X.

A complex-valued function f defined on an open set U in X is called a **holomorphic function** on U if for any i the function $f \circ \varphi_i^{-1}$ on $\varphi_i(U \cap U_i)$ is holomorphic. When we express the mapping φ_i as $\varphi_i(p) = (z^1(p), \ldots, z^n(p))$ on U_i in terms of the coordinates in $\mathbf{C}^n$, each z^α is a holomorphic function on U_i. We call $(z^1, \ldots, z^n)$ a **holomorphic local coordinate system** on U_i. Given two complex manifolds Y, X, a mapping $\varphi: Y \rightarrow X$ is said to be **holo-**

morphic if for any open set U in X and any holomorphic function f on U, $f \circ \varphi$ is holomorphic on $\varphi^{-1}(U) \subset Y$. When a mapping φ: $Y \to X$ is bijective and both φ and φ^{-1} are holomorphic, we say Y and X are **isomorphic** by φ as complex manifolds.

As in the case of †differentiable manifolds of class C^∞, we can define concepts such as **complex analytic submanifolds, holomorphic tangent vectors, holomorphic vector fields**, and **holomorphic differential forms of degree** k (or simply **holomorphic** k**-forms**). **Meromorphic functions** on complex manifolds can also be defined as in the theory of analytic functions of several complex variables (→ 23 Analytic Spaces D).

Let X be a complex manifold and p a point of X. Take a holomorphic local coordinate system $(z^1, \ldots, z^n)$ with center p (i.e., $z^\alpha(p)=0$ for all α). A holomorphic function defined on a neighborhood of p can be expressed as a holomorphic function in $(z^1, \ldots, z^n)$, hence as a power series in $(z^1, \ldots, z^n)$ absolutely convergent in a neighborhood of p. If we denote by $\mathcal{O} = \mathcal{O}_X$ the †sheaf of germs of holomorphic functions on X, the †stalk $\mathcal{O}_p$ of $\mathcal{O}$ at p is isomorphic to the †local ring of convergent power series in n variables $z^1, \ldots, z^n$. At a point p, $(\partial/\partial z^1)_p, \ldots, (\partial/\partial z^n)_p$ form a basis of the holomorphic tangent vector space at p. A holomorphic k-form ω defined on a neighborhood of p can be expressed as $\omega = \sum_{i_1 < \ldots < i_k} f_{i_1 \ldots i_k} dz^{i_1} \wedge \ldots \wedge dz^{i_k}$, where $f_{i_1 \ldots i_k}$ is a holomorphic function for each $(i_1, \ldots, i_k)$.

B. Almost Complex Structures

Let X be a complex manifold, and let $\{U_i, \varphi_i\}_{i \in I}$ be its complex analytic structure, i.e., a covering of X by holomorphic local coordinate systems with $\varphi_i = (z_i^1, \ldots, z_i^n)$. Express z_i^α in the form $z_i^\alpha = x_i^\alpha + \sqrt{-1}\, y_i^\alpha$, where x_i^α and y_i^α are the real and imaginary parts of z_i^α, respectively. Then x_i^α and y_i^α real-valued functions on the open set U_i of X, and the mapping $\psi_i : U \to \mathbf{R}^{2n}$ defined by $\psi_i(p) = (x_i^1(p), y_i^1(p), \ldots, x_i^n(p), y_i^n(p))$ is a homeomorphism of U_i onto an open set of $\mathbf{R}^{2n}$. This $\{U_i, \psi_i\}_{i \in I}$ defines on X a †differentiable structure of class C^∞ (in fact, a †real analytic structure). Thus a complex manifold of complex dimension n admits canonically a C^∞-structure of real dimension $2n$. For every point p of X there is a real coordinate system on a neighborhood of p, such as $(x^1, y^1, \ldots, x^n, y^n)$, where $(z^1, \ldots, z^n)$ $(z^\alpha = x^\alpha + \sqrt{-1}\, y^\alpha, \alpha = 1, \ldots, n)$ forms a holomorphic coordinate system in X. The real tangent vector space at a point p of X has $\{(\partial/\partial x^1)_p, (\partial/\partial y^1)_p, \ldots, (\partial/\partial x^n)_p, (\partial/\partial y^n)_p\}$ as its basis. Define a linear endomorphism J_p by $(\partial/\partial x^\alpha)_p \to (\partial/\partial y^\alpha)_p$, $(\partial/\partial y^\alpha)_p \to -(\partial/\partial x^\alpha)_p$ $(\alpha = 1, \ldots, n)$; then $J_p^2 = -1$, and J_p does not depend on the choice of holomorphic coordinate system at p. Considering J_p as a tensor of type $(1,1)$, we thus obtain a tensor field J of type $(1,1)$ of class C^∞ on X, which is called the **tensor field of almost complex structure** induced by the complex structure of X.

More generally, when a real differentiable manifold X is provided with a tensor field J of type $(1,1)$ of class C^∞ such that $J^2 = -1$ (considering J as a linear transformation of vector fields), we say that X admits an **almost complex structure** or that X is an **almost complex manifold**. In this case, for contravariant vector fields x and y on X, we define a tensor field S of type $(1,2)$ by $S(x,y) = -[x,y] + [J(x), J(y)] - J([J(x), y]) - J([x, J(y)])$. S is called a **Nijenhuis tensor**. An almost complex structure J is induced by a complex analytic structure if and only if its Nijenhuis tensor S vanishes identically [49]. A differentiable manifold X of dimension $2n$ admits an almost complex structure if and only if the structure group $GL(2n, \mathbf{R})$ of the bundle of †tangent $2n$-frames of X can be †reduced to $GL(n, \mathbf{C})$. Almost complex manifolds are †orientable.

C. Types of Differential Forms

Let X be a complex manifold, and let $(z^1, \ldots, z^n)$ be a holomorphic local coordinate system in a neighborhood of a point p with $z^\alpha = x^\alpha + \sqrt{-1}\, y^\alpha$ $(\alpha = 1, \ldots, n)$. On the complexified real tangent vector space $T_p(X) \otimes \mathbf{C}$ at p, we define $\partial/\partial z^\alpha$, $\partial/\partial \bar{z}^\alpha$ by

$$(\partial/\partial z^\alpha)_p = (1/2)\{(\partial/\partial x^\alpha)_p - \sqrt{-1}\,(\partial/\partial y^\alpha)_p\},$$

$$(\partial/\partial \bar{z}^\alpha)_p = (1/2)\{(\partial/\partial x^\alpha)_p + \sqrt{-1}\,(\partial/\partial y^\alpha)_p\}.$$

It is easy to see that the operation $\partial/\partial z^\alpha$ on holomorphic functions coincides with that of the holomorphic tangent vector $\partial/\partial z^\alpha$ defined in Section A. A function f is holomorphic at p if and only if $(\partial/\partial \bar{z}^\alpha) f = 0$ $(\alpha = 1, 2, \ldots, n)$. $T_p(X) \otimes \mathbf{C}$ is the direct sum of the subspace spanned by $\{\partial/\partial z^1, \ldots, \partial/\partial z^n\}$ and the subspace spanned by $\{\partial/\partial \bar{z}^1, \ldots, \partial/\partial \bar{z}^n\}$. Moreover, this decomposition is independent of the choice of holomorphic coordinate system. Elements of the two subspaces are respectively called **tangent vectors of type** $(1,0)$ and **tangent vectors of type** $(0,1)$. Similarly, the complexified space of real differentiable 1-forms can be decomposed into the direct sum of two subspaces spanned by $\{dz^1, \ldots, dz^n\}$ and $\{d\bar{z}^1, \ldots, d\bar{z}^n\}$, where $dz^\alpha = dx^\alpha + \sqrt{-1}\, dy^\alpha$ and $d\bar{z}^\alpha = dx^\alpha - \sqrt{-1}\, dy^\alpha$. We say that the elements of the former subspace are of type $(1,0)$ and those of the latter are of type $(0,1)$. Thus the space of differential forms of arbitrary degree can be written as

the direct sum of subspaces of type (r,s). Here the subspace of **differential forms of type** (r,s) has as a basis $\{dz^{\alpha_1}\wedge\ldots\wedge dz^{\alpha_r}\wedge d\bar{z}^{\beta_1}\wedge\ldots\wedge d\bar{z}^{\beta_s}$ $(1\leqslant\alpha_1<\ldots<\alpha_r\leqslant n,\ 1\leqslant\beta_1<\ldots<\beta_s\leqslant n)\}$. This decomposition is independent of the choice of a local coordinate system, hence the concept of type can be defined globally on X.

D. $\bar{\partial}$-Cohomology

In the rest of this article we consider only complex differential forms. For every differential form ω of type (r,s) on X, its †exterior derivative $d\omega$ decomposes into a sum of differential forms of types $(r+1,s)$ and $(r,s+1)$, which we denote by $\partial\omega$ (or $d'\omega$) and $\bar{\partial}\omega$ (or $d''\omega$), respectively. We have $d=\partial+\bar{\partial}$, $(d')^2=0$, $(\bar{\partial})^2=0$, and $\partial\bar{\partial}+\bar{\partial}\partial=0$. In terms of a local coordinate system, we write

$$\partial\omega=\sum_\gamma\frac{\partial f}{\partial z^\gamma}dz^\gamma\wedge dz^{\alpha_1}\wedge\ldots\wedge dz^{\alpha_r}\wedge d\bar{z}^{\beta_1}\wedge\ldots\wedge d\bar{z}^{\beta_s},$$

$$\bar{\partial}\omega=(-1)^r\sum_\gamma\frac{\partial f}{\partial\bar{z}^\gamma}dz^{\alpha_1}\wedge\ldots\wedge dz^{\alpha_r}\wedge d\bar{z}^\gamma\wedge d\bar{z}^{\beta_1}\wedge\ldots\wedge d\bar{z}^{\beta_s},$$

for $\omega=f dz^{\alpha_1}\wedge\ldots\wedge dz^{\alpha_r}\wedge d\bar{z}^{\beta_1}\wedge\ldots\wedge d\bar{z}^{\beta_s}$. A differential k-form ω is holomorphic if and only if ω is of type $(k,0)$ and $\bar{\partial}\omega=0$.

For the operator $\bar{\partial}$ **Dolbeault's lemma** holds: Let ω be a differential form on a neighborhood U of a point p. If $\bar{\partial}\omega=0$, there is a neighborhood V of p contained in U and a differential form θ on V such that $\omega=\bar{\partial}\theta$ on V.

Let $A^{(r,s)}$ and Ω^p be the †sheaf of germs of differential forms of type (r,s) and the sheaf of germs of holomorphic p-forms on X, respectively, and let $\Gamma(X,A^{(r,s)})$ be the set of †sections of $A^{(r,s)}$ on X. $\Gamma(X,A^{(r,s)})$ is the set of differential forms of type (r,s) on X, and $\sum_i\Gamma(X,A^{(p,i)})$ forms a †cochain complex with respect to $\bar{\partial}$. This complex is called the **$\bar{\partial}$-complex** or the **Dolbeault complex**, and its †cohomology groups are called the **$\bar{\partial}$-cohomology groups** or the **Dolbeault cohomology groups**. The qth cohomology group is denoted by $H^{p,q}(A,\bar{\partial})$. It follows easily from Dolbeault's lemma that $0\rightarrow\Omega^p\rightarrow A^{(p,0)}\xrightarrow{\bar{\partial}}A^{(p,1)}\rightarrow\ldots$ is an †exact sequence of sheaves. From this we get **Dolbeault's theorem**:

$$H^q(X,\Omega^p)\cong H^{p,q}(A,\bar{\partial}),$$

where the left-hand side is the cohomology group with coefficient sheaf Ω^p.

More generally, for any †complex analytic (holomorphic) vector bundle E on X, we can define the $\bar{\partial}$ cohomology groups of the differential forms on X with values in E, and they can be shown to be isomorphic to the cohomology groups with coefficient sheaves of germs of holomorphic forms with values in E (→ 194 Harmonic Integrals E).

E. Analytic Coherent Sheaves

The structure of a complex manifold X is determined by the †sheaf $\mathcal{O}$ of germs of holomorphic functions on X, and $\mathcal{O}$ is a †coherent sheaf (of rings) (**Oka's theorem**). Sheaves of $\mathcal{O}$-modules are called **analytic sheaves**, and coherent sheaves of $\mathcal{O}$-modules are called **coherent analytic sheaves**. Many properties of X can be expressed in terms of coherent analytic sheaves and their cohomology groups; some examples appear later in this article (also → 366 Riemann-Roch Theorems B).

It is important to know whether an analytic sheaf on a complex manifold is coherent. To this question, not only does Oka's theorem apply but so does **Cartan's theorem**: The †sheaf of ideals defined by an analytic subset of a complex manifold is coherent. (We say that a subset Y of X is an **analytic subset** if it is a closed subset and each point of Y has a neighborhood U such that $U\cap Y$ is the set of common zeros of a finite number of holomorphic functions on U.) Also relevant is **Grauert's theorem**: If $\pi: X\rightarrow Y$ is a †proper holomorphic mapping of complex manifolds (i.e., the inverse image of any compact subset of Y for a holomorphic mapping π is also compact), then for any coherent analytic sheaf F on X its †direct images $R^q\pi_*(F)$ $(q=0,1,2,\ldots)$ are also coherent [16]. (In fact, this theorem holds for analytic spaces; → 23 Analytic Spaces.)

For an analytic coherent sheaf F on a †Stein manifold X, we have the following **fundamental theorems of the Stein manifold. Theorem A**: $H^0(X,F)$ generates the stalk F_x (as an $\mathcal{O}_x$-module) at every point x of X. **Theorem B**: $H^q(X,F)=0$ for all $q>0$. Conversely, a Stein manifold X is characterized by the following property: For any coherent analytic sheaf F of ideals of $\mathcal{O}$, $H^1(X,F)=0$. If a complex manifold X is compact and F is a coherent analytic sheaf on X, then $H^q(X,F)$ is a complex vector space of finite dimension. If X is an open submanifold of another complex manifold and its closure is compact, then $H^q(X,F)$ is finite-dimensional for some q that depends on various properties (convexity or concavity) of the boundary of X [3].

Let E be a †complex analytic (holomorphic) vector bundle on a complex manifold X of dimension n, and let E^* be the †dual vector bundle of E. Then $H^q(X,\mathcal{O}^p(E))$ and $H_*^{n-q}(X,\mathcal{O}^{n-p}(E^*))$ (where H_* denotes the cohomology group with compact support) are dual as topological vector spaces under suit-

able conditions. The duality is given by the integration on X of the exterior products of the differential forms representing the respective elements of the cohomology groups. The duality holds, for example, when dim $H^q(X, \mathcal{O}^p(E)) < \infty$. If X is compact, we need not distinguish H_* from H (**Serre's duality theorem**) [50].

F. Compact Complex Manifolds

On a compact and connected complex manifold X, there are no holomorphic functions except constants (by the †maximum principle of holomorphic functions). The field $K(X)$ of meromorphic functions on X is finitely generated over the complex number field, and its †transcendence degree d does not exceed the complex dimension n of X. d is said to be the **algebraic dimension** of X and is denoted by $a(X)$. For elements of $K(X)$, functional independence and algebraic independence are equivalent [52]. When $n=1$, X is a compact †Riemann surface and the classical theory of algebraic functions shows that $K(X)$ is an †algebraic function field of one variable and X is a †projective algebraic variety. When $n=2$, $a(X)=2$, 1, and 0 can all occur. If $a(X)=2$, X is a projective †algebraic surface (**Chow-Kodaira theorem**). If $a(X)=1$, there exist an †algebraic curve Δ and a surjective holomorphic mapping $\varphi: X \rightarrow \Delta$ such that $K(\Delta)$ is isomorphic to $K(X)$ under φ^*, and $\varphi^{-1}(x)$ is an †elliptic curve for all but a finite number of $x \in \Delta$. K. Kodaira investigated the structure of compact complex surfaces in detail [34, III].

On a compact complex manifold X, the free Abelian group generated by the set of irreducible analytic subsets of codimension 1 is called the **divisor group** of X, and an element of it is called a **divisor** of X. For an analytic subset Y of codimension 1, the sheaf of ideals $\mathfrak{J}(Y)$ defined by Y is a sheaf of locally principal ideals of $\mathcal{O}$. For a divisor $D=\sum a_\alpha Y_\alpha$, the sheaf of locally principal fractional ideals $\mathfrak{J}(D)=\prod_\alpha \mathfrak{J}(Y_\alpha)^{a_\alpha}$ is called the **sheaf of ideals of** D. The set of nonzero coherent sheaves of locally principal fractional ideals corresponds bijectively to the set of divisors. An element $f \neq 0$ of $K(X)$ generates a sheaf of principal fractional ideals and therefore defines a divisor, which is denoted by (f). The divisor group has an ordering defined by $D=\sum a_\alpha Y_\alpha \succ 0$ if and only if all $a_\alpha \geqslant 0$, under which it becomes an †ordered group. For a divisor D, let

$$L(D)=\{f \in K(X) \mid f \neq 0, (f)+D \succ 0\} \cup \{0\}.$$

Then $L(D)$ is a $\mathbf{C}$-module of finite dimension. This submodule of $K(X)$ is easy to handle and exhibits various analytic properties of D. The †Riemann-Roch theorem is used to calculate the dimension of $L(D)$ in terms of other factors (→ 366 Riemann-Roch Theorems C). Here is an example of how $L(D)$ exhibits a property of D: We call two divisors D and D' **linearly equivalent** if there is a $0 \neq F \in K(X)$ such that $(F)=D-D'$. This is an equivalence relation finer than †homological equivalence. If D and D' are linearly equivalent, then $L(D)$ and $L(D')$ are isomorphic by the mapping $L(D) \ni f \rightarrow fF \in L(D')$; therefore $\dim L(D)=\dim L(D')$. (The latter equation does not follow from the homological equivalence.) A holomorphic vector bundle with fiber $\mathbf{C}$ and structure group $\mathbf{C}^*$ is called a **complex line bundle**. For the sheaf of ideals $\mathfrak{J}(D)$ of a divisor D, we can take a suitable open covering $\{U_j\}$ of X such that for each U_j there is an $R_j \in \Gamma(U_j, \mathfrak{J}(D))$ which generates $\mathfrak{J}(D)_x$ for any $x \in U_j$. Also $g_{jk}=R_j/R_k$ is a holomorphic function nowhere vanishing on $U_j \cap U_k$. With $\{g_{jk}\}$ as the system of †coordinate transformations, we define the **complex line bundle determined by** D and denote it by $[D]$. It is easy to see that $[D]$ is independent of the choice of $\{U_j\}$ or $\{R_j\}$. Moreover, $[D]$ is determined only by the linear equivalence class of D. If we denote by $\mathcal{O}([D])$ the sheaf of germs of holomorphic sections of $[D]$, the mapping $H^0(X, \mathcal{O}([D])) \ni \varphi=\{\varphi_j\} \rightarrow f=\varphi_j/R_j=\varphi_k/R_k \in L(D)$ is an isomorphism of these modules. On an algebraic variety in a projective space, any complex line bundle comes from a divisor (i.e., it can be expressed in the form $[D]$ for some divisor D) [35, 51], but this is not necessarily true on general compact complex manifolds. However, the importance of complex line bundles in the theory of complex manifolds lies in the relation $L(D) \cong H^0(X, \mathcal{O}([D]))$, which replaces "things with poles" with "things holomorphic."

Any analytic subvariety X of the projective space $\mathbf{P}^n$ is an algebraic variety (**Chow's theorem**) [51]. Suppose that X is an analytic submanifold of $\mathbf{P}^N$, and let Y be a general †hyperplane section of X. Then Y is a divisor on X, and the †Chern class of the complex line bundle $[Y]$ corresponds to the canonical Hodge metric on X (→ 232 Kähler Manifolds D). When $[Y]$ is represented by the system of coordinate transformations $\{g_{jk}\}$ with respect to an open covering $\{U_i\}$, we can associate with a coherent analytic sheaf F sheaves $F(n)$ $(n=0, \pm 1, \pm 2, \ldots)$ as follows. Denoting by F_j the restriction of F to U_j, we glue F_j and F_k together on $U_j \cap U_k$ with the relation $f_j \sim f_k \Leftrightarrow f_j = g_{kj}^n f_k$ (where $f_j \in F_j \subset F$, $f_k \in F_k \subset F$) and obtain a sheaf (denoted by $F(n)$) that is locally isomorphic to F. The following theorems for $F(n)$ hold. For each coherent analytic sheaf F there exists an integer n_0 such that for any $n \geqslant n_0$ the following **fundamental theorems A, B of projec-**

tive algebraic varieties hold [17]. **Theorem A**: $\Gamma(X, F(n))$ generates F_x (as an $\mathcal{O}_x$-module) for every $x \in X$. **Theorem B**: $H^q(X, F(n)) = 0$ for all $q > 0$. This means that if we permit "poles" on Y of sufficiently high order, then F has sufficiently many sections and the higher cohomology groups vanish.

On a (nonsingular) algebraic variety X in $\mathbf{P}^N$ we have the sheaf $\mathcal{O}$ of germs of holomorphic functions (the †structure sheaf as a complex manifold) and the sheaf $\mathcal{O}_{\mathrm{alg}}$ of germs of holomorphic rational functions (the structure sheaf as an algebraic variety). Therefore we have two kinds of coherent sheaves, coherent analytic sheaves and **coherent algebraic sheaves**. In fact, the cohomology theories derived from them are isomorphic. More precisely, for any coherent algebraic sheaf F, $\tilde{F} = F \otimes_{\mathcal{O}_{\mathrm{alg}}} \mathcal{O}$ is a coherent analytic sheaf. The correspondence $F \rightarrow \tilde{F}$ gives an equivalence between the †category of coherent algebraic sheaves and that of coherent analytic sheaves, thus giving an isomorphism of their cohomology groups. In other words, as far as the properties that can be expressed by cohomology-theoretic terms of coherent sheaves are concerned, there is no difference between the analytic and algebraic theories of projective algebraic varieties [51].

G. Deformations of Complex Structures

The deformation theory of complex structures was initiated by Kodaira and Spencer [36] in order to explain various phenomena in the theory of (compact) complex manifolds. A triple (X, π, S) is called a **family of compact complex manifolds** if X and S are connected †analytic spaces and π is a †proper holomorphic mapping of X onto S such that (i) it is smooth, i.e., is locally identified with the projection $S' \times U \rightarrow S'$, where S' and U are open in S and $\mathbf{C}^n$, respectively, and (ii) every fiber $V_s = \pi^{-1}(s)$ of π is connected. V_s is then a compact complex manifold. We sometimes write $\{V_s\}_{s \in S}$ instead of (X, π, S). S is called the **parameter space of the family**. Take a point $o \in S$ and fix it. We say that V_s is a **deformation of** V_o for any $s \in S$. There are a neighborhood S' of o in S and a diffeomorphism $\pi^{-1}(S') \rightarrow S' \times V_0$ (Kuranishi [39]). Thus every V_s, $s \in S'$, determines a †complex structure on a fixed differentiable manifold. We say that the complex structure of V_s, $s \in S'$, is a **deformation of** that of V_0.

Now $(X, \pi, S) = \{V_s\}_{s \in S}$ is said to be **complete at** o if for any family $(Y, \mu, T) = \{W_t\}_{t \in T}$ with a point $o' \in T$ and a holomorphic isomorphism $\iota: W_{o'} \rightarrow V_o$, there is a neighborhood T' of o' in T and holomorphic mappings $f: T' \rightarrow S$ and $h: \mu^{-1}(T') \rightarrow X$ such that (i) $f\mu = \pi h$, (ii) $f(o') = o$, and (iii) $h = \iota$ on $W_{o'}$. In this case, if T' is sufficiently small, then h induces a holomorphic isomorphism $h_t: W_t \rightarrow V_{f(t)}$, for any $t \in T'$. Hence $\{V_s\}_{s \in S}$ contains all small deformations of V_0. On the other hand, $\{V_s\}_{s \in S}$ is said to be **effectively parametrized at** o if the **Kodaira-Spencer mapping (Kodaira-Spencer map)** $\rho_o: T_o S \rightarrow H^1(V_o, \Theta)$ is injective, where $T_o S$ is the Zariski tangent space to S at o and Θ is the sheaf of germs of holomorphic vector fields on V_0. Here, the linear mapping ρ_o is defined by $\rho_o(\partial/\partial s) = \{(\partial g_{jk}/\partial s)_{(z_k, o)}\}$, where $\partial/\partial s \in T_o S$ and g_{jk} are the coordinate transformations $(z_j, s) = (g_{jk}(z_k, s), s)$, using the smoothness of π. $\rho_0(\partial/\partial s)$ is called the **infinitesimal deformation to the direction** $\partial/\partial s$.

Kuranishi's fundamental theorem [39, 40] states: For any compact complex manifold V, there exists a family $\{V_s\}_{s \in S}$ with a point $o \in S$ such that (i) it is complete at every point of S, (ii) is effectively parametrized at o, and (iii) $V_o = V$.

The parameter space S in the theorem is called the **Kuranishi space** or the **local moduli space of** V. It is given as the zeros of a holomorphic mapping $f: U \rightarrow H^2(V, \Theta)$ with $f(0) = 0$, where U is a neighborhood of 0 in $H^1(V, \Theta)$. Hence (1) if $H^2(V, \Theta) = 0$, then $S = U$ is nonsingular (Kodaira, Nirenberg, and Spencer [37]; (2) if $H^1(V, \Theta) = 0$, then $S = \{0\}$, i.e., one point (e.g., $V = \mathbf{P}^n(\mathbf{C})$, the complex projective space).

Kuranishi's theorem was generalized to a compact analytic space V by Grauert [18] and A. Douady [8].

There are Kuranishi-type theorems in the deformation theory of other objects: (1) compact analytic subvarieties of an analytic space (Douady [9]; → 23 Analytic Spaces G), (2) holomorphic mappings (Y. Miyajima [41]), (3) germs of analytic spaces with †isolated singularities [7], etc.

So far, only the local theory of deformations has been developed. The global theory is not yet in a satisfactory state. Its final purpose is to construct moduli spaces and to understand them. For a compact differentiable manifold $\mathbf{V}$, we denote by $M(\mathbf{V})$ the set of all isomorphism classes of complex structures on $\mathbf{V}$. It is difficult in general to determine the set $M(\mathbf{V})$. As yet unsolved problems are: (1) Is $M(S^6)$ nonempty? (2) Does $M(\mathbf{P}^n(\mathbf{C}))$ $(n \geqslant 3)$ consist of one point? ($M(\mathbf{P}^2(\mathbf{C}))$ is known to consist of one point (S. T. Yau; → 232 Kahler Manifolds C).

If $M(\mathbf{V})$ has a reasonable structure (e.g., an analytic space structure) and a universal property, then we call it the **moduli space** (→ 11 Algebraic Functions F, 16 Algebraic Varieties W). Kodaira [34, III] constructed the moduli space of Hopf surfaces. Only a few examples of moduli spaces are known. It is to be noted that moduli spaces cannot in general exist.

One of the reasons for this is that there may exist **jumpings of structures** (Kodaira and Spencer [36]).

As for †algebraic manifolds (varieties), there are two known methods for the moduli problem: (1) Griffiths' period mapping (→ 16 Algebraic Varieties V) and (2) Mumford's geometric invariant theory (→ 16 Algebraic Varieties V, W).

H. Monoidal Transformations

Let Y be an analytic subspace of a complex manifold X, which is defined by a nonzero coherent sheaf of ideals $\mathfrak{J}$. Then X has an open covering such that for each member U of the covering, there are elements $\varphi_1, \ldots, \varphi_m \in \Gamma(U, \mathfrak{J})$ that generate the stalks $\mathfrak{J}_x$ at all $x \in U$. Let W' be the graph of the holomorphic mapping $x \mapsto (\varphi_1(x) : \varphi_2(x) : \ldots : \varphi_m(x))$ from $U - U \cap Y$ to $\mathbf{P}^{m-1}$, and denote by W the closure of W' in $U \times \mathbf{P}^{m-1}$. Then W is an analytic space, possibly with singularities, which is independent of the choice of the generators $\{\varphi_j\}$ and is determined uniquely by U and $\mathfrak{J}$. Therefore all the W's can be glued together to form an analytic space $\tilde{X}$ and to determine a holomorphic mapping $p: \tilde{X} \rightarrow X$. We call p the **monoidal transformation** (or **blowing-up**) of X with center $\mathfrak{J}$ or with center Y. When Y is a point, p is also called a **locally quadratic transformation** or a **σ-process**. When Y is an analytic submanifold, $\tilde{X}$ also is a manifold, $\tilde{Y} = p^{-1}(Y)$ is a nonsingular divisor on $\tilde{X}$, and $p^{-1}(y) \cong p^{k-1}$ for all $y \in Y$, where k is the codimension of Y. Moreover, the line bundle $[\tilde{Y}]$ restricted to each fiber $p^{-1}(y)$ is isomorphic to the Hopf bundle L on $\mathbf{P}^{k-1}$, i.e., the line bundle associated to the natural $\mathbf{C}^*$ bundle $\mathbf{C}^k \rightarrow \mathbf{P}^{k-1}$. Conversely, suppose that we are given $\tilde{Y} \subseteq \tilde{X}$ and a holomorphic mapping $p_0: \tilde{Y} \rightarrow Y$ with the property as above so that, in particular, $p_0^{-1}(y) \cong \mathbf{P}^{k-1}$ and $[\tilde{Y}] \cong L$ on $p_0^{-1}(y)$ for all $y \in Y$. Then there exists a complex manifold containing Y as a submanifold such that $\tilde{X}$ is obtained by the monoidal transformation of X with center Y as above (S. Nakano [47]). Let $f: X \rightarrow Y$ be a proper modification of complex manifolds (or analytic spaces). Then there exists a proper holomorphic mapping $h: Y' \rightarrow Y$ such that $f^{-1} \circ h: Y' \rightarrow X$ is holomorphic, where over any relatively compact subdomain of Y, h is obtained by a finite succession of monoidal transformations with nonsingular centers (Hironaka). The result is called **Chow's lemma**.

I. Fiber Spaces

A triple of compact complex manifolds V, W, and a surjective holomorphic mapping f: $V \rightarrow W$ is called a **fiber space** if general fibers $f^{-1}(w)$ are irreducible. In addition, if both V and W are algebraic, it is called an **algebraic fiber space**.

For a compact complex manifold V, letting K_V denote the canonical line bundle of V, define a subfield $\mathfrak{K}$ of $\mathbf{K}(V)$ to be $\{\omega_1/\omega_2$, where $\omega_1, \omega_2 \in H^0(V, \Omega)$ for some $m > 0\}$. The transcendental degree of $\mathfrak{K}$ over $\mathbf{C}$ is denoted by $\kappa(V)$, which we call the **Kodaira dimension** of V.

If $\mathfrak{K} = 0$, then define $\kappa(V) = -\infty$. If $\kappa(V) = \dim V$, then V is said to be of **general type**, and there exist a projective manifold X and a bimeromorphic holomorphic mapping $\mu: X \rightarrow V$. If $\kappa(V) \geqslant 0$, there exists a fiber space $f: V^* \rightarrow W$ such that (1) V^* is bimeromorphically equivalent to V, (2) $\dim W = \kappa(V)$, and (3) some general fibers $f^{-1}(w)$ satisfy $\kappa(f^{-1}(w)) = 0$. In general, for a fiber space $f: V \rightarrow W$, we have $\kappa(V) \leqslant \kappa(f^{-1}(w)) + \dim W$, where $f^{-1}(w)$ is a general fiber. Moreover, if it is algebraic, an inequality of the form $\kappa(V) \geqslant \kappa(f^{-1}(w)) + \kappa(W)$ is called the **conjecture** C_n, n being $\dim V$. C_n has been verified by K. Ueno, E. Viehweg, T. Fujita, and Y. Kawamata [13, 14, 32, 33, 54–57] in the following cases: (1) $n \leqslant 3$, (2) when general fibers are curves, (3) $\dim w = 1$, $\kappa(w) = 1$, (4) $\kappa(V) \geqslant 0$, and W is of general types, etc. By using the results of case (4), Kawamata proves that an algebraic compact complex manifold V is birationally equivalent to an Abelian variety if and only if $\kappa(V) = 0$ and the irregularity of V equals the dimension of V [32].

J. Analytic Surfaces

In what follows, an analytic surface means a 2-dimensional compact complex manifold. For an analytic surface S, an exceptional curve on S and a (relatively) minimal model, etc., are defined with respect to bimeromorphic mappings in analogy with the corresponding concepts for an algebraic surface (→ 15 Algebraic Surfaces). Let $C \subset S$ be an irreducible curve on S. Then there exists a holomorphic mapping φ from S onto another surface S' such that $\varphi(C)$ is a point and such that φ induces the isomorphism $S - C \cong S' - \varphi(C)$ if and only if $C^2 = -1$ and C is a nonsingular rational curve (Grauert [17]). S has a minimal model if and only if S is not a ruled surface (Kodaira [34, III]). The irregularity $q = h^{0,1}$, the geometric genus p_g, i-genus P_i, etc., are also defined in the same way as in the case of algebraic surfaces. Note that, in general, $h^{0,1} \neq h^{1,0}$. The Riemann-Roch theorem and M. Noether's formula are valid also for an analytic surface (Atiyah and Singer; → 366 Riemann-Roch Theorems C).

K. Classification of Surfaces

The classification of analytic surfaces by the aid of their numerical invariants was completed by Kodaira and includes as a special case Enriques's classification of algebraic surfaces [34, III]. By an **elliptic surface** we mean a surface from which there exists a surjective holomorphic mapping φ onto an algebraic curve Δ, such that for a general point p on Δ, $\varphi^{-1}(p)$ is an irreducible nonsingular elliptic curve. If $a(S)=1$ or $\kappa(S)=1$, S has a unique structure as an elliptic surface. The image by φ of the points on S at which φ is not of maximal rank is a finite subset $\{a_1, \ldots, a_r\}$ of Δ. Let t_i be a local coordinate on Δ around a_i with $t_i(a_i)=0$. We call a singular fiber of φ the divisor on S defined by $\{t_i \circ \varphi = 0\}$. The structure and the construction of singular fibers of elliptic surfaces have been completely determined by Kodaira. By an elliptic surface of general type we mean a surface with Kodaira dimension 1. If $\kappa(S)=2$, S is projective algebraic and is called a **surface of general type**. If $a(S)=0$, then there exists only a finite number of irreducible curves on S. By a **Hopf surface** we mean a surface whose universal covering is $\mathbf{C}^2-(0,0)$. If an analytic surface S is homeomorphic to $S^1 \times S^3$, S is a Hopf surface. Let $b_\nu(S)$ be the νth Betti number of S. If $a(S)=0$, $b_1(S)=1$, $b_2(S)=0$, and S contains a curve, then S is a Hopf surface (Kodaira). By a surface of class VII_0 we mean a minimal surface S with $b_1(S)=1$. M. Inoue [26] constructed three families of surfaces of class VII_0 with $b_2=0$ which contain no curves. These are $S_M^{(+)}$ $(M \in SL(3,\mathbf{Z}))$ and $S_{N,p,q,r,t}^{(+)}$ and $S_{N,p,q,r}^{(-)}$ $(N \in SL(2,\mathbf{Z})$, $p, q, r \in Z$, $t \in \mathbf{C})$ which have $H \times \mathbf{C}$ as their universal covering surfaces, H being the complex upper half-plane. These have a line bundle L such that $H^0(\Omega^1 \otimes \mathcal{O}(L))=0$. This property characterizes these Inoue surfaces among VII_0 surfaces with $b_2=0$ which contain no curves. A couple of new surfaces of class VII_0 with $b_2>0$ were constructed and studied by Inoue, Ma. Kato, T. Oda, I. Nakamura, and I. Enoki [10, 27–29, 46]. Some of these have close connections with cusp singularities of Hilbert modular surfaces, torus embeddings, and global spherical shells. Enoki's surface is denoted by $S_{n,\alpha,t}$ $(n>0, 0<|\alpha|<1, t \in \mathbf{C}^n)$ and has the following properties: (1) class VII_0 and $b_2=n$, (2) there exists a connected curve D with $D^2=0$, (3) $S_{n,\alpha,t}-D$ is an affine bundle over an elliptic curve. Properties (1) and (2) characterize $S_{n,\alpha,t}$ (Enoki).

Let K_S denote the canonical line bundle of an analytic surface S. An analytic surface S is said to be a $K3$ **surface**, if K_S is trivial, i.e., if there exists a nonvanishing holomorphic 2-form ω and if $q=0$. $K3$ surfaces are simply connected and are deformations of a nonsingular quartic surface in $\mathbf{P}^3$ (Kodaira). We shall define the period mapping of $K3$ surfaces. Let L be a free Abelian group of rank 22 with the pairing $\langle\ \rangle$ which is the direct sum of the two copies of $-E_8$ and three copies of $U=\mathbf{Z}e_1+\mathbf{Z}e_2$ with $\langle e_1,e_1\rangle=\langle e_2,e_2\rangle=0$, $\langle e_1,e_2\rangle=\langle e_2,e_1\rangle=1$, where $-E_8$ denotes the lattice with the pairing corresponding to the Dynkin diagram E_8 with the opposite signs. $H^2(S,\mathbf{Z})$ with the intersection form is such a pair, consisting of S and $\langle\ \rangle$. A **marked** $K3$ **surface** is defined to be a pair (S,ψ), where S is a $K3$ surface and $\psi: H^2(S,Z) \rightarrow L$ is an isomorphism preserving $\langle\ \rangle$. The class $[\omega]$ of ω is a base of $H^{2,0}$ and satisfies $\langle[\omega],[\omega]\rangle=0$ and $\langle[\omega],[\bar{\omega}]\rangle>0$. Let $\mathbf{P}(L_C)$ denote the 21-dimensional projective space associated with $L_\mathbf{C}=L\otimes_\mathbf{Z}\mathbf{C}$. $\mathcal{D}=\{(a)\in\mathbf{P}(L_C) \mid \langle a,a\rangle=0, \langle a,\bar{a}\rangle>0\}$ is an open set of the quadric $\mathcal{D}_c=\{(a)\in\mathbf{P}(L_C)\mid\langle a,a\rangle=0\}$. $\mathcal{D}$ is the set of all Hodge structures on $L_\mathbf{C}$. The **period** of a marked $K3$ surface (S,ψ) is defined to be $(\psi_\mathbf{C}[\omega])$. Every point of $\mathcal{D}$ is the period of some marked Kähler $K3$ surface. If the periods of two marked Kähler $K3$ surfaces (S,ψ) and (S',ψ'), coincide then S is isomorphic to S'.

An analytic surface with $p_g=q=0$ and $K_S^{\otimes 2}\cong 0$ is called an **Enriques surface**, which has a $K3$ surface as its universal covering. An Enriques surface is an algebraic elliptic surface. An analytic surface with $q=1$ and $K_S^{\otimes 12}\cong 0$ is said to be a **hyperelliptic surface**, which has an Abelian surface as an unramified covering. It is an algebraic surface and is an elliptic bundle over an elliptic curve. The classification of minimal surfaces is given in Table 1. The following relations hold among these invariants: Let b^+ (b^-) be the number of positive (negative) eigenvalues (counted with multiplicities) of the intersection matrix on $H^2(S,\mathbf{R})$ and c_i be the ith Chern class of S. Then

(1) $b^+-b^-=\frac{1}{3}(c_1^2-2c_2)$ (**Hirzebruch signature theorem**);

(2) if b_1 is even, $q=h^{1,0}=\frac{1}{2}b_1$ and $b^+=2p_g+1$;

(3) if b_1 is odd, $q=h^{1,0}+1=\frac{1}{2}(b_1+1)$ and $b^+=2p_g$.

Let c_1^2 and c_2 denote the Chern numbers of an analytic surface S, i.e., $c_1^2=(K_S^2)$ and c_2 is the Euler number of S. Then $3c_2\geqslant c_1^2$ [43], and if equality holds, then K_S is ample (Y. Miyaoka) (→ 232 Kähler Manifolds C). If S is a surface of general type which is minimal, then $c_1^2>0$, the bigenus $P_2\geqslant 2$ and the m-genus P_m is $m(m-1)c_1^2/2+1-q+p_g$ for $m\geqslant 2$. For each $m\geqslant 5$, the mth canonical mapping of S is a birational holomorphic mapping onto its image (Kodaira, E. Bombieri [5]). In general, $c_1^2\geqslant 2p_g-4$, and if the canonical mapping, i.e., the

rational mapping associated with K_S, is birational, then $c_1^2 \geqslant 3p_g - 7$. For certain kinds of surfaces of general type, E. Horikawa succeeded in determining completely the structure of surfaces obtained as deformations of a given surface. Among others, every minimal surface with $p_g = 4$, $q = 0$, and $c_1^2 = 5$ is a deformation of a nonsingular quintic surface [22, 23].

Table 1. Classification of Minimal Surfaces

κ	p_g	P_{12}	q	b_1	Structure
2		>0			algebraic surface of general type
1					elliptic surface of general type
0	1	1	2	4	complex torus
	1	1	2	3	elliptic surface with a trivial canonical bundle
	0	1	1	2	hyperelliptic surface
	0	1	1	1	elliptic surface belonging to class VII_0
	1	1	0	0	$K3$ surface
	0	1	0	0	Enriques surface
$-\infty$	0	0	0	0	rational surface
			$\geqslant 1$	$2q$	ruled surface of genus q
			1	1	surface of class VII_0

For $a, b > 0$, let Σ_{ab} be the set of isomorphism classes of all minimal surfaces with $c_1^2 = a$ and $1 - q + p_g = b$. D. Gieseker [15] proved the existence of N and constructed the mapping $h: \Sigma_{ab} \to \mathbf{P}^N$ so that the following hold: (1) h is injective; (2) if $f: X \to Y$ is a smooth holomorphic mapping such that all $h^{-1}(y)$ represent some classes $\in \Sigma_{ab}$, then the set-theoretic mapping $f: Y \to \mathbf{P}^N$ induced from h is a morphism of schemes; (3) $h(\Sigma_{ab})$ is a locally closed subvariety of $\mathbf{P}^N$. Hence, $h(\Sigma_{ab})$ is a moduli variety of surfaces of general type with $c_1^2 = a$ and $1 - q + p_g = b$.

For each m with $1 \leqslant m \leqslant 9$, there exists a minimal surface of general type with $p_g = q = 0$ and $c_1^2 = m$.

If there exist a compact analytic surface S and a curve C on S such that $S - C$ is biholomorphic to a complex manifold M, then we say that S is a **compactification** of M with boundary C. Every compactification of $\mathbf{C}^2$ is a rational surface (Kodaira, J. Morrow [45]). Every compactification of $\mathbf{C} \times (\mathbf{C} - \{0\})$ is also rational (T. Ueda [53]). However, all compactifications of $(\mathbf{C} - \{0\})^2$ are rational surfaces, certain kinds of Hopf surfaces, or $\mathbf{P}^1$-bundles over elliptic curves as constructed by Serre (Ueda).

References

[1] T. Akahori, Complex analytic construction of the Kuranishi family on a normal strongly pseudo-complex manifold. Publ. Res. Inst. Math. Sci., 14 (1978), 789–847.
[2] I. Amemiya, K. Masuda, and K. Shiga, Lie algebras of differential operators, Osaka J. Math., 12 (1975), 139–172.
[3] A. Andreotti and H. Grauert, Théorèmes de finitude pour la cohomologie des espaces complexes, Bull. Soc. Math. France, 90 (1962), 193–259.
[4] A. Blanchard, Sur les varietes analytiques complexes, Ann. Sci. Ecole Norm. Sup., 73 (1956), 157–202.
[5] E. Bombieri, Canonical models of surfaces of general type, Publ. Math. Inst. HES, 42 (1973), 171–219.
[6] M. Commichau, Deformation kompakter komplexer Mannigfaltigkeiten, Math. Ann., 213 (1975), 43–96.
[7] I. F. Donin, Complete families of deformations of germs of complex spaces (in Russian), Mat. USSR Sb., 18 (1972), 397–406.
[8] A. Douady, Le problème des modules pour les variétés analytiques complexes, Sém. Bourbaki, 1964/65, no. 277.
[9] A. Douady, Le problème des modules locaux pour les espaces C-analytiques compacts, Ann. Sci. École Norm. Sup., sér. IV, 4 (1974), 569–602.
[10] I. Enoki, On surfaces of class VII_0 with curves, Proc. Japan Acad., 56 (1980), 275–279.
[11] O. Forster and K. Knorr, Konstruction verseller Familien kompakter komplexer Räume, Lecture notes in math. 705, Springer, 1979.
[12] A. Fujiki, On the blowing down of analytic spaces, Publ. Math. Res. Inst. Math. Sci., 10 (1975), 473–507.
[13] T. Fujita, On Kähler fiber spaces over curves, J. Math. Soc. Japan, 30 (1978), 779–794.
[14] T. Fujita, The sheaf of relative canonical forms of a Kähler fiber space over a curve, Proc. Japan Acad., 54 (1978), 183–184.
[15] D. Gieseker, Global moduli for surfaces of general type, Inventiones Math., 43 (1977), 233–282.
[16] H. Grauert, Ein Theorem der analytischen Garbentheorie und die Modulräume komplexer Struckturen, Publ. Math. Inst. HES, 5 (1960).
[17] H. Grauert, Über Modifikationen und

exzeptionelle analytische Mengen, Math. Ann., 146 (1962), 331-368.
[18] H. Grauert, Der Satz von Kuranishi für kompakte komplexe Räume, Inventiones Math., 25 (1974), 107-142.
[19] F. Hirzebruch, Hilbert modular surfaces, Enseignement Math., 19 (1973), 183-281.
[20] F. Hirzebruch and D. Zagier, Classification of Hilbert modular surfaces, Complex Analysis and Algebraic Geometry in honor of Kodaira, Iwanami Shoten (1977), 137-150.
[21] E. Horikawa, On deformations of holomorphic maps. I, II, J. Math. Soc. Japan, 25 (1973), 372-396; 26 (1974), 647-666. III, Math. Ann., 222 (1976), 275-282.
[22] E. Horikawa, On deformations of quintic surfaces, Inventiones Math., 31 (1975), 43-85.
[23] E. Horikawa, Algebraic surfaces of general type with small C_1^2. I, Ann. Math., (2) 104 (1976), 357-387. II, III, IV, Inventiones Math., 37 (1976), 121-155; 47 (1978), 209-248; 50 (1979), 103-128.
[24] E. Horikawa, Surjectivity of the period map of $K3$ surfaces of degree 2, Math. Ann., 228 (1977), 113-146.
[25] S. Iitaka, On D-dimensions of algebraic varieties, J. Math. Soc. Japan, 23 (1971), 356-373.
[26] M. Inoue, On surfaces of class VII_0, Inventiones Math., 24 (1974), 269-310.
[27] M. Inoue, New surfaces with no meromorphic functions, Proc. Int. Congr. Math., Vancouver, 1974, vol. 1, 423-426.
[28] M. Inoue, New surfaces with no meromorphic functions, Complex Analysis and Algebraic Geometry II, Iwanami Shoten and Cambridge Univ. Press, 1977, 91-106.
[29] M. Kato, Complex manifolds containing global spherical shells, Proc. Int. Symp. Algebraic Geom. Kyoto, 1977, 45-84.
[30] S. Kawai, On compact complex analytic manifolds of complex dimension 3. I, II, J. Math. Soc. Jap., 17 (1965), 438-442; 21 (1969), 604-616.
[31] Y. Kawamata, On deformations of compactifiable complex manifolds, Math. Ann., 235 (1978), 247-265.
[32] Y. Kawamata, Characterization of abelian varieties, Compositio Math., 43 (1981), 253-276.
[33] Y. Kawamata, Kodaira dimension of certain algebraic fiber spaces, J. Fac. Sci. Univ. Tokyo, sec. IA, 30 (1983), I-24.
[34] K. Kodaira, Collected works I-III, Iwanami, 1975.
[35] K. Kodaira and D. C. Spencer, Groups of complex line bundles over compact Kähler varieties, Divisor class groups on algebraic varieties, Proc. Nat. Acad. Sci. US, 39 (1953), 868-877.
[36] K. Kodaira and D. C. Spencer, On deformations of complex analytic structures: I, II, Ann. Math., (2) 67 (1958), 328-466; III, ibid., 71 (1960), 43-76.
[37] K. Kodaira, L. Nirenberg, and D. C. Spencer, On the existence of deformations of complex analytic structures, Ann. Math., (2) 68 (1958), 450-459.
[38] R. S. Kulkarni, On the principle of uniformization, J. Differential Geom., 13 (1978), 109-138.
[39] M. Kuranishi, On the locally complete families of complex analytic structures, Ann. Math., (2) 75 (1962), 536-577.
[40] M. Kuranishi, New proof for the existence of locally complete families of complex structures, Proc. Conf. Complex Analysis, Minneapolis, 1964, Springer, 1965, 142-154.
[41] K. Miyajima, On the existence of Kuranishi family for deformations of holomorphic maps, Sci. Rep. Kagoshima Univ., 27 (1978), 43-76.
[42] Y. Miyaoka, Kähler metrics on elliptic surfaces, Proc. Japan Acad., 50 (1974), 533-536.
[43] Y. Miyaoka, On the Chern numbers of surfaces of general type, Inventiones Math., 42 (1977), 225-237.
[44] B. G. Moishezon, The algebraic analog of compact complex spaces with a sufficiently large field of meromorphic functions I, II, III, Math. USSR-Izv., 3 (1969), 167-226; 305-343; 447-513.
[45] J. Morrow, Minimal normal compactification of $\mathbf{C}^2$, Rice Univ. Studies, 59, no. 1 (1973), 97-112.
[46] I. Nakamura, Inoue-Hirzebruch surfaces and a duality of hyperbolic unimodular singularities, I, Math. Ann., 252 (1980), 221-225.
[47] S. Nakano, On the inverse monoidal transformation, Publ. Res. Inst. Kyoto Univ., 6 (1971), 483-502.
[48] M. Namba, Moduli of open holomorphic maps of compact complex manifolds, Math. Ann., 220 (1976), 65-76.
[49] A. Newlander and L. Nirenberg, Complex analytic coordinates in almost complex manifolds, Ann. Math., (2) 65 (1957), 391-404.
[50] J.-P. Serre, Un théorème de dualité, Comment. Math. Helv., 29 (1955), 9-26.
[51] J.-P. Serre, Géométrie algébrique et géométrie analytique, Ann. Inst. Fourier, 6 (1955-1956), 1-42.
[52] C. L. Siegel, Meromorphe Funktionen auf kompacten analytischen Mannigfaltigkeiten, Nachr. Akad. Wiss. Göttingen, 1955, 71-77.
[53] T. Ueda, Compactification of $\mathbf{C} \times \mathbf{C}^*$ and $(\mathbf{C})^{*2}$, Tohoku Math. J., 31 (1979), 81-90.
[54] K. Ueno, Classification theory of algebraic varieties and compact complex spaces, Lecture notes in math. 439, Springer, 1975.

[55] K. Ueno, Classifacation of algebraic varieties II. Algebraic threefolds of parabolic type, Int. Symp. Algebraic Geometry, Kyoto, 1977, 693–708.
[56] K. Ueno, On algebraic fiber spaces of Abelian varieties, Math. Ann., 237 (1978), 1–22.
[57] E. Viehweg, Canonical divisors and the additivity of the Kodaira dimension for morphisms of relative dimension one, Compositio Math., 35 (1977), 197–223.
[58] E. Viehweg, Klassifikationstheorie algebraischer Varietäten der Dimension Drei, Compositio Math., 41 (1980), 361–400.
[59] K. Yano, Differential geometry on complex and almost complex spaces, Macmillan, 1965.

73 (V.15) Complex Multiplication

A. Classical Theory

If the ratio ω_1/ω_2 of two periods ω_1, ω_2 of an †elliptic function f belongs to an imaginary †quadratic field K, then there exists an algebraic relation between $f(z)$ and $f(\lambda z)$ for any λ in K, and such an f is said to have **complex multiplication**. This phenomenon for the †sn function with modulus $\sqrt{-1}$ was discovered by C. F. Gauss and was applied to the problem of dividing a †lemniscate into five arcs of equal length. More generally, N. H. Abel showed that the *special dividing equation* of an sn function with complex multiplication is algebraically solvable. From a number-theoretic point of view, L. Kronecker conjectured that every †Abelian extension of an imaginary quadratic number field K is determined by a *transform equation* of an elliptic function with complex multiplication by a number of K (1880). This is an analog of the fact, announced by Kronecker and proved by H. Weber, that every Abelian extension of the rational number field is a subfield of a †cyclotomic field. Kronecker's work was continued by Weber [2], and his conjecture was proved by T. Takagi (1903) for $K=\mathbf{Q}(\sqrt{-1})$, by T. Takenouchi (1916) for $K=\mathbf{Q}(e^{2\pi i/3})$, and by Takagi (1920) for the general case using †class field theory. H. Hasse [5] and M. Sugawara simplified the theory of complex multiplication, and Hasse noticed a relationship between complex multiplication and †congruence zeta functions. Working from Hasse's idea, M. Deuring constructed the theory of complex multiplication purely algebraically and determined †Hasse's zeta function of an elliptic curve with complex multiplication.

For the rest of this article, K always stands for an imaginary quadratic field. Let L be a †lattice group on the complex plane $\mathbf{C}$ generated by ω_1, ω_2, and let z be a complex variable. Define functions $\wp, g_2, g_3$ as follows: $\wp(z,L)=\wp(z;\omega_1,\omega_2)=z^{-2}+\sum((z-\omega)^{-2}-\omega^{-2})$, $g_2(L)=g_2(\omega_1,\omega_2)=60\sum\omega^{-4}$, $g_3(L)=g_3(\omega_1,\omega_2)=140\sum\omega^{-6}$, where the sum $\sum$ is over the elements ω of L except 0. Let $\wp'$ be the derivative of $\wp$; then $z\rightarrow(1,\wp(z),\wp'(z))$ is a one-to-one correspondence between the points on the complex torus $\mathbf{C}/L$ and those on the †elliptic curve E; $X_0X_2^2=4X_1^3-g_2X_0^2X_1-g_3X_0^3$ in the projective plane. If the quotient ω_1/ω_2 generates K, then the ring of analytic endomorphisms of $\mathbf{C}/L$ (i.e., the ring of endomorphisms of E) is isomorphic to a subring of the †principal order $\mathfrak{o}$ of K. In particular, if the lattice group L is an †ideal of K (for $K\subset\mathbf{C}$; → 347 Quadratic Fields), then the ring of endomorphisms coincides with $\mathfrak{o}$. The function $J(\tau)=J(E)=J(L)=2^6 3^3 g_2(L)^3/\Delta(L)$ ($\Delta(L)=g_2(L)^3-27g_3(L)^2$) of $\tau=\omega_1/\omega_2$, $\operatorname{Im}\tau>0$, is a †modular function of level 1, and $J(E)$ is called the **invariant** of the elliptic curve E. If E has a complex multiplication, then $J(E)$ is an algebraic integer. Then the three main theorems of the classical theory of complex multiplication can be stated.

Theorem 1. Let h be the †class number of K, and let $\mathfrak{a}_1,\ldots,\mathfrak{a}_h$ be a set of representatives of ideal classes of K. Then $J(\mathfrak{a}_1),\ldots,J(\mathfrak{a}_h)$ are exactly the conjugates of $J(\mathfrak{a}_1)$ over K, and $K(J(\mathfrak{a}_1))$ is the maximal unramified Abelian extension (the †absolute class field) of K (→ 59 Class Field Theory).

Next we define the function f by

$$\begin{aligned} f(z;L)&=g_2g_3\Delta^{-1}\cdot\wp(z;L), && K\neq\mathbf{Q}(\sqrt{-1}),\ \mathbf{Q}(e^{2\pi i/3}),\\ &=g_2^2\Delta^{-1}\cdot\wp(z;L)^2, && K=\mathbf{Q}(\sqrt{-1}),\\ &=g_3\Delta^{-1}\cdot\wp(z;L)^3, && K=\mathbf{Q}(e^{2\pi i/3}). \end{aligned}$$

Theorem 2. Let $\mathfrak{o}$ be the †principal order of K, $\mathfrak{m}$ be an integral ideal of K, and $\mathfrak{a}$ be an arbitrary ideal of K. Choose a number ξ of K such that $\mathfrak{m}=\{\lambda\in\mathfrak{o}\mid\lambda\xi\in\mathfrak{a}\}$. Then $K(J(\mathfrak{a}),f(\xi;\mathfrak{a}))$ is the †class field for the †ray modulo $\mathfrak{m}$.

The number ξ in this theorem is obtained by $\mathfrak{a}^{-1}\mathfrak{m}(\xi)=\mathfrak{b}$, where $\mathfrak{b}$ is an integral ideal belonging to the ideal class of $\mathfrak{a}^{-1}\mathfrak{m}$. Thus $f(\xi;\mathfrak{a})$ is a "special value" of an elliptic function as well as of a modular function.

The †general law of reciprocity, the †principal ideal theorem, and the ramification in the class field in theorem 2 can also be described

in terms of elliptic functions or elliptic curves. In general, the ring of endomorphisms of an elliptic curve defined over a field k of characteristic 0 is either **Z** or an order of an imaginary quadratic field. On the other hand, if the characteristic of k is not zero, then the ring of endomorphisms may be an order in a definite †quaternion algebra.

B. Complex Multiplication of an Abelian Variety

Following Kronecker's idea, D. Hilbert posed in his lecture at Paris (1900) the so-called 12th problem: to find an analytic function whose special values generate Abelian extensions over a given algebraic number field (→ 196 Hilbert). E. Hecke constructed unramified Abelian extensions of an imaginary biquadratic field using †Hilbert modular functions [10]. After this, there was no notable development concerning the problem until the theory of complex multiplication was generalized to the case of †Abelian varieties, which was made possible by progress in algebraic geometry, in particular A. Weil's geometric theory of Abelian varieties (G. Shimura and Y. Taniyama [11]). The following results have been obtained: Consider a triple $(A, \mathfrak{X}, t)$, consisting of an Abelian variety A defined over **C**, a †polarization $\mathfrak{X}$ of A, and a point t of A. Call two such triples $(A, \mathfrak{X}, t)$ and $(A', \mathfrak{X}', t')$ isomorphic if an isomorphism of A onto A' maps $\mathfrak{X}$ onto $\mathfrak{X}'$ and t onto t'. Then there exists one and only one subfield k_0 of **C** with the following property: In order for $(A, \mathfrak{X}, t)$ and $(A^\sigma, \mathfrak{X}^\sigma, t^\sigma)$ to be isomorphic for an automorphism σ of **C**, it is necessary and sufficient that σ fix all elements of k_0. We call k_0 the **field of moduli** of $(A, \mathfrak{X}, t)$. If A is an elliptic curve E and if $t=0$, then $k_0=\mathbf{Q}(J(E))$. In the higher-dimensional case, the field of moduli is generated by special values of a †Siegel modular function.

If F is a totally imaginary number field that is a quadratic extension of a totally real field F_0 of degree n, then there exists a set $\{\varphi_1, \ldots, \varphi_n\}$ of n different isomorphisms of F into **C** such that $\bar{\varphi}_i=\varphi_j$ never occurs for $i\neq j$, where the bar denotes complex conjugation. We call $(F; \{\varphi_\lambda\})$ a **CM-type**.

If $\mathfrak{a}$ is an ideal of F, then $L=\{(\varphi_1(\alpha), \ldots, \varphi_n(\alpha))\in\mathbf{C}^n \mid \alpha\in\mathfrak{a}\}$ is a lattice group in the n-dimensional complex linear space $\mathbf{C}^n$, and $\mathbf{C}^n/L$ is analytically isomorphic to an Abelian variety A of dimension n in a complex projective space. The †endomorphism ring $\mathfrak{A}(A)$ of A contains a ring that is isomorphic to the principal order $\mathfrak{o}$ of F. Conversely, every Abelian variety of dimension n such that $\mathfrak{A}(A)$ contains a ring isomorphic to $\mathfrak{o}$ can be constructed in this way.

Let M be a normal extension of **Q** containing F. Denote the Galois group of M by G and the subgroup of G corresponding to F by H, and put $S=\bigcup_\lambda \varphi_\lambda H$, where φ_λ stands for a prolongation of the previous φ_λ to G. Then the following three conditions are equivalent: (i) $\mathfrak{A}(A)\cong\mathfrak{o}$; (ii) A is simple; (iii) $H=\{\gamma\in G \mid S\gamma=S\}$. If $H^*=\{\delta\in G \mid \delta S=S\}$, then we can choose $\psi_n\in G$ such that $S=\bigcup_\mu H^*\psi_\mu$. If F^* is the subfield of M corresponding to H^*, then $(F^*; \{\psi_\mu\})$ is also a CM-type, and $F^*=\mathbf{Q}(\sum_\lambda \varphi_\lambda(\alpha) \mid \alpha\in F)$. Moreover, for an ideal $\mathfrak{r}$ of F^*, $\prod_\mu \psi_\mu(\mathfrak{r})$ is an ideal of F.

Now let $H_\mathfrak{m}$ be the group of ideals $\mathfrak{r}$ of F^* that are relatively prime to the norm $N(\mathfrak{m})$ of an integral ideal $\mathfrak{m}$ of F^* and for which there exists a number ξ of F with

$$\prod_\mu \psi_\mu(\mathfrak{r})=(\xi), \quad N(\mathfrak{r})=|\xi|^2, \quad \xi\equiv 1 \ (\mathrm{mod}^\times \mathfrak{m}).$$

Then $H_\mathfrak{m}$ is an †ideal group modulo $\mathfrak{m}$ in F^*.

Theorem 3. Assume that $\mathfrak{A}(A)\cong\mathfrak{o}$, and let $\mathfrak{X}$ be an arbitrary polarization of A. Denote a point of A with $\mathfrak{m}=\{\lambda\in\mathfrak{o} \mid \lambda t=0\}$ by t, and let $k_\mathfrak{m}$ be the field of moduli of $(A, \mathfrak{X}, t)$. Then $k_\mathfrak{m}F^*$ is the †class field over F^* corresponding to $H_\mathfrak{m}$.

The point t in theorem 3 always exists. In particular, if $\mathfrak{m}=\mathfrak{o}$, then $t=0$. When A is an elliptic curve E, F is an imaginary quadratic field and we have $F^*=K$, and theorem 3 coincides essentially with the content of theorems 1 and 2. If $n>1$, $F=F^*$ holds only in special cases.

The theory of complex multiplication of A is closely related to the †Hasse zeta function of A (→ 450 Zeta Functions).

References

[1] L. Kronecker, Zur Theorie der elliptischen Funktionen, Monatber. Königlich Preuss. Akad. Wiss. Berlin, (1881), 1165–1172. (Werke IV, Teubner, 1929, 309–318; Chelsea, 1969.)

[2] H. Weber, Lehrbuch der Algebra III, Vieweg, second edition, 1908 (Chelsea, 1961).

[3] T. Takagi, Über eine Theorie des relativ-Abelschen Zahlkörpers, J. Coll. Sci. Imp. Univ. Tokyo, 41 (1920), 1–132.

[4] R. Fueter, Vorlesungen über die singulären Moduln und die komplexe Multiplikation der elliptischen Funktionen, Teubner, I, 1924; II, 1927.

[5] H. Hasse, Neue Begründung der komplexen Multiplikation, J. Reine Angew. Math., I, 157 (1926), 115–139; II, 165 (1931), 64–88.

[6] R. Fricke, Lehrbuch der Algebra III, Vieweg, 1928.
[7] M. Deuring, Die Typen der Multiplikatorenringe elliptischer Funktionenkörper, Abh. Math. Sem. Univ. Hamburg, 14 (1941), 197–272.
[8] M. Deuring, Die Struktur der elliptischen Funktionenkörper und die Klassenkörper der imaginären quadratischen Zahlkörper, Math. Ann., 124 (1952), 393–426.
[9] M. Deuring, Die Klassenkörper der komplexen Multiplikation, Enzykl. Math., Bd. I_2, Heft 10, Teil II, 23, 1958.
[10] E. Hecke, Höhere Modulfunktionen und ihre anwendung auf die Zahlentheorie, Math. Ann., 71 (1912), 1–37; Über die Konstruktion relativ-Abelscher Zahlkörper durch Modulfunktionen von zwei Variabeln, Math. Ann., 74 (1913), 465–510. (Mathematische Werke, Vandenhoeck & Ruprecht, 1959, 21–58, 69–114.)
[11] G. Shimura and Y. Taniyama, Complex multiplication of Abelian varieties and its applications to number theory, Publ. Math. Soc. Japan, 1961.
[12] G. Shimura, Construction of class fields and zeta functions of algebraic curves, Ann. Math., (2) 85 (1967), 58–159.
[13] A. Borel et al., Seminar on complex multiplication, Lecture notes in math. 21, Springer, 1966.
[14] G. Shimura, Introduction to the arithmetic theory of automorphic functions, Publ. Math. Soc. Japan and Princeton Univ. Press, 1971.

74 (II.11) Complex Numbers

A. Algebraic Properties of Complex Numbers

A **complex number** is an expression of the form $a+ib$ with arbitrary real numbers a and b and the **imaginary unit** i. Writing $\alpha=a+ib$, $\beta=c+id$, we define $\alpha=\beta$ if and only if $a=c$ and $b=d$. As regards algebraic operations with complex numbers, we define $\alpha+\beta=(a+c)+i(b+d)$, $\alpha-\beta=(a-c)+i(b-d)$, $\alpha\beta=(ac-bd)+i(ad+bc)$, and for $\beta\neq 0$, i.e., $c^2+d^2\neq 0$, $\alpha/\beta=(ac+bd)/(c^2+d^2)+i((bc-ad)/(c^2+d^2))$. Then the addition and multiplication thus defined obey commutative, associative, and distributive laws, and complex numbers form a †commutative field with $0=0+i0$ and $1=1+i0$ as its zero element of addition and identity element of multiplication, respectively. The set of all complex numbers is usually denoted by **C**.

By assigning to each real number a a complex number $a+i0$, algebraic operations on real numbers are carried into those of the corresponding complex numbers. That is to say, the field **R** of all real numbers is mapped isomorphically into the field **C** of all complex numbers. By identifying a with $a+i0$, we are taking **R** as a subfield of **C**. Also, $0+i1$ will be denoted simply by i. From the previous definition of algebraic operations, it follows that $i^2=-1$. Furthermore, since $\alpha=a+ib=(a+i0)+(b+i0)(0+i1)$, $a+ib$ is not a mere symbolic expression but can also be regarded as the outcome of algebraic operations on a, b, i in **C**. The **real** and **imaginary parts** of a complex number $\alpha=a+ib$ are, by definition, a and b, denoted by $\mathrm{Re}\,\alpha$ and $\mathrm{Im}\,\alpha$, respectively. A complex number that is not a real number is sometimes called an **imaginary number**; in particular, a complex number α with $\mathrm{Re}\,\alpha=0$ is called a **purely imaginary number**. For each complex number $\alpha=a+ib$, we define its **conjugate complex number** as $a-ib$, and denote it by $\bar{\alpha}$. We then have $\overline{\alpha+\beta}=\bar{\alpha}+\bar{\beta}$ and $\overline{\alpha\beta}=\bar{\alpha}\bar{\beta}$. The mapping $\alpha\rightarrow\bar{\alpha}$ is an †automorphism of **C** which leaves each element of **R** invariant. Also, the following relations hold: $\mathrm{Re}\,\alpha=(\alpha+\bar{\alpha})/2$ and $\mathrm{Im}\,\alpha=(\alpha-\bar{\alpha})/(2i)$.

Regarded as an overfield of **R**, **C** is an †extension field of **R** of degree 2 obtained by the adjunction of i, which is a root of an irreducible equation $x^2+1=0$. The important algebraic property of **C** is that it is †algebraically closed. Namely, for any polynomial $f(x)$ with coefficients in **C**, the equation $f(x)=0$ possesses at least one root in **C** (†Gauss's fundamental theorem of algebra).

B. Topology of C

The **absolute value** (or **modulus**) of a complex number $\alpha=a+ib$, denoted by $|\alpha|$, is by definition $|\alpha|=\sqrt{a^2+b^2}=\sqrt{\alpha\bar{\alpha}}$. If α is real, then the absolute value of α in the sense of complex numbers is identical to the one in the sense of real numbers. It always holds that $|\alpha|\geqslant 0$ and $|\alpha|=0\Leftrightarrow\alpha=0$. It follows further that $|\alpha+\beta|\leqslant|\alpha|+|\beta|$, $|\alpha\beta|=|\alpha||\beta|$, and $|\alpha|=|\bar{\alpha}|$.

For each pair of complex numbers α and β, define $\rho(\alpha,\beta)=|\alpha-\beta|$. Then with $\rho(\alpha,\beta)$ as the †distance function, **C** satisfies the axioms for a †metric space, and in particular, $\lim_{n\to\infty}\alpha_n=\alpha_0\Leftrightarrow\lim\rho(\alpha_n,\alpha_0)=0\Leftrightarrow\lim|\alpha_n-\alpha_0|=0\Leftrightarrow$ ($\lim a_n=a_0$ and $\lim b_n=b_0$) (where $\alpha_n=a_n+ib_n$, $a_0=a_0+ib_0$). From this it is easily seen, as in the case of the set **R** of all real numbers, that **C** also becomes a †locally compact and †complete metric space.

With respect to this topology, the four oper-

ations (except for division by zero) are continuous: $\alpha_n \to \alpha_0$ and $\beta_n \to \beta_0$ imply $\alpha_n + \beta_n \to \alpha_0 + \beta_0$; $\alpha_n - \beta_n \to \alpha_0 - \beta_0$; $\alpha_n \beta_n \to \alpha_0 \beta_0$; and $\alpha_n / \beta_n \to \alpha_0 / \beta_0$ (where in the last case we assume $\beta_n \neq 0$ and $\beta_0 \neq 0$). Thus **C** becomes a topological field. Furthermore, the assignment $\alpha \to \bar{\alpha}$ gives a continuous mapping $\mathbf{C} \to \mathbf{C}$, a homeomorphic automorphism of **C**.

C. The Complex Plane

If in a plane to which is assigned rectangular coordinate axes a complex number $\alpha = a + ib$ is represented by a point (a, b), then the plane is called the **complex (number) plane (Gauss-Argand plane** or **Gaussian plane)** (Fig. 1), and the point representing α is called simply the point α. The abscissa and ordinate axes are called the **real** and **imaginary axes**, respectively. A point $\alpha = a + ib$ can be represented by †polar coordinates r, θ with the origin and the real axis as the pole and the generating line, respectively, where $r = \sqrt{a^2 + b^2}$ is the absolute value $|\alpha|$ of α and θ is the **argument** (or **amplitude**), denoted by $\arg \alpha$, of α. The argument of α is uniquely determined mod 2π if $\alpha \neq 0$ and is an arbitrary real number if $\alpha = 0$.

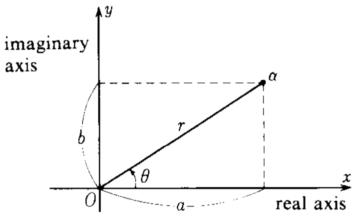


Fig. 1

The absolute value $|\alpha|$ of a complex number α, regarded as a vector from the origin to the point α, is the length of this vector. For complex numbers α and β, to the sum of the vectors α and β corresponds the sum $\alpha + \beta$ of the complex numbers. A complex number α, in terms of its absolute value r and argument θ, is expressed as $\alpha = r(\cos\theta + i\sin\theta)$, which is called the **polar form** of α. For polar forms the following hold: $\bar{\alpha} = r(\cos\theta - i\sin\theta) = r(\cos(-\theta) + i\sin(-\theta))$; $\alpha^{-1} = \bar{\alpha}/|\alpha|^2 = r^{-1}(\cos(-\theta) + i\sin(-\theta))$ $(\alpha \neq 0)$; and $\alpha_1\alpha_2 = r_1 r_2(\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2))$, where $|\alpha_j| = r_j$ and $\arg\alpha_j = \theta_j$ for $j = 1, 2$. This last relation when $r_1 = r_2 = 1$ is called **De Moivre's formula**. The nth roots of unity in **C** are given by $\rho_j = \cos 2\pi j/n + i \sin 2\pi j/n$ $(j = 0, 1, \ldots, n-1)$ (Fig. 2).

In the complex plane, the mapping $\alpha \to \bar{\alpha}$ corresponds to the †reflection of the plane in the real axis, $\alpha \to \alpha + \beta$ to the parallel †translation along a vector β, $\alpha \to \alpha\beta$ $(\beta \neq 0)$ to the †rotation through the angle $\arg\beta$ followed by a †homothetic transformation with center 0 and constant ratio $|\beta|$, and $\alpha \to \bar{\alpha}^{-1}$ to the †inversion with respect to the **unit circle** $\{\alpha \mid |\alpha| = 1\}$.

The distance $\rho(\alpha, \beta) = |\alpha - \beta|$ between α and β in **C** coincides with their Euclidean distance, provided that α and β are regarded as points in the Euclidean plane, so that the complex plane is †isometric, and accordingly homeomorphic, to the Euclidean plane.

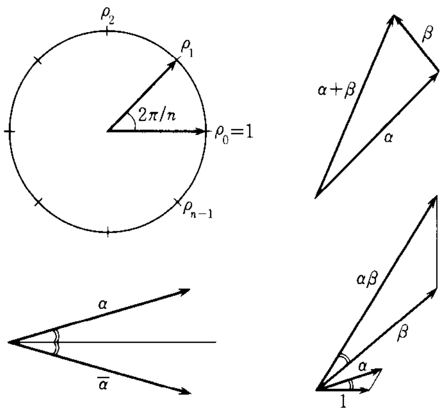


Fig. 2

D. The Complex Sphere

In the rest of this article, P denotes the complex plane and Σ denotes the sphere of radius 1, with 0, the origin of P, as its center. The points $N(0, 0, 1)$ and $S(0, 0, -1)$ of Σ will be called the **north** and **south pole**, respectively (Fig. 3), where the 1st and 2nd coordinate axes are the real and imaginary axes of P, respectively, and the 3rd coordinate axis is orthogonal to P. A straight line from N through a point z (a complex number) in P intersects Σ at a point $Z = (x_1, x_2, x_3)$ different from N, where $z = (x_1 + ix_2)/(1 - x_3)$, $x_1 = (z + \bar{z})/(1 + |z|^2)$, $x_2 = (z - \bar{z})/(i(1 + |z|^2))$, and $x_3 = (|z|^2 - 1)/(|z|^2 + 1)$. The mapping $z \to Z$ is called a **stereographic projection** from N, by means of which P and $\Sigma - \{N\}$ become †conformally equivalent to each other. Consequently z can be represented by a point Z of $\Sigma - \{N\}$, and Σ thus used is called a **complex sphere** (or **Riemann sphere**). Let us adjoin to the complex plane P a new element, denoted by ∞, called the **point at infinity**, which corresponds to the only exceptional point N of Σ. The topology of the complex plane with ∞ can be introduced by the corresponding topology of the Riemann sphere. Indeed, the family of all the sets $\{z \mid |z| > M\} \cup \{\infty\}$ for $M > 0$ forms a †local base around ∞. By introducing local complex coordinates $\zeta = 1/z$ into the neighborhoods of ∞, each element of this local base is represented as $\{\zeta \mid |\zeta| < M^{-1}\}$, in which the

convention $\zeta=0$ is adopted for $z=\infty$. The complex sphere thus defined can be regarded as a †Riemann surface (i.e., a 1-dimensional †complex manifold).

The complex plane (complex sphere) whose points are represented by a variable z or w is called a z-**plane** or a w-**plane** (a z-**sphere** or a w-**sphere**).

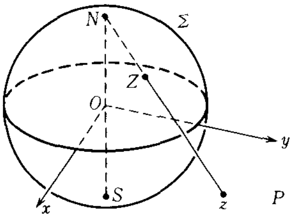


Fig. 3

E. Linear Fractional Functions

Given complex numbers a, b, c, and d with $ad-bc\neq 0$, we define a **linear fractional function** (or simply **linear function**) by

$$w=\frac{az+b}{cz+d}. \tag{1}$$

As a mapping from the z-sphere into the w-sphere, this linear function is called a **Möbius transformation (linear fractional** or simply **linear transformation)**. The usual linear transformation, i.e., the one with $c=0$ in the present case, is sometimes distinguished as an **entire linear transformation**. Since (1) depends only on the proportion $a:b:c:d$, we can assume $ad-bc=1$ without loss of generality.

The transformation (1) is †holomorphic and †univalent on the whole z-sphere with only one exceptional †pole at $-d/c$ (∞ if $c=0$) of order 1, and the inverse of (1) is also a linear fractional function. The set of all linear transformations forms a †group with composition of transformations as the group operation. One of its subgroups is the †modular group.

Linear transformations carry any circle of the complex plane (or of the Riemann sphere) into a circle of the same plane (or of the same sphere) if we adopt the convention that straight lines are a special kind of circle. (In the case of a Riemann sphere, no such convention is necessary.) Given on a plane a circle with center o and radius r and two points p and p' on a half-line issuing from o satisfying $op\cdot op'=r^2$, the points p and p' are called **symmetric points** (or **reflection points**) with respect to the circle. The transformation $p\rightarrow p'$ is called the **inversion** with respect to this circle. In the complex plane, let z and z' be symmetric points with respect to a circle C. Suppose that by a linear transformation, z, z' and C are carried to points w, w' and a circle D, respectively; then w and w' become symmetric with respect to the circle D (**principle of reflection**). Thus symmetricity is invariant under linear transformations. Also the †anharmonic ratio of any four points z_1, z_2, z_3, and z_4, $(z_1,z_2;z_3,z_4)=(z_1-z_3)/(z_1-z_4):(z_2-z_3)/(z_2-z_4)$, is invariant under a linear transformation; i.e., $(z_1,z_2;z_3,z_4)=(w_1,w_2;w_3,w_4)$ holds, where w_j is the image of z_j under a linear transformation ($j=1,2,3,4$).

F. Normal Forms of Linear Transformations

There exist fixed points of the transformation (1) on the Riemann sphere, i.e., points satisfying $z=(az+b)/(cz+d)$. The number of fixed points is 2 or 1, except when $w=z$. If the transformation has two fixed points, they will be denoted here by p and q. The natural convention $p=q$ is adopted if the transformation has one fixed point. If $c=0$, then p or q is ∞, and furthermore if $c=a-d=0$, then p and q are both ∞.

For unequal finite p and q, (1) can be rewritten in the following normal form:

$$\frac{w-p}{w-p}=\alpha\frac{z-p}{z-q}, \qquad \alpha=\frac{a-cp}{a-cq}\neq 1,$$

in which, according as $\arg\alpha=0$, $|\alpha|=1$, or otherwise, (1) is called a **hyperbolic** (Fig. 4), **elliptic** (Fig. 5), or **loxodromic transformation**, respectively. This classification can be applied also for finite p and infinite q, i.e., to the transformation $w-p=\alpha(z-p)$. Furthermore, for $p=q\neq\infty$, (1) is rewritten in the following form:

$$\frac{1}{w-p}=\frac{1}{z-p}+\beta, \qquad \beta=\frac{c}{a-cp}.$$

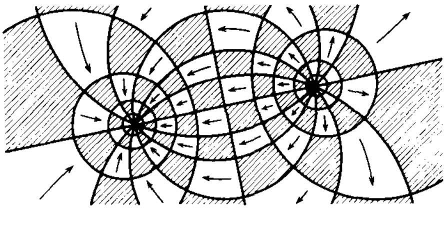


Fig. 4
Hyperbolic transformation.

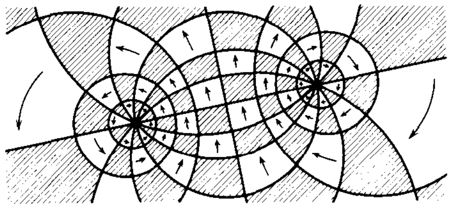


Fig. 5
Elliptic transformation.

In this case (1) is called a **parabolic transformation** (Fig. 6). For $p=q=\infty$, i.e., if $w=z+\beta$, (1) is also called parabolic. We can easily determine to which class (1) belongs from the discriminant $D=(a+d)^2-4$ of the quadratic equation $cz^2-(a-d)z-b=0$ obtained from $z=(az+b)/(cz+d)$ with $ad-bc=1$ by multiplying both sides by $cz+d$. If $a+d$ is real, then according as $D>0$, <0, or $=0$, (1) is hyperbolic, elliptic, or parabolic, respectively, and if $a+d$ is not real, then the transformation is loxodromic.

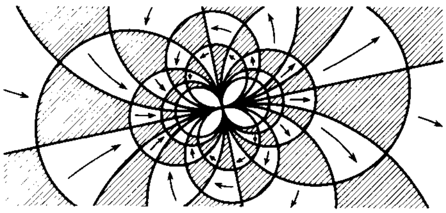


Fig. 6
Parabolic transformation.

Let D and D' be two arbitrary circular disks. Then there always exists a linear transformation which gives a one-to-one †conformal mapping from D onto D'. Conversely, any mapping with this property is given only by linear transformations (provided that the half-plane having a straight line together with the point at infinity as its boundary is regarded as a closed disk), which are uniquely determined by giving three points a, b, c from the boundary of D and as their corresponding points, three arbitrary points a', b', c' from the boundary of D'.

G. The Poincaré Metric

Since conformal mappings from the domain $|z|<1$ onto $|w|<1$ are given by the transformations $w=\varepsilon(z-z_0)/(1-\bar{z}_0z)$ ($|\varepsilon|=1$, $|z_0|<1$) (→ Appendix A, Table 13), for corresponding z and w it holds that

$$\frac{|dw|}{1-|w|^2}=\frac{|dz|}{1-|z|^2}. \qquad (2)$$

$|dz|/(1-|z|^2)$ is called **Poincaré's differential invariant**. With a metric having $ds=|dz|/(1-|z|^2)$ as its †line element, the unit disk $|z|<1$ becomes a †non-Euclidean space in the sense of Lobachevskiĭ, and the metric is called the **Poincaré metric**. Furthermore, since the transformations (2) leave the length of curves invariant, they can be regarded as †motions in this space, where the †geodesic through two points z_1 and z_2 is the circular arc orthogonal to the unit circle. If we denote the intersections of the arc with the unit circle by z_3 and z_4, then the †non-Euclidean distance between two points z_1 and z_2 along the geodesic is given by $(1/2)\log(z_1, z_2; z_3, z_4)$, provided that the points z_4, z_1, z_2, z_3 are arranged on the arc in this order (→ 285 Non-Euclidean Geometry).

References

[1] C. F. Gauss, Werke I–XII, Göttingen, 1863–1931.
[2] N. Bourbaki, Eléments de mathématique III. Topologie générale, ch. 8. Nombres complexes, Actualités Sci. Ind., 1235b, Hermann, third edition, 1963; English translation, General topology, pt. 2, Addison-Wesley, 1966.
[3] L. V. Ahlfors, Complex analysis, McGraw-Hill, second edition, 1966.
[4] C. Carathéodory, Funktionentheorie I, Birkhäuser, 1950; English translation, Theory of functions of a complex variable I, Chelsea, 1954.

75 (XVI.2) Computers

A. History

Since the beginning of civilization, people have utilized tools for aiding computation. In Japan, bamboo computing rods were used in the 7th century; before the close of the 16th century, the **abacus** was imported from China (→ 230 Japanese Mathematics (Wasan)). The first **calculator** capable of performing the four arithmetic operations automatically was designed by W. Schickard of Tübingen University (1623). After that, B. Pascal independently made his famous adding machine; this was improved by G. W. Leibniz so that it was able to execute multiplications and divisions.

Most calculators manufactored today are electronic; however, many of these machines are not "programmable" and hence require at every step of computation a manual operation for specifying the machine operations to be executed. On the other hand, modern automatic computers can execute automatically a sequence of operations according to a given program without any manual intervention.

The automatic computer was conceived by the English mathematician C. Babbage in the 19th century, but mechanical engineering at that time was not advanced enough to allow the construction of such a computer. His idea was first realized by the relay computers Z3 of K. Zuse (1941) and Mark I of Harvard University (1944). In 1947, the first **electronic computer**, ENIAC, appeared, in which vacuum

tubes were utilized instead of mechanical components. Since then, the capabilities of computers have increased rapidly and it can be said that we now live in the "computer age."

B. Principles of Modern Computers

Information processing in a computer is based on communication among its constituents by electric signals. In **digital computers**, information is encoded as a sequence of binary numbers 0, 1, whereas continuous values are allowed in †analog computers. The minimum quantity of information in digital computers is therefore a binary digit, called a **bit**. Since continuous values fluctuate on account of electric noise in the circuits, digital computers have the advantage of maintaining high precision during computation.

The binary numbers 0, 1 are represented in practice by two distinct electric signals: two distinct voltages, two distinct phases of alternating current, the existence or nonexistence of a pulse, etc. A system of circuits is called **synchronous** when it contains a clock, a generator of periodic pulses, which synchronizes the transmission and transformation of information. In an **asynchronous** system, circuits execute each step of information processing independently and advance to the next step after verifying the termination of the preceding step.

In ENIAC, a program was inserted into the computer by carefully connecting many control lines on plugboards. Since this caused much trouble in inserting the different programs needed for different jobs, J. von Neumann proposed encoding programs in sequences of binary digits and storing these programs in the memory unit of the computer. His principle, the **stored program principle**, was realized by EDSAC (1949) and has since been widely used.

An automatic computer in general consists of the following five units: the arithmetic unit, memory, control, input, and output (Fig. 1). These units are interconnected by wires that exchange information in the course of computation.

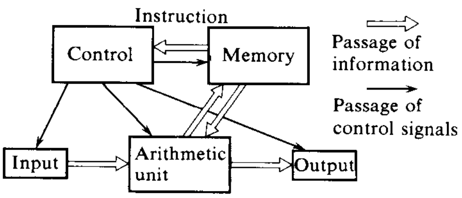


Fig. 1
Construction of electronic computers.

The **arithmetic unit** consists mainly of several memory registers and operational circuits associated with them. Each register stores a binary number of n bits (usually $16 \leqslant n \leqslant 64$). An important building block of the operational circuit is the basic adder of 1-bit numbers. A parallel adder of n-bit numbers can be obtained by connecting n copies of the basic adder. Alternatively, a sequential adder can be composed of a single basic adder which is utilized repeatedly to sum up bit by bit two binary numbers from the lowest bit, Subtraction is usually carried out by adding the complement of the subtrahend. Multiplication is realized by shifting the multiplicand to the left and adding it to the intermediate sum; this addition may be omitted, depending on the relevent bit of the multiplier. In division, we shift the dividend to the left and if possible subtract the divisor from it. By recording at each step whether or not the subtraction is possible, we obtain the quotient.

The **memory unit** stores the instructions and given data as well as the necessary data obtained in the course of computation. It is divided into many registers, each of which contains a number and is referred to by means of a serial number called an **address**. At present, high-speed memory units are usually made up of magnetic cores or integrated circuits. As auxiliary large-scale memory, there are also magnetic drums, disks, and magnetic tapes, etc. Other elements under investigation are extreme-low-temperature elements, chemical elements, optical elements, etc.

The **control unit** repeats the following operations consecutively: (1) takes an instruction from the memory location indicated by the sequential control counter, (2) gets a data word from the memory according to the address part of the instruction, (3) decodes the function part of the instruction and sends control signals to appropriate circuits, and (4) increases the content of the sequential control counter by 1 and, after receiving *end* signals from the arithmetic unit, returns to step (1). For these purposes the control unit contains a counter, a decoder, an encoder to send control signals, and a register to store the instruction to be executed.

The arithmetic, control, and memory units form the **central processor** of the computer. In contrast to the central processor, the input and output units are called **peripheral devices**. An input device receives the necessary information (a program and data); a card reader reads punched cards, and a teletypewriter sends signals directly into the computer. An output device presents the results obtained by the computer. The results are usually printed by a teletypewriter or line printer. Magnetic tape

units can be considered as auxiliary input-output devices, since tape reels are removable. The tape units can accept prepared input data as well as record the results, the output.

These devices are operated according to given input-output instructions under the control of the central processor. However, since every input-output device contains mechanical components and is extremely slow in comparison with the central processor, a large-scale computer is often accompanied by satellite computers which undertake the control of input-output devices.

C. Instructions and Programming

In stored-program operation, a computer performs a sequence of calculations according to given instructions. There are various types of instructions, but those most frequently used are **single-address instructions**, each of which contains a single-address part designating an operand.

A **program** is a finite sequence of instructions arranged suitably for the required computation. **Programming**, or making a program, is therefore the task of decomposing the required computation into elementary steps each of which corresponds to an instruction.

Every instruction is represented in a computer by a number, a numeric code, which is determined in a definite way by the construction of the control unit. Before starting computation, instructions thus encoded are stored in the memory. In this sense, a program is a sequence of numbers. This sequence is called a **machine-language program**.

A program is usually divided into several blocks, called subprograms or **subroutines**. Some subroutines are made in advance, especially those for frequently required jobs such as evaluating elementary functions and manipulating input-output devices, etc. The system of these ready-made programs is called **software**, in contrast to the **hardware** (i.e., mechanico-electronic equipment) of the computer. Quite often basic routines are **microprogrammed**, i.e., written in a simple code and stored in a fast read-only memory (ROM). In this case the set of built-in programs is called **firmware**. The handliness of a computer depends mainly on close matching of the software and hardware.

Programs are usually written in certain forms called **external languages** which are easy to master. A problem-oriented language is an advanced external language in which ordinary arithmetic expressions are available with slight modifications. Programs written in these languages are translated into machine languages by program input routines. The translator for a problem-oriented language is called the **compiler**.

A compiler accepts several macroinstructions. Moreover, it is equipped with the following facilities: (i) the ability to translate arithmetic expressions into machine language; (ii) the ability to generate linkages to various ready-made subroutines according to certain simple indications; (iii) the ability to automatically allocate programs, subroutines, and data in the memory; (iv) the ability to check automatically the syntactic correctness of programs. Thus it accepts an external form such as

if $x \geqslant 0$ **then** printreal (SQRT(x))

else printstring ('negative').

Another important translator is the **assembler**, which translates mnemonic codes of instructions (*add* for addition, etc.) into their numeric codes according to a given table. It allows us to utilize symbols for specifying addresses. It also converts decimal numbers into binary and generates certain segments of the program from rather simple indications. Compilers and assemblers are important constituents of software.

The large-scale high-speed computers that have recently appeared have made software systems inevitably more complex. There is now software, so-called monitors or **operating systems**, that supervises the uninterrupted processing of many programs. Some operating systems coordinate several assemblers and compilers so that several languages can be mixed in writing a program. Examples of other important software are mathematical (numerical) software and database management systems.

D. Mathematical Models of Computers

An operational circuit in the arithmetic unit is usually constructed from basic elements (logical gates), each of which performs a certain operation on the binary signals. Thus the construction of a circuit from basic elements is represented by a composition of a logical function defined over the set $\{0,1\}$ from a given set of basic functions. Post [3] considered functional composition without feedback loops and established a general criterion for a given set of logical functions to be complete, i.e., to be capable of generating the whole set of logical functions. Strictly, however, a logical gate takes a definite time delay to perform its operation. Therefore Kudryavtsev [4] proposed as a model of a logical gate a pair (f,d) of a logical function f and a non-

negative integer d: f represents the operation of the gate and d represents its delay. He defined the feedback-free composition of such functions with delays and gave a completeness criterion for a set of logical functions with delays. His study has been extended in various directions by Loomis [5], Nozaki [6], and Rosenberg [7].

When a circuit contains memory elements or feedback loops, its function is suitably represented by an †automaton. Hence the design and analysis of circuits with memory elements, state-minimization, and equivalence checking [8], decomposition into simpler components [9, 10], verification of completeness [11], etc., can be studied in terms of automata (→ 31 Automata). A computer itself can be considered to be a finite automaton, since it has a finite number of memory elements and its behavior is completely determined by the content of the memory (its internal state) and the inputs. However, since a modern computer has an enormous memory and contains replaceable parts, such as magnetic tapes, it is more adequately represented by an infinite model, such as a †Turing machine.

Turing machines are capable of simulating many intellectual activities governed by formal rules. It is believed that any well-defined finite algorithm can be simulated by a Turing machine (Church's thesis.) Thus any computer can be simulated by a Turing machine, and hence it is unable to resolve those decision problems which are unsolvable for Turing machines. For instance, no program can decide in finite steps whether a given program written in a computer language, say FORTRAN or PASCAL, eventually stops or not. On the other hand, a modern computer can simulate a universal Turing machine provided that its memory can be extended unboundedly by supplying magnetic tapes. Hence any well-defined finite algorithm can be simulated by a computer unless the limitation of memory capacity hampers its accomplishment.

E. Mathematical Theory of Programming Languages

Along with the development of software, programming techniques have gradually accumulated. For instance, we now have an almost satisfactory method of translating arithmetic expressions into machine language. However, we still lack a general theory to cover effectively a wide array of programming problems. The design of an adequate metalanguage is still an important problem, if the word "adequate" implies complete description of syntax and semantics of problem-oriented languages. An interesting problem is to define and suitably classify grammars with respect to their capabilities of forming languages. Such research is an important branch of **mathematical linguistics**. N. Chomsky [12] has investigated this problem, starting from research on natural languages, and has given formal definitions of the grammars (→ 31 Automata D).

†Context-free grammars (type-2 languages) play an important role in the theory of software as well as in Chomsky's syntax, since they are powerful enough to describe the parenthesis structure and simple and easy to manipulate. Some variants of context-free grammar have been proposed for attaining high efficiency in the automatic appraisal of programs [13, 14] or for increasing the programs' generative power [15].

F. Branches of Information Science

Many fields related to computers now make up the **information sciences** or **informatiques**. Important mathematical theories born or developed in the information sciences are:
(1) Design and analysis of hardware devices: †Boolean algebra [16], switching theory [17], and theory of †automata [18, 19].
(2) Design and analysis of programming language; theory of formal languages [14, 21].
(3) Design and analysis of algorithms; †numerical analysis, theory of †complexity of computation, theory of †data processing.
(4) Mathematical foundation of programming; logical verification of correctness and equivalence of programs [22], †recursive function theory, †decision problems.

Applications of the information sciences are found in diversified fields, such as statistics, operations research, mathematical psychology, econometrics, jurimetrics, and behaviorimetrics.

References

[1] H. H. Goldstine, The computer from Pascal to von Neumann, Princeton University Press, 1972.
[2] W. Eicken and E. H. Zuse, Elektronische Datenverarbeitung, 2 (1962), 74–75.
[3] E. L. Post, Introduction to a general theory of elementary propositions, Amer. J. Math., 43 (1921), 163–185.
[4] V. B. Kudryavtsev, A completeness theorem for a class of feedback-free automata (in Russian), Problemy Kibernet., 8 (1962), 91–115.

[5] H. H. Loomis, Jr., Completeness of sets of delayed-logic devices, IEEE Trans., EC-14 (1965), 157–172.
[6] T. Hikita and A. Nozaki, A completeness criterion for spectra, SIAM J. Comput., 6 (1977), 285–297.
[7] I. G. Rosenberg, Completeness properties of multiple-valued logic algebras, Computer Science and Multiple-Valued Logic, Theory and Applications, D. C. Rine (ed.), North-Holland, 1977, 142–186.
[8] S. Eilenberg, Automata, languages and machines, Academic Press, vol. A, 1974; vol. B, 1976.
[9] J. Hartmanis and R. E. Stearns, Algebraic structure theory of sequential machines, Prentice-Hall, 1966.
[10] A. Nozaki, Practical decomposition of automata, Information and Control, 36 (1978), 275–291.
[11] M. A. Arbib, Theories of abstract automata, Prentice-Hall, 1969.
[12] N. Chomsky, On certain formal properties of grammars, Information and Control, 2 (1959), 137–167.
[13] J. E. Hopcroft and J. D. Ullman, Formal languages and their relation to automata, Addison-Wesley, 1969.
[14] A. V. Aho and J. D. Ullman, The theory of parsing, translation and compiling, Prentice Hall, vol. 1, 1972; vol. 2, 1973.
[15] A. Salomaa, Formal languages, Academic Press, 1973.
[16] J. Kuntzmann, Algebra de Boole, Dunod, 1968.
[17] S. Muroga, Logic design and switching theory, Wiley, 1979.
[18] M. L. Minsky, Computation: Finite and infinite machines, Prentice-Hall, 1967.
[19] J. E. Hopcroft and J. D. Ullman, Introduction to automata theory, languages and computation, Addison-Wesley, 1979.
[20] Z. Manna, Mathematical theory of computation, McGraw-Hill, 1974.
[21] W. A. Wulf, M. Shaw, P. N. Hilfinger, and L. Flon, Fundamental structures of computer science, Addison-Wesley, 1981.

76 (VI.17) Conformal Geometry

A. Möbius Geometry

We represent an n-dimensional sphere S^n as the †quadric hypersurface $S^n: x_1^2+x_2^2+\ldots+x_n^2-2x_0x_\infty=0$ in an $(n+1)$-dimensional real †projective space $\mathbf{P}^{n+1}$, where the (x_α) are †homogeneous coordinates in $\mathbf{P}^{n+1}$. We denote by $M(n)$ the group of all †projective transformations of $\mathbf{P}^{n+1}$ that leave S^n invariant. Then the transformation group $M(n)$ acts on S^n. The pair $(S^n, M(n))$ is called the **conformal geometry** or **Möbius geometry**. We call S^n an n-dimensional **conformal space**, a transformation belonging to $M(n)$ a **Möbius transformation**, and $M(n)$ the **Möbius transformation group**. Every point of the projective space $\mathbf{P}^{n+1}$ corresponds to a **hypersphere** on S^n. For example, if a point A lies outside of S^n, then the intersection of S^n and the †polar hyperplane of A with respect to S^n is an $(n-1)$-dimensional sphere S^{n-1}, and the point A corresponds to this **real hypersphere** S^{n-1}. Similarly, if a point A lies on S^n, it corresponds to a **point hypersphere**, and if a point A lies inside of S^n, then A corresponds to an **imaginary hypersphere**. We sometimes identify the point A with the corresponding hypersphere. For any two points $A=(a_\alpha)$ and $B=(b_\alpha)$, we put $AB=a_1b_1+a_2b_2+\ldots+a_nb_n-(a_0b_\infty+a_\infty b_0)$ and call it the **inner product** of the two hyperspheres A and B. The **angle** θ between two intersecting real hyperspheres A and B is defined by $\cos\theta=AB/(\sqrt{A^2}\cdot\sqrt{B^2})$. This angle is invariant under the Möbius transformation.

In the projective space $\mathbf{P}^{n+1}$, we take a †frame $(A_0, A_1, \ldots, A_n, A_\infty)$ that satisfies the conditions

$$A_0^2=A_0A_j=A_iA_\infty=A_\infty^2=0, \quad A_0A_\infty=-1,$$

$$A_iA_j=g_{ij}, \quad i,j=1,2,\ldots,n,$$

where (g_{ij}) is a positive definite matrix. We see that A_0 and A_∞ are points on S^n and each A_i is a real hypersphere passing through these two points. Every hypersphere X of S^n can be written as a linear combination of the A_i: $X=u_0A_0+u_1A_1+\ldots+u_nA_n+u_\infty A_\infty$. That is, X is represented by †projective coordinates (u_α) with respect to the above frame. We call these homogeneous coordinates (u_α) $(n+2)$-**hyperspherical coordinates** of the hypersphere X. If we use these coordinates, the inner product of two hyperspheres $X=(u_\alpha)$ and $Y=(v_\alpha)$ is given by $XY=\sum_{i,j=1}^n g_{ij}u_iv_j-(u_0v_\infty+u_\infty v_0)$.

The Möbius transformation group $M(n)$ is a topological group with two †connected components. If we denote by $M_0(n)$ the maximal connected subgroup of $M(n)$ and by H the subgroup of $M(n)$ that leaves invariant a real hypersphere of S^n, then the set E of all real hyperspheres of S^n can be identified with the †homogeneous space $M_0(n)/H$. The group H also consists of two connected components. If we denote by H_0 the maximal connected subgroup of H, the homogeneous space $\tilde{E}=M_0(n)/H_0$ is a two-fold †covering space of E.

For each real hypersphere $A \in E$, an element $\tilde{A} \in \tilde{E}$ over A is called an **oriented real hypersphere**. The Möbius transformation group contains as its subgroups ones that are isomorphic to the group of †congruent transformations of a Euclidean space and ones that are isomorphic to the group of congruent transformations of a non-Euclidean space. That is, the subgroup of $M(n)$ that leaves a point hypersphere invariant is isomorphic to the group generated by congruent transformations and †homotheties of the Euclidean space E^n. The subgroup of †index 2 (the factor group by a †cyclic subgroup of †order 2) of the subgroup of $M(n)$ that leaves invariant a real (imaginary) hypersphere is isomorphic to the group of congruent transformation of n-dimensional hyperbolic (elliptic) †non-Euclidean space.

In the n-dimensional Euclidean space E^n, consider a hypersphere of radius r with center O. For each point P of E^n, mark a point Q on the †ray OP such that $\overline{OP} \cdot \overline{OQ} = r^2$. We call the point transformation that sends P to Q an **inversion** with respect to the hypersphere. A †symmetry with respect to a hyperplane, considered as an extreme case of inversions, is also called an inversion. We adjoin a point at infinity to the space E^n to construct an n-dimensional sphere S^n. Each inversion can be extended to a transformation of S^n, which we also call an inversion. Then each Möbius transformation is generated by a finite number of inversions. By a Möbius transformation of S^n, each hypersphere is transformed to a hypersphere. Any angle between two curves that intersect at a point of E^n is invariant under Möbius transformations. Conversely, if $n \geqslant 3$, each local transformation of E^n that leaves invariant the angle of each pair of intersecting curves is a †restriction of a Möbius transformation. However, for $n=2$ this is not true in general; any transformation that leaves angles invariant is called a †conformal mapping. Any Möbius transformation $z \rightarrow w$ on the †complex sphere $S^2 = \mathbf{C} \cup \{\infty\}$ can be expressed by an equation of the form $w = (\alpha z + \beta)/(\gamma z + \delta)$ or $w = (\alpha \bar{z} + \beta)/(\gamma \bar{z} + \delta)$, where α, β, γ, and δ are complex numbers such that $\alpha\delta - \beta\gamma \neq 0$ and $\bar{z}$ denotes the †complex conjugate of z.

B. Laguerre Geometry

Let Γ be an oriented smooth curve in a Euclidean plane E^2. The tangent line at a point p of Γ is supplied with an orientation that is induced by the orientation of the curve Γ in an obvious manner. The oriented line l thus obtained is called the **oriented tangent line** of Γ at p.

Let S be the set of oriented lines in E^2 and T be a given set of oriented smooth curves in E^2. Consider a bijection γ of the direct product $S \times T$ to itself satisfying the following condition: If l is an oriented tangent line of a curve Γ belonging to T and γ sends (l, Γ) to (l', Γ'), then l' is an oriented tangent of the curve Γ'. The set of such bijections forms a group G. Suppose that we have $l \in S$ for which there exist two curves Γ_1 and Γ_2 in T such that l is a common tangent line of Γ_1 and Γ_2 at p_1 and p_2, respectively. Further suppose that the element $\gamma \in G$ sends (l, Γ_i) to (l', Γ_i') $(i=1,2)$. Then the element γ of G is called an **equilong transformation** if the following conditions (i) and (ii) are satisfied: (i) l' is tangent to Γ_i' at points p_i' $(i=1,2)$, (ii) the distance between p_1 and p_2 is equal to the distance between p_1' and p_2'. The set of equilong transformations forms a subgroup H of G. In particular, if T is the set of oriented circles (including point circles), the elements of H are called **Laguerre transformations**. In an obvious manner, we can divide the set of oriented circles into two classes, those with "positive" and those with "negative" orientations. With an oriented circle Γ we associate the pair $\varphi(\Gamma) = (P, r)$, where P is the origin of the circle and r is a real number whose absolute value is equal to the radius of the circle and whose signature coincides with that of the orientation of Γ. The mapping $\varphi: T \rightarrow E^2 \times E^1$ thus defined is called **Lie's minimal projection**. An example of a Laguerre transformation, called a **dilatation**, is given by a bijection γ of $S \times T$ to itself satisfying the following condition: Let $\gamma(l, \Gamma) = (l', \Gamma')$; then l' is parallel to l, the distance between l and l' is a given number, and $\varphi(\Gamma') = (P, r+c)$, where $\varphi(\Gamma) = (P, r)$ and c is a given constant. We note here that the action of a Laguerre transformation $\gamma: (l, \Gamma) \rightarrow (l', \Gamma')$ is determined by its action on S (γ acts on S by $\gamma(l) = l'$). Another example of a Laguerre transformation γ, called a **Laguerre inversion**, is determined by means of a given oriented circle O and a line p that is not tangent to O; given an oriented line l, its image l' under the action of γ is determined as follows (here we describe the case where l is not parallel to p). There exists a uniquely determined oriented tangent line g of the circle O parallel to l. Now we have a uniquely determined oriented tangent line g' of O that passes through the point of intersection of the lines g and p such that $g' \neq g$. The image l' is the line parallel to g' and passing through the point of intersection of the lines l and p (Fig. 1). Each Laguerre transformation can be written as a product of a finite number of Laguerre inversions. We denote the group of Laguerre transformations by L. The pair (L, S) is, by defini-

tion, a model of **Laguerre geometry**. Notions such as Laguerre inversions, dilatations, and transformations can be generalized to cases of higher dimension by utilizing oriented hyperspheres and oriented hyperplanes.

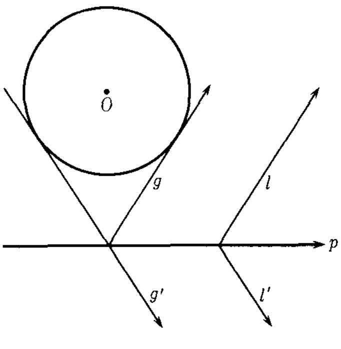


Fig. 1

C. Sphere Geometry

Let S be the set of oriented circles (including point circles and oriented lines) in the Euclidean plane E^2. Two oriented circles C_1 and C_2 are said to be in contact with each other if they have a point p and an oriented tangent line passing through p in common. (An oriented circle C and an oriented line l are in contact with each other if and only if l is an oriented tangent line of C.) In this case, we call the pair (C_1, C_2) a **contact pair**. A bijection γ of S to itself is called a **Lie transformation** if it sends any contact pair to another. (The Lie transformation is a special case of the †contact transformations.) An inversion with respect to a circle determines in an obvious manner a Lie transformation, which is also called an inversion. We denote the group of Lie transformations by G. Any element γ of G can be written as the product of a finite number of inversions and Laguerre inversions, and G contains the group of Möbius transformations and the group of Laguerre transformations as subgroups. The pair (G, S) is called a model of **circle geometry**. The notion of circle geometry can be generalized to that of **hypersphere geometry** for the case of higher dimensions. Specifically, when we replace E^2 by E^3 and circles by spheres, we have **sphere geometry**.

Let V be a complex 3-dimensional space, and let M, N be the sets of oriented lines and oriented spheres in V, respectively. Then M, N have the structure of 4-dimensional complex manifolds that are homeomorphic to each other. The homeomorphism is given by the **Lie line-sphere transformation** that induces a bijection from the set of pairs of intersecting oriented lines onto the set of pairs of oriented spheres that are in contact with each other.

D. Group-Theoretic Considerations

Here we discuss the preceding three kinds of geometries from the group-theoretic point of view (→ 137 Erlangen Program). Let us consider the quadratic form Q defined by $Q(x) = -x_0^2+x_1^2+x_2^2+x_3^2-x_4^2$ in a real projective space $\mathbf{P}^4$, where $x=(x_0, x_1, x_2, x_3, x_4)$ are homogeneous coordinates. We denote by G the set of all projective transformations of $\mathbf{P}^4$ that leave Q invariant. The group G consists of the set of matrices A of order 5 such that $\det A = 1$ and $Q(Ax)=Q(x)$ holds for all x in $\mathbf{P}^4$; we denote by L^3 the set of all points x in $\mathbf{P}^4$ that satisfy $Q(x)=0$. Then G acts transitively on L^3. Hence if we denote by H_a the set of all elements of G that fix a point a in L^3, for example $a=(-1,1,0,0,0)$, we may assume that $L^3 = G/H_a$. The circle geometry that belongs to the group of Lie transformations of circles is exactly the geometry of the homogeneous space G/H_a. The group G_b of all transformations of G that leave the hyperplane $x_4=0$ invariant acts transitively on $L^3 \cap \{x_4=0\}$ (G_b is isomorphic to $M(2)$). The geometry of the homogeneous space $G_b/H_a \cap G_b$ is plane conformal geometry. Next, the group G_a of all transformations of G that leave invariant $x_0+x_1=0$ coincides with H_a and acts transitively on $L^3 \cap \{x_0+x_1=0\}$. The geometry of the homogeneous space $L^3 \cap \{x_0+x_1=0\}$ (on which G_a acts transitively) is plane Laguerre geometry. In this sense, the circle geometry that belongs to the group of Lie transformations contains the other circle geometries as subgeometries.

We now describe how plane Laguerre geometry can be realized as the geometry of the space $L^3 \cap \{x_0+x_1=0\}$. Let E^2 be a plane in a Euclidean space E^3. We fix a Cartesian coordinate system (y_0, y_1, y_2) in E^3 so that E^2 is given by $y_0=0$. To each point y of E^3 we can associate an oriented circle in E^2 with center $(0, y_1, y_2)$, radius $|y_0|$, and positive (negative) orientation if y_0 is positive (negative). If y lies on E^2, the corresponding circle is the point circle y itself. Now let us consider the group G_a' of all affine transformations of E^3 whose rotation parts leave the quadratic form $Q'(y) = -y_0^2+y_1^2+y_2^2$ invariant. (G_a' is an isometry with respect to the †metric defined by Q'.) Each element of G_a' induces a transformation of the set of oriented circles in E^2 (including the oriented lines and point circles) onto itself. The mappings of E^3 into L^3 defined by $x_0 = (1+Q'(y))/2$, $x_1=(1-Q'(y))/2$, $x_2=y_1$, $x_3=y_2$, $x_4=y_0$ is a one-to-one correspondence of E^3 onto the subset of L^3 such that $x_0+x_1 \neq 0$. This correspondence induces an isomorphism of G_a' onto G_a. In Laguerre geometry, there are

no essential distinctions between points and oriented circles on E^2; and the group G_a acts on a 3-dimensional space of oriented circles (including point circles).

Reference

[1] W. Blaschke, Vorlesungen über Differentialgeometrie III, Springer, 1929.

77 (XI.14) Conformal Mappings

A. General Remarks

Let a function $w=f(z)$ that maps a †domain D on the †complex z-sphere homeomorphically onto a domain Δ on the complex w-sphere satisfy the following two conditions: (1) Every curve $C_z: z(t)$ $(0 \leqslant t \leqslant 1)$ that starts at any point z_0 in D and possesses a tangent there has an image curve $C_w: w=w(t)=f(z(t))$ $(0 \leqslant t \leqslant 1)$ that also possesses a tangent at the image point $w_0 = f(z_0)$. (2) The angle between any two curves $C_z^{(1)}$ and $C_z^{(2)}$ possessing tangents at z_0 is equal to the angle between their image curves $C_w^{(1)}$ and $C_w^{(2)}$, where the direction of the angle is also taken into account. Then the mapping from D onto Δ is said to be **conformal**, and Δ is called **conformally equivalent** to D. It has been proved that $w=f(z)$ is necessarily a function †analytic in D (D. Men'shov, 1931).

Consequently, the theory of conformal mapping is a branch of the theory of analytic functions. That a function $w=f(z)$ maps a domain D conformally onto a domain Δ means that it is a †meromorphic function †univalent in D and its range is Δ. Then $f'(z) \neq 0$ holds at every (finite) point in D, and the ratio of the lengths of the segments between two points $w_0 = w(0)$, $w(t)$ on C_w and between two points $z_0 = z(0)$, $z(t)$ on C_z tends to a fixed nonvanishing limit $|f'(z_0)|$ as $t \to 0$ independently of the choice of C_z. Hence if z_1 and z_2 lie on $C_z^{(1)}$ and $C_z^{(2)}$, respectively, and w_1 and w_2 are their image points lying on the image curves $C_w^{(1)}$ and $C_w^{(2)}$, respectively, then the two triangles $\Delta z_1 z_0 z_2$ and $\Delta w_1 w_0 w_2$ are nearly similar in the positive sense, provided that z_1 and z_2 are near enough to z_0. This justifies the word "conformal," which means "of the same form" (→ Appendix A, Table 13).

B. Conformal Mapping onto the Unit Disk

A fundamental theorem in the theory of conformal mapping is **Riemann's mapping theorem**, which states that any †simply connected domain D with at least two boundary points can be mapped conformally onto the interior Δ of the unit circle. This theorem is equivalent to the assertion of the existence of †Green's function of D and can be proved in various ways. B. Riemann (1851) gave a proof, based on an idea of C. F. Gauss, by assuming the existence of a solution for a variational problem minimizing the †Dirichlet integral. The logical incompleteness implied by this assumption was later removed by D. Hilbert and others. The proof that is now regarded as simplest is due to L. Fejér and F. Riesz's method (T. Radó, 1922, 1923), which applies †normal family theory. On the other hand, the **osculating process** due to P. Koebe (1912) is a purely constructive method of proving existence that is also applicable to the case of †multiply connected domains. The mapping function $w=f(z)$ in Riemann's mapping theorem is uniquely determined under the normalization condition $f(z_0)=0$, $\arg f'(z_0) = \theta_0$ at a point z_0 in D, where θ_0 is a given angle.

The other types of simply connected domains are the Riemann sphere less one point and the sphere itself. In both cases every conformal mapping of these domains is a linear transformation. A simply connected domain is called **hyperbolic**, **parabolic**, or **elliptic** if it is conformally equivalent to the unit disk, the complex plane, or the Riemann sphere, respectively.

Let $w=f(z)$ map a simply connected domain D on the z-sphere conformally onto a simply connected domain Δ on the w-sphere. If a sequence $\{z_\nu\}$ in D tends to a boundary point ζ of D, then the corresponding sequence $\{w_\nu\}$ $(w_\nu = f(z_\nu))$ has no †accumulation point in Δ and does not necessarily tend to a boundary point of Δ.

If for any sequence $\{z_\nu\}$ tending to ζ, $\{f(z_\nu)\}$ tends to a unique point ω on the boundary of Δ, it is said that $f(z)$ possesses a **boundary value** ω at ζ. To investigate the behavior of $\{w_\nu\}$ on a hyperbolic domain D, we can assume, by using a suitable mapping if necessary, that D is bounded. Then the problem is reduced, in view of Riemann's mapping theorem, to the case where D is a bounded simply connected domain and Δ is the unit disk $|w|<1$.

C. Correspondence between Boundaries

Concerning the correspondence between boundaries under the conformal mapping $w=f(z)$ of a bounded simply connected domain D onto the unit disk $|w|<1$, we have the following three theorems:

(1) To any †accessible boundary point z_C of D there corresponds a unique point on $|w|=1$, and to any distinct accessible boundary points z_{C_1} and z_{C_2} of D there correspond distinct points on the unit circumference. Furthermore, the set of all points on $|w|=1$ that correspond to accessible boundary points of D has †angular measure equal to 2π.

(2) There is a one-to-one correspondence between †boundary elements of D and points on $|w|=1$ (C. Carathéodory).

(3) Let $w=f(z)$ map the interior D of a †Jordan curve C conformally onto the unit disk $\Delta:|w|<1$. Then it possesses a boundary value, say $f(\zeta)$, at every point ζ on C that satisfies $|f(\zeta)|=1$. Hence $f(z)$ is continuous on the closed domain $\bar{D}=D\cup C$ and maps $\bar{D}$ bijectively onto the closed disk $\bar{\Delta}:|w|\leqslant 1$. Similarly, the inverse function $z=\varphi(w)$ has an analogous property and maps $\bar{\Delta}$ bijectively and continuously onto $\bar{D}$; that is, a conformal mapping of the interior D of a Jordan curve onto the unit disk Δ can be extended into a homeomorphism of the closure $\bar{D}$ to $\bar{\Delta}$ (Carathéodory).

In this case, if the Jordan curve C contains a †regular analytic arc Γ, then the mapping function $w=f(z)$ can be prolonged analytically beyond Γ (except at the endpoints of Γ). Hence the mapping $w=f(z)$ is conformal at interior points of Γ.

Problems on the correspondence of angles at the boundary are closely related to †angular derivatives. These problems have been attacked by Carathéodory, S. Warschawski, J. Wolff, and others.

D. Schwarz-Christoffel Transformation

The problem of determining the form of an analytic function that maps the interior of a circle or a half-plane conformally onto the interior of a polygon was first dealt with by H. A. Schwarz and E. B. Christoffel (1869) (→ Appendix A, Table 13).

Let P be a polygon in the complex w-plane with vertices b_μ $(\mu=1,\ldots,m)$ and interior angles $\alpha_\mu\pi$ at b_μ. Then a function $w=f(z)$ that maps a circular disk or a half-plane in the z-plane conformally onto the interior of P is given by

$$f(z)=C\int^z\prod_{\mu=1}^{m}(a_\mu-z)^{\alpha_\mu-1}\,dz+C'.$$

Here $b_\mu=f(a_\mu)$, and C, C' are constants depending on the position and magnitude of the polygon P. If instead of a disk we consider a half-plane in the z-plane, and if one of the a_μ is the point at infinity, then, modifying this formula by deleting the factor of the integrand corresponding to $a_\mu=\infty$, we obtain a formula for $w=f(z)$. This representation is called the **Schwarz-Christoffel transformation formula**. This formula was originally derived by Christoffel to solve a problem of 2-dimensional distribution of stationary temperature. It then found extensive application to several problems of conformal mapping concerning polygonal domains and to problems of determining force lines or stream lines and equipotential lines in 2-dimensional electrostatics or hydromechanics.

An analogous formula is also derived for a function that maps the interior of a circle or a half-plane conformally onto the exterior of a polygon. In connection with these formulas, a third-order differential equation that is satisfied by a function mapping the interior of a circle onto a domain bounded by a circular polygon is found useful in the theory of automorphic functions. Moreover, a representation analogous to the Schwarz-Christoffel transformation formula is obtained for a function that maps the interior of a circle onto a curvilinear polygonal domain bounded by arcs of logarithmic spirals with the origin as asymptotic point.

E. Conformal Mapping of Multiply Connected Domains

It is also important to consider problems concerning one-to-one conformal mapping of a multiply connected domain on the z-sphere onto a suitable multiply connected domain $\mathfrak{D}$ on the w-sphere. The two domains D and $\mathfrak{D}$ are then homeomorphic, but the converse is not true; i.e., there does not necessarily exist a one-to-one conformal mapping between D and $\mathfrak{D}$ even when they are homeomorphic. Now let D and $\mathfrak{D}$ be multiply connected domains on the z- and w-planes, respectively, both possessing at least three boundary points. The †universal covering surfaces $\hat{D}$ and $\hat{\mathfrak{D}}$ of D and $\mathfrak{D}$ are hyperbolic and they can be mapped onto the unit disks. The groups of their †covering transformations form the †Fuchsian groups G and $\mathfrak{G}$. Then in order for D to be mapped one-to-one and conformally onto $\mathfrak{D}$, it is necessary and sufficient that the group $\mathfrak{G}$ be transformed into the group G by a suitable linear transformation.

To a domain of finite †connectivity having only continua for its boundary components, we can associate **conformal invariants** (namely, **moduli**) expressed by one real parameter in the doubly connected case and by $3n-6$ real parameters in the $n(>2)$-connected case. A one-to-one conformal mapping is possible only within a class of domains having the same invariants (→ 416 Teichmüller Spaces).

While a circular disk is taken as a canonical domain in the simply connected case, an †annulus is often taken as a canonical domain in the doubly connected case. In the latter case, the logarithm of the ratio (>1) of the radii of two concentric boundary circles is usually called the **modulus**. There are various types of $n(\geqslant 2)$-connected canonical domains, for instance, the whole plane, a circular disk or annulus slit along concentric circular arcs or radial segments, a parallel slit plane, etc. The possibility of a one-to-one conformal mapping of a given domain onto a canonical domain of such a type was proved by Hilbert, Koebe, and others in a †potential-theoretic way and by E. Rengel, R. de Possel, H. Grunsky, and others in a purely function-theoretic way. Such canonical domains are characterized by some extremal properties. For example, the horizontal parallel slit mapping function $p_0(z, z_0) = (z-z_0)^{-1} + a[p_0](z-z_0) + \ldots$, $z_0 \in D$, has the extremal property that it is the unique function maximizing $\mathrm{Re}\, a[f]$ within the family of univalent functions $f(z) = (z-z_0)^{-1} + a[f](z-z_0) + \ldots$ in D. For a general domain D, $p_0(z, z_0)$, called the extremal horizontal slit mapping, is defined as the limit function of the sequence of horizontal slit mappings $p_0^{(n)}(z, z_0)$ of D_n, $z_0 \in D_n$. Here $\{D_n\}_{n=1}^{\infty}$ is a canonical exhaustion of D. It has the same extremal property as in the case of finite connectivity. The boundary components of the image domain of D under $p_0(z, z_0)$ consist of a horizontal slit and a point. The parallel slit mapping $p_\theta(z, z_0)$ in the direction of θ is defined similarly to the one that maximizes $\mathrm{Re}\, e^{-2i\theta} a[f]$ [8].

For a domain of infinite connectivity, by accumulation of boundary components, a pointlike boundary component can be mapped onto a continuum. A boundary component γ of a domain D is called **weak** if its image $f(\gamma)$ is a point under every conformal mapping f on D. γ is called **strong** if $f(\gamma)$ always consists of more than one point. A boundary component γ that is neither weak nor strong is called **unstable** (L. Sario, *J. Analyse Math.*, 5 (1956)). Criteria of weakness, etc., by means of extremal length are given in [8]. Moreover, in terms of extremal length, a generalization of a boundary element is given by Oikawa and Suita [9]. On the other hand, the existence of one-to-one and conformal mapping of a domain bounded by a finite number of curves onto the whole plane with mutually disjoint circular disks removed was proved by Koebe and later derived by J. Douglas and R. Courant as a particular case of the existence of a solution of †Plateau's problem [10]. M. Schiffer showed that the mapping function is a solution of an extremal problem involving Fredholm eigenvalues [11]. L. Bieberbach and H. Grunsky showed the possibility of mapping an n-connected domain onto an n-sheeted disk. Concerning doubly connected domains, detailed investigations were made by O. Teichmüller, Y. Komatu, and others. For a domain D bounded by n analytic Jordan curves, as an extension of the Schwarz lemma, Ahlfors showed that there exists a unique extremal function $f_0(z)$ maximizing $\mathrm{Re}\, f'(z_0)$, $z_0 \in D$, within the family of analytic functions $f(z)$ satisfying $|f(z)| < 1$ in D and that $f_0(z)$ maps D onto the n-sheeted unit disk (1947). $f_0(z)$ is called the **Ahlfors function** at z_0. For a general domain the uniqueness of the extremal function in nontrivial cases was established first by S. Havinson (1961) and later by L. Carleson (1967) and S. Fisher (1969).

Most †kernel functions of a plane domain D have connections with conformal mappings. For example, the Bergman kernel $U(z, \zeta)$ of exact differentials is equal to $(2\pi)^{-1}(p_0'(z, \zeta) - p_{\pi/2}'(z, \zeta))$. The adjoint kernel $V(z, \zeta)$ is defined by $(2\pi)^{-1}(p_0'(z, \zeta) + p_{\pi/2}'(z, \zeta))$. M. Schiffer showed that the integral of $V(z, \zeta)$, $S(z, \zeta) = \int_\zeta^z V(z, \zeta)\, dz$, is univalent and maps D onto a domain bounded by n analytic convex curves if D is bounded by n Jordan curves [13]. For a general domain D, this is true in the sense that each boundary component of the image domain under $S(z, \zeta)$ is a convex set (K. Oikawa and N. Suita, *Kôdai Math. Sem. Rep.*, 16 (1964)). It is known that $S(z, \zeta)$ maximizes the area of the complementary set of the image domain under $f(z)$ within the family of univalent functions $f(z) = (\pi(z-\zeta))^{-1} + b_1(z-\zeta) + \ldots$. The maximum value multiplied by 4π is called the **span** of the domain D; it is equal to $a[p_0] - a[p_{\pi/2}]$. For the †Szegö kernel function $k(z, \zeta)$ of a domain D bounded by n analytic curves, the adjoint kernel $l(z, \zeta)$ is defined, and the Ahlfors function $f_0(z)$ at ζ is expressed by $k(z, \zeta)/l(z, \zeta)$ (P. Garabedian, *Trans. Amer. Math. Soc.*, 67 (1949)).

F. Universal Constants

Among various universal constants appearing in the theory of conformal mapping, Bloch's constant is especially famous. A. Bloch (1924) showed that a covering surface over the w-plane obtained from a mapping $w = F(z) = z + \ldots$ that is one-to-one, conformal, and holomorphic in $|z| < 1$ always contains a †univalent (schlicht) disk whose radius B is a positive number independent of the function F (**Bloch's theorem**). The supremum $\mathfrak{B}$ of such constants B is called **Bloch's constant**. The true value of $\mathfrak{B}$ is yet unknown, but estimates have

been given by Ahlfors (*Trans. Amer. Math. Soc.*, 43 (1938)) and by Ahlfors and Grunsky (*Math. Z.*, 42 (1937)) in the form

$$\frac{\sqrt{3}}{4} \leqslant \mathfrak{B} \leqslant \sqrt{\pi}\, 2^{1/4} \frac{\Gamma(1/3)}{\Gamma(1/4)} \left(\frac{\Gamma(11/12)}{\Gamma(1/12)}\right)^{1/2} = 0.4719\ldots.$$

M. Heins showed that the equality sign can be deleted from the left-hand side (*Nagoya Math. J.*, 21 (1962)) and C. Pommerenke gave a simpler proof for the same fact (*J. London Math. Soc.*, 2 (1970)). It is conjectured that the correct value of $\mathfrak{B}$ is equal to the upper bound of Ahlfors and Grunsky. **Landau's constant** $\mathfrak{L}$ corresponding to the case where the image disks are not necessarily univalent satisfies $0.5 \leqslant \mathfrak{L} < 0.55$. The lower and upper bounds are due to Ahlfors (cited above) and Landau (*Math. Z.*, 30 (1929)). Pommerenke showed $0.5 < \mathfrak{L}$ (1970, cited above). For the family of univalent functions these constants coincide, and the value is called the **schlicht Bloch constant**, denoted by $\mathfrak{A}$. Its lower and upper bounds are known to be 0.5705 (J. Jenkins, *J. Math. Mech.*, 10 (1961)) and 0.6565 (R. Robinson, *Bull. Amer. Math. Soc.*, 41 (1935)). (For distortion theorems and coefficient problems → 438 Univalent and Multivalent Functions.)

References

[1] L. V. Ahlfors, Complex analysis, McGraw-Hill, second edition, 1979.
[2] Z. Nehari, Conformal mappings, McGraw-Hill, 1952.
[3] D. Gaier, Konstruktive Methoden der konformen Abbildung, Springer, 1964.
[4] C. Carathéodory, Conformal representation, Cambridge Univ. Press, 1932, second edition, 1952.
[5] N. Ohtsuka, Dirichlet problem, extremal length and prime ends, Van Nostrand, 1970.
[6] H. Grunsky, Lectures on theory of functions in multiply connected domains, Vandenhoeck & Ruprecht, 1978.
[7] G. M. Golusin, Geometric theory of functions of a complex variable, Amer. Math. Soc. Transl., 1969. (Original in Russian, 1952.)
[8] L. Sario and K. Oikawa, Capacity functions, Springer, 1969.
[9] K. Oikawa and N. Suita, On conformal mappings onto incised radial slit discs, Kôdai Math. Sem. Rep., 22 (1970), 45–72.
[10] R. Courant, Dirichlet's principle, conformal mapping, and minimal surfaces, with an appendix by M. Schiffer, Interscience, 1950.
[11] M. Schiffer, Fredholm eigenvalues of multiply connected domains, Pacific J. Math., 9 (1959), 211–269.
[12] S. Bergman, The kernel function and conformal mapping, Amer. Math. Soc. Math. Surveys, 5 (1970).
[13] M. Schiffer, The span of multiply connected domains, Duke Math. J., 10 (1943), 209–216.
[14] L. V. Ahlfors, Conformal invariants, McGraw-Hill, 1973.

78 (VI.9) Conic Sections

A. General Remarks

Suppose that we are given two straight lines l and m intersecting at V (but not orthogonally) in the 3-dimensional Euclidean space E^3. By rotating the line m around l, we obtain a surface $\mathfrak{F}$. We call this surface $\mathfrak{F}$ a **circular cone** with **vertex** V and **axis** l; a straight line on the surface passing through V is called a **generating line** of $\mathfrak{F}$.

A section C of $\mathfrak{F}$ by a plane π not passing through V $(C = \pi \cap \mathfrak{F})$ is called a **conic section** (or simply a **conic**). This C is a plane curve on the plane π. The point set $\mathfrak{F} - V$ consists of two †connected components $\mathfrak{F}_1$ and $\mathfrak{F}_2$. Let π_i $(i = 1, 2, 3)$ be planes not passing through V. If the conic section $C_1 = \pi_1 \cap \mathfrak{F}$ is †bounded, then C_1 is contained either in $\mathfrak{F}_1$ or in $\mathfrak{F}_2$ and is †connected. We call such a C_1 an **ellipse**. When $C_2 = \pi_2 \cap \mathfrak{F}$ is not bounded but is connected, then π_2 is parallel to one of the generating lines of $\mathfrak{F}$, and C_2 is contained either in $\mathfrak{F}_1$ or in $\mathfrak{F}_2$. We call such a C_2 a **parabola**. When π_3 intersects both of $\mathfrak{F}_1$ and $\mathfrak{F}_2$, then $C_3 = \pi_3 \cap F$ has two connected components and is not bounded. We call such a C_3 a **hyperbola**. These three types exhaust all possible types of conic sections. In particular, if the plane π is perpendicular to the axis l, then $C = \pi \cap \mathfrak{F}$ becomes a circle. Thus a circle is a special kind of ellipse.

B. Foci and Directrices

Let $C = \pi \cap \mathfrak{F}_1$ be an ellipse. The Euclidean space E^3 is divided by π into two †half-spaces E_1^3, E_2^3 (two "sides" of π). If we put $\mathfrak{F}_1 \cap E_1^3 = \mathfrak{F}_{11}$, $\mathfrak{F}_1 \cap E_2^3 = \mathfrak{F}_{12}$, we can construct a sphere S that is contained in E_1^3, tangent to $\mathfrak{F}_{11}$ along a circle K, and tangent to π at a point F. Similarly, we can construct a sphere S' that is in E_2^3, tangent to $\mathfrak{F}_{12}$ along a circle K', and tangent to π at a point F'. We call F, F' the **foci** of the ellipse (Fig. 1).

Let κ, κ' be the planes containing K, K'. Straight lines $d=\kappa\cap\pi$, $d'=\kappa'\cap\pi$ are called **directrices** of C. Unless C is a circle, we have $F\neq F'$, and κ, π (and κ', π) actually intersect; hence d, d' exist. When $C=\pi\cap\mathfrak{F}$ is a parabola or a hyperbola (Figs. 2 and 3), we can similarly define foci (a parabola has only one focus, F, while a hyperbola has two foci, F, F') and directrices (a parabola has only one directrix, d, while a hyperbola has two directrices, d, d').

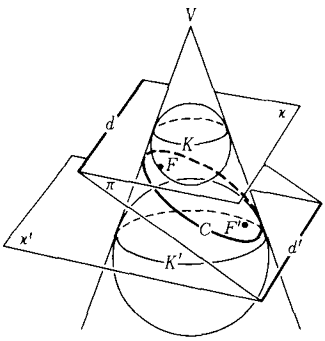


Fig. 1

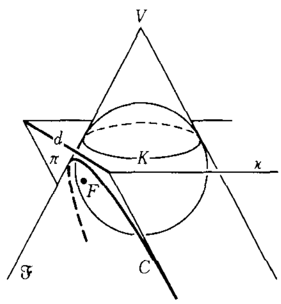


Fig. 2

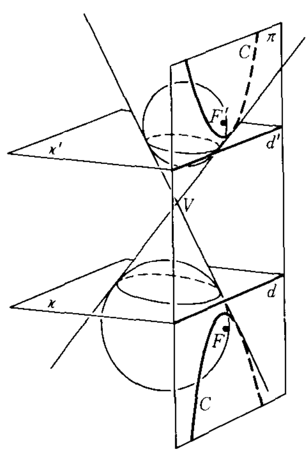


Fig. 3

Let X be a point on the plane π, let $D_F(X)$ be the distance between the point X and a focus F, and let $D_d(X)$ be the distance between X and a directrix d. Then the curve C is the locus of the points X satisfying the condition $D_F(X)=e\cdot D_d(X)$, where e is a constant. We call e the **eccentricity** of the conic section C. According as C is an ellipse, a parabola, or a hyperbola, we have $e<1$, $e=1$, or $e>1$. A circle is an ellipse whose eccentricity is zero. An ellipse is also characterized as the locus of X such that $FX+F'X=2a$; a hyperbola is the locus of X such that $|FX-F'X|=2a$, where a is a positive constant. When there are two foci, the straight line FF' is perpendicular to directrices d and d'.

C. Canonical Forms of Equations

When C is a hyperbola or an ellipse that is not a circle, C has two foci, F and F'. In this case, the midpoint O of the segment FF' is the center of symmetry of C (when C is a circle, its center O is, of course, the center of symmetry of C). We call O the **center** of C; an ellipse or a hyperbola is called a **central conic**. If we choose a rectangular coordinate system (x, y) having O as the origin and FF' as x-axis, then the equation of C can be expressed in the form

$$x^2/a^2\pm y^2/b^2=1, \qquad a,b>0. \tag{1}$$

According as C is an ellipse or a hyperbola, we take the + or − of the double sign. If C is an ellipse, we have $a>b$. Furthermore, $e=\sqrt{a^2-b^2}/a$ if C is an ellipse and $e=\sqrt{a^2+b^2}/a$ if C is a hyperbola. We also have $F=(ae,0)$ and $F'=(-ae,0)$; the equations of directrices are $x=\pm a/e$.

On the other hand, if C is a parabola, the straight line that is perpendicular to the directrix d and passes through F becomes the axis of symmetry of C. We call this straight line the **axis** of C; the intersection O of the axis and C is called the **vertex** of C. If we choose a rectangular coordinate system (x, y) having O as the origin and having the axis of C as the x-axis, the equation of C can be expressed in the form

$$y^2=4ax, \qquad a>0. \tag{2}$$

We call (1) and (2) the **canonical** (or **standard) forms of the equation** of C. We call the associated coordinate system the **canonical coordinate system**. Suppose that C is an ellipse (hence $a>b$). Let A, A' be points of intersection of the x-axis and the ellipse and B, B' be the points of intersection of the y-axis and the ellipse. We call AA' the **major axis** of C and BB' the **minor axis** of C. If C is a hyperbola and (x, y) is the canonical coordinate system, we call the x-axis the **transverse axis** and the y-axis the **conjugate axis**. If C is a central conic, the x- and y-axes of the canonical coordinate system are called the **principal axes**; if C is a parabola, the x-axis is sometimes called the **principal axis**.

D. Properties of Ellipses

An ellipse may be considered the image of a circle under a †parallel projection. Consequently, the section of a circular cylinder by a plane is an ellipse. Also, if we are given a circle C and a fixed diameter D of C, an ellipse is obtained as the locus of the points X lying on lines PM, which are perpendicular to D, with $P\in C$, $M\in D$, satisfying the condition that the ratio $PM:XM$ is constant.

Suppose that we are given two concentric circles having the center at the origin O and with radii a, b. Let P, Q be points of intersection of moving half-lines through O and the two circles. Then the locus of points X of intersection of the ordinates (lines parallel to the y-axis) passing through P and the abscissae (lines parallel to the x-axis) passing through Q is an ellipse (Fig. 4). Suppose that the equation of an ellipse is given by $x^2/a^2+y^2/b^2=1$, with $a>b$. Then the lengths of its major axis and minor axis are $2a$ and $2b$, respectively.

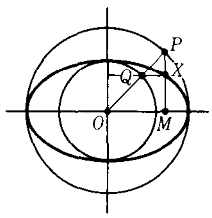


Fig. 4

Given an ellipse and its center O, the circle with center O and diameter equal to the major axis of the ellipse is called the **auxiliary circle** of the ellipse. Given an ellipse C and its focus F, the auxiliary circle of C is the locus of the points X satisfying the condition that the line FX is perpendicular to a tangent line to C passing through X. Suppose that X is a point on an ellipse with foci F, F'. Let TT' be the line tangent to the ellipse at X (X lies between T and T'). Then the angle $\angle TXF'$ is equal to $\angle T'XF$ (Fig. 5). Consequently, the rays starting from one focus of an ellipse and "reflected" by the ellipse converge on the other focus of the ellipse. Also, the product of the distances from two foci of an ellipse to an arbitrary tangent is constant and is equal to b^2.

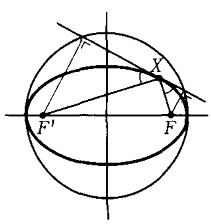


Fig. 5

The ellipse $C:x^2/a^2+y^2/b^2=1$ is expressed parametrically in the form

$$x=a\cos\theta,\qquad y=b\sin\theta. \tag{3}$$

We call the parameter θ the **eccentric angle** of a point (x,y) on C. Consequently, C is a †Jordan curve and divides the plane into two parts, the inside and the outside. The inside is the set of points (x,y) satisfying $x^2/a^2+y^2/b^2<1$, and the outside is the set of points (x,y) satisfying $x^2/a^2+y^2/b^2>1$. The inside is a †convex set. From a point Q outside C, two tangents to C can be drawn. The locus of points Q such that these two tangents are orthogonal is the circle $x^2+y^2=a^2+b^2$. We call this circle the **direct circle**. The area of the "sector" OAX formed by two points $A(a,0)$, $X(a\cos\theta, b\sin\theta)$ $(\theta>0)$ and the origin O is $ab\theta/2=(ab/2)\operatorname{Arccos}(x/a)$; the length of the arc $\widehat{AX}$ of the ellipse is represented by the value of the †elliptic integral

$$a\int_0^\theta\sqrt{1-e^2\cos^2\theta}\,d\theta=aE(\pi/2-\theta,e).$$

In particular, the area inside an ellipse is equal to πab, and the whole length of the ellipse is $4aE(0,e)$.

With respect to a polar coordinate system (r,θ) having the focus $F(ae,0)$ as the origin and the ray directed positively along the x-axis as the initial line, the equation of the ellipse C is

$$r=\frac{l}{1+l\cos\varphi},\qquad l=\frac{b^2}{a}. \tag{4}$$

Here l is equal to half of the length of the chord that is perpendicular to the major axis and passes through the focus. (This chord is called the **latus rectum** of the ellipse.) Suppose that F is a fixed point and that X is a moving particle attracted toward F by a †central force inversely proportional to the square of the length of FX. Suppose further that X begins with an initial velocity whose direction is tangent to the ellipse C with focus F. Then X always moves on C, and the areal velocity described by the radius FX is constant (Kepler's second law).

E. Properties of Hyperbolas

Two straight lines $x^2/a^2-y^2/b^2=0$, that is, $y/x=\pm b/a$, are †asymptotes of the hyperbola $C:x^2/a^2-y^2/b^2=1$. The hyperbola $C':x^2/a^2-y^2/b^2=-1$ is called the **conjugate hyperbola** of C. When $a=b$, the asymptotes are orthogonal to each other, and C, C' are congruent. In this case, we call C a **rectangular hyperbola** (or **equilateral hyperbola**). When we draw parallels to asymptotes from a point X on C, the area of the parallelogram formed by these

lines and two asymptotes is constant. (In particular, when C is a rectangular hyperbola, the equation of C becomes $xy=k^2/2$ if we take two asymptotes as coordinate axes. The segment cut off by the asymptotes on the tangent to C at X is divided equally at X. In the case of the hyperbola as well, the product of the distances from two foci to an arbitrary tangent is constant and is equal to b^2, and the angle between two straight lines joining two foci to a point X on C is divided equally by the tangent at X.

A hyperbola C is represented parametrically by

$$x=a\sec\theta, \qquad y=b\tan\theta. \tag{3'}$$

In this case also, we call θ the **eccentric angle** of (x,y). If we use the †hyperbolic functions and the parameter u, the equation of a hyperbola can be written as

$$x=a\cosh u, \qquad y=b\sinh u \tag{3''}$$

instead of (3′). The area of a "sector" OAX formed by two points $A(a,0)$, $X(x,y)$ on the hyperbola and the origin O is, in this case,

$$\frac{abu}{2}=\frac{ab}{2}\operatorname{Arccosh}\frac{x}{a}=\frac{ab}{2}\log\frac{x+\sqrt{x^2-a^2}}{a}.$$

The length of the arc $\widehat{AX}$ of the hyperbola is given by the elliptic integral

$$\int_0^x\sqrt{\frac{e^2x^2-a^2}{x^2-a^2}}\,dx.$$

With respect to a polar coordinate system (r,φ) having the focus $F(ae,0)$ as origin and the ray directed positively along the x-axis as the initial line, the equation of the hyperbola C becomes

$$r=\frac{l}{1-e\cos\varphi}, \qquad l=\frac{b^2}{a}, \tag{4'}$$

where l is equal to half the length of the chord passing through the focus and perpendicular to the principal axis. (This chord, too, is called the **latus rectum**.)

F. Properties of Parabolas

The curve described by a particle attracted by "gravitation" in a fixed direction and affected by no other force is a parabola (G. Galilei). The tangent at a point X on a parabola makes equal angles with the straight line joining F and X and the direction of the principal axis (Fig. 6). Consequently, if the rays starting from the focus of a parabola are "reflected" by the parabola, they all become rays parallel to the principal axis. Let X' be

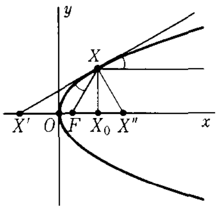


Fig. 6

the point of intersection of the tangent at $X(x_0,y_0)$ and the x-axis, X'' the point of intersection of the normal at X and the x-axis, and X_0 the foot of the perpendicular from X to the x-axis. Then $FX=FX'$, $\triangle FXX'$ is an isosceles triangle, and XX' is divided equally by the y-axis. Consequently, the locus of the foot of a perpendicular from F to a tangent is the y-axis. Also, the length of subtangent $X'X_0=2x_0$, and the length of subnormal $X''X_0=2a=2OF$. Conversely, a curve whose length of subnormal is constant is a parabola. The locus of the midpoints of parallel chords of a parabola is a straight line parallel to the principal axis.

With respect to a polar coordinate system having the focus as origin and the ray directed positively along the x-axis as the initial line, the equation of a parabola is

$$r=\frac{l}{1-\cos\varphi}, \qquad l=2a. \tag{4''}$$

The area bounded by a chord BC and an arc $\widehat{BC}$ of a parabola (Fig. 7) is equal to 4/3 the area of $\triangle ABC$, where A is the point of contact on the tangent of the parabola parallel to BC (Archimedes). Also, the length of the arc $\widehat{OX}$ of parabola (2) is

$$\frac{y_0}{4a}\sqrt{y_0^2+a^2}+a\log\frac{y_0+\sqrt{y_0^2+a^2}}{2a},$$

X having the coordinates (x_0,y_0).

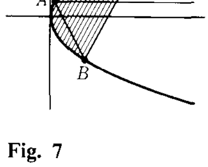


Fig. 7

G. Conjugate Diameters

The **diameter** is a straight line passing through the center of a central conic. The locus of the midpoints of the chords parallel to a diameter d is another diameter d', called **conjugate** to d.

Then the diameter conjugate to d' is d. The x-axis and y-axis of a canonical coordinate system form a set of conjugate diameters. Let $2a'$, $2b'$ be the lengths of the segments (sometimes called conjugate diameters) cut off from d, d' by the curve or by the curve and the conjugate one for hyperbolas and by ω, the angle between d and d'. Then the following relations hold (as to the double signs $\pm$, we take $+$ in the case of an ellipse and $-$ in the case of a hyperbola): $a'^2 \pm b'^2 = a^2 \pm b^2$, $a'b' \sin\omega = ab$. The product of the slopes of d, d' is equal to $\pm b^2/a^2$. With respect to an oblique coordinate system having d and d' as axes, the equation of the curve is $x^2/a'^2 \pm y^2/b'^2 = 1$.

H. Confocal Conic Sections

The set of ellipses and hyperbolas having two fixed points F, F' as foci is called the **family of confocal central conics** with foci F and F' (Fig. 8). The family of confocal central conics containing the ellipse $x^2/a^2 + y^2/b^2 = 1$ is represented parametrically by

$$\frac{x^2}{a^2+\lambda} + \frac{y^2}{b^2+\lambda} = 1.$$

There exist only one ellipse and only one hyperbola that pass through a point inside each quadrant (for example, a point (x_0, y_0), $x_0 > 0$, $y_0 > 0$, inside the first quadrant) and belong to the family of curves. The ellipses and hyperbolas of the same family cut each other orthogonally. Thus parameters corresponding to ellipses and hyperbolas belonging to a family of confocal central conics define an orthogonal curvilinear coordinate system, called an †elliptic coordinate system (→ 90 Coordinates).

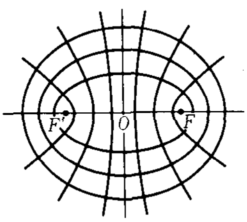


Fig. 8

The set of parabolas having a fixed point F as focus and a straight line passing through F as axis is called a **family of confocal parabolas** (Fig. 9). The family of confocal parabolas containing $y^2 = 4ax$ is the set of curves

$$y^2 = 4(a+\lambda)(x+\lambda).$$

Such families also give rise to orthogonal curvilinear coordinate systems.

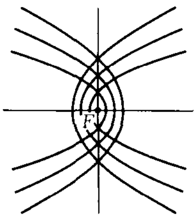


Fig. 9

I. Curves of the Second Order

With respect to a rectangular coordinate system, a curve represented by an equation with real coefficients of the second degree with two variables x, y,

$$ax^2 + 2hxy + by^2 + 2gx + 2fy + c = 0, \tag{5}$$

where $(a, h, b) \neq (0, 0, 0)$, is called a **curve of the second order**. A curve of the second order is either an empty set, one point, one or two straight lines, or a conic section. For equation (5), we put

$$D_0 = \begin{vmatrix} a & h \\ h & b \end{vmatrix}, \qquad D = \begin{vmatrix} a & h & g \\ h & b & f \\ g & f & c \end{vmatrix} \tag{6}$$

and call D the **discriminant** of the curve of the second order. If $D_0 \neq 0$, $D \neq 0$, and the curve is not an empty set, then the curve is a central conic. If $D_0 > 0$, then the curve is an ellipse or an empty set. If $D_0 < 0$, then the curve is a hyperbola. If $D_0 = 0$, $D \neq 0$, then the curve is a parabola. If $D = 0$, $D_0 > 0$, then the curve consists of one point. If $D = 0$, $D_0 < 0$, then the curve is two intersecting straight lines. If $D = D_0 = 0$, then the curve is an empty set, one straight line, or two parallel straight lines.

J. Poles and Polars

Let $F(x, y) = ax^2 + 2hxy + by^2 + 2gx + 2fy + c = 0$ be the equation of a conic C and (x_0, y_0) the coordinates of a point P on the plane. A straight line P^* having the equation

$$ax_0x + h(x_0y + xy_0) + by_0y + g(x + x_0) + f(y + y_0) + c = 0$$

is called the **polar** of P with respect to C (Fig. 10). When the polar of a point P is l, we call P the **pole** of l and denote it by l^*. In general, l^* is uniquely determined by l, and $P^{**} = P$, $l^{**} = l$. If $Q \in P^*$, then $P \in Q^*$. If $P' \in l$, then $l^* \in P'^*$. When a straight line passing through P intersects C at X, Y and intersects P^* at P', then P, P' are †harmonic conjugate with respect to X, Y. In particular, if $P \in P^*$, then $P \in C$, and P^* becomes the tangent of C at P. Given a

triangle $\triangle PQR$ on the plane of C, we call the triangle with sides P^*, Q^*, R^* the **polar triangle** of $\triangle PQR$. Let $Q^* \cap R^* = P'$, $R^* \cap P^* = Q'$, and $P^* \cap Q^* = R'$. Then the three straight lines $P \cup P'$, $Q \cup Q'$, $R \cup R'$ meet at a point (M. Chasles). When the polar triangle of $\triangle PQR$ coincides with itself, then $\triangle PQR$ is called a **self-polar triangle**. The polar of a focus is a directrix.

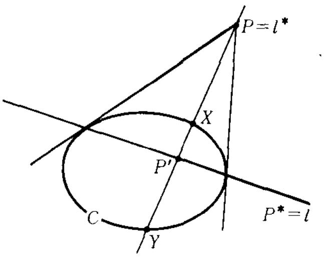


Fig. 10

K. Curves of the Second Class

When the coefficients u, v, w of a straight line $ux + vy + w = 0$ satisfy an equation with real coefficients of the second order,

$$Au^2 + 2Huv + Bv^2 + 2Guw + 2Fvw + Cw^2 = 0, \qquad (5')$$

where $(A, H, B) \neq (0, 0, 0)$, the curve enveloped by these straight lines is called **a curve of the second class**. Let Δ be the discriminant defined analogously to D in (6) by using $A, H, B, \ldots$ instead of $a, h, b, \ldots$. A curve of the second class (with $\Delta \neq 0$) is essentially the same as a curve of the second order with $D \neq 0$. In order for the curve (5′) with $\Delta \neq 0$ to coincide with the curve (5) with $D \neq 0$, it is necessary and sufficient that A, B, C, F, G, H be proportional to the †cofactors of a, b, c, f, g, h in the determinant D given by (6). If $\Delta = 0$, then (5′) represents either the empty set, or a point (regarded as the set of straight lines passing through the point), or two points.

From a projective point of view, a curve of the second order is defined as a locus of the point of intersection $l \cap l' = X$ of corresponding lines l and l' when two †pencils of lines $A(l, m, \ldots)$, $A'(l', m', \ldots)$ passing through two different centers A and A' are in correspondence under a †projective mapping f (J. Steiner) (Fig. 11). From this it can be proved that three points of intersection of three pairs of opposite sides (AB, DE), (BC, EF), (CD, FA) of a hexagon inscribed in a curve of the second order are on the same straight line (**Pascal's theorem**, Fig. 12). In particular, if the curve of the second order in this theorem consists of two straight lines, the theorem coincides with **Pappus's theorem** (Fig. 13). We call this straight line l the **Pascal line** of $ABCDEF$. Given a set of six points A, B, C, D, E, F on a curve of the second order, by considering all possible combinations of the points, we get 60 Pascal lines. A configuration consisting of these 60 lines is called **Pascal's configuration**, and has been studied by Steiner, Kirkman, and others. As a †dual to Pascal's theorem, **Brianchon's theorem** holds: Three diagonals of a hexagon with a curve of the second class inscribed meet at a point (Fig. 14).

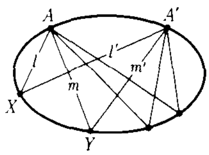


Fig. 11

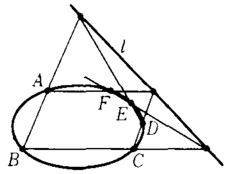


Fig. 12

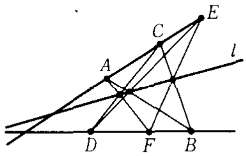


Fig. 13

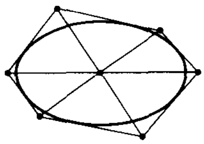


Fig. 14

References

[1] G. Salmon, A treatise on conic sections, Longmans, Green, sixth edition, 1879 (Chelsea 1962).
[2] H. F. Baker, Principles of geometry II. Plane geometry, Cambridge Univ. Press, 1922.
[3] M. Protter and C. Morrey, Jr., Calculus with analytic geometry, Addison-Wesley, 1962.

79 (II.20) Connectedness

A. General Remarks

A †topological space X is said to be **connected** if there are no proper closed subsets A and B of X such that $A \cap B = \varnothing$ and $A \cup B = X$ (C. Jordan, *Cours d'Analyse* I, 1983). A subset S of X is **connected** if S considered as a †subspace of X is connected.

If a subset S of X is connected, then the †closure $\bar{S}$ is also connected. Let $\{A_\alpha\}$ be a family of connected subsets of X such that either $\bar{A}_\alpha \cap A_\beta \neq \varnothing$ or $A_\alpha \cap \bar{A}_\beta \neq \varnothing$ for any pair A_α and A_β. Then the union $\bigcup_\alpha A_\alpha$ is connected. The continuous image of a connected set is connected. The †product space $\prod_\alpha X_\alpha$ of a family of connected spaces $\{X_\alpha\}$ is also con-

nected. Let $\{A_\alpha\}$ be a family of connected subsets of X and A_0 be a connected subset of X. If $A_0 \cap A_\alpha \neq \varnothing$ for every A_α, then the union $A_0 \cup \bigcup_\alpha A_\alpha$ is also connected. For a point p of a topological space X, the union of all connected subsets containing p is connected and is called the **connected component** of p (F. Hausdorff, 1927).

The n-dimensional Euclidean space E^n $(n \geqslant 0)$ and the n-dimensional unit sphere S^n $(n \geqslant 1)$ are connected, whereas S^0, consisting of two points, is not connected. $E^n \backslash \{0\}$ is connected for $n \geqslant 2$, whereas $E^1 \backslash \{0\}$ is not connected. This fact implies that E^1 is not homeomorphic to E^n $(n \geqslant 2)$. A connected open subset of a topological space X is called a **domain** (or **region**) in X.

A topological space X is said to be **locally connected** at a point p if for every open set U containing p, there is an open set V containing p and contained in the connected component of p in U. X is said to be **locally connected** if it is locally connected at each point of X. A space X is locally connected if and only if every connected component of every open subset is open in X.

There are connected spaces that are not locally connected. For example, the **comb space** $\{(x, y) \in \mathbf{R}^2 \mid x = 1/n$ and $0 \leqslant y \leqslant 1$ $(n = 1, 2, 3, \dots)$ or $0 \leqslant x \leqslant 1$ and $y = 0\}$ is not locally connected at $(0, 1)$ (see Fig. 1).

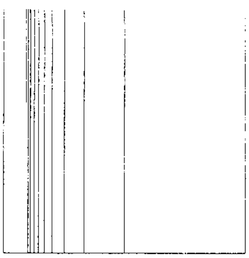


Fig. 1
Comb space.

B. Arcwise Connectedness

Two points a and b of a topological space X are said to be **joined by an arc** in X if there is a continuous map $f(t)$ of the closed interval $I = [0, 1]$ into X such that $f(0) = a$ and $f(1) = b$. A topological space X is said to be **arcwise connected** (or **path-connected**) if every two points of X are joined by an arc in X. For a point p of X, the union of all arcwise-connected subsets in X containing p is arcwise connected and is called the **arcwise-connected component** (or **path-component**) of p. A topological space X is said to be **locally arcwise connected** at a point p if for every open subset U containing p, there is an open subset V containing p such that every two points in V are joined by an arc in U. X is said to be **locally arcwise connected** if it is locally arcwise connected at each point of X.

An arcwise-connected space is connected, but the converse is not true. The †sinusoid $\{(x, y) \in \mathbf{R}^2 \mid y = \sin 1/x$ and $0 < x \leqslant 1$ or $x = 0$ and $-1 \leqslant y \leqslant 1\}$ is connected but not arcwise connected. A complete metric space is arcwise connected if it is connected and locally connected.

C. Simple Connectedness and n-Connectedness

Denote by S^n the n-dimensional unit sphere and by D^{n+1} the $(n+1)$-dimensional unit disk. A topological space X is said to be **n-connected** if every continuous map f from S^m to X is extendable over D^{m+1} $(m = 0, 1, 2, \dots, n)$. 0-Connectedness is equivalent to arcwise connectedness. A 1-connected space is also called **simply connected**. (For simple connectedness and n-ply connectedness of plane domains → 333 Plane Domains) A topological space X is said to be **locally n-connected** at a point p if for every open subset U containing p, there is an open subset V containing p and contained in U such that any continuous map $f: S^m \to V$ is extendable to a continuous map $\tilde{f}: D^{m+1} \to U$ $(m = 0, 1, 2, \dots, n)$. X is said to be **locally n-connected** if it is locally n-connected at each point of X. A space X is said to be **ω-connected** (or **locally ω-connected**) if X is n-connected (or locally n-connected) for every n.

S^n is $(n-1)$-connected but not n-connected. Similarly for $E^{n+1} \backslash \{0\}$. If X is n-connected $(n \geqslant 1)$, then the †suspension SX is $(n+1)$-connected and the †loop space ΩX is $(n-1)$-connected. The **Hawaiian earring** $\{(x, y) \in \mathbf{R}^2 \mid (x - 1/n)^2 + y^2 = 1/n^2$ $(n = 1, 2, 3, \dots)\}$ is arcwise connected but not locally simply connected. The †cone over the Hawaiian earring is simply connected, but not locally simply connected (Fig. 2).

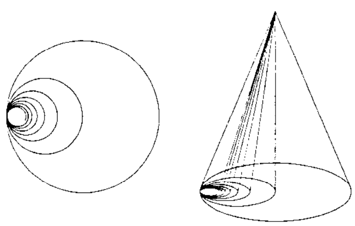


Fig. 2
Hawaiian earring and its cone.

A topological space X is said to be **contractible** if the identity map 1_X is †homotopic to a constant map c_{x_0} to a point x_0 of X (K. Borsuk, *Fund. Math.*, 24, 1935). A topological space X is said to be **locally contractible** at a

point p of X, if for every open subset U containing p, there is an open subset V containing p and contained in U such that the inclusion $i: V \subset U$ is homotopic to a constant map to a point in U. X is said to be **locally contractible** if it is locally contractible at each point of X.

A contractible or locally contractible space is ω-connected or locally ω-connected, respectively. The n-dimensional simplex, the n-dimensional disk D^n, and the n-dimensional Euclidean space E^n are contractible. More generally, a †convex set in E^n is contractible. †Topological manifolds and †polyhedra are locally contractible.

D. Continua and Discontinua

A topological space is said to be **totally disconnected** if each connected component consists of one point. Divide the closed interval $I = [0, 1]$ into three equal parts, and let I_{11} and I_{12} be the closed intervals obtained from I by removing the middle open interval. As the next step, divide I_{11} and I_{12} into three equal parts, respectively, and remove the middle open intervals. Inductively we obtain 2^{n+1} closed intervals $I_{n+1,i}$ $(i=1, \dots, 2^{n+1})$ from 2^n closed intervals $I_{n,j}$ by removing open intervals lying in their middles. Let $C^{(n)} = \bigcup_{i=1}^{2^n} I_{n,i}$ and $C = \bigcap_{n=1}^{\infty} C^{(n)}$. Then C is called the **Cantor discontinuum** or simply the **Cantor set** or the **ternary set** (G. Cantor, *Math. Ann.*, 21 (1883)) (Fig. 3).



Fig. 3
Cantor discontinuum.

C is a subset of I consisting of points with coordinates $t = (n_1/3) + (n_2/3^2) + \dots + (n_i/3^i) + \dots$, where $n_i = 0$ or 2. As a topological space C is homeomorphic to the Cartesian product of countably many copies of the discrete space $D = \{0, 1\}$. C has the power of the continuum $\mathfrak{c}$ and is a compact, totally disconnected, †perfect set. The Cartesian product of infinitely many (countably or not) copies of D is called the **general Cantor set**. The continuous image of a general Cantor set is called a **dyadic compactum**. Compact metric spaces and compact groups are examples of dyadic compacta.

A **continuum** is by definition a connected compact metric space consisting of more than one point. A metric space X is said to be **well-chained** if for every two points a, b and $\varepsilon > 0$ there are points $x_1, x_2, \dots, x_{n-1}$ such that $d(x_i, x_{i+1}) < \varepsilon$ $(x_0 = a, x_n = b)$. A well-chained compact metric space is a continum.

Let K be a continuum containing two points a and b. Then K is said to be **irreducible** between a and b if there is no proper subcontinuum of K containing a and b (L. Zoretti, *Ann. Sci. Ecole Norm. Sup.*, 26 (1909)) (→ 93 Curves).

A continuum K is said to be **indecomposable** if there are no proper subcontinua K_1, K_2 such that $K = K_1 \cup K_2$ (L. E. J. Brouwer, *Math. Ann.*, 66 (1910)). Simple examples of indecomposable continua have been given by A. Denjoy, *C. R. Acad.* Sci. Paris, 151 (1910); K. Yoneyama, *Tôhoku Math. J.*, 12 (1917); and B. Knaster, *Fund. Math.*, 3 (1922).

C. Kuratowski conjectured the following: If a plane continuum K is homogeneous (that is, for any two points $a, b \in K$ there exists a homeomorphism $h: K \to K$ with $h(a) = b$) then K is homeomorphic to the circle. A counterexample for this conjecture has been found by R. H. Bing and E. E. Moise. It is an indecomposable, homogeneous, plane continuum and is called the **pseudo-arc** [8].

References

[1] K. Kuratowski, Topologie II, Monograf. Mat. (1950).
[2] S. Lefschetz, Topics in topology, Ann. Math. Studies, Princeton Univ. Press, 1942.
[3] R. L. Wilder, Topology of manifolds, Amer. Math. Soc. Colloq. Publ., 1949.
[4] G. T. Whyburn, Analytic topology, Amer. Math. Soc. Colloq. Publ., 1942.
[5] S. T. Hu, Elements of general topology, Holden-Day, 1964.
[6] O. Hanner, Retraction and extension of mappings of metric and non-metric spaces, Ark. Mat., 2 (1952), 315–360.
[7] K. Borsuk, Theory of retracts, Monograf. Mat. (1967).
[8] J. G. Hocking and G. S. Young, Topology, Addison-Wesley, 1961.
Also → references to 425 Topological Spaces.

80 (VII.4) Connections

A. History

The geometric notion of connections originated with T. Levi-Civita's parallelism (*Rend. Circ. Mat. Palermo*, 42 (1917)) and was later generalized to the notion of connections of differentiable fiber bundles. Notions such as affine connections, Riemannian connections, projective connections, and conformal connections can be described in terms of bundles

constructed from the tangent bundles of differentiable manifolds. They are also standard examples of the Cartan connections formulated by E. Cartan and C. Ehresmann.

B. Connections in Principal Bundles

Let $P=(P,\pi,M,G)$ be a differentiable †principal fiber bundle. (For the sake of convenience, we assume that differentiability always means that of class C^∞.) The total space P and the base space M are †differentiable manifolds, and the projection π is a differentiable mapping. The †structure group G is a †Lie group and acts on P from the right as a transformation group. On each fiber, G acts transitively without fixed points. For elements a, x in G, P, we write $R_a(x)=xa$. The mappings induced on †tangent vector spaces by R_a and π will be denoted by the same letters, namely $R_a: T_x(P) \rightarrow T_{xa}(P)$, $\pi: T_x(P) \rightarrow T_{\pi(x)}(M)$. The tangent vector space $T_x(P)$ at each point x of P is mapped by the projection π onto the tangent vector space $T_p(M)$ at the point $p=\pi(x)$ of M. The kernel of this mapping is denoted by $V_x(P)$, and each vector in $V_x(P)$ is said to be **vertical**. The kernel $V_x(P)$ is the totality of elements of $T_x(P)$ that are tangent to the fiber.

C. Connections

We say that a **connection** is given in P if for each point $x \in P$, a subspace Q_x of the tangent space $T_x(P)$ is given in such a way that the following three conditions are satisfied: (i) $T_x(P)=V_x(P)+Q_x$ (direct sum); (ii) $R_a(Q_x)=Q_{xa}$ (Q is invariant under G); and (iii) the mapping $x \rightarrow Q_x$ is differentiable. A vector in Q_x is said to be **horizontal**.

Now suppose that X is an arbitrary †vector field on P. By condition (i), the value X_x of X at each point x of P can be expressed uniquely as $X_x=Y_x+Z_x$, where $Y_x \in V_x(P)$ and $Z_x \in Q_x$. The vector fields Y and Z defined by Y_x and Z_x ($x \in P$) are called the **vertical** and **horizontal components** of X, respectively. Condition (iii) implies that if X is a differentiable vector field, then its horizontal and vertical components are also differentiable vector fields. Let X be a vector field on the base space M. Since π defines an isomorphism of Q_x and $T_p(M)$ ($p=\pi(x)$), we have a unique vector field X^* on P such that (a) $\pi(X^*)=X$ and (b) $X^*_x \in Q_x$. We call X^* the **lift** of X, and it is invariant under G by condition (ii).

Suppose that a connection is given in P. If C is a piecewise differentiable curve in the base space M, we can define a mapping φ that maps the fiber over the initial point p of C onto the fiber over the endpoint q of C as follows: Take an arbitrary point x on the fiber at p. Then we have a unique curve C^*_x in P starting at x such that (a) $\pi(C^*_x)=C$, and (b) each tangent vector to C^*_x is horizontal. (C^*_x is called a **lift** of C that starts at x.) The endpoint y of the curve C^*_x belongs to the fiber over q. We set $\varphi(x)=y$. Because $C^*_{xa}=R_a(C^*_x)$, the mapping φ commutes with transformations of G. We call this mapping φ the **parallel displacement** or **parallel translation** along the curve C.

D. Holonomy Groups

Fix a point p in the base space. If C is a closed curve in M starting from p, the parallel displacement along C maps the fiber over p onto itself. So if we fix a point x on the fiber over p, x is transformed by the parallel displacement to a point xa ($a \in G$). Thus each closed curve C starting from p determines an element $a(x,C)$ of G. If C varies over the set of closed curves that start from p, the totality of such elements of G forms a subgroup of G. This subgroup is called the **holonomy group** of the connection defined over P with the reference point x. If M is connected, holonomy groups with different reference points are conjugate. In the above, if we choose as the closed curves C starting from p only those curves that are null-homotopic, the elements $a(x,C)$ form a subgroup of the holonomy group. This is called the **restricted holonomy group**. The holonomy group is a †Lie subgroup of the structure group, and its connected component containing the identity coincides with the restricted holonomy group. Holonomy groups are useful in the study of the behavior of connections.

E. Connection Forms

Let $\mathfrak{g}$ be the Lie algebra (→ 249 Lie Groups) of the structure group of G of a principal fiber bundle $P=(P,\pi,M,G)$. For each A in $\mathfrak{g}$, the 1-parameter subgroup $\exp tA$ $(-\infty<t<\infty)$ of G defines a †one-parameter group $R_{\exp tA}$ of transformations on P, and it determines a vector field A^* on P (→ 105 Differentiable Manifolds). Each element of the vector field A^* is vertical at each point x on P, and the A^* ($A \in \mathfrak{g}$) at x generate $V_x(P)$. Moreover, for each element a of G we have $R_a(A^*)=(\mathrm{ad}(a^{-1})A)^*$.

For a connection in P, we define the **connection form** ω on P with values in $\mathfrak{g}$ by the following: (i) $\omega_x(A^*_x)=A$ ($A \in \mathfrak{g}$), and (ii) $\omega_x(X)=0$ ($X \in Q_x$). The connection form ω thus defined satisfies (iii) $R^*_a(\omega)=\mathrm{ad}(a^{-1})\omega$ ($a \in G$), where $R^*_a(\omega)$ is the †differential form induced by the

transformation R_a from the differential form ω. Conversely, given a 1-form ω with values in $\mathfrak{g}$ that satisfies conditions (i) and (ii), we can define a connection in P by defining vectors X such that $\omega(X)=0$ as the horizontal, and its connection form coincides with ω. Thus giving a connection in P is equivalent to giving a connection form in P.

In particular, when a principal fiber bundle P is trivial, i.e., when $P=M\times G$, we can identify the tangent vector space $T_x(P)$ at a point $x=(p,g)$ of P with the direct sum of $T_p(M)$ and $T_g(G)$. If we set $Q_x=T_p(M)$, then Q defines a connection in $P=M\times G$. Such a connection is called **flat**. When a connection can always be expressed as above locally, it is called **locally flat**. Since each principal fiber bundle is locally a product fiber bundle, we see that locally there exists a connection. If the base space M is †paracompact, we can show the existence of connections on P.

F. Extension and Restriction of Connections

When a principal fiber bundle $P=(P,\pi,M,G)$ has a †reduced fiber bundle P', we shall consider the relation between the connections of P and of P'. Let G' be a Lie subgroup of G and $\mathfrak{g}'$ its Lie algebra. We shall denote by j both the injection of G' into G and also the injection of $\mathfrak{g}'$ into $\mathfrak{g}$. If there exist a differentiable principal fiber bundle $P'=(P',\pi',M,G')$ and a differentiable embedding f of P' into P such that $\pi\circ f=\pi'$ and $f\circ R_a=R_{j(a)}\circ f\,(a\in G')$ are satisfied, then (P',f) is said to be a reduced fiber bundle of P. Then we have $f_*(A_x^*)=j(A)_{f(x)}^*$ for each $A\in\mathfrak{g}'$ and $x\in P'$.

Suppose that a connection is given in P'; we denote the horizontal space at the point x of P' by Q_x'. At the point $f(x)$ of P, we take $f_*(Q_x')$ as the horizontal space and transform it by right translations of G. Thus we obtain a connection on P. Let ω' and ω be the corresponding connection forms. Then we have $j\circ\omega'=f^*(\omega)$ on P'. Conversely, suppose that we are given a connection in P with the connection form ω. If the induced form $f^*(\omega)$ on P' has values always in $j(\mathfrak{g}')$, we can write $f^*(\omega)=j\circ\omega'$, and ω' defines a connection in P'. In this case the connection in P is called an **extension** of the connection in P', and the connection in P' is called the **restriction** of the connection in P.

G. Curvature Forms

Suppose that a principal fiber bundle $P=(P,\pi,M,G)$ has a connection. Let F be a finite-dimensional vector space and α be a differential form of degree k on P with values in F. We define the **covariant differential** $D\alpha$ of α by

$$(D\alpha)(X_1,\dots,X_{k+1})=(d\alpha)(hX_1,\dots,hX_{k+1}),$$

where the X_i are vector fields on P and h denotes the projection to the horizontal component. $D\alpha$ is a differential form of degree $k+1$ on P with values in F.

Let $\rho: G\rightarrow GL(F)$ be a †representation of a Lie group G on F. A differential form α on P with values in F is called a **pseudotensorial form** of type ρ if α satisfies $R_a^*(\alpha)=\rho(a^{-1})\alpha$ $(a\in G)$. In particular, if a pseudotensorial form α satisfies $\iota(A^*)\alpha=0$ for any $A\in\mathfrak{g}$ (→ 105 Differentiable Manifolds Q), it is called a **tensorial form** of type ρ. For each representation ρ of G, we can construct an associated vector bundle E over M with fiber F. A tensorial form of type ρ is identified with a differential form on M with values in E. If α is a pseudotensorial form of type ρ, then $D\alpha$ is a tensorial form of type ρ.

For a connection form ω on P, the covariant differential $D\omega=\Omega$ of ω is called the **curvature form** of the connection. Since ω is a pseudotensorial form of type ad, Ω is a tensorial form of type ad. For the connection form we have the **structure equation** $d\omega=-[\omega,\omega]+\Omega$ [4, 6]. Let X and Y be vector fields on M, and let X^* and Y^* be their lifts, respectively. Then we have $\omega([X^*,Y^*])=\Omega(X^*,Y^*)$, which shows that the curvature form Ω for X^*, Y^* gives the vertical component of $[X^*,Y^*]$.

It is known that the following three conditions for a connection are equivalent: (i) The connection is locally flat. (ii) The curvature form vanishes. (iii) The restricted holonomy group is trivial (i.e., the identity group).

The following two theorems are fundamental:

(1) Suppose that a connection is given in a principal fiber bundle $P=(P,\pi,M,G)$. Then the structure group of P can be reduced to the holonomy group [4, 6]. In fact, for $x\in P$, let $P(x)$ be the set of points y in P that can be connected to x by a piecewise horizontal curve in P. Then $P(x)$ gives a reduced fiber bundle of P, and the connection in P is an extension of a connection in $P(x)$ [4, 6].

(2) The Lie algebra of the holonomy group with a reference point x in P coincides with the vector subspace of $\mathfrak{g}$ spanned by $\{\Omega_y(X,Y)\mid y\in P(x), X,Y\in T_y(P)\}$ [4, 6].

The curvature form Ω is used to express the †characteristic classes of the bundle P [1, 2] (→ 56 Characteristic Classes).

In some cases, a connection in the principal fiber bundle induces a connection in an †associated fiber bundle. In particular, when G is $GL(n,\mathbf{R})$ or $GL(n,\mathbf{C})$, we can define a connection in any associated vector bundle. The notion of connections in vector bundles can be

defined more algebraically (M. F. Atiyah, *Trans. Amer. Math. Soc.*, 85 (1957)) and can also be defined as a kind of differential operator on vector bundles [8].

H. Affine Connections

Let M be a differentiable manifold of dimension n and P be the †bundle of tangent n-frames over M. Then P has the structure group $GL(n, \mathbf{R})$, and it is the principal bundle associated with the tangent vector bundle of M, which consists of all tangent vectors of M. A connection in the bundle of tangent n-frames is called an **affine connection** (or **linear connection**) on M. An affine connection on M defines (as well as the curvature form Ω) a new form Θ called the torsion form on P, which is given as follows: Let F be an n-dimensional vector space with a fixed basis $(\xi_1, \xi_2, \ldots, \xi_n)$. Since the bundle of tangent n-frames P is the set of all bases (i.e., n-frames) $(e_1, \ldots, e_n)$ in $T_p(M)$ at each point p of M, every point $x=(e_1, \ldots, e_n)$ of P is given as a mapping $\bar{x}$ of F onto $T_p(M)$ $(p=\pi(x))$ defined by $\bar{x}: \xi_i \rightarrow e_i$. We define differential form θ of degree 1 with values in F on P by $\theta_x(X)=\bar{x}^{-1}(\pi_x(X))$ $(X \in T_x(P))$. θ is called a **canonical 1-form** of the bundle of tangent n-frames of the manifold M and has the following property: Any diffeomorphism φ of M onto itself induces a bundle automorphism $\tilde{\varphi}$ of P onto itself, and $\tilde{\varphi}$ preserves θ, that is, $\tilde{\varphi}^*(\theta)=\theta$. Conversely, we can show that any bundle automorphism of P that preserves θ is induced by a diffeomorphism of the base space M.

For an affine connection on M, we define the **torsion form** Θ by $\Theta=D\theta$. Θ is a differential form of degree 2 on P with values in F and satisfies $R_a\Theta=a^{-1}\cdot\Theta$ $(a \in GL(n, \mathbf{R}))$. Furthermore, we have the **structure equation** for Θ, $d\theta=[\omega, \theta]+\Theta$ [2, 4, 6].

For each element ξ in F, there exists a unique horizontal vector field $B(\xi)$ on P such that $\theta(B(\xi))=\xi$. $B(\xi)$ is called the **basic vector field** corresponding to ξ. At each point $x \in P$, $B(\xi_1)_x, \ldots, B(\xi_n)_x$ form a basis of Q_x. Let $\{A_1, \ldots, A_m\}$ $(m=n^2)$ be a basis of $\mathfrak{g}=\mathfrak{gl}(n, \mathbf{R})$. Then at each point $x \in P$, $\{(A_1^*)_x, \ldots, (A_m^*)_x, B(\xi_1)_x, \ldots, B(\xi_n)_x\}$ is a basis of the tangent vector space $T_x(P)$. Thus the bundle P of frames is a †parallelizable manifold. The projection to M of any †integral curve of a basic vector field is a geodesic, which is defined in Section I [4].

An affine connection on M gives a parallel displacement of the tangent vector space of M as follows: Let $C=p_t$ $(0 \leqslant t \leqslant 1)$ be a curve in M and $C^*=x_t$ be a lift of C to P. The parallel displacement of the tangent n-frame x_0 at p_0 along the curve C is x_t, and the mapping $\bar{x}_t \circ \bar{x}_0^{-1}: T_{p_0}(M) \rightarrow T_{p_t}(M)$ is called the **parallel displacement** of the tangent space $T_{p_0}(M)$ onto $T_{p_t}(M)$ along C. It is easily seen that the mapping is independent of the choice of lifts.

I. Covariant Differentials

Let $C=\{p_t\}$ $(0 \leqslant t \leqslant 1)$ be a differentiable curve in M. If we have a vector Y_t in $T_{p_t}(M)$ for each t and the correspondence $t \rightarrow Y_t$ is differentiable, then $\{Y_t\}$ is called a vector field along the curve C. For $\{Y_t\}$ we set

$$Y_t'=\lim_{h \rightarrow 0}(1/t)(\varphi_{t,h}^{-1}(Y_{t+h})-Y_t),$$

where $\varphi_{t,h}$ is the parallel displacement of $T_{p_t}(M)$ onto $T_{p_{t+h}}(M)$ along the curve C. The vector field $\{Y_t'\}$ along C thus obtained is called the **covariant derivative** of $\{Y_t\}$. $\{Y_t\}$ is parallel along C; that is, $Y_t=\varphi_{0,t}(Y_0)$ if and only if $Y_t' \equiv 0$. In particular, if the tangent vectors to a curve C are parallel along C itself, then C is said to be a **geodesic**.

Let X and Y be vector fields on a manifold M with an affine connection. The **covariant derivative** $\nabla_X Y$ of the vector field Y in the direction of the vector field X is defined as follows: Let p_0 be a point in M, $C=\{p_t\}$ $(-\varepsilon \leqslant t \leqslant \varepsilon)$ be an integral curve of X through p_0, and $\{\varphi_t\}$ be the parallel displacement along C. We set

$$(\nabla_X Y)_{p_0}=\lim_{t \rightarrow 0}(1/t)(\varphi_t^{-1}(Y_{p_t})-Y_{p_0}).$$

Then $\nabla_X Y$ is also a vector field on M.

The mapping $(X, Y) \rightarrow \nabla_X Y$ satisfies the following three conditions: (i) $\nabla_X Y$ is linear with respect to X and Y; (ii) $\nabla_{fX} Y=f\cdot\nabla_X Y$; and (iii) $\nabla_X(fY)=f\cdot\nabla_X Y+(Xf)\cdot Y$, where f is a differentiable function on M. Conversely, if a mapping satisfying conditions (i)–(iii) above is given, then there exists a unique affine connection on M whose covariant derivative coincides with the given mapping [4, 6].

Fix a vector field Y. Then the mapping $X \rightarrow \nabla_X Y$ defines a †tensor field of type (1, 1). This tensor field is called the **covariant differential** of Y and is denoted by ∇Y. Now fix a vector field X. Then the mapping $Y \rightarrow \nabla_X Y$ can be naturally extended to tensor fields of arbitrary type, and it commutes with the †contraction of the tensors. For a tensor field K this is denoted by $K \rightarrow \nabla_X K$. Furthermore, the mapping $X \rightarrow \nabla_X K$ is called the **covariant differential** of K and is denoted by ∇K. We call $\nabla_X K$ the **covariant derivative** of K in the direction of X. A tensor field K is invariant under parallel displacements if and only if $\nabla K=0$ (→ 417 Tensor Calculus).

J. Curvature Tensors and Torsion Tensors

For an affine connection on M, the **curvature tensor** R and the **torsion tensor** T are defined by

$$R(X, Y)(Z)=\nabla_X(\nabla_Y Z)-\nabla_Y(\nabla_X Z)-\nabla_{[X,Y]}(Z),$$

$$T(X\cdot Y)=\nabla_X Y-\nabla_Y X-[X, Y],$$

where X, Y, and Z are vector fields on M, and R and T are tensors of types (1,3) and (1,2), respectively. Also, in terms of the curvature form Ω and the torsion form Θ on the bundle of tangent n-frames P over M, they can be defined by

$$R_p(X, Y)(Z)=\bar{x}^{-1}\cdot\Omega_x(X^*, Y^*)\cdot\bar{x}(Z),$$

$$T_p(X, Y)=\bar{x}^{-1}(\Theta_x(X^*, Y^*)),$$

where $\pi(x)=p$, X, $Y\in T_p(M)$, and X^*, Y^* are lifts of X, Y, respectively. The curvature tensor and the torsion tensor satisfy the relations $R(X, Y)=-R(Y,X)$, $T(X, Y)=-T(Y,X)$. Moreover, **Bianchi's identities** hold:

$$\mathfrak{S}(R(X, Y)(Z))=\mathfrak{S}(T(T(X, Y),Z)+(\nabla_X T)(Y,Z))$$

and

$$\mathfrak{S}((\nabla_X R)(Y,Z)+R(T(X, Y),Z))=0,$$

where $\mathfrak{S}$ denotes the sum of terms that are obtained by cyclic permutations of X, Y, Z [4]. For instance, in the case of a Riemannian connection (Section K), we have $T=0$, and Bianchi's identities reduce to

$$R(X, Y)(Z)+R(Y,Z)(X)+R(Z,X)(Y)=0,$$

$$(\nabla_X R)(Y,Z)+(\nabla_Y R)(Z,X)+(\nabla_Z R)(X, Y)=0.$$

We now consider a system of coordinates $(x^1,\dots,x^n)$ in an n-dimensional linear space $M=\mathbf{R}^n$. The vector fields $(X_1,\dots,X_n)$ $(X_i=\partial/\partial x^i)$ form a basis for vector fields on M. If we set $\nabla_{X_i}X_j=0$, we get an affine connection on $\mathbf{R}^n$. For such a connection we have $R=0$, $T=0$, and any straight line in $\mathbf{R}^n$ is a geodesic with respect to this connection. The connection is called the **canonical affine connection** on $\mathbf{R}^n$. An affine connection on a manifold M satisfies $R=0$ and $T=0$ if and only if the connection on M is locally isomorphic to the canonical affine connection of $\mathbf{R}^n$.

Let φ be a diffeomorphism of a manifold M with an affine connection onto itself. We call φ an **affine transformation** of M if the induced automorphism $\tilde{\varphi}$ on the bundle of tangent n-frames preserves the connection. In terms of covariant differentials, this condition is equivalent to the condition $\nabla_{\varphi(X)}\varphi(Y)=\varphi(\nabla_X Y)$ for any vector fields X, Y. An affine transformation of the canonical affine connection on $\mathbf{R}^n$ is an ordinary affine transformation. For an affine connection on a manifold M, the set of all affine transformations forms a Lie group and acts on M as a Lie transformation group [4,5].

Let M be a manifold with an affine connection. M is called an affine **locally symmetric space** if $\nabla R=0$ and $\nabla T=0$ are satisfied. These conditions are satisfied if and only if at each point p of M, there exist a neighborhood U of p and an affine transformation φ of U such that $\varphi^2=1$ and p is the isolated fixed point of φ. If for each point p of M there exists an affine transformation of M such that $\varphi^2=1$ and p is an isolated fixed point, then M is called an **affine symmetric space**. A symmetric Riemannian space is a special case of this type (→ 413 Symmetric Spaces).

At each point p in a manifold M with an affine connection, we can choose local coordinates $(x^1,\dots,x^n)$ such that $x^i(p)=0$ and the curve $x^i=a^i t$ $(-\delta<t<\delta)$ is a geodesic for each $(a^1,\dots,a^n)$ with $\Sigma(a^i)^2=1$. Such local coordinates are called **geodesic coordinates** at p [4]. With respect to geodesic coordinates, we have $(\nabla_{X_i}(X_j))_p=0$ $(\partial/\partial x^i=X_i)$.

K. Riemannian Connections

When a Riemannian metric g (→ 364 Riemannian Manifolds) is given on a manifold M, it defines a metric on the tangent space $T_p(M)$ at each point p of M, and we can take orthonormal bases in $T_p(M)$. The set P' of all orthonormal bases of tangent spaces is a subset of the bundle P of tangent n-frames of M and forms a subbundle of P; its structure group is the †orthogonal group $O(n)$, and P' gives a reduction of the bundle of tangent n-frames. Conversely, when a reduction of the bundle of frames to $O(n)$ is given, we can define a Riemannian metric on M such that the reduced bundle consists of all orthonormal frames.

For a Riemannian metric g on M, there exists a unique affine connection on M such that (i) $\nabla g=0$, and (ii) the torsion tensor T vanishes [4]. This connection is called the **Riemannian connection** corresponding to g. The first condition is equivalent to the invariance of the Riemannian metric g under parallel displacement. Thus the affine connection transforms orthonormal bases on M to orthonormal bases and induces a connection in the bundle P'. It is known that the restricted holonomy group of any Riemannian connection is a closed subgroup of $O(n)$ [4]. An affine connection on a manifold M is called a **metric**

connection if it preserves a Riemannian metric g on M, i.e., if it satisfies the condition (i).

L. Representations in Local Coordinates

(1) Let $(x^1, \ldots, x^n)$ be a local coordinate system in a manifold M and consider vector fields $X_i = \partial/\partial x_i$. For an affine connection on M, the covariant derivative can be expressed as

$$\nabla_{X_j}(X_k) = \sum_i \Gamma^i_{jk} X_i.$$

The Γ^i_{jk} are called **coefficients of the affine connection** with respect to the local coordinate system $(x^1, \ldots, x^n)$. We denote by $\bar{\Gamma}^i_{jk}$ the coefficients of connection with respect to another local coordinate system $(y^1, \ldots, y^n)$. Then on the intersection of their coordinate neighborhoods, we have

$$\bar{\Gamma}^i_{jk} = \sum_{\alpha,\beta,\gamma} \frac{\partial y^i}{\partial x^\alpha}\left(\frac{\partial x^\beta}{\partial y^j}\frac{\partial x^\gamma}{\partial y^k}\Gamma^\alpha_{\beta\gamma} + \frac{\partial^2 x^\alpha}{\partial y^j \partial y^k}\right).$$

Conversely, if the Γ^i_{jk} are given in each local coordinate system of M and satisfy this relation on each intersection of their coordinate neighborhoods, then there exists a unique affine connection such that the coefficients of the connection are given by Γ^i_{jk}.

(2) The **coefficients of the Riemannian connection** corresponding to a Riemannian metric $g = \sum g_{ij} dx^i dx^j$ on a manifold M are given by

$$\Gamma^i_{jk} = \frac{1}{2}\sum_l g^{il}\left\{\frac{\partial g_{jl}}{\partial x^k} + \frac{\partial g_{lk}}{\partial x^j} - \frac{\partial g_{jk}}{\partial x^l}\right\}$$

and are called the **Christoffel symbols**.

(3) With respect to each local coordinate system $(x^1, \ldots, x^n)$, we express the **torsion tensor** T and the **curvature tensor** R of an affine connection by

$$T = \sum_{ijk} T^i_{jk} dx^j \otimes dx^k \otimes X_i,$$

$$R = \sum_{ijkl} R^i_{jkl} dx^j \otimes dx^k \otimes dx^l \otimes X_i.$$

The components T^i_{jk} and R^i_{jkl} are given by

$$T^i_{jk} = \Gamma^i_{jk} - \Gamma^i_{kj},$$

$$R^i_{jkl} = (\partial\Gamma^i_{lj}/\partial x^k - \partial\Gamma^i_{kj}/\partial x^l) + \sum_m (\Gamma^m_{lj}\Gamma^i_{km} - \Gamma^m_{kj}\Gamma^i_{lm}).$$

(4) Let $K = (K^{i_1 \ldots i_r}_{j_1 \ldots j_s})$ be a tensor field of type (r, s). Then the **covariant differential** $\nabla K = (K^{i_1 \ldots i_r}_{j_1 \ldots j_s; k})$ is given by

$$K^{i_1 \ldots i_r}_{j_1 \ldots j_s; k} = \partial K^{i_1 \ldots i_r}_{j_1 \ldots j_s}/\partial x^k + \sum_{\alpha=1}^{r}\left(\sum_l \Gamma^{i_\alpha}_{kl} K^{i_1 \ldots l \ldots i_r}_{j_1 \ldots j_s}\right) - \sum_{\beta=1}^{s}\left(\sum_m \Gamma^m_{kj_\beta} K^{i_1 \ldots i_r}_{j_1 \ldots m \ldots j_s}\right).$$

(5) A curve $x^i = x^i(t)$ is a **geodesic** if and only if

$$\frac{d^2 x^i}{dt^2} + \sum_{j,k} \Gamma^i_{jk} \frac{dx^j}{dt}\frac{dx^k}{dt} = 0, \qquad i = 1, 2, \ldots, n$$

(→ 178 Geodesics, 417 Tensor Calculus).

M. Cartan Connections

Let M be a differentiable manifold of dimension n. Consider a homogeneous space $F = G/G'$ of the same dimension n, where G is a Lie group and G' is a closed subgroup of G (→ 199 Homogeneous Spaces). Let $B = (B, M, F, G)$ be a fiber bundle over M with fiber F and structure group G, and $P = (P, M, G)$ be the principal fiber bundle associated with B. Suppose that there exists a cross section f over M to B. Then the structure group of P can be reduced to G'. We denote this reduced fiber bundle by $P' = (P', M, G')$ and the injection of P' into P by j (→ 147 Fiber Bundles).

Suppose that a connection is given in P. Its connection form ω is a differential form of degree 1 on P with values in $\mathfrak{g}$, and the induced form $\omega' = j^*(\omega)$ is also a differential form of degree 1 on P' with values in $\mathfrak{g}$. We call the connection in P a **Cartan connection** on M with the fiber $F = G/G'$ if at each point x of P', ω'_x gives an isomorphism of $T_x(P')$ onto $\mathfrak{g}$ as linear spaces. Such a connection in P is equivalently defined as a 1-form ω' on P' with values in $\mathfrak{g}$ satisfying the following three conditions: (i) $\omega'(A^*) = A$ ($A \in \mathfrak{g}'$ (Lie algebra of G')); (ii) $R^*_a(\omega') = \mathrm{ad}(a^{-1})\omega'$ ($a \in G'$); and (iii) ω'_x gives an isomorphism of $T_x(P')$ onto $\mathfrak{g}$ at each point $x \in P'$. For such ω', we can take a connection form ω in P such that $\omega' = j^*(\omega)$; ω defines a Càrtan connection.

N. Soudures

A cross section f over M to B gives a vector bundle $T'(B)$ on M defined as follows: For each point p of M, the projection $B \to M$ defines a mapping $T_{f(p)}(B) \to T_p(M)$. The kernel of this mapping is denoted by $V_{f(p)}(B)$. Then $T'(B) = \bigcup_p V_{f(p)}(B)$ forms a vector bundle over M, and the dimension of its fibers is equal to $n = \dim F$.

A Cartan connection in P gives a bundle isomorphism between $T'(B)$ and the tangent vector bundle $T(M)$ of M as follows: Let x be an arbitrary point in P', and put $p = \pi(x)$. The projection $\pi: P' \to M$ induces an isomorphism of $T_x(P')/V_x(P')$ onto $T_p(M)$. On the other hand, ω'_x gives an isomorphism of $T_x(P')/V_x(P')$ onto $\mathfrak{g}/\mathfrak{g}'$. As a point in P', x gives a mapping of F

$= G/G'$ onto the fiber in B over p and sends the point $\{G'\}$ in F to $f(p)$. By this mapping, $T_0(F) = \mathfrak{g}/\mathfrak{g}'$ is mapped isomorphically onto $V_{f(p)}(B)$. Combining these isomorphisms, we get an isomorphism between $T_p(M)$ and $V_{f(p)}(B)$ that is independent of the choice of $x \in P'$ over p. The set of such isomorphisms for $p \in M$ defines a bundle isomorphism of $T(M)$ and $T'(B)$. If a fiber bundle B over M has an isomorphism such as above through a cross section, then B is said to have a **soudure** [3].

Conversely, if a fiber bundle B over M has a soudure with respect to a cross section f, then there exists a Cartan connection in P such that the soudure given by the connection coincides with the original one [3]. There are many Cartan connections in P with the given soudure. However, when $F = G/G'$ is a symmetric space of compact type such that G is noncompact and contains the identity component of the group of isometries, we can determine uniquely the so-called **normal** Cartan connection among the Cartan connections which gives rise to the given soudure [10].

For the tangent vector bundle $T(M)$ of M, the fiber F is an n-dimensional linear space and can be expressed as $F = G/G'$, where G is the †affine transformation group of F and $G' = GL(n, \mathbf{R})$. Then $T(M)$ has the 0-section over M, and there exists a natural soudure. Furthermore, an affine connection on M canonically induces a Cartan connection on M with the fiber $F = G/G'$ [3].

For a Cartan connection on M, we can introduce the notion of **development** of a curve in M into the fiber and also the notion of **completeness** [3].

O. Projective Connections

Let $F_1 = G_1/G_1'$ and $F_2 = G_2/G_2'$ be homogeneous spaces with $\dim F_1 = \dim F_2 = n$. Suppose that G_1 is a Lie subgroup of G_2 and G_1' is contained in G_2' by the injection. Then we have a canonical injection $F_1 \to F_2$ (F_1 is an open subset of F_2 by the assumption).

Suppose that a fiber bundle B_1 with fiber F_1 over M has a cross section f_1. Using f_1, we can construct a bundle B_2 with fiber F_2 over M which also has a cross section f_2. The principal bundle of B_2 is given by extending the structure groups from the principal bundle of B_1. We can show that if B_1 has a soudure with respect to the cross section f_1, then B_2 also has a soudure with respect to the cross section f_2. A Cartan connection in the principal fiber bundle P_1 associated with B_1 that is compatible with a given soudure on B_1 induces a Cartan connection in the principal fiber bundle P_2 associated with B_2, which then induces a soudure on B_2. The latter is called a Cartan connection **induced** from the former.

Let F_1 be an n-dimensional linear space and F_2 be the real †projective space of dimension n. Then the affine transformation group of F_1 can be embedded into the projective transformation group of F_2. Thus the tangent vector bundle of a manifold M induces a fiber bundle over M with the n-dimensional projective space as its fiber. A Cartan connection in this fiber bundle is called a **projective connection** on M. By the argument in this section, we see that every affine connection on M induces a projective connection on M.

Given two affine connections on M, we denote by ∇ and ∇' their corresponding covariant differentials. The two affine connections on M induce the same projective connection on M if and only if there exists a differential form φ of degree 1 on M such that $\nabla'_X Y - \nabla_X Y = \varphi(X)Y + \varphi(Y)X$ for any vector fields X, Y[7]. If a diffeomorphism φ of M preserves the projective connection induced by an affine connection in M, then φ maps geodesics of M into geodesics.

P. Conformal Connections

Let F_1 be an n-dimensional Euclidean space and F_2 an n-dimensional sphere (a †conformal space). We can embed the group of †isometries of F_1 canonically into the group of †conformal transformations of F_2. A Riemannian metric of the tangent vector bundle of a manifold M of dimension n gives a fiber bundle over M with fiber F_2. A Cartan connection in this fiber bundle is called a **conformal connection** on M; a Riemannian connection on M induces a conformal connection on M.

Two Riemannian metrics g_1, g_2 on M induce the same conformal connection on M if and only if there exists a positive function f on M such that $g_2 = fg_1$. Thus for a Riemannian manifold M with metric tensor g_1, a diffeomorphism φ of M such that $\varphi^*(g) = fg_1$ leaves invariant the conformal connection induced by g_1. Such a φ is called a **conformal transformation** of M with respect to the given Riemannian metric g_1.

For a Riemannian manifold M with metric tensor g, we define **Weyl's conformal curvature tensor** W by

$$W^i_{jkl} = R^i_{jkl} + \frac{1}{n-2}(R_{jk}\delta^i_l - R_{jl}\delta^i_k + g_{jk}R^i_l - g_{jl}R^i_k)$$

$$-\frac{R}{(n-1)(n-2)}(g_{jk}\delta^i_j - g_{jl}\delta^i_k),$$

where the R^i_{jkl} and R_{jk} are components of the curvature tensor and Ricci tensor, respectively,

and R is the scalar curvature (→ 364 Riemannian Manifolds, 417 Tensor Calculus). When $\dim M \geqslant 3$, the conformal connection induced by g on M is locally flat if and only if the conformal curvature tensor vanishes [7].

Q. Yang-Mills G-Connection

Let $P=(P,\pi,M,G)$ be a differentiable principal fiber bundle over a compact oriented Riemannian manifold M with group G. Then fiber bundles $G_P=P\times_c G$ and $\mathfrak{g}_P=P\times_{\mathrm{Ad}}\mathfrak{g}$ associated with P are induced naturally from the group conjugation $c: G\to \mathrm{Aut}(G)$ and the adjoint representation $\mathrm{Ad}: G\to \mathrm{Aut}(\mathfrak{g})$, respectively. A (local) section of G_P is called a (local) **gauge transformation** of P. The set of all global gauge transformations, which is denoted by $\mathscr{G}_P$, has a group structure.

A locally faithful representation ρ of G to an n-dimensional complex vector space F with a fixed basis $(\xi_1,\ldots,\xi_n)$ defines a differentiable complex vector bundle $E=P\times_\rho F$ associated with P. Every point x of P is identified with a linear mapping $\bar{x}$ of F onto the fiber $E_{\pi(x)}$ defined by $\bar{x}:\xi_i\to e_i$, where e_i denotes the equivalence class of $\{x,\xi_i\}\in P\times F$, $1\leqslant i\leqslant n$.

In a manner similar to the case of an affine connection, a connection in P with connection form ω gives a notion of parallel displacement of E as follows: Let $c=p_t$ $(0\leqslant t\leqslant 1)$ be a curve in M and $c^*=x_t$ be a lift of c to P. The mapping $\bar{x}_t\circ\bar{x}_0^{-1}: E_{p_0}\to E_{p_t}$ is called the parallel displacement of E_{p_0} onto E_{p_t} along c.

Let X be a vector field on M and φ be a differentiable section of E. The covariant derivative $\nabla_X\varphi$ of φ in the direction of X is defined as follows: Let p_0 be a point of M, $c=p_t$ $(-\varepsilon\leqslant t\leqslant -\varepsilon)$ be an integral curve of X through p_0 and $c^*=x_t$ be a lift of c to P. We set

$$(\nabla_X\varphi)_{p_0}=\lim_{t\to 0}\frac{1}{t}(\bar{x}_t\circ\bar{x}_0^{-1}(\varphi_{p_t})-\varphi_{p_0}).$$

Then ∇_X is also a differentiable section of E.

The mapping $(X,\varphi)\to\nabla_X\varphi$ satisfies the following conditions: (i) $\nabla_X\varphi$ is linear with respect to X and φ; (ii) $\nabla_{fX}\varphi=f\nabla_X\varphi$; and (iii) $\nabla_X(f\varphi)=(Xf)\varphi+f\nabla_X\varphi$, where f is a differentiable function on M. From these conditions it is seen that the mapping $X\to\nabla_X\varphi$ for a fixed section φ of E defines a differential form of degree one with values in E, denoted by $\nabla\varphi$. ∇ is called a G-connection on E (induced from the connection in P). A linear operator $d^\nabla:\Gamma(\Lambda^p\otimes E)\to\Gamma(\Lambda^{p+1}\otimes E)$, defined by $d^\nabla(\alpha\otimes\varphi)=d\alpha\otimes\varphi+(-1)^p\alpha\wedge\nabla\varphi$ for a differential form α of degree p and a differentiable section φ of E is called a covariant exterior differentiation. Here $\Gamma(\Lambda^p\otimes E)$ denotes the set of all differential forms of degree p with values in E.

The curvature form R^∇ of ∇ is defined by $R^\nabla(X,Y)\varphi=\nabla_X(\nabla_Y\varphi)-\nabla_Y(\nabla_X\varphi)-\nabla_{[X,Y]}\varphi$, where X and Y are vector fields on M and φ is a differentiable section of E. In terms of the curvature form Ω of the connection in P, it can be defined also by

$$R^\nabla_p(X,Y)\varphi=\bar{x}\circ\Omega_x(X^*,Y^*)\cdot\bar{x}^{-1}(\varphi),$$

where X and Y are in T_pM and $x\in\pi^{-1}(p)$ and X^*, Y^* are lifts of X, Y to P, respectively, and φ is an element of E_p. Since Ω is a tensorial form of type ad and the differential ρ_* of ρ induces a faithful representation of $\mathfrak{g}$ to $\mathfrak{gl}(F)$, R^∇ can be regarded as a differential form of degree two with values in the bundle $\mathfrak{g}_P=P\times_{\mathrm{Ad}}\mathfrak{g}$.

The curvature form R^∇ satisfies Bianchi's identity: $d^\nabla R^\nabla=0$, where d^∇ is a covariant exterior differentiation with respect to the G-connection on $\mathfrak{g}_P$ canonically induced from the connection in P.

We denote by $\mathscr{C}$ the set of all G-connections on E. Now let G be a compact semisimple Lie group. A functional $\mathscr{S}:\mathscr{C}\to\mathbf{R}$ defined by $\nabla\to\mathscr{S}(\nabla)=-\frac{1}{2}\int_M\mathrm{tr}(R^\nabla\wedge *R^\nabla)$ is called the **action integral (Yang-Mills functional)** of ∇, where $*$ is Hodge's star operator, given by the fixed orientation of M.

The group $\mathscr{G}_P$ acts on $\mathscr{C}$ by $\nabla\to f^{-1}\circ\nabla\circ f$ for $f\circ\mathscr{G}_P$ and then the curvature form is transformed by this action as $R^\nabla\to\mathrm{Ad}(f^{-1})R^\nabla$. Thus $\mathscr{S}$ is $\mathscr{G}_P$-invariant.

A connection ∇ is called a **Yang-Mills G-connection** if ∇ is a critical point of $\mathscr{S}$. The Euler-Lagrange equation of $\mathscr{S}$ is given by $d^{\nabla *}R^\nabla=0$ by the aid of the formal adjoint operator $d^{\nabla *}$ of d^∇. This equation, called the **Yang-Mills equation**, is a system of nonlinear second-order elliptic partial differential equations.

When M is a 4-dimensional vector space $\mathbf{R}^4$ with Minkowskian metric and G is the Abelian group $U(1)$, the Yang-Mills equation coincides with Maxwell's equations for an electromagnetic field. Thus the Yang-Mills equation for a non-Abelian group G is a natural extension of Maxwell's equations. In fact, the theory of Yang-Mills connections has its origin in the field theory of physics [11].

In the case of $\dim M=4$, special Yang-Mills G-connections occur. A G-connection ∇ satisfying the condition $*R^\nabla=R^\nabla$ (resp. $*R^\nabla=-R^\nabla$) as a differential form of degree two is called a **self-dual** (resp. **anti-self-dual**) G-connection. From Bianchi's identity and the expression of $d^{\nabla *}: d^{\nabla *}=-*\circ d^\nabla\circ *$, it follows that every self-dual (anti-self-dual) G-connection gives a solution to the Yang-Mills equation. Since the first Pontryagin number $p_1(E)$ is given by $p_1(E)=-(1/4\pi^2)\int_M\mathrm{tr}(R^\nabla\wedge R^\nabla)$ by virtue of the Chern-Weil theorem, the

action integral satisfies $\mathscr{S}(\nabla) \geqslant 2\pi^2 |p_1(E)|$ for every ∇ in $\mathscr{C}$ and the equality holds if and only if ∇ is self-dual or anti-self-dual.

Explicit forms have been obtained for (anti-) self-dual connections over the 4-sphere S^4 by many interesting methods [12–15]. And it has been shown that moduli space of self-dual G-connections (i.e., the set of all solutions to the Yang-Mills equation modulo $\mathscr{G}_P$) over S^4 has the structure of a Hausdorff manifold of dimension $p_1(\mathfrak{g}_P) - \dim G$ for every principal bundle P with group G [16, 17]. It is not yet known whether there exists a Yang-Mills G-connection over S^4 whose holonomy group is an open subgroup of G and which is neither self-dual nor anti-self-dual [18]. The following is one of the few known facts concerning the properties of Yang-Mills G-connections: If a Yang-Mills G-connection, $G = SU(2)$, $SU(3)$, or $O(3)$, over S^4 is weakly stable, i.e., if the second variation of $\mathscr{S}$ is positive semidefinite, then it is self-dual or anti-self-dual [19].

References

[1] H. Cartan, Notions d'algèbre différentielle; application aux groupes de Lie et aux variétés où opère un groupe de Lie, Colloque de Topologie, Brussels, 1950, 15–27.
[2] S. S. Chern, Topics in differential geometry, Lecture notes, Institute for Advanced Study, Princeton, 1951.
[3] C. Ehresmann, Les connexions infinitésimales dans un espace fibré différentiable, Colloque de Topologie, Brussels, 1950, 29–55.
[4] S. Kobayashi and K. Nomizu, Foundations of differential geometry, Interscience, I, 1963; II, 1969.
[5] A. Lichnerowicz, Théorie globale des connexions et des groupes d'holonomie, Cremona, 1955.
[6] K. Nomizu, Lie groups and differential geometry, Publ. Math. Soc. Japan, 1956.
[7] K. Yano and S. Bochner, Curvature and Betti numbers, Ann. Math. Studies, Princeton Univ. Press, 1953.
[8] R. S. Palais, Seminar on the Atiyah-Singer index theorem, Ann. Math. Studies, Princeton Univ. Press, 1965.
[9] S. S. Chern, Differentiable manifolds, Lecture notes, Univ. of Chicago, 1959.
[10] N. Tanaka, On the equivalence problem associated with a certain class of homogeneous spaces, J. Math. Soc. Japan, 17 (1965), 103–139.
[11] C. N. Yang and R. L. Mills, Conservation of isotopic spin and isotopic gauge invariance, Phys. Rev., 96 (1954), 191–195.
[12] A. A. Belavin, A. M. Polyakov, A. S. Schwartz, and Yu. Tyupkin, Pseudo-particle solutions of the Yang-Mills equations, Phys. Lett., 59B, 1 (1975), 85–87.
[13] A. A. Belavin and V. E. Zakharov, Yang-Mills equations as inverse scattering problem, Phys. Lett. 73B, (1978), 53–57.
[14] R. Jackiw, C. Nohl, and C. Rebbi, Conformal properties of pseudo particle configurations, Phys. Rev. D15, 6 (1977), 1642–1646.
[15] V. G. Manin and Yu. I. Manin, A description of instantons, Comm. Math. Phys., 63 (1978), 177–192.
[16] M. F. Atiyah, N. J. Hitchin, and I. M. Singer, Self-duality in four-dimensional Riemannian geometry, Proc. Royal Soc. London A, 362 (1978), 425–461.
[17] M. Itoh, Moduli of anti-self-dual connections on Kähler manifolds, Proc. Japan Acad., 57A (1981), 176–180.
[18] M. F. Atiyah and J. D. S. Jones, Topological aspects of Yang-Mills theory, Comm. Math. Phys., 61 (1978) 97–118.
[19] J. P. Bourguignon and H. B. Lawson, Stability and isolation phenomena for Yang-Mills fields, Comm. Math. Phys., 79 (1981), 189–230.

81 (I.11) Constructive Ordinal Numbers

A. General Remarks

To extend the theory of †recursive functions to transfinite ordinal numbers, A. Church and S. C. Kleene [1] considered the set of effectively accessible ordinal numbers and defined the concept of constructive ordinal numbers as explained later in this article. Their work became the basis of fruitful research by Kleene, W. Markwald, C. Spector, and others [2–5]. A constructive ordinal number was originally introduced as an "expression" in a †formal system utilizing the λ-notation. Since such a system is "effective," we can arithmetize it utilizing †Gödel numbers and assume from the outset that each ordinal number is representable by a natural number. The notations, terminology, and theorems mentioned in this article are mainly those for constructive ordinal numbers of the †second number class.

B. Definition and Fundamental Properties

We call a set of natural numbers satisfying conditions (I) and (II) a **system of notations for ordinal numbers**, and an ordinal number a

constructive ordinal number when it is representable by a natural number belonging to such a system of notations: (I) No natural number represents two distinct ordinal numbers. (II) There are three †partial recursive functions $K(x)$, $P(x)$, and $Q(x,n)$ defined as follows: (i) for any natural number x representing X, $K(x)$ takes the value 0, 1, or 2 according as X is zero, an †isolated ordinal number, or a †limit ordinal number, respectively; (ii) when X is the ordinal number †immediately after an ordinal number Y, $P(x)$ represents Y for any natural number x representing X; (iii) when X is a limit ordinal number, for any natural number x representing X there exists an increasing sequence $\{Y_n\}$ of ordinal numbers such that $X=\lim_n Y_n$ and $Q(x,n)$ represents Y_n for each natural number n.

The system called S_3 by Kleene is the most useful and convenient among systems of notations for ordinal numbers. Let n_O be a †primitive recursive function of the variable n defined by $0_O=1$, $(n+1)_O=2^{n_O}$. The fundamental notion $a\in O$ and relation $a<_O b$ of the system S_3 are introduced by the following inductive definition: (1) $1\in O$; (2) if $y\in O$, then $2^y\in O$ and $y<_O 2^y$; (3) if a sequence $\{y_n\}$ of natural numbers has the property that for each n, $y_n\in O$ and $y_n<_O y_{n+1}$, and if y is a Gödel number that defines y_n recursively as a function of n_O (i.e., $y_n\cong\{y\}(n_O)$) (→ 356 Recursive Functions)), then $3\cdot 5^y\in O$, and for each n, $y_n<_O 3\cdot 5^y$; (4) if x, y, $z\in O$, $x<_O y$, and $y<_O z$, then $x<_O z$; (5) $a\in O$, $a<_O b$ hold only when they follow from (1)–(4).

Now, for brevity, we write $a\leqslant_O b$ for $(a<_O b)\vee(a=b)$. The following propositions hold for S_3: (1) If $a<_O b$, then $b\neq 1$; (2) If $a<_O b$, then a, $b\in O$; (3) If $a<_O 2^y$, then $a\leqslant_O y$; (4) If $a<_O 3\cdot 5^y$, then there is a natural number n such that $a\leqslant_O y_n$, where $y_n=\{y\}(n_O)$; (5) If $a\in O$, then $1\leqslant_O a$; (6) If $a\in O$, then for any †number-theoretic function α such that $\alpha(0)=a$ and $\forall n(\alpha(n)\neq 1\rightarrow\alpha(n+1)<_O\alpha(n))$, there is a k such that $\alpha(k)=1$; (7) For each a, $\neg(a<_O a)$; (8) If $c\in O$, $a\leqslant_O c$, and $b\leqslant_O c$, then $a<_O b$ or $a=b$ or $b<_O a$.

Each member a of O **represents** an ordinal number $|a|$ as follows: $|1|=0$; $|2^y|=|y|+1$ for $y\in O$; $|3\cdot 5^y|=\lim_n|y_n|$ for $3\cdot 5^y\in O$, where $y_n=\{y\}(n_O)$. Let b be a member of O. Then $|a|<|b|$ when $a<_O b$; and conversely, for each $\alpha<|b|$, there is a number a such that $|a|=\alpha$ and $a<_O b$. Hence the set $\{a\mid a<_O b\}$ is a †well-ordered set with respect to $<_O$, and its †order type is $|b|$. The least number ξ greater than $|a|$ for every member a of O is the least ordinal number that is not constructive. It is denoted by ω_1^{CK} (Church and Kleene denoted it by ω_1). There is a subsystem of S_3 that is well ordered with respect to $\leqslant_O$ and contains a unique notation for each constructive ordinal number α. For such a subsystem we can take a Π_1^1 set K such that K is †recursive in O (S. Feferman and Spector, R. O. Gandy) (→ 356 Recursive Functions H).

C. Constructive Ordinals and the Kleene Hierarchy

Let $R(x,y)$ be a †predicate on natural numbers. We write $x\leqslant_R y$ for any natural numbers x, y for which $R(x,y)$ holds. We consider only the case where $\leqslant_R$ is a †linear ordering on the set $D_R=\{x\mid\exists y(R(x,y)\vee R(y,x))\}$. If D_R is a well-ordered set with respect to $\leqslant_R$, we denote its †order type by $|R|$. (1) For each (constructive) ordinal number $\alpha<\omega_1^{\mathrm{CK}}$, there is a †general recursive (more strictly, †primitive recursive) predicate R such that $|R|=\alpha$ (Markwald, Spector, Kleene [3]). (2) Conversely, if R is a †hyperarithmetic predicate (e.g., R is general recursive), then $|R|<\omega_1^{\mathrm{CK}}$ (Markwald, Spector). The following theorems are the most fruitful ones in the theory of constructive ordinal numbers, and they fully support the validity of Kleene's idea of †analytic hierarchy. (3) The set O is Π_1^1 (→ 356 Recursive Functions H), and so is the predicate $a<_O b$. Namely, for O, there is a primitive recursive predicate $R(a,x,\alpha)$ such that

$$a\in O\Leftrightarrow\forall\alpha\exists x R(a,x,\alpha)$$

(Kleene [3]). (4) For each ordinal number $\alpha<\omega_1^{\mathrm{CK}}$, the set $\{a\mid\alpha\geqslant|a|\}$ is a hyperarithmetic set (Spector). (5) O is a †complete set for Π_1^1. That is, for any Π_1^1 set E, there is a primitive recursive function φ such that $a\in E\Leftrightarrow\varphi(a)\in O$ (Kleene [3]). Accordingly, O is not a Σ_1^1 set (→ 356 Recursive Functions H).

D. Relativization and Extension

Given a (number-theoretic) predicate Q of one variable (or a set of natural numbers), we can †relativize to Q the notion of constructive ordinal numbers. The least ordinal that is not constructive relative to Q is denoted by ω_1^Q. The relativization to Q of the fundamental notion $a\in O$ and relation $a<_O b$ of the system S_3 of notations are denoted by $a\in O^Q$ and $a<_O^Q b$, respectively. Then we can relativize the results of the preceding paragraphs to Q. For example, as the relativization of (3), we have the following: There is a predicate $R^Q(a,x,\alpha)$ which is primitive recursive †uniformly in Q such that

$$a\in O^Q\Leftrightarrow\forall\alpha\exists x R^Q(a,x,\alpha).$$

When Q is hyperarithmetic, we have no generalization of the constructive ordinal numbers

by relativizing them to Q, that is, $\omega_1^Q = \omega_1^{CK}$ holds (Spector). Now by relativizing to O the concept of constructive ordinal numbers, we obtain a proper extension of it ($\omega_1^{CK} < \omega_1^O$), and then, performing such extensions successively, we have a (transfinite) sequence O, O^O, O^{O^O}, ... On the other hand, we can extend the constructive ordinal numbers to those corresponding to any like higher number class, beyond the second, in which partial recursive functions are used at limit levels to provide an "accessibility" mapping from previously defined number classes. There are several extensions done by Church and Kleene, H. C. Wang, D. L. Kreider and H. Rogers, Jr., H. Putnam, W. Richter, A. Kino and G. Takeuti, and others. Richter [9] has shown that these two ways of extending the constructive ordinals are equivalent, provided the sets of notations for the higher number classes satisfy certain natural conditions. Specifically, the ordinals of the Addison and Kleene [6] constructive third number class are exactly the ordinals less than ω_1^O and the set O_{2_0} of notations for those is recursively isomorphic to O^O (Richter).

E. Constructive Ordinals in Higher Number Classes

Putnum [8] has defined a system C of notations for constructive ordinals of the Cantor higher number classes, improving that of Kreider and Rogers [7], as follows: N_α is the set of notations for the ordinal number α; if for some α, $x \in N_\alpha$, let $|x| = \mu\xi[x \in N_\xi]$, which is called a **hyperconstructive ordinal**; and let $C_\alpha = \cup\{N_\xi \mid \xi < \alpha\}$. Define i to be a Gödel number of the †identity function, and n to be an index in C_α if $3^n \cdot 5^t \in C_\alpha$ for some t. There are four cases:

Case 1. $\alpha = 0$. Then $N_\alpha = \{1\}$.

Case 2. $\alpha = \beta + 1$, where N_β is already defined. Then $N_\alpha = \{2^x \mid x \in N_\beta\}$.

Case 3. α is a limit ordinal such that N_ξ is already defined for all $\xi < \alpha$, and there exists an ordinal $\beta < \alpha$ such that $3^a \cdot 5^i \in N_\beta$ for some a, and a partial recursive function f giving an order-preserving †cofinal mapping (o.p.c.m.) from C_β into C_α. Then N_α is taken to be the set of all numbers $3^a \cdot 5^n$ such that $3^a \cdot 5^i \in N_\beta$ (where β is any ordinal with the above property) and $\{n\}$ is an o.p.c.m. from C_β into C_α.

Case 4. α is a limit ordinal such that N_ξ is already defined for all $\xi < \alpha$, there is no β as in Case 3, but there is a number $a \in C_\alpha$ which is not an index in C_α. Let β_0 be the least β for which such a's belong to N_β. Then $N_\alpha = \{3^a \cdot 5^n \mid a \in N_{\beta_0}$ and $\{n\}$ is an o.p.c.m. from C_α into $C_\alpha\}$.

We have that $x \in N_\alpha$ only as required by Cases 1–4. Finally, let ω_C be the least ordinal number which is not hyperconstructive, and let $C = C_{\omega_C}$.

Extending Kleene's H_y for $y \in O$ (→ 356 Recursive Functions H), Putnum has defined a hierarchy H_x for $x \in C$, and shown that hyperconstructive ordinal numbers are uniqueness ordinals, i.e., for $x, y \in C$, if $|x| = |y|$, then H_x and H_y are of the same †degree of (recursive) unsolvability. $\mathbf{E}_1$ is the type-2 object (→ 356 Recursive Functions F) introduced by T. Tugué such that $\mathbf{E}_1(\alpha) = 0$ if $\forall\beta\exists x[\alpha(\beta(x)) = 0]$; otherwise, $\mathbf{E}_1(\alpha) = 1$. It is known that $\omega_C = \omega^{\mathbf{E}_1}$, where $\omega^{\mathbf{E}_1}$ is the least ordinal which is not the order type of any well-ordering recursive in $\mathbf{E}_1$ (Richter [10]).

References

[1] A. Church and S. C. Kleene, Formal definitions in the theory of ordinal numbers, Fund. Math., 28 (1937), 11–21.

[2] S. C. Kleene, On notation for ordinal numbers, J. Symbolic Logic, 3 (1938), 150–155.

[3] S. C. Kleene, On the forms of the predicates in the theory of constructive ordinals II, Amer. J. Math., 77 (1955), 405–428.

[4] W. Markwald, Zur Theorie der konstruktiven Wohlordnungen, Math. Ann., 127 (1954), 135–149.

[5] C. Spector, Recursive well-orderings, J. Symbolic Logic, 20 (1955), 151–163.

[6] J. W. Addison and S. C. Kleene, A note on function quantification, Proc. Amer. Math. Soc., 8 (1957), 1002–1006.

[7] D. L. Kreider and H. Rogers, Jr., Constructive versions of ordinal number classes, Trans. Amer. Math. Soc., 100 (1961), 325–369.

[8] H. Putnum, Uniqueness ordinals in higher constructive number classes, Essays on the Foundations of Mathematics, Magnes Press, Jerusalem, 1961, and North-Holland, 1962, 190–206.

[9] W. Richter, Extensions of the constructive ordinals, J. Symbolic Logic, 30 (1965), 193–211.

[10] W. Richter, Constructively accessible ordinal numbers, J. Symbolic Logic, 33 (1968), 43–55.

82 (XIII.18) Contact Transformations

A. General Remarks

A transformation of $2n+1$ variables z, x_j, p_j $(j = 1, 2, \ldots, n)$,

$$Z=Z(z,x_1,\dots,x_n,p_1,p_2,\dots,p_n),$$

$$X_j=X_j(z,x_1,x_2,\dots,x_n,p_1,p_2,\dots,p_n),\quad j=1,2,\dots,n,$$

$$P_j=P_j(z,x_1,x_2,\dots,x_n,p_1,p_2,\dots,p_n),\quad j=1,2,\dots,n, \qquad (1)$$

is called a **contact transformation** in the $(n+1)$-dimensional space $\mathbf{R}^{n+1}$ with the coordinate system $(z,x_1,\dots,x_n)$ if the †total differential equation

$$dz-p_1\,dx_1-p_2\,dx_2-\dots-p_n\,dx_n=0 \qquad (2)$$

is invariant under the transformation, i.e., if the equality

$$dZ-P_1\,dX_1-P_2\,dX_2-\dots-P_n\,dX_n$$
$$=\rho(dz-p_1\,dx_1-p_2\,dx_2-\dots-p_n\,dx_n) \qquad (3)$$

holds identically for a suitable nonzero function ρ of z,x_j,p_j. Here we assume that (1) has an inverse transformation. Using **Lagrange's bracket**

$$[f,g]=\sum_{j=1}^{n}\left(\frac{\partial f}{\partial p_j}\left(\frac{dg}{dx_j}\right)-\frac{\partial g}{\partial p_j}\left(\frac{df}{dx_j}\right)\right),$$

$$\left(\frac{df}{dx_j}\right)\equiv\frac{\partial f}{\partial x_j}+p_j\frac{\partial f}{\partial z},$$

we see that (1) is a contact transformation if and only if $[X_j,X_k]=[X_j,Z]=[P_j,P_k]=0$, $[P_j,X_k]=\rho\delta_{jk}$, $[P_j,Z]=\rho P_j$, where δ_{jk} is †Kronecker's delta.

From this fact it follows that the composite of two contact transformations and the inverse transformation of a contact transformation are also contact transformations. Since the identity transformation $Z=z$, $X_j=x_j$, $P_j=p_j$ is a contact transformation, the set of all contact transformations forms an infinite-dimensional †topological group. Given a set of scalars $p_1,\dots,p_n$, a pair consisting of a point (z,x_j) and an n-dimensional hyperplane $z^*-z=\sum_{j=1}^{n}p_j(x_j^*-x_j)$ in an $(n+1)$-dimensional space is called a **hypersurface element**, and the set of hypersurface elements satisfying (2) is called a **union of hypersurface elements**. Using these concepts, we can state that a transformation (1) of coordinates z,x_j,p_j $(j=1,2,\dots,n)$ is a contact transformation if it transforms each union of hypersurface elements into another one. Consequently, if two n-dimensional hypersurfaces are tangent at a point (z,x_j) in the $(n+1)$-dimensional space, their images under a contact transformation, which are again two n-dimensional hypersurfaces, are tangent at the image point of (z,x_j). The name "contact transformation" is derived from this fact.

For instance, the †correlation with respect to a hypersurface of the second order gives a contact transformation. In fact, from the relation between †poles and †polar lines with respect to the parabola $x^2+2y=0$ in a plane, we have **Legendre's transformation** $X=-p$, $Y=xp-y$, $P=-x$ $(\rho=-x)$.

In general, an invertible transformation defined by the three relations $\Omega(x,y,X,Y)=0$, $\partial\Omega/\partial X+P\partial\Omega/\partial Y=0$, $\partial\Omega/\partial x+p\partial\Omega/\partial y=0$ derived from a function $\Omega(x,y,X,Y)$ is a contact transformation. The function Ω is called the **generating function** of this transformation. In this transformation, to each point (x_0,y_0) there corresponds a curve $\Omega(x_0,y_0,X,Y)=0$. These results are valid also in the case of several variables. For instance, in an $(n+1)$-dimensional space, a transformation $Z=z-x_1p_1-\dots-x_\nu p_\nu$; $X_1=p_1,\dots,X_\nu=p_\nu$, $X_{\nu+1}=x_{\nu+1},\dots,X_n=x_n$; $P_1=-x_1,\dots,P_\nu=-x_\nu$, $P_{\nu+1}=p_{\nu+1},\dots,P_n=p_n$ represents a contact transformation. Here ν is an integer between 1 and n. In the case $n=2$, $\nu=2$, this transformation reduces to a Legendre transformation; and in the case $n=2$, $\nu=1$, it is called **Ampère's transformation** (→ Appendix A, Table 15.IV).

B. Canonical Transformations

A transformation of $2n$ variables x_j, p_j $(j=1,2,\dots,n)$

$$X_j=X_j(x_1,x_2,\dots,x_n,p_1,p_2,\dots,p_n),$$

$$P_j=P_j(x_1,x_2,\dots,x_n,p_1,p_2,\dots,p_n),\quad j=1,2,\dots,n, \qquad (4)$$

is called a **canonical transformation** if the †differential form $\sum_{j=1}^{n}(P_j\,dX_j-p_j\,dx_j)$ is †exact in x_j, p_j, i.e., if there exists a function U of x, p such that

$$\sum_{j=1}^{n}(P_j\,dX_j-p_j\,dx_j)=dU. \qquad (5)$$

Let (1) denote a contact transformation, and let λ denote a new variable which is different from zero. Set $\bar{x}_0=z$, $\bar{p}_0=\lambda$, $\bar{x}_j=x_j$, $\bar{p}_j=-\lambda p_j$ $(j=1,2,\dots,n)$ and define a transformation of $(2n+2)$ variables by

$$\bar{X}_0(\bar{x},\bar{p})=Z(z,x,p),$$

$$\bar{P}_0(\bar{x},\bar{p})=\bar{P}_0/\rho(z,x,p),\ \bar{X}_j(\bar{x},\bar{p})=X_j(x,p),$$

$$\bar{P}_j(\bar{x},\bar{p})=-\bar{p}_0P_j(z,x,p)/\rho(z,x,p),\quad j=1,2,\dots,n, \qquad (6)$$

where x, p, $\bar{x}$, and $\bar{p}$ denote points $(x_1,x_2,\dots,x_n)$, $(p_1,p_2,\dots,p_n)$, $(\bar{x}_0,\bar{x}_1,\dots,\bar{x}_n)$, and $(\bar{p}_0,\bar{p}_1,\dots,\bar{p}_n)$, respectively. Then the identity (3) when multiplied by λ on both sides becomes

$$\bar{P}_0\,d\bar{X}_0+\bar{P}_1\,d\bar{X}_1+\dots+\bar{P}_n\,d\bar{X}_n$$
$$=\bar{p}_0\,d\bar{x}_0+\bar{p}_1\,d\bar{x}_1+\dots+\bar{p}_n\,d\bar{x}_n. \qquad (7)$$

Thus the transformation (6) represents a canonical transformation (U = constant). Moreover, from the definition the relation (6) is homogeneous, i.e., for all $\mu \neq 0$,

$$\bar{X}_j(\bar{x}, \mu\bar{p}) = \bar{X}_j(\bar{x}, \bar{p}), \qquad \bar{P}_j(\bar{x}, \mu\bar{p}) = \mu\bar{P}_j(\bar{x}, \bar{p}), \qquad j = 0, 1, \ldots, n.$$

Conversely, let a transformation of $(2n+2)$ variables

$$\bar{X}_j = \bar{X}_j(\bar{x}, \bar{p}), \qquad \bar{P}_j = \bar{P}_j(\bar{x}, \bar{p}), \qquad j = 0, 1, \ldots, n, \quad (8)$$

be a **homogeneous canonical transformation**. Define a transformation of $(2n+1)$ variables z, x_j, p_j $(j = 1, 2, \ldots, n)$ by

$$Z(z, x, p) = \bar{X}_0(z, x_1, \ldots, x_n, \lambda, -\lambda p_1, \ldots, -\lambda p_n),$$

$$1/\rho(z, x, p) = \frac{1}{\lambda}\bar{P}_0(z, x_1, \ldots, x_n, \lambda, -\lambda p_1, \ldots, -\lambda p_n),$$

$$X_j(z, x, p) = \bar{X}_j(z, x_1, \ldots, x_n, \lambda, -\lambda p_1, \ldots, -\lambda p_n),$$

$$P_j(z, x, p) = -\frac{\rho}{\lambda}\bar{P}_j(z, x_1, \ldots, x_n, \lambda, -\lambda p_1, \ldots, -\lambda p_n), \qquad j = 1, 2, \ldots, n. \quad (9)$$

Since the necessary and sufficient condition for (8) to be a homogeneous canonical transformation is that (7) hold, it follows that

$$\frac{\lambda}{\rho}(dZ - P_1\,dX_1 - \ldots - P_n\,dX_n) = \lambda(dz - p_1\,dx_1 - \ldots - p_n\,dx_n).$$

Thus the transformation (9) represents a contact transformation. Therefore the most general contact transformation of $(2n+1)$ variables and the most general homogeneous canonical transformation are identical concepts, differing only in the choice of notation.

The necessary and sufficient condition for (4) to be a canonical transformation is expressed by the relations

$$(X_j, X_k) = 0, \qquad (X_j, P_k) = \delta_{jk}, \qquad (P_j, P_k) = 0, \qquad j, k = 1, 2, \ldots, n, \quad (10)$$

where $(\cdot, \cdot)$ is **Poisson's bracket**, i.e., for a pair of functions $f(x, p)$, $g(x, p)$,

$$(f, g) = \sum_{j=1}^{n}\left(\frac{\partial f}{\partial x_j}\frac{\partial g}{\partial p_j} - \frac{\partial f}{\partial x_j}\frac{\partial h}{\partial p_j}\right).$$

Denote Poisson's bracket in the coordinate system $X_1, X_2, \ldots, X_n, P_1, P_2, \ldots, P_n$ by $(\cdot, \cdot)'$, and for $f(x, p)$ and $g(x, p)$ denote their transformations by $f'(X, P)$ and $g'(X, P)$. Then the relation (10) is equivalent to

$$(f, g) = (f', g')' \text{ for all } f \text{ and } g.$$

From this fact it follows that the set of all canonical transformations forms an infinite-dimensional topological group.

Suppose that we can take $x_1, x_2, \ldots, x_n, X_1, X_2, \ldots, X_n$ as $2n$ independent variables. Denote by $\Omega_1(x, X)$ the function U of (5) represented by x, X. By means of this Ω_1, $2n$ functions which give a canonical transformation (4) are derived from the relations

$$p_j = -\frac{\partial\Omega_1}{\partial x_j}, \qquad P_j = \frac{\partial\Omega_1}{\partial X_j}, \qquad j = 1, 2, \ldots, n.$$

Conversely, for $\Omega(x, X)$ such that $\det\left(\frac{\partial^2\Omega}{\partial x_j \partial X_k}\right) \neq 0$, a transformation determined by the relations

$$p_j = -\frac{\partial\Omega}{\partial x_j}, \qquad P_j = \frac{\partial\Omega}{\partial x_j}, \qquad j = 1, 2, \ldots, n,$$

represents a canonical transformation. For a general canonical transformation, $x_1, x_2, \ldots, x_n, X_1, X_2, \ldots, X_n$ are not necessarily $2n$ independent variables. But for a canonical transformation there exists n new variables $X_{i_1}, X_{i_2}, \ldots, X_{i_k}, P_{j_1}, \ldots, P_{j_{n-k}}$ such that $x_1, x_2, \ldots, x_n, X_{i_1}, X_{i_2}, \ldots, X_{i_k}, P_{j_1}, P_{j_2}, \ldots, P_{j_{n-k}}$ are $2n$ independent variables, where $(i_1, i_2, \ldots, i_k)$, $(j_1, j_2, \ldots, j_{n-k})$ is a partition of the set $(1, 2, \ldots, n)$ to two disjoint subsets. Denote by Ω_2 the function $U - \sum_{l=1}^{n-k} X_{j_l}P_{j_l}$. Then the transformation (4) is derived from relations

$$p_j = -\frac{\partial\Omega_2}{\partial x_j}, \qquad j = 1, 2, \ldots, n,$$

$$X_{j_l} = -\frac{\partial\Omega_2}{\partial P_{j_l}}, \qquad l = 1, 2, \ldots, n-k,$$

$$P_{i_l} = \frac{\partial\Omega_2}{\partial X_{i_l}}, \qquad l = 1, 2, \ldots, k.$$

These functions Ω_1, Ω_2 are called the **generating functions** of a canonical transformation.

Consider a transformation of $2n$ variables depending smoothly on a parameter t:

$$X_j = X_j(t, x_1, x_2, \ldots, x_n, p_1, p_2, \ldots, p_n),$$

$$P_j = P_j(t, x_1, x_2, \ldots, x_n, p_1, p_2, \ldots, p_n), \qquad j = 1, 2, \ldots, n. \quad (11)$$

If (11) represents a canonical transformation of $2n$ variables for each t, then X_j and P_j, $j = 1, 2, \ldots, n$, satisfy †Hamilton's canonical equations

$$\frac{dX_j}{dt} = \frac{\partial H}{\partial P_j}(t, X, P), \qquad \frac{dP_j}{dt} = -\frac{\partial H}{\partial X_j}(t, X, P), \qquad j = 1, 2, \ldots, n, \quad (12)$$

for some †Hamiltonian function $H(t, x, p)$. Conversely, for any solution X_j, P_j of Hamilton's canonical equations (12) depending on

parameters $x_1, x_2, \ldots, x_n, p_1, p_2, \ldots, p_n$, if the transformation (11) is a canonical transformation for $t=0$, (11) represents a canonical transformation for each t. By a canonical transformation (4), the solutions of Hamilton's canonical equations

$$\frac{dx_j}{dt}=\frac{\partial H}{\partial p_j}(t,x,p), \qquad \frac{dp_j}{dt}=-\frac{\partial H}{\partial x_j}(t,x,p), \qquad j=1,2,\ldots,n, \quad (13)$$

are transformed to the solutions of

$$\frac{dX_j}{dt}=\frac{\partial K}{\partial P_j}(t,X,P), \qquad \frac{dP_j}{dt}=-\frac{\partial K}{\partial X_j}(t,X,P), \qquad j=1,2,\ldots,n, \quad (14)$$

where the new Hamiltonian function is given by $K(t,X,P)=H(t,x,p)$.

C. Applications to the Integration of Differential Equations

Contact transformations have applications to the integration of differential equations since they transform each union of surface elements into another one.

As an example, we shall describe an outline of their application to a partial differential equation of the first order

$$F(x,y,z,p,q)=0;$$

$$p\equiv\partial z/\partial x, \qquad q\equiv\partial z/\partial y. \quad (15)$$

For this purpose, we regard (15) as an equation defining unions of surface elements and transform it into a simpler equation by means of a contact transformation. If the transformed equation can be solved, then the solution of the original equation can be obtained by means of the inverse transformation. Now, let $z=\omega(x,y,a,b)$ be a †complete solution of (15). Then (15) is reduced to $Z-c=0$ by the transformation generated by the function

$$\Omega\equiv Z-z+\omega(x,y,X,Y)-c=0,$$

where c is a constant. In this equation the solution $X=a$, $Y=b$, $Z=c$, $\alpha P+\beta Q=0$ (a,b,c,α,β are constants) plays an important role, and this line element will be called a **characteristic line element**. The characteristic line element satisfies equations that can be transformed by means of the inverse transformation into **Charpit's subsidiary equations** for (15):

$$\frac{dx}{\partial F/\partial p}=\frac{dy}{\partial F/\partial q}=\frac{dz}{p\partial F/\partial p+q\partial F/\partial q}$$

$$=\frac{-dp}{\partial F/\partial x+p\partial F/\partial z}=\frac{-dg}{\partial F/\partial y+q\partial F/\partial z}. \quad (16)$$

Consequently, if we have $p=p(x,y,z;a)$, $q=q(x,y,z;a)$ from the solution $G(x,y,z,p,q)=a$ of (16) and $F=0$, then the total differential equation $dz=pdx+qdy$ is †completely integrable, and the †general solution of this equation is a complete solution of (15). Also, if we know two independent solutions $G(x,y,z,p,q)=a$, $H(x,y,z,p,q)=b$ of (16) such that $[G,H]=0$, we can obtain a complete integral of (15) by eliminating p, q among the three equations $F=0$, $G=a$, $H=b$. This method is called the †Lagrange-Charpit method, which is applicable also to the equation $F(z,x_1,\ldots,x_n;p_1,\ldots,p_n)=0$ (→ 322 Partial Differential Equations (Methods of Integration)).

D. Applications to Analytical Dynamics

Consider Hamilton's canonical equations (13). An example of system (13) is the equation of motion for a dynamical system which is derived from $\delta\int L(x_j,\dot{x}_j)dt=0$. Here the x_j denote generalized coordinates, $p_j=\partial L/\partial\dot{x}_j$ and $H=\sum_{j=1}^{n}p_j\dot{x}_j-L$.

Let $S(x_1,x_2,\ldots,x_n,X_1,X_2,\ldots,X_n)$ be a solution depending on parameters $X_1,X_2,\ldots,X_n$ of the †Hamilton-Jacobi equation

$$H\left(t,x_1,x_2,\ldots,x_n,-\frac{\partial S}{\partial x_1},-\frac{\partial S}{\partial x_2},\ldots,-\frac{\partial S}{\partial x_n}\right)$$

$$=K(X_1,X_2,\ldots,X_n) \quad (17)$$

such that $\det\left(\frac{\partial^2 S}{\partial x_j\partial X_k}\right)\neq 0$. Then by a canonical transformation of a generating function S, i.e., by a transformation determined by the relations

$$p_j=-\frac{\partial S}{\partial x_j}, \qquad P_j=\frac{\partial S}{\partial X_j}, \qquad j=1,2,\ldots,n,$$

the system (13) is transformed to (14). Here

$$K(t,X,P)=H\left(t,x,-\frac{\partial S}{\partial x}\right)=K(X).$$

Then (14) becomes

$$\frac{dX_j}{dt}=0, \qquad \frac{dP_j}{dt}=-\frac{\partial K}{\partial x_j}(X), \qquad j=1,2,\ldots,n,$$

and the solutions are

$$X_j(t)=X_j(0), \qquad P_j(t)=P_j(0)-t\left.\frac{\partial K}{\partial X_j}\right|_{X=X(0)}$$

In particular, a transformation that makes $K\equiv 0$ is called a **transformation to an equilibrium system**. Thus solving Hamilton's canonical equations reduces to finding a †general solution of the Hamilton-Jacobi equation (17).

When the variables of the Hamilton-Jacobi equation (17) are totally separable we can obtain a general solution of (17) by †quadra-

tures. Then one can integrate by quadratures the system of equations (13). For many important problems exact solutions are obtained by this method (→ 271 Mechanics).

Furthermore, every mapping in an optical system is a canonical transformation; in optics, the quantity corresponding to S is called an **eikonal** (→ 180 Geometric Optics).

References

[1] R. Courant and D. Hilbert, Methods of mathematical physics II, Interscience, 1962.
[2] C. Carathéodory, Calculus of variations and partial differential equations of the first order, Holden-Day, 1965.
[3] V. Arnold, Les méthodes mathématiques de la mécanique classique, Mir, 1976.
[4] I. M. Gel'fand and S. V. Fomin, Calculus of variations, Prentice-Hall, 1963.

83 (V.3) Continued Fractions

A. The Notion of Continued Fractions

Let $\{b_n\}$ $(n=0,\ldots,m)$ and $\{c_n\}$ $(n=1,\ldots,m)$ be finite sequences of elements in a †field F. A fraction of the form

$$b_0+\cfrac{c_1}{b_1+\cfrac{c_2}{b_2+\ldots\cfrac{\cdots}{\cdots+\cfrac{c_{m-1}}{b_{m-1}+\cfrac{c_m}{b_m}}}}}$$

is called a **finite continued fraction**. It expresses an element in the field F unless devision by 0 occurs in the process of reduction. Symbolically, it is also written in the forms

$$b_0+\frac{c_1}{b_1}+\frac{c_2}{b_2}+\cdots+\frac{c_{m-1}}{b_{m-1}}+\frac{c_m}{b_m},$$

$$b_0+\frac{c_1|}{|b_1}+\frac{c_2|}{|b_2}+\cdots+\frac{c_{m-1}|}{|b_{m-1}}+\frac{c_m|}{|b_m},$$

$$b_0+\frac{c_1}{b_1}\dot{+}\frac{c_2}{b_2}\dot{+}\cdots\dot{+}\frac{c_{m-1}}{b_{m-1}}\dot{+}\frac{c_m}{b_m},$$

$$\left[b_0,\frac{c_1}{b_1},\frac{c_2}{b_2},\ldots,\frac{c_{m-1}}{b_{m-1}},\frac{c_m}{b_m}\right],$$

etc., or more briefly,

$$b_0+\left[\frac{c_n}{b_n}\right]_{n=1}^{m}.$$

If $\{b_n\}$ $(n=0,1,\ldots)$ and $\{c_n\}$ $(n=1,2,\ldots)$ are infinite sequences, the expression

$$b_0+\cfrac{c_1}{b_1+\cfrac{c_2}{b_2+\cdots\cfrac{\cdots}{\cdots+\cfrac{c_n}{b_n+\ldots}}}}$$

is called an **infinite continued fraction**. By analogy with the finite case, it is expressed by

$$b_0+\frac{c_1}{b_1}+\frac{c_2}{b_2}+\cdots+\frac{c_n}{b_n}+\ldots$$

or by $b_0+\left[\frac{c_n}{b_n}\right]_{n=1}^{\infty}$.

For an infinite continued fraction, the quantity

$$k_n=b_0+\frac{c_1}{b_1}+\cdots+\frac{c_n}{b_n},\qquad k_0=b_0$$

is called its nth **convergent**, b_0 is called the **initial term**, and c_n and b_n $(n\geqslant 1)$ are called the **partial numerator** and **partial denominator**, respectively. If F is a †topological field (e.g., the real or complex number field) and the sequence $\{k_n\}$ of its elements converges, then the infinite continued fraction is said to **converge**, and the limit is called its **value**.

A finite or infinite continued fraction in which c_n $(n\geqslant 1)$ are all equal to 1, b_0 is a rational integer, and b_n $(n\geqslant 1)$ are all positive rational integers is called a **simple continued fraction**. It is expressed by $[b_0,b_1,\ldots]$. b_n is often called the n**th partial quotient**. In the following paragraphs we shall mostly discuss simple continued fractions.

For a real number x we mean by $[x]$ the greatest integer not exceeding x. [ ] is called the **Gauss symbol**. Let ω be any given real number, and put

$$\omega=b_0+\frac{1}{\omega_1},\quad b_0=[\omega];$$

$$\omega_n=b_n+\frac{1}{\omega_{n+1}},\quad b_n=[\omega_n],\quad n=1,2,\ldots.$$

Then an expansion of ω into a simple continued fraction

$$\omega=b_0+\frac{1}{b_1}+\cdots+\frac{1}{b_n}+\cdots$$

is obtained. If ω is irrational, this expansion is determined uniquely. Conversely, any infinite simple continued fraction converges to an irrational number. If ω is rational, the process is interrupted at a finite step $(\omega_m=b_m)$, resulting in

$$\omega=b_0+\frac{1}{b_1}+\cdots+\frac{1}{b_m}.$$

An alternative representation of a rational number by a simple continued fraction is given by replacing b_m above by $(b_m-1)+1/1$.

Examples of infinite simple continued fractions are

$$\frac{e^{2/p}+1}{e^{2/p}-1}=p+\frac{1}{3p+}\cdots\frac{1}{+(2n+1)p+}\cdots,$$

where p is a natural number (J. H. Lambert), and

$$e=2+\frac{1}{1+}\frac{1}{2+}\frac{1}{1+}\cdots\frac{1}{+1+}\frac{1}{2n+}\frac{1}{1+}\cdots$$

(L. Euler).

B. Convergents

Let the nth convergent of a simple continued fraction be expressed in the form of an irreducible fraction

$$\frac{P_n}{Q_n}=b_0+\frac{1}{b_1+}\cdots\frac{1}{+b_n},\quad n\geqslant 0,$$

and for convenience put $P_{-2}=0$, $P_{-1}=1$, $Q_{-2}=1$, $Q_{-1}=0$. Then we have the recurrence relations

$$P_n=b_nP_{n-1}+P_{n-2},\quad Q_n=b_nQ_{n-1}+Q_{n-2},\qquad n\geqslant 0,$$

whence follows

$$P_nQ_{n-1}-P_{n-1}Q_n=(-1)^{n+1},\quad n\geqslant -1.$$

Any simple continued fraction represents a real number ω which satisfies

$$\omega=(\omega_{n+1}P_n+P_{n-1})/(\omega_{n+1}Q_n+Q_{n-1})$$

in terms of the notation defined in this and the previous section.

In particular, let ω be an irrational number. Then each of the fractions

$$\frac{P_n^{(k)}}{Q_n^{(k)}}=\frac{P_{n-2}+kP_{n-1}}{Q_{n-2}+kQ_{n-1}},\quad k=1,2,\ldots,b_n-1,$$

which are inserted between two convergents P_{n-2}/Q_{n-2} and $P_n/Q_n=(P_{n-2}+b_nP_{n-1})/(Q_{n-2}+b_nQ_{n-1})$, is called an **intermediate convergent**, while the original convergent P_n/Q_n is called a **principal convergent**.

If a fraction P/Q approximating an irrational number ω satisfies $|\omega-P/Q|<|\omega-p/q|$ for any other fraction p/q with $q\leqslant Q$, then it is said to give the **best approximation**. The fraction giving the best approximation of ω is always a principal or intermediate convergent $P_n^{(k)}/Q_n^{(k)}$ of ω with $k>b_n/2$ or $k=b_n/2$, $Q_n>Q_{n-1}\omega_n$.

The convergents satisfy the relation $P_n/Q_n-P_{n-1}/Q_{n-1}=(-1)^{n-1}Q_nQ_{n-1}$; hence the sequence $\{P_{2n}/Q_{2n}\}$ (resp. $\{P_{2n+1}/Q_{2n+1}\}$) is monotonically increasing (decreasing). Approximation by convergents is shown in the relations

$$\left|\omega-\frac{P_n}{Q_n}\right|\leqslant\frac{1}{Q_{n+1}Q_n},\quad\left|\omega-\frac{P_n}{Q_n}\right|<\left|\omega-\frac{P_{n-1}}{Q_{n-1}}\right|,$$

$$\omega=\lim_{n\to\infty}\frac{P_n}{Q_n}=b_0+\sum_{n=0}^{\infty}\frac{(-1)^n}{Q_{n+1}Q_n}.$$

In particular, $|\omega-P_n/Q_n|<1/Q_n^2$. On the other hand, if P, Q are relatively prime positive integers with $|\omega-P/Q|<1/2Q^2$, then P/Q is a convergent to ω (A. M. Legendre). There are several results concerning the measure of approximation by convergents. Any ω satisfies $|\omega-P_n/Q_n|<1/(\sqrt{5}Q_n^2)$ for infinitely many n, while there exists an ω which satisfies $|\omega-P_n/Q_n|<1/(\lambda Q_n^2)$ for only a finite number of n provided that $\lambda>\sqrt{5}$ (e.g., $\omega=\sqrt{5}$, A. Hurwitz); at least one of two adjacent convergents satisfies the inequality $|\omega-P_n/Q_n|<1/(2Q_n^2)$ (K. T. Vahlen); at least one of three consecutive convergents satisfies the inequality $|\omega-P_n/Q_n|<1/(\sqrt{5}Q_n^2)$ (E. Borel); $|\omega-P/Q|<1/\sqrt{8}Q^2$ has infinitely many rational solutions P/Q whenever ω is not †equivalent to $(\sqrt{5}+1)/2$ (Hurwitz). One calls an irrational number ω **badly approximable** if there is a constant $c=c(\omega)>0$ such that $|\omega-P/Q|>c/Q^2$ for every rational P/Q. ω is badly approximable precisely if the partial quotients of the continued fraction expansion of ω are bounded. In particular, real quadratic irrationals have bounded partial quotients (→ Section C; 182 Geometry of Numbers F).

C. Quadratic Irrationals

If an infinite simple continued fraction $[b_0,b_1,\ldots]$ satisfies $b_{m+k+v}=b_{m+v}$ ($v=0,1,2,\ldots$), it is called a **periodic continued fraction** and is denoted by the symbol $[b_0,b_1,\ldots,\overline{b_m,\ldots,b_{m+k-1}}]$. According as $m=0$ or $m>1$, the periodic continued fraction is said to be **pure** or **mixed**, and the sequence b_m, $b_{m+1},\ldots,b_{m+k-1}$ is called a **period**. In order for the continued fraction of ω to be periodic, it is necessary and sufficient that ω be a quadratic irrational number, i.e., a root of $ax^2+bx+c=0$ with rational integral coefficients a, b, c and nonsquare discriminant b^2-4ac (J. L. Lagrange). In order for ω to be represented by a purely periodic continued fraction, it is necessary and sufficient that ω be an **irreducible quadratic irrational number**, i.e., that ω satisfy $\omega>1$ and $0>\omega'>-1$, where ω' is the conjugate root of ω (E. Galois). In order for ω to be equal to the square root of a nonsquare rational number, it is necessary and sufficient

that its continued fraction be of the form $[b_0, b_1, \ldots, b_{k-1}, 2b_0]$ (Legendre).

D. Application to Pell's Equation

Let $ax - by = 1$ $((a, b) = 1)$ be a †Diophantine equation of the first degree, and $a/b = [b_0, b_1, \ldots, b_m] = P_m/Q_m$. Since $P_m Q_{m-1} - P_{m-1} Q_m = (-1)^{m-1}$, a solution of the equation is given by $x_0 = (-1)^{m-1} Q_m$, $y_0 = (-1)^{m-1} P_m$. The general solution is then represented in the form $x_0 + bt$, $y_0 + at$ $(t \in \mathbf{Z})$. This method of obtaining a solution is essentially the same as the method which uses the †Euclidean algorithm.

Pell's equation $x^2 - Dy^2 = 1$ (D is a nonsquare integer > 1) was solved by Lagrange in terms of continued fractions. If the length of the period of D is k, all positive solutions of Pell's equation are given by $x = P_{2vk-1}$, $y = Q_{2vk-1}$ if k is odd, and by $x = P_{vk-1}$, $y = Q_{vk-1}$ if k is even $(v = 1, 2, \ldots)$, where P_n/Q_n denotes the nth convergent of the continued fraction expansion of $\sqrt{D}$. Incidentally, $x = P_{(2v-1)k-1}$, $y = Q_{(2v-1)k-1}$ $(v = 1, 2, \ldots)$ are the positive solutions of $x^2 - Dy^2 = -1$ provided that k is odd. There are no solutions of $x^2 - Dy^2 = \pm 1$ other than $x_v, y_v (v = 1, 2, \ldots)$ given by $(x_1 + \sqrt{D}\, y_1)^v = x_v + \sqrt{D}\, y_v$, where x_1, y_1 is the least positive solution. For instance, the least positive solution of $x^2 - 211y^2 = 1$ is $x = 278{,}354{,}373{,}650$, $y = 19{,}162{,}705{,}353$.

Lagrange made further use of continued fractions in order to obtain approximate values of roots of algebraic equations. The method is especially useful for precise computation of neighboring roots.

The theory of continued fractions may be investigated geometrically making use of lattices ([2, 3]; → 182 Geometry of Numbers) (F. Klein, G. Humbert). For instance, a measure of approximation of P_n/Q_n to ω in Diophantine approximation is represented by the closeness of a lattice point (P_n, Q_n) to the straight line $y = \omega x$ on the plane.

E. Continued Fractions with Variable Terms

There are few results on continued fractions with variable terms. It is noteworthy that from the expansion of $\tan z$ into a continued fraction

$$\tan z = \frac{z}{1+}\,\frac{-z^2}{3}\,+\,\frac{-z^2}{5}\,+\,\frac{-z^2}{7}\,+\cdots$$

(Lambert), the irrationality of π and of $\tan z$ for rational z $(\neq 0)$ can be deduced (A. Pringsheim).

Among continued fractions with variable terms, those of the form

$$[a_0, a_n z]_1^\infty = \frac{a_0}{1} + \frac{a_1 z}{1} + \cdots + \frac{a_n z}{1} + \cdots$$

are called **normal continued fractions**. Let the convergent of such a continued fraction be $P_n(z)/Q_n(z)$, and for convenience put $P_{-1}(z) = 0$, $Q_{-1}(z) = 1$. Then we have the recurrence formulas

$$P_n(z) = P_{n-1}(z) + a_n z P_{n-2}(z),$$

$$Q_n(z) = Q_{n-1}(z) + a_n z Q_{n-2}(z), \quad n \geqslant 1.$$

There are the further relations

$$P_n(z) Q_{n-1}(z) - P_{n-1}(z) Q_n(z) = (-1)^n z^n \prod_{v=0}^{n} a_v,$$

$$[a_0, a_n z]_1^\infty = \sum_{n=0}^{\infty} \frac{(-1)^n z^n}{Q_{n-1}(z) Q_n(z)} \prod_{v=0}^{n} a_v,$$

where the latter is formal. Let the †power series expansion of the nth convergent of $[a_0, a_n z]_1^\infty$ be

$$\frac{P_n(z)}{Q_n(z)} = \sum_{v=0}^{\infty} b_{nv} z^v, \quad n \geqslant 1.$$

Then $b_{mv} = b_{nv}$ $(0 \leqslant v \leqslant m \leqslant n)$. If $[a_0, a_n z]_1^\infty$ has a power series expansion about the origin, then

$$[a_0, a_n z]_1^\infty = \sum_{n=0}^{\infty} b_{nn} z^n.$$

If the supremum $\bar{g}$ of $\{|a_n|\}_1^\infty$ is finite, then $[a_0, a_n z]_1^\infty$ converges uniformly for $|z| \leqslant (1/4)\,\bar{g}$, and hence it represents an †analytic function which is holomorphic in $|z| < (1/4)\bar{g}$.

References

[1] G. H. Hardy and E. M. Wright, An introduction to the theory of numbers, Oxford Univ. Press, 1938.
[2] H. M. Stark, An introduction to number theory, Markham, 1970.
[3] W. M. Schmidt, Diophantine approximation, Lecture notes in math. 785, Springer, 1980.
[4] O. Perron, Die Lehre von den Kettenbrüchen, Teubner, third edition, 1954 (Chelsea, 1950).
For functions defined by continued fractions and applications to moment problems,
[5] H. S. Wall, Analytic theory of continued fractions, Chelsea, 1967.
[6] A. Ya. Khinchin (Hinčin), Continued fractions, Noordhoff, 1963. (Original in Russian, 1949.)
[7] A. N. Khovanskiĭ (Hovanskiĭ), The application of continued fractions and their generalizations to problems in approximation theory, Noordhoff, 1963. (Original in Russian, 1956.)

84 (X.2) Continuous Functions

A. General Remarks

The notion of continuity is defined for a mapping or a function $f: X \rightarrow Y$ from a topological space X to a topological space Y (→ 425 Topological Spaces G). In the present article, however, we are concerned mainly with the case where both X and Y are †metric spaces with the distances ρ_X and ρ_Y, respectively. The most usual case is $X = \mathbf{R}^n$ (Euclidean space), $Y = \mathbf{R}$ (real numbers).

A function $f: X \rightarrow Y$ is said to be **continuous** at a point $x_0 \in X$ if for every positive number ε, we can select a suitable positive number δ (depending on ε and also on x_0) such that $\rho_X(x, x_0) < \delta$ implies $\rho_Y(f(x), f(x_0)) < \varepsilon$. This is equivalent to the condition that $x \rightarrow x_0$ implies $f(x) \rightarrow f(x_0)$ (→ 87 Convergence). We call f **continuous** (on X) if it is continuous at every point x_0 of X. If for every positive number ε, we can select a suitable positive number δ independent of x and y such that $\rho_X(x, y) < \delta$ implies $\rho_Y(f(x), f(y)) < \varepsilon$ for all $x, y \in X$, we call f **uniformly continuous** on X.

The †supremum $\omega(\delta)$ of $\rho_Y(f(x), f(y))$ for $x, y \in X$ satisfying $\rho_X(x, y) < \delta$ is called the **modulus of continuity** of the function f in X. Uniform continuity means that $\omega(\delta) \rightarrow 0$ for $\delta \rightarrow 0$.

If $\omega(\delta) \leqslant M\delta^\alpha$ for suitable constants $M, \alpha > 0$, that is, if the inequality $\rho_Y(f(x), f(y)) \leqslant M(\rho_X(x, y))^\alpha$ holds for $x, y \in X$, then f is said to satisfy the **Hölder condition of order** α, also known as the **Lipschitz condition of order** α. If $\alpha = 1$, this condition is called simply the **Lipschitz condition**. A function satisfying one of these conditions is uniformly continuous. The family of functions satisfying the Lipschitz condition of order α is sometimes denoted by **Lip** α.

In general, the composite function $g \circ f: X \rightarrow Z$ is continuous if both functions $f: X \rightarrow Y$ and $g: Y \rightarrow Z$ are continuous. If the ranges of f, g are both the real field $\mathbf{R}$ (or the complex field $\mathbf{C}$, or more generally a †topological field), then $f \pm g$, fg are continuous if f and g are continuous; and f/g is continuous provided that $g(x) \neq 0$. If $\mathbf{R}$ is the range of both f and g, then $\min(f, g)$ and $\max(f, g)$ are continuous when f and g are continuous. If X is †connected (for example, an interval I in $\mathbf{R}$) and if f is continuous, the image $f(X)$ is also connected.

B. Continuity from One Side

In this section, we always assume that the domain X is an interval I in $\mathbf{R}$ and f is a function from I to a metric space Y. A point x_0 of X is called a **discontinuity (point) of the first kind** of f if both limits $\lim_{x \uparrow x_0} f(x)$ and $\lim_{x \downarrow x_0} f(x)$ exist in Y and are different. Then we say also that f has a **jump** (or **gap**) at x_0. If these two limits exist and have the same value, then f is continuous at x_0.

We say that f has a discontinuity **of at most the first kind** at x_0 if f is continuous at x_0 or if x_0 is a discontinuity of the first kind for f. (Sometimes the phrase "discontinuity of the first kind" is used to mean "discontinuity of at most the first kind.") A **discontinuity point** of f (i.e., a point at which f is not continuous) that is not of the first kind is called a **discontinuity point of the second kind**. When $\lim_{x \downarrow x_0} f(x) = f(x_0)$, we call f **right continuous** (or **continuous from the right**) at x_0. In this case, $\lim_{x \uparrow x_0} f(x)$ need not exist. Replacing $x \downarrow x_0$ by $x \uparrow x_0$, we can similarly define the concept of being **left continuous** (or **continuous from the left**). If a function f has a finite number of discontinuity points of the first kind in the interval $[a, b]$ and is continuous at all other points, we call f a **piecewise continuous function** in $[a, b]$.

C. Semicontinuous Functions

In this section, we assume that the domain of the functions is a subset E of a metric space X, and that the range is the set of real numbers extended to include $\pm\infty$. Let x be a point in the closure of E. We denote by $M(x, \delta)$ and $m(x, \delta)$, respectively, the supremum and the infimum of the values of a given function f in the δ-neighborhood of x. We put

$$M(x) = \lim_{\delta \rightarrow 0} M(x, \delta), \qquad m(x) = \lim_{\delta \rightarrow 0} m(x, \delta),$$

and call them the **upper limit function** and **lower limit function**, respectively. We have $-\infty \leqslant m(x) \leqslant M(x) \leqslant +\infty$. If $M(x_0) = f(x_0)$ at $x_0 \in E$, then f is called **upper semicontinuous** at x_0. If $m(x_0) = f(x_0)$ at $x_0 \in E$ (i.e., if $-f$ is upper semicontinuous at x_0), then f is called **lower semicontinuous** at x_0. The function with one of these two properties is said to be **semicontinuous** at x_0.

Either of the following two conditions is necessary and sufficient for the function f to be upper semicontinuous at $x_0 \in E$: (1) $f(x_0) = +\infty$, or for every constant λ such that $f(x_0) < \lambda$, there exists a δ-neighborhood such that $M(x_0, \delta) < \lambda$. (2) For every sequence x_n of E converging to x_0 we have $\limsup_{n \rightarrow \infty} f(x_n) = f(x_0)$.

A function f is called **upper (lower) semicontinuous in** E if it is upper (lower) semicontinuous at every point $x \in E$. A necessary and sufficient condition for the upper semicontinu-

ity of the function $f(x)$ in E is that $\{x \mid f(x) < \alpha\}$ be a †relatively open set in E for every real number α. We can define semicontinuity for functions on a topological space by using this latter property.

A real-valued function $f(x)$ is continuous at $x_0 \in E$ if and only if it is upper and lower semicontinuous at x_0 and $f(x_0)$ is finite. A function $f(x)$ is continuous on E if and only if it takes finite real values on E and both $\{x \mid f(x) < \alpha\}$ and $\{x \mid f(x) > \alpha\}$ are relatively open in E for any real number α. When E is †compact, an upper (lower) semicontinuous function on E attains its (supremum) infimum at a point in E. In particular, a continuous function on a compact set E is bounded and assumes its maximum and minimum on E (**Weierstrass's theorem**). Furthermore, if E is connected (e.g., the interval I in $\mathbf{R}$), it follows from the connectedness of the image $f(E)$ that if $\alpha, \beta \in f(E)$ and γ lies between α and β, then $\gamma \in f(E)$ (**intermediate-value theorem**).

A real-valued function $f(x)$ on a set E of $\mathbf{R}$ satisfies the Lipschitz condition if it is †differentiable and the derivative is bounded. Such a function is also †absolutely continuous, continuous, and of †bounded variation. (For the polynomial approximation of real continuous functions → 336 Polynomial Approximation.)

The limit function $f(x)$ of a monotone decreasing sequence of upper semicontinuous functions $f_n(x)$ is also upper semicontinuous. The limit function $f(x)$ of a uniformly converging sequence of continuous functions is continuous. (Regarding the †equicontinuous family of functions → 435 Uniform Convergence.)

D. Baire Functions

The limit function of a pointwise converging sequence of continuous functions defined on a metric space X is not necessarily continuous. R. Baire (*Ann. Mat. Pura Appl.*, (1899)) introduced the notion of Baire functions as follows: He named continuous functions the **functions of class** 0. Then he called a function that is a pointwise limit of a sequence of continuous functions a **function of at most class** 1. A function is said to be of class 1 if it is of at most class 1 and is not of class 0. He similarly defined the notion of **class** n for arbitrary natural number n.

Further, a function is said to be of at most class ω if it is a pointwise limit of a sequence of functions of class n_ν for a sequence of natural numbers n_ν. A function is said to be of **class** ω if it is of at most class ω and is not of class n for any finite number n. In general, using †transfinite induction, we can define the notion of functions of **class** ξ for an arbitrary †ordinal number ξ [4].

All these functions are called **Baire functions**. Actually, there is no function of class ξ for an uncountable ordinal number ξ. If X is a †perfect set in Euclidean space, then there is actually a function defined on X of class ξ for an arbitrary countable ordinal number ξ. Hereafter, we shall be concerned with this case only.

If X has the cardinality of the †continuum, then the set of all Baire functions on X has the cardinality of the continuum. On the other hand, the cardinality of all functions on X is actually greater than that of the continuum. Hence there exist functions that are not Baire functions on X. A function is a Baire function if and only if it is †Borel measurable (H. Lebesgue). Therefore a necessary and sufficient condition for a function f to be a Baire function is that the set $(x \mid f(x) > \alpha, x \in X\}$ be a Borel set for any real number α (→ 270 Measure Theory J). The limit of a countable sequence of Baire functions is also a Baire function. If $f(x)$ and $g(x)$ are of at most class α on X, then the following functions are also of at most class α: $|f(x)|$, $f(x) \pm g(x)$, $f(x) \cdot g(x)$, and $f(x)/g(x)$ (provided that $g(x) \neq 0$ on X).

The condition that a function f is of at most class 1 on X is equivalent to either of the following two conditions: (1) For any closed subset F of X, the restriction f^* of f to F has a continuity point in F. (2) For every real number α, the set $\{x \mid f(x) < \alpha, x \in X\}$ is an †F_σ set (Baire). In a †complete metric space, a necessary and sufficient condition for a function f to be of at most class 1 is that the set of continuity points be dense in X.

For example, the **Dirichlet function**, which takes the value 1 at rational points and 0 at irrational points, is expressed as

$$\lim_{\nu \to \infty} \left(\lim_{k \to \infty} (\cos \nu! \, \pi x)^{2k} \right),$$

which is of class 2. A function $f(x, y)$ of two real variables that is continuous in each variable x and y separately is a function of at most class 1.

References

[1] S. Saks, Theory of the integral, Warsaw, 1937.
[2] N. Bourbaki, Eléments de mathématique, Topologie générale, ch. 10, Actualités Sci. Ind., 1084b, Hermann, second edition, 1967.
For Baire functions,
[3] R. Baire, Leçons sur les fonctions discontinues, Gauthier-Villars, 1905.
[4] C. J. de La Vallée-Poussin, Intégrales de Lebesgue, fonctions d'ensemble, classes de Baire, Gauthier-Villars, second edition, 1934.

[5] W. Rudin, Principles of mathematical analysis, McGraw-Hill, second edition, 1964.

85 (VI.19) Continuous Geometry

A. General Remarks

The structure of a †projective geometry is determined by the †lattice (lattice-ordered set) of subspaces of the projective space. For this reason, this lattice itself is sometimes called a projective geometry. The concept of continuous geometry was introduced by J. von Neumann as an abstraction of lattice-theoretic properties from a special class of †lattices (lattice-ordered sets) which he encountered in his research on †operator rings in Hilbert spaces [1]. Continuous geometry contains projective geometry as a special case when the dimension is discrete; but more usually the lattices are of continuous dimension.

A **continuous geometry** is a †complete and †complemented †modular lattice L (→ 243 Lattices F) that satisfies the following property and its dual (both called **properties of continuity**): For any element a of L and any subset W of L which is †well ordered with respect to the ordering in L, we have $a \cap \sup w = \sup(a \cap w)$ $(w \in W)$. The †center Z of the lattice L is called the **center** of the continuous geometry L, which is said to be **irreducible** when Z has no elements other than the †least element 0 and the †greatest element I; otherwise L is said to be **reducible**. A reducible continuous geometry is isomorphic to a sublattice of a †direct product of irreducible continuous geometries.

On any continuous geometry L, there can be defined a function $d(x)$ whose values belong to a complete lattice-ordered linear space M and which satisfies the following four conditions: (1) $d(x) \geqslant 0$; (2) $d(x) = d(y)$ implies the existence of a common complement of x and y; (3) $d(x \cup y) + d(x \cap y) = d(x) + d(y)$; (4) $\sup d(w) = d(\sup w)$ $(w \in W)$ for any subset W of L which is well ordered with respect to the ordering in L. Such a function $d(x)$ is called a **dimension function** on L. Irreducibility of L is equivalent to the property that a real-valued dimension function can be introduced; in this case, if $d(w)$ takes only a finite number of values, then L is a finite-dimensional projective geometry; on the other hand, if $d(w)$ takes every number in the closed interval $[0, 1]$, then L is called a continuous geometry in the strict sense. An example of the latter can be constructed as a limit of a sequence of projective geometries of increasing dimensions; another example is the lattice of projection operators of the †von Neumann algebra of type II_1. If a group G of †automorphisms of L is given, there can be introduced a generalized dimension function which is invariant under G satisfying slightly weaker conditions (T. Iwamura [3]).

B. Representation of Continuous Geometry

A ring R is called a **regular ring** if it has a unity element and if, for any element a of R, there exists an element x in R such that $axa = a$. A continuous geometry L is isomorphic to the lattice (with $\supset$ as its ordering) of †principal left ideals of a regular ring R provided that $d(I) = n \cdot d(x)$ $(n \geqslant 4)$ for some natural number n and some element x of L. The decomposition of R into a †direct sum of ideals corresponds to the decomposition of L into a direct product of lattices. The condition that L is irreducible and finite-dimensional is equivalent to the condition that R is a matrix ring over a †skew field; when these conditions hold and L is considered as a projective geometry, then the coordinates are given by this skew field. In continuous geometries, join and meet are often denoted by the symbols for sum and product, respectively, and sometimes a direct product of continuous geometries is called their direct sum. Sometimes, the requirement that a continuous geometry be complete is weakened to the requirement that it be †conditionally σ-complete.

References

[1] J. von Neumann, Continuous geometry, Princeton Univ. Press, 1960 (Lecture notes, 1935–1937).
[2] F. Maeda, Continuous geometry (in Japanese), Iwanami, 1952; German translation, Kontinuierliche Geometrien, Springer, 1958.
[3] T. Iwamura, On continuous geometries I, II, Japan. J. Math., 19 (1944), 57–71; J. Math. Soc. Japan, 2 (1950), 148–164.
[4] L. A. Skornyakov, Complemented modular lattices and regular rings, Oliver & Boyd, 1964. (Original in Russian, 1961.)

86 (XIX.11) Control Theory

A. General Remarks

The classical theory of automatic control mostly deals with linear feedback control

systems with single input and single output. Mathematical structures of such systems must be, in principle, described in terms of ordinary linear differential equations with constant coefficients. Hence control engineers use block diagrams to describe systems, and operational calculus based on †Laplace transforms to obtain response characteristics (→ 240 Laplace Transform). Thus the input/output relation of a system is described in terms of transfer functions. The main objectives of control theory are to ensure system stability, to maximize closed-loop response characteristics by choosing the best feedback, and to obtain a desirable transient response to an impulse or step input. One of the remarkable contributions to classical control theory is Nyquist's criterion for stability testing of linear feedback systems [1]. The test consists of plotting the Nyquist diagram of a transfer function in the frequency domain (complex plane), and differs essentially from the †Routh-Hurwitz stability test for linear differential equations with constant coefficient. Classical control theory was almost complete by the end of World War II.

Revolutionary technological innovations in electronics and computers and the invention of new control instruments and systems developed after World War II have opened the way to modern control theory. Around 1960, three remarkable contributions were made concurrently; they are dynamic programming (proposed by R. E. Bellman [2]), Pontryagin's maximum principle (L. S. Pontryagin et al. [3]), and linear system theory (R. E. Kalman [4]). The first two give rise to mathematical tools to solve optimal control problems and to design optimal controllers or regulators. In contrast to the classical theory of control, optimal control problems are formulated in terms of the system of linear or nonlinear multivariable differential equations with multi-input forcing terms called control variables. This leads to the **state-space approach**, which has become ubiquitous in modern control theory. Linear system theory derives from the concepts of controllability and observability. The former was introduced by Pontryagin et al. [3] and later modified by Kalman [5], and the latter was introduced by Kalman [5]. Controllability and observability are concerned with the interrelation between internal states of a system and its inputs and outputs. Hence linear system theory, profoundly related to the state-space approach, remains a principal theme of modern control theory. More advanced control theories, such as adaptive control or learning control, are mostly described and formulated in the framework of linear system theory.

Modern control theory has stimulated the development of †cybernetics, the science of control and communication, and control theory is important in the information sciences. In fact, current control theory has many features that might be considered to belong as well to other areas, such as mathematical programming (→ 264 Mathematical Programming), operations research (→ 307 Operations Research), game theory (→ 173 Game Theory), prediction and filtering theory, digital signal processing, circuit theory, and computer or microprocessor technology.

Current control theory embodied in modern instrumentation and computer technology has a vast range of applications. We do not consider these applications here; we confine ourselves to some important theories that continue to be major topics in control theory.

B. Linear Dynamical System

Let the state of a physical process to be controlled be represented by a real n-dimensional column vector $x(t)=(x_1(t),\ldots,x_n(t))^T\in\mathbf{R}^n$, where ^{T} denotes the transpose of a vector or matrix. If the state is determined by a differential system with input $u(t)=(u_1(t),\ldots,u_r(t))^T\in\mathbf{R}^r$ and output $y(t)=(y_1(t),\ldots,y_m(t))^T\in\mathbf{R}^m$,

$$\frac{d}{dt}x(t)=A(t)x(t)+B(t)u(t),$$

$$y(t)=C(t)x(t), \tag{1}$$

then the process is called a **linear dynamical system** or simply a linear system. The dimensions of the coefficient matrices are given by

$$A(t)\in\mathbf{R}^{n\times n},\quad B(t)\in\mathbf{R}^{n\times r},\quad C(t)\in\mathbf{R}^{m\times n}.$$

If $A(t)$, $B(t)$, and $C(t)$ are constant matrices, the system described by

$$\dot{x}=\frac{d}{dt}x=Ax+Bu,$$

$$y=Cx \tag{2}$$

is called a **linear time-invariant** system. Correspondingly, the system described by equation (1) is called a **linear time-varying system**.

Given an initial state $x(t_0)$ and input function $u(t)$ for $t\geqslant t_0$, the state of system (1) is represented by

$$x(t)=\Phi(t,t_0)x(t_0)+\int_{t_0}^{t}\Phi(t,\tau)B(\tau)u(\tau)d\tau,$$

where $\Phi(t,\tau)$ is a †fundamental solution to the matrix differential equation

$$\frac{\partial}{\partial t}\Phi(t,\tau)=A(t)\Phi(t,\tau),$$

$$\Phi(\tau,\tau)=I\quad(n\times n\text{ identity matrix}).$$

Hence the output of the system is described by

$$y(t)=C(t)\Phi(t,t_0)x(t_0)+\int_{t_0}^{t}C(t)\Phi(t,\tau)B(\tau)u(\tau)\,d\tau.$$

It is seen from this equation that the input-output relation of the system is governed by

$$W(t,\tau)=C(t)\Phi(t,\tau)B(\tau)\in\mathbf{R}^{m\times r},$$

which is called a **weighting matrix**. As a special case, the weighting matrix for the time-invariant system (2) becomes

$$W(t,\tau)=W(t-\tau)=Ce^{A(t-\tau)}B,$$

where e^{At} is a matrix exponential function defined by

$$e^{At}=I+tA+\frac{t^2}{2!}A^2+\frac{t^3}{3!}A^3+\dots.$$

The Laplace transform of the weighting matrix is

$$G(s)=\int_0^\infty e^{-st}W(t)\,dt=C(sI-A)^{-1}B,$$

which is called a **transfer function matrix** of the system (2). First we note that the weighting matrix $W(t,\tau)$ for the system (1) is invariant under any linear transformation $\tilde{x}(t)=P(t)x(t)$ of the state, and also that the transfer function matrix $G(s)$ for the system (2) is invariant under $\tilde{x}=Px$. For example, the transformation $\tilde{x}=Px$ yields

$$\dot{\tilde{x}}=PAP^{-1}\tilde{x}+PBu,$$

$$y=CP^{-1}\tilde{x}. \qquad (3)$$

In fact, the transfer function matrix for the system (3) becomes equivalent to that for the system (2), that is,

$$\tilde{G}(s)=CP^{-1}(sI-PAP^{-1})^{-1}PB=C(sI-A)^{-1}B=G(s).$$

As a converse to this, there arises the problem of whether any linear time-invariant systems with the same transfer function matrix are similar to each other. To solve this problem, it is necessary to introduce the concept of controllability and observability for the system (1) or (2).

C. Controllability and Observability

The concept of controllability and observability plays a fundamental role in linear system theory. Roughly, controllability implies the possibility of steering the state from the input.

Definition of **controllability**: The linear system (1) is said to be controllable at time t_0 if for any state $x(t_0)\in\mathbf{R}^n$ and $x^1\in\mathbf{R}^n$ there exists a control input $u(t)$, $t\in[t_0,t_1]$, that transfers the state $x(t_0)$ to the state x^1 at a finite time $t_1>t_0$. Otherwise, the system (1) is said to be uncontrollable at time t_0.

The controllability theorem states: The system (1) is controllable at time t_0 if and only if there exists a finite $t_1(>t_0)$ such that the matrix

$$D(t_0,t_1)=\int_{t_0}^{t_1}\Phi(t_0,t)B(t)B^{\mathrm{T}}(t)\Phi^{\mathrm{T}}(t_0,t)\,dt \qquad (4)$$

is nonsingular. Similarly, the linear time-invariant system (2) is controllable if and only if the $n\times(nr)$ matrix

$$D=[B,AB,\dots,A^{n-1}B] \qquad (5)$$

has rank n. The matrix D is called the controllability matrix of the system (2).

The concept of observability is dual to that of controllability. It implies the possibility of estimating the state from the output.

Definition of **observability**: The linear system (1) is said to be observable at t_0 if for any state $x(t_0)\in\mathbf{R}^n$ there exists a finite $t_1(>t_0)$ such that the knowledge of the input $u(t)$ and the output $y(t)$ over the time interval $[t_0,t_1]$ suffices to determine the state $x(t_0)$. Otherwise, the system (1) is said to be unobservable at t_0.

The observability theorem states: The system (1) is observable at time t_0 if and only if there exists a finite $t_1(>t_0)$ such that the matrix

$$M(t_0,t_1)=\int_{t_0}^{t_1}\Phi^{\mathrm{T}}(t,t_0)C^{\mathrm{T}}(t)C(t)\Phi(t,t_0)\,dt \qquad (6)$$

is nonsingular. Similarly, the linear time-invariant system (2) is observable if and only if the $n\times(nm)$ matrix

$$M=[C^{\mathrm{T}},A^{\mathrm{T}}C^{\mathrm{T}},\dots,(A^{\mathrm{T}})^{n-1}C^{\mathrm{T}}] \qquad (7)$$

has rank n. The matrix M is called the observability matrix of the system (2).

We deal only with linear time-invariant systems for simplicity, although some of the main results to be stated here can be extended to linear time-varying systems. If the controllability matrix D has rank n_c $(<n)$, there exists a linear transformation $\bar{x}=Px$ with nonsingular matrix P which transforms the system (2) into

$$\begin{bmatrix}\dot{x}_c\\ \dot{x}_{c'}\end{bmatrix}=\begin{bmatrix}A_c & A_{12}\\ 0 & A_{c'}\end{bmatrix}\begin{bmatrix}x_c\\ x_{c'}\end{bmatrix}+\begin{bmatrix}B_c\\ 0\end{bmatrix}u,\qquad y=[C_c\quad C_{c'}]\begin{bmatrix}x_c\\ x_{c'}\end{bmatrix} \qquad (8)$$

so that the n_c-dimensional subsystem

$$\dot{x}_c=A_cx_c+B_cu,\qquad y=C_cx_c \qquad (9)$$

is controllable. Clearly this subsystem has the same transfer function matrix as the system (2).

If the observability matrix M has rank n_o ($<n$), there exists a linear transformation $\bar{\bar{x}}=Qx$ that transforms the system (2) into

$$\begin{bmatrix}\dot{x}_o\\ \dot{x}_{o'}\end{bmatrix}=\begin{bmatrix}A_o & 0\\ A_{21} & A_{o'}\end{bmatrix}\begin{bmatrix}x_o\\ x_{o'}\end{bmatrix}+\begin{bmatrix}B_o\\ B_{o'}\end{bmatrix}u,$$
$$y=[C_o \quad 0]\begin{bmatrix}x_o\\ x_{o'}\end{bmatrix}, \tag{10}$$

so that the n_o-dimensional subsystem

$$\dot{x}_o=A_o x_o+B_o u,$$
$$y=C_o x_o \tag{11}$$

is observable. This subsystem also has the same transfer function matrix as the system (2). Combination of these two properties yields the **canonical decomposition theorem** (found by Kalman [6]), which states: There is a nonsingular transformation $\tilde{x}=Px$ by which the system (2) can be transformed into the canonical form

$$\begin{bmatrix}\dot{x}_{co}\\ \dot{x}_{co'}\\ \dot{x}_{c'o}\\ \dot{x}_{c'o'}\end{bmatrix}=\begin{bmatrix}A_{co} & 0 & A_{13} & 0\\ A_{21} & A_{22} & A_{23} & A_{24}\\ 0 & 0 & A_{33} & 0\\ 0 & 0 & A_{43} & A_{44}\end{bmatrix}\begin{bmatrix}x_{co}\\ x_{co'}\\ x_{c'o}\\ x_{c'o'}\end{bmatrix}$$
$$+\begin{bmatrix}B_{co}\\ B_2\\ 0\\ 0\end{bmatrix}u, \tag{12}$$
$$y=[C_{co} \quad 0 \quad C_3 \quad 0]\begin{bmatrix}x_{co}\\ x_{co'}\\ x_{c'o}\\ x_{c'o'}\end{bmatrix},$$

where the subvector x_{co} of $\tilde{x}$ is controllable and observable, $x_{co'}$ is controllable but not observable, $x_{c'o}$ is not controllable but observable, and $x_{c'o'}$ is neither controllable nor observable. Furthermore, the transfer function matrix of the system (2) is equal to that of the system (12), which is described by

$$G(s)=C_{co}(sI-A_{co})^{-1}B_{co},$$

that is, the transfer function matrix of the system (2) depends solely on the controllable and observable part of the state.

A linear system (2) is said to be **reducible** if and only if there exists a linear time-invariant system of smaller dimension that has the same transfer function matrix. Otherwise, the system is said to be **irreducible**. The canonical decomposition theorem implies that a linear time-invariant system is irreducible if and only if it is controllable and observable.

D. Realization Theory

Realization theory is concerned with determining (1) a linear time-invariant dynamical system that has a prescribed **rational function matrix** or (2) a linear time-varying system that has a prescribed weighting matrix. Here we consider only the former, although there are many interesting results for the latter as well.

The rational function matrix $\tilde{G}(s)$ is said to be strictly proper if and only if $\tilde{G}(\infty)=0$. A linear time-invariant system with a prescribed $\tilde{G}(s)$ as a transfer function matrix is called a **realization** of $\tilde{G}(s)$. A realization of $\tilde{G}(s)$ with the least possible dimension is said to be minimal. A minimal realization should be irreducible by definition. The fundamental theorem on minimal realization, which answers the question raised at the end of Section B, states: For a given strictly proper rational matrix $\tilde{G}(s)$, any two minimal realizations of $\tilde{G}(s)$ are mutually similar in the sense that there exists a linear transformation $\tilde{x}=Px$ that transforms one of the systems into the other.

A **minimal realization** of a scalar **transfer function**

$$\tilde{g}(s)=\frac{b_1 s^{n-1}+\ldots+b_{n-1}s+b_n}{s^n+a_1 s^{n-1}+\ldots+a_{n-1}s+a_n} \tag{13}$$

is given by

$$\dot{x}=\begin{bmatrix}0 & 1 & 0 & \ldots & 0\\ 0 & 0 & 1 & \ldots & 0\\ & & & \ldots & \\ 0 & 0 & 0 & \ldots & 1\\ -a_n & -a_{n-1} & -a_{n-2} & \ldots & -a_1\end{bmatrix}x$$
$$+\begin{bmatrix}0\\ 0\\ \vdots\\ 0\\ 1\end{bmatrix}u, \tag{14}$$
$$y=[b_n \quad b_{n-1} \quad b_{n-2} \quad \ldots \quad b_1]x.$$

It is easy to see that this system is controllable. Hence system (14) is called a controllable canonical realization. Kalman [4] showed that this realization is observable and therefore minimal if and only if there is no common factor between the denominator and the numerator of $\tilde{g}(s)$. An observable canonical realization of $\tilde{g}(s)$ can be similarly written down as a dual of the system (14).

For a given strictly proper matrix $\tilde{G}(s)$, the degree of the least common denominator of all entries of $\tilde{G}(s)$ is called the degree of $\tilde{G}(s)$, provided that there is no common factor between the denominator and the numerator of each entry. It was shown by Kalman [7], together with a realization procedure, that the dimension of minimal realization of $\tilde{G}(s)$ is equal to the degree of $\tilde{G}(s)$. Other important realization procedures for rational function matrices were proposed by B. L. Ho and Kalman [8], W. A. Wolovich and P. L. Falb [9, 10], J. Rissanen and T. Kailath [11], and

B. Dickinson et al. [12]. Procedures for time-varying system realization for prescribed weighting matrices were presented by Kalman [4], D. C. Youla [15], and L. M. Silverman [16]. Realization problems for a given impulse response matrix were first discussed by C. A. Desoer and P. P. Varaiya [17], and a fast algorithm for discrete-time system realization was proposed by C. T. Mullis and R. A. Roberts [18].

E. State Estimation

When a linear system is observable, it is possible to construct a device, called a **state estimator**, that approximates the state vector. For the given system (2), define a linear dynamical system

$$\dot{z} = Dz + Ey + Gu,$$

$$\hat{x} = L_1 z + L_2 y \qquad (15)$$

with imputs $y \in \mathbf{R}^m$ and $u \in \mathbf{R}^r$, state $z \in \mathbf{R}^q$, and output $\hat{x} \in \mathbf{R}^n$. If the system (2) is observable, rank $(C) = m$, and $q \geqslant n - m$, there exist matrices D, E, G, L_1, L_2, and M that satisfy

$$MA - DM = EC,$$

$$L_1 M + L_2 C = I_n \quad (n \times n \text{ identity matrix}),$$

$$G = MB.$$

Then it follows that

$$\hat{x}(t) - x(t) = L_1 e^{Dt}[z(0) - Mx(0)].$$

Furthermore, it is possible to choose a matrix D which is asymptotically stable, that is, every eigenvalue of D has negative real part. Thus $\hat{x}(t) \to x(t)$ as $t \to \infty$, and system (15) yields a state estimator. This result was first shown by D. G. Luenberger [19, 20], and thus the system (15) is called a **Luenberger observer**.

The **Kalman filter** proposed by Kalman [21] for discrete-time processes and by Kalman and R. S. Bucy [22] for continuous-time processes is an optimal state estimator for a stochastic system. Let

$$\dot{x}(t) = A(t)x(t) + B(t)u(t),$$

$$y(t) = C(t)x(t) + v(t), \qquad (16)$$

in which $u(t)$ is a white Gaussian noise (†Brownian motion) (→ 45 Brownian Motion) such that

$$Eu(t) \equiv 0, \qquad Eu(t)u^{\mathrm{T}}(s) = U(t)\delta(t-s),$$

and $v(t)$ is also a white Gaussian noise such that

$$Ev(t) \equiv 0, \qquad Ev(t)v^{\mathrm{T}}(s) = V(t)\delta(t-s),$$

where E and $^{\mathrm{T}}$ denote the †expectation and the transpose, respectively. In addition, it is assumed that

$$Ex(t_0) = 0, \qquad Ex(t_0)x^{\mathrm{T}}(t_0) = X_0,$$

and that $x(t_0)$, $u(t)$, and $v(s)$ are uncorrelated. Under these conditions, the process $x(t)$ becomes a wide-sense Gauss-Markov process (J. L. Doob [23]). The **Kalman-Bucy filter** is a linear dynamical system described by

$$\dot{\hat{x}}(t) = [A(t) - L(t)C(t)]\hat{x}(t) + L(t)y(t),$$

where

$$L(t) = P(t)C^{\mathrm{T}}(t)V^{-1}(t)$$

and $P(t)$ is a solution to the **matrix Riccati differential equation**

$$\dot{P}(t) = A(t)P(t) + P(t)A^{\mathrm{T}}(t) + B(t)U(t)B^{\mathrm{T}}(t)$$
$$\qquad - P(t)C^{\mathrm{T}}(t)V^{-1}(t)C(t)P(t),$$

$$P(t_0) = X_0.$$

The **Wiener filter** proposed in 1949 [24] is a special case of the Kalman filter in which the signal process is stationary and of single input and single output. Extensions to nonlinear filtering were investigated first by H. J. Kushner [25] and subsequently by W. M. Wonham [26], R. S. Liptzer and A. N. Shiryaev [27], and M. Fuzisaki et al. [28].

F. Optimal Control

Consider a nonlinear dynamical system described by a system of differential equations

$$\dot{x} = f(x, u), \qquad (17)$$

where $x \in \mathbf{R}^n$ and $u \in \mathbf{R}^r$. For a given set U in $\mathbf{R}^r$, a piecewise continuous function $u(t)$ defined over $[t_0, t_1]$ is said to be an admissible control function if $u(t) \in U$ for every $t \in [t_0, t_1]$. When a functional

$$J = \int_{t_0}^{t_1} f_0(x(t), u(t))\, dt$$

is given as a performance index of control and two points x^0 and x^1 are given in $\mathbf{R}^n$, the optimal control problem consists of finding an admissible function $u(t)$ that minimizes J and simultaneously transforms the state from $x(t_0) = x^0$ to $x(t_1) = x^1$. To solve this, a necessary condition for optimality was developed by L. S. Pontryagin and his colleagues [3] and termed the **maximum principle**. It states: In order for an admissible control $u(t)$ and a corresponding solution trajectory of equation (17) to be optimal, it is necessary that there exist a nonzero vector $\psi(t) = (\psi_1(t), \ldots, \psi_n(t))^{\mathrm{T}}$ such that (i) $\psi(t)$ satisfies

$$\dot{\psi}(t) = -\frac{\partial H(\psi(t), x(t), u(t))}{\partial x(t)},$$

where

$H(\psi, x, u) = -f_0(x, u) + \psi^T f(x, u),$

(ii) the function $H(\psi(t), x(t), u)$ of the variable $u \in U$ attains its maximum at the point $u = u(t)$ almost everywhere in $[t_0, t_1]$, and (iii) at the terminal time t_1 it holds that

$H(\psi(t_1), x(t_1), u(t_1)) = 0.$

When $f(x, u)$ is linear in x and u, $f_0(x, u) \equiv 1$, and t_1 is free, we have the **time-optimal control** problem investigated by Bellman et al. [29] and J. P. LaSalle [30]. When $f_0(x, u)$ is instead a quadratic function of x and u, we have the **optimal regulator problem** solved by Kalman [31]. The relation between the maximum principle and the †calculus of variations (→ 46 Calculus of Variations) has been pointed out by L. D. Berkovitz [32]. Generalizations of the maximum principle and other results on optimal control problems are found in [33–35].

G. Miscellany

There are many other areas of control theory. For details, we refer the reader to [36] for adaptive control, [37] for stability theory, and [38] for system identification.

References

[1] H. Nyquist, Regeneration theory, Bell Syst. Tech. J., 11 (1932), 126–147.
[2] R. E. Bellman, Dynamic programming, Princeton Univ. Press, 1957.
[3] L. S. Pontryagin, V. G. Boltyanskiĭ, R. V. Gamkrelidze, and E. F. Mishchenko, The mathematical theory of optimal processes, Interscience, 1962.
[4] R. E. Kalman, Mathematical description of linear dynamical systems, SIAM J. Control, 1 (1963), 152–192.
[5] R. E. Kalman, On the general theory of control systems, Proc. First Int. Congress Automatic Control, Butterworth, 1960, 481–493.
[6] R. E. Kalman, Canonical structure of linear dynamical systems, Proc. Nat. Acad. Sci. US, 48 (1962), 596–600.
[7] R. E. Kalman, Irreducible realization and the degree of a rational matrix, SIAM J. Appl. Math., 13 (1965), 520–544.
[8] B. L. Ho and R. E. Kalman, Effective construction of linear state variable models from input/output data, Regelungstechnik, 12 (1966), 545–548.
[9] W. A. Wolovich and P. L. Falb, On the structure of multivariable systems, SIAM J. Control, 7 (1969), 437–451.
[10] W. A. Wolovich, The determination of state-space representations for linear multivariable systems, Automatica, 9 (1973), 97–106.
[11] J. Rissanen and T. Kailath, Partial realization of random systems, Automatica, 8 (1972), 389–396.
[12] B. Dickinson, M. Morf, and T. Kailath, A minimal realization algorithm for matrix sequences, IEEE Trans. Automatic Control, AC-19 (1974), 31–38.
[13] B. D. O. Anderson, A system theory criterion for positive real matrices, SIAM J. Control, 5 (1967), 171–182.
[14] B. D. O. Anderson, An algebraic solution to the spectral factorization problems, IEEE Trans. Automatic Control, AC-12 (1967), 410–414.
[15] D. C. Youla, The synthesis of linear dynamical systems from prescribed weighting patterns, SIAM J. Appl. Math., 14 (1966), 527–549.
[16] L. M. Silverman, Synthesis of impulse response matrices by internally stable and passive realizations, IEEE Trans. Circuit Theory, CT-15 (1968), 238–245.
[17] C. A. Desoer and P. P. Varaiya, The minimal realization of a nonanticipative impulse response matrix, SIAM J. Appl. Math., 15 (1967), 754–764.
[18] C. T. Mullis and R. A. Roberts, The use of second-order information in the approximation of discrete-time linear systems, IEEE Trans. Acoust. Speech Signal Process., ASSP-24 (1976), 226–238.
[19] D. G. Luenberger, Observing the state of a linear system, IEEE Trans. Mil. Electron., MIL-8 (1964), 74–80.
[20] D. G. Luenberger, Observers for multivariable systems, IEEE Trans. Automatic Control, AC-11 (1966), 190–197.
[21] R. E. Kalman, A new approach to linear filtering and prediction problems, Trans. ASME, J. Basic Eng., 82 (1960), 35–45.
[22] R. E. Kalman and R. S. Bucy, New results in linear filtering and prediction theory, Trans. ASME, J. Basic Eng., 83 (1961), 95–108.
[23] J. L. Doob, Stochastic processes, Wiley, 1953.
[24] N. Wiener, Extrapolation, interpolation and smoothing of stationary time series, with engineering applications, Wiley, 1949.
[25] H. J. Kushner, On the dynamical equations of the conditional density functions with applications to optimal stochastic control theory, J. Math. Appl., 8 (1964), 332–344.
[26] W. M. Wonham, Some applications of stochastic differential equations to optimal nonlinear filtering, SIAM J. Control, 2 (1965), 347–369.
[27] R. S. Liptzer and A. N. Shiryaev, Non-

linear filtering of Markov diffusion processes (in Russian), Trudy Mat. Inst. Steklov., 104 (1968), 135–180.
[28] M. Fuzisaki, G. Kallianpur, and H. Kunita, Stochastic differential equations for the nonlinear filtering problem, Osaka J. Math., 9 (1972), 19–40.
[29] R. Bellman, I. Glicksberg, and O. Gross, On the "bang-bang" control problem, Quart. Appl. Math., 14 (1956), 11–18.
[30] J. P. LaSalle, The time optimal control problem, Contributions to the Theory of Nonlinear Oscillations V, Ann. Math. Studies, Princeton Univ. Press (1960), 1–24.
[31] R. E. Kalman, Contributions to the theory of optimal control, Bol. Soc. Mat. Mexicana, second ser., 5 (1960), 102–119.
[32] L. D. Berkovitz, Variational methods in problems of control and programming, J. Math. Anal. Appl., 3 (1961), 145–169.
[33] E. B. Lee and L. Markus, Foundations of optimal control theory, Wiley, 1967.
[34] A. G. Butkovskiĭ, Theory of optimal control of distributed parameter systems, American Elsevier, 1969.
[35] J. L. Lions, Optimal control of systems governed by partial differential equations, Springer, 1971.
[36] Y. D. Landau, Adaptive control, Dekker, 1979.
[37] N. P. Bhatia and G. P. Szegö, Stability theory of dynamical systems, Springer, 1970.
[38] R. K. Mehra and D. G. Lainiotis (eds.), System identification: Advances and case studies, Academic Press, 1976.

87 (II.19) Convergence

A. Introduction

The notion of convergence was first introduced in the real number system to deal with sequences of numbers, functions, series, or definite integrals (→ 379 Series; 216 Integral Calculus). The notion was then extended to the case of generalized sequences where the index moves over a directed set, and the terms are in a topological space.

B. Convergence of Sequences of Numbers

A sequence $\{a_n\}$ of numbers is said to be **convergent** to a number a or to **converge** to a, written $\lim_{n\to\infty} a_n = a$ or $a_n \to a$ as $n\to\infty$, if for any positive number ε we can choose a (sufficiently large) natural number n_0 such that for every n larger than n_0 the inequality $|a_n - a| < \varepsilon$ holds. Then a is called the **limit** (or **limit point**) of the sequence $\{a_n\}$. If $\{a_n\}$ has a limit, it is unique. A sequence which is not convergent is said to be **divergent** or to **diverge**.

A set A of real numbers is said to be **bounded from above** if there is a real number b such that $a \leqslant b$ for all $a\in A$, **bounded from below** if there is a real number c such that $a \geqslant c$ for all $a\in A$, and **bounded** if it is bounded from above and below. A sequence $\{a_n\}$ of real numbers is said to be **monotonically increasing (monotonically decreasing)**, written $a_n\uparrow$ ($a_n\downarrow$), if $a_1\leqslant a_2\leqslant\ldots\leqslant a_n\leqslant a_{n+1}\leqslant\ldots$ ($a_1\geqslant a_2\geqslant\ldots\geqslant a_n\geqslant a_{n+1}\geqslant\ldots$). A monotonically increasing or decreasing sequence is called a **monotone sequence**.

C. Criteria for the Convergence of Sequences of Numbers

Every bounded monotone sequence of real numbers is convergent; its limit is $\sup\{a_n\}$ ($\inf\{a_n\}$) (→ 355 Real Numbers B) if it is monotonically increasing (decreasing). For any bounded sequence $\{a_n\}$ of real numbers, setting $\alpha_n = \inf\{a_n, a_{n+1}, \ldots\}$ and $\beta_n = \sup\{a_n, a_{n+1}, \ldots\}$, we have $\alpha_n\uparrow$, $\beta_n\downarrow$, and $\alpha_n\leqslant a_n\leqslant\beta_n$. Hence $\lim_{n\to\infty}\alpha_n=\alpha$ ($=\sup\{\alpha_n\}$) and $\lim_{n\to\infty}\beta_n=\beta$ ($=\inf\{\beta_n\}$) exist. α is called the **inferior limit** (or **limit inferior**) of $\{a_n\}$, written $\liminf_{n\to\infty} a_n$ or $\underline{\lim}_{n\to\infty} a_n$, while β is called the **superior limit** (or **limit superior**), written $\limsup_{n\to\infty} a_n$ or $\overline{\lim}_{n\to\infty} a_n$. If $\limsup_{n\to\infty} a_n = \liminf_{n\to\infty} a_n = \alpha$, then $\lim_{n\to\infty} a_n$ exists and equals α. The limit of a convergent subsequence of a sequence $\{a_n\}$ of numbers is called an **accumulation point** of the sequence. Here we should distinguish an accumulation point of a sequence $\{a_n\}$ from an accumulation point of $\{a_n\}$ viewed as a set (→ 425 Topological Spaces O); for example, if $a_n=1$ for every n, then the former is 1 and the latter does not exist. For any bounded sequence of real numbers, its superior (inferior) limit is the maximum (minimum) of its accumulation points. Moreover, if β is the superior limit of a sequence $\{a_n\}$, then for any positive number ε, there exist only a finite number of n's for which a_n is greater than $\beta+\varepsilon$, while there may exist an infinite number of n's for which a_n is less than $\beta-\varepsilon$. The inferior limit of the sequence has a similar property.

Suppose that we are given a sequence $\{a_n\}$ of real numbers and that there exist two sequences $\{u_n\}$ and $\{v_n\}$ such that $u_n\leqslant a_n\leqslant v_n$, $\lim(u_n - v_n)=0$, $\{u_n\}$ is monotonically increasing, and $\{v_n\}$ is monotonically decreasing. Then $\lim a_n$ exists and is equal to $\lim u_n =$

$\lim v_n$ (**principle of nested intervals**). In particular, if $\limsup a_n = \liminf a_n$, then $\lim a_n$ exists. The converse also holds.

If $\{a_n\}$ is convergent, then $|a_n - a_m| \to 0$ as $n, m \to \infty$, and vice versa; that is, $\{a_n\}$ is convergent if and only if for any positive number ε there exists a positive integer n_0 such that $|a_n - a_m| < \varepsilon$ for all $n, m \geqslant n_0$ (**Cauchy's criterion**).

D. Infinity

For a set A of real numbers, the expression $\sup A = +\infty$ means that A is not bounded from above; $\inf A = -\infty$ means that A is not bounded from below. For a sequence $\{a_n\}$ of real numbers, $\lim a_n = +\infty$ means that for any real number b there exists a positive integer n_0 such that $a_n > b$ for all $n \geqslant n_0$; the notation $\lim a_n = -\infty$ has a similar meaning. The symbols $+\infty$ and $-\infty$ are called **positive** (or **plus**) **infinity** and **negative** (or **minus**) **infinity**, respectively. We say that the limit of $\{a_n\}$ is $+\infty$ $(-\infty)$ if $\lim a_n = +\infty$ $(-\infty)$. In these cases, we customarily say that $\{a_n\}$ **diverges** (or is **divergent**) **to** $+\infty$ $(-\infty)$, or that a_n becomes positively (negatively) infinite as $n \to \infty$. We also define $\limsup a_n = +\infty$ $(\liminf a_n = -\infty)$ to mean $\sup\{a_n\} = +\infty$ $(\inf\{a_n\} = -\infty)$. A sequence $\{a_n\}$ is said to **oscillate** if $\limsup a_n > \liminf a_n$.

We now have the following propositions concerning sequences of numbers: If $\lim a_n = a$ and $\lim b_n = b$, then $\lim(\alpha a_n + \beta b_n) = \alpha a + \beta b$, $\lim(a_n b_n) = ab$, and $\lim(a_n/b_n) = a/b$ (provided that $b_n \neq 0, b \neq 0$). For sequences of real numbers, these formulas also hold when a or b is infinity. In those cases we set $\alpha \cdot (\pm\infty) = \pm\infty$ $(\alpha > 0)$, $\alpha \cdot (\pm\infty) = \mp\infty$ $(\alpha < 0)$, $\alpha \pm \infty = \pm\infty$, $\alpha/(\pm\infty) = 0$ for a real number α. The cases $0 \cdot (\pm\infty)$, $+\infty + (-\infty)$, $\pm\infty/(\pm\infty)$ are excluded.

E. Convergence of Sequences of Points in a Topological Space

A sequence $\{a_n\}$ of points in a topological space (→ 425 Topological Spaces) is said to **converge** to a point a if for any †neighborhood U of a there exists a positive integer n_0 such that $a_n \in U$ for all $n \geqslant n_0$. The point a is called a **limit** (or **limit point**) of $\{a_n\}$ and we write $\lim_{n\to\infty} a_n = a$ or $a_n \to a$ as $n \to \infty$. A sequence is said to **diverge** if it does not converge to any point.

In particular, the set $\mathbf{R}$ of all real numbers is a topological space in which the set of intervals $(a-\varepsilon, a+\varepsilon)$ for some $\varepsilon > 0$ is a †base for the neighborhood system of a point a, so that the notion of limit in $\mathbf{R}$ explained previously is a special case of the same notion in a topological space. By adding the symbols $+\infty$ and $-\infty$ to $\mathbf{R}$, we obtain the topological space $\bar{\mathbf{R}}$, in which any set containing $\{x \mid x > \alpha, x \in \bar{\mathbf{R}}\}$ $(\{x \mid x < \alpha, x \in \bar{\mathbf{R}}\})$ for some $\alpha \in \mathbf{R}$ is a neighborhood of $+\infty$ $(-\infty)$, where the ordering is defined as $-\infty < \alpha < +\infty$ $(\alpha \in \mathbf{R})$. Then $\lim a_n = +\infty$ $(-\infty)$ is interpreted as convergence in the topological space $\bar{\mathbf{R}}$. The elements of $\bar{\mathbf{R}}$ are called **extended real numbers**.

In the case where the topological space is a †metric space (→ 273 Metric Spaces) with metric ρ, $a_n \to a$ is equivalent to $\rho(a_n, a) \to 0$.

For convergence of sequences of points in a topological space, the following properties (S) hold: (S) (1). If $a_n = a$ for all n, then $\lim a_n = a$; (2) if $a_n \to a$, then $a_{n_k} \to a$ for any subsequence $\{a_{n_k}\}$; (3) if there is a point a such that any subsequence $\{a_{n_k}\}$ of $\{a_n\}$ has a suitable subsequence converging to a, then $a_n \to a$. In a †Hausdorff space (e.g., a metric space), the additional property (S*) holds: If a sequence $\{a_n\}$ has a limit, this limit is unique, and can be denoted $\lim a_n$.

F. Limits of Functions

Let a real-valued function $f(x)$ of a real variable x be defined for $x \neq a$ belonging to a neighborhood of the point a. We say that the **limit** of $f(x)$ is b as x tends to a, and write $\lim_{x\to a} f(x) = b$ or $f(x) \to b$ as $x \to a$, if for any positive number ε there exists a positive number δ such that $0 \neq |x-a| < \delta$ implies $|f(x) - b| < \varepsilon$. Replacing $0 \neq |x-a| < \delta$ by $a < x < a+\delta$ $(a-\delta < x < a)$, we define $f(x) \to b$ as $x \to a+0$ $(x \to a-0)$ and say that b is the **limit on the right** (**left**) of $f(x)$ as x tends to a. We define $f(x) \to +\infty$ or $f(x) \to -\infty$ as $x \to a$ analogously to the case of sequences. The expression $f(x) \to b$ as $x \to +\infty$ means that for any positive number ε there is a real number k such that $|f(x) - b| < \varepsilon$ for any $x > k$. There are similar definitions for $x \to -\infty$ and $b = \pm\infty$. When $f(x) \to \pm\infty$ as $x \to a$, we often say that f diverges definitely at a.

In general, for a mapping f from a subset D of a topological space X into a topological space Y, with a point a in the closure of D and a point b in Y, $\lim_{x\to a} f(x) = b$ or $f(x) \to b$ as $x \to a$ means that any neighborhood V of b contains $f(U \cap D - \{a\})$ for some neighborhood U of a. If Y is a Hausdorff space, b is unique (if it exists) for given f and a. This point b is called the **limit** (or **limit value**) of $f(x)$ as $x \to a$.

It is easy to see that this definition of $\lim_{x\to a} f(x) = b$ is a generalization of the cases where the topological spaces are $\bar{\mathbf{R}}$ or $\mathbf{R}$. Let $\mathbf{N}$ be the set of all natural numbers, and let $\bar{\mathbf{N}} =$

$\mathbf{N} \cup \{+\infty\}$ be supplied with the †relative topology as a subspace of $\bar{\mathbf{R}}$. A sequence $\{a_n\}$ of real numbers or points can be identified with a mapping f from $\mathbf{N}$ into $\mathbf{R}$ ($\mathbf{N} \to \mathbf{R}$) defined by $f(n) = a_n$. It converges to a if and only if $\lim_{n\to\infty} f(n) = a$.

Suppose, in particular, that f is a mapping from a metric space (X, ρ) into a metric space (Y, σ). Then, $f(x) \to b$ as $x \to a$ means that for any $\varepsilon > 0$ there exists a $\delta > 0$ such that $\sigma(f(x), b) < \varepsilon$ for all $x \in D$ such that $0 < \rho(x, a) < \delta$. Thus $f(x) \to b$ as $x \to a$ if and only if $f(x_n) \to b$ for any $\{x_n\}$ in D with $x_n \to a$. If we set $\rho(z_1, z_2) = |z_1 - z_2|$ for complex numbers z_1, z_2, the function ρ supplies the set of all complex numbers $\mathbf{C}$ with a metric and $\mathbf{C}$ becomes isometric to the plane $\mathbf{R}^2$ (→ 74 Complex Numbers). Thus the cases $X = \mathbf{C}$ or $Y = \mathbf{C}$ are particular cases of the above generalization. Furthermore, we introduce the †Riemann sphere $\bar{\mathbf{C}} = \mathbf{C} \cup \{\infty\}$ by adding the †point at infinity ∞ to $\mathbf{C}$. We can define a topology on $\bar{\mathbf{C}}$ such that any set containing $\{\infty\} \cup \{z \mid |z| > r\}$ for some positive number r is a neighborhood of ∞. Thus we can define the notions $f(z) \to b$ as $z \to \infty$, $f(x) \to \infty$ as $x \to a$, etc., for a complex-valued function f by considering f as a mapping from the topological space $\bar{\mathbf{C}}$ into itself. Then $f(x) \to \infty$ is equivalent to $1/f(x) \to 0$.

G. Orders of Infinities and Infinitesimals

Let f be a complex-valued function defined on a topological space X and a a point of X. Then f is called an **infinity** (at a) or an **infinitesimal** (at a) if $f(x) \to \infty$ as $x \to a$ or $f(x) \to 0$ as $x \to a$, respectively. Suppose that f and g are infinities and f/g is an infinitesimal. Then f is said to be **of lower order** than g, and g is said to be **of higher order** than f. If both f/g and g/f are bounded, then f is said to be **of the same order** as g. This last relation, written $f \sim g$, is an equivalence relation. An infinity f is said to be **of the nth order** with g if $f \sim g^n$. For two infinitesimals f and g, f is called **of higher order** than g and g **of lower order** than f if f/g is an infinitesimal. For infinitesimals, the terms **of the same order** and **of the nth order** are defined similarly as above. In particular, when $X = \mathbf{C}$ and $a = \infty$, we usually omit the phrase "at ∞." Also, for such a function f, we customarily say that the **order of an infinity (infinitesimal)** is n if $f \sim z^n$ (z^{-n}).

To describe the order of an infinity or an infinitesimal simply, the following notions, due to E. Landau [10], are in common use. Let f and g be two functions. If $|f(x)/g(x)|$ is bounded as $x \to a$, then f is called **at most of the order of** g as $x \to a$, and we write $f(x) = O(g(x))$ as $x \to a$. Second, if $f(x)/g(x)$ is an infinitesimal at a, then f is said to be of lower order than g as $x \to a$, and we write $f(x) = o(g(x))$ as $x \to a$. The symbols O, o, indicating the word "order," are called **Landau's symbols**. The notation $f(x) = h(x) + O(g(x))$ means $f(x) - h(x) = O(g(x))$. When we use the symbols O, o we should indicate clearly the phrase "as $x \to a$," which is sometimes omitted when no confusion is to be feared (e.g., for the case of a complex variable with $a = \infty$). These symbols are employed for sequences as well, to describe their behavior as $n \to \infty$.

H. Convergence of Nets (Moore-Smith Convergence) [7, 8]

Let $\mathfrak{A}$ be a (preordered) †directed set. A family of points in a set X with index set $\mathfrak{A}$ (namely, a mapping from $\mathfrak{A}$ to X) is called a **net** in X. A net is denoted by $\{x_\alpha\}_{\alpha\in\mathfrak{A}}$ ($\{x_\alpha\}_{\mathfrak{A}}$ or $\{x_\alpha\}$). A net $\{y_\beta\}_{\mathfrak{B}}$ in X is called a **subnet** of $\{x_\alpha\}_{\mathfrak{A}}$ if there exists a mapping $\varphi: \mathfrak{B} \to \mathfrak{A}$ such that (1) $y_\beta = x_{\varphi(\beta)}$ and (2) for any $\alpha_0 \in \mathfrak{A}$ there exists a $\beta_0 \in \mathfrak{B}$ such that $\beta \geqslant \beta_0$ implies $\varphi(\beta) \geqslant \alpha_0$. In particular, if $\mathfrak{B}$ is a cofinal directed subset of $\mathfrak{A}$, $\{x_\beta\}_{\mathfrak{B}}$ is called a cofinal subnet of $\{x_\alpha\}_{\mathfrak{A}}$. A net is called a **universal net** if either $\{\alpha \mid x_\alpha \in Y\}$ or $\{\alpha \mid x_\alpha \in X - Y\}$ is residual in $\mathfrak{A}$ for any subset Y of X. For any net there is a universal subnet.

For a net $\{x_\alpha\}_{\mathfrak{A}}$ in a topological space X, $\{x_\alpha\}$ is said to **converge** to a point x in X if for any neighborhood U of x there is an α_0 such that $\{x_\alpha \mid \alpha \geqslant \alpha_0\} \subset U$. Then a is called a **limit** of the net $\{x_\alpha\}$. We then write $x_\alpha \to x$ ($\alpha \in \mathfrak{A}$) (or simply $x_\alpha \to x$). The convergence of sequences of points is the special case where $\mathfrak{A} = \mathbf{N}$. The notion of convergence using nets was introduced by E. H. Moore and H. E. Smith.

Concerning this convergence, we have the following propositions (**D**): (**D**) (1) If $x_\alpha = x$ for all α, then $x_\alpha \to x$. (2) If $x_\alpha \to x$ and $\{y_\beta\}$ is a subnet of $\{x\}$, then $y_\beta \to x$. (3) If for a net $\{x_\alpha\}$ there is a point x such that any subnet $\{y_\beta\}$ of $\{x_\alpha\}$ has a suitable subnet converging to x, then $x_\alpha \to x$. (4) Suppose that $x_\alpha \to x$ ($\alpha \in \mathfrak{A}$) and $y_{\alpha\beta} \to x_\alpha$ ($\beta \in \mathfrak{B}_\alpha$) for each α. Then let $\mathfrak{C} = \mathfrak{A} \times \prod \mathfrak{B}_\alpha$ be the direct product of directed sets with projections $p: \mathfrak{C} \to \mathfrak{A}$ and $p_\alpha: \mathfrak{C} \to \mathfrak{B}_\alpha$, and for $\gamma \in \mathfrak{C}$ define $z_\gamma = y_{\alpha\beta}$, where $\alpha = p(\gamma)$ and $\beta = p_\alpha(\gamma)$. Then $z_\gamma \to x$ ($\gamma \in \mathfrak{C}$). Furthermore, the space X is a †Hausdorff space if and only if we have the condition (**D***): Any net in X has at most one limit.

A limit of $\{x_\alpha\}$ is denoted by $\lim x_\alpha$ or $\lim_{\alpha\in\mathfrak{A}} x_\alpha$. Then $x_\alpha \to x$ if and only if x is contained in the closure of any subnet $\{y_\beta \mid \beta \in \mathfrak{B}\}$ of $\{x_\alpha\}$. (We may consider this to be a definition of $x_\alpha \to x$.)

I. Convergence of Filters [9]

Let X be a set. A set Φ of subsets of X is called a **filter** if the following conditions are satisfied: (i) $\varnothing \notin \Phi$ ($\varnothing$ is the empty set); (ii) $A \subset B \subset X$ and $A \in \Phi$ imply $B \in \Phi$; (iii) $A, B \in \Phi$ imply $A \cap B \in \Phi$. Let $\mathfrak{B}$ be a set of subsets of X and Φ be the collection of subsets of X such that each element A of Φ contains a subset belonging to $\mathfrak{B}$. If Φ is a filter, then $\mathfrak{B}$ is called a **filter base** which **generates** Φ. $\mathfrak{B}$ is a filter base if and only if (i) $\varnothing \notin \mathfrak{B}$; (ii) $A, B \in \mathfrak{B}$ implies that there is a $C \in \mathfrak{B}$ with $A \cap B \supset C$. A filter Φ is called an **ultrafilter** (or **maximal filter**) if there exists no filter which contains Φ properly. For any filter there exists an ultrafilter containing it. If $\{\Phi_\lambda\}_\Lambda$ is a family of filters, then the intersection $\bigcap \Phi_\lambda$ is a filter. If $\mathfrak{F}$ is a filter base in the index set Λ, then $\bigcup_{M \in \mathfrak{F}} (\bigcap_{\lambda \in M} \Phi_\lambda)$ is also a filter.

We denote by $\mathfrak{U}(x)$ the †neighborhood system of a point x in a topological space X. A filter Φ in X is said to **converge** to a point a, written $\Phi \to a$, if $\mathfrak{U}(x) \subset \Phi$. A filter base $\mathfrak{B}$ is said to **converge** to a if the filter generated by $\mathfrak{B}$ converges to a.

The convergence of filters just defined has the following fundamental properties (**L**): (1) for a point a in X the filter $\Phi_a = \{A \mid a \in A \subset X\}$ converges to a; (2) for two filters Φ and Ψ, $\Phi \to a$ and $\Phi \subset \Psi$ imply $\Psi \to a$; (3) if $\Phi_\lambda \to a$ for all members in a family $\{\Phi_\lambda\}$ of filters, then $\bigcap \Phi_\lambda = \Phi \to a$; (4) suppose that filters $\Phi_y \to y$ are assigned for all points y in a subset Y of X, and that we are given a filter Ψ in X which converges to a, generated by a filter base $\mathfrak{B}$ in Y; then $\bigcup_{B \in \mathfrak{B}} (\bigcap_{y \in B} \Phi_y) \to a$. Furthermore, the space X is Hausdorff if and only if we have the condition (**L***): Each filter in X has at most one limit.

J. Relations among Various Definitions of Convergence

Convergence of sequences of points is a special case of that of nets. Properties (1), (2), and (3) of (**D**) imply (1), (2), and (3) of (**S**), respectively, and (**D***) implies (**S***). Consider a net $\{x_\alpha\}_{\mathfrak{A}}$ in X. Then the set $\{\{x_\alpha \mid \alpha \in \mathfrak{A}, \alpha \geqslant \alpha_0\} \mid \alpha_0 \in \mathfrak{A}\}$ of subsets of X is a filter base in X which generates a filter Φ, and $\Phi \to x$ if and only if $x_\alpha \to x$. In this situation, (**L**) implies (**D**), and (**L***) implies (**D***). Suppose that we are given a function $f: X \to Y$ with the domain D and a point $a \in X$. Let $\mathfrak{U}(a)$ be the neighborhood system of a and assume that, for any $U \in \mathfrak{U}(a)$, $U \cap D - \{a\} \neq \varnothing$. Then the set $\{f(U \cap D - \{a\}) \mid U \in \mathfrak{U}(a)\}$ is a filter base. Let Φ be the filter generated by it. Then $f(x) \to b$ as $x \to a$ if and only if $\Phi \to b$. Consequently, the various types of convergence described previously can be expressed by means of convergence of filters.

K. Convergence and Topology

In a topological space X, the concept of convergence of nets and that of filters can be defined. Conversely, convergence of nets in X defines a topology of X. In fact, let us assume that we are given a set X and a definition of convergence of filters which satisfies the properties (**L**). Then convergence of nets that satisfies (**D**) can be introduced as above. If $\bar{A}$ is defined as the set of limits of all nets contained in a subset A of X, then $\bar{A}$ satisfies the axiom of closures (→ 425 Topological Spaces), and a topology can be defined on X. Then we have the following propositions: (i) $a \in \bar{A}$ if and only if there is a net $\{x_\alpha\}$ with $x_\alpha \in A$ converging to a; (ii) U is a neighborhood of a if and only if $x_\alpha \to a$ implies that there exists an α_0 such that $x_\alpha \in U$ for all $\alpha \geqslant \alpha_0$. Thus, if X is a topological space, it carries a "new" topology defined by way of convergence of nets. But this "new" topology coincides with the original one. Similarly, starting from convergence of filters (or nets), we can obtain a "new" definition of convergence of filters (or nets), which coincides with the initial one. In conclusion, defining a topology on a space X is the same thing as defining convergence of filters in X or of nets in X.

We shall describe here a few notions on topological spaces in terms of convergence. The fact that a topological space X is compact is equivalent to the fact that every universal net in X converges, and to the fact that every ultrafilter in X converges. Also equivalent is the fact that every net has a convergent subnet. A mapping f from a topological space X into a topological space Y is continuous at a if and only if one of the following conditions is satisfied: (1) for any net $\{x_\alpha\}$ in X converging to $a \in X$, we have $f(x_\alpha) \to f(a)$ in Y; (2) for any filter Φ in X converging to $a \in X$, we have $f(\Phi) = \{f(M) \mid M \in \Phi\} \to f(a)$ in Y; (3) $f(x) \to f(a)$ in Y as $x \to a$ in X (in the sense of the limit of a function at a).

M. Fréchet [6] gave a definition of a topology on a space using the notion of convergence as a foundation. A set is called an **L-space** (or **Fréchet L-space**) if convergence of sequences of points in it is defined so as to satisfy conditions (1) and (2) of (**S**) and (**S***) (1906). Such convergence is called **star convergence** if it also satisfies (3) of (**S**), and in that case the space is called an **L^*-space**. For any subset A of an L-space X, define $\bar{A}$ as the set of all points a such that $x_n \to a$ for some sequence $\{x_n\}$ contained in A. Then the axioms

$\bar{A} \supset A$, $\overline{A \cup B} = \bar{A} \cup \bar{B}$, and $\bar{\varnothing} = \varnothing$ are satisfied, so that X is a †generalized topological space (the axiom $\bar{\bar{A}} = \bar{A}$ is not necessarily satisfied). For a Hausdorff space X with the †first countability axiom and convergence of sequences defined by means of its topology, the closure operation defined above gives the same topology as the initial one.

L. (o)-Convergence

A sequence $\{a_n\}$ of elements of an ordered set S is said to be (o)-**convergent** to an element a of S if there exist two sequences $\{u_n\}$ and $\{v_n\}$ such that $u_n \leqslant a_n \leqslant v_n$, $u_n \leqslant u_{n+1}$, and $v_n \geqslant v_{n+1}$, and $a = \sup u_n = \inf v_n$. When we write this $a_n \to a$, properties (1) and (2) of (S) and (S*) concerning the convergence of sequences hold. Next, a sequence $\{a_n\}$ is said to be (o)-**star convergent** to a if any subsequence of $\{a_n\}$ has a suitable subsequence which converges to a. Then (o)-star convergence satisfies the properties (S) and (S*).

For any set X the set $\mathfrak{P}(X)$ of all subsets of X is an ordered set under the inclusion relation. The fact that a sequence $\{A_n\}$ of subsets is (o)-convergent to a subset A is equivalent to:

$$A = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n = \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} A_n.$$

The set A is also equal to $\lim A_n$, which is the †limit of a sequence $\{A_n\}$ of subsets.

References

[1] W. Rudin, Principles of mathematical analysis, second edition, McGraw-Hill, 1964.
[2] J. L. Kelley, General topology, Van Nostrand, 1955.
[3] G. Birkhoff, Lattice theory, Amer. Math. Soc. Colloq. Publ., revised edition, 1948.
[4] N. Bourbaki, Eléments de mathématique III, Topologie générale, ch. 1, Actualités Sci. Ind., 1142d, Hermann, fourth edition, 1965; English translation, General topology, Addison-Wesley, 1966.
[5] J. W. Tukey, Convergence and uniformity in topology, Ann. Math. Studies, Princeton Univ. Press, 1940.
[6] M. Fréchet, Sur quelques points du calcul fonctionnel, Rend. Circ. Mat. Palermo, 22 (1906), 1–74.
[7] E. H. Moore and H. L. Smith, A general theory of limits, Amer. J. Math., 44 (1922), 102–121.
[8] G. Birkhoff, Moore-Smith convergence in general topology, Ann. Math., (2) 38 (1937), 39–56.
[9] H. Cartan, Théorie des filtres, C. R. Acad. Sci. Paris, 205 (1937), 595–598; Filtres et ultrafiltres, ibid., 777–779.
For Landau's symbol,
[10] E. G. H. Landau, Handbuch der Lehre von der Verteilung der Primzahlen, Teubner, 1909 (Chelsea, 1953).

88 (X.4) Convex Analysis

A. Convex Functions

A real-valued function $f(x)$ defined on a †convex set D in a linear space over $\mathbf{R}$ is called a **convex function** if for every $x, y \in D$ and $0 \leqslant \lambda \leqslant 1$ we have

$$f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda) f(y). \qquad (1)$$

The function $f(x)$ is called a **strictly convex function** if the sign $\leqslant$ in (1) is replaced by $<$ for $x \neq y$ except when λ is either 0 or 1. If $-\psi(x)$ is convex (strictly convex), the function $\psi(x)$ is called a **concave function (strictly concave function)**. The notion of convex function was introduced by J. L. W. V. Jensen [1] for the case where the domain D is an interval on the real line $\mathbf{R}$.

Sometimes the condition for a convex function is weakened in such a way that (1) is assumed only for $\lambda = 1/2$. However, if D is a †topological linear space and f is continuous, then the weakened condition implies the original one. Hereafter, we mainly consider the case where D is an interval on the real line. In this case, a convex function $f(x)$ (in the weaker sense) is continuous in the interior of the interval if $f(x)$ is †measurable or bounded from above on a set of positive measure (the latter was proved by A. Ostrowski [2]). In particular, suppose that $f(x)$ is defined in the interval I and is bounded from below. Then either $f(x)$ is continuous or its graph is dense in the set $\{(x, y) \mid x \in I, y \geqslant g(x)\}$, where $g(x)$ is a suitable convex continuous function (Hukuhara [3]). We note here that the original definition of a convex function $f(x)$ implies the continuity of $f(x)$ in the interior of the interval. In such a case, $f(x)$ always has †right and †left derivatives and satisfies $f'_-(x) \leqslant f'_+(x) \leqslant f'_-(y) \leqslant f'_+(y)$ for $x < y$. Hence it is differentiable except for at most countably many points.

A function $f(x)$ is a continuous convex function in $a \leqslant x \leqslant b$ if and only if it is expressible in the form

$$f(x) = f(a) + \int_a^x \varphi(t)\, dt,$$

where $\varphi(t)$ is a monotone increasing function. If $f(x)$ is twice differentiable, then $f''(x) \geqslant 0$ $(a < x < b)$ is a necessary and sufficient condition for $f(x)$ to be convex in (a, b).

B. Convex Functions and Inequalities

If $f(x)$ is convex (in the original sense), we have, for $a_\nu > 0$,

$$f\left(\sum a_\nu x_\nu \Big/ \sum a_\nu\right) \leqslant \sum a_\nu f(x_\nu) \Big/ \sum a_\nu. \qquad (2)$$

Similarly, we have, for $\varphi > 0$,

$$f\left(\int \varphi\psi\, dx \Big/ \int \varphi\, dx\right) \leqslant \int \varphi f(\psi)\, dx \Big/ \int \varphi\, dx. \qquad (3)$$

The functions x^a $(a > 1$ or $a < 0)$, $-x^a$ $(0 < a < 1)$, $-\log x$, $x \log x$ are strictly convex for $x > 0$, and the functions x^{2n} $(n \geqslant 1)$, $\exp x$, $\log(1 + e^x)$, $\sqrt{a^2 + x^2}$ $(a \neq 0)$ are strictly convex in $-\infty < x < +\infty$. Applying the inequalities (2) or (3) to these functions, we obtain various inequalities, including the inequalities on means (→ 211 Inequalities).

A continuous convex function $f(x)$ over a topological linear space satisfying the relation $f(\alpha x) = \alpha f(x)$ for an arbitrary positive number α is called a **subadditive functional** and is often utilized in functional analysis.

C. M. Riesz's Convexity Theorem

Let $x = (\xi_1, \ldots, \xi_n)$ be an n-tuple of complex numbers, and let $\nu \geqslant 0$. We put $N_\nu(x) = (\sum_{j=1}^n |\xi_j|^{1/\nu})^\nu$ for $\nu > 0$ and $N_0(x) = \sup |\xi_j|$. Let (α_{ij}) be an $m \times n$ complex matrix, $x = (\xi_1, \ldots, \xi_n)$, $z = (\zeta_1, \ldots, \zeta_m)$, $\nu \geqslant 0$, and $\mu \geqslant 0$. We put

$$M(\nu, \mu) = \sup_{N_\nu(x) \leqslant 1,\, N_\mu(z) \leqslant 1} \left| \sum_{i=1}^m \sum_{j=1}^n \alpha_{ij} \zeta_i \xi_j \right|.$$

Then $\log M(\nu, \mu)$ is a convex function of (ν, μ) in the following sense: Let $0 < \nu_i \leqslant 1$, $0 < \mu_i \leqslant 1$, and $\nu_i + \mu_i \geqslant 1$ $(i = 1, 2)$. Then $\log M((1-t)\nu_1 + t\nu_2, (1-t)\mu_1 + t\mu_2)$ is a convex function with respect to t for $0 \leqslant t \leqslant 1$ [4, 5]. These results are called **M. Riesz's convexity theorem**. Famous inequalities such as the †Hölder inequality or the †Minkowski inequality follow from this theorem. For example, let T be an †additive operator from the †function space $L_p(\Omega)$ into $L_p(\Omega)$ for all $1 \leqslant p \leqslant \infty$. If T is a continuous operator for $p = 1$ and $p = \infty$, and the norm of T is $\leqslant C$ for $p = 1$ and $p = \infty$, then T is continuous for all p $(1 < p < \infty)$, and its norm is always $\leqslant C$.

D. Subdifferentials of Convex Functions

Currently, **convex analysis** is playing an important role in the study of nonlinear evolution equations. This analysis treats convex functions on infinite-dimensional spaces [9].

In convex analysis, it is sometimes convenient to consider a proper convex function φ on a Hilbert space X. A mapping φ from X into $(-\infty, +\infty]$ is called a **proper convex function** if φ is not identically equal to $+\infty$ and if

$$\varphi(\lambda x + (1-\lambda) y) \leqslant \lambda \varphi(x) + (1-\lambda)\varphi(y)$$

for every $x, y \in X$ and $0 \leqslant \lambda \leqslant 1$. The convex set

$$D(\varphi) = \{x \in X \mid \varphi(x) < \infty\} \subset X$$

is called the **effective domain** of φ. The **subdifferential** $\partial\varphi$ of φ is defined to be a multivalued function which assigns to each x in $D(\varphi)$ all elements $y \in X$ satisfying $\varphi(\xi) \geqslant \varphi(x) + \langle y, \xi - x\rangle$ for any $\xi \in X$. Then $\partial\varphi$ is monotone in X in the following sense: If $y_1 \in \partial\varphi(x_1)$ and $y_2 \in \partial\varphi(x_2)$, then

$$\langle y_1 - y_2, x_1 - x_2 \rangle \geqslant 0.$$

Under some conditions on φ, $\partial\varphi$ becomes a single-valued function, as is seen in the following example: Let A be a nonnegative self-adjoint operator in a Hilbert space X. For the †fractional power $\sqrt{A}$ of A, define φ by

$$\varphi(x) = \begin{cases} \frac{1}{2}\|\sqrt{A}x\|^2, & x \in D(\sqrt{A}), \\ +\infty, & \text{otherwise}. \end{cases}$$

Then $D(\varphi) = D(\sqrt{A})$ and $\partial\varphi = A$.

E. Convex Functions and Nonlinear Semigroups

Let φ be a lower semicontinuous proper convex function on a Hilbert space X. Since $A = -\partial\varphi$ is a maximal †dissipative operator in X, it generates a nonlinear semigroup $\{T_t \mid t \geqslant 0\}$ on the closed convex set $\overline{D(A)}$:

$$T_t x = \lim_{\tau \downarrow 0} (I - \tau A)^{-[t/\tau]} x$$

(→ 286 Nonlinear Functional Analysis, 378 Semigroups of Operators and Evolution Equations). It can be shown that for any $a \in \overline{D(A)}$, $T_t a \in D(A)$ whenever $t > 0$. Thus the †abstract Cauchy problem

$$\frac{du}{dt} + \partial\varphi(u) \ni 0 \quad (t > 0),$$

$$u(+0) = a,$$

is considered to be "parabolic." Furthermore,

we have

$$\left\| \frac{d}{dt} T_t a \right\| \leqslant \frac{1}{t} \|a - T_t a\| \quad (t > 0)$$

for all $a \in \overline{D(A)}$ (H. Brézis [10]).

References

[1] J. L. W. V. Jensen, Sur les fonctions convexes et les inégalités entre les valeurs moyennes, Acta Math., 30 (1906), 175–193.
[2] A. Ostrowski, Mathematische Miszellen XIV, Über die Funktionalgleichung der Exponentialfunktion und verwandte Funktionalgleichungen, Jber. Deutsch. Math. Verein., 38 (1929), 54–62.
[3] M. Hukuhata, Sur la fonction convexe, Proc. Japan Acad., 30 (1954), 683–685.
[4] M. Riesz, Sur les maxima des formes bilinéaires et sur les fonctionelles linéaires, Acta Math., 49 (1927), 465–497.
[5] G. O. Thorin, An extention of a convexity theorem due to M. Riesz, Medd. Lunds Univ. Mat. Sem., 4 (1939), 1–5.
[6] G. H. Hardy, J. E. Littlewood, and G. Pólya, Inequalities, ch. 3, Cambridge Univ. Press, 1934, revised edition, 1952.
[7] N. Bourbaki, Eléments de mathématique, Espaces vectoriels topologiques, ch. 2, Actualité Sci. Ind., 1189a, Hermann, second edition, 1966.
[8] F. A. Valentine, Convex sets, McGraw-Hill, 1964.
[9] V. Barbu, Nonlinear semigroups and differential equations in Banach spaces, Noordhoff, 1976.
[10] H. Brézis, Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert, Math. Studies 5, North-Holland, 1973.

89 (VI.11) Convex Sets

A. General Remarks

A nonempty subset X of the n-dimensional Euclidean space $\mathbf{R}^n$ is called a **convex set** if for any elements x, y in X and any number a such that $0 \leqslant a \leqslant 1$, the element $ax + (1-a)y$ of $\mathbf{R}^n$ is also contained in X. The †interior and the †closure of a convex set are also convex. A point x of a convex set X is called an **extreme point** of X if x cannot be expressed as $(x_1 + x_2)/2$ in terms of a pair of distinct points x_1, x_2 in X. A bounded closed convex set is called a **convex body** if it has †interior points.

Given an arbitrary nonempty subset X of $\mathbf{R}^n$, the minimum convex set containing X exists, called the **convex hull** of X and denoted by $[X]$. Each point x of $[X]$ can be expressed as $x = \sum_{i=1}^{n+1} a_i x_i$, where x_i belongs to X and the a_i are nonnegative numbers such that $\sum_{i=1}^{n+1} a_i = 1$. When X is a finite set, $[X]$ is called a **convex polyhedron**. If $\dot{X}$ denotes the set of extreme points (also called **vertices**) of a convex polyhedron X, then $X = [\dot{X}]$.

For elements x, y of $\mathbf{R}^n$, denote the inner product by (x, y). Given a nonzero element v of $\mathbf{R}^n$ and a fixed number a, the †hyperplane $H = \{x \mid (v, x) = a\}$ divides the space $\mathbf{R}^n$ into two †half-spaces $\{x \mid (v, x) \leqslant a\}$ and $\{x \mid (v, x) \geqslant a\}$, each of which is a closed convex set. If a convex set X is contained in one of the half-spaces S determined by the hyperplane H and the boundaries of X and H intersect, then we say that H is a **supporting hyperplane** of X and S is a **supporting half-space** of X. A closed convex set X is the intersection of its supporting half-spaces. A boundary point of a convex set X is contained in a supporting hyperplane of X. Given mutually disjoint convex sets X and Y, the **separation theorems** (1) and (2) hold.

(1) If X has inner points, then there exist a nonzero element v of $\mathbf{R}^n$ and a number a such that X is contained in the set $\{x \mid (x, v) \geqslant a\}$ and Y is contained in the set $\{x \mid (x, v) \leqslant a\}$.

(2) If X and Y are closed and X is bounded, then we can replace the signs $\leqslant$ and $\geqslant$ in (1) by $<$ and $>$, respectively (when the separation of convex sets X and Y is described by strict inequalities, we say that X and Y are **strongly separated**).

As an immediate consequence of the separation theorems, we obtain the following proposition: Suppose that A is an $m \times n$ matrix with real entries. For an element z in a Euclidean space, we write $z \geqslant 0$ (>0) if each component of z is $\geqslant 0$ (>0). Now if ${}^tAy > 0$ never holds for an m-dimensional vector $y > 0$, then there exists a nonzero n-dimensional vector $x \geqslant 0$ such that $Ax \leqslant 0$ (→ 173 Game Theory).

The definitions given previously for subsets of $\mathbf{R}^n$ can be naturally extended to the case of †real topological linear spaces (→ Section G). Also, in the theory of analytic functions of several complex variables, various notions of convexity of the subsets of $\mathbf{C}^n$ are considered (→ 21 Analytic Functions of Several Complex Variables).

B. Helly's Theorem

Suppose that we are given an index set Λ of cardinality greater than $n+1$, and bounded closed convex sets C_λ ($\lambda \in \Lambda$) in $\mathbf{R}^n$. If any $n+1$

sets of the C_λ have nonempty intersection, then the intersection of all the C_λ is nonempty (**Helly's theorem**).

This theorem has a wide range of applications. For example, we have propositions (1)–(4).

(1) If a convex set X of $\mathbf{R}^n$ is covered by a finite number of half-spaces, then X can be covered by no more than $n+1$ half-spaces among them. (2) Let X and Y be finite subsets of $\mathbf{R}^n$. X and Y are strongly separated by a hyperplane if for an arbitrarily chosen subset S of $X \cup Y$ consisting of at most $n+2$ points, the sets $S \cap X$ and $S \cap Y$ are strongly separated by a hyperplane. (3) If the †diameter of a subset X of $\mathbf{R}^n$ is not greater than 2, then X is contained in a †ball of radius $(2n/(n+1))^{1/2}$. (4) Let X be a convex body in $\mathbf{R}^n$. There exists a point x in X such that $\|x-u\|/\|v-u\| \leqslant n/(n+1)$, where u, v are points of intersection of an arbitrary straight line passing through x with the boundary of X, and $\|x\|$ denotes the length $(x,x)^{1/2}$ of x.

Helly's theorem can also be applied to problems of approximation of functions.

C. Ovals

The boundaries of convex bodies in $\mathbf{R}^2$ and $\mathbf{R}^3$ are called **ovals** and **ovaloids**, respectively. An oval E is a †Jordan curve $C(t)$ which admits at every point $P_0 = C(t_0)$ left and right "tangents" $l^+_{P_0}$ and $l^-_{P_0}$, where $l^\pm_{P_0}$ is a straight line expressed as the set of points $P(\lambda)$, $\lambda \in \mathbf{R}$ such that $P(\lambda) - P_0 = \lambda a^\pm$, with $a^\pm = \lim_{t \to \pm 0}(C(t)-P_0)/|C(t) - P_0|$). There may exist exceptional points P on E for which left and right tangents do not coincide, but the set of such points is at most countable. Each tangent $l^\pm_{P_0}$ shares a point or a segment with E. A line satisfying this condition is called a **supporting line** of the oval. It is also a supporting hyperplane in $\mathbf{R}^2$ of the convex body $[E]$ in the sense of Section A. If we fix an interior point O of a convex body X and take an arbitrary point P different from O, then the boundary E of X admits one and only one supporting line $l(P)$ which is perpendicular to the line OP and meets the half-line OP. Take a rectangular coordinate system with the origin O, and let (x,y) denote the coordinates of P. Then the points (ξ,η) on $l(P)$ satisfy the equation $\xi x + \eta y = H(x,y)$, where $H(x,y)$ is a function determined for all (x,y) in $\mathbf{R}^2$ and satisfying the following conditions: (i) $H(0,0)=0$; (ii) $H(tx,ty)=tH(x,y)$, for $t \geqslant 0$; (iii) $H(x_1+x_2, y_1+y_2) \leqslant H(x_1,y_1)+H(x_2,y_2)$. The function $H(x,y)$ is called the **supporting line function** of E. The magnitude and shape of E are determined by H, and any function satisfying conditions (i)–(iii) is a supporting line function of an oval. An oval E has a finite length $L=L(E)$, and the convex body $[E]$ has a finite area $F=F(E)$. If OP' denotes the half-line with direction opposite to that of OP and l' denotes the supporting line $l(P')$, the distance between the parallel lines l and l' is called the **breadth** of E in the direction PP'. Let $D=D(E)$ and $\Delta=\Delta(E)$ be the maximum and minimum of the breadth of E, respectively. D is the †diameter of E (or $[E]$), and Δ is called the **thickness** of E (or $[E]$). In particular, if $D=\Delta$, then the oval E is called a **curve of constant breadth**. In the following inequalities, equality holds only when E is one of the figures mentioned in parentheses: (1) $L^2 \geqslant 4\pi F$ (circles, J. Steiner (1838)) (→ 228 Isoperimetric Problems); (2) $\pi D^2 \geqslant 4F$ (circles, L. Bieberbach (1915)); (3) $L \leqslant \pi D$ (curves of constant breadth, W. Blaschke (1916)); (4) $F \geqslant \Delta^2/\sqrt{3}$ (regular triangles, J. Pál (1921)). See T. Kubota, *Tôhoku Sci. Bull.*, I, 12, 13; *Tôhoku Math. J.*, 24, 49.

D. Linear Combinations of Ovals

Let H_1 and H_2 be supporting line functions of ovals E_1 and E_2, and let t_1 and t_2 be positive numbers. Since the function $t_1 H_1 + t_2 H_2$ satisfies conditions (i)–(iii), given before, it is a supporting line function of an oval $E(t_1,t_2)$. In this case, we can also write $E(t_1,t_2) = t_1 E_1 + t_2 E_2$ and call it a **linear combination** of E_1 and E_2. In particular, the oval $(E_1+E_2)/2$ is called the **mean oval** of E_1 and E_2. In general, there exists a quantity M, called the **mixed area** of E_1 and E_2, such that $F(E(t_1,t_2)) = F(E_1)t_1^2 + 2Mt_1t_2 + F(E_2)t_2^2$. M does not depend on the choice of t_1 and t_2, and $M^2 \geqslant F(E_1)F(E_2)$. Here, the equality holds if and only if E_1 and E_2 are homothetic and situated in a position of homothety. Furthermore, if $0 \leqslant t \leqslant 1$, then the square root of $F(E(t,1-t))$ is a †convex function of t (H. Minkowski).

E. Specific Ovals

Suppose that we are given an equilateral triangle ABC. Draw three circles C_1, C_2, and C_3 with centers A, B, and C and radii equal to the length of the sides of ABC. The minor arcs AB, BC, and CA of the circles form an oval which is called a **Reuleaux triangle**. This oval is of constant breadth. Furthermore, given a fixed breadth D, the area $F(E)$ of an oval E of constant breadth D attains its minimum when E is a Reuleaux triangle. A Reuleaux triangle obtained from a triangle ABC revolves freely within the square of side AB and touches each side. In general, an oval which revolves touch-

ing the sides of a convex polygon from the inside is called an **inrevolvable** oval. Any such oval revolves inside some regular polygon (M. Fujiwara, S. Kakeya).

Various properties of an oval already described can be generalized to the case of a boundary of a convex body in $\mathbf{R}^n$. For the volumes of subsets A, B in $\mathbf{R}^n$ and $A+B=\{x+y \mid x\in A, y\in B\}$ in $\mathbf{R}^n$, the Brunn-Minkowski inequality

$$[\mathrm{vol}(A+B)]^{1/n} \geqslant [\mathrm{vol}(A)]^{1/n} + [\mathrm{vol}(B)]^{1/n}$$

holds.

F. Convex Cones

A nonempty subset X of $\mathbf{R}^n$ is called a convex cone if for any elements x, y of X and a nonnegative number a, ax and $x+y$ are contained in X. A convex cone is a convex set. Given any nonempty subset X of $\mathbf{R}^n$, the minimum convex cone $K(X)$ containing X exists. Given two convex cones X and Y, a convex cone $X+Y$, called the **sum** of X and Y, is defined as the set of elements $x+y$, where x, y are elements of X, Y. The intersection of convex cones X and Y is also a convex cone. Given a convex cone X, the subset of $\mathbf{R}^n$ consisting of the elements y such that $(x,y)\leqslant 0$ for any element x in X is a convex cone which is called the **dual convex cone** (or **conjugate convex cone**) of X, denoted by X^*. If X is a finite set, $K(X)$ is called a **convex polyhedral cone**. For example, if v is a nonzero vector, then the half-line $(v)=\{x \mid x=av, a\geqslant 0\}$ or the half-space $(v)^*=\{x \mid (v,x)\leqslant 0\}$ is a convex polyhedral cone. A convex polyhedral cone is closed. A convex cone X is a convex polyhedral cone if and only if X is the sum of a finite number of half-lines. Given convex cones X and Y, we have propositions (1)–(3): (1) If $X_1\subset X_2$, then $X_2^*\subset X_1^*$; (2) $(X_1+X_2)^*=X_1^*\cap X_2^*$; (3) $X_1^*+X_2^*\subset (X_1\cap X_2)^*$. If X_1 and X_2 are convex polyhedral cones, then $X_1^*+X_2^*=(X_1\cap X_2)^*$. Generally, $X\subset (X^*)^*=X^{**}$ for a convex cone X. If X is a closed convex cone, then $X=X^{**}$. Namely, the **duality principle** holds for closed convex cones. A linear subspace of $\mathbf{R}^n$ is a convex polyhedral cone. Also, if A is an $m\times n$ real matrix, the subsets $\{x \mid Ax=0, x\geqslant 0\}$ and $\{x \mid Ax\geqslant 0\}$ are convex polyhedral cones. Since the duality principle holds for convex polyhedral cones, we obtain the †Minkowski-Farkas theorem (i.e., if A is an $m\times n$ real matrix and v is an element of $\mathbf{R}^m$, then the equation $Ay=v$ has a solution $y\geqslant 0$ in $\mathbf{R}^n$ if and only if $(v,x)\geqslant 0$ for all $x\in\mathbf{R}^m$ such that ${}^tAx\geqslant 0$). (For linear inequalities → 255 Linear Programming.)

G. Convex Sets in Function Spaces

The definitions of convex sets and convex cones in $\mathbf{R}^n$ can be naturally extended to the case of any real linear space. Some of their properties can be generalized and applied to the case of function spaces.

(1) Let E be a †locally convex real topological linear space satisfying Hausdorff's †separation axiom. Let A and B be convex sets in E, and assume that B has interior points and $A\cap B$ is empty. Then A and B are separated by a hyperplane. Namely, there exists a nonzero †continuous linear functional f on E such that $\sup f(A)\leqslant \inf f(B)$.

(2) Let E be as in (1), and let C be a convex set in E. If a boundary point x of C admits a nonzero continuous linear functional f such that $f(x)=\sup f(C)$, we call such a point x a **supporting point** of C, and f a **supporting functional** of C. If C has interior points, then any boundary point x of C is a supporting point of C.

(3) Let C be a closed convex set of a †Banach space E. The set of supporting points of C is dense in its boundary.

A convex set C contained in the dual space E^* of a real topological linear space E is called a **regularly convex set** if for any f_0 in E^* not contained in C, there exists a point x_0 in E such that $\sup\{f(x_0) \mid f\in C\}<f_0(x_0)$.

Let E be a real topological linear space satisfying Hausdorff's separation axiom, and let C be a closed convex cone having 0 as its extreme point. Furthermore, assume that $C\cap(-C)=\{0\}$. If we set $x\leqslant y$ when $y-x\in C$, a partial ordering $\leqslant$ is defined in E. For example, if E is $\mathbf{R}^n$, then the **positive orthant** $C=\{x=(x_i) \mid x_i\geqslant 0, i=1,\ldots,n\}$ satisfies these requirements, and the partial ordering $x\leqslant y$ defined by means of C is equivalent to the relation $x_i\leqslant y_i$ for all i.

Some of the properties of matrices of positive entries or †integral operators whose †kernel functions are positive-valued can be generalized to properties of mappings $f: E\to E$ such that $f(C)\subset C$ (→ 255 Linear Programming; for the †Kreĭn-Milman theorem → 424 Topological Linear Spaces).

References

[1] W. Blaschke, Kreis und Kugel, Veit, 1916 (Chelsea, 1949).
[2] T. Bonnesen and W. Fenchel, Theorie der konvexen Körper, Erg. Math., Springer, 1934 (Chelsea, 1948).
[3] W. Fenchel, Convex cones, sets and functions, Lecture notes, Princeton Univ. Press, 1953.

[4] Convexity, Amer. Math. Soc. Proc. Symp. in Pure Math. VII (1963).
[5] M. G. Kreĭn and M. A. Rutman, Linear operators leaving invariant a cone in a Banach space, Amer. Math. Soc. Transl., 26 (1950). (Original in Russian, 1948.)
[6] H. Nikaido, Convex structures and economic theory, Academic Press, 1968.
[7] B. Grünbaum, Convex polytopes, Interscience, 1967.
[8] I. M. Yaglom (Jaglom) and V. G. Boltyanskiĭ, Convex figures, Holt, Rinehart and Winston, 1961. (Original in Russian, 1951.)
[9] G. Choquet, Lectures on analysis, vols. I, II, Benjamin, 1969.
[10] R. T. Rockafellar, Convex analysis, Princeton Univ. Press, 1970.

90 (VI.13) Coordinates

A. General Remarks

Suppose that we are given a Euclidean plane E^2 and two lines $X'X$ and $Y'Y$ in E^2 perpendicular to each other. Let O be the point of intersection of $X'X$ and $Y'Y$. We identify each of the straight lines $X'X$ and $Y'Y$ with the set of real numbers $\mathbf{R}$; the point O on each line is identified with zero. Let P be an arbitrary point in E^2. We draw lines PQ, PR parallel to $Y'Y$, $X'X$, where Q, R are on $X'X$, $Y'Y$, respectively. Let x and y be the real numbers corresponding to Q and R. Thus we obtain a mapping sending the point P to the ordered pair (x, y) of real numbers. This mapping gives a one-to-one correspondence between the points P of E^2 and the ordered pairs (x, y) of real numbers in $\mathbf{R}^2$. The numbers x and y are called the **coordinates** of P.

In general, given a set of mathematical objects, if we have a mechanism that assigns quantities to each element of the set, then such a mechanism is called a **coordinate system** on the set, and the quantities corresponding to each element are called its **coordinates**. In the previous example, the mechanism is called a rectangular coordinate system. Coordinate systems are also useful in expressing quantitative concepts by geometric ones which are intuitively easier to grasp, e.g., diagrams of train schedules and †nomograms. †Map projection, †graphical calculation, †descriptive geometry, etc., may be viewed as applications of the concept of coordinate systems.

In many cases, when we introduce a coordinate system in a space, it is determined uniquely by fixing a basic figure in the space. In the case of a rectangular coordinate system on a plane E^2, the basic figure consists of $X'X$ and $Y'Y$, which are called coordinate axes (the point O is called the origin). Sometimes it is convenient to consider real-valued functions f and g on $\mathbf{R}$ and a coordinate system on the plane E^2 determined by the function that sends an arbitrary point P to $(f(x), g(y))$, where (x, y) are the coordinates of P in the rectangular coordinate system. Logarithmic papers, †probability papers, and †stochastic papers (binomial probability papers), etc., are constructed in this way to fit their respective purposes.

In various branches of mathematics there are many varieties of coordinate systems. In this article we deal with frames and coordinates, curvilinear coordinates, and local coordinates.

B. Frames and Coordinates

Suppose that we are given a space M and a †transformation group G acting on M. It is desirable to introduce a coordinate system that best represents the geometric structure of M. Let G_* be a set of figures in M such that G acts †simply transitively on G_*. Each element of G_* is called a **frame**. Utilizing each frame as basic figure, we introduce a coordinate system that is "G-invariant" in the following sense: Let $R \in G_*$, $X \in M$, and $C_R(X)$ be the coordinates of X with R as basic figure. Then the coordinate system is G-invariant if $C_R(X) = C_{gR}(gX)$ for any element g in G. If we have such a coordinate system for each $R \in G_*$, then the expressions of geometric properties of M in terms of the coordinates are independent of the choice of frames.

(1) Projective Coordinates. Let M be an n-dimensional †projective space $\mathbf{P}^n$ over a field K, and let G be the †projective transformation group of $\mathbf{P}^n$. As a frame we can take the system of $n+1$ points $(A_0, A_1, \ldots, A_n)$ in general position. The †homogeneous coordinates of an arbitrary point $X \in \mathbf{P}^n$ are given by the $(n+1)$-tuple $(x_0, x_1, \ldots, x_n)$ satisfying the equation $X = \sum_{j=0}^n x_j A_j$, $x_j \in K$. They are called †projective coordinates. In fact, if $(x_0, x_1, \ldots, x_n) \neq (0, 0, \ldots, 0)$, then $(x_0, x_1, \ldots, x_n)$ and $(\lambda x_0, \lambda x_1, \ldots, \lambda x_n)$ $(\lambda \neq 0)$ represent the same point in $\mathbf{P}^n$. A †hyperplane π of $\mathbf{P}^n$ is expressed as the set of points whose coordinates $(x_0, x_1, \ldots, x_n)$ satisfy a linear homogeneous equation $\sum_{j=0}^n x_j u_j = 0$, $u_j \in K$. Therefore, the hyperplane π is represented by the homogeneous coordinates $(u_0, u_1, \ldots, u_n)$, called †hyperplane coordinates of π.

(2) Affine Coordinates. Let M be an n-dimensional †affine space E^n, and let G be the †group of affine transformations of E^n. As a frame we can take the system $(O; \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$, where O, called the origin, is a point in E^n, and the set of vectors $\{\mathbf{e}_i\}$ is a basis of the †standard vector space of E^n. Then †inhomogeneous coordinates of an arbitrary point $X \in E^n$ are given by the n-tuple $(x_1, x_2, \ldots, x_n)$, where $X = O + \sum_{i=1}^n x_i \mathbf{e}_i$. They are also called †affine coordinates of X in E^n. Sometimes we replace G above by the group of †equivalent affinities and consider the frames $(O; \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$ such that the volume of $[\mathbf{e}_1, \ldots, \mathbf{e}_n] = 1$.

Furthermore, if E^n has the structure of a †Euclidean space, we sometimes replace G by the †group of motions and consider a system of **rectangular coordinates** determined by an †orthogonal frame, that is, a frame $(O; \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$ such that the inner product $(\mathbf{e}_i, \mathbf{e}_j) = \delta_{ij}$, where δ is the †Kronecker delta. By contrast, the general affine coordinate system of a Euclidean space is called a system of **oblique coordinates**. In this case the inner products $(\mathbf{e}_i, \mathbf{e}_j) = g_{ij}$ are invariants of Euclidean geometry, and the distance ρ between two points (x_i) and (y_i) is given by $\rho = (\sum_{i,j=1}^n g_{ij}(y_i - x_i)(y_j - x_j))^{1/2}$. We sometimes consider an oblique coordinate system satisfying $(\mathbf{e}_i, \mathbf{e}_i) = 1$ $(i = 1, \ldots, n)$. In such cases the angle θ_{ij} between two basis vectors $\mathbf{e}_i$ and $\mathbf{e}_j$ is determined by $g_{ij} = \cos\theta_{ij}$.

(3) Barycentric Coordinates. In an n-dimensional affine space E^n, we take $n+1$ linearly independent points $A_0, A_1, \ldots, A_n$ and denote the position vectors from a point O to these points by $\mathbf{a}_0, \mathbf{a}_1, \ldots, \mathbf{a}_n$, respectively. Then for any point $X \in E^n$ there exists a unique set of numbers $(\lambda_0, \lambda_1, \ldots, \lambda_n)$ such that $X = O + \sum_{j=0}^n \lambda_j \mathbf{a}_j$, $\sum_{j=0}^n \lambda_j = 1$. We call these numbers **barycentric coordinates** of X in E^n. They are independent of the choice of the point O.

(4) Plücker Coordinates. Let $V(n, m)$ be the set of all m-dimensional subspaces in an n-dimensional projective space $\mathbf{P}^n$. Then $V(n, m)$ has the structure of a †Grassmann manifold. In order to introduce a coordinate system on $V(n, m)$, we fix a projective coordinate system on $\mathbf{P}^n$. An m-dimensional subspace $\pi \in V(n, m)$ in $\mathbf{P}^n$ is spanned by $m+1$ independent points $B_0, B_1, \ldots, B_m \in \mathbf{P}^n$. We denote projective coordinates of these points by $(b_{0j}), (b_{1j}), \ldots, (b_{mj})$ and construct the determinants

$$p_{j_0 j_1 \ldots j_m} = \begin{vmatrix} b_{0j_0} b_{0j_1} \ldots b_{0j_m} \\ \cdots\cdots\cdots \\ b_{mj_0} b_{mj_1} \ldots b_{mj_m} \end{vmatrix}, \quad 0 \leqslant j_0, \ldots, j_m \leqslant n.$$

Then the subspace π can be represented by homogeneous coordinates $(\ldots, p_{j_0 j_1 \ldots j_m}, \ldots)$. These coordinates are independent of the choice of $m+1$ points that span π and are called **Plücker coordinates** (or **Grassmann coordinates**) of π in $V(n, m)$. In these coordinates, the $p_{j_0 j_1 \ldots j_m}$ are alternating and satisfy the **Plücker relations**

$$\sum_{k=0}^{m+1} (-1)^k p_{i_1 i_2 \ldots i_m j_k} p_{j_0 \ldots \hat{j}_k \ldots j_{m+1}} = 0,$$

where $\hat{j}_k$ means that j_k is removed. In particular, when $n = 3$ and $m = 1$, we have only one Plücker relation $Q: p_{01}p_{23} - p_{02}p_{13} + p_{03}p_{12} = 0$, which is a homogeneous equation of the second degree. In other words, the set $V(3, 1)$ of all lines in a 3-dimensional projective space $\mathbf{P}^3$ is realized as a quadric surface Q in a 5-dimensional projective space $\mathbf{P}^5$ that has $(p_{01}, p_{02}, p_{03}, p_{12}, p_{13}, p_{23})$ as projective coordinates. Moreover, when $\mathbf{P}^3$ is a complex projective space, we put

$$p_{01} = \xi_0 + i\xi_3, \quad p_{02} = \xi_1 + i\xi_4, \quad p_{03} = \xi_2 + i\xi_5,$$

$$p_{23} = \xi_0 - i\xi_3, \quad p_{13} = -\xi_1 + i\xi_4, \quad p_{12} = \xi_2 - i\xi_5$$

and obtain a relation

$$\xi_0^2 + \xi_1^2 + \xi_2^2 + \xi_3^2 + \xi_4^2 + \xi_5^2 = 0$$

corresponding to the Plücker relation. Thus every line in $\mathbf{P}^3$ can be represented by homogeneous coordinates $(\xi_0, \xi_1, \ldots, \xi_5)$, which we call **Klein's line coordinates**.

(5) $(n+2)$-Hyperspherical Coordinates. Let $(x_0, x_1, \ldots, x_n, x_\infty)$ be projective coordinates in an $(n+1)$-dimensional real projective space $\mathbf{P}^{n+1}$. An n-dimensional †conformal space S^n is realized as a quadric hypersurface S^n in $\mathbf{P}^{n+1}: \sum_{i,j=1}^n g_{ij} x_i x_j - 2x_0 x_\infty = 0$, where (g_{ij}) is a positive definite symmetric matrix. A general point in $\mathbf{P}^{n+1}$ represents a †hypersphere of S^n. That is, a hypersphere represented by a point $X \in \mathbf{P}^{n+1}$ is realized as the intersection of S^n with the †polar hyperplane of X with respect to S^n; according as X lies outside of S^n, on S^n, or inside of S^n, it represents a real hypersphere, a point hypersphere, or an imaginary hypersphere. Therefore any hypersphere of S^n in $\mathbf{P}^{n+1}$ is expressed by homogeneous coordinates $(x_0, x_1, \ldots, x_n, x_\infty)$, called **$(n+2)$-hyperspherical coordinates**. When $n = 2$, they are called **tetracyclic coordinates**, and when $n = 3$, **pentaspherical coordinates**. Therefore, if we restrict $(n+2)$-hyperspherical coordinates for points on S^n, then they satisfy the quadratic relation stated before. In the frame $(A_0, A_1, \ldots, A_n, A_\infty)$ of $\mathbf{P}^{n+1}$ which defines the $(n+2)$-hyperspherical coordinates, A_0 and A_∞ are points on S^n, and the other A_i are real hyperspheres passing through the points A_0 and A_∞. It is possible to choose a frame $(A_0, A_1, \ldots, A_n, A_\infty)$ such that the equation for S^n becomes $\sum_{i=1}^n x_i^2 - 2x_0 x_\infty = 0$ (i.e.,

$g_{ij}=\delta_{ij}$). Among hypersurfaces in S^n, one that is expressed by a homogeneous equation of the second degree with respect to $(n+2)$-hyperspherical coordinates is called a **cyclide**. It is an algebraic surface of the fourth order and is an enveloping surface of the family of hyperspheres that are tangent to n fixed hyperspheres.

(6) Moving Coordinates. When we study the differential geometry of an m-dimensional surface W in a space M on which a transformation group G acts, it is often preferable to take a frame or frames at each point of W and consider a †connection among them. These frames are called **moving frames**, and the set of coordinate systems with respect to moving frames is called a **moving coordinate system** (→ 111 Differential Geometry of Curves and Surfaces).

C. Curvilinear Coordinates

Let $(x_1, x_2, \ldots, x_n)$ be a rectangular coordinate system on an n-dimensional Euclidean space E^n. If $x_i=x_i(u_1,u_2,\ldots,u_n)$, $i=1,\ldots,n$, are functions of n variables $(u_1,u_2,\ldots,u_n)$ of class C^r $(r\geqslant 1)$ and the †functional determinant $D(x_1,\ldots,x_n)/D(u_1,\ldots,u_n)$ is not equal to zero in some open domain, then $(u_1,u_2,\ldots,u_n)$ are considered local coordinates in E^n. We call them **curvilinear coordinates** of E^n. A hypersurface $u_i=$constant (obtained by fixing the value of one of the variables u_i) is called a **coordinate hypersurface**, and a curve $u_j=$constant $(j\neq i)$ is called a **coordinate curve**. The line element ds of a Euclidean space E^n is given by

$$ds^2=\sum_{k=1}^{n} dx_k^2=\sum_{i,j=1}^{n} g_{ij}du_idu_j,$$

$$g_{ij}=\sum_{k=1}^{n}\frac{\partial x_k}{\partial u_i}\frac{\partial x_k}{\partial u_j}.$$

Thus E^n is equipped with a †Riemannian metric. However, as E^n is †flat, its †curvature tensor satisfies $R^i_{jkl}=0$. If the metric is diagonal, namely, if $ds^2=\sum_{i=1}^n g_i du_i^2$, the coordinates are called **orthogonal curvilinear coordinates**. Moreover, if $g_1=\ldots=g_n$, the coordinates are called **isothermal coordinates**. The metric is diagonal if and only if coordinate curves are mutually perpendicular at the points of intersection. Actually, curvilinear coordinate systems that are often used practically are diagonal. The concept of curvilinear coordinates has been generalized to the case of †differentiable manifolds and is utilized to determine their local coordinates.

On any 2-dimensional Riemannian manifold there always exist isothermal coordinates in a neighborhood of any point [7].

(1) Curvilinear Coordinates on Planes or Spaces (→ Appendix A, Table 3.V). Let (x,y) be rectangular coordinates of a point in a Euclidean plane E^2. We have the following coordinate systems on E^2:

Polar coordinates (r,θ), where

$$x=r\cos\theta,\qquad y=r\sin\theta.$$

Elliptic coordinates (λ,μ), where

$$x^2=(\lambda+a^2)(\mu+a^2)/(a^2-b^2),$$

$$y^2=(\lambda+b^2)(\mu+b^2)/(b^2-a^2),$$

$$a>b>0,\qquad \lambda>-b^2>\mu>-a^2.$$

Parabolic coordinates (α,β), where

$$x=-(\alpha+\beta),\qquad y=\sqrt{-4\alpha\beta},\qquad \alpha>0>\beta.$$

Equilateral (or rectangular) hyperbolic coordinates (u,v), where

$$x=uv,\qquad y=(u^2-v^2)/2.$$

Bipolar coordinates (ξ,η), where

$$x=a\sin\xi/(\cosh\xi+\cos\eta),$$

$$y=a\sin\eta/(\cosh\xi+\cos\eta),$$

$$-\infty<\xi<\infty,\qquad 0\leqslant\eta\leqslant 2\pi.$$

Next we consider the case of a 3-dimensional Euclidean space E^3 and let (x,y,z) be rectangular coordinates on E^3. The following systems of coordinates on E^3 are sometimes useful.

Cylindrical coordinates (r,θ,z), where

$$x=r\cos\theta,\qquad y=r\sin\theta,\qquad z=z.$$

Spherical coordinates (r,θ,φ), where

$$x=r\sin\theta\cos\varphi,\quad y=r\sin\theta\sin\varphi,\quad z=r\cos\theta.$$

Ellipsoidal coordinates (λ,μ,ν), where

$$x^2=(\lambda+a^2)(\mu+a^2)(\nu+a^2)/(a^2-b^2)(a^2-c^2),$$

$$y^2=(\lambda+b^2)(\mu+b^2)(\nu+b^2)/(b^2-c^2)(b^2-a^2),$$

$$z^2=(\lambda+c^2)(\mu+c^2)(\nu+c^2)/(c^2-a^2)(c^2-b^2),$$

$$a>b>c>0,\quad \lambda>-c^2>\mu>-b^2>\nu>-a^2.$$

These coordinate systems are all systems of orthogonal curvilinear coordinates. Suppose that we are given two rectangular coordinate systems (ξ,η,ζ) and (x,y,z) sharing the same origin. The correlation of the two is given by **Euler's angles** (θ,φ,ψ), where θ, φ, and ψ are the angles between the z-axis and ζ-axis, zx-plane and $z\zeta$-plane, and $\zeta\xi$-plane and ζz-plane, respectively. The Euler angles θ, φ, and ψ are subject to the inequalities $0\leqslant\theta\leqslant\pi$ and $0\leqslant\varphi$, $\psi<2\pi$. They are often utilized in the dynamics of rigid bodies.

(2) Multipolar Coordinates. Let $P_1, P_2, \ldots, P_m$ be m points in general position in an n-dimensional Euclidean space E^n, $m\leqslant n$. If we

denote by ρ_i ($\geqslant 0$) the distance between a point X of E^n and P_i, then $(\rho_1, \rho_2, \ldots \rho_m)$ can be regarded as coordinates of a point X contained in a suitable domain of E^n. They are called **multipolar coordinates**. In particular, if $m=2$ they are called **bipolar coordinates**, and if $m=3$, **tripolar coordinates**. When $m>n$, these coordinates satisfy $m-n$ relations. Next let α_1, $\alpha_2, \ldots, \alpha_m$ be m hyperplanes in general position in E^n, $m \leqslant n$. For an arbitrary point X of E^n, we denote by ξ_i the directed distance of X from each hyperplane α_i. The m-tuple $(\xi_1, \xi_2, \ldots, \xi_m)$ provides coordinates of X that are called **multiplanar coordinates** in E^n. When $m>n$, these coordinates satisfy $m-n$ relations. In particular, when $n=2$ and $m=3$, they are called **trilinear coordinates**. In this case, if we denote by S the area of the triangle defined by three lines $\alpha_1, \alpha_2, \alpha_3$, and by a_1, a_2, a_3 the lengths of the three sides of the triangle, then the trilinear coordinates (ξ_1, ξ_2, ξ_3) satisfy a linear relation $a_1\xi_1 + a_2\xi_2 + a_3\xi_3 = 2S$.

(3) Tangential Polar Coordinates. In a Euclidean plane E^2, we take a directed line l_0 passing through a point O. For an arbitrary directed line g, let p be the directed distance between O and g, and let θ be the angle between l_0 and g. Then (p, θ) are called **tangential polar coordinates** (Fig. 1). They are useful for representing tangent lines to curves in E^2. Let C be an †oval in E^2. A line is called a †supporting line of C if its intersection with C consists of a point or a line segment. In this case we take the origin O inside C and consider the coordinates (p, θ) of the supporting lines of C. Then the equation of C can be represented as $p = p(\theta)$, where $p(\theta)$ is a periodic function of period 2π. The coordinates (p, θ) are especially useful when the function $p(\theta)$ can be expanded in a †Fourier series. In the case of Euclidean space E^3, the notion of tangential polar coordinate system can also be defined by using tangent planes.

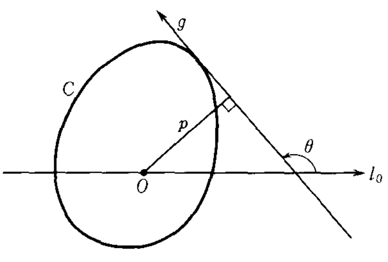


Fig. 1
Tangential polar coordinates.

(4) Normal Coordinates. Let M be an n-dimensional †Riemannian manifold, and let $T_A(M)$ be the tangent space to M at a point A. For each tangent vector $\mathbf{v} \in T_A(M)$, we draw a †geodesic through A with the initial direction $\mathbf{v}$ and take a point P on the geodesic such that the distance from A to P is equal to the length of $\mathbf{v}$. Then the correspondence that sends $\mathbf{v}$ to P is a †diffeomorphism of a neighborhood of the zero vector 0 of $T_A(M)$ with a neighborhood of A in M. Therefore the components $(v^1, v^2, \ldots, v^n)$ of $\mathbf{v}$ with respect to a basis of $T_A(M)$ give the coordinates of the points P contained in a suitable neighborhood of A. We call them **normal coordinates** about the point A of M. In these coordinates, each geodesic passing through A is given by equations $v^i = \alpha^i r$ $(i = 1, 2, \ldots, n)$, where the (α^i) are components of the unit vector in the direction of $\mathbf{v}$ and r is the parameter that represents the arc length from A to the point $(v^1, \ldots, v^n)$. In particular, when $n=2$, we fix a tangent vector $\mathbf{v}_0$ at A and denote the angle between $\mathbf{v}$ and $\mathbf{v}_0$ by θ. Then (r, θ) are coordinates of P called **geodesic polar coordinates**. The notion of normal coordinates can also be defined for †Lie groups or differentiable manifolds with †affine connections.

D. Local Coordinates

Suppose that we have a space M that has a covering by a family of open neighborhoods with coordinate systems. If, for each pair of neighborhoods with nonempty intersection, the coordinate transformation in the intersection satisfies certain specified conditions, then a mathematical structure on M can be defined.

Let E be a †topological space. Suppose that Φ is a family of open sets in E such that the union of any number of open sets in Φ and the intersection of any finite number of open sets in Φ also belong to Φ. A set Γ of †homeomorphisms is called a **pseudogroup of transformations** on E if Γ satisfies the following three conditions: (i) Any homeomorphism $f \in \Gamma$ is defined on an open set $U \in \Phi$, and $f(U) \in \Phi$. (ii) When an open set $U \in \Phi$ is expressed as the union of a family $\{U_i\}$ of open sets $U_i \in \Phi$, a homeomorphism f defined on U belongs to Γ if and only if its restriction to each U_i belongs to Γ. (iii) For any open set $U \in \Phi$, the identity mapping on U belongs to Γ, and if $f, g \in \Gamma$, then the inverse f^{-1} and the composition $g \circ f$, if it exists, belong to Γ.

Let E and M be topological spaces. A homeomorphism $\varphi: U \rightarrow V$ of an open set U in E to an open set V in M is called a **local coordinate system** of M with respect to E. For two local coordinate systems $\varphi_1: U_1 \rightarrow V_1$ and $\varphi_2: U_2 \rightarrow V_2$, the homeomorphism

$$\varphi_2^{-1} \circ \varphi_1 : \varphi_1^{-1}(V_1 \cap V_2) \rightarrow \varphi_2^{-1}(V_1 \cap V_2)$$

is called a **transformation of local coordinates**.

Let Γ be a pseudogroup of transformations

on E. A set Σ of local coordinate systems of M with respect to E is said to define a **Γ-structure** on M if Σ satisfies the following two conditions: (1) the totality of the images of local coordinate systems belonging to Σ covers M; (2) if two local coordinate systems φ_1 and φ_2 of Σ have a transformation of local coordinates, it belongs to Γ. Now suppose that two sets Σ and Σ' of local coordinate systems define Γ-structures on M. If the union of Σ and Σ' defines a Γ-structure, then we say that the first two Γ-structures are equivalent. Let Γ be a pseudogroup of †diffeomorphisms, each defined from an open subset of the n-dimensional space $\mathbf{R}^n$ onto another open set. If a Γ-structure is defined on a space M, then M is an n-dimensional †differentiable manifold.

On the other hand, let Γ be a pseudogroup of complex analytic homeomorphisms in an n-dimensional complex number space $\mathbf{C}^n$. If a Γ-structure is defined on a space M, then M is an n-dimensional †complex analytic manifold. Locally homogeneous spaces, †foliated manifolds, the †fiber bundles are all equipped with local coordinate systems with suitable Γ-structures.

References

[1] E. Cartan, La méthode du repère mobile, la théorie des groupes continus et les espaces généralisés, Actualités Sci. Ind., Hermann, 1935.

[2] C. Ehresmann, Sur la théorie des espaces fibrés, Colloques Intern. Topologie algébrique, Paris, 1949, C. N. R. S. XII, 3–15.

[3] F. Klein, Vorlesungen über höhere Geometrie, Springer, third edition, 1926 (Chelsea, 1957).

[4] B. L. van der Waerden, Einführung in die algebraische Geometrie, Springer, 1939.

[5] O. Veblen and J. H. C. Whitehead, The foundations of differential geometry, Cambridge Univ. Press, 1932.

[6] Wu Wen-Tsün and G. Reeb, Sur les espaces fibrés et les variétés feuilletées, Actualités Sci. Ind., Hermann, 1952.

[7] S. S. Chern, An elementary proof of the existence of isothermal parameters on a surface, Proc. Amer. Math. Soc., 6 (1955), 771–782.

91 (IX.3) Covering Spaces

A. General Remarks

A continuous mapping $p: \tilde{Y} \to Y$ of an †arcwise connected topological space $\tilde{Y}$ onto a connected topological space Y is called a **covering mapping (covering map)** if the following condition (C) is satisfied: (C) Each point of Y has an open neighborhood V such that every †connected component of $p^{-1}(V)$ is mapped homeomorphically onto V by p. Here we note that Y is in fact arcwise connected.

If there is a covering mapping $p: \tilde{Y} \to Y$, we call $\tilde{Y}$ a **covering space** of Y and $(\tilde{Y}, p, Y)$ a **covering**. In particular, for a †differentiable manifold Y, if $\tilde{Y}$ is also a differentiable manifold and p is differentiable, then $\tilde{Y}$ is called a **covering (differentiable) manifold** of Y. (In the theory of †Riemann surfaces, a †covering surface may have some †branch points violating condition (C). Upon removing such points, we obtain a covering space as defined above.)

For each †path $w: I \to Y$ $(I=[0,1])$ of Y, a path $\tilde{w}: I \to \tilde{Y}$ with $p \circ \tilde{w} = w$ is uniquely determined by the point $\tilde{w}(1) \in p^{-1}(w(1))$, and a bijection $w_{\#}: p^{-1}(w(1)) \to p^{-1}(w(0))$ is determined by $w_{\#}(\tilde{w}(1)) = \tilde{w}(0)$. Thus there exists a one-to-one correspondence between $p^{-1}(y)$ and $p^{-1}(y')$ for every pair of points y, y' of Y, and $(\tilde{Y}, p, Y, p^{-1}(y_0))$ is a †locally trivial fiber space with discrete fiber $p^{-1}(y_0)$. When the cardinal number of $p^{-1}(y)$ is a finite number n, we call $(\tilde{Y}, p, Y)$ an ***n*-fold covering**. In this case, for *a* †loop $w(I, \dot{I}) \to (Y, y_0)$ with base point y_0, $w_{\#}: p^{-1}(y_0) \to p^{-1}(y_0)$ is a permutation of the n elements in $p^{-1}(y_0)$, and we obtain a homomorphism of the †fundamental group $\pi_1(Y) = \pi_1(Y, y_0)$ of Y into the †symmetric group $\mathfrak{S}_n$, given by the correspondence $w \to w_{\#}$. The permutation group $\mathfrak{M}$, which is the image of this homomorphism, is called the **monodromy group** of the n-fold covering.

Two coverings $(\tilde{Y}_i, p_i, Y)$ $(i=1,2)$ are said to be **equivalent** if there is a homeomorphism $\varphi: \tilde{Y}_1 \to \tilde{Y}_2$ with $p_2 \circ \varphi = p_1$; such a φ is called an **equivalence**. In particular, a self-equivalence $\varphi: \tilde{Y} \to \tilde{Y}$ of a covering $(\tilde{Y}, p, Y)$ is called a **covering transformation**. The set π of all covering transformations forms a group by the composition of mappings, which is called the **covering transformation group** of $\tilde{Y}$. We call $(\tilde{Y}, p, Y)$ a **regular covering** if for each $y \in Y$ and $\tilde{y}_1$, $\tilde{y}_2 \in p^{-1}(y)$, there exists a unique covering transformation that maps $\tilde{y}_1$ to $\tilde{y}_2$. In this case, the †orbit space $\tilde{Y}/\pi$ is homeomorphic to Y, $(\tilde{Y}, p, Y, \pi)$ is a †principal bundle, and the monodromy group $\mathfrak{M}$ is isomorphic to π.

For a covering $(\tilde{Y}, p, Y)$, we call $\tilde{Y}$ a **covering group** of Y if $\tilde{Y}$ and Y are topological groups and p is a homomorphism. Then $(\tilde{Y}, p, Y)$ is a regular covering, and its covering transformation group is isomorphic to $p^{-1}(e)$ (e is the identity element of Y), which is a discrete subgroup lying in the center of $\tilde{Y}$ (→ 423 Topological Groups O).

B. Universal Covering Spaces

For a covering $(\tilde{Y}, p, Y)$, we have the following relations of the †homotopy groups: $p_*: \pi_i(\tilde{Y}) \to \pi_i(Y)$ is isomorphic (monomorphic) for $i \geqslant 2$ $(i=1)$, and $\pi_1(Y)/p_*(\pi_1(\tilde{Y}))$ is in one-to-one correspondence with $p^{-1}(y_0)$ $(y_0 \in Y)$. If $\tilde{Y}$ is †simply connected, $\tilde{Y}$ is called a **universal covering space** of Y; if in addition $\tilde{Y}$ is a covering group, $\tilde{Y}$ is called a **universal covering group** of Y.

For a †locally arcwise connected space Y, a covering $(\tilde{Y}, p, Y)$ is regular if and only if $p_*(\pi_1(\tilde{Y}))$ is a normal subgroup of $\pi_1(Y)$, and a universal covering space of Y is a covering space of any covering space of Y. Moreover, if Y is a topological group, any covering space $\tilde{Y}$ of Y can be given a unique topological group structure with which $\tilde{Y}$ is a covering group of Y.

Let Y be an arcwise connected, locally arcwise connected, and †locally simply connected space. Then the following classification theorem of coverings holds: The set of equivalence classes of coverings of Y is in one-to-one correspondence with the set of conjugate classes of subgroups of the fundamental group $\pi_1(Y)$; in particular, the equivalence class of a covering $(\tilde{Y}, p, Y)$ corresponds to the conjugate class of the subgroup $p_*(\pi_1(\tilde{Y}))$. Also, there is a unique universal covering space $\tilde{Y}$ of Y up to homeomorphism. If in addition Y is a topological group, then $\tilde{Y}$ is a unique universal covering group of Y up to isomorphism of topological groups. Such a space $\tilde{Y}$ is obtained as follows: Consider the †path space $\Omega(Y; y_0, Y)$ of all paths in Y starting from a fixed point $y_0 \in Y$, and define two paths $w_0, w_1: (I, 0) \to (Y, y_0)$ such that $w_0(1) = w_1(1)$ to be equivalent if and only if there is a †homotopy w_t: $(I, 0) \to (Y, y_0)$ with $w_t(1) = w_0(1)$ $(0 \leqslant t \leqslant 1)$. Then we obtain the †identification space $\tilde{Y}$ of $\Omega(Y; y_0, Y)$ by this equivalence relation and the mapping $p: \tilde{Y} \to Y$ by $p\{w\} = w(1)$; this $\tilde{Y}$ is the universal covering space of Y.

Let $(\tilde{Y}, p, Y)$ be a regular covering with the covering transformation group π. Then there is a †locally trivial fiber space $(Y', q, B, \tilde{Y})$ such that the total space Y' has the same †(co) homology groups as Y, and the base space B is an †Eilenberg-MacLane space $K(\pi, 1)$. The (co) homology †spectral sequence of this fiber space is called that of the given regular covering $(\tilde{Y}, p, Y)$, E_∞ is a bigraded module †associated with a certain filtration of the †singular (co) homology module $H(Y)$, and E_2 is the (co) homology module $H(\pi; H(\tilde{Y}))$ of the group π, where π operates on the coefficient module $H(\tilde{Y})$ via the induced homomorphisms of covering transformations (→ 148 Fiber Spaces).

For any group π, there is a regular covering (E_π, p, B_π) with the covering transformation group π such that E_π is †contractible; B_π is an Eilenberg-MacLane space $K(\pi, 1)$. We can take S^1 (1-sphere) as $B_{\mathbf{Z}}$ of the infinite cyclic group $\mathbf{Z}$, and the following infinite lens space as $B_{\mathbf{Z}_k}$ of the finite cyclic group $\mathbf{Z}_k$ (→ 70 Complexes).

C. Lens Spaces

Let k be a positive integer and $l_1, \ldots, l_n$ be integers prime to k. Let $S^{2n+1} = \{(z_0, \ldots, z_n) \in \mathbf{C}^{n+1} \mid |z_0|^2 + \ldots + |z_n|^2 = 1\}$ be the unit sphere in the $(n+1)$-dimensional complex linear space $\mathbf{C}^{n+1}$, and define the rotation γ by $\gamma(z_0, z_1, \ldots, z_n) = (z_0 \exp 2\pi i/k, z_1 \exp 2\pi l_1 i/k, \ldots, z_n \exp 2\pi l_n i/k)$. Then the †orbit space $S^{2n+1}/\mathbf{Z}_k = L(k; l_1, \ldots, l_n)$, where $\mathbf{Z}_k = \mathbf{Z}/k\mathbf{Z}$ is interpreted as the cyclic group generated by γ, is called a **lens space**. It is an orientable $(2n+1)$-dimensional †differentiable manifold. Also, the **infinite lens space** $L^\infty(k) = L(k; 1, \ldots, 1, \ldots)$ is defined by taking $n = \infty$; the infinite sphere S^∞ is a k-fold covering space of $L^\infty(k)$, and $L^\infty(k) = B_{\mathbf{Z}_k} = K(\mathbf{Z}_k, 1)$. Its †cohomology ring is given as follows.

(1) For integral coefficients, $H^{2i+1}(L^\infty(k)) = 0$, $H^{2i}(L^\infty(k)) = \mathbf{Z}_k$ $(i > 0)$, and the †cup product of generators of degree $2i$ and $2j$ is a generator of degree $2(i+j)$.

(2) Let $k = pk'$ (p is a prime). If $p \neq 2$, or $p = 2$ and k' is even, $H^*(L^\infty(k); \mathbf{Z}_p) = \wedge(e_1) \otimes \mathbf{Z}_p[e_2]$. If $p = 2$ and k' is odd, $H^*(L^\infty(k); \mathbf{Z}_2) = \mathbf{Z}_2[e_1]$ (e_i is an element of degree i). Here $\wedge$ indicates the †exterior algebra over $\mathbf{Z}_p$, and $\mathbf{Z}_p[\ ]$ the †polynomial ring over $\mathbf{Z}_p$.

Two lens spaces $L(k; l_1, \ldots, l_n)$ and $L(k'; l'_1, \ldots, l'_n)$ are of the same homotopy type if and only if $k = k'$ and there is an integer m prime to k with

$$l_1 \ldots l_n \equiv \pm m^{n+1} l'_1 \ldots l'_n \pmod{k}$$

[8]. Furthermore, the condition $k = k'$, $l \equiv \pm l'^{\pm 1} \pmod{k}$ holds if and only if $L(k; l)$ and $L(k'; l')$ are homeomorphic. (Sufficiency is shown in [1]; necessity follows from the fact that the †Hauptvermutung is valid for combinatorial 3-manifolds and that the condition holds if the polyhedra $L(k; l)$ and $L(k; l')$ have isomorphic subdivisions [6].) Also → [7, 11] and 65 Combinatorial Manifolds.

References

[1] H. Seifert and W. Threlfall, Lehrbuch der Topologie, Teubner, 1934 (Chelsea, 1965).
[2] N. Steenrod, The topology of fiber bundles, Princeton Univ. Press, 1951.
[3] S.-T. Hu, Homotopy theory, Academic Press, 1959.

[4] S. Eilenberg and S. MacLane, Homology of spaces with operators II, Trans. Amer. Math. Soc., 65 (1949), 49–99.
[5] J. H. C. Whitehead, On incidence matrices, nuclei and homotopy types, Ann. Math., (2) 42 (1941), 1197–1239.
[6] K. Reidemeister, Homotopieringe und Linsenräume, Abh. Math. Sem. Univ. Hamburg, 11 (1935), 102–109.
[7] W. Franz, Über die Torsion einer Überdeckung, J. Reine Angew. Math., 173 (1935), 245–254.
[8] P. Olum, Mappings of manifolds and the notion of degree, Ann. Math., (2) 58 (1953), 458–480.
[9] W. S. Massey, Algebraic topology; An introduction, Harcourt, Brace & World, 1967.
[10] I. M. Singer and J. A. Thorpe, Lecture notes on elementary topology and geometry, Scott, Foresman, 1967.
[11] M. M. Cohen, A course in simple homotopy theory, Springer, 1972.

92 (IV.15) Crystallographic Groups

A. Space Groups and Point Groups

Let G be a discrete subgroup of the group of †motions $\mathfrak{G}$ in the real n-dimensional Euclidean space V. If G contains n linearly independent translations, then G is called an n-dimensional **crystallographic space group**, a **space group**, or a **crystallographic group**. The following assertions are known to be equivalent to each other: (i) A discrete subgroup G of $\mathfrak{G}$ is a space group; (ii) $\mathfrak{G}/G$ is compact; (iii) V/G is compact; (iv) There is a compact subset P of V such that $V=GP=\bigcup_{g\in G} gP$ [1]. The subgroup T consisting of all †translations in G is called the **lattice group** of G. T is a normal subgroup of G and is generated by n linearly independent translations, say, $t_1, t_2, \ldots, t_n$. Take a point x of V. Then the T-orbit of the point x is called a **lattice** of G (or T). $t_1, \ldots, t_n$ can be identified with a basis of V. Let K be the quotient group G/T and $k=gT$ an element of K. Then k gives rise to a linear transformation $\bar{k}$ on T via the formula $\bar{k}(t)=gtg^{-1}$ (the product in G), $t\in T$, and K can be regarded as a subgroup of the †orthogonal group of V via the above identification of T with a lattice in V. That is, the set of orthogonal transformations stabilizing x, $\{\bar{g} \mid (\bar{g}(y)=x+g(y)-g(x), y\in V), g\in G\}$ forms a group isomorphic to K. The quotient group K is a finite group and is called the **point group** of G. A group isomorphic to a point group of an n-dimensional space group is sometimes called an n-dimensional **crystallographic group**. An algebraic characterization of space groups among abstract groups is given by "A group G is isomorphic to an n-dimensional space group $\Leftrightarrow$ G has a normal †free Abelian subgroup of †rank n, which is maximal Abelian and has finite index" [1].

By representing the action of K on T in terms of a basis of T, K can be regarded as a subgroup of $GL(n, \mathbf{Z})$. Hence we obtain the following **crystallographic restriction**: If K contains an element of order m, then $n \geqslant \varphi(m)$. Here φ is the †Euler function. For example, if $n=2, 3$, then $m=1, 2, 3, 4$ or 6; and if $n=4, 5$, then $m=1, 2, 3, 4, 5, 6, 8, 10$, or 12. The number of nonconjugate finite subgroups of $GL(n, \mathbf{Z})$ is finite (Jordan-Zassenhaus theorem; → [2, 3, 21]).

Two space groups G_1 and G_2 are called **equivalent** if they are conjugate via an †affine transformation f of V, i.e., $G_2=fG_1f^{-1}$. Let T_i and K_i be the lattice group and the point group of G_i $(i=1,2)$, respectively. Then $T_2=hT_1$ and $K_2=hK_1h^{-1}$ for the linear transformation h induced by f. Moreover, there is an orthogonal transformation k such that $K_2=kK_1k^{-1}$; this k can naturally be considered as an element of $GL(n, \mathbf{Z})$. G_1 and G_2 are equivalent $\Leftrightarrow$ G_1 and G_2 are isomorphic (as abstract groups) [1]. In applications it is often required that in the definition of equivalence f be an †orientation-preserving affine transformation. Under this definition, there can be a pair G_1, G_2 such that G_1 and G_2 are mutually isomorphic but not equivalent, since they are conjugate only by means of an †orientation-reversing affine transformation. In this case, they are called **enantiomorphous** to each other, or the pair is called an **enantiomorphic** pair (Table 1 below). For a given dimension n, there is only a finite number of equivalence classes of space groups [1]. We also conclude from this that point groups have finitely many equivalence classes up to conjugacy in the orthogonal group $O(V)$ or in $GL(n, \mathbf{Z})$.

B. Crystal Classes, Bravais Types

Let T be an n-dimensional lattice in n-dimensional Euclidean space V, and K a finite subgroup of the orthogonal group $O(V)$. Denote by a pair (T, K) a †faithful linear †representation of K on T, i.e., a †monomorphism of groups $K \to \mathrm{Aut}(T)$. A space group G determines a pair $(T, G/T)$. Conversely, for any pair (T, K) the following holds: "Any group †extension of K over the †kernel T is isomorphic to a space group" (→ the algebraic characterizations in Sections A and C). Thus to each pair (T, K) there corresponds a set of space groups.

Two pairs (T_1, K_1) and (T_2, K_2) are called **arithmetically equivalent** (this is denoted by $(T_1, K_1) \approx (T_2, K_2)$) if there exists an invertible linear transformation $g \in GL(V)$ such that $T_2 = gT_1$, $K_2 = gK_1g^{-1}$. These pairs are called **geometrically equivalent** or simply **equivalent** (this is denoted by $(T_1, K_1) \sim (T_2, K_2)$) if there exists a $g \in GL(V)$ such that $K_2 = gK_1g^{-1}$. The relation $\approx (\sim)$ is an equivalence relation, and equivalence classes are called **arithmetic (geometric) crystal classes**. The set of geometric crystal classes is in one-to-one correspondence with the set of equivalence classes of point groups up to conjugacy, and is sometimes called the set of **crystal classes**. The geometric crystal class of (T, K) is usually denoted by K.

Now let T be a lattice. Then the group of all orthogonal transformations that leave T invariant is called the **Bravais group** of T and is denoted by $B(T)$, i.e., $B(T) = \{g \in O(V) \mid gT = T\}$. The group $B(T)$ is finite and determines a pair $(T, B(T))$. Two lattices T_1 and T_2 are called **arithmetically equivalent** (this is denoted by $T_1 \approx T_2$) if $(T_1, B(T_1))$ and $(T_2, B(T_2))$ are arithmetically equivalent. An equivalence class is called a **Bravais type**, and an arithmetic crystal class determined by $(T, B(T))$ is called a **Bravais class**. For each Bravais type, a representative is said to be its **Bravais lattice** (Fig. 3 below).

Define an †order relation on the set of lattices belonging to an arithmetic crystal class as follows. When $(T_1, K_1) \approx (T_2, K_2)$, define $T_2 \leqslant T_1$ if there exists a $g \in GL(V)$ such that $T_2 = gT_1$, $K_2 = gK_1g^{-1}$, and $B(T_2) \subseteq gB(T_1)g^{-1}$. Note that $T_1 \approx T_2 \Leftrightarrow T_1 \leqslant T_2$ and $T_2 \leqslant T_1$. In each arithmetic crystal class, there is a pair (T, K) with a †minimal T with respect to this order relation. In this case the Bravais type of T is referred to as the **Bravais type** of the class of (T, K). This lattice T is, intuitively, of the most general type appearing in the class. Now define a relation $T_1 \sim_f T_2$ if there are pairs (T_i, K_i), $i = 1, 2$, such that $(T_1, K_1) \sim (T_2, K_2)$ and T_i is minimal in the class of (T_i, K_i). The equivalence class of the equivalence relation generated by the relation $\sim_f$ is called a **crystal family**.

Summarizing, we have defined the following. Let $\mathscr{S}, \mathscr{A}, \mathscr{G}, \mathscr{B}, \mathscr{CF}$ be the set of equivalence classes of space groups, arithmetic crystal classes, geometric crystal classes, Bravais types, and crystal families, respectively. Then we have the relations shown in Fig. 1. (All arrows in Fig. 1 are surjective.) $\mathscr{CF}$ is nothing but the †coproduct of $\mathscr{G}$ and $\mathscr{B}$ over $\mathscr{A}$, $\mathscr{CF} \simeq \mathscr{G} \coprod_A \mathscr{B}$.

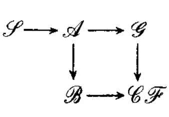


Fig. 1

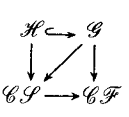


Fig. 2

Moreover, there are injective mappings $\mathscr{B} \hookrightarrow \mathscr{A}$ and $\mathscr{A} \hookrightarrow \mathscr{S}$ such that the composites $\mathscr{B} \hookrightarrow \mathscr{A} \to \mathscr{B}$ and $\mathscr{A} \hookrightarrow \mathscr{S} \to \mathscr{A}$ are identities. The first is the mapping sending a Bravais type T to the arithmetic crystal class defined by $(T, B(T))$. The second one sends a class of (T, K) to a †semidirect product of T and K (→ 190 Groups N). A space group belonging to a class of the image of $\mathscr{A} \hookrightarrow \mathscr{S}$ is said to be **symmorphic** or **symmorphous**. A class which belongs to the image $\mathscr{H}$ of the composite $\mathscr{B} \hookrightarrow \mathscr{A} \to \mathscr{G}$ is called a **holohedral** or **holosymmetric class**, or a **holohedry**. Define a mapping M of $\mathscr{G}$ to the power set $\mathfrak{P}(\mathscr{B})$ by $M = (\mathscr{B} \leftarrow \mathscr{A}) \cdot (\mathscr{A} \to \mathscr{G})^{-1}$. For two elements C_1 and C_2 of $\mathscr{G}$, define $C_1 \sim_s C_2$ if $M(C_1) = M(C_2)$. Each equivalence class is called a **crystal system**, and the set of crystal systems is denoted by $\mathscr{CS}$. Two elements C and C' of $\mathscr{G}$ belong to the same crystal family if and only if there exists a sequence of elements of $\mathscr{G}$, $C_1, C_2, \ldots, C_k$ such that $C_1 = C$ and $C_k = C'$ and $M(C_i) \cap M(C_{i+1}) \neq \varnothing$ for $i = 1, \ldots, k-1$. Therefore there is a surjective mapping $\mathscr{CS} \to \mathscr{CF}$ (Fig. 2). When $n \leqslant 4$, the composite $\mathscr{H} \hookrightarrow \mathscr{G} \to \mathscr{CS}$ is bijective. It should be remarked that when $n \geqslant 3$, there is no mapping $\mathscr{B} \to \mathscr{CS}$ such that $\mathscr{A} \to \mathscr{G} \to \mathscr{CS}$ and $\mathscr{A} \to \mathscr{B} \to \mathscr{CS}$ coincide.

The numbers of elements of these sets are shown in Table 1.

Table 1

n =	1	2	3	4
$\mathscr{CF}$, crystal families	1	4	6	23
$\mathscr{CS}$,[a] crystal systems	1	4	7	33
$\mathscr{B}$, Bravais types	1	5	14	64
$\mathscr{G}$, point groups (isomorphic classes)[b]	2 (2)	10 (9)	32 (18)	227 (118)
$\mathscr{A}$, arithmetic crystal classes	2	13	73	710
$\mathscr{S}$, space groups	2	17	219[c]	4783[d]

[a] When $n \leqslant 4$, the number of elements of $\mathscr{H}$ coincides with that of $\mathscr{CS}$.
[b] The number in parentheses denotes the number of equivalence classes under algebraic isomorphism.
[c,d] The number of equivalence classes under orientation-preserving transformations is 230 and 4895, respectively.

C. Construction of Space Groups

Take an element (T, K) of an arithmetic crystal class. Since K is finite and T is †finitely generated, the †cohomology groups $H^i(K, T)$ are finite groups (→ 200 Homological Algebra G). Let α be a †second cocycle representing an element $[\alpha]$ of $H^2(K, T)$. Let $(T, K)_\alpha$ be the †extension of K with kernel T corresponding to $[\alpha]$ (→ 190 Groups N). Then the set of elements of $(T, K)_\alpha$ is given by $\{\langle t, k \rangle \mid$

$t \in T, k \in K\}$, and the product is defined by $\langle t_1, k_1\rangle\langle t_2, k_2\rangle = \langle t_1 + k_1 t_2 + \alpha(k_1, k_2), k_1 k_2\rangle$. Since $H^i(K, V) = 0$, $i \geqslant 1$, there is a †first cochain $\beta \in C^1(K, V)$ such that $\alpha(k_1, k_2) = \beta(k_1) + k_1\beta(k_2) - \beta(k_1 k_2)$, and the †first cocycle $\bar{\beta}$ on V/T defined by β induces an isomorphism $H^1(K, V/T) \cong H^2(K, T)$, $[\bar{\beta}] \mapsto [\alpha]$. Define an action of $\langle t, k\rangle$ on V by

$$\langle t, k\rangle(x) = kx + t + \beta(k), \qquad x \in V.$$

Then $(T, K)_\alpha$ acts on V as a space group. Concerning their equivalence, this result holds: "The equivalence classes of $(T, K)_\alpha$, $[\alpha] \in H^2(K, T)$ are in one-to-one correspondence with the orbits of $N(K)$, the normalizer of K in $\mathrm{Aut}(T)$ acting on $H^2(K, T)$" [4]. The action of $h \in N(K)$ is given by ${}^h\alpha(\underline{k}) = h\alpha(h^{-1}kh)$, $\alpha \in C^2(K, T)$.

For a pair (T, K) the **weight lattice** or the **weight group** T^* of (T, K) is defined by $\{v \in V \mid v - kv \in T, \forall k \in K\}$. Let R be an irreducible reduced †root system, $Q(R)$ its lattice, and $W(R)$ its †Weyl group. Then for a pair $(Q(R), W(R))$, $Q(R)^*$ is nothing but the usual weight lattice $P(R)$. (→ 248 Lie Algebras). Suppose that K of (T, K) does not contain the central inversion $-I_n: v \mapsto -v$, $v \in V$. Denote the group $K \cup (-I_n)K$ by $\pm K$. Then there is an isomorphism

$$H^1(\pm K, V/T) \cong [H^1(K, V/T)]_2 \times (T^*/T)/2(T^*/T).$$

Here $[\cdot]_2$ denotes the subgroup of elements of order not greater than 2 [5]. This isomorphism provides powerful means for the construction of space groups.

In constructing a pair (T, K), it is useful to take a basis of T in a special form. Let $\{t_1, t_2, \ldots, t_n\}$ be a basis of a lattice T, and set $a_{ij} = (t_i, t_j)$, where $(,)$ is the inner product in V. The $n \times n$ matrix $A = (a_{ij})$ is †symmetric and †positive definite, and the numbers a_{ij} are called the **lattice constants** of T. A basis $\{t_1, \ldots, t_n\}$ is called a **reduced basis** if the †quadratic form $(Ax, x) = \sum a_{ij} x_i x_j$ is reduced (→ 348 Quadratic Forms). A lattice admits at least one reduced basis, and in this case $\pm 2a_{ij} \leqslant a_{ii} \leqslant a_{jj}$, $i < j$ [6, Table 5.1; 7]. The totality of an n-dimensional reduced basis forms in a natural manner a semialgebraic set of dimension $(n+2)(n-1)/2$ in $\mathbf{R}^{(n+1)n/2}$. Its closure, say Q_n, is compact. On the other hand, the set $L_n = (\mathbf{R}_+ \times O(\mathbf{R}^n)) \backslash GL(n, \mathbf{R})/GL(n, \mathbf{Z})$ is identified with the set of equivalence classes of n-dimensional lattices under the relation of "same shape." There is a natural inclusion $L_1 \cup \ldots \cup L_n \hookrightarrow Q_n$, and L_n is dense in Q_n. Moreover, the inclusion gives a bijective homeomorphism when $n \leqslant 4$ [8].

A finite subgroup K of the orthogonal group $O(V)$ is said to be **fully transitive** if there is a set $S = \{e_1, \ldots, e_s\}$ that spans V on which K acts †transitively and if K has no †invariant subspace in V. In this case one can choose S as either of the following: (i) the **primitive hypercubic** type, $S = \{e_1, \ldots, e_n\}$, $(e_i, e_j) = \delta_{ij}$ (†Kronecker delta); (ii) the **primitive hyperbolic** type, $S = \{f_1, \ldots, f_{n+1}\}$, $(f_i, f_i) = 1$, $i = 1, \ldots, n+1$, $(f_i, f_j) = -1/n$, $i, j = 1, \ldots, n+1$, $i \neq j$ (especially, $\sum_{i=1}^{n+1} f_i = 0$) [10].

D. Color Symmetry Groups, Twinning

A **color symmetry group**, or a **colored symmetry group**, associated with a space group G is a pair (G, G') such that G' is a subgroup of G with finite index. The index $r = [G : G']$ is called the **number of colors**. We call (G, G') a **white group** when $r = 1$, a **black and white group** or a **magnetic group** when $r = 2$, and a **polychromatic group** when $r \geqslant 3$. G' is also a space group with the lattice group $T' = G' \cap T$, where T is the lattice group of G. Two color symmetry groups (G_1, G_1') and (G_2, G_2') are called **equivalent** if there exists an †affine transformation f such that $G_2 = fG_1f^{-1}$ and $G_2' = fG_1'f^{-1}$. Then (G_1, G_1') and (G_2, G_2') are equivalent if and only if there is an isomorphism $\sigma: G_1 \to G_2$ such that $\sigma(G_1') = G_2'$. A color symmetry group (G, G') is called **lattice equivalent** if $T' = T$, $K' \subset K$, and is called **class equivalent** if $T' \subset T$, $K' = K$. Take a pair (G, G'), and let G'' be the †inverse image of K' under the canonical †epimorphism $G \to K$. Then (G, G'') is lattice equivalent and (G'', G') is class equivalent. Fix the color number r. Then (i) the equivalence classes of lattice equivalent color symmetry groups are in one-to-one correspondence with the conjugacy classes of subgroups K_1 of index r, and (ii) the number of equivalence classes of class equivalent color symmetry groups is finite. When r is prime to the order of K, the classes are in one-to-one correspondence with K-invariant sublattices T_1 of index r in T. Therefore, when the number of colors is given, the set of equivalence classes of n-dimensional color symmetry groups is finite.

In case (i), a pair (K, K_1) is called a **color point group**, and in case (ii), (T, T_1) is called a **color lattice**. In particular, when $r = 2$, they are called a **black and white point group** and a **black and white lattice**, respectively. Their equivalence is defined in a similar way.

Take a color lattice (T, T_1) of color number r. Let m be the least natural number such that $mT \subset T_1$, then m is a divisor of r. Let $m = p_1^{n_1} \ldots p_k^{n_k}$, $p_i \neq p_j$ $(i \neq j)$ be the decomposition of m into its prime factors. Then there is a sequence of K-invariant sublattices of T, T_i, $i = 2, \ldots, k-1$, such that (i) $T_1 \subset T_2 \ldots \subset T_k \subset T_{k+1} = T$; (ii) $p_i^{n_i} T_{i+1} \subset T_i$, $i = 1, \ldots, k$; (iii) the

submodules $T_i/p_i^{n_i}T_{i+1}$ of $i=1,\ldots,k$, $\mathbf{Z}^n\otimes\mathbf{Z}/p_i^{n_i}\mathbf{Z}$ are uniquely determined independently of the choice of such a sequence $\{T_i\}$. Thus the existence of color lattices is related to the †modular representation of K (→ 362 Representations G). For example, if r is a prime number, then the representation of K on $T/rT\simeq\mathbf{Z}^n\otimes\mathbf{Z}/r\mathbf{Z}$ is the reduction modulo r of that of K on T.

A mathematical treatment of twinning is given as follows. Let V be an n-dimensional Euclidean space and T an n-dimensional lattice in V. The volume of T, vol(T), is defined by $\mathrm{vol}(T)=(|\det A|)^{1/2}$, where $A=(a_{ij})$ is the matrix of lattice constants. This definition is independent of the choice of a basis of T. If T' is a sublattice of index r in T, $\mathrm{vol}(T')=r\cdot\mathrm{vol}(T)$ follows. Let K be a finite subgroup of $O(V)$, T_1 and T_2 K-invariant lattices, and U a hyperplane in V. Then the quadruple (T_1,T_2,K,U) is called a **twinning structure** if it satisfies the following: (i) $\mathrm{vol}(T_1)=\mathrm{vol}(T_2)$; (ii) the lattice T_i is minimal in the arithmetic crystal class of (T_i,K), $i=1,2$; and (iii) $T_1\cap U=T_2\cap U=T_0$, and T_0 is an $(n-1)$-dimensional lattice in U. Two arithmetic crystal classes are said to be **twinable** if there exist representatives (T_1,K), (T_2,K), and a hyperplane U such that (T_1,T_2,K,U) is a twinning structure. Let (T_1,K) and (T_2,K) be two arithmetic crystal classes. If there exists a K-invariant hypersurface U such that $T_0=T_1\cap U=T_2\cap U$ is an $(n-1)$-dimensional lattice in U, then two classes (T_1,K) and (T_2,K) are twinable.

E. Three- and Two-Dimensional Crystallography

Three-dimensional crystal classes are listed in Appendix B, Table 5.IV. To name these crystal classes, both Schoenflies' notation and the international notation are used.

Fix an †orthogonal †right-handed system (x,y,z) in V. For each crystal system, define three axes as follows. The first axis is the z-axis. The second is the line $x=y=z$ for the cubic system and the x-axis for the others. The third is the y-axis for orthorhombic, trigonal, and hexagonal systems, and the line $x=y$, $z=0$ for the others. Then a crystal class is expressed by its generators with respect to these axes. The symbol k $(k=1,2,3,4,6)$ denotes the rotation by the angle $2\pi/k$ around the axis; $\bar{k}$ is a composition of the rotation k and the central inversion $(v\mapsto -v, v\in V)$, and k/m $(k=2,4,6)$ means the composition of k and the reflection about the hyperplane perpendicular to the above axis. The symbol m denotes $\bar{2}$. Usually, 1 is omitted. This is the **full international notation** and a set of generators can be read from it. The **short international notation** is also used.

Schoenflies' notation consists of the letters C_n, D_n, S_n, T, O, and the subscripts h, v, d. The letters denote the group and order: C_n is the †cyclic group of order n; D_n is the dihedral group of order $2n$; T is the †tetrahedral group isomorphic to the fourth †alternating group; O is the †octahedral group isomorphic to the fourth †symmetric group; and $\mathrm{S}_2=\bar{1}$, $\mathrm{S}_4=\bar{4}$, $\mathrm{S}_6=\bar{3}$. The subscripts h, v, d mean that the group is generated by reflections in the hyperplane which is horizontal, vertical, or perpendicular to a diagonal, respectively. The orientation is appropriately given. For example, in $\mathrm{C}_{n\mathrm{h}}$ and $\mathrm{D}_{n\mathrm{h}}$ there is an nth order rotation around the z-axis and the horizontal plane is the xy-plane.

The above notations are also applied to noncrystallographic finite subgroups of the 3-dimensional orthogonal group [13].

The symbol for Bravais type is specified by the name of its crystal system and one of the letters P, A, B, C, F, I. Six kinds of **simple Bravais lattices** are given as follows. A simple lattice is generated by $\{\mathbf{a},\mathbf{b},\mathbf{c}\}$ and satisfies the following condition. Denote the lengths of vectors by $a=\|\mathbf{a}\|$, $b=\|\mathbf{b}\|$, $c=\|\mathbf{c}\|$, and let $\alpha=\angle(\mathbf{b},\mathbf{c})$, $\beta=\angle(\mathbf{c},\mathbf{a})$, $\gamma=\angle(\mathbf{a},\mathbf{b})$ be the angles between the indicated pairs of vectors. For **triclinic P**, $a\neq b\neq c\neq a$, $\alpha\neq\beta\neq\gamma\neq\alpha$; for **monoclinic P**, $a\neq b\neq c\neq a$, $\alpha=\gamma=90°\neq\beta$; for **orthorhombic P**, $a\neq b\neq c\neq a$, $\alpha=\beta=\gamma=90°$; for **tetragonal P**, $a=b\neq c$, $\alpha=\beta=\gamma=90°$; for **hexagonal P** or **trigonal P**, $a=b\neq c$, $\alpha=\beta=90°$, $\gamma=120°$; for **cubic P**, $a=b=c$, $\alpha=\beta=\gamma=90°$. Let T be one of the above lattices generated by $\{\mathbf{a},\mathbf{b},\mathbf{c}\}$. Then the lattices generated by $\{\mathbf{a},\mathbf{b},(\mathbf{b}+\mathbf{c})/2\}$, $\{(\mathbf{c}+\mathbf{a})/2,\mathbf{b},\mathbf{c}\}$, $\{\mathbf{a},(\mathbf{a}+\mathbf{b})/2,\mathbf{c}\}$ $\{(\mathbf{a}+\mathbf{b}+\mathbf{c})/2,\mathbf{b},\mathbf{c}\}$, $\{(\mathbf{a}+\mathbf{b})/2,(\mathbf{b}+\mathbf{c})/2,(\mathbf{c}+\mathbf{a})/2\}$ are called **A, B, C, I, F lattices** determined by T. The original lattice T is called the **primitive lattice**. The **trigonal R** lattice is generated by $\{\mathbf{a}',\mathbf{b}',\mathbf{c}'\}$ satisfying $a'=b'=c'$, $\alpha'=\beta'=\gamma'<120°$, $\neq 90°$. Set $\mathbf{a}=\mathbf{a}'-\mathbf{b}'$, $\mathbf{b}=\mathbf{b}'-\mathbf{c}'$, $\mathbf{c}=\mathbf{a}'+\mathbf{b}'+\mathbf{c}'$. Then the hexagonal P lattice generated by $\{\mathbf{a},\mathbf{b},\mathbf{c}\}$ is defined to be the primitive lattice of the trigonal R lattice. The fourteen Bravais lattices are illustrated in Fig. 3.

An arithmetic crystal class (T,K) is denoted by the symbol of the Bravais lattice T and the symbol of K with respect to the action of K on T. Let the primitive lattice of T be generated by $\{\mathbf{a},\mathbf{b},\mathbf{c}\}$. Put the vector $\mathbf{a}$ on the x-axis, and $\mathbf{b}$ on the xy-plane. The action of K is then determined by the international notation for K, except when $K=\bar{4}2m$, 32, $3m$, $\bar{3}m$, $\bar{6}m2$, and T is the primitive or I lattice. In these cases two kinds of actions interchanging the second and the third symbols are not equivalent. If such is the case, 1 is inserted into the third position for trigonal systems. Let (K) be

the symbol determined as above and (T) be the symbol of T. Then the action (T, K) is expressed by $((T), (K))$. For example, when $K = 32$, there are three arithmetic crystal classes $(P, 321)$, $(P, 312)$ $(R, 32)$.

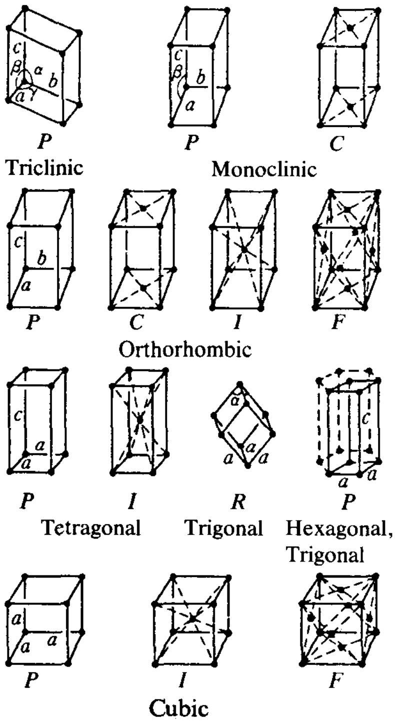


Fig. 3
The conventional unit cell for each of the 14 Bravais lattices.

The notation of space groups determined by (T, K) is given by writing both the symbol (T), and the symbols of actions of elements on V which appear in the symbol (K) defined above. The symbol $(T)(K)$ denotes the †semidirect product of T and K by the action (T, K), and is the symmorphic space group determined by (T, K). Other nonsymmorphic groups are designated by replacing symbols in (K). If k is replaced by k_j $(j = 1, \ldots, k-1)$, then k_j is a **k-fold screw glide with pitch** j/k, that is, the composite of the rotation k and the translation $j\mathbf{f}/k$, where $\mathbf{f}$ is the vector of minimum length of T along the axis of k. If m is replaced by a, b, c, n, d, then it is a **glide reflection**, that is, the composite of the reflection m and the translation in the reflecting hyperplane with the direction of **a**, **b**, **c**, a face diagonal and a diamond, respectively. See [6] for the precise meaning of the notation discussed in this paragraph.

The notations for 2-dimensional point groups and space groups can be adopted from the ones for 3-dimensional ones by setting the reference plane as the xy-plane and the z-axis perpendicular to that plane. In this setting, the notation of a plane point group is given by the corresponding point group in 3-space, and similarly for plane space groups. There is only one **glide operation**, denoted by g. The motion g is the composite of the line reflection about an axis and half the translation along the axis with the vector of minimum length of T. In [13], the precise group-theoretic description of space groups can be found. It should be noted that the notations for the two groups $p3m1$ and $p31m$ are frequently interchanged in the literature (including earlier editions of [13]).

F. History

The first mathematical study of the structure of crystals was done by the mineralogist J. F. C. Hessel (1830). He enumerated the finite subgroups of the 3-dimensional orthogonal group. Afterward, his result was rediscovered many times. A. Bravais (1850) showed that in 3-dimensional Euclidean space there are only 14 different lattices. All space groups consisting of orientation-preserving transformations were determined by C. Jordan (1868, 1869) and L. Sohnke (1879). Finally, almost all the space groups were determined independently by E. S. von Fedorov (1885, 1889) and A. Schoenflies (1887, 1889). Probably as a result of comparison of each other's lists, they established in 1891 the existence of the 230 space groups. Then A. Barlow (1894, 1896) derived these space groups by adding reflection operations to Sohnke's 65 groups.

In 1900 in his eighteenth problem D. Hilbert raised the question of whether the number of equivalence classes of space groups of a given dimension is finite. This was answered affirmatively by L. Bieberbach (1910 [1]). An algorithm for determining space groups was given by H. Zassenhaus (1948 [4]). Their results, including [2, 3], and work by C. Hermann (1949 [9], 1952 [10]) gave a solid foundation for n-dimensional crystallography.

Then followed a concrete treatment of 4-dimensional crystallography. A. C. Hurley (1951), using an earlier work of M. E. Goursat (1889) on finite subgroups of the 4-dimensional special orthogonal group, determined 221 of the 4-dimensional point groups. In the 1960s A. L. Mackay and G. S. Pawley (1963) and others gave 56 of the 4-dimensional Bravais types. After E. C. Dade (1965) determined 9 maximal finite subgroups of $GL(4, \mathbf{Z})$, R. Bülow (1967) and H. Brown, J. Neubüser, and H. Zassenhaus (1968) independently determined 710 arithmetic crystal classes. They reconfirmed the result with H. Wondratschek, and as a result, 64 Bravais types and 227 geometric crystal classes were established (1971 [15]). They ran Brown's computer

program for the determination of space groups based on [4] and obtained 4,783 space groups. Their work culminated in a book (1978 [16]). There have been several attempts to establish a unified treatment of n-dimensional crystallography [5, 8, 14, 17] and crystallography in complex Euclidean spaces.

Two-dimensional space groups were empirically known in ancient times, as demonstrated by artistic decorations. All the 17 2-dimensional space groups appear in the tile patterns of the Alhambra in Granada. A discrete subgroup of the plane motion group whose translation subgroup is only rank one is called a **frieze group**.

The notion of black and white groups was introduced by H. Heesch (1930), H. J. Woods (1935), and A. V. Shubnikov (1951). Only with the introduction of the use of neutron diffraction techniques did it become apparent that these groups could be used in the description of magnetically ordered structures [12]. Color symmetry groups were defined by B. L. van der Waerden and J. J. Bruckhardt (1961 [11]) for an arbitrary number of colors. A mathematical treatment of the twinning structure is found in papers by T. Ito (1938) and R. Sadanaga (1959).

References

[1] L. Bieberbach, Über die Bewegungsgruppen der Euklidischen Räume, Nachr. Akad. Wiss. Göttingen (1910), 75–84; Math. Ann. 70 (1910), 297–336 and 72 (1912), 400–412.
[2] L. Bieberbach, Über die Minkowskische Reduktion der positiven quadratischen Formen und die endlichen Gruppen linear ganzzahliger Substitutionen, Nachr. Akad. Wiss. Göttingen (1912), 207–216.
[3] H. Zassenhaus, Neuer Beweis der Endlichkeit der Klassenzahl bei unimidularer Äquivalenz endlicher ganzzahliger Substitutionsgruppen, Abh. Math. Sem. Univ. Hamburg, 12 (1938), 278–288.
[4] H. Zassenhaus, Über einen Algorithmus zur Bestimmung der Raumgruppen, Comment. Math. Helv., 21 (1948), 117–141.
[5] G. Maxwell (and N. Broderick), The crystallography of Coxeter groups, J. Algebra, 35 (1975), 159–177 and 44 (1977), 290–318.
[6] N. F. M. Henry and K. Lonsdale, International tables for X-ray crystallography, vol. 1, Symmetry groups, Kynoch, Birmingham, 1952, 1965, 1969.
[7] B. L. van der Waerden, Die Reduktionstheorie der positiven quadratischen Formen, Acta Math., 96 (1956), 265–293.
[8] R. L. E. Schwarzenberger, Classification of crystal lattices, Proc. Cambridge Philos. Soc., 72 (1972), 325–349.
[9] C. Hermann, Kristallographie in Räumen beliebiger Dimensionzahl I, Die Symmetrieoperationen, Acta Cryst., 2 (1949), 139–145.
[10] C. Hermann, Translationsgruppen in n-Dimensionen, Zur Struktur und Materie der Festkörper, Springer, 1952.
[11] B. L. van der Waerden and J. J. Burckhardt, Farbgruppen, Z. Krist., 115 (1961), 231–234.
[12] A. V. Shubnikov and N. V. Belov, Colored symmetry, Pergamon, 1964.
[13] H. S. M. Coxeter and W. O. Moser, Generators and relations for discrete groups, Springer, 1957, 1964, 1972, 1979.
[14] E. Ascher and A. Janner, Algebraic aspects of crystallography, Helv. Phys. Acta, 38 (1965), 551–571, and Comm. Math. Phys., 11 (1968), 138–167.
[15] R. Bülow, J. Neubüser, and H. Wondratschek, On crystallography in higher dimensions, Acta Cryst., A27 (1971), 517–535.
[16] H. Brown, R. Bülow, J. Neubüser, H. Wondratschek, and H. Zassenhaus, Crystallographic groups of 4-dimensional space, Wiley, 1978.
[17] R. L. E. Schwarzenberger, N-dimensional crystallography, Research notes in math., Pitman, 1980.
[18] A. Kelly and G. W. Groves, Crystallography and crystal defects, Longman, 1970.
[19] M. J. Buerger, Introduction to crystal geometry, McGraw-Hill, 1971.
[20] G. J. Bradley and A. P. Cracknell, The mathematical theory of symmetry in solids, Clarendon Press, 1972.
[21] C. W. Curtis and I. Reiner, Representation theory of finite groups and associative algebras, Wiley, 1962.

93 (VI.20) Curves

A. Introduction

In the beginning of his *Elements*, Euclid gave definitions such as: A line is a length having no width; an end of a line is a point. However, he left notions such as width and length undefined. Thus his definitions were far from satisfactory. Actually, it was only during the latter half of the 19th century that efforts were made to obtain exact definitions of lines and curves. Euclid, among others, distinguished two kinds of curves: **straight lines** and **curves**. Nowadays, however, **lines** in the sense of Euclid are called curves, and a straight line is considered a curve. A first effort to give an exact definition

of a curve using analytic methods was made by C. Jordan in his *Cours d'analyse* I (1893).

B. Jordan Arcs and Jordan Curves

Following Jordan, we define a **continuous plane curve** C to be the image of a †continuous mapping sending the interval $[0,1]$ into the Euclidean plane E^2. Namely, C is the set of points (x, y) in E^2 such that

$$x=f(t), \quad y=g(t), \quad 0\leqslant t\leqslant 1,$$

with continuous functions f, g defined on $[0,1]$. A continuous curve is also called a **continuous arc**. We call $(f(0), g(0))$ and $(f(1), g(1))$ the **ends** of the arc. Given continuous functions f, g defined on $(0,1)$, the set $\{(x,y) \mid x=f(t), y=g(t), 0<t<1\}$ is called an **open arc**. More generally, the image of a continuous mapping of $[0,1], (0,1), [0,1)$, or $(0,1]$ is called an **arc** (or **curve**). Suppose that C is an arc that is the image of an interval I and $P=(x,y)$ is a point on C to which there correspond two elements t_1, t_2 $(t_1<t_2)$ of I such that P is the image of both t_1, t_2. In this case, the point P is called a **multiple point** on C. An arc having no multiple point is called a **simple arc** or **Jordan arc**.

An arc with one and only one multiple point $P=(f(0), g(0))=(f(1), g(1))$ is called a **Jordan curve** or **simple closed curve** (→ Section K). A Jordan curve can be regarded as a topological image in a plane of a circle. Let C be a curve that is the image $\varphi(I)$ of an interval. Then C is said to be **of class** C^k **(analytic)** if the mapping φ is of †class C^k (†analytic). In general, if S is a topological space, then the image $\varphi(I)$ in S of an interval is called a **curve** in S. In particular, if S has the structure of a differentiable (analytic) manifold, we can define the notion of **curve of class** C^k **(analytic curve)** in S.

C. Ordinary Curves

A †connected subset of E^2 that is the union of a finite number of simple arcs meeting at a finite number of points is called an **ordinary curve**. An ordinary curve is called a tree if it does not contain a subset that is homeomorphic to a Jordan curve. Let p be a point on an ordinary curve C. The †boundary of a sufficiently small †neighborhood of p meets C at a finite number of points, and this number is independent of the choice of the small neighborhood. We call it the **order** of p in C. A point of order 1 is called an **endpoint** of C, a point of order 2 an **ordinary point**, and a point of order $\geqslant 3$ a **branch point**. If we can represent an ordinary curve C as a continuous curve tracing each simple arc of C just once, we say that C is **unicursal** (Fig. 1). A necessary and sufficient condition for C to be unicursal is that the number of points of odd orders in C be less than or equal to 2 (L. Euler).

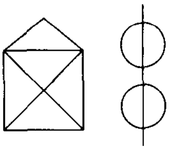


Fig. 1

D. Further Consideration of Definitions

Although the set of ordinary curves as defined above contains most familiar curves, it does not contain a point set defined by $y=\sin 1/x$ in $0<x\leqslant 1$ and $-1\leqslant x<0$, and by $-1\leqslant y\leqslant 1$ at $x=0$ (Fig. 2). This point set is called a **sinusoid**, and it is desirable to obtain a definition of curves wide enough to contain a sinusoid. On the other hand, the notion of continuous curves is, in a sense, too wide, because a curve such as a Peano curve (→ Section J), which covers a whose square, is among such curves. The notion of simple arcs is too narrow, because even a circle is not a simple arc. As a point set in E^2, a continuous arc is characterized as a †locally connected †continuum and is sometimes called a **Peano continuum** (H. Hahn, S. Mazurkiewicz). On the other hand, A. Schoenflies, inspired by the statement of the Jordan curve theorem (→ Section K), considered a closed set that divides the plane into two parts, forming the common boundary of both domains, and called it a closed curve. According to this definition, however, a simple curve is not a closed curve. Thus as a general definition of curves it is not appropriate.

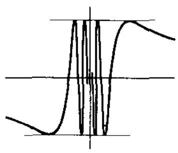


Fig. 2

To give a general notion of curves on a plane (containing sinusoids), we may define a curve as a continuum that is †nowhere dense in E^2 (i.e., a continuum that is a boundary of open sets on the plane) (G. Cantor). Furthermore, to deal with the curves on a topological space, P. S. Uryson and K. Menger defined a **general curve** to be a 1-dimensional continuum (Menger, 1921–1922 [1]). In E^2, the latter notion coincides with the notion of curves defined by Cantor, while in E^3 a general curve

is a continuum that does not divide any †domain (→ 79 Connectedness).

E. Universal Curve

Consider a 3-dimensional cube I^3 ($I=[0,1]$). Draw two planes parallel to each face so that the two planes parallel to a face trisect the edges of the cube meeting the planes. Thus I^3 is divided into 3^3 cubes. Let M_1 be the closure of the subset of I^3 that is obtained from I^3 by deleting the cube I_1^3 ($I_1=[1/3,2/3]$) and the 6 cubes having common faces with I_1^3. Then M_1 consists of 20 cubes (Fig. 3). We apply to each of the 20 cubes forming M_1 the same operation that we applied to I^3 and denote by M_2 the union of point sets thus obtained (consisting of 20^2 cubes, the length of whose edges equals $1/3^2$). Repeating this process, we obtain a point set M_n consisting of 20^n cubes, the length of whose edges equals $1/3^n$. Thus we obtain the sequence $M_1\supset M_2\supset M_3\supset\ldots$. The set $U=\bigcap_{n=1}^{\infty}M_n$ is a general curve in the sense of Uryson and Menger. Moreover, we can prove that an arbitrary general curve is homeomorphic to a subset of U. Hence we call U the **universal curve**.

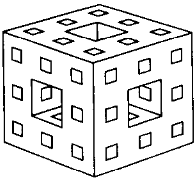


Fig. 3

F. Length of a Curve

In this section, by a curve we mean a continuous curve in a Euclidean space E^n. Let C be a curve in E^n defined by $x_i=f_i(t)$ ($i=1,2,\ldots,n$; $a\leqslant t\leqslant b$; the f_i are real-valued continuous functions defined in $[a,b]$). (We sometimes write this simply as $X=\mathfrak{f}(t)$, where $X=(x_1,\ldots,x_n)$.) We divide $[a,b]$ arbitrarily and denote the dividing points by $a=t_0<t_1<t_2<\ldots<t_r=b$. Let $X_k=\mathfrak{f}(t_k)$, $k=0,\ldots,r$, and let $\overline{X_{k-1}X_k}$ be the length of the straight line segment joining X_{k-1} and X_k. If the length $l=\sum_{k=1}^{r}\overline{X_{k-1}X_k}$ of the broken line $(X_0X_1\ldots X_r)$ (Fig. 4) is bounded for any subdivision of $[a,b]$, C is called a **rectifiable curve**, and the upper limit of l with respect to the subdivisions is called the **length** of C. For C to be rectifiable, it is necessary and sufficient that the f_i ($i=1,2,\ldots,n$) be of †bounded variation (Jordan). Thus if C is a rectifiable curve, then each $f_i(t)$ ($i=1,2,\ldots,n$) is almost everywhere differentiable (H. Lebesgue). In particular, if C is of class C^1, then C is rectifiable, and its length can be represented by

$$s=\int_a^b\left(\sum_{i=1}^{n}\left(\frac{df_i}{dt}\right)^2\right)^{1/2}dt$$

(→ 246 Length and Area).

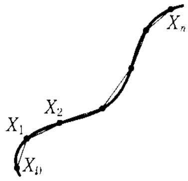


Fig. 4

G. Shapes of Curves

In this section by a curve we mean the image $\varphi(I)$ in a Euclidean space E^n of an interval I (bounded or unbounded), where φ is a continuous mapping. When a curve C of class C^k is given in a Euclidean space E^n, it often becomes necessary to examine the shape of C globally. The determination of the global shape of C from the equation of the curve is called the **curve tracing** of C. The problem has been thoroughly studied in the particular case in which $n=2$ and the equation of C is given by $f(x,y)=0$ in a rectangular coordinate system (by $F(r,\theta)=0$ in a †polar coordinate system), where f (or F) is an analytic function. The problems in the case of a rectangular coordinate system are as follows:

Let φ be a single-valued analytic function and I a (bounded or unbounded) interval on the x-axis. If a curve C_0 represented by $y=\varphi(x)$ ($x\in I$) is a subset of C, then C_0 is called a **branch** of C. According as I is bounded or unbounded, C_0 is said to be a **finite branch** or **infinite branch** of C. When a curve represented by $x=\psi(y)$ ($y\in J$) (ψ is a single-valued analytic function and J is an interval on the y-axis) is a subset of C, it is also called a branch of C. If it is necessary to distinguish these two branches, we call the former the *x-branch* and the latter the *y-branch*. C consists of an at most denumerable set of branches. If $P(x_0,y_0)\in C$ and $f_y=\partial f/\partial y\neq 0$ at P, then there exists an x-branch containing P; if $f_x=\partial f/\partial x\neq 0$, there exists a y-branch of C. If $\partial f/\partial x=0$, $\partial f/\partial y=0$ at P, then P is called a **singular point** of C. Points on C that are not singular points are called **ordinary points** of C.

When P is an ordinary point of C, a branch C_0 of C containing P is determined, and the **tangent line** and **normal line** to C at P are the same as those to C_0 and are uniquely determined. The equations of these are $(x-x_0)f_x(x_0,y_0)+(y-y_0)f_y(x_0,y_0)=0$ and $(x-x_0)f_y(x_0,y_0)-(y-y_0)f_x(x_0,y_0)=0$, respec-

tively. If we choose a coordinate system with P as origin and the tangent line and normal line as ξ-axis and η-axis, respectively, then the equation of C with respect to this coordinate system is of the form $\eta = c_2\xi^2 + c_3\xi^3 + \ldots$ in the neighborhood of P. If we denote by ρ the †curvature of C at P, then $\rho = 2c_2 = -(f_{xx}f_y^2 - 2f_{xy}f_xf_y + f_{yy}f_x^2)/(f_x^2 + f_y^2)^{3/2}$. When $c_2 = \rho = 0$, P is a †stationary point. A stationary point of a curve of class C^2 on C is also called a **point of inflection**. When P is not a point of inflection and (ξ, η) are points on C in a neighborhood V of P, the sign of η is definite if V is small enough (Fig. 5). However, if P is a point of inflection and $c_3 \neq 0$, then C is of the shape shown in Fig. 6. At a point of inflection, if $c_3 = \ldots = c_{\nu-1} = 0$, $c_\nu \neq 0$, and ν is even, C is of the shape shown in Fig. 5, and if ν is odd, C is of the shape shown in Fig. 6.

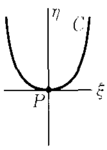


Fig. 5

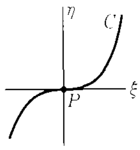


Fig. 6

In a neighborhood of a singular point, C takes various shapes. For example, consider a curve represented by $y^2 = x^2(x+a)$, and let P be the origin $(0,0)$. If $a > 0$, then there are two branches of C passing through P, and they have different tangents at P (Fig. 7). As in this case, if there are a finite number of different branches passing through P with different tangents, P is called a **node** of C. If $a < 0$, then $P \in C$, but there is no other point of C in the neighborhood of P (Fig. 8). Such a point is called an **isolated point** of C. If $a = 0$, then there are two branches of C starting from P, and the tangents to these at P are the same (Fig. 9). Such a point is called a **cusp** of C. When C is an †algebraic curve, we can examine the shape of a curve in a neighborhood of a singular point using the †Puiseux series.

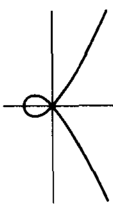


Fig. 7

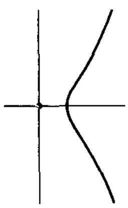


Fig. 8

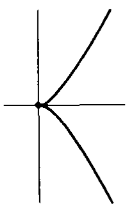


Fig. 9

When C has an infinite branch C_0 (for example, when $C_0: y = \varphi(x)$ $(x \in I)$ is an x-branch of C and $I = [a, \infty)$), if the tangent to C_0 at $P(x_0, y_0)$ $(x_0 \in I)$ has a limiting line for $x_0 \to \infty$, then the limiting line l is called an **asymptote** of C_0. In this case, the distance from a point $P(x_0, y_0)$ of C_0 to l converges to zero when $x_0 \to \infty$. An asymptote of an infinite branch of C is also called an asymptote of C.

H. Special Plane Curves

The following are the well-known curves (for ellipses, parabolas, and hyperbolas → 78 Conic Sections).

Among curves of the third order, those having an equation of the form

$$y^2 = f(x)/(x-a) \qquad (1)$$

($f(x)$ is a rational expression of at most the third order in x) are symmetric with respect to the x-axis and have $x = a$ as an asymptote. In particular, if $a > 0$, $f(x) = -x^3$ in (1), then the curve is as shown in Fig. 10 and the origin is a cusp. Let a half-line starting from the origin meet the curve, the circle with diameter $[0, a]$, and the straight line $x = a$ at points X, Y, and A, respectively. Then we have $OX = YA$. This curve is called a **cissoid of Diocles**.

If $a = 0$, $f(x) = c^2(c-x)$ $(c > 0)$ in (1), then the curve takes the shape shown in Fig. 11. Let A, C be the points whose coordinates are $(a, 0)$, $(c, 0)$ $(0 < a < c)$, respectively, and let X, Y be the points in the first quadrant at which the straight line parallel to the y-axis and passing through A meets the curve and the circle with diameter OC, respectively (Fig. 11). Then we have $AX : AY = OC : OA$. This curve is called a **witch of Agnesi**.

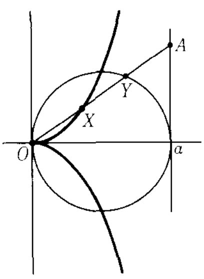


Fig. 10

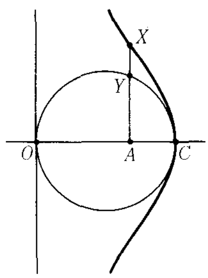


Fig. 11

If $a < 0$, $f(x) = -x^2(x/3 + a)$ in (1), then the curve takes the shape shown in Fig. 12a. If we rotate it by $\pi/4$ and put it in the position shown in Fig. 12b, then the equation of the curve takes the form $x^3 + y^3 = 3cxy$ $(c = -\sqrt{2}$

a). If we take as parameter $t=y/x$, then we get the parametric representation $x=3ct/(1+t^3)$, $y=3ct^2/(1+t^3)$. This curve is called a **folium cartesii** (or **folium of Descartes**). A curve that has a parametric representation of the form $x=\varphi(t)$, $y=\psi(t)$, where φ, ψ are rational functions, is called an (algebraic) **unicursal curve** (or **rational curve**). Such a curve is an algebraic curve of †genus 0.

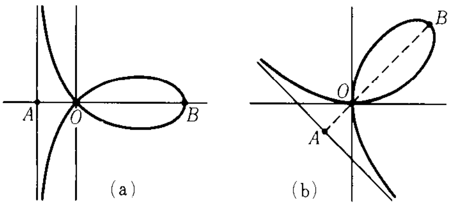


Fig. 12

Let $r=f_1(\theta), r=f_2(\theta), \ldots, r=f_k(\theta)$ be equations of curves $C_1, C_2, \ldots, C_k$ with respect to a polar coordinate system with origin O. A curve C having equation $r=\lambda_1 f_1(\theta)+\ldots+\lambda_k f_k(\theta)$ (the λ_i are constants, usually $+1$ or -1) in the same coordinate system is called a **cissoidal curve** with respect to O. (Fig. 13: $r=-f_1(\theta)+f_2(\theta)$). In Fig. 10, let C_1 be the circumference of the circle with the diameter $[0,a]$, let C_2 be the straight line $x=a$, and put $\lambda_1=-1$, $\lambda_2=1$. Then we have a cissoid of Diocles. We can regard the folium cartesii as a cissoidal curve obtained from a straight line and an ellipse. When C_1 is a circle with center at O, we call C a **conchoidal curve** of C_2 with respect to O. In particular, when C_2 is a straight line and O is not on C_2, the conchoidal curve is called a **conchoid of Nicomedes**.

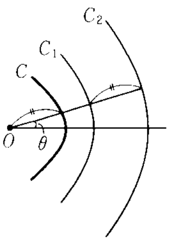


Fig. 13

As shown in Fig. 14, when C_2 is perpendicular to the initial line of the polar coordinate system, the equation of the conchoid is $r=a\sec\theta\pm b$ (b is the radius of C_1), and the Cartesian equation of the curve is $(x-a)^2(x^2+y^2)=b^2x^2$. According as $a>b$, $a=b$, or $a<b$, the curve has a node, cusp, or isolated point, respectively. When C_2 is a circle and O is on C_2, the conchoidal curve of C_2 with respect to O is called a **limaçon** (or **limaçon of Pascal**) (Fig. 15). The equation of a limaçon C with respect to a polar coordinate system having the diameter of a circle passing through O as its initial line is $r=a\cos\theta\pm b$, while the equation of C with respect to a Cartesian coordinate system is $(x^2+y^2-ax)^2=b^2(x^2+y^2)$. In this case, if $a>b$, O is a node of the curve; if $a=b$, O is a cusp. When $a=b$, the curve is called a **cardioid** (the curve shown by a dashed line in Fig. 15; see also Fig. 26 below).

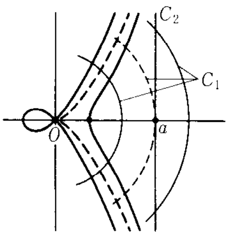


Fig. 14

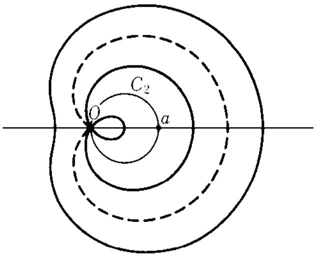


Fig. 15

The locus of a point X having a constant product of its distances from two fixed points A, B is called **Cassini's oval** (Fig. 16). The equation of this curve with respect to the Cartesian coordinate system whose origin O is the midpoint of the segment AB and whose x-axis is the straight line AB is $(x^2+y^2)^2-2a^2(x^2-y^2)=k^4-a^4$ (where $AB=2a$, $k^2=AX\cdot BX$). In particular, if $a^2=k^2$, then O is a nodal point of the curve. In this case, the curve (shown by the dashed line in Fig. 16) is called a **lemniscate** (or **Bernoulli's lemniscate**) (Jakob Bernoulli).

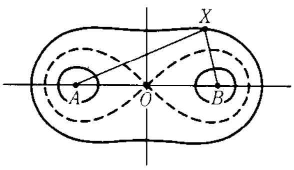


Fig. 16

The locus of the foot of the perpendicular drawn from a fixed point O to the tangent of a fixed curve C at each point of the curve is called the **pedal curve** of C with respect to O. The pedal curve of a rectangular hyperbola with respect to its center is a lemniscate (Fig. 17), and the pedal curve of a circle with respect to a point is a limaçon.

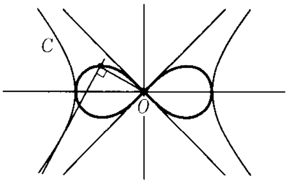


Fig. 17

When a curve C' rolls on a fixed curve C without slipping and is always tangent to C, the locus Γ of a point X kept fixed with respect to the curve C' is called a **roulette** whose **base** is C, **rolling curve** is C', and **pole** is X. In particular, when C is a straight line, C' is a circle, and X is on C', Γ is called a **cycloid** (Fig. 18). When X is not on C', Γ is called a **trochoid** (Fig. 19). A trochoid is represented parametrically by the equations $x=a\theta-b\sin\theta$, $y=a-b\cos\theta$, where the parameter θ is the angle of rotation of C'. When $a=b$, the equation represents a cycloid. The †evolute and †involute of a cycloid are also cycloids (Fig. 20).

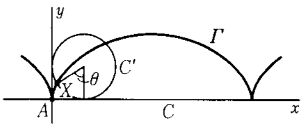


Fig. 18

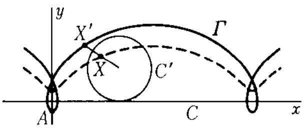


Fig. 19

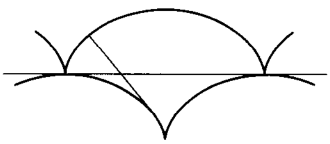


Fig. 20

Suppose that we are in a gravitational field with a given path represented as a cycloid, as is shown in Fig. 21. Assuming that there is no friction, the time necessary for a particle to slide down the path from a point X on the curve to the lowest point C of the curve is independent of the initial position X (C. Huygens). Because of this property, the cycloid is also called a **tautochrone**. Suppose that a particle starts from a point A in the space and slides down to a lower point B along a curve Γ (without friction) under the effect of a gravitational force. To minimize the elapsed time, we simply take Γ as a cycloid that lies in the vertical plane containing AB and has a horizontal line through A as the base (Fig. 22). Because of this property, the cycloid is called also a **brachistochrone**, i.e., the **line of swiftest descent** (Johann Bernoulli and others).

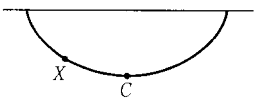


Fig. 21

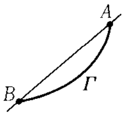


Fig. 22

When the base curve C and the rolling curve C' are both circles and X is on C', we call Γ an **epicycloid** if C, C' are externally tangent (Fig. 23), and a **hypocycloid** if C, C' are internally tangent (Fig. 24). When X is not on C', corresponding to these two cases, we have an **epitrochoid** and a **hypotrochoid**, respectively. Let a, b be radii of C, C', respectively, c the distance from the center of C' to X, and θ the angle of rotation of C'. Then the parametric equations of these curves are $x=(a\pm b)\cos\theta\mp c\cos((a\pm b)/b)\theta$, $y=(a\pm b)\sin\theta-c\sin((a\pm b)/b)\theta$. (Take the upper signs when the curve is an epicycloid and the lower signs when the curve is a hypotrochoid. When $b=c$ the equations are equations of epicycloids and hypocycloids.) When the ratio $a:b$ is a rational number p/q (p, q are mutually prime), then C' returns to its initial position after rotating q times around C; in this case each Γ becomes

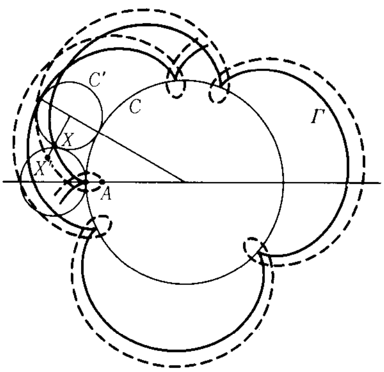


Fig. 23

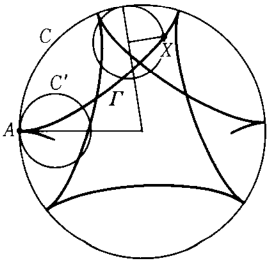


Fig. 24

an algebraic curve. In particular, when $a=4b=4c$, the hypocycloid is called an **astroid** (Fig. 25). Its equation (with respect to a Cartesian coordinate system) is $x^{2/3}+y^{2/3}=a^{2/3}$. The envelope of line segments of length a whose endpoints are on the x-axis and y-axis, respectively, is an astroid (Fig. 25). When $a=b=c$ an epicycloid is a **cardioid** (Fig. 26).

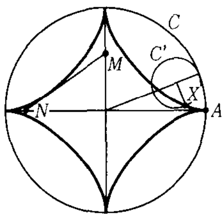


Fig. 25

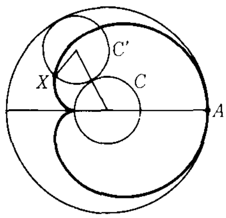


Fig. 26

When the base C is a straight line, the rolling curve C' is an ellipse or a hyperbola, and the pole is a focus of C', then the roulette is called a **Delaunay curve** (Fig. 27).

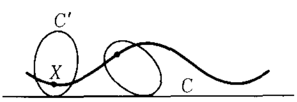


Fig. 27

When C is a straight line, C' is a parabola, and X is the focus of C', the roulette is called a **catenary** (Fig. 28). When we hold two ends of a string of homogeneous density in the gravitational field, the string takes the form of this curve. The equation of the catenary with respect to a Cartesian coordinate system is $y=a\cosh x/a=a(e^{x/a}+e^{-x/a})/2$. The involute starting at the point $A(0,a)$ of this curve is called a **tractrix** (Fig. 29). Let Q be the point of intersection of the tangent at P to the tractrix and the x-axis; then the length of PQ is constant and is equal to a. Consequently, when we drag a weight at A by a string of length OA along the x-axis, the curve described by the weight is the tractrix. The parametric equations of the tractrix are $x=a(\log\tan t/2+\cos t)$, $y=a\sin t$.

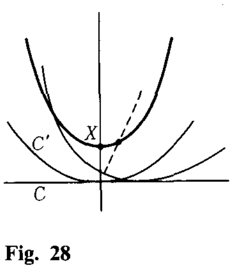


Fig. 28

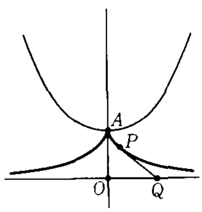


Fig. 29

Suppose that a point Q moves with constant velocity on the x-axis and another point P also moves with constant velocity always toward Q. The locus of the point P is called a **curve of pursuit** (Fig. 30). When the velocity of Q is α times that of P, the equation of the curve of pursuit is $2(x-a)=y^{1-\alpha}/c(1-\alpha)-cy^{1+\alpha}/(1+\alpha)$ if $\alpha\neq 1$ and $2(x-a)=(1/c)\log y-cy^2/2$ if $\alpha=1$. We can consider similar problems when Q moves on a general curve instead of on the x-axis.

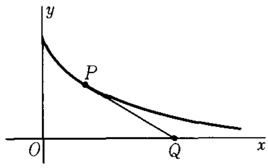


Fig. 30

Many plane curves that are called **spirals** can be expressed by $r=f(\theta)$ (f monotonic) in polar coordinates (r,θ). An **Archimedes spiral** is a curve having the equation $r=a\theta$ (Fig. 31). Archimedes found that the area bounded by two straight lines $\theta=\theta_1$, $\theta=\theta_2$ $(\theta_1<\theta_2)$ and the curve is $a^2(\theta_2^3-\theta_1^3)/6$. A **logarithmic spiral** (**equiangular spiral** or **Bernoulli spiral**) is a curve having the equation $r=ke^{a\theta}$ (Fig. 32). The angle between the straight line $\theta=$constant and the tangent to the curve is constant. Johann Bernoulli found that the involute and evolute of this curve are congruent to the original curve. A curve having the equation $r=a/\theta$ is called a **hyperbolic spiral** (or **reciprocal spiral**), and the one having the equation $r^2\theta=a$ is called a **lituus**. These two spirals are shown in Figs. 33 and 34, respectively. Let $\rho=\varphi(s)$ be the †natural equation of a curve. Properties of the curves for which the functions $\varphi(s)$ are simple have been investigated. Specifically, a curve having the natural equation $\rho=ks$ (k is a constant) is a logarithmic spiral. A curve having the equation $\rho=a^2/s$ is called a **Cornu spiral** (or **clothoid**; → 167 Functions of Con-

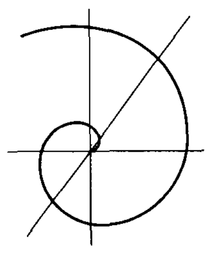


Fig. 31

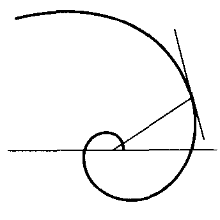


Fig. 32

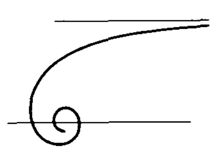


Fig. 33

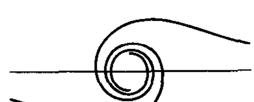


Fig. 34

fluent Type). Its parametric representation is

$$x = a\sqrt{\pi}\int_0^t \cos\frac{\pi t^2}{2}dt,$$

$$y = a\sqrt{\pi}\int_0^t \sin\frac{\pi t^2}{2}dt$$

(†Fresnel integral). M. A. Cornu used this curve in the representation of diffraction in physical optics.

There are also curves that appear as graphs of †elementary functions. For example, a curve having the equation $y=\sin x$ is called the **sine curve**, and the graphs of equations $y=e^x$ and $y=\log x$ are called the **exponential curve** and the **logarithmic curve**, respectively, although they are congruent. In contrast to algebraic curves, these analytic curves that are not algebraic are called **transcendental curves**. (Regarding the differential geometric properties of plane and spaces curves → 111 Differential Geometry of Curves and Surfaces; for plane algebraic curves → 9 Algebraic Curves.)

I. Envelopes

Let $\mathfrak{f}(s,t)$ be a function of class C^1 of real variables s, t. If we fix $t=t_0$, then $\mathfrak{r}=\mathfrak{f}(s,t_0)$ is the equation of a curve C_{t_0} with a parameter s. If $s(t)$ is a function of t, then $\mathfrak{f}(s(t_0),t_0)$ represents a point P_{t_0} on C_{t_0}. Let E be the locus of P_{t_0} when t_0 moves. If $s(t)$ is a function of class C^1, then E is a curve of class C^1. If E and C_{t_0} are always tangent at each point P_{t_0}, we call E the **envelope** of the family of curves C_t. When $\mathfrak{f}(s,t)$ is given, to find E we need only determine the function $s(t)$. We note that $\partial\mathfrak{f}/\partial s=\lambda\partial\mathfrak{f}/\partial t$ is a condition that must be satisfied by the function $s(t)$. When $n=2$ and the equation of C_{t_0} is given in the form $f(x,y,t_0)=0$, the point of intersection of C_{t_0} and $f_t(x,y,t_0)=0$ ($f_t=\partial f/\partial t$) is P_{t_0}. The equation $R(x,y)=0$ obtained by eliminating t from $f(x,y,t)=0$ and $f_t(x,y,t)=0$ is called the **discriminant** of $f(x,y,t)=0$. The set of points (x,y) satisfying the discriminant $R=0$ is the union of E and the locus of the singular points of C_{t_0}.

J. Peano Curve

A †continuous curve in the Euclidean plane E^2 (i.e., the image $f(I)$ of a segment $I=[0,1]$ under a continuous mapping $f: I\rightarrow E^2$) may cover a square. We call such a curve a **Peano curve** after G. Peano, who gave the first example. D. Hilbert simplified the example and constructed a Peano curve as follows (*Math. Ann.*, 36 (1890), 38 (1891)).

Divide a square and a segment into four equal parts (Fig. 35) and let each square D_i correspond to the segment T_i ($i=0,1,2,3$). Then divide each D_i into four equal parts (Fig. 36) and let D_{ij} correspond to T_{ij} ($j=0,1,2,3$), and continue this process (Fig. 37). A sequence of squares $D_i\supset D_{ij}\supset D_{ijk}\supset\ldots$ has the unique common point $p_{ijk\ldots}$, and we let this point correspond to the unique common point $t_{ijk\ldots}$ of the sequence of segments $T_i\supset T_{ij}\supset T_{ijk}\supset\ldots$. The correspondence $t_{ijk\ldots}\rightarrow p_{ijk\ldots}$ is a continuous mapping of the segment $[0,1]$ onto the square D, and this continuous curve has double points, triple points, and quadruple points. The set of multiple points has the cardinal number of the continuum and is a †dense set in the square. We can improve this method so that there are no multiple points other than double and triple points, but it is impossible to eliminate triple points altogether.

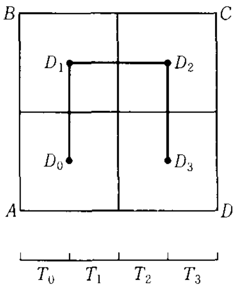


Fig. 35

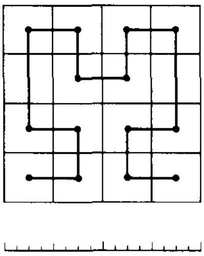


Fig. 36

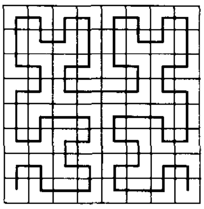


Fig. 37

We can also construct Peano curves as shown in Figs. 38, 39, and 40. That is, we "bisect" a triangle and a segment $[0,1]$ successively and build a correspondence as follows: Let a point represented by a binary number $t=0.ijk\ldots$ ($i,j,k,\ldots=0,1$) correspond to the unique common point of the sequence

$\Delta_i \supset \Delta_{ij} \supset \Delta_{ijk} \supset \ldots$ and obtain a continuous mapping from the segment $[0, 1]$ onto the triangle.

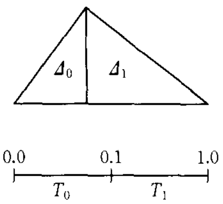


Fig. 38

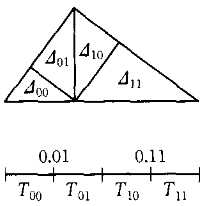


Fig. 39

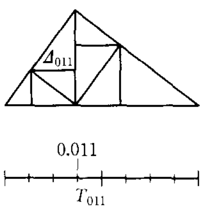


Fig. 40

K. The Jordan Curve Theorem

The **Jordan curve theorem** states: A Jordan curve J in the plane $\mathbf{R}^2$ separates $\mathbf{R}^2$ into inner and outer †regions (C. Jordan, *Cours d'analyse*, 2nd ed., 1893). More precisely, $\mathbf{R}^2 - J$ is the disjoint union $G_1 \cup G_2$ of two regions G_1 and G_2 whose common †boundary is J. Let p be a point of J. Then there is a Jordan arc with p as an endpoint such that all points of the Jordan arc are contained in G_i $(i = 1$ or $2)$ except for p (A. Schönflies), that is, J is **accessible** from G_i. Conversely, let J be a compact subset of $\mathbf{R}^2$ such that $\mathbf{R}^2 - J = G_1 \cup G_2$ and $G_1 \cap G_2 = \varnothing$, where the G_i are regions such that J is accessible from both G_1 and G_2. Then J is a Jordan curve (Schönflies, 1908). A homeomorphism between a Jordan curve J and the circle C extends to a homeomorphism (more precisely to a †conformal mapping) between a plane containing J and a plane containing the circle C (→ 65 Combinatorial Manifolds G; 77 Conformal Mappings).

References

[1] K. Menger, Kurventheorie, Teubner, 1932 (Chelsea, 1967).

[2] R. L. Wilder, Topology of manifolds, Amer. Math. Soc. Colloq. Publ., 1949.

[3] G. T. Whyburn, Analytic topology, Amer. Math. Soc. Colloq. Publ., 1942.

[4] G. Loria, Spezielle algebraische und transzendente ebene Kurven. Theorie und Geschichte I, II, Teubner, 1910, 1911.

[5] F. G. Teixeira, Traité des courbes spéciales remarquables planes et gauches I–III, Coïmbre, 1908–1915.

[6] P. S. Alexandrov and H. Hopf, Topologie I, Springer, 1935.

[7] S. Lefschetz, Introduction to topology, Princeton Univ. Press, 1949.

[8] M. H. A. Newman, Elements of the topology of plane sets of points, Cambridge Univ. Press, second edition, 1951.

[9] M. Greenberg, Lectures on algebraic topology, Benjamin, 1966.

[10] E. E. Moise, Geometric topology in dimensions 2 and 3, Springer, 1977.

94 (X.11) Curvilinear Integrals and Surface Integrals

A. General Remarks

The integral of a function (or more precisely, a †differential form) along a †curve (†surface) is called a **curvilinear integral (surface integral)**. Because a curvilinear integral is a special case of the Stieltjes integral, we shall first explain this notion, formulated by T. J. Stieltjes (1894) as a generalization of the †Riemann integral. The notion was introduced in connection with Stieltjes's study of †continued fractions, and led to the idea of integrals with respect to general measures.

B. The Riemann-Stieltjes Integral

Suppose that $f(x)$, $\alpha(x)$ are real-valued bounded functions defined on $[a, b]$. Take a partition of the interval $a = x_0 < x_1 < x_2 < \ldots < x_{n-1} < x_n = b$ (→ 216 Integral Calculus) and consider the Riemann sum with respect to $\alpha(x)$:

$$\sum_{i=0}^{n-1} f(\xi_i)(\alpha(x_{i+1}) - \alpha(x_i)), \quad \xi_i \in [x_i, x_{i+1}].$$

Suppose that the Riemann sum tends to a fixed number as $\max(x_{i+1} - x_i)$ tends to zero. Then the limit is called the **Riemann-Stieltjes integral** (or simply **Stieltjes integral**) of $f(x)$ with respect to $\alpha(x)$ and is denoted by

$\int_a^b f(x)\,d\alpha(x)$. The Riemann integral of $f(x)$ is a special case, where $\alpha(x)=x$.

The Riemann-Stieltjes integral has the elementary properties, such as linearity, of the usual Riemann integral. We also have the following theorem: The integral $\int f(x)\,d\alpha(x)$ exists for every continuous function $f(x)$ if and only if $\alpha(x)$ is of †bounded variation. Hence when we consider the Stieltjes integral of $f(x)$ with respect to $\alpha(x)$, we usually assume that $f(x)$ is continuous and $\alpha(x)$ is of bounded variation. However, the Stieltjes integral can be defined if $f(x)$ is of bounded variation (not necessarily continuous) and $\alpha(x)$ is continuous (not necessarily of bounded variation). If a sequence $f_n(n)$ $(n=1,2,\dots)$ of uniformly bounded continuous functions defined on the interval $[a,b]$ converges to a continuous function $f(x)$ on the interval $[a,b]$, we have

$$\lim_{n\to\infty}\int_a^b f_n(x)\,d\alpha(x)=\int_a^b f(x)\,d\alpha(x),$$

where $\alpha(x)$ is a function of bounded variation. Furthermore, if $\alpha(x)$ and $\alpha_n(x)$ $(n=1,2,\dots)$ are functions of bounded variation whose †total variations are uniformly bounded and $\lim_{n\to\infty}\alpha_n(x)=\alpha(x)$ at every point of continuity of $\alpha(x)$, then we have

$$\lim_{n\to\infty}\int_a^b f(x)\,d\alpha_n(x)=\int_a^b f(x)\,d\alpha(x)$$

for every continuous function $f(x)$ on $[a,b]$ **(Helly's theorem)**.

Let $\alpha(x)$ be a †strictly monotone increasing continuous function, and let $\beta(y)$ be its inverse function. Then we have

$$\int_a^b f(x)\,d\alpha(x)=\int_{\alpha(a)}^{\alpha(b)} f(\beta(y))\,dy, \qquad (1)$$

where the right-hand side is the usual Riemann integral. A function $\alpha(x)$ of bounded variation is represented as the difference of two strictly monotone increasing functions $\alpha_1(x)$ and $\alpha_2(x)$. If we denote by $\beta_i(y)$ the inverse function of $\alpha_i(x)$ $(i=1,2)$, we have

$$\int_a^b f(x)\,d\alpha(x)=\int_{\alpha_1(a)}^{\alpha_1(b)} f(\beta_1(y))\,dy - \int_{\alpha_2(a)}^{\alpha_2(b)} f(\beta_2(y))\,dy. \qquad (2)$$

It $\alpha'(x)$ exists and is continuous, we have

$$\int_\alpha^b f(x)\,d\alpha(x)=\int_a^b f(x)\alpha'(x)\,dx. \qquad (3)$$

C. The Lebesgue-Stieltjes Integral

Suppose that $\alpha(x)$ is a monotone increasing and right continuous function and $I=(x_1,x_2]$. We define an †interval function $U(I)=\alpha(x_2)-\alpha(x_1)$. It is nonnegative and countably additive. Hence by utilizing $U(I)$ we can construct the outer measure and also a completely additive measure (→ 270 Measure Theory; 380 Set Functions). The Lebesgue integral with respect to this measure is called the **Lebesgue-Stieltjes integral** (or **Lebesgue-Radon integral**) and is denoted by $\int_a^b f(x)\,d\alpha(x)$. If $\alpha(x)$ is a strictly monotone increasing continuous function and $\beta(y)$ its inverse function, then formula (1) is true if the left-hand side is a Lebesgue-Stieltjes integral and the right-hand side is a Lebesgue integral. If $\alpha(x)$ is a function of bounded variation, decomposing $\alpha(x)$ into the difference of two strictly monotone increasing functions, we also have formula (2). If $\alpha(x)$ is †absolutely continuous, formula (3) is valid, where the right-hand side is a Lebesgue integral.

The Stieltjes integral has the following two properties.

Integration by parts: In the interval $[a,b]$, we have

$$\int_a^b U\,dV+\int_a^b V\,dU=U(b)V(b)-U(a)V(a)$$

if one of $U(x)$, $V(x)$ is continuous and the other is of bounded variation.

Second mean value theorem: If $U(x)$ is monotone increasing and $V(x)$ is continuous, then there exists a ξ in $[a,b]$ such that

$$\int_a^b U\,dV$$
$$=U(a)(V(\xi)-V(a))+U(b)(V(b)-V(\xi)).$$

D. The Curvilinear Integral

A continuous mapping from an interval $a\leqslant t\leqslant b$ in $\mathbf{R}^1$ into $\mathbf{R}^n$: $\varphi(t)=(\varphi_1(t),\dots,\varphi_n(t))$ is an oriented curve. Suppose that a function $f(x_1,\dots,x_n)$ is defined in a neighborhood U of the image C of the mapping $\varphi(t)$ or merely on C. The Stieltjes integral

$$\int_a^b f(\varphi_1(t),\dots,\varphi_n(t))\,d\varphi_i(t), \qquad i=1,\dots,n, \qquad (4)$$

is called the **curvilinear integral** of the function $f(x_1,\dots,x_n)$ along the curve C with respect to x_i and is denoted by $\int_C f\,dx_i$. The curve C is called the **contour** (or **path**) **of the integration**, $\varphi(a)$ is called the **initial point** (or **lower end**), and $\varphi(b)$ is called the **terminal point** (or **upper end**) of the integration. Let C be a rectifiable curve, and denote by $s(t)$ the arc length of C from the initial point to the point $\varphi(t)$. Then the Stieltjes integral

$$\int_a^b f(\varphi(t))\,ds(t)$$

or simply $\int_C f ds$ is called the **curvilinear integral with respect to the line element**. Here, the line element, denoted by ds, means $\sqrt{(\varphi_1'(t))^2+\ldots+(\varphi_n'(t))^2}\,dt$ when $\varphi(t)$ is of class C^1. If the integrand in (4) is of bounded variation as a function of t, the curvilinear integral is well defined. If C is a †rectifiable curve, the curvilinear integral is defined for an arbitrary continuous function. In the usual case, we are concerned mainly with this sort of situation. For a differential form $\omega=f_1\,dx_1+\ldots+f_n\,dx_n$ defined on U, the curvilinear integral $\int_C\omega$ is defined by $\sum_{i=1}^n\int_C f_i dx_i$.

The curvilinear integral is linear with respect to its integrand. If the terminal point of C_1 is the initial point of C_2, we can construct the joint curve $C=C_1+C_2$, and we have **additivity for the contours**

$$\int_C f dx_i=\int_{C_1} f dx_i+\int_{C_2} f dx_i$$

(a similar formula holds if we replace dx_i by ds). Monotonicity, which asserts that $\int_C f dx_i\leqslant\int_C g\,dx_i$ whenever $f\leqslant g$, holds if $\varphi_i(t)$ is monotone increasing, and monotonicity also holds for the curvilinear integral with respect to the line element.

If $n=2$, $\mathbf{R}^2$ can be identified with the †complex plane $\mathbf{C}=\{z=x+iy\}$, and we define $\int_C f(z)dz$ by

$$\left\{\int_C u(x)\,dx-\int_C v(z)\,dy\right\}+i\left\{\int_C v(z)\,dx+\int_C u(z)\,dy\right\},$$

where $f(z)=u(z)+iv(z)$. The integral is then said to be an integral in the complex domain. (For the application of integrals in the complex domain to complex analysis → 198 Holomorphic Functions.)

E. The Surface Integral

By an m-dimensional smooth surface S we mean the image S of a †regular mapping of class C^1 from a domain G in $\mathbf{R}^m$ into $\mathbf{R}^n$ $(m<n)$, $\xi(u)=(\xi_1(u_1,\ldots,u_m),\ldots,\xi_n(u_1,\ldots,u_m))$. Given a continuous function $f(x_1,\ldots,x_n)$ defined in a neighborhood U of S in $\mathbf{R}^n$, the multiple integral

$$\int\cdots\int_G f(\xi_1(u),\ldots,\xi_n(u))\times\frac{D(x_{i_1},\ldots,x_{i_m})}{D(u_1,\ldots,u_m)}du_1\ldots du_m,$$

$$\{i_1,\ldots,i_m\}\subset\{1,\ldots,n\}, \qquad (5)$$

is called the **surface integral** of f along S with respect to $x_{i_1},\ldots,x_{i_m}$ and is denoted by $\int_S f dx_{i_1}\ldots dx_{i_m}$ or $\int\ldots\int_S f dx_{i_1}\ldots dx_{i_m}$. This definition depends on the choice of the parameters $(u_1,\ldots,u_m)$ in the following way. Let $(u_1',\ldots,u_m')$ be another parametrization of S. Since

$$\frac{D(x_{i_1},\ldots,x_{i_m})}{D(u_1,\ldots,u_m)}=\frac{D(x_{i_1},\ldots,x_{i_m})}{D(u_1',\ldots,u_m')}\frac{D(u_1',\ldots,u_m')}{D(u_1,\ldots,u_m)}$$

and

$$du_1'\ldots du_m'=\left|\frac{D(u_1',\ldots,u_m')}{D(u_1,\ldots,u_m)}\right|du_1\ldots du_m,$$

thus according as

$$\frac{D(u_1,\ldots,u_m)}{D(u_1',\ldots,u_m')}\gtrless 0,$$

the foregoing expression remains the same or changes its sign. Usually we assign to the parameters $(u_1,\ldots,u_m)$ positive or negative orientation, in which case S is called oriented. If we replace the Jacobian $D(x_{i_1},\ldots,x_{i_m})/D(u_1,\ldots,u_m)$ in (5) by the quantity

$$\left(\sum_{i_1<\ldots<i_m}\left(\frac{D(x_{i_1},\ldots,x_{i_m})}{D(u_1,\ldots,u_m)}\right)^2\right)^{1/2},$$

which corresponds to the surface element of S, the integral is called a **surface integral with respect to the surface element** and is denoted by $\int_S f dS$ or $\int_S f d\sigma$. The surface integral of a differential form of degree m in $\mathbf{R}^n$ is similarly defined. In the case $m=1$, the surface integral reduces to a curvilinear integral. Just as the Stieltjes integral is a generalization of the curvilinear integral, there are several ways to generalize the notion of surface integral without assuming that the mapping $\xi(u)$ is of class C^1.

F. The Stokes Formula

Let S be an m-dimensional smooth surface in $\mathbf{R}^n$ $(m\leqslant n)$ and ∂S be the $(m-1)$-dimensional surface corresponding to the boundary of S. Let ω be a differential form of class C^1 of degree $(m-1)$ and $d\omega$ be its †exterior derivative. Then we have $\int_{\partial S}\omega=\int_S d\omega$, which is called the **Stokes formula** (or the **Green-Stokes formula**). (For the Stokes formula on a general differentiable manifold → 105 Differentiable Manifolds.) As special cases of the Stokes formula, we have the following three classical theorems:

(1) The case of a plane domain: Let D be a bounded domain on the xy-plane bounded by a finite number of smooth curves C with positive directions. If $\omega=P\,dx+Q\,dy$ is a differential form of class C^1 on $\bar{D}$, we have

$$\int_C P\,dx+Q\,dy=\iint_D\left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right)dx\,dy, \qquad (6)$$

since $d\omega=(\partial Q/\partial x-\partial P/\partial y)dx\wedge dy$. This is called **Green's formula** (or **Green's formula on the plane**). Equality (6) is true if P, Q are totally differentiable and the integrand on the right-hand side is continuous (even if the functions $\partial Q/\partial x$, $\partial P/\partial y$ themselves are not continuous) (E. Goursat). This formula remains true under the following weaker assumptions: (i) C is rectifiable; (ii) P and Q are continuous in $\bar{D}$ and $\partial P/\partial y$ and $\partial Q/\partial x$ are continuous and summable (in the sense of Lebesgue) in D.

(2) The case of a domain in a 3-dimensional space: Let D be a bounded domain in xyz-space surrounded by a finite number of smooth surfaces S. For a †vector field $\mathbf{V}=\langle P,Q,R\rangle$ of class C^1 on $\bar{D}$, we put $\omega=P\,dy\wedge dz+Q\,dz\wedge dx+R\,dx\wedge dy$. Then since $d\omega=(\partial P/\partial x+\partial Q/\partial y+\partial R/\partial z)dx\wedge dy\wedge dz$, we have

$$\iint_S P\,dy\,dz+Q\,dz\,dx+R\,dx\,dy$$
$$=\iiint_D \operatorname{div}\mathbf{V}\,dx\,dy\,dz$$
$$=\iiint_D\left(\frac{\partial P}{\partial x}+\frac{\partial Q}{\partial y}+\frac{\partial R}{\partial z}\right)dx\,dy\,dz. \qquad (7)$$

Equality (7) is called the **Gauss formula (Ostrogradskiĭ's formula** or the **divergence theorem)**. The left-hand side of (7) is equal to the surface integral $\iint_S(\mathbf{V},\mathbf{n})d\sigma$, which is the †vector flux through S (where $\mathbf{n}$ means the outer unit normal vector of the surface S). This formula remains true under the following weaker assumptions: (i) S is piecewise of class C^1; (ii) P, Q, R are continuous in $\bar{D}$ and $\partial P/\partial x$, $\partial Q/\partial y$, and $\partial R/\partial z$ are continuous and summable in D.

(3) The case of a bordered surface in a 3-dimensional space: Let $\bar{S}: x=x(u,v)$, $y=y(u,v)$, $z=x(u,v)$ $((u,v)\in\bar{G})$ be a smooth surface in xyz-space, and suppose that the boundary Γ of the domain of the parameters $\bar{G}$ consists of a finite number of smooth curves with positive direction. The boundary C of the surface S is the image curve of Γ. Now let $\mathbf{V}=\langle P,Q,R\rangle$ be a vector field of class C^1 on S, $\mathbf{n}$ be the unit normal vector of S (its direction being canonically assigned by the parameter (u,v)), and $\mathbf{t}$ be the unit tangent vector, and set $\omega=P\,dx+Q\,dy+R\,dz$. Then since $d\omega=(\partial R/\partial y-\partial Q/\partial z)dy\wedge dz+(\partial P/\partial z-\partial R/\partial x)dz\wedge dx+(\partial Q/\partial x-\partial P/\partial y)\cdot dx\wedge dy$, we have

$$\iint_S(\operatorname{rot}\mathbf{V},\mathbf{n})d\sigma=\iint_S\left(\left(\frac{\partial R}{\partial y}-\frac{\partial Q}{\partial z}\right)dy\,dz\right.$$
$$\left.+\left(\frac{\partial P}{\partial z}-\frac{\partial R}{\partial x}\right)dz\,dx+\left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right)dx\,dy\right)$$
$$=\int_C(P\,dx+Q\,dy+R\,dz)=\int_C(\mathbf{V},\mathbf{t})ds. \qquad (8)$$

Equation (8) is called the **Stokes formula** (→ Appendix A, Table 3.III).

References

[1] H. Lebesgue, Leçons sur l'intégration et la recherche des fonctions primitives, Gauthier-Villars, second edition, 1928, ch. II.
[2] S. Saks, Theory of the integral, Warsaw, 1937.
[3] D. V. Widder, The Laplace transform, Princeton Univ. Press, 1941, ch. I.
[4] H. Flanders, Differential forms, with applications to the physical sciences, Academic Press, 1963.
[5] H. K. Nickerson, D. C. Spencer, and N. E. Steenrod, Advanced calculus, Van Nostrand, 1959.
[6] H. Okamura, On the surface integral and Gauss-Green's theorem, Mem. Coll. Sci. Univ. Kyôto, 26 (1950), 5–14.
[7] L. Schwartz, Cours d'analyse, Hermann, 1967.
[8] H. Cartan, Formes différentielles, Hermann, 1967.

95 (XVI.4) Cybernetics

The term **cybernetics** was invented in 1947 by Norbert Wiener [1] to denote a field of science that treats the system of control and communication in animals and machine. The term was derived from a Greek word κυβερνήτης, also the source of the word "governor." The stimulus for establishing such a new area of science came from studies of automatic computation, automatic control, and information processing. These fields had given rise to technological innovations such as the high-speed electronic computer and automatic control instruments. Such machines have had a profound influence on the information sciences, as well as on theoretical investigations in biology. However, even now, we can hardly say that cybernetics has been formly established as a systematic branch of science or of applied mathematics. Nevertheless, it has had far-reaching influence on both biology and machine engineering as a methodology and as a philosophy. In the Soviet Union and in European countrics, the word corresponding to cybernetics is still widely used for the "Grenzgebiete" between biology and machine engineering in the wider sense. Nonetheless, it is difficult to say that cybernetics as a whole has undergone systematic development. Systematic theories have

been established and developed for separate parts of the field, and these have become independent disciplines; they are, for example, †control theory, †information theory, the theory of artificial intelligence, the theory of †automata, mathematical biology, genetics, and ecology.

The contributions of Wiener to cybernetics include not only the invention of the concept itself, but also prediction theory or the theory of the Wiener filter [2]. The latter was subsequently formalized by R. E. Kalman from a different point of view as an estimation problem within a linear system, working from finite observation data, and including the nonstationary case (→ 405 Stochastic Control and Stochastic Filtering G). Another contribution by Wiener is the input-output identification of a nonlinear system using statistical time series analysis of the outputs when the inputs are white noises [4]. This is achieved by expansion of the output function in terms of the convolution of input functions, which corresponds to the expansion of functions in terms of †Hermite interpolation polynomials. The kernel of the expansion is called a **Wiener kernel**. Recently, this method has been applied to input-output identification for nervous systems; it also finds application in nonlinear system theory [5]. Wiener also studied brain waves [6, 9].

Cybernetics in the wider sense, although not necessarily called by this name, may include investigations of the following "Grenzgebiete": information processing of the nervous system [7]; self-organizing systems in the nonequilibrium thermodynamics of Prigogine and Haken; the self-reproducing machines of von Neumann; and the theory of pattern formation in biological or chemical systems [8].

For related topics → 176 Gaussian Process I, 395 Stationary Processes D.

References

[1] N. Wiener, Cybernetics; or, control and communication in the animal and the machine, Wiley, 1947; second edition, 1961.
[2] N. Wiener, Extrapolation, interpolation and smoothing of stationary time series, with engineering applications, Wiley, 1949.
[3] R. E. Kalman and R. S. Bucy, New results in linear filtering and prediction theory, Trans. ASME, J. Basic Eng., 83 (1961), 95–107.
[4] N. Wiener, Nonlinear problems in random theory, MIT Press, 1958.
[5] Y. W. Lee and S. Schetzen, Measurement of the Wiener kernels of a non-linear system by cross correlation, Int. J. Control, 2 (1965), 237–254.
[6] N. Wiener and A. Rosenblueth, The mathematical formulation of the problem of conduction of impulses in a network of connected excitable elements, specifically in cardiac muscle, Arch. Inst. Cardiology, Mexico, 16 (1946), 205–265.
[7] M. A. Arbib, Metaphorical brain, Interscience, 1972.
[8] P. C. Fife, Mathematical aspects of reacting and diffusing systems, Lecture notes in math. 28, Springer, 1979.
[9] N. Wiener and J. P. Schadé, Cybernetics of the nervous system, Progress in brain research XVII, Elsevier, 1965.
[10] K. Steinbuch and S. W. Wagner, Neuere Ergebnisse der Kybernetik, R. Oldenbourg, 1964.
[11] S. Beer, Decision and control, Wiley, 1966.

D

96 (XVI.7) Data Processing

A. General Remarks

With the development of electronic computers, effective systems for data transmission and processing have been created on a large scale, and there is now a large literature concerning data-processing methods and techniques for such systems.

Research on data processing encompasses both specific techniques and entire systems of processing, for example, the system of programming techniques, as well as mathematical problems originating from them, such as the complexity of computations (→ 71 Complexity of Computations). As domains of application, we have, for example, information retrieval, stock management, program evaluation, and review techniques. We denote the set of data by D and assume that D is finite. Depending on the properties of D, we have various suitable representations and processing problems.

B. The Notion of Data

In recent applications of computers, the main task has been arranging and searching for items or attributes in storage, rather than numerical computation. Individual information is called a **record**. Records arranged and stored in the memories of computers are usually called **data**. A collection of such data is called a **file**. A complex, large-scale data collection is often called a **data base**.

In an abstract sense, the record consists of a string of letters, but it is usually convenient to view it as consisting of its identifying mark followed by a finite sequence of items. The contents can be classified according to their properties, such as topological relations, order relations, or items representing numerical values. For each case, there may be different suitable representations and operations.

We often store information after a suitable process of **information compression**. In some cases the operations are reversible, and complete recovery is possible. In other cases the operations are not completely reversible, and we must throw away part of the information in order to compress the rest. An example of recoverable information is the replacement of a run of 1's (or 0's) in a binary code by its length. An example of lost information often occurs in the graphical manipulation of numbers by means of **hashing**. In this process we select a suitable function $s=f(a_1, \ldots, a_n)$, where the vector $(a_1, \ldots, a_n)$ corresponds to the representation of the record in such a way that the value s can be monomorphic for the universe of the records as often as possible, and the data $(a_1, \ldots, a_n)$ is kept in the storage corresponding to the value s. Generally, the variable representing the location of data is called a **pointer**.

In the abstract sense, data bases are simply sets of items, but in most cases a base will also exhibit some mathematical structure, such as order relations. In order to handle a data base efficiently, we must take its particular structure into account, and suitably represent that structure. Such structures and their representations are called **data structures**.

C. Linear Structures

The component of a vector or the stations along a single railway line form linearly ordered sets. Their most essential feature is the notion of immediate predecessor or immediate successor. Such a data structure is called a **linear structure**, and the aligned sequence is called an **array**. Multidimensional arrays, such as the elements of a matrix, are stored in the form of 1-dimensional arrays in the memory of a computer.

When the set of data D is a linearly ordered set, it is usually represented as a suitable array. Here, the problems of **ordering** and **table look-up** are fundamental. Ordering in this sense means putting the given elements of D into the order defined for D. Since, historically, sorting machines were used to put punched cards in order, this process is also called **sorting**. The process of arranging several individually sorted data packs into one sorted pack is called **merging**.

A fundamental data-sorting operation is "comparison" with respect to the order for D. There have been many investigations of the estimation of the lower or upper bounds for the number of comparisons and of developing efficient algorithms. Asymptotically, $O(n \log n)$ is the theoretical lower bound for n elements, and some algorithms are known to achieve this bound (→ 71 Complexity of Computations).

If there is given a univalent correspondence $f: D \to D'$ and the correspondence table is stored in the memory in a suitable form, the problem of table look-up arises, which requires finding $f(d)$ for a given $d \in D$. In this process, a fundamental operation is a comparison of $d \in D$ with some $x \in D$ in the table. There have been many investigations of efficient arrange-

ments and of algorithms with minimal numbers of comparisons.

D. List Representation

If the given finite set D has an order satisfying the reflexive and transitive laws, addressing is often used as a medium for representing the order in the memory. Let $d_1, \ldots, d_n$ be all the immediate successors to an element $d \in D$. We can represent them by the following sequence of triplets:

If $n=0$, we take $(d,0,0)$;
If $n=1$, we take $(d,d_1^*,0)$;
If $n=2$, we take (d,d_1^*,d_2^*);
If $n=3$, we take $(d,d_1^*,e_1^*),(e_1,d_2^*,d_3^*)$;
If $n \geqslant 4$, we take $(d,d_1^*,e_1^*),(e_1,d_2^*,e_2^*),$
$(e_2,d_3^*,e_3^*),\ldots,(e_{n-2},d_{n-1}^*,d_n^*)$,

where d_i^* means the pointer to d_i, e_i is an element introduced for convenience, and e_i^* is the pointer to e_i. This is called a **list** representation. If we denote by $d \rightarrow d_1$, the fact that d_1 is an immediate successor of d, then the total set can be regarded as a †direct graph. Usually, this graph is a †tree, and then the graph is called a **tree structure** and its representation a **tree representation**. The advantage of this representation is that addition or deletion is quite easy.

The set of logical formulas is partially ordered, with the order given by the rules of inference. The set consisting of a series of inferences forms an ordered subset. Thus, if a tree representation can be automatically treated, so can the process of inference.

In dealing with linguistic data (words or sentences, for example), it is often natural to consider a noncommutative †free semigroup D generated by a finite number of generators (the alphabet or vocabulary). In this case, if there exists a natural order for the generators, it determines in D a lexicographic partial ordering. Then the tree representation can be used for representing a dictionary whose entires are elements of D. This method is not efficient with respect to speed of table look-up and economy of memory, but it sometimes has the advantage of simplifying the treatment of complicated data.

E. Memory Devices for Processing

In dealing with algebraic formulas or languages with parentheses, data maintenance methods such as tree representation or push-down storage are often convenient as auxiliary memory-controlling methods. The characteristic of the **push-down storage** method is that it returns the data in reverse order with respect to the time of acceptance and remittance.
It is sometimes called a **stack** or a **first-in–last-out memory**. Contrary to a stack, there is a storage that returns the data in the same order as the acceptance. It is called a **queue** or a **first-in–first-out memory**.

F. Information Retrieval

A request to obtain all the records qualified by some property in a data base Ω is called a **query**. The part of the collection of records Ω qualified for a query, therefore, can be identified by its characteristic subset of the space of all possible records. A procedure that determines the subset Ω_q of Ω composed of all the records relevant to a query q is called the **information retrieval** of the query. A system which provides such a procedure for every query $q \in \mathbf{Q}$ is called an **information retrieval system** organized for Ω with respect to $\mathbf{Q}$.

It is essential to design the system in such a way that records can be retrieved quickly for queries in a certain class. Such a selected set of queries can be composed of all the first-order queries specifying an item and asking for all the records containing the attribute value or the key characterizing the item. It can, in some cases, include second- or higher-order queries specifying a number of items and asking for all the records in which attribute values or keys characterizing the item occur simultaneously.

In an information retrieval system, the master file of a data base Ω is usually organized by way of auxiliary memories on data structures such as sequential files, indexed sequential files, virtual storage files, or direct access files, using the magnitude of the accession number or primary key of each record. Various management systems, called SAM (Sequential Access Method), ISAM (Index Sequential Access Method), VSAM (Virtual Storage Access Method), or DAM (Direct Access Method), exist.

In addition to the organization of such a master file, various directory files or indexes are organized in order to retrieve each query $q \in \mathbf{Q}$ quickly, because there are many relevant keys or combination of keys other than the primary key. An **inverted filing scheme** (IFS) is a typical scheme for organizing such indexes or directory files. A bucket B_t or an addressable set of secondary memories is provided for each canonical query $q_t \in \mathbf{Q}$ in a one-to-one way. An index or a list of accession numbers of pertinent records is organized in each of the buckets contiguously so as to make it possible

to retrieve all accession numbers of relevant records quickly. The essentials of an IFS is to define a one-to-one QAT (Query to Address Transformation) from the inverted set of queries $\mathbf{Q}=\{q_t\}$ to the set of buckets $\mathbf{B}=\{B_t\}$.

An IFS has efficient retrieval performance with respect to the inverted canonical queries. The scheme, however, may in some cases require a large number of buckets. Moreover, the scheme requires quite redundant storage of accession numbers by a number of buckets because a record may be pertinent to a number of canonical queries simultaneously. Although higher retrieval performance can be expected by including higher-order queries in the inverted set, the space and machine time needed for such organization is prohibitive. This is one of the reasons why an IFS for first-order queries is preferable in almost all practical cases. Another reason is that if the first-order queries are inverted, every retrieval can logically be performed by certain Boolean operations executed among a certain number of retrieved sets of accession numbers. A trade-off of space and time needed for the organization of a scheme and its retrieval performance might be the determining factor for the selection of the set of canonical queries to be inverted.

An attempt to overcome the limitations inherent in the inverted scheme can be found in the work on the balanced file organization scheme due to Abraham et al. [7]. By extracting the essentials, Yamamoto et al. [9, 10] defined a BFS (Balanced File-organization Scheme) in a wider sense as follows:

(i) Buckets are organized in such a manner that every bucket is associated with more than one query.

(ii) Every canonical query is associated with a unique bucket.

(iii) The accession number of a record with some additional information is stored in a bucket once if and only if it is pertinent to at least one of the associated queries.

The essentials of a BFS is to define a many-to-one transformation from the set of canonical queries $\mathbf{Q}$ to the set of bucket addresses $\mathbf{B}$, or an MQAT (Multiple Queries to Address Transformation). An MQAT defines a partition of $\mathbf{Q}$ into mutually disjoint subsets. It is a generalization of a QAT which defines an IFS. If c is the number of queries to be associated simultaneously with a unique bucket in a BFS, then the number of buckets to be prepared is $1/c$, which is a drastic reduction from that for an IFS. This reduction makes it possible to extend the feasible range of canonical queries. Actually, in the system HUNDRED (Hiroshima University New Documents REtrieval and Dissemination), a 51-to-1 MQAT is used for the organization of indexes.

Although some space overload may occur by storing some additional information in order to tell which record is pertinent to which query in a bucket, reduction of redundancy can be expected, because a record may be pertinent to more than one query simultaneously. The reduction in the number of buckets and redundancy would contribute much in saving time and space needed for the organization of indexes.

Among those BFS's defined by MQAT's with a given data base Ω, $\mathbf{Q}$, and the number c of queries associated with the same bucket, one BFS can be called the best if its redundancy is the least under some reasonable assumptions imposed on the distribution of keys. Several combinatorial problems have been raised and solved in this connection [10].

References

[1] A. T. Berztiss, Data structure—Theory and practice, Academic Press, 1971.

[2] H. S. Stone, Introduction to computer organization and data structures, McGraw-Hill, 1972.

[3] A. K. Scidmore and B. L. Weinberg, Storage and seach properties of a tree-organized memory system, Comm. ACM, 6 (1963), 28–31.

[4] R. D. Luce, R. R. Bush, and E. Galanter, Handbook of mathematical psychology II, Wiley, 1963.

[5] Y. Bar-Hillel, Language and information, Addison-Wesley, 1964.

[6] J. A. Fodor and J. J. Katz, The structure of language: Readings in the philosophy of language, Prentice-Hall, 1964.

[7] C. T. Abraham, S. P. Ghosh, and D. K. Ray-Chaudhuri, File organization schemes based on finite geometries, Information and Control, 12 (1968), 143–163.

[8] S. P. Ghosh, Data base organization for data management, Academic Press, 1977.

[9] S. Yamamoto, H. Ikeda, S. Shige-eda, K. Ushio, and N. Hamada, Design of a new balanced file organization scheme with the least redudance, Information and Control, 21 (1975), 72–91.

[10] S. Yamamoto, Design of optimal balanced filing schemes, Proc. IBM Working Conference on Database Engineering, in Database Engineering, Springer 1981, vol. II, 23–35.

[11] D. E. Knuth, The art of computer programming, Addison-Wesley, vol. 1, revised edition, 1976; vol. 2, revised edition, 1980; vol. 3, 1973.

97 (I.10) Decision Problem

Suppose that we are given a set S and a proposition $P(x_1, x_2, \ldots, x_n)$ for elements x_i of S. Then we have the **problem of universal validity of** P, which is the problem of finding a general **algorithm** (i.e., a finitary procedure) by which we can discern whether $P(x_1, \ldots, x_n)$ is true for all n-tuples $(x_1, \ldots, x_n)$. The problem of finding an algorithm by which we can discern the validity of $P(x_1, \ldots, x_n)$ for some specifically chosen n-tuples $(x_1, \ldots, x_n)$ is called the **problem of satisfiability of** P. These two problems are customarily called **decision problems**. The problems are such that affirmative solution of one of them implies negative solution of the other.

To give a precise definition of decision problems, let us note that a †free semigroup with countable generators can be identified with a subset of the set $\mathbf{N}$ of natural numbers (by virtue of †Gödel numbering; → 185 Gödel numbers). On the other hand, if $\mathfrak{S}$ is a given †formal system with countably many symbols, the set of all †formulas in $\mathfrak{S}$ is a subset of the free semigroup generated by the symbols in $\mathfrak{S}$. Thus the set of all formulas in $\mathfrak{S}$ is identified with a subset of $\mathbf{N}$. A subset M of $\mathbf{N}$ (or $\mathbf{N} \times \mathbf{N} \times \ldots \times \mathbf{N}$) is **(general) recursive** if its †representing function is general recursive (→ 356 Recursive Functions). By using the concept of recursive function, a precise definition of the decision problem can be given as follows: **The decision problem of** M **is solved affirmatively** if and only if we can obtain effectively a procedure defining the representing function of M, and the function is general recursive. **The decision problem of** M **is solved negatively** if and only if we can obtain a proof that M is not recursive.

For a set A of formulas in $\mathfrak{S}$, we let $g(A)$ be the set of all Gödel numbers corresponding to the elements of A. Let A' be the set of formulas in A that are deducible in $\mathfrak{S}$, and let $\tau(A) = g(A')$. The decision problem of the set A of formulas is said to be solved affirmatively (negatively) if the decision problem of $\tau(A)$ is solved affirmatively (negatively). By refining this concept we arrive at the notion of the **degree of (recursive) unsolvability**. Let A and B be subsets of $\mathbf{N}$. The relation "A is recursive in B and B is recursive in A" (→ 356 Recursive Functions) is reflexive, symmetric, and transitive. Hence this relation decomposes the class of all subsets in $\mathbf{N}$ into disjoint nonempty equivalence classes. A and B are defined to have the same degree of unsolvability if they belong to the same equivalence class. Thus the degrees of unsolvability can be identified with the equivalence classes. The degree of recursive sets is $\mathbf{0}$.

The relation $\mathbf{a} \leqslant \mathbf{b}$ is defined between the degrees $\mathbf{a}$ of A and $\mathbf{b}$ of B to mean "A is recursive in B." Clearly, for any degree $\mathbf{a}$, we have $\mathbf{0} \leqslant \mathbf{a}$. The partially ordered system of degrees constitutes an †upper semilattice.

Research on the decision problem has been done mostly in areas related to the †first-order predicate calculus L^1 and the formal systems on it. We now list some important results.

(I) Results concerning L^1. The decision problem has been solved negatively for the sets of formulas of the following forms. (Here it is assumed that no function symbols appear and that $\mathfrak{A}$ represents a formula involving no occurrence of $\forall$, $\exists$, or free variables.)

(1) All formulas in L^1 (A. Church, A. M. Turning),
(2) $\exists x_1 \exists x_2 \ldots \exists x_m \forall y_1 \forall y_2 \ldots \forall y_n \, \mathfrak{A}$ (T. Skolem),
(3) $\exists x_1 \exists x_2 \exists x_3 \forall y_1 \forall y_2 \ldots \forall y_n \exists z \, \mathfrak{A}$ (K. Gödel),
(4) $\exists x_1 \exists x_2 \forall y_1 \forall y_2 \ldots \forall y_n \exists z \, \mathfrak{A}$ (L. Kalmar),
(5) $\exists x_1 \exists x_2 \forall y \exists z_1 \exists z_2 \ldots \exists z_n \, \mathfrak{A}$ (J. Pepis),
(6) $\forall x \exists y \forall z \exists u_1 \exists u_2 \ldots \exists u_n \, \mathfrak{A}$ (W. Ackermann),
(7) $\exists x_1 \exists x_2 \exists x_3 \forall y \, \mathfrak{A}$ or $\exists x_1 \exists x_2 \forall y \exists z \, \mathfrak{A}$ (J. Surányi),
(8) $\exists x \forall y_1 \forall y_2 \exists z_1 \exists z_2 \, \mathfrak{A}$ or $\forall x \exists y \forall z \exists u_1 \exists u_2 \, \mathfrak{A}$ (Surányi).

The decision problem has been solved affirmatively for the sets of formulas of the following forms, where it is assumed again that no function symbols appear and that $\mathfrak{A}$ is as above.

(1) All formulas involving variables only on predicates with one argument. (L. Löwenheim, Skolem, H. Behmann),
(2) $\forall x_1 \forall x_2 \ldots \forall x_m \, \mathfrak{A}$ (P. Bernays, M. Schönfinkel, Ackermann),
(3) $\forall x_1 \forall x_2 \ldots \forall x_m \exists y_1 \exists y_2 \ldots \exists y_n \, \mathfrak{A}$ (Bernays, Schönfinkel, Ackermann),
(4) $\forall x_1 \forall x_2 \ldots \forall x_m \exists y_1 \exists y_2 \forall z_1 \forall z_2 \ldots \forall z_n \, \mathfrak{A}$ (Gödel, Kalmar, K. Schutte).

(II) Results concerning formal systems on L^1. Throughout the rest of this article we assume that no †function variables appear. Predicate constants, function constants, and object constants may appear. By the decision problem for a formal system $\mathfrak{L}$ we mean the decision problem for all †closed formulas in $\mathfrak{L}$. Most of the results obtained so far concerning the decision problem for formal systems have been negative. Such results include those for formal systems formalizing natural number theory, the theory of rational integers, the elementary theory of †groups, †rings, †fields, †lattices, and the like, and axiomatic set theory (A. Tarski et al.).

The word problem for groups was solved negatively by P. S. Novikov (→ 161 Free Groups B). In connection with this decision problem, there are some investigations

by W. W. Boone, G. Higman, and others [14–17].

The decision problem for a formal system formalizing the elementary theory of †Abelian groups has been solved affirmatively (W. Szmielew). Little is known about the decision problem concerning partial systems of formulas of given formal systems except the following: (1) the decision problem for the set of formulas of the form $\forall x_1 \forall x_2 \dots \forall x_m \mathfrak{A}$ in a formal system (→ 161 Free Groups); (2) the Hilbert-type problem, which is the decision problem for the set of formulas of the form $\exists x_1 \exists x_2 \dots \exists x_m (t=s)$ in a formal system. In particular, the Hilbert-type problem in a formal system formalizing natural number theory is called Hilbert's tenth problem (→ 196 Hilbert). The latter is the problem of finding an algorithm for deciding whether a †Diophantine equation has an integral solution.

This decision problem was studied by M. Davis, H. Putnam, J. Robinson, and others, and finally Yu. V. Matiyasevich solved it negatively by showing that every recursively enumerable relation is Diophantine [9]. (A relation $R(m_1, \dots, m_j)$ is called **Diophantine** if there is a polynomial $P(x_1, \dots, x_j, y_1, \dots, y_k)$ with integer coefficients such that $R(m_1, \dots, m_j)$ holds if and only if $P(m_1, \dots, m_j, y_1, \dots, y_k) = 0$ has a solution for $y_1, \dots, y_k$ in natural numbers.) In addition, some investigations have been made about the †second-order predicate calculus L^2, †intuitionistic logic, etc. [1, 2].

References

[1] W. Ackermann, Solvable cases of the decision problem, North-Holland, 1954.
[2] S. C. Kleene, Introduction to metamathematics, Van Nostrand, 1952.
[3] A. Tarski, Undecidable theories, North-Holland, 1953.
[4] A. Church, An unsolvable problem of elementary number theory, Amer. J. Math., 58 (1936), 345–363.
[5] S. C. Kleene and E. L. Post, The upper semi-lattice of degrees of recursive unsolvability, Ann. Math., (2) 59 (1954), 379–407.
[6] G. E. Sacks, Degrees of unsolvability, Ann. Math. Studies, Princeton Univ. Press, 1963.
[7] M. Davis, H. Putnam, and J. Robinson, The decision problem for exponential diophantine equations, Ann. Math., (2) 74 (1961), 425–436.
[8] M. Davis and H. Putnam, Reductions of Hilbert's tenth problem, J. Symbolic Logic, 23 (1958), 183–187.
[9] Yu. V. Matiyasevich, Diophantine enumerable sets (in Russian), Dokl. Akad. Nauk SSSR, 191 (1970), 279–282.
[10] H. Hermes, Enumerability, decidability, computability, Springer, 1965.
[11] H. Rogers, Theory of recursive functions and effective computability, McGraw-Hill, 1967.
[12] J. R. Shoenfield, Mathematical logic, Addison-Wesley, 1967.
[13] P. S. Novikov, On the algorithmic unsolvability of the word problem in groups (in Russian), Trudy Mat. Inst. Steklov., 44 (1955).
[14] G. Higman, Subgroups of finitely presented groups, Proc. Roy. Soc. London, (A) 262 (1961), 455–475.
[15] W. W. Boone, The word problem, Ann. Math., 77 (1963), 16–32.
[16] C. F. Miller, On group-theoretic decision problems and their classification, Ann. Math. Studies, Princeton Univ. Press, 1971.
[17] W. W. Boone, F. B. Cannonito, and R. C. Lyndon (eds.), Word problems, North-Holland, 1973.

98 (XXI.18) Dedekind, Julius Wilhelm Richard

Julius Wilhelm Richard Dedekind (October 6, 1831–February 12, 1916) was born in the city of Braunschweig in central Germany and studied at the University of Göttingen under C. F. †Gauss, who was then in his later years. He received his doctorate at Göttingen with a thesis on the †Euler integral. He was professor of mathematics from 1858 to 1862 at Zürich and from 1863 to 1894 at the Technische Hochschule in Braunschweig. During his early twenties, he wrote papers concerning analysis and the theory of probability, but in 1857 he began publishing papers on the theory of numbers. He edited †Dirichlet's lectures on number theory (*Vorlesungen über Zahlentheorie*, first edition 1863, fourth edition 1899) and concentrated on research in arithmetic and algebra. The theory of †ideals, which he founded, was originally set out in a supplement (1863) to Dirichlet's *Vorlesungen.*

Dedekind treated subjects ranging from the axiomatic foundations of the theory of ideals to †lattices and †groups as algebraic systems. He was a pioneer of the abstract algebra of the 20th century. Among his notable achievements are the †Dedekind zeta functions of †algebraic number fields, †Dedekind cuts in the theory of real numbers, the algebraic theory of †algebraic functions (of which he was a coauthor with H. Weber), and the theory of natural numbers. He was one of the first to support

†Cantor's set theory. His theory of natural numbers was founded on the concept of sets and included the idea of †recursive functions.

References

[1] R. Dedekind, Gesammelte mathematische Werke I–III, Vieweg, 1930–1932 (Chelsea, 1969).
[2] M. Cantor and R. Dedekind, Briefwechsel, Cantor-Dedekind, Actualités Sci. Ind., Hermann, 1937.

99 (IX.4) Degree of Mapping

A. Degree of Mapping

Let M^n and N^n be n-dimensional †closed †oriented †C^r manifolds (or †combinatorial manifolds). For example, $M^n = N^n = S^n$ (the †n-sphere). Their nth †homology groups with integral coefficients $H_n(M^n; \mathbf{Z})$ and $H_n(N^n; \mathbf{Z})$ are infinite cyclic groups generated by the fundamental homology classes $[M^n]$ and $[N^n]$, respectively (→ 201 Homology Theory). A continuous mapping $f: M^n \to N^n$ induces a homomorphism $f_*: H_n(M^n; \mathbf{Z}) \to H_n(N^n; \mathbf{Z})$, and there exists an integer d_f such that $f_*([M^n]) = d_f[N^n]$. This integer d_f is called the **degree of mapping** (or the **mapping degree**) of f. When $M^n = N^n$, d_f does not depend on the orientation of M^n.

If a continuous mapping $g: M^n \to N^n$ is †homotopic to f ($f \simeq g$), then we have $d_f = d_g$. If f is homotopic to a †constant mapping ($f \simeq 0$), then $d_f = 0$, while if f is a homeomorphism, then $d_f = \pm 1$. When $M^n = N^n$, a homeomorphism $f: M^n \to M^n$ is called an **orientation-preserving mapping** if $d_f = 1$ and an **orientation-reversing mapping** if $d_f = -1$.

Suppose that M^n and N^n are closed oriented n-dimensional combinatorial manifolds and that $f: M^n \to N^n$ is a †simplicial mapping. Let $\sum_i \sigma_i^n$ and $\sum_j s_j^n$ (σ_i^n, s_j^n are n-simplexes of M^n, N^n, respectively) represent $[M^n]$, $[N^n]$, and let p_j (resp. q_j) be the number of n-simplexes σ_i^n such that $f(\sigma_i^n)$ is equal to s_j^n (resp. $-s_j^n$). Then $p_j - q_j$ is independent of the index j and equal to d_f.

Suppose that $f, g: S^n \to S^n$ are continuous mappings ($n \geq 1$). Then $f \simeq g$ if and only if $d_f = d_g$ (**Brouwer mapping theorem**). This implies that $\pi_n(S^n) \cong \mathbf{Z}$ (→ 202 Homotopy Theory).

B. Local Degree of Mapping

Suppose that M^n and N^n are n-dimensional oriented C^r (or combinatorial) manifolds and $f: M^n \to N^n$ is a continuous mapping. Suppose further that a point p of M^n has a neighborhood U such that $f(p) \neq f(q)$ for any point q contained in $U - \{p\}$. Then f induces a homomorphism $f_*: H_n(U, U - \{p\}) \to H_n(N, N - \{f(p)\})$ of n-dimensional †local homology groups with integral coefficients that are both isomorphic to $\mathbf{Z}$. If u and v are generators of the groups $H_n(U, U - \{p\})$ and $H_n(N, N - \{f(p)\})$ corresponding to orientations, respectively, then there exists an integer k such that $f_*(u) = kv$. We call this integer k the **local degree of mapping** f at p. If M^n and N^n are closed oriented C^r manifolds ($r \geq 1$) and $f: M^n \to N^n$ is a C^r mapping, then there exists a point r of N^n such that the set $f^{-1}(r)$ is a discrete subset $\{p_1, \ldots, p_t\}$ of M^n, and each p_i has a neighborhood U_i satisfying the foregoing condition (†Sard's theorem). If k_i is the local degree of f at p_i, then $d_f = \sum k_i$.

C. Linking Numbers

Given two mutually disjoint smooth closed curves C_1 and C_2 in Euclidean 3-space, a quantity $\mathrm{Lk}(C_1, C_2)$ indicating how closely they are interlinked with each other was given by Gauss as follows: Let C_i be expressed by the parameters $x_i = x_i(t_i)$ ($i = 1, 2$), where $x_i(t_i)$ are †continuously differentiable. Then the quantity

$$\mathrm{Lk}(C_1, C_2) = -\frac{1}{4\pi} \int_{C_1} \int_{C_2} \frac{1}{|x_2 - x_1|^3} \times \det\left(x_2 - x_1, \frac{dx_1}{dt_1}, \frac{dx_2}{dt_2} \right) dt_2 \, dt_1$$

is an integer called the **linking number** of C_1 and C_2.

More generally, let M^n be an n-dimensional oriented †combinatorial manifold (or C^r manifold ($r \geq 1$)) and K and K^* its †cellular decompositions such that K^* is dual to K. Let z_1^r and z_2^s ($r + s = n - 1$) be †boundaries belonging to the complex K and K^*. Suppose that C^{r+1} is any †chain of K whose boundary is z_1^r. Then the †intersection number $[C^{r+1}] \cdot [z_2^s]$ does not depend on the choice of such a chain C^{r+1}. We set $\mathrm{Lk}(z_1^r, z_2^s) = [C^{r+1}] \cdot [z_2^s]$ and call it the **linking number** of z_1^r and z_2^s. The **linking number** $\mathrm{Lk}(\tilde{z}_1^r, \tilde{z}_2^s)$ of †singular boundaries $\tilde{z}_1^r$, $\tilde{z}_2^s$ ($r + s = n - 1$) of M^n is similarly defined by considering the approximations z_1^r, z_2^s of $\tilde{z}_1^r$, $\tilde{z}_2^s$ belonging to a suitable cellular decomposition K and its dual K^*. The number $\mathrm{Lk}(\tilde{z}_1^r, \tilde{z}_2^s)$ is bilinear with respect to $\tilde{z}_1^r$, $\tilde{z}_2^s$, and we have $\mathrm{Lk}(\tilde{z}_1^r, \tilde{z}_2^s) = (-1)^{rs+1} \mathrm{Lk}(\tilde{z}_2^s, \tilde{z}_1^r)$. In the example in 3-dimen-

sional Euclidean space $\mathbf{R}^3$ shown in the left half of Fig. 1, we have $\mathrm{Lk}(\tilde{z}_1^1, \tilde{z}_2^1) = 1$, while $\mathrm{Lk}(\tilde{z}_1^1, \tilde{z}_2^2) = 2$ for the example shown in the right half of the same figure. In particular, if $\tilde{z}_1^r$ is homologous to 0 in $M^n - |\tilde{z}_2^s|$, then we have $\mathrm{Lk}(\tilde{z}_1^r, \tilde{z}_2^s) = 0$ (Fig. 2). Generally, if $\tilde{z}_1^r$ and $\tilde{z}_1'^r$ are homologous in $M^n - |\tilde{z}_2^s|$, then $\mathrm{Lk}(\tilde{z}_1^r, \tilde{z}_2^s) = \mathrm{Lk}(\tilde{z}_1'^r, \tilde{z}_2^s)$.

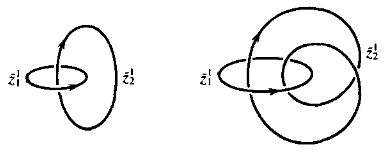


Fig. 1

D. Order of a Point with Respect to a Cycle

Let M^n be an n-dimensional oriented combinatorial manifold (or C^r manifold $(r \geqslant 1)$) with the nth Betti number $b_n = 0$, $\tilde{z}^{n-1}$ an $(n-1)$-dimensional singular boundary of M^n, and o a point of M^n that is not contained in $|\tilde{z}^{n-1}|$. We set $\mathrm{ord}(\tilde{z}^{n-1}, o) = \mathrm{Lk}(\tilde{z}^{n-1}, o)$ and call it the **order of the point** o with respect to $\tilde{z}^{n-1}$. For example, when $M^n = \mathbf{R}^2$ and $\tilde{z}^1 = \{f(t) \mid 0 \leqslant t \leqslant 1, f(0) = f(1)\}$, where f is a continuous function, the order $\mathrm{ord}(\tilde{z}^1, o)$ is equal to the **rotation number** around o of a moving vector $\overrightarrow{of(t)}$ as t varies from 0 to 1. This $\mathrm{ord}(\tilde{z}^1, o)$ stays invariant as the point o moves in a †connected component of the complement $\mathbf{R}^2 - |\tilde{z}^1|$ (Fig. 3). On the other hand, if $\tilde{z}_i^1 = \{f_i(t) \mid 0 \leqslant t \leqslant 1, f_i(0) = f_i(1)\}$ $(i = 0, 1)$ are closed curves in $\mathbf{R}^2$ and the distance $\rho(f_0(t), f_1(t))$ is smaller than $\rho(f_0(t), o)$ for all t in the interval $[0, 1]$, then we have $\mathrm{ord}(\tilde{z}_0^1, o) = \mathrm{ord}(\tilde{z}_1^1, o)$ (**Rouché's theorem**).

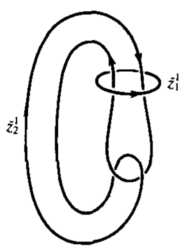


Fig. 2

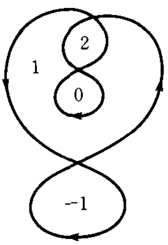


Fig. 3

References

[1] P. S. Aleksandrov and H. Hopf, Topologie I, Springer, 1935 (Chelsea, 1965).
[2] H. Seifert and W. Threlfall, Lehrbuch der Topologie, Teubner, 1934 (Chelsea, 1968); English translation, A textbook of topology, Academic Press, 1980.
[3] J. W. Milnor, Topology from the differentiable viewpoint, Univ. Press of Virginia, 1965.

100 (X.17) Denjoy Integrals

A. History

For a real-valued function $f(x)$ of a real variable to be †Lebesgue integrable, it is necessary and sufficient that there exist an †absolutely continuous function $F(x)$ such that $F'(x) = f(x)$ at †almost all points x (→ 221 Integration Theory D). In general, the derivative of a function is not necessarily Lebesgue integrable. A function $f(x)$ is Lebesgue integrable if and only if $|f(x)|$ is integrable. Hence a function which has an improper Riemann integral is not necessarily Lebesgue integrable (→ 221 Integration Theory A). For this reason, it is desirable to extend the concept of Lebesgue integrals. In 1912, A. Denjoy constructively defined a new concept of integrals (Denjoy integral in the restricted sense; → Section D), which is an extension of both Lebesgue and Riemann integrals. Later, N. N. Luzin provided the descriptive theory of this integral. Independently, and nearly simultaneously, A. J. Khinchin and Denjoy defined a more general integration (Denjoy integral in the wide sense (1916); → Section D).

In 1914, O. Perron, independently of Denjoy, defined a concept of integrals (Perron integrals) that is equivalent to that of Denjoy integrals in the restricted sense. To establish this concept, Perron considered the differential equation $y' = f(x)$ and utilized a method similar to the one used in the proof of the existence theorem for the solution of the differential equation $y' = f(x, y)$. However, the concept of Denjoy integrals is inadequate to treat unbounded functions. Thus to extend the concepts of Riemann and Lebesgue integrals, various ideas have been introduced; for example, Denjoy (1921), J. C. Burkill (1951), and R. D. James (1950) introduced new concepts as byproducts of investigations concerning the coefficients of trigonometric series [2, 3]. The A-integral concept devised by A. N. Kolmogorov was meant to deal with the problem of the conjugate function of Fourier series [4]. As a certain completion of the space of functionals of step functions, K. Kunugi defined the notion of E. R. integrals (1956), which coincides with that of A-integrals in a special case [5, 6].

What has been stated so far deals only with functions of a real variable. Concerning the extension of Denjoy integrals to the case of several variables, research has been done by M. Loomis, S. Kempisty, S. Nakanishi (née Enomoto), and others [7, 8].

B. Approximate Derivative

If we have $\lim_{h\to 0, k\to 0} m\{E\cap(\xi-h, \xi+k)\}/(h+k)=1$ at a point ξ of a †measurable subset E of the real line, where m is the 1-dimensional Lebesgue measure, the point ξ is called a **point of density** for E. Almost all the points of E are points of density for E (**Lebesgue's density theorem**). Let E be a measurable set having x_0 as a point of density, and let $F(x)$ be a measurable function on E. If there exists a number l such that for each $\varepsilon>0$, x_0 is a point of density for the set $\{x \mid l-\varepsilon \leqslant (F(x)-F(x_0))/(x-x_0) \leqslant l+\varepsilon, x\in E\}$, then l is called the **approximate derivative** of $F(x)$ at x_0 and is denoted by $\mathrm{AD}\,F(x_0)$. If $\mathrm{AD}\,F(x_0)$ exists, $F(x)$ is said to be **approximately derivable** at x_0. If $F(x)$ is approximately derivable at each point of E, then $F(x)$ is said to be approximately derivable in E. If $F'(x)$ exists at a point x, then $\mathrm{AD}\,F(x)$ exists at x, and we have $\mathrm{AD}\,F(x)=F'(x)$. However, there exists a continuous function $F(x)$ that is approximately derivable at almost all points of an interval and yet not differentiable at any point of a set of positive measure.

C. Generalized Absolute Continuity

Let E be a set in $\mathbf{R}$, and let $F(x)$ be a real-valued function whose domain contains E. If for each $\varepsilon>0$, there is a $\delta>0$ such that for every sequence $\{[a_n, b_n]\}$ of nonoverlapping intervals whose endpoints belong to E the inequality $\sum(b_n-a_n)<\delta$ implies $\sum|F(b_n)-F(a_n)|<\varepsilon$, then the function $F(x)$ is said to be **absolutely continuous** on E. We denote by AC the set of all functions that are absolutely continuous on E. If $F(x)$ is continuous on E and E is the union of a countable sequence of sets E_n on each of which $F\in AC$, then $F(x)$ is called a **generalized absolutely continuous function**, and we write $F\in GAC$. If $F\in GAC$, $\mathrm{AD}\,F(x)$ exists almost everywhere.

If, for each $\varepsilon>0$, there is a $\delta>0$ such that for every sequence $\{[a_n, b_n]\}$ of nonoverlapping intervals whose endpoints belong to E the inequality $\sum(b_n-a_n)<\varepsilon$ implies $\sum_n O\{F;[a_n,b_n]\}<\varepsilon$ ($O\{F;[a_n,b_n]\}$ denotes the oscillation of the function $F(x)$ in $[a_n,b_n]$, i.e., the difference between the least upper bound and the greatest lower bound of the values assumed by $F(x)$ on $[a_n,b_n]$), then $F(x)$ is said to be **absolutely continuous in the restricted sense** (or **absolutely continuous** $(*)$) on E; and we write $F\in AC(*)$. Just as we defined the notions of generalized absolute continuity and absolute continuity, so we define the notions of **generalized absolute continuity in the restricted sense** and **generalized absolute continuity** $(*)$. Thus $F\in GAC(*)$ means that $F(x)$ is a generalized absolute continuous $(*)$ function on E. If $F\in GAC(*)$, then $F'(x)$ exists almost everywhere.

D. Definitions of Denjoy Integrals

Let $f(x)$ be a real-valued function defined on $I=[a,b]$. If for $f(x)$ there exists a function $F(x)$ that belongs to GAC on I and for which $\mathrm{AD}\,F(x)=f(x)$ holds almost everywhere, then $f(x)$ is said to be **Denjoy integrable in the wide sense** (or **D-integrable**) on I. We call $F(b)-F(a)$ the **definite D-integral** of $f(x)$ over I, and denote the value by $(D)\int_a^b f(x)\,dx$. The function $F(x)$ is called an **indefinite D-integral** of $f(x)$ on I. Similarly, we obtain the definition of **Denjoy integral in the restricted sense** (or **D$(*)$-integral**) by replacing GAC by $GAC(*)$ and $\mathrm{AD}\,F(x)$ by $F'(x)$ in the definition of the D-integral. If a continuous function $F(x)$ satisfies the equality $\mathrm{AD}\,F(x)=f(x)\neq\pm\infty$ ($F'(x)=f(x)\neq\pm\infty$) for all except countably many points in I, then $F(x)$ is an indefinite D-integral (D$(*)$-integral) of $f(x)$. A Lebesgue-integrable function is D$(*)$-integrable, a D$(*)$-integrable function is D-integrable, and a D-integrable function that is almost everywhere nonnegative is Lebesgue integrable.

E. Constructive Definition of Integrals

Let S be a functional whose domain $\bigcup_I K(S;I)$ consists of the union of sets $K(S;I)$ of real-valued functions defined on closed intervals $I=[a,b]$. If f belongs to $K(S;I)$, we denote the value $S(f)$ by $S(f;I)$. Such a functional S is called an **integral operator** if the following three conditions are satisfied: (1) If $f\in K(S;I_0)$ and I is an arbitrary interval contained in I_0, then the †restriction f_I to I of f also belongs to $K(S;I)$. Also, $S(f;I)$ is a †continuous additive function of the interval $I\subset I_0$. (2) Let $I_1=[a,b]$, $I_2=[b,c]$, and $I=[a,c]$ $(a<b<c)$. If for a function f defined on I, $f_1\in K(S;I_1)$ and $f_2\in K(S;I_2)$, where $f_1=f_{I_1}$ and $f_2=f_{I_2}$, then $f\in K(S;I)$. (3) If f is identically 0 on I, then $f\in K(S;I)$ and $S(f;I)=0$. For two integral operators S_1 and S_2, we say that S_2 includes S_1 (or S_1 is weaker than S_2) if $K(S_1;I)\subset K(S_2;I)$ for every I and $S_1(f;I)=S_2(f;I)$ for every $f\in K(S_1;I)$. The D-integral (D$(*)$-integral) is the weakest integral operator containing the Lebesgue integral and satisfying the following two conditions, (C) and (H) (resp. H$(*)$): (C) **Cauchy's condition**. If, for every function f defined on I_0, we have $f_I\in K(S;I)$ for any $I=[a+\delta, b-\varepsilon]\subsetneqq I_0=[a,b]$, and also if the finite limit $\lim_{\delta\to 0,\varepsilon\to 0} S(f;I)$ exists, then $f\in K(S;I_0)$ and $S(f;I_0)$ coincides with the foregoing limit value. (H) **Harnack's condition**. Let E be a closed subset of I_0, $\{I_k\}$ be a sequence of inter-

vals contiguous to the set consisting of the points of E and the endpoints of I_0, and f be a function on I_0 satisfying the following three conditions: (i) $f_E \in K(S; I_0)$, where $f_E(x) = f(x)$ whenever $x \in E$ and $f_E = 0$ otherwise; (ii) $f_k = f_{I_k} \in K(S; I_k)$ for each k; and (iii) $\Sigma_k |S(f_k; I_k)| < +\infty$ and $\lim_{k \to \infty} O(S; f_k; I_k) = 0$ when the sequence $\{I_k\}$ is infinite. Then it follows that $f \in K(S; I_0)$ and $S(f; I_0) = S(f_E; I_0) + \Sigma_k S(f_k; I_k)$. (Here $O(S; f_k; I_k)$ denotes the **variation** of $S(f_k)$ on I_k, that is, the least upper bound of the numbers $|S(f_J; J)|$, where J denotes any subinterval of I_k.) We obtain condition (H(*)) by replacing condition (iii) in (H) with a more restrictive condition: $\Sigma_k O(S; f_k; I_k) < +\infty$. The constructive definition of the Denjoy integral in the wide sense (the Denjoy integral in the restricted sense) is obtained by †transfinite induction starting with the Lebesgue integral and using two methods, (C) and (H) (resp. (H(*))), of extensions.

F. Perron Integrals

Given a function $f(x)$ defined on an interval $[a, b]$, suppose that $F(x)$ is a function defined on the same interval such that (1) $\underline{F}(x) \geqslant f(x)$; (2) $\underline{F}(x) \neq -\infty$ (resp. (1′) $\bar{F}(x) \leqslant f(x)$; (2′) $\bar{F}(x) \neq +\infty$) at every point x, where $\underline{F}(x)$ (resp. $\bar{F}(x)$) denotes the †lower (upper) derivative of $F(x)$. In this case, $F(x)$ is called a **major (minor) function** of $f(x)$. If for any $\varepsilon > 0$ there is a major function $\psi(x)$ and a minor function $\varphi(x)$ of $f(x)$ such that $\psi(b) - \varphi(b) < \varepsilon$, then $f(x)$ is said to be **Perron integrable**. We denote by $(P)\int_a^b f(x)\,dx$ the value $\inf_\psi\{\psi(b) - \psi(a)\} = \sup_\varphi\{\varphi(b) - \varphi(a)\}$.

G. Properties of Integrals

If $\{f_n\}$ is a nondecreasing sequence of functions that are D-integrable on an interval $[a, b]$ and whose D-integrals over $[a, b]$ constitute a sequence bounded from above, then the function $f(x) = \lim_{n \to \infty} f_n(x)$ is D-integrable on $[a, b]$, and we have

$$(D)\int_a^b f(x)\,dx = \lim_{n \to \infty} (D)\int_a^b f_n(x)\,dx.$$

If $F(x)$ is a function of †bounded variation and $g(x)$ is a D-integrable function on an interval $[a, b]$, then $F(x)g(x)$ is D-integrable on $[a, b]$; moreover, denoting by $G(x)$ the indefinite D-integral of $g(x)$, the following formula is valid:

$$(D)\int_a^b F(x)g(x)\,dx$$

$$= G(b)F(b) - G(a)F(a) - \int_a^b G(x)\,dF(x),$$

where the last term is the †Stieltjes integral **(integration by parts)**.

If $F(x)$ is a nondecreasing function and $g(x)$ is D-integrable on $[a, b]$, there is a point ξ in $[a, b]$ for which the following formula is valid:

$$(D)\int_a^b g(x)F(x)\,dx$$

$$= F(a) \cdot (D)\int_a^\xi g(x)\,dx + F(b) \cdot (D)\int_\xi^b g(x)\,dx$$

(the second mean value theorem).

The foregoing theorems remain valid if D is replaced by $D(*)$ in the hypotheses and conclusions.

References

[1] S. Saks, Theory of the integral, Warsaw, 1937.

[2] R. D. James, Integrals and summable trigonometric series, Bull. Amer. Math. Soc., 61 (1955), 1–15.

[3] R. Henstock, Theory of integration, Butterworth, 1963.

[4] Yu. S. Ochan, Generalized integrals (in Russian), Mat. Sb., 28 (70), no. 2 (1951), 293–336.

[5] K. Kunugui, Sur une généralisation de l'intégrale, Fundamental and Applied Aspects of Math., 1 (1959), 1–30.

[6] H. Okano, Sur une généralisation de l'intégral (E. R.) et un théorème général de l'intégration par parties, J. Math. Soc. Japan, 14 (1962), 430–442.

[7] S. Kempisty, Sur les fonctions absolument continue d'intervalle, Fund. Math., 27 (1936), 10–37.

[8] S. Enomoto, Sur une totalisation dans les espaces de plusieurs dimensions I, II, Osaka Math. J., 7 (1955), 69–102, 157–187.

101 (XXI.19)
Descartes, René

René Descartes, (March 31, 1596–February 11, 1650), philosopher, mathematician, and natural scientist, was born in the province of Touraine in France. He became dissatisfied with his studies of scholastic philosophy in the Jesuit Academy in La Flèche, and later, in 1619, while stationed in Ulm during a tour of duty in the army, he underwent a philosophical conversion. He had an idea of methodologically unifying the various fields of interest to him using mathematics as a model. He returned to Paris in 1621, but moved to Holland in 1628 to concentrate on his work. Swe-

den's Queen Christina invited him in 1649 to that country, where he died the next year, evidently from a combination of cold and overwork.

Descartes, often considered the founder of modern philosophy, discarded early the traditional theological view of the world and stated that all knowledge should be recognized as logical only after it has been submitted to rational criticism. This ushered in the modern view of the world based on mathematics and physics. In 1637, he published *Géométrie* as an appendix to his *Discours de la méthode*, which also contained his works on optics and meteorology. In it, he promoted F. †Viète's symbolic algebra, which he applied to geometric problems. His idea that algebra could be used as a general method for geometry established him as the founder of †analytic geometry.

References

[1] R. Descartes, Oeuvres I–XII, C. Adam, P. Tannery, and Léopold Cerf (eds.), 1897–1910.
[2] H. Lefèbvre, Descartes, Paris, 1947.

102 (XVIII.12) Design of Experiments

A. General Remarks

The **design of experiments** is a part of the **statistical planning** required to collect the data appropriate to the purpose of statistical inference (→ 401 Statistical Inference) in various fields of scientific research and application. The main purposes of the design of experiments are (1) to analyze a given statistical linear model (→ 403 Statistical Models) and (2) to devise a "good" statistical linear model. Sometimes this term also refers to a statistical method including the †analysis of variances. Thus the purpose of designing an experiment is to provide the most efficient and economical methods of reaching valid and relevant conclusions from that experiment (→ 403 Statistical Models).

R. A. Fisher, whose contributions to statistical theory were remarkable and far-ranging, propounded three required principles to control the experimental field in order to guarantee the validity of statistical methods and to increase the sensitivity of experiments: (i) **replication**, or the repetition of the set of all the phenomena to be compared in the experiment, for the evaluation of experimental error variance; (ii) **randomization**, or the procedure allotting various experimental treatments at random, in order to change systematic errors into random errors; and (iii) **local control**, or the procedure which makes the variation within each experimental block as small as possible, in order to minimize or remove systematic errors. These are called **Fisher's three principles**. A design satisfying principles (i) and (ii) is called a **completely randomized design**; it enables us to attach a probability statement to estimated treatment differences by obtaining valid estimate of experimental error variance.

Let an n-dimensional †random variable $\mathbf{Y}=(Y_1,\ldots,Y_n)'$ be represented by a linear model

$$\mathbf{Y}=X\xi+\mathbf{W}, \tag{1}$$

where X is a given $n\times s$ real matrix, $\xi=(\xi_1,\ldots,\xi_s)'$ is an s-vector, and $\mathbf{W}=(W_1,\ldots,W_n)'$ is a random vector with the †expectation $E(\mathbf{W})=\mathbf{0}$. Then $\mathbf{Y}$ is called the **observation vector,** $\mathbf{W}$ the **error vector,** ξ the **effect vector** of $\mathbf{Y}$, and X the **design matrix**.

According to the properties of the effect vector ξ, the linear model (1) is separated into three classes: (i) The class of **fixed-effects models** for which ξ is a fixed unknown parameter. In this case, the component ξ_i of ξ is called a **fixed effect**, and a linear function $\pi=\mathbf{F}'\xi$ of ξ, with a given coefficient vector $\mathbf{F}$ is called a **linear parameter** or **parametric function**. (ii) The class of **random-effects models** for which the components Ξ_i of ξ are random variables. In this case, each component Ξ_i is called a **random effect**, and ξ is denoted by $\mathbf{\Xi}$. (iii) The class of **mixed models** for which there are both fixed effects ξ_i and random effects Ξ_j in ξ. In this case, the model (1) becomes

$$\mathbf{Y}=X_1\xi^1+X_2\mathbf{\Xi}^2+\mathbf{W}, \tag{2}$$

where $\xi^1=(\xi_1,\ldots,\xi_r)'$ is a fixed-effect vector and $\mathbf{\Xi}^2=(\Xi_{r+1},\ldots,\Xi_s)'$ is a random-effect vector. The conditions frequently assumed for the †distribution law of $\mathbf{Y}$ are: (a) The errors W_i $(i=1,\ldots,n)$ are uncorrelated and $E(W_1)=\ldots=E(W_n)=0$. (b) The errors W_i $(i=1,\ldots,n)$ have a common unknown †variance σ^2. (c) The errors W_i $(i=1,\ldots,n)$ have the †normal distribution. (d) The random effects Ξ_j are uncorrelated and independent of the error vector $\mathbf{W}$. (e) The random effects Ξ_j have a common unknown variance σ_1^2. (f) The random effects Ξ_j have the normal distribution.

Let $L(X)$ be a †linear subspace of $\mathbf{R}^n$ spanned by the column vectors of X. The linear model

$$\mathbf{Y}=\tilde{X}\xi+\mathbf{W} \tag{3}$$

is called a hypothesis on the linear model (1) if $L(\tilde{X})\subset L(X)$.

The main issues of the theory of design of experiments are concerned with (I) statis-

tical inferences, such as estimation or testing hypotheses, under models (1), (2), (3), (i), (ii), (iii); (II) determination of the design matrix X satisfying certain optimal conditions; (III) construction of a theoretical foundation that can explain the validity of the statistical treatment of the observed data by means of the above models.

B. Block Design

The design of experiments is described here in terms of the so-called **block design**. There are n experimental units $\alpha=1,\ldots,n$ called **plots**, and an observation Y_α is assigned to each plot α. A **block** is constructed with several plots under Fisher's principle (iii), and the number of plots in a block is called the **block size**, the jth one being denoted by $k_j, j=1,\ldots,b$, with $\sum_j k_j=n$. One of v operations, called **treatments** or **varieties**, is applied to each plot. It is assumed that the observation Y_α at the plot α in the jth block under the ith treatment has the structure

$$Y_\alpha=\xi_i+\eta_j+W_\alpha.$$

The ξ_i, $i=1,\ldots,v$, are called **treatment effects**, and the $\eta_j, j,\ldots,b$, **block effects**. It is also assumed that $\sum_i\xi_i=0$. In this case, $\mathbf{Y}$ is represented in matrix notation as

$$\mathbf{Y}=\Phi\xi+\Psi\boldsymbol{\eta}+\mathbf{W}, \tag{4}$$

where $\Phi=(\varphi_{\alpha i})$, $\alpha=1,\ldots,n$, $i=1,\ldots,v$, with

$$\varphi_{\alpha i}=\begin{cases}1 & \text{when the } i\text{th treatment is applied to the plot } \alpha,\\ 0 & \text{otherwise,}\end{cases}$$

and $\Psi=(\psi_{\alpha j})$, $\alpha=1,\ldots,n, j=1,\ldots,b$, with

$$\psi_{\alpha j}=\begin{cases}1 & \text{when the plot } \alpha \text{ belongs to the } j\text{th block,}\\ 0 & \text{otherwise.}\end{cases}$$

Here it is assumed that $\sum_i\varphi_{\alpha i}=1$, $\sum_\alpha\varphi_{\alpha i}=r_i\geqslant 1$, $\sum_i r_i=n$, $\sum_j\psi_{\alpha j}=1$, and $\sum_\alpha\psi_{\alpha j}=k_j\geqslant 1$. We call r_i the **number of replications** of the ith treatment. We set $N=(n_{ij})=\Phi'\Psi$. Then n_{ij} is the number of observations in the jth block to which the ith treatment is applied. The matrix N is called the **incidence matrix** of the block design.

In any experiment, each plot has its own effect. The blocks are constructed so that this plot effect in each block becomes as homogeneous as possible, although it is impossible to eliminate the effect completely. For this purpose, randomization is adopted. Suppose that we are given k plots in a block and v treatments ($k\leqslant v$). Then randomization is utilized to select a treatment out of v treatments to be allocated to each plot so that the selection is "at random." Then the plot effects are random, and the error term in model (4) can be considered to be the sum of a plot effect and an original error.

A block design satisfying Fisher's three principles is called a **randomized block design**. Blocks that can accommodate all the treatments to be studied are called **complete blocks**. Those that do not contain all the treatments are called **incomplete blocks**. Blocking can be considered to be an extension of pairing. In the terminology of block design, we can say that many experiments are block experiments, or even (tautologically, if we allow those with only one block) that all experiments are block experiments.

In a block design $N=(n_{ij})$, treatments i_0 and i_l are said to be **connected** if there exists a chain

$$i_0 j_1 i_1 j_2\ldots i_{l-1} j_l i_l$$

of integers such that $1\leqslant i_p\leqslant v$ $(p=0,1,2,\ldots,l)$, $1\leqslant j_q\leqslant b$ $(q=1,2,\ldots,l)$, and $n_{i_0j_1}>0$, $n_{i_1j_1}>0$, $n_{i_1j_2}>0,\ldots,n_{i_{l-1}j_l}>0$, $n_{i_lj_l}>0$. If all pairs of treatments are mutually connected, then the design is said to be **connected**. In this case, the rank of the matrix C defined in Section C is $v-1$. If the design is disconnected, then the incidence matrix N can be partitioned into two or more connected portions, e.g.,

$$N=\begin{pmatrix}N_1 & 0\\ 0 & N_2\end{pmatrix}.$$

Thus without loss of generality we can restrict ourselves to the connected case.

C. Estimation under the Fixed-Effects Model

Conditions (a) and (b) of Section A are assumed here. The †normal equation which gives the †least square estimates $\hat{\xi}$ and $\hat{\boldsymbol{\eta}}$ of ξ and $\boldsymbol{\eta}$, respectively, is

$$\begin{pmatrix}\Phi'\\ \Psi'\end{pmatrix}(\Phi,\Psi)\begin{pmatrix}\hat{\xi}\\ \hat{\boldsymbol{\eta}}\end{pmatrix}=\begin{pmatrix}\Phi'\\ \Psi'\end{pmatrix}\mathbf{Y}. \tag{5}$$

Set $\Phi'\Phi=\operatorname{diag}(r_1,\ldots,r_v)=D_r$, $\Psi'\Psi=\operatorname{diag}(k_1,\ldots,k_b)=D_k$, $C=D_r-ND_k^{-1}N'$, $\mathbf{Q}=(\Phi'-ND_k^{-1}\Psi')\mathbf{Y}$, where diag(...) means a diagonal matrix with the diagonal elements .... Then (5) reduces to

$$C\hat{\xi}=\mathbf{Q},\qquad \hat{\boldsymbol{\eta}}=D_k^{-1}(\Psi'\mathbf{Y}-N'\hat{\xi}). \tag{6}$$

Let L be an orthogonal matrix that transforms the matrix C to a diagonal form; that is, $L'CL=\operatorname{diag}(\rho_1,\ldots,\rho_{v-1},0)=\Lambda$, $\rho_i>0$ for all i. Set $\Lambda^*=\operatorname{diag}(\rho_1^{-1},\ldots,\rho_{v-1}^{-1},0)$, $C^*=L\Lambda^*L'$. Then $\hat{\xi}$

$= C^*\mathbf{Q}$ is a particular solution of (6) and $\sum_i \hat{\xi}_i = 0$. A parametric function $\pi = \mathbf{F}'\xi$ with coefficient vector $\mathbf{F} = (F_1, F_2, \ldots, F_v)'$ is called a **treatment contrast** if the sum $\sum_i F_i$ of coefficients vanishes. A treatment contrast $\pi = \mathbf{F}'\xi$ is called a **normalized contrast** if $\mathbf{F}'\mathbf{F} = 1$. It is called the **elementary contrast** if $\mathbf{F}$ has only two nonzero elements 1 and -1. Elementary contrasts of treatment effects show the comparison of treatments involved in them. When a design is connected, any contrast π is †estimable, and the †best linear unbiased estimate of π is $\hat{\pi} = \mathbf{F}'\hat{\xi}$. Furthermore, if $\mathbf{f}_i$ is the eigenvector of the matrix C with unit length corresponding to an eigenvalue ρ_i and $\mathbf{F} = \sum_i a_i \mathbf{f}_i$, then the variance of the estimate $\hat{\pi}$ is given by $\sigma^2 \sum_i a_i^2/\rho_i$. The following properties are equivalent: (i) A design is connected. (ii) Any treatment contrast is estimable. (iii) The rank of the matrix C is $v-1$. (iv) The minimum eigenvalue 0 of $D_r^{-1/2} C D_r^{-1/2}$ is simple and other eigenvalues θ_1 satisfy $0 < \theta_1 \leqslant 1$. (v) The maximum eigenvalue 1 of $D_r^{-1/2} N D_k^{-1} N' D_r^{-1/2}$ is simple and other eigenvalues θ_2 satisfy $0 \leqslant \theta_2 < 1$. (vi) There exists a positive integer p such that each element of $(D_r^{-1/2} N D_k^{-1} N' D_r^{-1/2})^p$ is positive. It holds that $\theta_2 = 1 - \theta_1$.

D. Test of a Hypothesis $H: \xi_1 = \ldots = \xi_v$ in the Fixed-Effects Model

Conditions (a), (b), and (c) of Section A are assumed here. The hypothesis $H: \xi_1 = \ldots = \xi_v$ is represented by

$$\mathbf{Y} = \Psi \boldsymbol{\eta} + \mathbf{W}. \tag{7}$$

Consider a direct sum decomposition $\mathbf{R}^n = L(\Gamma) + L_\Gamma^\perp(\Psi) + L_\Psi^\perp(\Phi, \Psi) + L_{\Phi,\Psi}^\perp$ of $\mathbf{R}^n$, where $L_B^\perp(A)$ and $L_B^\perp$ stand for the †orthocomplements of $L(B)$ with respect to $L(A)$ and $\mathbf{R}^n$, respectively, and $\Gamma = (1, 1, \ldots, 1)' \in \mathbf{R}^n$. The †projection operator matrices for the decomposed subspaces $L(\Gamma)$, $L_\Gamma^\perp(\Psi)$, $L_\Psi^\perp(\Phi, \Psi)$, and $L_{\Phi,\Psi}^\perp$ are denoted by P_1, P_2, P_3, and P_4, respectively. Then we have

$$P_1 = n^{-1} E_{nn}, \qquad P_2 = \Psi D_k^{-1} \Psi' - n^{-1} E_{nn},$$

$$P_3 = (I_n - \Psi D_k^{-1} \Psi') \Phi C^* \Phi' (I_n - \Psi D_k^{-1} \Psi'),$$

$$P_4 = I_n - P_1 - P_2 - P_3,$$

where E_{ab} is an $a \times b$ matrix whose entries are all unity and I_n is the $n \times n$ identity matrix.

The analysis of variance for the hypothesis (7) in model (4) is given by

$$\mathbf{Y}'\mathbf{Y} = \mathbf{Y}'P_1\mathbf{Y} + \mathbf{Y}'P_2\mathbf{Y} + \mathbf{Y}'P_3\mathbf{Y} + \mathbf{Y}'P_4\mathbf{Y}.$$

This is called an **intrablock analysis**. A usual test for the hypothesis H is given by a critical region with

$$F = \frac{n-v-b+1}{v-1} \frac{\mathbf{Y}'P_3\mathbf{Y}}{\mathbf{Y}'P_4\mathbf{Y}} > \text{constant}$$

(→ 403 Statistical Models).

E. Optimal Block Design

A block design is said to be **optimal** when it minimizes the variance of the estimate $\hat{\pi}$ of a normalized contrast π. Suppose that the number v of treatments, the number b of blocks, and each block size $k_j, j = 1, \ldots, b$, are given. Under each of the following criteria the corresponding block design is optimal in the sense indicated: For positive eigenvalues, $\rho_1, \rho_2, \ldots, \rho_{v-1}$, of the matrix C in Section C, (I) $\prod_{i=1}^{v-1} \rho_i$ is maximal (D-**optimality**); (II) $\min \rho_i$ is maximal (E-**optimality**); (III) $\sum_{i=1}^{v-1} \rho_i^{-1}$ is minimal (or the average variance of the estimates of all normalized contrasts of the parameters ξ_i is minimal) (A-**optimality**).

If $\rho_1 = \ldots = \rho_{v-1} = (n-b)/(v-1)$ $(= \rho$, say) and n_{ij} is either 1 or 0, then the design is optimal for each of the optimality criteria (I), (II), and (III). In this case, we have

$$C = \rho(I_v - v^{-1} E_{vv}).$$

Such a design is called a **variance-balanced block design** in which every normalized contrast is estimable with the same variance. If $D_r^{-1} N D_k^{-1} N' - n^{-1} E_{v1} \mathbf{r}' = \mu(I_v - n^{-1} E_{v1} \mathbf{r}')$ (or $C = (1-\mu)(D_r - n^{-1}\mathbf{r}\mathbf{r}')$), every normalized contrast is estimable with the same efficiency $1-\mu$, where $r = D_r E_{v1}$ and $0 < \mu < 1$. Such a design is called an **efficiency-balanced block design**. For a block design, any two of the following properties imply the third: (i) The design is variance-balanced; (ii) the design is efficiency-balanced; (iii) the design is equireplicated.

When all block sizes k_j equal some number k independent of j, all numbers r_i of replications equal some number r independent of i, and $\lambda_{ii'} = \sum_j n_{ij} n_{i'j}$ (number of times that both treatments i and i' are applied to the same block) equals some number λ independent of i and i', then the design is combinatorially balanced. The design is usually called the **balanced incomplete block design (BIBD)** if these three conditions are fulfilled and $k < v$. A BIBD is both variance-balanced and efficiency-balanced. If all the treatments are replicated the same number of times and the blocks are of the same size in a block design, then the only variance-balanced design is a BIBD, provided such a design exists. A BIBD is often denoted by BIBD(v, b, r, k, λ) and we have the relations $vr = bk$, $\lambda(v-1) = r(k-1)$ and $v \leqslant b$, among the parameters; the last

relation is called **Fisher's inequality**. The design is said to be **symmetric** when $v=b$. Furthermore, if $v=b$ is even, then $r-\lambda$ must be a perfect square. If $v=b$ is odd, then the equation $x^2=(r-\lambda)y^2+(-1)^{(v-1)/2}\lambda z^2$ must have a solution in integers x, y, z not all zero. This is called the **Bruck-Ryser-Chowla theorem**. Necessary conditions for the existence of a BIBD have been obtained. One of these conditions is stated in terms of the Hasse-Minkowski p-invariant $C_p(A)=(-1,-1)_p\prod_{i=1}^{n}(D_i,-D_{i-1})_p$, where $(m,m')_p$ is the †Hilbert norm-residue symbol, $D_0=1$, and D_i is the principal minor of the $n\times n$ matrix A. Another is described in terms of the embedding of a quasiresidual design into the corresponding symmetric BIBD. No effective necessary and sufficient condition for the existence of a BIBD has been obtained. In general, it is conjectured that for a positive integer k, with finitely many exceptions, BIBD(v,b,r,k,λ)'s exist for all pairs v, λ of positive integers satisfying $v>k$, $\lambda(v-1)\equiv 0 \pmod{k-1}$ and $\lambda v(v-1)\equiv 0 \pmod{k(k-1)}$.

The known methods of constructing block designs are of two main types, direct and recursive. Recursive methods are a way of constructing designs from smaller ones. Direct methods yield easier constructions, but are applicable only for special values of the parameters. A direct method of constructing a BIBD designates the treatments and blocks, respectively, as the points and subspaces of the †projective space and the †affine space over a †finite field. To explain another method of constructing a BIBD, we let G be an †additive group of order n and $x_1^{(i)},\dots,x_m^{(i)}$ be m treatments corresponding to each element $x^{(i)}$ of the group $(i=1,2,\dots,n)$. The treatment $x_\alpha^{(i)}$ is said to belong to the αth class $(\alpha=1,2,\dots,m)$, and a pair $(x_\alpha^{(i)},x_\beta^{(j)})$ of treatments is called a **difference** of type $(\alpha,\beta,x^{(p)})$ if $x^{(i)}-x^{(j)}=x^{(p)}$ $(\neq 0)$. We can form t blocks of size k

$$B_1=\{x_{\alpha_1}^{(i_1)},\dots,x_{\alpha_k}^{(i_k)}\},\qquad \dots,$$

$$B_t=\{x_{\beta_1}^{(j_1)},\dots,x_{\beta_k}^{(j_k)}\},$$

such that each block B_i contains exactly r treatments belonging to the αth class $(\alpha=1,\dots,m)$ and among all pairs of treatments in the same block there are λ differences of each type $(\alpha,\beta,x^{(p)})$. Such a set of t blocks is called a **difference set**. The t blocks in a difference set are called **initial blocks**. Given such a difference set, we can obtain nt blocks by joining elements of G to the elements of each B_s $(s=1,2,\dots,t)$. These nt blocks form a BIBD$(v=mn,b=nt,r,k=rm/t,\lambda)$ (→ 66 Combinatorics).

As a generalization of Fisher's inequality, for an unequal-replicated block design with unequal block sizes, $b\geqslant v-\delta$ holds, where δ is the multiplicity of the maximum eigenvalue 1 of the matrix $D_r^{-1/2}CD_r^{-1/2}$. The equality sign holds if and only if the projection operator corresponding to zero eigenvalue of $D_r^{-1/2}CD_r^{-1/2}$ is a zero matrix.

F. Estimation in a Mixed Model

Consider a block design (4), where every block has the same size k and every treatment has the same number r of replications. Let ξ be a fixed effect and $\boldsymbol{\eta}$ a random effect denoted by $\mathbf{H}$, and assume that $\mathbf{W}$ satisfies conditions (a) and (b) of Section A and that $\mathbf{H}$ satisfies conditions (d), (e), and (f) (where the Ξ_j are replaced by the coordinates H_j of $\mathbf{H}$). If $E(H_j)=\gamma$, $j=1,\dots,b$, then, changing the notation $\mathbf{H}-E_{v1}\gamma$ to $\mathbf{H}$, we can rewrite (4) as

$$\mathbf{Y}=\Gamma\gamma+\Phi\xi+\Psi\mathbf{H}+\mathbf{W}, \tag{8}$$

with $E(\mathbf{H})=\mathbf{0}$. The normal equation that gives the least square estimate of ξ is

$$(C+\sigma^2(\sigma^2+k\sigma_1^2)^{-1}C_1)\hat{\xi}$$
$$=\mathbf{Q}+\sigma^2(\sigma^2+k\sigma_1^2)^{-1}\mathbf{Q}_1, \tag{9}$$

where C and $\mathbf{Q}$ are the same as in (6) and $C_1=ND_k^{-1}N'-rv^{-1}E_{vv}$, $\mathbf{Q}_1=(ND_k^{-1}\Psi'-v^{-1}E_{v1}\Gamma')\mathbf{Y}$. Equation (9) cannot be solved unless the ratio $\sigma^2:\sigma_1^2$ is given. When $\sigma^2:\sigma_1^2$ is not known, substituting in $\sigma^2+k\sigma_1^2$ its †unbiased estimate given by analysis of variance, one obtains a solution of (9) which tends to a †consistent estimate of ξ as the number of blocks tends to infinity.

G. Estimation in a Random-Effects Model

Let ξ (denoted by Ξ) and $\mathbf{H}$ in the model (8) be random effects. Suppose that Ξ, $\mathbf{H}$, and $\mathbf{W}$ are mutually independent, and that the distributions of Ξ, $\mathbf{H}$, and $\mathbf{W}$ are $N(\mathbf{0},\sigma_2^2I_v)$, $N(\mathbf{0},\sigma_1^2I_b)$, and $N(\mathbf{0},\sigma^2I_n)$, respectively. The distribution of $\mathbf{Y}$ in (8) contains four parameters $\gamma,\sigma^2,\sigma_1^2,\sigma_2^2$. When $k_j<v$, the †minimal sufficient statistic is generally incomplete, and therefore the optimal estimate of γ, σ^2, σ_1^2, and σ_2^2 cannot be determined. As an example, the minimal sufficient statistic for the random-effects model of a BIBD(v,b,r,k,λ) is

$$(\textstyle\sum_i Y_i,\mathbf{Y}'P_{21}\mathbf{Y},\mathbf{Y}'P_{22}\mathbf{Y},\mathbf{Y}'P_3\mathbf{Y},\mathbf{Y}'P_4\mathbf{Y},\mathbf{Y}'P_5\mathbf{Y}),$$

where

$$P_{21}=k^{-1}(r-\lambda)^{-1}BTB-kr(r-\lambda)^{-1}n^{-1}E_{nn},$$

$$P_{22}=k((k-1)r+\lambda)^{-1}r^{-1}$$
$$\times(T-k^{-1}BT)(T-k^{-1}TB),$$

$$P_3=k^{-1}B-k^{-1}(r-\lambda)^{-1}BTB$$
$$+v\lambda(r-\lambda)^{-1}n^{-1}E_{nn},$$

$$P_4=I_n-k^{-1}B-P_{22},\qquad P_5=r^{-1}T-n^{-1}E_{nn},$$

with $T=\Phi\Phi'$ and $B=\Psi\Psi'$. In this case, $E(\mathbf{Y}'P_4\mathbf{Y})=(n-v-b+1)\sigma^2$, $E(\mathbf{Y}'P_3\mathbf{Y})=(b-v)(\sigma^2+k\sigma_1^2)$. From these equations unbiased estimates of σ^2 and σ_1^2 can be derived, but their optimality is not guaranteed (→ 396 Statistic).

H. Factorial Experiments

Suppose that there are h **factors** $F_1,\ldots,F_h$ which affect $\mathbf{Y}$, and each factor F_1 has s_i **levels** $(i=1,\ldots,h)$. It is assumed that $v=s_1\times s_2\times\ldots\times s_h$ treatment combinations are derived by all the combinations of the levels of h factors, and that v treatment effects are represented by the sum of subeffects called **main effects** and **interactions**, or **factorial effects** covering both. Such an experiment is called a **factorial experiment** or **factorial design**; it allows us to examine the effects of two or more factors, each factor being applied at two or more levels, by testing all possible treatment combinations formed from the factors under study. Specifically, suppose that we have the case $h=2$, called a **two-way layout**. Let main effects be denoted by $\xi^1=(\xi_1^1,\ldots,\xi_{s_1}^1)'$, $\xi^2=(\xi_1^2,\ldots,\xi_{s_2}^2)'$, and interaction by $\xi^{12}=(\xi_{11}^{12},\xi_{12}^{12},\ldots,\xi_{s_1s_2}^{12})'$, where $\sum_i\xi_i^1=0$, $\sum_j\xi_j^2=0$, $\sum_i\xi_{ij}^{12}=\sum_j\xi_{ij}^{12}=0$. When there is no restriction on the number of observations, each of $v=s_1\times s_2$ treatment combinations is replicated t times. Components Y_{ijk} of the observation vector $\mathbf{Y}$ are represented by a linear model

$$Y_{ijk}=\gamma+\xi_i^1+\xi_j^2+\xi_{ij}^{12}+W_{ijk},$$

$$i=1,\ldots,s_1;\quad j=1,\ldots,s_2;\quad k=1,\ldots,t,$$

or, in vector notation,

$$\mathbf{Y}=\mathbf{\Gamma}\gamma+X_1\xi^1+X_2\xi^2+X_{12}\xi^{12}+\mathbf{W}.$$

The analysis of variance in this case is given by

$$\mathbf{Y}'\mathbf{Y}=\sum_i\mathbf{Y}'P_i\mathbf{Y},$$

where P_i is the projection operator matrix on the subspace derived by a direct sum decomposition

$$\mathbf{R}^n=L(\mathbf{\Gamma})+L_{\Gamma}^{\perp}(X_1)+L_{\Gamma}^{\perp}(X_2)+L_{X_1,X_2}^{\perp}(X_{12})$$
$$+L_{X_{12}}^{\perp},$$

where $n=s_1s_2t$. Denoting by $\bar{Y}_{ij\cdot}$ the arithmetic mean of Y_{ijk} over the subscript k and using similar notation $\bar{Y}_{i\cdot\cdot}$, $\bar{Y}_{\cdot j\cdot}$, and $\bar{Y}_{\cdot\cdot\cdot}$, we have $\mathbf{Y}'P_1\mathbf{Y}=n\bar{Y}_{\cdot\cdot\cdot}^2$, $\mathbf{Y}'P_2\mathbf{Y}=s_2t\sum(\bar{Y}_{i\cdot\cdot}-\bar{Y}_{\cdot\cdot\cdot})^2$, $\mathbf{Y}'P_3\mathbf{Y}=s_1t\sum(\bar{Y}_{\cdot j\cdot}-\bar{Y}_{\cdot\cdot\cdot})^2$, $\mathbf{Y}'P_4\mathbf{Y}=t\sum(\bar{Y}_{ij\cdot}+\bar{Y}_{\cdot\cdot\cdot}-\bar{Y}_{i\cdot\cdot}-\bar{Y}_{\cdot j\cdot})^2$, $\mathbf{Y}'P_5\mathbf{Y}=\sum(Y_{ijk}-Y_{ij\cdot})^2$. The analysis of variance table is given in Table 1. For $h\geqslant 3$ similar models can also be considered for dealing with interactions up to h-factor. The factorial experiment is said to be **symmetric** if $s_1=s_2=\ldots=s_h=s$ and is called the s^h **factorial experiment**. Otherwise it is said to be **asymmetric** and is called the $s_1\times s_2\times\ldots\times s_h$ **factorial experiment**.

Table 1 Analysis of Variance

Source	Sum of Squares	Degrees of Freedom
F_1 main effects	$\mathbf{Y}'P_2\mathbf{Y}$	s_1-1
F_2 main effects	$\mathbf{Y}'P_3\mathbf{Y}$	s_2-1
F_1F_2 interactions	$\mathbf{Y}'P_4\mathbf{Y}$	$(s_1-1)(s_2-1)$
Error	$\mathbf{Y}'P_5\mathbf{Y}$	$s_1s_2(t-1)$

I. Fractional Factorial Designs

Factorial designs require at least an experiment for all the combinations of levels of factors under consideration. When there are a large number of treatment combinations resulting from a large number of factors to be tested, it will often be beyond the resources of the investigator to experiment with all of them. For such cases Finney (*Ann. Eugen.*, 12 (1945)) proposed a method in which only a fraction of the treatment combinations are experimented with. A design of this type is called a **fractional factorial design**. One reason for the usefulness of fractional factorial designs in preference to factorial designs is that they involve a smaller number of treatment combinations, since, in most scientific experiments, it is usually found that a large number of the higher-order interactions are negligible. The crucial part of the specification of fractional factorial designs is the suitable choice of the defining or identity relationship. Equating the nonestimable factorial effects for the selected fraction of treatment combinations with I gives the **identity relation**. Each factorial effect is not estimable after selecting a fraction of treatment combinations, and any contrast of the selected treatment combinations represents more than one factorial effect. All factorial effects represented by the same treatment combinations are called **aliases**. In aliases, by assuming that other interactions are negligible when compared to the one of interest, estimation can be made by means of the corresponding contrast of the selected treatment combinations. When all factorial effects of order higher than l are assumed to be zero, a fractional factorial design is said to be of **resolution** $2l+1$ if it satisfies the condition that under the usual model all factorial effects up to order l are estimable, whereas a fractional factorial design is said to be of **resolution** $2l$ if it satisfies the condition that under the usual model all factorial effects up to order $l-1$ are estimable.

In the beginning, the theory was developed

for an **orthogonal fractional factorial design** in which the estimates of various effects of interest are all uncorrelated. However, these are available only for special values of the number n of treatment combinations and are in general uneconomical in the sense that they require a large value of n in comparison with the number of unknown effects. As a generalization of orthogonal fractional factorial designs, I. M. Chakravarti (*Sankhyā*, 17 (1956)) introduced the concept of **balanced fractional factorial designs**. In these designs, the estimates of effects are not always uncorrelated. However, the †covariance matrix of the estimates is invariant under any permutation of factors. Balanced fractional factorial designs are flexible in the number of treatment combinations, with the result that more experimental situations can be handled. These two kinds of fractional factorial designs can be constructed by using orthogonal arrays and balanced arrays, defined in Section L. All alias relations can be derived from the identity relation in orthogonal designs.

In the model (1) for a design, the variance of the estimates of estimable linear functions of ξ depends on the matrix $X'X$ which is called the **information matrix**. For eigenvalues of $X'X$, the D-, E-, and A-optimalities of fractional factorial designs can be defined similarly to (I), (II), and (III) given in Section E. An orthogonal fractional 2^h factorial design is D-, E- and A-optimal. A fractional factorial design is said to be **saturated** if the number of treatment combinations is equal to that of parameters in the model to be estimated. However, since saturated designs do not provide an estimate of experimental error, their use should, in general, be confined to those situations in which a prior estimate of experimental error is available. The existence of a symmetric BIBD$(4t-1, 4t-1, 2t-1, 2t-1, t-1)$ implies that of an orthogonal saturated fractional 2^{4t-1} factorial design of resolution III.

J. Application of Algebras

In the theory of design of experiments, the ideas of association algebra and relationship algebra play an important role. Let A_i be a $v \times v$ symmetric matrix with entries 0 or 1 ($i = 0, 1, \ldots, m$). If a set $\{A_i \mid i = 0, 1, \ldots, m\}$ satisfies the conditions

$$A_0 = I_v, \qquad \sum_{i=0}^{m} A_i = E_{vv},$$

and there is a nonnegative integer p_{jk}^i for every i, j, k such that

$$A_j A_k = \sum_{i=0}^{m} p_{jk}^i A_i,$$

then the A_i are called **association matrices**. In this case, there exists a natural number n_i for every i such that $A_i E_{vv} = n_i E_{vv}$ holds. If an (α, β) component $a_{\alpha i}^{\beta}$ of A_i is unity, then the treatments α and β are said to be the ith associates. The †algebra $\mathscr{A}$ over the real number field generated by the matrices $A_0, A_1, \ldots, A_m$ is called the **association algebra**. $\mathscr{A}$ is commutative and †completely reducible. $A_k \to \mathscr{P}_k = (p_{jk}^i)$ is the †regular representation of $\mathscr{A}$. There is a nonsingular matrix $U = (u_{ij})$ that transforms all $\mathscr{P}_i$ into diagonal matrices simultaneously:

$$U \mathscr{P}_i U^{-1} = \operatorname{diag}(z_{0i}, \ldots, z_{mi}), \qquad i = 0, 1, \ldots, m.$$

$$A_i^{\#} = \left(\sum_{f=0}^{m} u_{if} z_{if} \right)^{-1} \sum_{j=0}^{m} u_{ij} A_j, \qquad i = 0, 1, \ldots, m,$$

are mutually orthogonal idempotent elements of $\mathscr{A}$. For example, consider the case where $A_0 = I_{s_1} \otimes I_{s_2}$, $A_1 = (E_{s_1 s_1} - I_{s_1}) \otimes I_{s_2}$, $A_2 = I_{s_1} \otimes (E_{s_2 s_2} - I_{s_2})$, $A_3 = (E_{s_1 s_1} - I_{s_1}) \otimes (E_{s_2 s_2} - I_{s_2})$ with $m = 3$, where $\otimes$ means the †Kronecker product. The algebra $\mathscr{A}$ generated by A_0, A_1, A_2, A_3 is called an F_2 type association algebra. The association matrices A_0, A_1, A_2, and A_3 correspond to the relationships between the treatments in a two-way layout. The mutually orthogonal idempotent elements in this case are $A_0^{\#} = s_1^{-1} E_{s_1 s_1} \otimes s_2^{-1} E_{s_2 s_2}$, $A_1^{\#} = (I_{s_1} - s_1^{-1} E_{s_1 s_1}) \otimes s_2^{-1} E_{s_2 s_2}$, $A_2^{\#} = s_1^{-1} E_{s_1 s_1} \otimes (I_{s_2} - s_2^{-1} E_{s_2 s_2})$, and $A_3^{\#} = (I_{s_1} - s_1^{-1} E_{s_1 s_1}) \otimes (I_{s_2} - s_2^{-1} E_{s_2 s_2})$. For the factorial experiment with h factors ($h \geqslant 3$), association matrices can be constructed in a similar way. If $h = 3$, the number m of association matrices is 7. Many types of association schemes are known, and some of them for $m = 2$ are classified as group divisible, triangular, Latin square, cyclic type, and so on. For a group divisible type of $v = s_1 s_2$, $A_0 = I_v$, $A_1 = I_{s_1} \otimes E_{s_2 s_2} - A_0$, $A_2 = E_{vv} - A_0 - A_1$; $A_0^{\#} = v^{-1} E_{vv}$, $A_1^{\#} = v^{-1} \{(s_1 - 1)(A_0 + A_1) - A_2\}$, $A_2^{\#} = s_2^{-1} \{(s_2 - 1) A_0 - A_1\}$.

An experimental design consists of a set of n experimental units called plots. Define a relationship R between the plots as a set of ordered pairs (i, j) of plots. A relationship R among a set of n plots can be expressed as a symmetric $n \times n$ matrix (r_{ij}) of 0's and 1's:

$$r_{ij} = \begin{cases} 1 & \text{if the } i\text{th plot is related to the } j\text{th} \\ & \text{plot by the relationship } R, \\ 0 & \text{otherwise,} \end{cases}$$

and this matrix is also denoted by R. If there are k types of relationships $R_1, \ldots, R_k$ among n plots, the algebra $\mathscr{R}$ over the real number field generated by the matrices $R_1, \ldots, R_k$ is called the **relationship algebra** of the design. $\mathscr{R}$ is a †semisimple algebra.

Example (1). The relationship algebra $\mathscr{R}$ of the factorial experiment with $h = 2$ and t replications is generated by the following matrices

over the real number field: $B_1 = A_0 \otimes I_t$, $B_2 = A_0 \otimes (E_{tt} - I_t)$, $B_3 = A_1 \otimes E_{tt}$, $B_4 = A_2 \otimes E_{tt}$, $B_5 = A_3 \otimes E_{tt}$, where A_i, $i = 0, 1, 2, 3$, are the association matrices of F_2 type. The mutually orthogonal idempotents that correspond to the †two-sided ideal decomposition of $\mathscr{R}$ are $B_1^\# = A_0^\# \otimes t^{-1}E_{tt}$, $B_2^\# = A_1^\# \otimes t^{-1}E_{tt}$, $B_3^\# = A_2^\# \otimes t^{-1}E_{tt}$, $B_4^\# = A_3^\# \otimes t^{-1}E_{tt}$, and $B_5^\# = A_0 \otimes (I_t - t^{-1}E_{tt})$. These are the same as the projection P_i in the two-way layout design given in Section H.

Example (2). Consider a block design with v treatments, each having the same number r of replications, and with b blocks, each having the same size k ($<v$). Suppose that association matrices A_i, $i = 0, 1, \ldots, m$, are given, by which the associations among the treatments are defined. Let $\lambda^i_{\alpha\beta}$ be the number of blocks to which the ith associate treatments α and β are applied. The design is called the **partially balanced incomplete block design (PBIBD)** if $\lambda^i_{\alpha\beta} = \lambda_i \geqslant 0$ independently of α and β. When $m = 1$, the design is a BIBD. Let the observation vector be represented by (4). The relationship algebra $\mathscr{R}$ of a PBIBD is generated by $n \times n$ matrices I_n, E_{nn}, $B = \Psi\Psi'$, $T_i = \Phi A_i \Phi'$, $i = 1, 2, \ldots, m$, where $NN' = \sum_{i=0}^m \lambda_i A_i = \sum_{i=0}^m \rho_i A_i^\#$, $\rho_i = \sum_{j=0}^m \lambda_j z_{ij}$ $(0 \leqslant \rho_i \leqslant rk;\ i = 0, 1, \ldots, m)$, and $\rho_0 = rk = \sum_{i=0}^m n_i \lambda_i$. Write $T_i^\# = \Phi A_i^\# \Phi'$. According as $\rho_i = rk$, $0 < \rho_i < rk$, or $\rho_i = 0$, $L(T_i^\#)$ is said to be **confounded** with the blocks, **partially confounded** with the blocks, or **orthogonal** to the blocks. $\mathscr{R}$ is noncommutative, completely reducible, and isomorphic to the algebra of matrices of the type shown in Fig. 1.

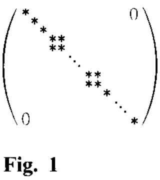


Fig. 1

The analysis of variance of a PBIBD is given by a decomposition of I_n into mutually orthogonal idempotent elements of $\mathscr{R}$. For a PBIBD, the matrix C in (6) is of the form $C = \sum_i \tau_i A_i^\#$, where $\tau_i = r - k^{-1}\rho_i$. For a connected PBIBD, all $A_i^\# \xi$, $i = 1, 2, \ldots, m$, are estimable. If there exists a group divisible PBIBD($m = 2$) with $\lambda_2 = \lambda_1 + 1$, then it is E- and A-optimal.

K. Design for Two-Way Elimination of Heterogeneity

Consider a design with v treatments in a $u \times w$ rectangular block. The row effect and the column effect of this block are denoted by $\boldsymbol{\eta}$ and $\mathbf{v}$, respectively. Thus the observation vector $\mathbf{Y}$ is of the form

$$\mathbf{Y} = \Gamma\gamma + \Phi\xi + \Psi\boldsymbol{\eta} + \Pi\mathbf{v} + \mathbf{W}, \tag{10}$$

where the definitions of Γ, Φ, Ψ and Π are similar to those for block designs. Set $L = \Phi'\Pi$, $M = \Phi'\Psi$, $D_r = \Phi'\Phi$, $F = D_r - w^{-1}LL' - u^{-1}MM' + (uw)^{-1}LE_{uu}L'$. The matrix F plays a role similar to that of the matrix C in (6). When the rank of F equals $v - 1$, the design is said to be **connected.** If

$$F = \tau(I_v - v^{-1}E_{vv}), \tag{11}$$

then the design satisfies the optimum criteria (I), (II), and (III) given in Section E. When $u = w = v$ and (11) holds, the design is called the **Latin square**. When $u = v$ and (11) holds, the design is called the **Youden square**. When $u > v$ and (11) holds, the design is called the **Shrikhande square**. In a **Youden square design** treatments are grouped into replications in two different ways, i.e., rows and columns, where rows constitute a BIBD, whereas columns are complete blocks. The existence of a Youden square design is equivalent to that of a symmetric BIBD.

If the associations among v treatments are defined in terms of association matrices, the partially balanced design for **two-way elimination of heterogeneity** can be defined in a way similar to the PBIBD. In this case, the equation $F = \sum_i \tau_i A_i^\#$ holds, and if $\mathscr{A} = \{I_v, E_{vv} - I_v\}$, then (11) holds, hence the optimum criteria are fulfilled. The definition of the relationship algebra $\mathscr{R}$ of a partially balanced design for two-way elimination of heterogeneity is similar to that used for the PBIBD. $\mathscr{R}$ is isomorphic to the algebra of matrices shown in Fig. 2.

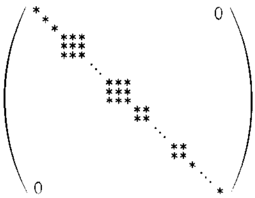


Fig. 2

The analysis of variance of this design is given by a decomposition of I_n into mutually orthogonal idempotent elements of $\mathscr{R}$.

L. Balanced Array and Orthogonal Array

Suppose that T is an $n \times h$ matrix with entries from a set A of s ($\geqslant 2$) distinct elements. Consider the s^t $1 \times t$ matrices $X = (x_1, x_2, \ldots, x_t)$ that can be formed by giving different values to the $x_i (\in A)$, $i = 1, 2, \ldots, t$. Suppose that associated with each $1 \times t$ matrix X there is a nonnegative integer $\lambda(x_1, x_2, \ldots, x_t)$ which is invariant under any permutation of a given set $\{x_1, x_2, \ldots, x_t\}$. If, for every t-columned submatrix of T, the s^t $1 \times t$ matrices X occur as

rows $\lambda(x_1, x_2, \ldots, x_t)$ times, then the matrix T is called the **balanced array (BA)** of **size** n, h **constraints**, s **levels**, and **strength** t having the **index set** $\{\lambda(x_1, \ldots, x_t)\}$. Such an array is denoted by $BA(n, h, s, t)$; n is also called the **number of treatment combinations**. In particular, when $\lambda(x_1, \ldots, x_t) = \lambda$ for every $x_1, \ldots, x_t$, such an array is called an **orthogonal array (OA)** of size n, h constraints, s levels, strength t, and **index** λ, which is denoted by $OA(n, h, s, t)$. We have $n = \lambda s^t$. Balanced arrays have the advantage that they can be constructed with fewer treatment combinations than the orthogonal arrays for given h, s, t parameters. The transpose of the incidence matrix of a $BIBD(v, b, r, k, \lambda)$ is, for $A = \{0, 1\}$, a $BA(b, v, 2, 2)$ with $\lambda(0,0) = b - 2r + \lambda$, $\lambda(1,0) = \lambda(0,1) = r - \lambda$, and $\lambda(1,1) = \lambda$. There is a close relation between the existence of an OA and that of a BIBD. For example, the existence of an $OA(s^2, h, s, 2)$ is equivalent to that of $h-2$ mutually †orthogonal Latin squares of order s that is for $h = s+1$ equivalent to the existence of a $BIBD(s^2, s(s+1), s+1, s, 1)$. The existence of an $OA(4t, 4t-1, 2, 2)$ is equivalent to that of a symmetric $BIBD(4t-1, 4t-1, 2t-1, 2t-1, t-1)$ which is also equivalent to the existence of an $OA(8t, 4t, 2, 3)$.

Balanced arrays and orthogonal arrays play a vital role in the construction of symmetric and asymmetric confounded factorial experiments and fractional factorial designs. In an $OA(n, h, s, t)$, regarding columns and entries as h factors and levels of a factor of the column, respectively, each row corresponds to a treatment combination. In this case, this orthogonal array gives an orthogonal fractional s^t factorial design of resolution $t+1$ with n treatment combinations. A necessary and sufficient condition for a fractional s^h factorial design of resolution $2l+1$ to be orthogonal (resp. balanced) is that the design be an orthogonal (resp. balanced) array of strength $2l$, provided that the information matrix of the design is nonsingular.

M. Response Surface

If all the factors represent quantitative variables, such as time, temperature, amount of ingredients, etc., it is natural to think of the yield or response $\mathbf{Y}$ of the experimental results as a continuous function of the levels of these factors. We can write a functional relationship

$$Y_\alpha = f(x_{1\alpha}, x_{2\alpha}, \ldots, x_{h\alpha}) + W_\alpha, \quad \alpha = 1, 2, \ldots, n,$$

where Y_α represents the αth observation in the designed experiment and $x_{i\alpha}$ represents the level of the ith factor in the αth observation. The function f or the surface defined in $(h+1)$-dimensional †Euclidean space by f is called the **response surface**. The residual W_α measures the experimental error of the αth observation. A knowledge of the function f gives a complete summary of experimental results and also enables us to predict the response for values of the $x_{i\alpha}$ that were not tested in the experiment, or to determine such combinations of the values of variables $x_{i\alpha}$ that give the maximum (or minimum) value of f. When the mathematical form of f is not known, the function can sometimes be approximated satisfactorily within the experimental region, which is relatively small, by flexible graduating functions, such as polynomials of degrees 1 and 2 in the variables $x_{i\alpha}$:

$$Y_\alpha = \beta_0 + \sum_{i=1}^{h} \beta_i x_{i\alpha} + W_\alpha,$$

$$Y_\alpha = \beta_0 + \sum_{i=1}^{h} \beta_i x_{i\alpha} + \sum_{i=1}^{h} \beta_{ii} x_{i\alpha}^2 + \sum_{i<j}^{h} \beta_{ij} x_{i\alpha} x_{j\alpha} + W_\alpha.$$

The coefficients $\beta_0, \beta_1, \ldots$ are parameters to be estimated from the data. Thus, (i) when the form of the true f is assumed known, the object is to estimate the parameters; (ii) when the form of the true f is unknown, the object is to approximate the f by some graduating function. Designs appropriate for (i) and (ii) are called **designs for estimating parameters** and **designs for exploring a response surface**, respectively. Some experimental designs that have been developed for fitting polynomials of the first and second degrees are called **first-order designs** and **second-order designs**, respectively. The problem here is to increase the precision of fitting response surfaces by appropriately choosing n points in a given experimental region S of variable x. That is, how do we find the experimental region that interests us in the sense of finding optimum conditions and, having found it, how do we design experiments to map f over the region? As optimum criteria of allocations, corresponding to (I) and (II) given in Section E, consider the following: (IV) the †generalized variance of the estimates of coefficients is minimal; (V) the supremum of the variance of estimates of the expectation $E(\mathbf{Y})$ of response $\mathbf{Y}$ is minimal for all variables $x \in S$. For example, when an observation $\mathbf{Y}$ is given by a polynomial regression of a variable x as $Y_\alpha = \beta_0 + \beta_1 x_\alpha + \beta_2 x_\alpha^2 + \ldots + \beta_h x_\alpha^h + W_\alpha$, if for the †Legendre polynomial $P_h(x)$ we allocate an experiment replicated at $x = \pm 1$ and $h-1$ roots of $P_h'(x) = 0$ the same number of times, then it is an optimum design in the sense of (IV) and (V).

As a design for fitting first-order response planes, we can use a 2^h factorial design (or its fraction) of resolution III, since it is sufficient to consider only linear effects for each of h

factors. In comparing first-order designs, we take as an optimum criterion the minimum of the average variance of the estimates of the coefficients, corresponding to (III) of Section E. For a design for fitting second-order response surfaces, we need at least three levels of each factor for estimating all effects up to quadratic ones; then we can use a 3^h factorial design (or its fraction) of resolution V for estimating the linear $\times$ linear component of interactions of two factors. In this case, treatment combinations are too large in number, so we can adopt **central composite designs**, which are constructed by adding further treatment combinations to those obtained from a 2^h factorial design (or its fraction). If the coded levels of each factor are -1 and $+1$ in the 2^h factorial design, the $(2h+1)$ additional combinations are $(0,0,\ldots,0)$, $(\pm d, 0,\ldots,0)$, $(0, \pm d, 0, \ldots, 0)$, $\ldots, (0,\ldots,0, \pm d)$. The total number of treatment combinations to be tested is 2^h+2h+1 $(<3^h$ for $h\geqslant 3)$. The value of d can be chosen to make the coefficients in the quadratic polynomials as orthogonal as possible to each other or to minimize the bias that is created if the true form of the response surface is not quadratic or to give the design with the property of being rotatable. If $\hat{Y}_0$ is the estimated response at $(x_{10}, x_{20}, \ldots, x_{h0})\in S$, the response surface design is said to be **rotatable** if and only if the variance of $\hat{Y}_0$ is a function only of the distance $\rho=(x_{10}^2+x_{20}^2+\ldots+x_{h0}^2)^{1/2}$ of $(x_{10}, x_{20}, \ldots, x_{h0})$ from the origin, so that the variance contours in the experimental region of the variables are circles, spheres, or hyperspheres centered at the origin. Now, when $Y=\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_{11}x_1^2+\beta_{22}x_2^2+2\beta_{12}x_1x_2+W$ with $|x_i|\leqslant 1$ for $i=1, 2$, an optimum experiment in the sense of (V) is given by allocating 0.0960 at the origin, 0.0802 at experimental points $(1,0)$, $(-1,0)$, $(0,1)$, $(0,-1)$, and 0.1458 at experimental points $(1,1)$, $(1,-1)$, $(-1,1)$, $(-1,-1)$, respectively. However, this design is not rotatable. Instead of an optimum criterion (V), we can consider a region $\tilde{S}$ which is not always equal to the original experimental region S and a criterion: (V′) The supremum of the variance of estimatesof the expectation $E(Y)$ of response Y is minimal for all variables $x\in\tilde{S}$. Within $x_1^2+x_2^2\leqslant c^2$, an optimum allocation depends on the value of c, and it is shown that a rotatable design is obtained only if c is in a certain range.

There are many topics in the theory of experimental design besides the ones mentioned in this article ($\rightarrow$ [4] for multiple comparison, [9] for confounding designs that are factorial experiments in which the block size is reduced and in which for each block a fractional of all the treatment combinations are tested, [10] for split-plot designs in which certain main effects are confounded with blocks, [15] for weighing designs).

References

[1] "Experimental design" (in Japanese), Gendai tôkeigaku daiziten (Encyclopedia of modern statistics), Tôyôkeizai simpôsya, 1962.
[2] R. A. Fisher, The design of experiments, Oliver & Boyd, seventh edition, 1960.
[3] P. W. M. John, Statistical design and analysis of experiments, Macmillan, 1971.
[4] H. Scheffé, The analysis of variance, Wiley, 1959.
[5] J. Kiefer, Optimum experimental designs, J. Roy. Statist. Soc., ser. B, 21 (1959), 273–319.
[6] V. V. Fedorov, Theory of optimal experiments, Academic Press, 1972.
[7] D. Silvey, Optimal design, Chapman, 1980.
[8] D. Raghavarao, Constructions and combinatorial problems in design of experiments, Wiley, 1971.
[9] O. Kempthorne, The design and analysis of experiments, Wiley, 1952.
[10] W. G. Cochran and G. M. Cox, Experimental designs, Wiley, 1950.
[11] A. T. James, The relationship algebra of an experimental design, Ann. Math. Statist., 28 (1957), 993–1002.
[12] G. E. P. Box and N. R. Draper, Evolutionary operation, Wiley, 1968.
[13] M. Kendall and A. Stuart, The advanced theory of statistics, vol. 3, Design and analysis, and time-series, Griffin, third edition, 1976.
[14] G. E. P. Box and K. B. Wilson, On the experimental attainment of optimum conditions, J. Roy. Statist. Soc., ser. B, 13 (1952), 1–45.
[15] K. S. Banerjee, Weighing designs for chemistry, medicine, economics, operation research, statistics, Dekker, 1975.

103 (III.3) Determinants

A. Definition

The **determinant** of an $n\times n$ †matrix $A=(a_{ik})$ in a †commutative ring R is defined to be the following element of R:

$$\sum(\operatorname{sgn} P)a_{1p_1}a_{2p_2}\ldots a_{np_n},$$

where

$$P=\begin{pmatrix}1 & 2 & \ldots & n\\ p_1 & p_2 & \ldots & p_n\end{pmatrix}$$

is a permutation of the numbers $1,2,\ldots,n$,

sgn P denotes the **sign** of the permutation P (that is, sgn $P=1$ if P is an even permutation and sgn $P=-1$ if P is an odd permutation), and the summation extends over all $n!$ permutations of $1,2,\ldots,n$. The determinant of A is denoted by

$$\begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ & & \ldots & \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{vmatrix}.$$

It is written $|a_{ik}|$ or $|A|$ and is also denoted by $\det A$. Usually we suppose that R is the field $\mathbf{R}$ of real numbers or the field $\mathbf{C}$ of complex numbers, but the following theorems are also valid for cases in which R is any commutative ring, unless otherwise stated.

B. Relation to Exterior Algebras

Consider an †exterior algebra (†Grassmann algebra) of a linear space (†free module) of dimension n over R with a basis $(e_1,e_2,\ldots,e_n)$. Set

$$e_i'=a_{i1}e_1+a_{i2}e_2+\ldots+a_{in}e_n,$$

where $a_{ij}\in R$. Then we have $e_1'\wedge e_2'\wedge\ldots\wedge e_n'=|a_{ik}|e_1\wedge e_2\wedge\ldots\wedge e_n$. Conversely, we can define the determinant $|a_{ik}|$ by this relation. The properties of determinants can be easily deduced from those of exterior algebras.

C. Fundamental Properties of Determinants

(1) The determinant of the †transposed matrix ${}^tA$ of a matrix A is equal to the determinant of A. Hence the theorems stated for rows are also valid for columns.

(2) If the elements of one row (column) of a matrix are multiplied by a factor c, the determinant of the matrix is also multiplied by c. If the elements of one row (column) of a matrix are zero, its determinant is equal to zero.

(3) If from a matrix $A=(a_{ik})$, we obtain two matrices A' and A'' by replacing one row, for instance the ith row, by $a_{i1}',\ldots,a_{in}'$ and by $a_{i1}+a_{i1}',\ldots,a_{in}+a_{in}'$, respectively, then $|A''|=|A|+|A'|$. This relation is equally valid for a column.

(4) If we obtain A_Q by a permutation Q on the rows of a matrix A, then $|A_Q|=(\operatorname{sgn}Q)|A|$. In particular, if two rows (columns) of a matrix are interchanged, then the determinant changes sign.

(5) The determinant of a matrix is zero if two rows (or columns) are identical.

(6) The determinant of a matrix is not changed if the elements of any row (column), each multiplied by the same factor, are added to the corresponding elements of another row (column).

(7) Suppose that R has unity element. Let x_{ik} $(i,k=1,\ldots,n)$ be n^2 variables in R, and denote a function (having its values in R) of these variables by $\varphi(X)$, $X=(x_{ik})$. Assume that $\varphi(X)$ has the following properties: (i) if the elements of one row of X are multiplied by a factor λ, the value of φ is also multiplied by λ; (ii) if we obtain two matrices X' and X'' by replacing one row of X, for instance the ith row, by $x_{i1}',\ldots,x_{in}'$ and by $x_{i1}+x_{i1}',\ldots,x_{in}+x_{in}'$, respectively, then $\varphi(X'')=\varphi(X)+\varphi(X')$; and (iii) if two rows of X are equal, $\varphi(X)=0$. Then $\varphi(X)=c|X|$ for some constant c (in R).

(8) Suppose now that R is a field K. Assume that a function $\varphi(X)$ (in K) has the following properties: (i) if the elements of one row in X are multiplied by λ, the value of φ is also multiplied by λ; and (ii) the value of φ is not changed if the elements of any row are added to the corresponding elements of another row. Then $\varphi(X)=c|X|$ for some constant c (in K).

D. The Laplace Expansion Theorem

Let $A=(a_{ik})$ be an $n\times n$ matrix. Take r-tuples $(i_1,\ldots,i_r)$ and $(k_1,\ldots,k_r)$, where i_a and k_b belong to $\{1,\ldots,n\}$ and $i_1<\ldots<i_r$, $k_1<\ldots<k_r$. Let $(i_{r+1},\ldots,i_n)$ and $(k_{r+1},\ldots,k_n)$ be $(n-r)$-tuples such that $i_{r+1}<\ldots<i_n$, $k_{r+1}<\ldots<k_n$ and $\{i_1,\ldots,i_r,i_{r+1},\ldots,i_n\}=\{k_1,\ldots,k_r,k_{r+1},\ldots,k_n\}=\{1,\ldots,n\}$. Let $a_{(i_1,\ldots,i_r)(k_1,\ldots,k_r)}$ be the determinant of an $r\times r$ matrix whose (p,q)-component is the (i_p,k_q)-component of A for each p and q. We call this determinant a **minor** of degree r of the matrix A. (The corresponding submatrix of A is sometimes also called a minor of A.) In particular, if $(i_1,\ldots,i_r)=(k_1,\ldots,k_r)$, then it is called a **principal minor**. Furthermore, we define the **cofactor** of the minor $a_{(i_1,\ldots,i_r)(k_1,\ldots,k_r)}$ of A to be

$$\tilde{a}_{(i_1,\ldots,i_r)(k_1,\ldots,k_r)}=(-1)^{\lambda+\mu}a_{(i_{r+1},\ldots,i_n)(k_{r+1},\ldots,k_n)},$$

where $\lambda=i_1+\ldots+i_r$ and $\mu=k_1+\ldots+k_r$. In the particular case $r=1$, the cofactor of a_{ik} is $\tilde{a}_{ik}=(-1)^{i+k}\Delta_{ik}$, where Δ_{ik} is the determinant of the $(n-1)\times(n-1)$ matrix obtained from A by eliminating its ith row and kth column. For simplicity, we abbreviate $(i_1,\ldots,i_r)$, $(k_1,\ldots,k_r)$, and $(j_1,\ldots,j_r)$ as (i), (k), and (j), respectively. Then we have

$$\sum_{(j)}a_{(i)(j)}\tilde{a}_{(k)(j)}=\begin{cases}|A| & \text{if } (i)=(k),\\ 0 & \text{if } (i)\neq(k),\end{cases}$$

$$\sum_{(j)}a_{(j)(k)}\tilde{a}_{(j)(i)}=\begin{cases}|A| & \text{if } (i)=(k),\\ 0 & \text{if } (i)\neq(k),\end{cases}$$

where $\Sigma_{(j)}$ means that the sum is taken over all combinations (j). This is called the **Laplace expansion theorem**. If a matrix A has the form

$$A=\begin{pmatrix} B & 0 \\ * & C \end{pmatrix} \quad \text{or} \quad A=\begin{pmatrix} B & * \\ 0 & C \end{pmatrix},$$

and B and C are square matrices, then by this theorem we have $|A|=|B||C|$. If we number the combinations $(i)=(i_1,\dots,i_r)$ and $(k)=(k_1,\dots,k_r)$ appropriately (for instance, in lexicographical order) and regard the numbers assigned to them as row numbers and column numbers, respectively, to form a matrix $(a_{(i)(k)})$, then the Laplace theorem can be expressed as

$$(a_{(i)(k)})(\tilde{a}_{(k)(i)})=(\tilde{a}_{(k)(i)})(a_{(i)(k)}) = \begin{pmatrix} |A| & & 0 \\ & \ddots & \\ 0 & & |A| \end{pmatrix}.$$

In the particular case $r=1$, we have

$$\sum_{j=1}^{n} a_{ij}\tilde{a}_{kj}=\begin{cases} |A| & \text{if } i=k, \\ 0 & \text{if } i\neq k, \end{cases}$$

$$\sum_{j=1}^{n} a_{jk}\tilde{a}_{ji}=\begin{cases} |A| & \text{if } i=k, \\ 0 & \text{if } i\neq k. \end{cases}$$

E. Product of Determinants

Let $A=(a_{ik})$ and $B=(b_{ik})$ be two $n\times n$ matrices. For the product $AB=C=(c_{ik})$, where $c_{ik}=\sum_{j=1}^{n} a_{ij}b_{jk}$ $(i,k=1,\dots,n)$, we have $|AB|=|A||B|$. The †inverse matrix A^{-1} exists for an $n\times n$ matrix $A=(a_{ik})$ if and only if $|A|\neq 0$, and then $A^{-1}=(b_{ik})$ with elements $b_{ik}=\tilde{a}_{ki}/|A|$. Moreover, we have $|A^{-1}|=|A|^{-1}$. (In the case where the elements a_{ik} are in the commutative ring R with unity element, A^{-1} exists if and only if $|A|$ is a †regular element of R.)

F. Theorems on Determinants

(1) Let $\tilde{a}_{ik}$ be the cofactor of a_{ik} in the determinant of an $n\times n$ matrix $A=(a_{ik})$. Then the determinant $|\tilde{a}_{ik}|$ is equal to $|A|^{n-1}$. In general,

$$|a_{(i_1,\dots,i_r)(k_1,\dots,k_r)}|=|A|^{\binom{n-1}{r-1}},$$

$$|\tilde{a}_{(i_1,\dots,i_r)(k_1,\dots,k_r)}|=|A|^{\binom{n-1}{r}}.$$

(2) The determinant of a submatrix of the matrix $(\tilde{a}_{ik})$, composed of the i_1th, $\dots$, i_rth rows and k_1th, $\dots$, k_rth columns of $(\tilde{a}_{ik})$ is equal to

$$|A|^{r-1}\tilde{a}_{(i_1,\dots,i_r)(k_1,\dots,k_r)}.$$

(3) Let $\Delta\begin{pmatrix} i_1,\dots,i_r \\ k_1,\dots,k_r \end{pmatrix}$ be the determinant of the $(n-r)\times(n-r)$ matrix obtained from an $n\times n$ matrix A by eliminating its i_1th, $\dots$, i_rth rows and k_1th, $\dots$, k_rth columns. Then

$$|A|\Delta\begin{pmatrix} i & j \\ k & l \end{pmatrix}=\Delta\begin{pmatrix} i \\ k \end{pmatrix}\Delta\begin{pmatrix} j \\ l \end{pmatrix}-\Delta\begin{pmatrix} i \\ l \end{pmatrix}\Delta\begin{pmatrix} j \\ k \end{pmatrix},$$

$$i<j, \quad k<l.$$

(4) **Sylvester's theorem**. Let b_{ik} $(i,k=1,\dots,n-r)$ denote the minor $a_{(1,\dots,r,r+i)(1,\dots,r,r+k)}$ of an $n\times n$ matrix $A=(a_{ji})$. Then

$$|b_{ik}|=|A|\begin{vmatrix} a_{11} & \dots & a_{1r} \\ & \dots & \\ a_{r1} & \dots & a_{rr} \end{vmatrix}^{n-r-1}.$$

(5) Let A be an $n\times m$ matrix and B an $m\times n$ matrix. Then AB is an $n\times n$ matrix. If $n>m$, then $|AB|=0$. If $n\leqslant m$, let $(i)=(i_1,\dots,i_n)$ $(i_1<\dots<i_n)$ be a combination of $1,2,\dots,m$, taken n at a time. Let $A_{(i)}$ be the $n\times n$ matrix composed of the i_1th, $\dots$, i_nth columns of A, and $B_{(i)}$ the $n\times n$ matrix composed of the i_1th, $\dots$, i_nth rows of B. Then $|AB|=\sum_{(i)}|A_{(i)}||B_{(i)}|$, where the summation extends over all possible combinations (i).

(6) Determinant of a †Kronecker product. If A is an $m\times m$ matrix and B is an $n\times n$ matrix, then $|A\otimes B|=|A|^n|B|^m$.

(7) Let H be an $n\times n$ †Hermitian matrix, and let H_k denote the matrix composed of its first k rows and columns. Then H is positive definite if and only if $|H_k|>0$ for all $k=1,\dots,n$.

G. Special Determinants

(1) A determinant of the form

$$\begin{vmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \\ x_1^2 & x_2^2 & \dots & x_n^2 \\ & & \dots & \\ x_1^{n-1} & x_2^{n-1} & \dots & x_n^{n-1} \end{vmatrix}$$

is called a **Vandermonde determinant**. It is equal to the †simplest alternating function $\prod_{i>k}(x_i-x_k)$.

(2) A **cyclic determinant** is one of the form

$$\begin{vmatrix} x_0 & x_1 & x_2 & \dots & x_{n-1} \\ x_{n-1} & x_0 & x_1 & \dots & x_{n-2} \\ & & & \dots & \\ x_1 & x_2 & x_3 & \dots & x_0 \end{vmatrix}$$

$$=\prod_{i=0}^{n-1}(x_0+\zeta^i x_1+\zeta^{2i}x_2+\dots+\zeta^{(n-1)i}x_{n-1}),$$

where ζ is a †primitive nth root of unity.

(3) Consider the vectors $\alpha_i=(a_{i1},a_{i2},\dots,a_{in})$ $(i=1,2,\dots,n)$, and let (α_i,α_j) denote the †inner product of α_i and α_j. Then the following determinant is called the **Gramian** of these vectors:

$$\begin{vmatrix} (\alpha_1,\alpha_1) & (\alpha_1,\alpha_2) & \dots & (\alpha_1,\alpha_n) \\ (\alpha_2,\alpha_1) & (\alpha_2,\alpha_2) & \dots & (\alpha_2,\alpha_n) \\ & & \dots & \\ (\alpha_n,\alpha_1) & (\alpha_n,\alpha_2) & \dots & (\alpha_n,\alpha_n) \end{vmatrix}$$

$$=\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ & & \dots & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}^2.$$

(4) For an †alternating matrix (namely, a square matrix X such that ${}^tX=-X$), we have the identity

$$\begin{vmatrix} 0 & x_{12} & x_{13} & \cdots & x_{1n} \\ -x_{12} & 0 & x_{23} & \cdots & x_{2n} \\ -x_{13} & -x_{23} & 0 & \cdots & x_{3n} \\ & & & \cdots & \\ -x_{1n} & -x_{2n} & -x_{3n} & \cdots & 0 \end{vmatrix} = \begin{cases} P_n(\ldots, x_{ij}, \ldots)^2 & \text{if } n \text{ is even,} \\ 0 & \text{if } n \text{ is odd,} \end{cases}$$

where $P_n(\ldots, x_{ij}, \ldots)$ is a polynomial of variables x_{ij}, which (equipped with appropriate sign) is called the **Pfaffian** of these variables.

References

[1] I. Satake, Linear algebra, Dekker, 1975.
[2] G. W. H. Kowalewski, Einführung in die Determinantentheorie, de Gruyter, second revised edition, 1925 (Chelsea, 1948).
[3] N. Bourbaki, Eléments de mathématique, Algèbre, ch. 3, Actualités Sci. Ind., 1044, Hermann, new edition, 1958.
[4] F. R. Gantmakher, The theory of matrices, Chelsea, I, II, 1959. (Original in Russian, 1953.)
[5] A. G. Kurosh, Lectures on general algebra, Chelsea, 1970. (Original in Russian, 1955.)
[6] R. Godement, Cours d'algèbre, Hermann, 1963.
[7] S. MacLane and G. Birkhoff, Algebra, Macmillan, 1967.

104 (XIII.15) Difference Equations

A. General Remarks

Let y be a function of a real variable x defined on an interval I and let Δx be a fixed quantity. When two points x and $x+\Delta x$ are in I, we define the **difference** $\Delta y(x)$ of y at x by $\Delta y(x) = y(x+\Delta x)-y(x)$ and the difference quotient by $\Delta y(x)/\Delta x$; Δx is called the difference of x. Without loss of generality, we can take $\Delta x = 1$, for otherwise there is a constant b such that $\Delta x'=1$ for the new independent variable $bx=x'$. If $\Delta x=1$, the **second difference** $\Delta^2 y(x)=\Delta(\Delta y(x))$ is given by

$$\Delta^2 y(x)=\Delta y(x+1)-\Delta y(x)$$
$$=y(x+2)-2y(x+1)+y(x).$$

Similarly, the **difference of the** n**th order** is defined by $\Delta^n y(x)=\Delta(\Delta^{n-1}y(x))$, and

$$\Delta^n y(x)=\sum_{k=0}^{n}(-1)^{n-k}\binom{n}{k}y(x+k).$$

Conversely, $y(x+n)$ is expressed by differences as

$$y(x+n)=\sum_{k=0}^{n}\binom{n}{k}\Delta^k y(x)$$

(→ 223 Interpolation).

B. Summation

Given a function $g(x)$ and Δx, a function $y(x)$ that satisfies $\Delta y(x)/\Delta x = g(x)$ is called a **sum** of $g(x)$. **Summation** of $g(x)$ is to find a sum of $g(x)$. Given a sum $y(x)$ of $g(x)$, an **indefinite sum** of $g(x)$, written as $Sg(x)\Delta x$, is given by $Sg(x)\Delta x = y(x)+c(x)$, where $c(x)$ is an arbitrary periodic function of period Δx. In many cases, $c(x)$, which corresponds to an arbitrary constant in an indefinite integral, is omitted. For example, a sum of $g(x)=nx^{n-1}$ for $\Delta x=1$ is the nth-order †Bernoulli polynomial $B_n(x)$ for $n\neq 0$; a sum of x^{-1} is $\psi(x)$, given by $\psi(x)= d\log\Gamma(x)/dx$ (→ 174 Gamma Function).

When the series $-\Delta x\sum_{k=0}^{\infty}g(x+k\Delta x)$ or $\Delta x\sum_{k=1}^{\infty}g(x-k\Delta x)$ converges, both can be sums of $g(x)$. Since the requirement of convergence for these series was found to be too strict, the following requirement was given by N. E. Nörlund instead: Let x be a real variable and $g(x)$ be continuous for $x\geqslant b$. Define $\lambda(x)$ by $\lambda(x)=x^p(\log x)^q$ ($p\geqslant 1, q\geqslant 0$). Then if for a positive η

$$F(x,\Delta x,\eta)=\int_a^\infty g(z)e^{-\eta\lambda(z)}dz$$
$$-\Delta x\sum_{k=0}^{\infty}g(x+k\Delta x)e^{-\eta\lambda(x+k\Delta x)}$$

is convergent for $a>b$, F satisfies $\Delta F(x,\Delta x,\eta)/\Delta x = g(x)\exp(-\eta\lambda(x))$. Accordingly, if $F(x,\Delta x,\eta)$ approaches a limit $F(x)$ as $\eta\to 0$, $F(x)$ is a solution of $\Delta F(x)/\Delta x=g(x)$. $F(x)$ is called the **principal solution** of $\Delta F(x)/\Delta x = g(x)$.

C. Difference Equations

Let $\Delta x=1$. An equation $F(x,y(x),\Delta y(x), \ldots,\Delta^n y(x))=0$ in x and differences of an unknown function $y(x)$ is called a **difference equation**. If the substitution $y=\varphi(x)$ satisfies the equation for x in some interval, $\varphi(x)$ is a **solution** of the equation. Because of the relation between $y(x)$, $y(x+1),\ldots,y(x+n)$ and the differences of y at x, we can transform the given difference equation in the form $G(x,y(x), y(x+1),\ldots,y(x+n))=0$. This form appears more often in applications and is called the **standard form** of a difference equation.

If the equation is linear in $y(x)$, $y(x+1)$, $\dots, y(x+n)$, namely, if it is given by

$$\sum_{i=0}^{n} p_i(x)y(x+i)=q(x),$$

the difference equation is said to be **linear**. When $q(x)\equiv 0$, it is **homogeneous**; otherwise, it is **inhomogeneous** (or **nonhomogeneous**).

D. Linear Difference Equations

Assume that $p_0(x), \dots, p_n(x)$ are single-valued analytic functions without poles and common zeros in some domain. Consider the linear difference equation

$$\sum_{i=0}^{n} p_i(x)y(x+i)=0. \tag{1}$$

If $\varphi_1(x), \varphi_2(x), \dots, \varphi_m(x)$ are solutions of (1), then a linear combination $a_1(x)\varphi_1(x)+a_2(x)\varphi_2(x)+\dots+a_m(x)\varphi_m(x)$ with arbitrary periodic functions $a_1(x), a_2(x), \dots, a_m(x)$ of period 1 is also a solution of (1).

Let $\beta_1, \beta_2, \dots$ be singular points of $p_1(x), p_2(x), \dots, p_n(x)$, $\alpha_1, \alpha_2, \dots$ be the zeros of $p_0(x)$, and $\gamma_1, \gamma_2, \dots$ be the zeros of $p_n(x+n)$. Then the set of **singular points** of the linear difference equation (1) is the set $\{\alpha_i, \beta_i, \gamma_i\}$.

A function $\varphi_m(x)$ is said to be **linearly dependent** on the functions $\varphi_1(x), \varphi_2(x), \dots, \varphi_{m-1}(x)$ with respect to the difference equation (1) if $\varphi_m(x)=a_1(x)\varphi_1(x)+a_2(x)\varphi_2(x)+\dots+a_{m-1}(x)\varphi_{m-1}(x)$, where $a_1(x), a_2(x), \dots, a_{m-1}(x)$ are functions of period 1, every one of which takes a nonzero finite value at least at one point not congruent (mod **Z**) to any of the singular points, where **Z** is the additive group of integers.

A set of m functions is called linearly independent if none of the functions is dependent on the other $m-1$ functions. When a set of n solutions of equation (1) is linearly independent, it is a **fundamental system** for (1). Any solution of (1) can be expressed as a linear combination of n solutions of a fundamental system.

The determinant

$$\begin{vmatrix} \varphi_1(x) & \varphi_2(x) & \dots & \varphi_n(x) \\ \varphi_1(x+1) & \varphi_2(x+1) & \dots & \varphi_n(x+1) \\ & & \dots & \\ \varphi_1(x+n-1) & \varphi_2(x+n-1) & \dots & \varphi_n(x+n-1) \end{vmatrix}$$

formed from n functions $\varphi_1(x), \varphi_2(x), \dots, \varphi_n(x)$ is called **Casorati's determinant** and is denoted by $D(\varphi_1(x), \varphi_2(x), \dots, \varphi_n(x))$. A necessary and sufficient condition for a given set of n functions to be independent is that Casorati's determinant be nonzero at every point except those which are congruent to singular points of (1). Casorati's determinant is used to determine whether a given set of solutions is fundamental.

Let $\psi(x)$ be a solution of a nonhomogeneous linear difference equation

$$P_x(y)=\sum_{i=0}^{n} p_i(x)y(x+i)=q(x). \tag{2}$$

If $\varphi_1(x), \varphi_2(x), \dots, \varphi_n(x)$ are n linearly independent solutions of (1), then an arbitrary solution of (2) is given by

$$y=a_1(x)\varphi_1(x)+a_2(x)\varphi_2(x)+\dots$$
$$+a_n(x)\varphi_n(x)+\psi(x),$$

where $a_1(x), \dots, a_n(x)$ are arbitrary periodic functions of period 1. Then the expression for y is called a **general solution** of (2). If we abbreviate Casorati's determinant of a fundamental system of solutions $\varphi_1(x), \varphi_2(x), \dots, \varphi_n(x)$ of (1) by $D(x)$ and write $\mu_i(x)$ as the quotient of the cofactor of $\varphi_i(x+n)$ of $D(x+1)$ by $D(x+1)$, we have

$$\psi(x)=\sum_{i=1}^{n} \varphi_i(x)\mathbf{S}q(z)\mu_i(z)\Delta z,$$

assuming that the summation **S** on the right-hand side is known. This is the analog of Lagrange's †method of variation of constants in the theory of linear ordinary differential equations.

E. Linear Difference Equations with Constant Coefficients

If all the coefficients in

$$\sum_{i=0}^{n} p_i y(x+i)=0, \qquad p_0\neq 0, \quad p_n\neq 0, \tag{3}$$

are constants, n linearly independent solutions are obtained easily. Indeed, if λ is a root of the algebraic equation $\sum_{i=0}^{n} p_i\lambda^i=0$, λ^x is a solution of (3). This algebraic equation is called the **characteristic equation** of (3). If it has n distinct roots $\lambda_1, \lambda_2, \dots, \lambda_n$, then $\lambda_1^x, \lambda_2^x, \dots, \lambda_n^x$ are n linearly independent solutions. In general, if λ is an m-tuple root of the characteristic equation, then $\lambda^x, x\lambda^x, \dots, x^{m-1}\lambda^x$ are solutions of (3). Accordingly, if λ_j is a root of multiplicity m_j ($\sum_{j=1}^{s} m_j=n, j=1,2,\dots,s$), then $\lambda_j^x, x\lambda_j^x, \dots, x^{m_j-1}\lambda_j^x$ ($j=1,\dots,s$) constitute a set of n linearly independent solutions.

Even if all the p_i are real, the characteristic equation may have complex roots. In such a case real solutions are obtained as follows: When $\lambda=\mu+iv$ is a root of multiplicity m, $\bar{\lambda}=\mu-iv$ is also a root of the same multiplicity. If we write $\rho=\sqrt{\mu^2+v^2}$, $\tan\varphi=v/\mu$, then $\rho^x\cos\varphi x, \rho^x\sin\varphi x, x\rho^x\cos\varphi x, x\rho^x\sin\varphi x, \dots, x^{m-1}\rho^x\cos\varphi x, x^{m-1}\rho^x\sin\varphi x$ are $2m$ independent real solutions.

Nonhomogeneous equations with con-

stant coefficients can be generally solved by Lagrange's method with these solutions. However, when the nonhomogeneous term has a special form such as

$$\sum_{i=0}^{n} p_i y(x+i) = p(x)\lambda^x,$$

where $p(x)$ is a polynomial in x and λ is a root of multiplicity m of the characteristic equation, we can use the method of undetermined coefficients. In this particular case, the substitution $(A_0 + A_1 x + \ldots + A_k x^k) x^m \lambda^x$ with undetermined coefficients $A_0, A_1, \ldots, A_k$ gives solutions, k being the degree of $p(x)$.

F. Difference and Differential Equations

The differential operator d/dx acts on the family of functions $\{x^m \mid m=0, \pm 1, \ldots\}$ according to $dx^m/dx = mx^{m-1}$, just as the difference operator Δ acts on the family $\{x^{(m)} = \Gamma(x+1)/\Gamma(x-m+1) \mid m=0, \pm 1, \ldots\}$ according to $\Delta x^{(m)} = mx^{(m-1)}$. Hence by using the **factorial series** $\sum a_m x^{(m)}$ and its similarity to the power series $\sum a_m x^m$, we may obtain some analogies with the theory of differential equations. For example, the †Frobenius method in the theory of †regular singular points can be applied to the system of difference equations

$$(z-1)\Delta_{-1} w_k(z) = \sum_{j=1}^{n} a_{jk}(z) w_j(z),$$

$k = 1, 2, \ldots, n.$

However, there are certain essential differences between functions defined as solutions of differential and difference equations. For example, **Hölder's theorem** states that no solution of the simple difference equation $y(x+1) - y(x) = x^{-1}$ satisfies any †algebraic differential equation. Consequently, the gamma function, which is related to a solution of the equation $\psi(x) = d\log\Gamma(x)/dx$, cannot be a solution of any algebraic differential equation. For the numerical solution of ordinary differential equations by difference equation approximation → 303 Numerical Solution of Ordinary Differential Equations.

G. Geometric Difference Equations

For an arbitrary complex number q, an equation of the form $y(qx) = f(x, y(x))$ is called a **geometric difference equation**. For example, the ordinary difference equation (1) can be transformed into

$$\sum_{k=0}^{n} p_k(z) U(zq^k) = B(z) \qquad (1')$$

by the change of variable $z = q^x$. Although it is possible to transform an equation of the type (1′) into that of the type (1), there are theories developed specifically for the type (1′), since the coefficients of the equation may become more complicated by such a transformation.

References

[1] P. M. Batchelder, An introduction to linear difference equations, Dover, 1967.
[2] G. Boole, A treatise on the calculus of finite differences, Dover, 1960.
[3] R. M. Cohn, Difference algebra, Interscience, 1965.
[4] T. Fort, Finite differences and difference equations in the real domain, Oxford Univ. Press, 1948.
[5] C. Jordan, Calculus of finite differences, Chelsea, 1933.
[6] K. S. Miller, Linear difference equations, Benjamin, 1968.
[7] L. M. Milne-Thomson, The calculus of finite differences, Chelsea, 1980.
[8] N. E. Nörlund, Vorlesungen über Differenzenrechnung, Springer, 1924.
[9] N. E. Nörlund, Leçons sur les équations linéaires aux différences finies, Gauthier-Villars, 1929.
[10] C. H. Richardson, An introduction to calculus of finite differences, Van Nostrand, 1954.

105 (VII.2) Differentiable Manifolds

A. General Remarks

The rudimentary concept of n-dimensional manifolds can already be seen in J. Lagrange's dynamics. In the middle of the 19th century n-dimensional †Euclidean space was known as a continuum of n real parameters (A. Cayley; H. Grassmann, 1844, 1861; L. Schläfli, 1852). The notion of general n-dimensional manifolds was introduced by B. Riemann as a result of his differential geometric observations (1854). He considered an n-dimensional manifold to be a set formed by a 1-parameter family of $(n-1)$-dimensional manifolds, just as a surface is formed by the motion of a curve. Analytical studies of topological structures of manifolds and their local properties were initiated and developed by Riemann, E. Betti, H. Poincaré, and others. To avoid the difficulties and disadvantages of analytical methods, Poincaré restricted his consideration to those topological spaces X that are †connected, †triangulable, and such that each point of X

has a neighborhood homeomorphic to an n-dimensional Euclidean space. We often refer to such spaces as **Poincaré manifolds**; Poincaré called them n-dimensional manifolds. In 1936, H. Whitney published a monumental paper [14] on differentiable manifolds in which the various fundamental concepts on differentiable manifolds were established. This and subsequent papers written by Whitney during nearly twenty years greatly influenced the rapid advance of the theory of differentiable manifolds since 1950.

B. Topological Manifolds

An n-dimensional **topological manifold** M is by definition a †Hausdorff space in which each point p has a neighborhood $U(p)$ homeomorphic to an open set of $\mathbf{R}^n$.

Let M' be a Hausdorff space in which each point p of M' has a neighborhood $U(p)$ homeomorphic to an open set of H^n, where H^n is the half-space $\{(x_1, x_2, \ldots, x_n) \in \mathbf{R}^n \mid x_n \geqslant 0\}$. Let $\partial M'$ denote the set consisting of points p of M' such that p corresponds to a point of $H_0^n = \{(x_1, \ldots, x_n) \in H^n \mid x_n = 0\} \subset H^n$ under the homeomorphism from $U(p)$ to an open set of H^n. M' is called an n-dimensional **topological manifold with boundary** if $\partial M' \neq \emptyset$, and $\partial M'$ is called the **boundary** of M'. On the other hand, M defined as above or M' with $\partial M' = \emptyset$ is called an n-dimensional **topological manifold without boundary**. The **interior** of M' is the complement $M_0' = M' - \partial M'$ of the boundary. The boundary of an n-dimensional topological manifold is an $(n-1)$-dimensional topological manifold. A topological manifold without boundary is called **closed** or **open** according as it is compact or has no connected component which is compact. There exist connected topological manifolds that are not †paracompact; among them, the 1-dimensional ones are called **long lines**. A connected paracompact topological manifold M has a †countable open base and is †metrizable.

C. Local Coordinates

Let M be an n-dimensional topological manifold. A pair (U, ψ) consisting of an open set U of M and a homeomorphism ψ of U onto an open set of $\mathbf{R}^n$ is called a **coordinate neighborhood** of M. If we denote by $(x^1(p), \ldots, x^n(p))$ $(p \in U)$ the coordinates of the point $\psi(p)$ of $\mathbf{R}^n$, then $x^1, x^2, \ldots, x^n$ are real-valued continuous functions defined on U. We call these n functions the **local coordinate system** in the coordinate neighborhood (U, ψ) and the n real numbers $x^1(p), \ldots, x^n(p)$ the **local coordinates** of the point $p \in U$ (with respect to (U, ψ)).

A set $S = \{(U_\alpha, \psi_\alpha)\}_{\alpha \in A}$ of coordinate neighborhoods is called an **atlas** of M if $\{U_\alpha\}_{\alpha \in A}$ forms an †open covering of M.

D. Differentiable Manifolds

Let $S = \{(U_\alpha, \psi_\alpha)\}_{\alpha \in A}$ be an atlas of an n-dimensional topological manifold M. For each pair of coordinate neighborhoods (U_α, ψ_α) and (U_β, ψ_β) in S such that $U_\alpha \cap U_\beta \neq \emptyset$, $\psi_\beta \circ \psi_\alpha^{-1}$ is a homeomorphism of the open set $\psi_\alpha(U_\alpha \cap U_\beta)$ of $\mathbf{R}^n$ onto the open set $\psi_\beta(U_\alpha \cap U_\beta)$ of $\mathbf{R}^n$. Let $x = (x^1, \ldots, x^n) \in \psi_\alpha(U_\alpha \cap U_\beta)$. Then we can write $(\psi_\beta \circ \psi_\alpha^{-1})(x) = (f_{\beta\alpha}^1(x), \ldots, f_{\beta\alpha}^n(x))$. If the n real-valued functions $f_{\beta\alpha}^1, \ldots, f_{\beta\alpha}^n$ defined in $\psi_\alpha(U_\alpha \cap U_\beta)$ are of †class C^r $(1 \leqslant r \leqslant \infty)$ (resp. †real analytic) for any α, β in A such that $U_\alpha \cap U_\beta \neq \emptyset$, then we call S an **atlas of class C^r** (resp. C^ω) of M. When an n-dimensional topological manifold M has an atlas S of class C^r $(1 \leqslant r \leqslant \omega)$, we call the pair (M, S) an n-dimensional **differentiable manifold of class C^r** (or **C^r-manifold**). A C^∞-manifold is also called a **smooth manifold**, while a C^ω-manifold is called a **real analytic manifold**. We call M the **underlying topological space** of (M, S), and we say that S defines a **differentiable structure of class C^r** (or **C^r-structure**) in M.

In particular, a C^ω-structure is called a **real analytic structure**. A C^r-manifold whose underlying topological space is compact (†paracompact) is called a **compact (paracompact) C^r-manifold**. A coordinate neighborhood (U, ψ) of M is called a **coordinate neighborhood of class C^r** of (M, S) if the union $S \cup \{(U, \psi)\}$ is also an atlas of class C^r of M. In particular, each coordinate neighborhood of M belonging to S is of class C^r. The set $\tilde{S}$ of all coordinate neighborhoods of class C^r of (M, S) is an atlas of M containing S, and we call $\tilde{S}$ the maximal atlas containing S. Let S and S' be two atlases of class C^r of M. If $\tilde{S} = \tilde{S}'$, then we say that S and S' define the same differentiable structure of class C^r on M and that the differentiable manifolds (M, S) and (M, S') of class C^r are equivalent. In particular, (M, S) and $(M, \tilde{S})$ are equivalent C^r-manifolds. Let S and S' be atlases of class C^r and class C^s, respectively, where $1 \leqslant r < s \leqslant \omega$. Since $s > r$, we can consider S' an atlas of class C^r. If S and S' define the same C^r-structure in M, then we say that the C^s-structure defined by S' is **subordinate** to the C^r-structure defined by S. If M is paracompact, then there exists a C^ω-structure subordinate to a C^r-structure of M (Whitney [14]).

E. Differentiable Manifolds with Boundaries

Let U and U' be open sets in the half-space H^n, and let $\varphi: U \to U'$ be a continuous mapping. If

there exist open sets W and W' in $\mathbf{R}^n$ containing U and U', respectively, and a mapping $\psi: W \rightarrow W'$ of class C^r that extends φ, we call φ a mapping of U into U' of class C^r. Let M be a Hausdorff topological space. A structure of a C^r-manifold on M is defined by a set $S = \{(U_\alpha, \psi_\alpha)\}_{\alpha \in A}$, where $\{U_\alpha\}_{\alpha \in A}$ is an open covering of M, and, for each α, ψ_α is a homeomorphism of U_α onto an open set of H^n such that for any $\alpha, \beta \in A$ with $U_\alpha \cap U_\beta \neq \emptyset$, $\psi_\beta \circ \psi_\alpha^{-1}$ is a mapping of class C^r from $\psi_\alpha(U_\alpha \cap U_\beta)$ onto $\psi_\beta(U_\alpha \cap U_\beta)$. Let ∂M denote the set consisting of points p of M such that $p \in U_\alpha$ and $\psi_\alpha(p) \in H_0^n = \{(x_1, \ldots, x_n) \in H^n \mid x_n = 0\}$ for some $\alpha \in A$. If $\partial M \neq \emptyset$, the pair (M, S) is called an n-dimensional **differentiable manifold with boundary of class** C^r (or C^r**-manifold with boundary**), and ∂M is called the **boundary** of M. ∂M forms an $(n-1)$-dimensional C^r-manifold. If we put $U'_\alpha = U_\alpha \cap \partial M$ and denote the restriction of ψ_α to U'_α by ψ'_α, then $S' = \{(U'_\alpha, \psi'_\alpha)\}_{\alpha \in A}$ is an atlas of class C^r of ∂M. If ∂M is empty, then (M, S) is a C^r-manifold. In this sense a C^r-manifold is sometimes called a C^r**-manifold without boundary**.

F. Orientation of a Manifold

Let $S = \{(U_\alpha, \psi_\alpha)\}_{\alpha \in A}$ be an atlas of class C^r in M, and for each α let $\{x_\alpha^1, \ldots, x_\alpha^n\}$ be the local coordinate system in a coordinate neighborhood (U_α, ψ_α). If U_α and U_β intersect, then there exist n real-valued functions F^i $(i = 1, \ldots, n)$ defined on $\psi_\alpha(U_\alpha \cap U_\beta)$ such that $x_\beta^i(p) = F^i(x_\alpha^1(p), \ldots, x_\alpha^n(p))$ for $p \in U_\alpha \cap U_\beta$ and $i = 1, \ldots, n$. The †Jacobian $D_{\alpha\beta} = D(F^1, \ldots, F^n)/D(x_\alpha^1, \ldots, x_\alpha^n)$ is different from zero at each point $(x_\alpha^1, \ldots, x_\alpha^n)$ of $\psi_\alpha(U_\alpha \cap U_\beta)$. If we can choose an atlas S of M so that, for any α, β such that $U_\alpha \cap U_\beta$ is nonempty, the Jacobian $D_{\alpha\beta}$ is always positive, then we say that the C^r-manifold M is **orientable**, and we call S an **oriented atlas**.

Let $S = \{(U_\alpha, \psi_\alpha)\}_{\alpha \in A}$ and $S' = \{(V_\lambda, \varphi_\lambda)\}_{\lambda \in \Lambda}$ be two oriented atlases of a connected C^r-manifold M. If M is connected, then the sign of the Jacobian $D_{\alpha\lambda}(p)$ of the transformation of local coordinates is independent of the choice of $\alpha \in A$, $\lambda \in \Lambda$, and $p \in U_\alpha \cap V_\lambda$. We say that S and S' define the **same (opposite) orientation** if $D_{\alpha\lambda}$ is always positive (negative). Hence if M is connected, the set of all oriented atlases of class C^r is composed of two subsets such that atlases belonging to one of them have the same orientation, while two atlases belonging to different ones have the opposite orientation. Each of these subsets is called an **orientation** of the connected C^r-manifold M. When we assign to M one of two possible orientations, M is called an **oriented manifold**; the assigned orientation is called its **positive orientation** and the other its **negative orientation**. If $S = \{(U_\alpha, \psi_\alpha)\}_{\alpha \in A}$ belongs to the positive orientation, S and (U_α, ψ_α) are called an atlas and local coordinate system, respectively, compatible with the positive orientation.

G. Differentiable Functions

Let f be a real-valued function defined in a neighborhood of a point p of a C^∞-manifold M. Let (U, ψ) be a coordinate neighborhood of class C^∞ such that $p \in U$. If the function $f \circ \psi^{-1}$ is of class C^r $(1 \leqslant r \leqslant \infty)$ in a neighborhood of the point $\psi(p)$ in $\mathbf{R}^n$, then the function f is called a **function of class** C^r **at** p. This definition is independent of the choice of a coordinate neighborhood of class C^∞. If we denote the local coordinate system in (U, ψ) by $(x^1, \ldots, x^n)$, there exists a function $f(x^1, \ldots, x^n)$ of n variables defined in a neighborhood of $\psi(p)$ in $\mathbf{R}^n$ such that $f(q) = f(x^1(q), \ldots, x^n(q))$ for each point q in the neighborhood of p. Here we use the same symbol f for the function f defined in a neighborhood of p in M and for the function $f \circ \psi^{-1}$ defined in the image of the neighborhood by ψ in $\mathbf{R}^n$. The function f is of class C^r at p if and only if $f(x^1, \ldots, x^n)$ is of class C^r in a neighborhood of the point $(x^1(p), \ldots, x^n(p))$ of $\mathbf{R}^n$. A **function of class** C^r (or C^r**-function**) **in** M is a real-valued function in M that is of class C^r at every point of M.

H. Tangent Vectors

Let M be a C^∞-manifold, and let $\mathfrak{F}(M)$ be the real vector space consisting of all C^∞-functions in M. (For the sake of simplicity, we denote a manifold (M, S) by M.) A **tangent vector** L at a point p of M is a linear mapping $L: \mathfrak{F}(M) \rightarrow \mathbf{R}$ such that $L(fg) = L(f)g(p) + f(p)L(g)$ for any f and g in $\mathfrak{F}(M)$. For any two tangent vectors L_1, L_2 and any pair of real numbers λ_1, λ_2 we define $\lambda_1 L_1 + \lambda_2 L_2$ by $(\lambda_1 L_1 + \lambda_2 L_2)(f) = \lambda_1 L_1(f) + \lambda_2 L_2(f)$, $f \in \mathfrak{F}(M)$.

Thus tangent vectors at p form a real vector space T_p, which we call the **tangent vector space** (or simply the **tangent space**) of M at the point p. The dimension of the tangent vector space T_p equals the dimension of M. The set of all tangent vectors of M forms a †vector bundle over the base space M, called the **tangent vector bundle** (or **tangent bundle**) of M.

By a **tangent** r**-frame** $(r \leqslant n)$ at p we mean an ordered set of r linearly independent tangent vectors at p. The set of all tangent r-frames also forms a fiber bundle over M called the **tangent** r**-frame bundle** (or **bundle of tangent** r**-frames**) (→ 147 Fiber Bundles F).

I. Differentials of Functions

For a C^∞-function f in M and a point p of M we can define a linear mapping $df_p: T_p \to \mathbf{R}$ by $df_p(L) = L(f)$ for all $L \in T_p$, and we call df_p the **differential of** f **at** p. The totality of differentials at p of C^∞-functions in M forms the †dual vector space of the tangent vector space T_p.

J. Differentiable Mappings

Let φ be a continuous mapping of a C^∞-manifold M into a C^∞-manifold M'. We call φ a **differentiable mapping of class** C^r (or simply a C^r**-mapping**) ($1 \leqslant r \leqslant \infty$) if the function $f \circ \varphi$ is of class C^r for any C^r-function f on M'. If φ is a homeomorphism of M onto M' and φ and φ^{-1} are both of class C^r, then we call φ a **diffeomorphism of class** C^r. If there exists a diffeomorphism of class C^∞ of a C^∞-manifold M onto a C^∞-manifold M', then M and M' are said to be **diffeomorphic**.

Let M and M' be C^∞-manifolds and φ be a C^∞-mapping of M into M'. For a tangent vector L of M at p, a tangent vector L' of M' at $\varphi(p)$ is defined by $L'(g) = L(g \circ \varphi)$, $g \in \mathfrak{F}(M')$. The mapping $L \to L'$ defines a linear mapping $(d\varphi)_p$ of the tangent vector space T_p of M at p into the tangent vector space $T_{\varphi(p)}$ of M' at $\varphi(p)$. The linear mapping $(d\varphi)_p$ is called the **differential of the differentiable mapping** φ **at** p. If $(d\varphi)_p$ is surjective, p is called a **regular point** of φ. A point on M which is not a regular point is called a **critical point** of φ. A point q on M' which is an image of a critical point is called a **critical value** of φ, and a point on M' which is not a critical value is called a **regular value**. In $\mathbf{R}^n$, the diffeomorphic image of a set of Lebesgue measure zero has Lebesgue measure zero. So the set of Lebesgue measure zero is well defined on a (paracompact) C^∞-manifold. Then **Sard's theorem** states: Let $\varphi: M \to M'$ be a C^∞-mapping; then the set of critical values of φ has Lebesgue measure zero in M'.

K. Immersions and Embeddings

Let M and M' be C^∞-manifolds and φ be a C^∞-mapping of M into M'. If $(d\varphi)_p$ is injective at every point p of M, then φ is called an **immersion** of M into M'. If φ is an immersion, then for some neighborhood U_p of any point p of M the restriction $\varphi \mid U_p$ gives rise to a homeomorphism from U_p into M'. If an immersion φ is injective, then φ is called an embedding (or an **imbedding**) of M into M'. An alternative definition is often used, which says that φ is an **embedding** if, in addition to the above conditions, φ gives a homeomorphism from M onto $\varphi(M)$, where $\varphi(M)$ has the relative topology of M'. If the former definition is adopted as embedding, then a mapping φ satisfying the conditions of the alternative definition is sometimes referred to as a **regular embedding**. If M is compact, the two definitions coincide. In Sections L and M, embedding always means regular embeddings.

The theory of embeddings and immersions is mainly concerned with ways to embed and immerse a given manifold M into a manifold M' of a particular type with lowest possible dimension. M' is usually the Euclidean space $\mathbf{R}^n$, the projective space $P^N\mathbf{R}$ or a certain standard manifold. The theory was initiated by Whitney (1936). He proved by "general position" argument that an n-dimensional C^∞-manifold M with countable basis can always be immersed in the $2n$-dimensional Euclidean space and can always be embedded in the $(2n+1)$-dimensional Euclidean space as a closed set (**Whitney's theorem**).

L. Submanifolds

A C^∞-manifold M is said to be a **submanifold** of a C^∞-manifold M' if M is a subset of M' and the identity mapping of M into M' is an immersion. If the identity mapping of M into M' is an embedding, then M is called a **regular submanifold** of M'. A regular submanifold M of M' is called a **closed submanifold** if M is a closed subset of M'.

Let φ be a C^∞-mapping from M into M' and M'' be a submanifold of M'. Then for each $q \in M''$ the tangent space T_q'' of M'' at q is a linear subspace of the tangent space T_q' of M' at q. Denote by π_q the projection of quotient vector space onto T_q'/T_q''. A C^∞-mapping φ is called **transverse** to M'' if for each $p \in \varphi^{-1}(M'')$ the composite $\pi_{\varphi(p)} \circ d\varphi_p: T_p \to T'_{\varphi(p)}/T''_{\varphi(p)}$ is surjective. If φ is transverse to M'' then $\varphi^{-1}(M'')$ is a submanifold of M. For any C^∞-mapping $\varphi: M \to M'$ and any submanifold $M'' \subset M'$, we can find a C^∞-mapping $\varphi': M \to M'$ which is transverse to M'' and arbitrarily close to φ (**transversality theorem**). Let M_1 and M_2 be submanifolds of M'. Then we say M_1 **intersects transversely** to M_2 if the inclusion $M_1 \subset M'$ is transverse to M_2.

A C^∞-mapping is called a **submersion** if it has no critical point. Let φ be a submersion from M into M'. Then for each point $q \in M'$, $\varphi^{-1}(q)$ is a regular submanifold of M, and M is covered by a mutually disjoint family of submanifolds: $M = \bigcup_{q \in M'} \varphi^{-1}(q)$.

Let M be a submanifold of an n-dimensional Euclidean space $\mathbf{R}^m$. We can identify the tangent vector space T_p of M at p with the geom-

etric tangent space of M at p in the Euclidean space $\mathbf{R}^m$. A vector in $\mathbf{R}^m$ that is orthogonal to the tangent vector space T_p of M at p is called a **normal vector** to M at p. The set of all vectors normal to M forms a vector bundle over M, which we call the **normal vector bundle** (or **normal bundle**) of M. If M is compact, then the totality of vectors normal to M whose length is $\leqslant \varepsilon$ (where ε is a sufficiently small positive real number) forms a neighborhood $N(M)$ of M in $\mathbf{R}^m$ which we call a **tubular neighborhood** of M. Int $N(M)$ is called an **open tubular neighborhood**.

M. Vector Fields

Let N be a subset of a C^∞-manifold M. By a **vector field** on N we mean a mapping X that assigns to each point p of N a tangent vector X_p of M at p. We can consider X as a †cross section over N of the tangent vector bundle of M. Let X be a vector field in M, and let f be a C^∞-function in M. Then we can define a function Xf in M by $(Xf)(p)=X_pf$. We call X a **vector field of class** C^r if the function Xf is of class C^r for any C^∞-function f in M. Let $(x^1,\ldots,x^n)$ be the local coordinate system in a coordinate neighborhood (U,ψ), and let $(\partial/\partial x^i)_pf=(\partial f/\partial x^i)(p)$ for $p\in U$ and $f\in\mathfrak{F}(M)$. Then $\partial/\partial x^i$ $(i=1,\ldots,n)$ are vector fields in U, and the $(\partial/\partial x^i)_p$ form a basis of T_p at every point $p\in U$. A vector field X in U is written uniquely as $X_p=\sum_i\xi^i(p)(\partial/\partial x^i)_p$ at each point $p\in U$. Then $\xi^1,\ldots,\xi^n$ are real-valued functions defined in U, called the **components** of X with respect to the local coordinate system $(x^1,\ldots,x^n)$. A vector field X in M is of class C^r if and only if its components ξ^i with respect to each coordinate system are functions of class C^r $(0\leqslant r\leqslant\infty)$. Let $(\bar{x}^1,\ldots,\bar{x}^n)$ be another local coordinate system in a neighborhood U of p, and let $(\bar{\xi}^1,\ldots,\bar{\xi}^n)$ be the components of X with respect to $(\bar{x}^1,\ldots,\bar{x}^n)$. Then we have $\bar{\xi}^i(q)=\sum_j(\partial\bar{x}^i/\partial x^j)(q)\xi^j(q)$ at each point $q\in U$.

For the rest of this article we mean by a vector field in M a vector field of class C^∞, and we denote by $\mathfrak{X}(M)$ the set of all vector fields in M. Then $\mathfrak{X}(M)$ is an $\mathfrak{F}(M)$-†module, where $\mathfrak{F}(M)$ denotes the algebra of all C^∞-functions in M. In fact, for $f,g\in\mathfrak{F}(M)$ and $X, Y\in\mathfrak{X}(M)$, we can define a vector field $fX+gY$ by $(fX+gY)_p=f(p)X_p+g(p)Y_p$, and this defines an $\mathfrak{F}(M)$-module structure in $\mathfrak{X}(M)$.

In a coordinate neighborhood (U,ψ), we can write $X=\sum_i\xi^i(\partial/\partial x^i)$. The right-hand side of this equation is sometimes called the **symbol** of the vector field X. A vector field X can also be interpreted as a linear differential operator that acts on $\mathfrak{F}(M)$.

Let X and Y be vector fields in M. Then there exists a unique vector field Z in M such that $Zf=X(Yf)-Y(Xf)$ for any C^∞-function f in M. We denote Z by $[X,Y]$ and call it the **Poisson bracket** (or simply **bracket**) of X and Y. If ξ^i and η^i denote the components of X and Y, respectively, in a coordinate neighborhood (U,ψ), then the components ζ^i of $[X,Y]$ are given by $\zeta^i=\sum_k\{\xi^k(\partial\eta^i/\partial x^k)-\eta^k(\partial\xi^i/\partial x^k)\}$. The bracket of vector fields has the following properties: (i) $[X,Y]f=X(Yf)-Y(Xf)$, (ii) $[fX,gY]=fg[X,Y]+f(Xg)Y-g(Yf)X$, (iii) $[X+Y,Z]=[X,Z]+[Y,Z]$, (iv) $[X,Y]=-[Y,X]$, and (v) $[[X,Y],Z]+[[Y,Z],X]+[[Z,X],Y]=0$ (**Jacobi identity**). These identities show that $\mathfrak{X}(M)$ is a Lie algebra (→ 248 Lie Algebras) over $\mathbf{R}$.

If φ is a diffeomorphism of M onto M', then for any vector field X in M we can define a vector field φ_*X in M' by the condition $(\varphi_*X)_p=d\varphi_q(X_q)$, $p=\varphi(q)$. Then φ_* is an isomorphism of the Lie algebra $\mathfrak{X}(M)$ onto the Lie algebra $\mathfrak{X}(M')$.

N. Vector Fields and One-Parameter Groups of Transformations

A **one-parameter group of transformations** of M is a family φ_t $(t\in\mathbf{R})$ of diffeomorphisms satisfying the following two conditions: (i) the mapping of $\mathbf{R}\times M$ into M defined by $(t,p)\rightarrow\varphi_t(p)$ is of class C^∞; and (ii) $\varphi_s\circ\varphi_t=\varphi_{s+t}$ for $s,t\in\mathbf{R}$.

Let φ_t be a one-parameter group of transformations of M. Then we can define a vector field X by $X_pf=\lim_{t\rightarrow0}(f(\varphi_t(p))-f(p))/t$, where $p\in M$ and $f\in\mathfrak{F}(M)$. The vector field X thus defined is called the **infinitesimal transformation** of φ_t. We also say that φ_t is generated by X, and sometimes we denote φ_t by the symbol $\exp tX$. In this case, if $(x^1,\ldots,x^n)$ is a local coordinate system, then at each point p of the coordinate neighborhood, we have $X_p=\sum_i(dx^i(\varphi_t(p))/dt)_{t=0}(\partial/\partial x^i)_p$.

If M is compact, then every vector field in M is the infinitesimal transformation of a one-parameter group of transformations; that is, every vector field generates a one-parameter group of transformations. For M noncompact, this is not always true. Nevertheless, for each vector field X we have the following result concerning local properties of X: For each point p of M, there exist a neighborhood U of p, a positive real number ε, and a family $\varphi_t(|t|<\varepsilon)$ of mappings of U into M satisfying the following three conditions. (1) The mapping of $(-\varepsilon,\varepsilon)\times U$ into M defined by $(t,q)\rightarrow\varphi_t(q)$ is of class C^∞, and for each fixed t, φ_t is a diffeomorphism of U onto an open set $\varphi_t(U)$ of M. (2) If $|s|$, $|t|$, and $|s+t|$ are all smaller than ε and q and $\varphi_t(q)$ both belong to U, then

$\varphi_s(\varphi_t(q))=\varphi_{s+t}(q)$. (3) $X(q)f=\lim_{t\to 0}(f(\varphi_t(q))-f(q))/t$ for $q\in U$ and $f\in\mathfrak{F}(M)$. We call φ_t the **local one-parameter group of local transformations** around p generated by X.

Let X and Y be vector fields in M, and let φ_t be the local one-parameter group of local transformations around p generated by X. Then $[X,Y]_p=\lim_{t\to 0}(Y_p-((\varphi_t)_*Y)_p)/t$, where $(\varphi_t)_*Y$ is a vector field defined as follows: Let U be a neighborhood of the point p where φ_t ($|t|<\varepsilon$) is defined. Then $(\varphi_t)_*Y$ is the vector field on $\varphi_t(U)$ that is the image of Y under the diffeomorphism φ_t. In particular, if X generates a one-parameter group of transformations of M, then we have $[X,Y]=\lim_{t\to 0}(Y-(\varphi_t)_*Y)/t$ for any vector field Y in M.

O. Tensor Fields

Let $T_s^r(p)$ be the vector space consisting of all r-times contravariant and s-times covariant tensors over the tangent vector space T_p of a C^∞-manifold M, that is,

$$T_s^r(p)=\left(\overset{r}{\bigotimes}T_p\right)\otimes\left(\overset{s}{\bigotimes}T_p^*\right),$$

where T_p^* denotes the dual linear space of T_p (→ 256 Linear Spaces). A **tensor field** (more precisely, **contravariant of order** r **and covariant of order** s, or simply a **tensor field of type** (r,s)) on a subset N of M is a mapping K that assigns to each point p of N an element K_p of the vector space $T_s^r(p)$. In particular, if $r=s=0$, K is a real-valued function on N, and we call K a **scalar field**. If $r=1$ and $s=0$, K is a vector field, called a **contravariant vector field**. When $r=0$ and $s=1$, we call K a **covariant vector field** (or **differential form of degree** 1). If $r\neq 0$, $s=0$ or $r=0$, $s\neq 0$, we call K a **contravariant tensor field of order** r or a **covariant tensor field of order** s, respectively. A contravariant or covariant tensor field K is said to be **symmetric (alternating)** if K_p is a symmetric (alternating) tensor at every point p of M.

Let $(x^1,\ldots,x^n)$ be the local coordinate system in a coordinate neighborhood (U,ψ). Then at each point p of U, the $(\partial/\partial x^i)_p$ $(i=1,\ldots,n)$ form a basis of the tangent vector space T_p, the differentials $(dx^i)_p$ $(i=1,\ldots,n)$ form a basis of the dual space T_p^*, and these bases are dual to each other. A tensor field K of type (r,s) defined on M is written at any point p of U in the following form:

$$K_p=\sum K^{i_1\ldots i_r}_{j_1\ldots j_s}(p)(\partial/\partial x^{i_1})_p\otimes\ldots\otimes(\partial/\partial x^{i_r})_p \otimes(dx^{j_1})_p\otimes\ldots\otimes(dx^{j_s})_p.$$

The functions $K^{i_1\ldots i_r}_{j_1\ldots j_s}$ defined in U are called the **components** of the tensor field K of type (r,s) with respect to the local coordinate system $(x^1,\ldots,x^n)$. If $\bar{K}^{i_1\ldots i_r}_{j_1\ldots j_s}$ are the components of K with respect to the local coordinate system $(\bar{x}^1,\ldots,\bar{x}^n)$ in another coordinate neighborhood (U',ψ') such that $U\cap U'\neq\varnothing$, then for each $q\in U\cap U'$, the following relations hold:

$$\bar{K}^{i_1\ldots i_r}_{j_1\ldots j_s}(q)=\sum_{k,l}(\partial\bar{x}^{i_1}/\partial x^{k_1})_q\ldots(\partial\bar{x}^{i_r}/\partial x^{k_r})_q \times(\partial x^{l_1}/\partial\bar{x}^{j_1})_q\ldots(\partial x^{l_s}/\partial\bar{x}^{j_s})_q K^{k_1\ldots k_r}_{l_1\ldots l_s}.$$

A tensor field K in M is called a **tensor field of class** C^t $(0\leqslant t\leqslant\infty)$ if the components are functions of class C^t for any coordinate neighborhood of M.

The sum $K+L$ and the tensor product $K\otimes L$ of two tensor fields K and L are defined by the rules $(K+L)_p=K_p+L_p$ and $(K\otimes L)_p=K_p\otimes L_p$, respectively. The contraction of two tensor fields is also defined by taking the contraction pointwise.

Let φ be a diffeomorphism of M into M'. Then the differential $(d\varphi)_q$ is an isomorphism of T_q onto T_p $(p=\varphi(q))$ for each $q\in M$ and hence induces an isomorphism $\tilde{\varphi}_q$ of the vector space $T_s^r(q)$ onto the vector space $T_s^r(p)$ (→ 256 Linear Spaces). For any tensor field K in M we can define a tensor field $\tilde{\varphi}K$ in M' by $(\tilde{\varphi}K)_p=\tilde{\varphi}_q(K_q)$, $p=\varphi(q)$, $q\in M$. Then $\tilde{\varphi}(K+L)=\tilde{\varphi}K+\tilde{\varphi}L$, $\tilde{\varphi}(K\otimes L)=\tilde{\varphi}K\otimes\tilde{\varphi}L$, and the mapping $\tilde{\varphi}$ commutes with contraction.

Let K be a tensor field and X be a vector field (both of class C^∞) in M. We define a tensor field L_XK by $(L_XK)_p=\lim_{t\to 0}(K_p-(\tilde{\varphi}_tK)_p)/t$, where φ_t denotes the local one-parameter group of local transformations around p generated by X. We call L_XK the **Lie derivative** of K with respect to the vector field X. The operator $L_X:K\to L_XK$ has the following six properties: (i) $L_X(K+K')=L_XK+L_XK'$; (ii) $L_X(K\otimes K')=(L_XK)\otimes K'+K\otimes L_XK'$; (iii) the operator L_X commutes with contraction; (iv) $L_Xf=Xf$ for a scalar field f and $L_XY=[X,Y]$ for a vector field Y; (v) $L_{[X,Y]}=L_XL_Y-L_YL_X$, that is, $L_{[X,Y]}K=L_X(L_YK)-L_Y(L_XK)$; and (vi) K is invariant under φ_t, i.e., $\tilde{\varphi}_tK=K$ for all t, if and only if $L_XK=0$.

Let K be a covariant tensor field of order r in M. We always assume that K is of class C^∞. The value K_p of K at $p\in M$ is an element of the vector space $T_p^*\otimes\ldots\otimes T_p^*$ (r times tensor product of T_p^*); hence we may consider K_p an r-linear mapping of T_p into $\mathbf{R}$ (→ 256 Linear Spaces). If $X_1,\ldots,X_s$ are vector fields in M, we define a C^∞-function $K(X_1,\ldots,X_r)$ by $K(X_1,\ldots,X_r)(p)=K_p((X_1)_p,\ldots,(X_r)_p)$. Then the mapping that assigns to each r-tuple $(X_1,\ldots,X_r)$ of vector fields the C^∞-function $K(X_1,\ldots,X_r)$ is an r-linear mapping on the $\mathfrak{F}(M)$-module $\mathfrak{X}(M)$ consisting of all vector fields of class C^∞ in M into $\mathfrak{F}(M)$; that is, $K(X_1,\ldots,fX_i+gY_i,\ldots,X_r)=fK(X_1,\ldots,X_i,$

$\ldots, X_r) + gK(X_1, \ldots, Y_i, \ldots, X_r)$ $(i=1, \ldots, r)$ for $f, g \in \mathfrak{F}(M)$. Conversely any r-linear mapping of the $\mathfrak{F}(M)$-module $\mathfrak{X}(M)$ into $\mathfrak{F}(M)$ can be interpreted as a covariant tensor field of order r in M. If the tensor field K is symmetric (alternating), the corresponding r-linear mapping $K(X_1, \ldots, X_r)$ is symmetric (alternating) with respect to $X_1, \ldots, X_r$. For the Lie derivative $L_X K$ of a covariant tensor field of order r of K, we have the following formula: $(L_X K)(X_1, \ldots, X_r) = X(K(X_1, \ldots, X_r)) - \sum_{i=1}^r K(X_1, \ldots, [X, X_i], \ldots, X_r)$.

P. Riemannian Metrics

A symmetric covariant tensor field g of order 2 and of class C^∞ in M is called a **pseudo-Riemannian metric** if the symmetric bilinear form g_p on the tangent vector space T_p is nondegenerate at each point $p \in M$; and g is called a **Riemannian metric** if g_p is positive definite for all p. If g is a Riemannian metric, the length $\|L\|$ of a tangent vector $L \in T_p$ is defined by $\|L\|^2 = g_p(L, L)$. On a paracompact C^∞-manifold there always exists a Riemannian metric. A pair consisting of a differentiable manifold and a Riemannian metric on it is called a **Riemannian manifold** (→ 364 Riemannian Manifolds).

Q. Differential Forms

An alternating covariant tensor field in M of order r and of class C^t $(0 \leqslant t \leqslant \infty)$ is also called a **differential form** (or **exterior differential form**) **of degree** r. A differential form of degree 1 is sometimes called a **Pfaffian form**. Let ω be a differential form of degree r. Since each alternating covariant tensor of order r at a point p is an element of $\bigwedge^r T_p^*$, the r-fold †exterior product of T_p^*, the form ω is a mapping that sends each point p of M to an element ω_p of $\bigwedge^r T_p^*$. We can also regard ω as an alternating r-linear mapping of $\mathfrak{X}(M)$ into $\mathfrak{F}(M)$. Let $(x^1, \ldots, x^n)$ be the local coordinate system in a local coordinate neighborhood (U, ψ). Since $(dx^i)_p$ $(i=1, \ldots, n)$ is a basis of T_p^* at each point p of U, we can express ω_p $(p \in U)$ uniquely in the form

$$\omega_p = \sum_{i_1 < \ldots < i_r} a_{i_1 \ldots i_r}(p)(dx^{i_1})_p \wedge \ldots \wedge (dx^{i_r})_p,$$

where the sum extends over all ordered r-tuples $(i_1, \ldots, i_r)$ of indices such that $1 \leqslant i_1 < i_2 < \ldots < i_r \leqslant n$. For an ordered r-tuple $(i_1, \ldots, i_r)$ of indices with repeated indices we put $a_{i_1 \ldots i_r} = 0$, and for $(i_1, \ldots, i_r)$ with r distinct indices we put $a_{i_1 \ldots i_r} = (\operatorname{sgn} \sigma) a_{j_1 \ldots j_r}$, where $(j_1, \ldots, j_r)$ $(j_1 < \ldots < j_r)$ is a permutation of $(i_1, \ldots, i_r)$ and $\operatorname{sgn} \sigma$ denotes the sign of the permutation $\sigma: i_k \to j_k$ $(k=1, \ldots, r)$. Then we can write

$$\omega_p = \frac{1}{r!} \sum_{i_1, \ldots, i_r = 1}^{n} a_{i_1 \ldots i_r}(p)(dx^{i_1})_p \wedge \ldots \wedge (dx^{i_r})_p.$$

The functions $a_{i_1 \ldots i_r}$ are the components of the tensor field ω, and ω is of class C^t if these components are of class C^t for any coordinate neighborhood. By the **support** (or **carrier**) of a differential form ω we mean the closure of the subset of M consisting of all p such that $\omega_p \neq 0$.

In the rest of this article, differential forms are always of class C^∞, and we denote by $\mathfrak{D}^r(M)$ the real vector space consisting of all differential forms of degree r and of class C^∞. In particular, $\mathfrak{D}^0(M) = \mathfrak{F}(M)$ and $\mathfrak{D}^r(M) = \{0\}$ for $r > n$, $n = \dim M$.

For differential forms we have the following five important operations.

(1) Exterior product. Let ω and θ be differential forms of degree r and s, respectively. The **exterior product** $\omega \wedge \theta$ of ω and θ is the differential form of degree $r+s$ defined by $(\omega \wedge \theta)_p = \omega_p \wedge \theta_p$, $p \in M$. Let $X_1, \ldots, X_{r+s}$ be $r+s$ vector fields in M. Then we have

$$(\omega \wedge \theta)(X_1, \ldots, X_{r+s})$$
$$= \sum \operatorname{sgn}(i;j) \omega(X_{i_1}, \ldots, X_{i_r}) \theta(X_{j_1}, \ldots, X_{j_s}),$$

where the summation runs over all possible partitions of $(1, 2, \ldots, r+s)$ such that $i_1 < i_2 < \ldots < i_r$ and $j_1 < j_2 < \ldots < j_s$, and $\operatorname{sgn}(i;j)$ means the sign of the permutation $(1, 2, \ldots, r+s) \to (i_1, \ldots, i_r, j_1, \ldots, j_s)$. In particular, if $\omega_1, \ldots, \omega_r$ are differential forms of degree 1, then we have $(\omega_1 \wedge \ldots \wedge \omega_r)(X_1, \ldots, X_r) = \det(\omega_i(X_j))$.

(2) Exterior differentiation. Let ω be a differential form of degree r, and let $\omega = (1/r!) \sum a_{i_1 \ldots i_r} dx^{i_1} \wedge \ldots \wedge dx^{i_r}$ in a coordinate neighborhood (U, ψ). Then we can define a differential form $d\omega$ of degree $r+1$ by the condition $d\omega = (1/r!) \sum da_{i_1 \ldots i_r} \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_r}$ in U. The differential form $d\omega$ is called the **exterior derivative** (or **exterior differential**) of ω. A differential form ω satisfying the condition $d\omega = 0$ is called a **closed differential form**, and a differential form η that can be expressed as $\eta = d\omega$ for some ω is called an **exact differential form**. If $a, b \in \mathbf{R}$ and $\omega, \omega' \in \mathfrak{D}^r(M)$, then we have $d(a\omega + b\omega') = a d\omega + b d\omega'$. Therefore the set $\mathfrak{C}^r(M)$ of all closed differential forms of degree r and the set $\mathfrak{E}^r(M)$ of all exact differential forms of degree r are linear subspaces of $\mathfrak{D}^r(M)$.

For the exterior derivative $d\omega$, we have the following formula:

$$(d\omega)(X_1, \ldots, X_r)$$
$$= \sum_{i=1}^{r+1} (-1)^{i+1} X_i(\omega(X_1, \ldots, \hat{X}_i, \ldots, X_{r+1}))$$
$$+ \sum_{i<j} (-1)^{i+j} \omega([X_i X_j], X_1, \ldots, \hat{X}_i, \ldots, \hat{X}_j, \ldots, X_{r+1}),$$

where the variables under the sign ˆ are to be omitted.

(3) Interior product with vector field. Let ω be a differential form of degree r and X a vector field. When $r \geqslant 1$, we can define a differential form $\iota(X)\omega$ of degree $r-1$ by the formula $(\iota(X)\omega)(X_1, \ldots, X_{r-1}) = \omega(X, X_1, \ldots, X_{r-1})$ for any $r-1$ vector fields $X_1, \ldots, X_{r-1}$; if $r=0$, we put $\iota(X)\omega = 0$. The differential form $\iota(X)\omega$ is called the **interior product** of ω with X.

(4) Lie derivative. The **Lie derivative** $L_X\omega$ of a differential form of degree r with respect to a vector field X is a differential form of the same degree. For any r vector fields $X_1, \ldots, X_r$ we have, by definition, $(L_X\omega)(X_1, \ldots, X_r) = X(\omega(X_1, \ldots, X_r)) - \sum_{i=1}^r \omega(X_1, \ldots, [X, X_i], \ldots, X_r)$.

(5) Let φ be a C^∞-mapping of M into M', and let $\varphi_p^*: T_{\varphi(p)}^* \to T_p^*$ be the †transpose (or †dual) of the linear mapping $(d\varphi)_p: T_p \to T_{\varphi(p)}$ for $p \in M$, i.e., the mapping defined by the condition $((d\varphi)_p L, \alpha) = (L, \varphi_p^* \alpha)$ for each $L \in T_p$, $\alpha \in T_{\varphi(p)}^*$. We denote the linear mapping of $\bigwedge T_{\varphi(p)}^*$ into $\bigwedge T_p^*$ induced by φ_p^* by φ_p^* also. Let ω be a differential form of degree r in M. Then a differential form $\varphi^*\omega$ in M, the **pullback** by φ of ω, is defined by $(\varphi^*\omega)_p = \varphi_p^* \omega_{\varphi(p)}$, $p \in M$.

The operations defined previously satisfy the following six important relations: (i) $d^2 = 0$, that is, $d(d\omega) = 0$; (ii) $d(\omega \wedge \theta) = d\omega \wedge \theta + (-1)^r \omega \wedge d\theta$, where ω is of degree r; (iii) $\varphi^*(\omega \wedge \theta) = \varphi^*\omega \wedge \varphi^*\theta$, $\varphi^*(d\omega) = d(\varphi^*\omega)$; (iv) $L_X(\omega \wedge \theta) = (L_X\omega) \wedge \theta + \omega \wedge (L_X\theta)$; (v) $L_X = \iota(X) \cdot d + d \cdot \iota(X)$, $L_X(d\omega) = d(L_X\omega)$; and (vi) $L_{[X,Y]} = L_X \cdot L_Y - L_Y \cdot L_X$, $\iota([X, Y]) = L_X \cdot \iota(Y) - \iota(Y) \cdot L_X$.

R. De Rham Cohomology

Let $\mathfrak{D}(M) = \sum_{r=0}^n \mathfrak{D}^r(M)$, where $n = \dim M$. Then $\mathfrak{D}(M)$ is a †cochain complex with †coboundary operator d. We denote by $H^r(\mathfrak{D})$ the r-dimensional cohomology group of this cochain complex, and we call it the r-dimensional **de Rham cohomology group** of the differentiable manifold M. If we denote by $\mathfrak{C}^r(M)$ and $\mathfrak{E}^r(M)$ the subspaces of $\mathfrak{D}^r(M)$ consisting of closed differential forms and exact differential forms, respectively, then $H^r(\mathfrak{D}) = \mathfrak{C}^r(M)/\mathfrak{E}^r(M)$ $(0 \leqslant r \leqslant n)$ by definition. If $\omega \in \mathfrak{C}^i(M)$ and $\theta \in \mathfrak{C}^r(M)$, then $\omega \wedge \theta \in \mathfrak{C}^{i+r}(M)$, and if $\omega \in \mathfrak{C}^i(M)$ and $\theta \in \mathfrak{E}^r(M)$ (or $\omega \in \mathfrak{E}^i(M)$ and $\theta \in \mathfrak{C}^r(M)$), then $\omega \wedge \theta \in \mathfrak{E}^{i+r}(M)$. So if we put $H(\mathfrak{D}) = \sum_{r=1}^n H^r(\mathfrak{D})$ (direct sum), we can define a product in $H(\mathfrak{D})$ by $[\omega] \cdot [\theta] = [\omega \wedge \theta]$ for each $[\omega] \in H^i(\mathfrak{D})$, $[\theta] \in H^r(\mathfrak{D})$. With respect to this product, $H(\mathfrak{D})$ forms an algebra over $\mathbf{R}$ called the **de Rham cohomology ring** of M.

S. Partitions of Unity

Let M be a paracompact C^∞-manifold, and let $\{V_i\}_{i \in I}$ be a †locally finite open covering of M such that the closure of V_i is compact for each index i. Then there exists a C^∞-function f_i in M for each i satisfying the following three conditions: (i) $0 \leqslant f_i \leqslant 1$, (ii) the support of f_i is contained in V_i, and (iii) $\sum_{i \in I} f_i(x) = 1$ for every $x \in M$. The family of C^∞-functions f_i, $i \in I$, is called a **partition of unity of class C^∞** subordinate to the open covering $\{V_i\}_{i \in I}$.

T. Integrals of Differential Forms

Integrals over an Oriented Manifold. Let M be an n-dimensional paracompact oriented C^∞-manifold and ω a differential form of degree n in M with compact support. We can choose a positively oriented atlas $S = \{(U_\alpha, \psi_\alpha)\}_{\alpha \in A}$ such that $\{U_\alpha\}_{\alpha \in A}$ is a locally finite covering; $\bar{U}_\alpha$ is compact for each α. Suppose first that the support of ω is contained in U_α for some index α. Then $(\psi_\alpha^{-1})^*\omega_\alpha$ is a differential form of degree n in $\psi_\alpha(U_\alpha)$, and we can express $(\psi_\alpha^{-1})^*\omega_\alpha$ in the form $a\,dx^1 \wedge \ldots \wedge dx^n$, where $(x^1, \ldots, x^n)$ are the coordinates in $\mathbf{R}^n$ such that $x_\alpha^i = x^i \circ \psi_\alpha$ $(i = 1, \ldots, n)$ give a local coordinate system compatible with the orientation of M and a is a C^∞-function with compact support. Then we define the integral of ω over M by

$$\int_M \omega = \int_{\psi_\alpha(U_\alpha)} a\,dx^1 \ldots dx^n.$$

For the general case let $\{f_\alpha\}_{\alpha \in A}$ be a partition of unity of class C^∞ subordinate to $\{U_\alpha\}_{\alpha \in A}$. Then the support of $f_\alpha\omega$ is contained in U_α, and except for a finite number of the indices α, $f_\alpha\omega$ vanishes identically. Therefore we may define the integral of ω over M by

$$\int_M \omega = \sum_\alpha \int_M f_\alpha \omega,$$

and we can show that this definition of the integral is independent of the choice of oriented atlas S and of a partition of unity subordinate to S.

Integrals over a Singular Chain. We fix rectangular coordinates in $\mathbf{R}^r$. Let d_0 be the origin and d_i be the unit point on the ith coordinate axis. Let S^r denote the oriented r-simplex $(d_0, d_1, \ldots, d_r)$ with vertices $d_0, d_1, \ldots, d_r$. When we regard S^r as a point set, we denote it by $|S^r|$. An **oriented singular r-simplex of class C^∞** in M is, by definition, a pair (S^r, φ) consisting of S^r and a C^∞-mapping φ of an open neighborhood of $|S^r|$ into M. An element of the free $\mathbf{Z}$-module generated by singular r-simplexes of class C^∞ is called an integral **singular r-chain**

of class C^∞ in M. We define a real singular r-chain of class C^∞ analogously. Let ω be a differential form of degree r and (S^r, φ) be an oriented singular r-simplex of class C^∞ in M. Then $\varphi^*\omega$ is a differential form of degree r in a neighborhood of $|S^r|$, and we can express $\varphi^*\omega$ in the form $\varphi^*\omega = a\,dx^1 \wedge dx^2 \wedge \ldots \wedge dx^r$. We define the integral of ω over (S^r, φ) by

$$\int_{(S^r,\varphi)} \omega = \int_{|S^r|} a\,dx^1 \ldots dx^r,$$

and the integral of ω over a singular r-chain C of class C^∞ by

$$\int_C \omega = \sum_i m_i \int_{(S^r,\varphi_i)} \omega,$$

where $C = \sum_i m_i(S^r, \varphi_i)$, $m_i \in \mathbf{Z}$ (or $m_i \in \mathbf{R}$). When $r=0$, then ω is a function in M, and S^0 is a point o. In this case we put $\int_C \omega = \sum_i m_i \omega(\varphi_i(o))$.

Let $C_r(S, \mathbf{Z})$ ($C_r(S, \mathbf{R})$) be the $\mathbf{Z}$-module (vector space over $\mathbf{R}$) of integral (real) singular r-chains of class C^∞ in M, and let $\omega(C)$ be the value of the integral of ω over a chain C. Then ω is a linear function in the vector space $C_r(S, \mathbf{R})$, and hence we can consider ω a **singular r-cochain of class** C^∞.

U. Stokes's Formulas

(1) Let D be a †domain in an n-dimensional C^∞-manifold M, and let ∂D and $\bar{D}$ be the boundary and the closure of D, respectively. Let $S = \{(U_\alpha, \psi_\alpha)\}_{\alpha \in A}$ be an atlas of class C^∞ of M, $U'_\alpha = U_\alpha \cap \bar{D}$, ψ'_α be the restriction of ψ_α to U'_α, and $T = \{(U'_\alpha, \psi'_\alpha)\}_{\alpha \in A}$. If the pair $(\bar{D}, T)$ is a C^∞-manifold with boundary under a suitable choice of S, then the domain D is called a **domain with regular** (or **smooth**) **boundary**. The boundary ∂D of $(\bar{D}, T)$ is then an $(n-1)$-dimensional closed submanifold of M, and if M is orientable, ∂D is also orientable. Now let M be a paracompact and oriented manifold and D be a domain with regular boundary. Let C be a characteristic function of D in M, i.e., a function defined by the condition $C(p) = 1$ for $p \in D$ and $C(p) = 0$ for $p \notin D$. Let θ be a differential form of degree n in M with compact support. We define the integral of θ over D by

$$\int_D \theta = \int_M C \cdot \theta.$$

Let ω be a differential form of degree $n-1$ in M with compact support. We then have **Stokes's formula**:

$$\int_D d\omega = \int_{\partial D} i^*\omega,$$

where i denotes the identity mapping of the submanifold ∂D into M with ∂D having the orientation induced naturally from that of M.

(2) Let C be a singular r-chain of class C^∞ in M, and let ∂C be the boundary of C. Then for any differential form ω of degree $r-1$, we have

$$\int_C d\omega = \int_{\partial C} \omega.$$

This formula is also called **Stokes's formula**.

V. De Rham's Theorem

Let M be a connected paracompact C^∞-manifold. If we consider ω as a singular cochain, we have $(d\omega)(C) = \omega(\partial C)$; by Stokes's formula, this means that the exterior differential $d\omega$ of ω is equal to the coboundary of the singular cochain ω. Let ω and C be a closed differential form of degree r and a singular r-cycle of class C^∞, respectively, and let $[\omega]$ and $[C]$ be the de Rham cohomology class and the singular homology class represented by ω and C. Using Stokes's formula, we can define the inner product $([\omega], [C])$ by

$$([\omega], [C]) = \int_C \omega.$$

Through this inner product, it follows that the de Rham cohomology group $H^r(\mathfrak{D})$ is isomorphic to the rth singular cohomology group $H^r(M, \mathbf{R})$, the dual space of the rth homology group of the complex of real singular chains of class C^∞. Moreover, the de Rham cohomology ring $H(\mathfrak{D})$ is isomorphic to the singular cohomology ring $H^*(M, R)$ (**de Rham's theorem**).

W. Divergence of a Vector Field

Let M be an n-dimensional oriented C^∞-manifold, and let S be an oriented atlas. Let ω be a differential form of degree n, and let $(x^1, \ldots, x^n)$ be the local coordinate system in a coordinate neighborhood in S. Then we can express ω in the coordinate neighborhood uniquely in the form $\omega = a\,dx^1 \wedge \ldots \wedge dx^n$. If the function a is positive for any coordinate neighborhood in S, we call ω a **volume element** of M. In a paracompact oriented manifold, there always exists a volume element. (We remark that an n-dimensional differentiable manifold M is orientable if and only if there exists an everywhere nonvanishing differential form of degree n.) Let f be a C^∞-function in M with compact support. Then $f \cdot \omega$ is a differential form of degree n with compact support, and so the integral

$$\int_M f \cdot \omega$$

is defined. We call this integral the **integral of the function f with respect to the volume element ω.**

Let g be a Riemannian metric in M and g_{ij} the components of g with respect to the local coordinate system $(x^1, \ldots, x^n)$ as before. Then we can define a volume element ω in M by putting $\omega = \sqrt{G}\, dx^1 \wedge \ldots \wedge dx^n$, $G = \det(g_{ij})$ in each coordinate neighborhood. The volume element thus defined is called the **volume element associated with the Riemannian metric g.**

Let ω be a volume element and X a vector field in M. Then the Lie derivative $L_X\omega$ is also a differential form of degree n, and we can express $L_X\omega$ in the form $L_X\omega = f_X \cdot \omega$, where f_X is a scalar field, i.e., a function in M. We call f_X the **divergence** of the vector field X with respect to the volume element ω and denote it by $\operatorname{div} X$. If ω is associated with a Riemannian metric, then $\operatorname{div} X$ is called the **divergence** of X with respect to the Riemannian metric.

If M is compact, we have

$$\int_M \operatorname{div} X \cdot \omega = 0$$

for any vector field X. This result is known as **Green's theorem.**

X. Jets

Let M and N be C^∞-manifolds. We define an equivalence relation in the set of all C^∞-mappings of M into N. Let f and g be such mappings and p be a point of M. Choosing local coordinate systems, we write $f(p) = (f_1(x), \ldots, f_n(x))$, $g(p) = (g_1(x), \ldots, g_n(x))$, $x = (x_1, \ldots, x_m)$. We say f and g are equivalent at p if $f(p)$, $g(p)$, and the values at p of all the partial derivatives of f_i and g_i up to the order r (r an integer, $r \geqslant 0$) are equal ($i = 1, \ldots, n$). An equivalence class with respect to this equivalence relation is called a **jet of order** r at p. A jet of order r at p represented by a function f is denoted by $j_p^r f$, and the points p and $f(p)$ are called the **source** and the **target** of the jet $j_p^r f$, respectively. We denote by $J_p^r(M, N)$ the set of all jets of order r with source at p and target in N and let $J^r(M, N) = \bigcup_{p \in M} J_p^r(M, N)$. For any jet j, let $\pi_s(j)$ and $\pi_t(j)$ denote the source and the target of j, respectively. We can introduce the structure of a C^∞-manifold in $J^r(M, N)$ in a natural way such that the projections $\pi_s : J^r(M, N) \to M$ and $\pi_t : J^r(M, N) \to N$ are both of class C^∞ and $J^r(M, N)$ is a fiber bundle over $M(N)$ with projection $\pi_s(\pi_t)$. As examples, we have:

(1) $J_0^1(\mathbf{R}, N)$ is identified with the tangent vector bundle of N.

(2) The set $J^r(\xi)$ of all jets of order r determined by the sections of a vector bundle ξ of class C^∞ is also a vector bundle of class C^∞.

Let $f: M \to N$ and $g: N \to L$ be C^∞-mappings. We define a composition of jets by $j_{f(p)}^r g \cdot j_p^r f = j_p^r(g \circ f)$. A jet $j_p^r f \in J^r(M, N)$ is **invertible** if there exists a mapping $g: N \to M$ such that $j_{f(p)}^r g \cdot j_p^r f = j_p^r(1_M)$, where 1_M denotes the identity mapping of M onto itself. We denote by $I^r(M, N)$ the set of all invertible jets in $J^r(M, N)$, and put $I_p^r(M, N) = I^r(M, N) \cap J_p^r(M, N)$.

(3) Let $G^r(n)$ be the set of all invertible jets in $I^r(\mathbf{R}^n, \mathbf{R}^n)$ whose source and target are the origin of $\mathbf{R}^n$. Then with respect to the composition of jets, $G^r(n)$ forms a Lie group that is an †extension of $G^1(n) = GL(n, \mathbf{R})$ by a simply connected nilpotent Lie group. The projection $G^r(n) \to G^1(n)$ is a special example of the natural projection $J^r(M, N) \to J^s(M, N)$ ($r \geqslant s$), which is defined in general.

(4) We can identify $I_0^1(\mathbf{R}^m, M)$ ($m = \dim M$) with the tangent m-frame bundle over M. More generally, $I_0^r(\mathbf{R}^m, M)$ is a $G^r(m)$-bundle over M.

Y. Pseudogroup Structure

Let X be a topological space, and let Γ be a set consisting of homeomorphisms $f: U_f \to V_f$, where U_f, V_f are open subsets of X. We call Γ a **pseudogroup** of topological transformations if Γ satisfies the following four conditions: (i) Γ contains the identity mapping of X onto X; (ii) if $f \in \Gamma$, then the restriction of f onto any open subset U of U_f is also contained in Γ; (iii) if f and g are in Γ and $V_f \subset U_g$, then $g \circ f$ is contained in Γ; and (iv) if $f \in \Gamma$, then $f^{-1}: V_f \to U_f$ is also in Γ.

Following the definition of differentiable manifolds we define a **pseudogroup structure** of M (or, more precisely, a **Γ-structure** of M) as a set A of bijections, with each member α defined on a subset U_α of M onto an open set V_α of X, satisfying the following three conditions: (i) $\bigcup_\alpha U_\alpha = M$; (ii) if $\alpha, \beta \in A$, then $\alpha \circ \beta^{-1} \in \Gamma$, where the domain of definition of $\alpha \circ \beta^{-1}$ is $\beta(U_\alpha \cap U_\beta)$; and (iii) A is the maximal set of bijections that satisfies conditions (i) and (ii). We introduce in M the weakest (coarsest) topology such that every bijection α is a homeomorphism. If Γ' is a pseudogroup of X containing Γ and A' is a Γ'-structure of M such that $A \subset A'$, then we say that A' is subordinate to A. If $X = \mathbf{R}^n$ (or H^n, a half-space of $\mathbf{R}^n$), Γ is the totality of diffeomorphisms of class C^r of open sets of X onto open sets of X, and M is a space with Hausdorff topology, then the Γ-structure is the C^r-structure with or without boundary which we have already defined. We give three examples of Γ-

structures subordinate to Γ', where Γ' is the totality of local transformations of class C^r $(r \geqslant 1)$ in $\mathbf{R}^n$.

(1) When n is even, we identify $\mathbf{R}^n$ with $\mathbf{C}^{n/2}$ and denote the totality of holomorphic transformations of connected open domains by Γ. The Γ-structure in this case is called a **complex structure**.

(2) When n is odd, we define Γ as the totality of transformations of connected open domains in $\mathbf{R}^n$ that leave invariant a Pfaffian form $\sum_{i=1}^{m} x^i dx^{m+i} + dx^{2m+1}$ $(n=2m+1)$ up to scalar factors. The Γ-structure in this case is called a **contact structure**.

(3) We consider $\mathbf{R}^n = \mathbf{R}^p \times \mathbf{R}^{n-p}$ and define Γ as the set of all diffeomorphisms $U \rightarrow V$ (where U, V are open in $\mathbf{R}^n$) such that each set of the form $U \cap (\mathbf{R}^p \times \{y\})$ is mapped onto a set of the form $V \cap (\mathbf{R}^p \times \{y'\})$. The Γ-structure in this case is called a **foliated structure**.

The problem of determining whether there exists a Γ-structure for given Γ and M involves widely ranging problems of topology and analysis. The classification of Γ with reasonable conditions is another important open problem.

Haefliger has constructed the classifying space $B\Gamma$ for Γ-structures.

Z. Infinite-Dimensional Manifolds

Let B and B' be Banach spaces and φ be a mapping from an open set of B to B'. Then, using the notion of Fréchet derivatives, we can define φ to be smooth. For a smooth mapping φ, the Jacobian $J(\varphi)$ at each point x is a linear mapping from B to B', for which the inverse function theorem holds true as a straightforward extension of the corresponding one in the finite-dimensional case. Similar extension also holds for the existence and the uniqueness theorems of solution of ordinary differential equations with value in B. These facts permit us to generalize the notion of differentiable manifolds to infinite-dimensional ones. Actually, infinite-dimensional manifolds can be defined in a way similar to the finite-dimensional case, taking open sets of a certain Banach space as local coordinate neighborhoods. Such a manifold is called a **Banach manifold**. Various formal definitions of differentiable manifolds can also be stated for Banach manifolds in extended form. However, while differentiable manifolds are locally compact and admit a partition of unity by smooth functions subordinate to a locally finite open covering, Banach manifolds generally lack these properties. Actually, local compactness gives a criterion for whether a manifold is finite-dimensional or not. Banach manifolds often provide a basic functional-analytic view to nonlinear analysis and global analysis.

The most important category of Banach manifolds is provided by the **Hilbert manifold**, that is, a Banach manifold whose local coordinates are modeled on a Hilbert space. For a Hilbert manifold, a partition of unity subordinate to a locally finite open covering can be taken from smooth functions. In the following, we refer to a separable Hilbert manifold simply as a Hilbert manifold. An infinite-dimensional Hilbert manifold M can be smoothly embedded as an open set of a Hilbert space and thus covered by a single coordinate neighborhood. Hence, in particular, the tangent bundle of M is trivial. Historically, this fact was the first instance showing that the distinguishing properties are shared by infinite-dimensional manifolds; this was first recognized as a consequence of the theorem stating that the unitary group of an infinite-dimensional Hilbert space is contractible. Two infinite-dimensional Hilbert manifolds are diffeomorphic if and only if they are homotopically equivalent. A typical example of a Hilbert manifold is provided by the space of L^2-loops on a compact Riemannian manifold. Morse theory can be extended in a suitable way to a Riemannian Hilbert manifold under the Palais-Smale condition, which makes it possible for the integral curve of $\operatorname{grad} f$ to tend to a critical point, where f is a Morse function.

AA. Gel'fand-Fuks Cohomology

The space $\mathfrak{X}(M)$ consisting of all the smooth vector fields on a smooth manifold M has the structure of a Lie algebra under the bracket operation $[X, Y] = XY - YX$, where the vector fields X and Y are regarded as derivations on the algebra $C^\infty(M)$ of smooth functions on M. $\mathfrak{X}(M)$ is a topological Lie algebra when endowed with the topology defined by uniform convergence of the components of vector fields and all their partial derivatives on each compact set of M. When $\mathfrak{X}(M)$ acts continuously on a topological vector space V, the continuous cohomology $H^*(\mathfrak{X}(M), V)$ of $\mathfrak{X}(M)$ with coefficients in V is the cohomology of the cochain complex $\bigoplus \{C^p = C^p(\mathfrak{X}(M), V), d\}$. Here $C^0 = V$, $C^p (p \geqslant 1)$ is the space of all the alternating p-linear continuous mapping φ of $\mathfrak{X}(M) \times \ldots \times \mathfrak{X}(M)$ (p-times) into V, and $d: C^p \rightarrow C^{p+1}$ is defined for $\varphi \in C^p$ and $X_i \in \mathfrak{X}(M)$ by $d\varphi(X_1) = X_1 \varphi$ $(p=0)$ and $d\varphi(X_1, \ldots, X_{p+1}) = \sum_{i<j} (-1)^{i+j} \varphi([X_i, X_j]; X_1, \ldots, \check{X}_i, \ldots, \check{X}_j, \ldots, X_{p+1}) + \sum_i (-1)^{i+1} X_i \varphi(X_1, \ldots, \check{X}_i, \ldots, X_{p+1})$ $(p \geqslant 1)$. When V is a topological algebra and the elements of $\mathfrak{X}(M)$ act on V as derivations, the exterior multiplication of cochains induces

a graded algebra structure in $H^*(\mathfrak{X}(M), V)$. When $V=\mathbf{R}$ is the trivial $\mathfrak{X}(M)$-module, then $H^*(\mathfrak{X}(M))=H^*(\mathfrak{X}(M),\mathbf{R})$ is called the **Gel'fand-Fuks cohomology** of M. Gel'fand and Fuks proved that, for any compact oriented manifold M, we have $\dim H^p(\mathfrak{X}(M))<+\infty$ for all p and $H^p(\mathfrak{X}(M))=0$ for $1\leqslant p\leqslant n$ ($n=\dim M$). For example, if M is the circle S^1, then the algebra $H^*(\mathfrak{X}(S^1))$ is generated by two generators $\alpha\in H^2$, $\beta\in H^3$, which are explicitly described as cochains in the following way:

$$\alpha\left(f\frac{d}{dt},\ g\frac{d}{dt}\right)=\int_{S^1}\begin{vmatrix} f' & f'' \\ g' & g''\end{vmatrix}dt,$$

$$\beta\left(f\frac{d}{dt},\ g\frac{d}{dt},\ h\frac{d}{dt}\right)=\int_{S^1}\begin{vmatrix} f & f' & f'' \\ g & g' & g'' \\ h' & h' & h''\end{vmatrix}dt.$$

Localization of the concept of Gel'fand-Fuks cohomology naturally yields the cohomology of **formal vector fields**. Here a formal vector field means the expression $\Sigma f_\mu(x_1,\dots,x_n)\partial/\partial x_\mu$, f_μ being formal power series in $x_1,\dots,x_n$. The set of all the formal vector fields forms a Lie algebra $\mathfrak{a}_n$ and the continuous cohomology of $\mathfrak{a}_n$ with respect to the Krull topology is denoted by $H^*(\mathfrak{a}_n)$. Let B_U be the universal †classifying space of the unitary group $U(n)$, let $(B_U)_{2n}$ be its $2n$-skeleton, and let P_{2n} be the canonical principal $U(n)$-bundle restricted to $(B_U)_{2n}$. Then there is an algebra isomorphism $H^*(\mathfrak{a}_n)\cong H^*(P_{2n};\mathbf{R})$. This cohomology and its variants play important roles in the theory of foliations.

An important subcomplex $\bigoplus\{C^p_\Delta,d\}$ of $\bigoplus\{C^p,d\}$, the diagonal complex, is defined as $C^p_\Delta=\{\varphi\in C^p\mid\varphi(X_1,\dots,X_p)=0$ if $\operatorname{supp}X_1\cap\dots\cap\operatorname{supp}X_p=\varnothing\}$. Here, $\operatorname{supp}X_i$ denotes the support of X_i, that is, $\overline{\{x\mid X_i(x)\neq0\}}$. Let P_M be the principal $U(n)$-bundle associated with the complexified tangent bundle of M. $U(n)$ acts freely on the product $P_M\times P_{2n}$ and the quotient space $E_M=P_M\times P_{2n}/U(n)$ is a fiber bundle over M with fiber P_{2n}. Then, if M is a compact oriented manifold, the cohomology $H^*_\Delta(\mathfrak{X}(M))$ of the diagonal complex is completely determined by the isomorphisms $H^p_\Delta(\mathfrak{X}(M))\cong H^{p+n}(E_M;\mathbf{R})$ for all p. In particular, if all the Pontryagin classes of M vanish, then $H^p_\Delta(\mathfrak{X}(M))=\Sigma_{i+j=p+n}H^i(M;\mathbf{R})\otimes H^j(\mathfrak{a}_n)$.

The Gel'fand-Fuks cohomology has a topological interpretation: $H^*(\mathfrak{X}(M))\cong H^*(\Gamma(E_M),\mathbf{R})$ as graded algebras, where $\Gamma(E_M)$ denotes the space of all the continuous cross-sections of $E_M\to M$ with the compact open topology. Moreover the differential graded algebra $C^*(\mathfrak{X}(M))$ has a model in the sense of Sullivan constructed purely algebraically from a model of the de Rham algebra of M and the Pontryagin classes of M. This model shows in particular that $H^*(\mathfrak{X}(M))$ is not necessarily finitely generated as a graded algebra.

The cohomology theory of $\mathfrak{X}(M)$ in the case where the representation is nontrivial has also been investigated. The natural representation on $C^\infty(M)$ is a typical example. There is also a topological interpretation: $H^*(\mathfrak{X}(M), C^\infty(M))\cong H^*(Y_M,\mathbf{R})$, where Y_M is the fiber product of the evaluation mapping $M\times\Gamma(E_M)\to E_M$ and the inclusion $P_M\hookrightarrow E_M$ corresponding to a fiber inclusion $U(n)\hookrightarrow P_{2n}$.

References

[1] L. Auslander, Differential geometry, Harper, 1967.
[2] R. L. Bishop and S. I. Goldberg, Tensor analysis on manifolds, Macmillan, 1968.
[3] N. Bourbaki, Variétés différentielles et analytiques, Hermann, 1967.
[4] E. Cartan, Les systèmes différentiels extérieurs et leurs applications géométriques, Actualités Sci. Ind., Hermann, 1945.
[5] H. Cartan, Formes différentielles, Hermann, 1967.
[6] C. Chevalley, Theory of Lie groups I, Princeton Univ. Press, 1946.
[7] G. de Rham, Variétés différentiables, Actualités Sci. Ind., Hermann, 1955.
[8] S. Kobayashi and K. Nomizu, Foundations of differential geometry, Interscience, I, 1963; II, 1969.
[9] S. Lang, Introduction to differentiable manifolds, Interscience, 1962.
[10] Y. Matsushima, Differentiable manifolds, Dekker, 1972.
[11] J. R. Munkres, Elementary differential topology, Ann. Math. Studies 54, Princeton Univ. Press, 1963.
[12] K. Nomizu, Lie groups and differential geometry, Publ. Math. Soc. Japan, 1956.
[13] S. Sternberg, Lectures on differential geometry, Prentice-Hall, 1964.
[14] H. Whitney, Differentiable manifolds, Ann. Math., (2) 37 (1936), 645–680.
[15] H. Whitney, Geometric integration theory, Princeton Univ. Press, 1957.
[16] J. Eells, A setting for global analysis, Bull. Amer. Math. Soc., 72 (1966), 751–807.
[17] D. Burghelea and N. Kuiper, Hilbert manifolds, Ann. Math., (2) 90 (1969), 335–352.
[18] N. Kuiper, The homotopy type of the unitary group of Hilbert space, Topology, 3 (1965), 19–30.
[19] R. Palais and S. Smale, A generalized Morse theory, Bull. Amer. Math. Soc., 70 (1964), 156–172.
[20] I. M. Gel'fand, The cohomology of infinite-dimensional Lie algebras; some ques-

tions of integral geometry, Actes Congr. Int. Math., Nice, 1970, Gauthier-Villars, vol. 1, 95–111.
[21] I. M. Gel'fand and D. B. Fuks, Cohomologies of the Lie algebra of tangent vector fields of a smooth manifold I, II, Functional Anal. Appl., 3 (1969), 194–210; 4 (1970), 110–116. (Original in Russian, 1969, 1970.)
[22] I. M. Gel'fand and D. B. Fuks, The cohomology of the Lie algebra of formal vector fields, Math. USSR-Izv. (1970), 322–337. (Original in Russian, 1970.)
[23] V. W. Guillemin, Cohomology of vector fields on a manifold, Advances in Math., 10 (1973), 192–220.
[24] M. V. Losik, On the cohomologies of infinite-dimensional Lie algebras of vector fields, Functional Anal. Appl., 4 (1970), 127–135. (Original in Russian, 1970.)
[25] R. Bott and G. Segal, The cohomology of the vector fields on a manifold, Topology, 16 (1977), 285–298.
[26] A. Haefliger, Sur la cohomologie de l'algèbre de Lie des champs de vecteurs, Ann. Sci. Ecole Norm. Sup., 9 (1976), 503–532.

106 (X.6) Differential Calculus

A. First-Order Derivatives

Let $y=f(x)$ be a real-valued function of x defined on an interval I of the real line $\mathbf{R}$. If for a fixed $x_0 \in I$, the limit

$$\lim_{\substack{h\to 0\\ x_0+h\in I}} \frac{f(x_0+h)-f(x_0)}{h}$$

exists and is finite, then f is called **differentiable at the point** x_0, and the limit is the **derivative** (**differential coefficient** or **differential quotient**) of f at the point x_0. If f is differentiable at every point of a set $A \subset I$, then f is said to be **differentiable on** A. The function that assigns the derivative of f at x to $x \in A$ is called the **derivative** (or **derived function**) of $f(x)$, which is denoted by dy/dx, y', $\dot{y}$, $df(x)/dx$, $(d/dx)f(x)$, $f'(x)$, or $D_x f(x)$. The process of determining f' is known as the **differentiation** of f. The derivative of f at the point x_0 is written $f'(x_0)$, $(df/dx)(x_0)$, $D_x f(x_0)$, $[dy/dx]_{x=x_0}$, etc. We say that f is **right (left) differentiable** or **differentiable on the right (left)** if the limit on the right, $\lim_{h\to+0}(f(x_0+h)-f(x_0))/h$ (the limit on the left, $\lim_{h\to+0}(f(x_0-h)-f(x_0))/h$), exists and is finite. This limit is called the **right (left) derivative** or **derivative on the right (left)** and is denoted by $D_x^+ f(x_0)$ or $f'_+(x_0)$ $(D_x^- f(x_0)$ or $f'_-(x_0))$. For instance, if f is defined on $I=[a,b)$, then $D_x f(a)$ is identical to $D_x^+ f(a)$.

B. Differentials

In the definitions given above, neither dx nor dy in dy/dx has a meaning by itself. In the following, however, we give a definition of dx and dy, using the concept of increment, so that $dy=f'(x)dx$. Let Δy denote the **increment** $f(x+\Delta x)-f(x)$ of f corresponding to the increment Δx of x. Suppose that $f(x)$ is differentiable at x. We set $\Delta y/\Delta x=f'(x)+\varepsilon$. Then we have $\lim_{\Delta x\to 0}\varepsilon=0$. This can be written utilizing †Landau's notation as $\Delta y=f'(x)\Delta x+o(|\Delta x|)$ $(\Delta x\to 0)$; in other words, Δy is the sum of two terms, of which the first, $f'(x)\Delta x$, is proportional to Δx and the second is an †infinitesimal of an order higher than Δx. Here the principal part $f'(x)\Delta x$ of Δy is called the **differential** of $y=f(x)$ and is denoted by dy. The differential dy thus defined is a function of two independent variables x and Δx. In particular, if $f(x)=x$, from the definition we get $dx=1\cdot\Delta x=\Delta x$. Hence, in general, we have $dy=f'(x)dx$ and $f'(x)=dy/dx$.

With respect to the rectangular coordinates (x,y), the straight line with slope $f'(x_0)$ through a point $(x_0,f(x_0))$ on the graph of $y=f(x)$ is the †tangent line of the graph at the point $(x_0,f(x_0))$. A function is continuous at a point where the function is differentiable, but the converse of this proposition does not hold. In fact, Weierstrass showed that the function defined by the infinite series $\sum_{n=0}^{\infty}a^n\cos b^n\pi x$, where $0<a<1$ and b is an odd integer with $ab>(3/2)\pi+1$, is continuous everywhere and nowhere differentiable on $(-\infty,\infty)$ [3].

C. Differentiation

For two differentiable functions f and g defined on the interval I, the following formulas hold: $(\alpha f+\beta g)'=\alpha f'+\beta g'$, where α and β are constants; $(fg)'=f'g+fg'$; and $(f/g)'=(f'g-fg')/g^2$ (at every point where $g\neq 0$). Let $y=f(x)$ be a function of x defined on the interval (a,b) and $x=\varphi(t)$ a function of t defined on (α,β). If $\varphi(t)\in(a,b)$ whenever $t\in(\alpha,\beta)$, then the composite function $y=F(t)=f(\varphi(t))=(f\circ\varphi)(t)$ is well defined. Assume further that f and φ are differentiable on (a,b) and (α,β), respectively. Then the composite function $F(t)=(f\circ\varphi)(t)$ is differentiable on (α,β), and we have the **chain rule**, $F'(t)=f'(x)\varphi'(t)$ $(x=\varphi(t))$, or $dy/dt=(dy/dx)(dx/dt)$. Assume that a function $y=f(x)$ is †strictly increasing or decreasing and differentiable at x_0. If furthermore we have

$f'(x) \neq 0$, then the inverse function $x = f^{-1}(y)$ is also differentiable at y_0 $(=f(x_0))$ and satisfies $(dx/dy)_{y=y_0}(dy/dx)_{x=x_0} = 1$. However, if $f'(x_0) = 0$, then even though $f^{-1}(y)$ is not differentiable at y_0, $\lim_{\Delta y \to 0}(f^{-1}(y_0+\Delta y) - f^{-1}(y_0))/\Delta y$ exists and is $+\infty$ or $-\infty$.

D. Higher-Order Derivatives

If the derivative $f'(x)$ of a function $y = f(x)$ is again differentiable on I, then $(f'(x))' = f''(x)$ is well defined as a function of x on I. In general, if $f^{(n-1)}(x)$ is differentiable on I, then $f(x)$ is called **n-times differentiable** on I, and the **nth derivative** (or **nth derived function**) $f^{(n)}(x)$ of $f(x)$ is defined by $f^{(n)}(x) = (f^{(n-1)}(x))'$ and is also denoted by $d^n y/dx^n$ or $D^{(n)}y$. The nth derivative for $n \geqslant 2$ is called a **higher-order derivative**.

Concerning the nth derivative of the product of two functions, **Leibniz's formula** holds:

$$(fg)^{(n)} = f^{(n)}g + \binom{n}{1} f^{(n-1)}g' + \ldots$$

$$+ \binom{n}{k} f^{(n-k)}g^{(k)} + \ldots + fg^{(n)}.$$

Analogous to $dy = y'\Delta x$, which is a function of x and Δx, we can define d^2y in the notation d^2y/dx^2 by $d^2y = d(dy) = d(y'\Delta x) = (y'\Delta x)'\Delta x = y''\Delta x^2$. Since $\Delta x = dx$, it follows from the above that $d^2y = y''\,dx^2$. Similarly, $d^n y = y^{(n)}dx^n$ and is called the **nth differential** (or **differential of nth order**) of $f(x)$.

E. The Mean Value Theorem

Let $f(x)$ be a continuous function defined on $[a,b]$, and suppose that for every point x_0 on (a,b) there exists a limit $\lim_{h\to 0}(f(x_0+h) - f(x_0))/h$, which may be infinite. (These conditions are satisfied if $f(x)$ is differentiable on $[a,b]$.) Then there exists a point ξ such that

$$\frac{f(b)-f(a)}{b-a} = f'(\xi), \qquad a < \xi < b.$$

This proposition is called the **mean value theorem**. A special case of the theorem under the further condition that $f(a) = f(b)$ is called **Rolle's theorem**. If we put $b - a = h$, $\xi = a + \theta h$, then the conclusion of the theorem may be written as $f(a+h) = f(a) + h \cdot f'(a+\theta h)$ $(0 < \theta < 1)$.

This theorem implies the following: Let $f(x)$ be a function as in the hypothesis of the mean value theorem, and assume further that $A \leqslant f'(x) \leqslant B$ holds for all x with $a < x < b$. Then $A \leqslant (f(b)-f(a))/(b-a) \leqslant B$. (French mathematicians sometimes call this the "théorème des accroissements finis.") Using the mean value theorem, the following theorem can be proved: If $f(x)$ is continuous on $[a,b]$ and $f'(x)$ exists and is positive on (a,b), then $f(b) > f(a)$. Accordingly, if $f'(x) > 0$ at every point x of an interval I, then $f(x)$ is †strictly increasing on that interval. (If $f'(x) < 0$ on I, then f is strictly decreasing.) The converse of the previous statement does not always hold ($f'(x) = x^3$ is a counterexample, since $f'(0) = 0$). Furthermore, from the mean value theorem it follows that if $f'(x) = 0$ everywhere in an interval, then $f(x)$ is constant on that interval. Consequently, two functions with the same derivative on an interval differ only by a constant.

Suppose that $f(x)$ is n-times differentiable on an open interval I. For a fixed $a \in I$ and an arbitrary $x \in I$, we put

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \ldots$$

$$+ \frac{f^{(n-1)}(a)}{(n-1)!}(x-a)^{n-1} + R_n.$$

Then $R_n = f^{(n)}(\xi)(x-a)^n/n!$ for some ξ between a and x. This is called **Taylor's formula**, where R_n is the **remainder** of the nth order given by Lagrange. We also have several other forms for R_n (→ Appendix A, Table 9). If $f^{(n)}(x)$ is continuous at $x = a$, then $\xi \to a$ as $x \to a$, and accordingly, $f^{(n)}(\xi) \to f^{(n)}(a)$. Hence $f(x) = \sum_{k=0}^{n}(f^{(k)}(a)/k!)(x-a)^k + o((x-a)^n)$. If $f^{(n)}(x)$ is continuous at $x = a$, then, by Taylor's formula, the value of the polynomial $\sum_{k=0}^{n}(f^{(k)}(a)/k!)(x-a)^k$ can be considered an approximate value of $f(x)$ for x near a. This approximation is called the **nth approximation** of $f(x)$, and its error is given by $|R_{n+1}|$. By applying this formula, it is sometimes possible to calculate a limit such as $A = \lim_{x\to a} f(x)/g(x)$, where $f(x) \to 0$ and $g(x) \to 0$ as $x \to a$. For instance, if $f'(x)$ and $g'(x)$ are both continuous at $x = a$ and $g'(a) \neq 0$, then by taking the first approximations of $f(x)$ and $g(x)$ it is easily seen that $A = f'(a)/g'(a)$. A limit of this type is often called a **limit of an indeterminate form** $0/0$. Similarly, we can calculate limits of such indeterminate forms as $0 \cdot \infty$ or 0^∞ (for limits of indeterminate forms → [6]).

F. Partial Derivatives

Let $w = f(x, y, \ldots, z)$ be a real-valued function of n independent real variables $x, y, \ldots, z$ defined on a domain G contained in n-dimensional Euclidean space $\mathbf{R}^n$. We obtain a function of a single variable from f by keeping $n-1$ variables (say, $(x, \hat{y}, \ldots, z)$, i.e., all except y) fixed. If

such a function $\varphi(y)=f(x_0, y, \ldots, z_0)$ is differentiable, that is, if

$$\varphi'(y_0)=\lim_{\Delta y\to 0}\frac{\varphi(y_0+\Delta y)-\varphi(y_0)}{\Delta y}$$

$$=\lim_{\Delta y\to 0}\frac{f(x_0, y_0+\Delta y, \ldots, z_0)-f(x_0, y_0, \ldots, z_0)}{\Delta y}$$

exists and is finite, then f is called **partially differentiable** with respect to y at $(x_0, y_0, \ldots, z_0)$, and the derivative is called the **partial derivative** (or **partial differential coefficient**) of $f(x, y, \ldots, z)$ with respect to y at $(x_0, y_0, \ldots, z_0)$. It is denoted by $[\partial w/\partial y]_{x=x_0,\ldots,z=z_0}$, $(\partial/\partial y)f(x_0, y_0, \ldots, z_0)$, $f_y(x_0, y_0, \ldots, z_0)$, or $D_y f(x_0, y_0, \ldots, z_0)$, etc. We usually assume that the point $(x, y, \ldots, z)$ where partial derivatives are considered is an †interior point of the †domain of the function. Since in a space of dimension higher than 1, the †boundary of a domain may be complicated, partial derivatives at boundary points are usually not considered. If a function f possesses a partial derivative with respect to x at every point of an open set G, then f_x is a function on G and is called a **partial derivative** of f with respect to x. The process of determining partial derivatives of f is called the **partial differentiation** of f.

G. Total Differential

Let $w=f(x, y, \ldots, z)$ be a function defined on a domain G, and let $P=(x, y, \ldots, z)$ be an interior point of the domain G of a function $w=f(x, y, \ldots, z)$. Put $\Delta w=f(x_0+\Delta x, y_0+\Delta y, \ldots, z_0+\Delta z)-f(x_0, y_0, \ldots, z_0)$. If there exist constants $\alpha, \beta, \ldots, \gamma$ such that $\Delta w=\alpha\Delta x+\beta\Delta y+\ldots+\gamma\Delta z+o(\rho)$ $(\rho\to 0)$, where $\rho=\sqrt{\Delta x^2+\Delta y^2+\ldots+\Delta z^2}$, then f is called **totally differentiable** (or **differentiable in the sense of Stolz**) at P. In this case, f is partially differentiable at P with respect to each of the variables $x, y, \ldots, z$, and $\alpha=f_x(x_0, y_0, \ldots, z_0)$, $\beta=f_y(x_0, y_0, \ldots, z_0), \ldots, \gamma=f_z(x_0, y_0, \ldots, z_0)$. The principal part of Δw as $\rho\to 0$ is $\alpha\Delta x+\beta\Delta y+\ldots+\gamma\Delta z$, which is called the **total differential** of w at P. If f is totally differentiable at every point of G, then f is said to be totally differentiable on G. The total differential of w is denoted by dw. Since the total differentials of $x, y, \ldots, z$ are $dx=\Delta x$, $dy=\Delta y, \ldots, dz=\Delta z$, respectively, we can write $dw=f_x dx+f_y dy+\ldots+f_z dz$; and dw is a function of independent variables $x, y, \ldots, z, dx, dy, \ldots, dz$. The total differentiability of f implies the continuity of f, whereas the partial differentiability of f with respect to each variable does not imply that f is continuous. (Example: Define $f(x, y)=xy/(x^2+y^2)$ for $(x, y)\neq(0,0)$, and $f(0,0)=0$; then the function f is not continuous at $(0,0)$, even though both f_x and f_y exist at $(0,0)$.) The function f is totally differentiable on G if all $f_x, f_y, \ldots, f_z$ exist and are continuous on G, or, more weakly, if all $f_x, f_y, \ldots, f_z$ exist and, with possibly one exception, are continuous. Suppose that $w=f(x, y)$ is totally differentiable at (x, y), and let $\Delta x=\rho\cos\theta$, $\Delta y=\rho\sin\theta$. As $\rho\to 0$, for a fixed θ there exists the limit $\lim_{\rho\to 0}(\Delta w/\rho)=f_x(x, y)\cos\theta+f_y(x, y)\sin\theta$. This limit is called the **directional derivative** in the direction θ at (x, y). The partial derivatives f_x and f_y are special cases of the directional derivative for $\theta=0$ and $\pi/2$, respectively. Suppose that we are given a curve lying in the †interior of the domain of f and that the curve passes through the point (x, y), where the curve is differentiable. Then the partial derivative of $w=f(x, y)$ in the direction of the normal to the curve at (x, y) is called the **normal derivative** of w at the point (x, y) on the curve and is denoted by $\partial w/\partial n$. Analogous definitions and notations have been introduced for functions of more than two variables.

To see the geometric significance of the total differentiability of $w=f(x, y)$, we consider the graph of the function $w=f(x, y)$ and a point $(a, b, f(a, b))$ on the graph. Then the plane represented by $w-f(a, b)=\alpha(x-a)+\beta(y-b)$ is the †tangent plane to the surface at $(a, b, f(a, b))$ if and only if $\alpha=f_x(a, b)$ and $\beta=f_y(a, b)$. The existence of f_x and f_y depends on the choice of coordinate axes, while the total differentiability of f does not.

H. Higher-Order Partial Derivatives

Suppose that a partial derivative of a function $w=f(x, y, \ldots, z)$ defined on an open set G again admits partial differentiation. The latter partial derivative is called a second-order partial derivative of f. We can similarly define the ***n*th order partial derivatives**. Higher-order partial derivatives are denoted as follows:

$$\frac{\partial}{\partial x}\left(\frac{\partial w}{\partial x}\right)=\frac{\partial^2 w}{\partial x^2}=f_{xx}(x, y, \ldots, z),$$

$$\frac{\partial}{\partial y}\left(\frac{\partial w}{\partial x}\right)=\frac{\partial^2 w}{\partial x\partial y}=f_{xy}(x, y, \ldots, z),$$

$$\frac{\partial}{\partial x}\left(\frac{\partial^2 w}{\partial x\partial y}\right)=\frac{\partial^3 w}{\partial x\partial y\partial x}=f_{xyx}(x, y, \ldots, z), \ldots.$$

In general, f_{xy} and f_{yx} are not equal. (Peano's example: Let $f(x, y)=xy(x^2-y^2)/(x^2+y^2)$ for $(x, y)\neq(0,0)$ and $f(0,0)=0$. Then $f_{xy}=-1$, $f_{yx}=1$ at $(0,0)$.) However, if both f_{xy} and f_{yx} are continuous on an open set G', then they coincide in G'. Furthermore, if f_x, f_y, and f_{xy} exist in a neighborhood U of a point P belonging to the domain of f and f_{xy} is continuous

at P, then f_{yx} exists at P and $f_{xy}=f_{yx}$ (H. A. Schwarz). If f_x and f_y exist in U and are totally differentiable at P, then $f_{xy}=f_{yx}$ at P (W. H. Young). Similarly, if the partial derivatives of order $\geqslant 3$ $f_{\dots xy\dots}$ and $f_{\dots yx\dots}$ are all continuous, then $f_{\dots xy\dots}=f_{\dots yx\dots}$. Hence we can change the order of differentiation if all the derivatives concerned are continuous.

I. Composite Functions of Several Variables

Let w be a function of $x, y, \dots, z$, and let each $x, y, \dots, z$ be a function of t. Suppose that the range of $(x(t), y(t), \dots, z(t))$ is contained in the domain of w. Then w is a function of t. If further w is totally differentiable and $x, y, \dots, z$ are all differentiable, then w as a function of t is differentiable, and we have

$$\frac{dw}{dt}=\frac{\partial w}{\partial x}\frac{dx}{dt}+\frac{\partial w}{\partial y}\frac{dy}{dt}+\dots+\frac{\partial w}{\partial z}\frac{dz}{dt}.$$

If partial derivatives of order $\geqslant 2$ are totally differentiable, then d^2w/dt^2, $d^3w/dt^3, \dots$ are obtained by repeating the above procedure. A similar consideration is valid when $x, y, \dots, z$ are functions of several variables.

J. Taylor's Formula for Functions of Several Variables

Suppose that $f(x,y)$ is defined on an open domain G, $f(x,y)$ has continuous partial derivatives of orders up to n, and the line segment $(a+(x-a)t, b+(y-b)t)$ $(0\leqslant t\leqslant 1)$ is contained in the domain G. Then there exists a number θ $(0<\theta<1)$ such that

$$\begin{aligned}
&f(x,y)\\
&=f(a,b)+\left((x-a)\frac{\partial}{\partial x}+(y-b)\frac{\partial}{\partial y}\right)f(a,b)\\
&\quad+\frac{1}{2!}\left((x-a)\frac{\partial}{\partial x}+(y-b)\frac{\partial}{\partial y}\right)^2 f(a,b)+\dots\\
&\quad+\frac{1}{(n-1)!}\left((x-a)\frac{\partial}{\partial x}+(y-b)\frac{\partial}{\partial y}\right)^{n-1} f(a,b)\\
&\quad+\frac{1}{n!}\left((x-a)\frac{\partial}{\partial x}+(y-b)\frac{\partial}{\partial y}\right)^{n}\\
&\qquad\times f(a+(x-a)\theta, b+(y-b)\theta),
\end{aligned}$$

where, for instance, the third term $((x-a)\cdot(\partial/\partial x)+(y-b)(\partial/\partial y))^2 f(a,b)$ means $(x-a)^2(\partial^2 f/\partial x^2)(a,b)+2(x-a)(y-b)\cdot(\partial^2 f/\partial x\partial y)(a,b)+(y-b)^2(\partial^2 f/\partial y^2)(a,b)$, with $(\partial^2 f/\partial x^2)(a,b)$, $(\partial^2 f/\partial x\partial y)(a,b)$, and $(\partial^2 f/\partial y^2)(a,b)$ denoting the values of $(\partial^2 f/\partial x^2)$, $(\partial^2 f/\partial x\partial y)$, and $(\partial^2 f/\partial y^2)$ at (a,b), respectively. The displayed formula is called **Taylor's formula** for a function of two variables. A similar formula is valid for a function of n variables $(n\geqslant 3)$. As in the case of functions of one variable, we can derive approximation formulas for f from Taylor's formula.

K. Classes of Functions

If all the partial derivatives of order n of $f(P)$ are continuous on an open set G, then f is said to be a **function of class** C^n (or **n-times continuously differentiable**) on G. The set of all n-times continuously differentiable functions is denoted by C^n $(n=1,2,\dots)$. A continuous function is of **class** C^0. A **function of class** C^1 is also called a **smooth** function. It is obvious that $C^0\supset C^1\supset C^2\supset\dots$. A partial derivative of order $r\leqslant s$ of a function belonging to class C^s does not depend on the order of the differentiation. A function belonging to $C^\infty=\bigcap_{r=1}^{\infty}C^r$ is said to be of **class** C^∞ or **infinitely differentiable**. We sometimes say that a function has a certain "nice" property or is "well behaved" if it belongs to some C^r $(r\geqslant 1)$.

Let $w=f(x,y,\dots,z)$ be a function defined on an open set G in $\mathbf{R}^n$ and $P=(a,b,\dots,c)\in G$. If

$$f(x,y,\dots,z)=f(a,b,\dots,c)+\sum_{r_1=1}^{\infty}\dots\sum_{r_n=1}^{\infty}\alpha_{r_1 r_2\dots r_n}(x-a)^{r_1}(y-b)^{r_2}\dots(z-c)^{r_n}$$

holds in some open neighborhood U of P, where the right-hand side of the equality is an absolutely convergent series, then f is said to be **real analytic** at P. In this case, f is r-times differentiable at P for any r, and we have

$$\alpha_{r_1 r_2\dots r_n}=\frac{r_1!r_2!\dots r_n!}{(r_1+r_2+\dots+r_n)!}\times\frac{\partial^{r_1+r_2+\dots+r_n}f}{\partial x^{r_1}\partial y^{r_2}\dots\partial z^{r_n}}(a,b,\dots,c).$$

If f is real analytic at every point P of the domain G, then f is called a **real analytic function** on G. Sometimes, a real analytic function is called a function of **class** C^ω. A real analytic function belongs to C^∞, but the converse is not true (→ 58 C^∞-Functions and Quasi-Analytic Functions E).

L. Extrema

Let f be a real-valued function defined on a domain G in an n-dimensional Euclidean space $\mathbf{R}^n$ that has the point P_0 in its interor. If there exists a neighborhood U of P_0 such that for every point P $(\neq P_0)$ of U we have $f(P)\geqslant f(P_0)$, then we say that f has a **relative minimum** at P_0, and $f(P_0)$ is a **relative minimum** of f. Replacing $\geqslant$ by $\leqslant$, we obtain the definition of a **relative maximum**. $f(P_0)$ is

called a **relative extremum** if it is either a relative maximum or a relative minimum.

To find relative extrema, the following facts concerning the sign of the derivative are useful. Suppose that a function of a single variable is differentiable on an interval I. Then we have the following: (1) If f has a relative extremum at an interior point x_0 of I, then $f'(x_0)=0$. (2) If $f'(x_0)=0$ and $f'(x)$ changes its sign at x_0 from positive (negative) to negative (positive), then f has a relative maximum (minimum) at x_0. (3) If $f'(x_0)=0$ and f is twice differentiable on some neighborhood of x_0, then f has a relative maximum or minimum according as $f''(x_0)<0$ or >0. If $f''(x_0)=0$, then nothing definite can be concluded about a relative extremum of f at x_0. In general, if there exists a neighborhood of x_0 in which f is r-times differentiable (r is even) and $f^{(r)}$ is continuous, and if $f'(x_0)=f''(x_0)=\ldots=f^{(r-1)}(x_0)=0$, $f^{(r)}(x_0)>0$ (or <0), then f has a relative minimum (maximum) at x_0. On the other hand, if this condition holds with odd r, then $f(x_0)$ is not a relative extremum. If $f'(x_0)=0$, then $f(x_0)$ is called a **stationary value** of f.

If a function f on n variables $x, y, \ldots, z$ has a relative extremum at $(x_0, y_0, \ldots, z_0)$, then we have $f_x(x_0, y_0, \ldots, z_0)=0$, $f_y(x_0, y_0, \ldots, z_0)=0, \ldots, f_z(x_0, y_0, \ldots, z_0)=0$, provided that the partial derivatives of f exist. Assume that for a function f of class C^2 of two variables x and y, we have $f_x(x_0, y_0)=0$ and $f_y(x_0, y_0)=0$, and let $\delta=f_{xx}(x_0, y_0)f_{yy}(x_0, y_0)-f_{xy}^2(x_0, y_0)$. Then we have the following: (1) If $\delta>0$, then according as $f_{xx}(x_0, y_0)<0$ or >0, f has a relative maximum or minimum at (x_0, y_0). (2) If $\delta<0$, then f does not have a relative extremum at (x_0, y_0). (3) If $\delta=0$, then without further information nothing definite can be said about a relative extremum of f at the point.

Let $x_1, \ldots, x_n$ be independent variables. If a function f of variables $x_1, \ldots, x_n$ has a relative extremum at a point $P_0=(x_1^0, \ldots, x_n^0)$, then $f_i=f_{x_i}(P_0)=0$ $(i=1, \ldots, n)$, provided that all the partial derivatives of f exist. In general, a point P_0 where f is totally differentiable and these conditions are satisfied is called a **critical point** of f. The value $f(P_0)$ at a critical point is called a **stationary value**. If further f is of class C^2, then consider a †quadratic form of n variables $Q=Q(X_1, \ldots, X_n)=\sum_{i,k} f_{ik}X_iX_k$, where $f_{ik}=f_{x_ix_k}(P_0)$. Suppose that $|f_{ik}|\neq 0$. Then according to whether Q is †positive definite, †negative definite, or †indefinite, f has a relative minimum, relative maximum, or no relative extremum at P_0. If $|f_{ik}|=0$, then nothing can be said in general. A critical point P of f is said to be **nondegenerate** if $|f_{ik}|\neq 0$ and **degenerate** if $|f_{ik}|=0$.

We can also apply the method of differentiation of †implicit functions to find relative extrema of functions defined implicitly. Given functions $\varphi_1, \ldots, \varphi_m$ $(m<n)$, the problem of finding a relative extremum of $f(x_1, \ldots, x_n)$ under the condition that $\varphi_1(x_1, \ldots, x_n)=0, \ldots, \varphi_m(x_1, \ldots, x_n)=0$ is called the problem of finding a **conditional relative extremum**. This problem can be reduced to the problem of finding a relative extremum of an implicit function. Actually, if the functions $f, \varphi_1, \ldots, \varphi_m$ are of class C^1 and the †Jacobian $\partial(\varphi_1, \ldots, \varphi_m)/\partial(x_{n-m+1}, \ldots, x_n)$ does not vanish in the domain considered, then $y_1=x_{n-m+1}, \ldots, y_m=x_n$ can be regarded as implicit functions of $x_1, \ldots, x_l$ $(l=n-m)$. Hence we can set $f(x_1, \ldots, x_l, y_1, \ldots, y_m)=f^*(x_1, \ldots, x_l)$. Then f has a relative extremum at $(x_1^0, \ldots, x_n^0)$ under the condition $\varphi_1=\ldots=\varphi_m=0$ if and only if f^* has a relative extremum at $P_0=(x_1^0, \ldots, x_l^0)$. The latter condition implies that all $\partial f^*/\partial x_j$ $(j=1, \ldots, l)$ vanish at P_0, which holds if and only if for arbitrary constants $\lambda_1, \ldots, \lambda_m$ the function $F(x_1, \ldots, x_n)=f+\lambda_1\varphi_1+\ldots+\lambda_m\varphi_m$ satisfies $\partial F/\partial x_i=0$ $(i=1, \ldots, n)$, and further $\varphi_1=0, \ldots, \varphi_m=0$ at $(x_1^0, \ldots, x_n^0)$. From this system of equations we can often find the values of $x_1^0, \ldots, x_n^0$. This method of finding conditional relative extrema is called **Lagrange's method of indeterminate coefficients** or the **method of Lagrange multipliers** (→ 208 Implicit Functions; 216 Integral Calculus H; 379 Series H).

References

[1] T. M. Apostol, Mathematical analysis, Addison-Wesley, 1957.
[2] N. Bourbaki, Eléments de mathématique, Fonctions d'une variable réelle, Actualités Sci. Ind., 1074b, 1132a, Hermann, second edition, 1958, 1961.
[3] R. C. Buck, Advanced calculus, McGraw-Hill, second edition, 1965.
[4] R. Courant, Differential and integral calculus I, II, Nordemann, 1938.
[5] G. H. Hardy, A course of pure mathematics, Cambridge Univ. Press, seventh edition, 1938.
[6] E. Hille, Analysis, Blaisdell, I, 1964; II, 1966.
[7] W. Kaplan, Advanced calculus, Addison-Wesley, 1952.
[8] E. G. H. Landau, Einführung in die Differentialrechnung und Integralrechnung, Noordhoff, 1934; English translation, Differential and integral calculus, Chelsea, 1965.
[9] J. M. H. Olmsted, Advanced calculus, Appleton-Century-Crofts, 1961.
[10] A. Ostrowski, Vorlesungen über

Differential-und Integralrechnung I, II, III, Birkhäuser, second edition, 1960–1961.
[11] M. H. Protter and C. B. Morrey, Modern mathematical analysis, Addison-Wesley, 1964.
[12] W. Rudin, Principles of mathematical analysis, McGraw-Hill, second edition, 1964.
[13] V. I. Smirnov, A course of higher mathematics. I, Elementary calculus; II, Advanced calculus, Addison-Wesley, 1964.
[14] A. E. Taylor, Advanced calculus, Ginn, 1955.

107 (XIII.1) Differential Equations

A. Ordinary Differential Equations

It was Galileo who found that the acceleration of a falling body is a constant and thence derived his law of a falling body $x(t)=gt^2/2$ as what we would now view as a solution of the differential equation $x''(t)=g$, where $x(t)$ denotes the distance the body has fallen during the time interval t and g is the constant gravitational acceleration. This pioneering work may be regarded as the first example of solution of a differential equation. Also, the †equations of motion, proposed by I. †Newton as the mathematical formulation of the law of motion, including Galileo's law as a special case, are differential equations of the second order. Thus differential equations appeared, simultaneously with differential and integral calculus, as an indispensable tool for the unified and concise expression of the laws of nature. Such laws are generally called **differential laws**.

Newton completely solved the equations of the †two-body problem proposed by himself; G. W. †Leibniz also succeeded in solving many simple differential equations.

In the 18th century, many mathematicians, such as the †Bernoullis, A. C. Clairaut, J. F. Riccati, L. †Euler, and J. L. †Lagrange, attacked and solved differential equations of various types independently. In that period, the emphasis was on solution by **quadrature**, that is, applying to †elementary functions a finite number of algebraic operations, transformations of variables, and indefinite integrations. It was toward the end of the 18th century that new methods, such as integration by infinite series, came to be discussed. A method of variation of constants for the solution of linear ordinary differential equations was invented by Lagrange in 1775. At the beginning of the 19th century, C. F. †Gauss initiated the study of differential equations satisfied by †hypergeometric series.

The problem of existence of solutions, which supplies a foundation of modern differential equation theory, was first treated by A. L. †Cauchy. His proof of the existence theorem was later improved by R. L. Lipschitz (1869).

Pioneers in the function-theoretic treatment of differential equations were C. A. A. Briot and J. C. Bouquet, who investigated the singular points of a function defined by an analytic differential equation. Also, B. †Riemann proposed a new viewpoint which influenced L. Fuchs in his development of the theory of linear ordinary differential equations in the complex domain (1865). Works of A. M. Legendre on †elliptic functions and of H. †Poincaré on †automorphic functions should also be mentioned in this connection.

After the Cauchy-Lipschitz existence theorem for the equation $y'=f(x,y)$ was known, efforts were directed toward weakening the conditions imposed on $f(x,y)$. G. Peano first succeeded in giving a proof under the continuity assumption only (1890), and his results were sharpened by O. Perron (1915).

Regarding the uniqueness of solutions of †initial value problems, there are various results by W. F. Osgood (1898), Perron (1925), and many Japanese mathematicians. In the course of this work the necessary and sufficient condition for uniqueness was successfully formulated in a concise form (→ 316 Ordinary Differential Equations (Initial Value Problems)).

For linear differential equations with periodic coefficients, investigations were carried out by C. Hermite (1877), E. Picard (1881), G. Floquet (1883), G. W. Hill (1886), and others. For instance, solutions satisfying $y(x+\omega)=\lambda y(x)$ were found to exist, where ω is the period of the coefficients. Analogous results followed in the case of doubly periodic coefficients.

Techniques of factorization of linear differential equations developed by G. Frobenius (1873) and E. Landau (1920) should also be noted. Picard (1883), J. Drach (1898), and E. Vessiot (1903, 1904) established a remarkable result on the solvability (in the sense of solution by quadrature) of linear differential equations, successfully extending the †Galois theory in this new direction.

The concept of †asymptotic series, which in a sense approximate the solution of differential equations, was introduced by Poincaré (1886) and extended by M. A. Lyapunov (1892), J. C. C. Kneser (1896), J. Horn (1897), C. E. Love (1914), and others. Poincaré was also the founder of topological methods in differential

equation theory, and his ideas were developed extensively by I. Bendixson (1900), Perron (1922, 1923), G. D. Birkhoff, and others (→ 126 Dynamical Systems).

In 1890, Picard invented an ingenious technique of †successive approximation for the proof of existence theorems, and his technique is now widely used in every application of functional equations. The technique of reducing linear differential equations to linear †integral equations of Volterra type was also developed.

On the †boundary value problems and †eigenvalue problems that appear in many areas of physics, there was extensive research by mathematicians such as J. C. F. Sturm (1836), J. Liouville, L. Tonelli, Picard, M. Bôcher (1898, 1921), Birkhoff (1901, 1911), and others. In this connection the problem arises of expanding a given function by an †orthogonal system of functions obtained as †eigenfunctions of a given boundary value problem. Those problems were brought into unified form by D. †Hilbert (1904) in his theory of †integral equations. Subsequently boundary value problems of ordinary and partial differential equations came to be discussed in this framework.

Finally, it should be mentioned that the †calculus of variations created by Euler and Lagrange gave rise to the study of a certain class of differential equations bearing the name of Euler (→ 46 Calculus of Variations).

B. Partial Differential Equations

The origin of partial differential equations can be traced back to the study of hydrodynamic problems by J. d'Alembert (1744) and Euler. However, perhaps Lagrange and P. S. †Laplace were the first to investigate the general theory. Subsequently, during the 18th and 19th centuries, it was developed by G. Monge, A.-M. Ampère, J. F. Pfaff, C. G. †Jacobi, Cauchy, S. †Lie, and many other mathematicians. The fundamental existence theorem for the initial value problem, now called the Cauchy-Kovalevskaya theorem, was proved by S. Kovalevskaya in 1875 (→ 321 Partial Differential Equations (Initial Value Problems) B).

Because of their close connection with problems of physics, linear equations of the second order have been a chief object of research. Up to the 19th century, classification into †elliptic, †hyperbolic, and †parabolic types and the study of boundary and initial value problems for each of these types constituted the main part of the theory.

In the 20th century, more complicated problems—†nonlinear problems appearing in the study of viscous or compressible fluids, or the study of equations of †mixed type in connection with supersonic flow—have emerged as important topics; and the newly developed techniques of functional analysis have brought about remarkable changes. Especially in the study of the †Schrödinger equations of quantum mechanics and of more general †evolution equations , this method has proved to be a powerful tool.

Finally, we should not fail to mention that the development of electronic computers has made it possible to obtain numerical solutions and to discover many important facts. †Numerical analysis is now becoming an indispensable part of the theory (→ 304 Numerical Solution of Partial Differential Equations).

References

[1] E. L. Ince, Ordinary differential equations, Dover, 1956.
[2] L. S. Pontryagin, Ordinary differential equations, Addison-Wesley, 1962. (Original in Russian, 1961.)
[3] E. A. Coddington and N. Levinson, Theory of ordinary differential equations, McGraw-Hill, 1955.
[4] P. Hartman, Ordinary differential equations, Wiley, 1964.
[5] E. Hille, Lectures on ordinary differential equations, Addison-Wesley, 1969.
[6] L. Bieberbach, Theorie der gewöhnlichen Differentialgleichungen auf funktionen theoretischen Grundlage dargestellt, Springer, 1953.
[7] E. Hille, Ordinary differential equations in the complex domain, Wiley, 1976.
[8] W. Wasow, Asymptotic expansions for ordinary differential equations, Interscience, 1965.
[9] I. Kaplansky, An introduction to differential algebra, Hermann, 1957.
[10] M. A. Naĭmark, Linear differential operators, I, II, Ungar, 1967, 1968. (Original in Russian, 1954.)
[11] R. Courant and D. Hilbert, Methods of mathematical physics I, II, Interscience, 1953, 1962.
[12] P. R. Garabedian, Partial differential equations, Wiley, 1964.
[13] I. G. Petrovskiĭ, Lectures on partial differential equations, Interscience, 1954. (Original in Russian, 1950.)
[14] L. Hörmander, Linear partial differential operators, Springer, 1963.
[15] K. Yosida, Lectures on differential and integral equations, Wiley, 1960. (Original in Japanese, 1950.)

108 (XIX.10) Differential Games

A. Introduction

The study of differential games arose from the study of pursuit and evasion problems and various tactical problems. The first work was done by R. P. Isaacs [1] in a series of RAND Corporation memoranda that appeared in 1954. He applied to many illustrative examples the method of Hamilton-Jacobi differential equations (→ eq. (4) below). The heuristic results of Isaacs were made rigorous by W. H. Fleming [2], L. D. Berkovitz [3], A. Friedman [4, 5], and others.

B. Zero-Sum Two-Person Games

Suppose that there are two antagonists, each exerting partial control over the state of a system. One wishes to maximize a given payoff that is a functional of the state and the control exerted, while the other wishes to minimize this payoff. Let the state of a differential game at time t be represented by an n-dimensional vector $x(t)\in \mathbf{R}^n$. In a zero-sum differential game between two players I and II, we are given a system of n differential equations

$$dx/dt=f(t,x,u,v) \qquad (1)$$

with an initial condition $x(\tau)=\xi\in\mathbf{R}^n$, where $u\in\mathbf{R}^p$ is chosen at each instant of time by player I and $v\in\mathbf{R}^q$ is chosen by player II. We assume that the function $f(t,x,u,v)$ is continuous in t and continuously differentiable on the entire (x,u,v)-space.

It is usually assumed that both players know the present state of the game and that they know how the game proceeds; that is, they know the system (1). Each player can take the state of the game into account in making his choice. Thus player I can let his choice of u be governed by a vector function $u(t,x)$ defined on D, where $D\subset[0,\infty)\times\mathbf{R}^n$ is a fixed region of the (t,x)-space. Similarly, player II can let his choice of v be governed by a vector function $v(t,x)$ defined on D.

A finite collection of subregions $D_1,\dots,D_r$ of a region D is said to constitute a decomposition of D whenever the following conditions hold: (i) Each D_i $(i=1,\dots,r)$ is connected and has a piecewise smooth boundary; (ii) $D_i\cap D_j=\varphi$ if $i\neq j$. A function defined on $\bar{D}$ is said to be piecewise C^1 in x on $\bar{D}$ if there is a decomposition of D such that on each D_i the function and all its derivatives with respect to x are continuous in (t,x) on $\bar{D}_i$. Let U and V be fixed closed subsets of $\mathbf{R}^p$ and $\mathbf{R}^q$, respectively.

Let S_u denote the class of functions $u(t,x)$ that are piecewise C^1 in x on $\bar{D}$ and have their range in U. Similarly, let S_v denote the class of functions $v(t,x)$ that are piecewise C^1 in x on $\bar{D}$ and have their range in V.

Let $u\in S_u$ and $v\in S_v$, and consider the differential equation

$$dx/dt=f(t,x,u(t,x),v(t,x)), \qquad (2)$$

subject to the initial condition

$$x(\tau)=\xi. \qquad (3)$$

We say that a pair $(u,v)\in S_u\times S_v$ is **playable** if for every (τ,ξ) in D every solution of (2) satisfying (3) stays in D and reaches a terminal manifold F in finite time, where F is a smooth manifold contained in $\bar{D}$. Let $\Omega_u\subset S_u$, $\Omega_u\subset S_v$ be the maximal pair of subclasses such that each pair $(u,v)\in\Omega_u\times\Omega_v$ is playable. We call the functions in Ω_u and Ω_v the **strategies** for the players.

For each strategy pair (u,v) we can define a functional

$$J(\tau,\xi,u,v)$$
$$=g(t_1,x(t_1))+\int_{t_0}^{t_1}h(t,x(t),u(t,x(t)),v(t,x(t)))\,dt,$$

where $x(t)$ is a solution of (2) and (3) and t_1 is the first time that $(t,x(t))$ reaches the terminal manifold F. The functional J is called the **payoff**.

Let $(u^*,v^*)\in\Omega_u\times\Omega_v$ be a strategy pair. Suppose that for any $u\in\Omega_u$ and $v\in\Omega_v$, the inequalities

$$J(\tau,\xi,u,v^*)\leqslant J(\tau,\xi,u^*,v^*)\leqslant J(\tau,\xi,u^*,v)$$

hold for all $(\tau,\xi)\in D$. We say that (u^*,v^*) is a **saddle point** relative to the classes Ω_u and Ω_v. The function

$$W(t,x)=J(t,x,u^*,v^*)$$

defined on D is called the **value function** of the game. Berkovitz [3] proved that the value function $W(t,x)$ is continuous on D and continuously differentiable on each D_i and satisfies

$$\max_{u\in U}[h(t,x,u,v^*)+W_x(t,x)f(t,x,u,v^*)]$$
$$=\min_{v\in V}[h(t,x,u^*,v)+W_x(t,x)f(t,x,u^*,v)]$$
$$=h(t,x,u^*,v^*)+W_x(t,x)f(t,x,u^*,v^*)$$
$$=-W_t(t,x). \qquad (4)$$

Equation (4) is called the **Hamilton-Jacobi equation**.

Let $x^*(t;\tau,\xi)$ be the optimal trajectory corresponding to the saddle-point strategies (u^*,v^*) and resulting from an initial point $(\tau,\xi)\in D$. Then there exists an n-dimensional continuous

vector function $\lambda(t;\tau,\xi)$ such that the following hold [3]:

(i) The functions x^* and λ satisfy the system of differential equations

$$dx/dt = f(t, x, u^*(t,x), v^*(t,x)),$$

$$d\lambda/dt = -H_x(t, x, \lambda, u^*(t,x), v^*(t,x)),$$

where

$$H(t,x,\lambda,u,v) = h(t,x,u,v) + \lambda f(t,x,u,v).$$

(ii) If $x = x^*(t;\tau,\xi)$, $t \geqslant \tau$, then

$$W_x(t,x) = \lambda(t;\tau,\xi).$$

(iii) At $t = t_1$, the **transversality condition**

$$H\frac{\partial T}{\partial \sigma} + \frac{\partial g}{\partial t}\frac{\partial T}{\partial \sigma} + \frac{\partial g}{\partial x}\frac{\partial X}{\partial \sigma} - \lambda\frac{\partial X}{\partial \sigma} = 0$$

holds, where the terminal manifold F is given parametrically by the relations

$$t = T(\sigma), \qquad x = X(\sigma),$$

σ ranging over a cube in some finite-dimensional Euclidean space.

(iv) For all $\tau \leqslant t \leqslant t_1$,

$$\max_{u\in U}\min_{v\in V} H(t, x^*(t), \lambda(t), u, v)$$

$$= \min_{v\in V}\max_{u\in U} H(t, x^*(t), \lambda(t), u, v)$$

$$= H(t, x^*(t), \lambda(t), u^*(t, x^*(t)), v^*(t, x^*(t))).$$

P. Varaiya and J. Lin [6] and Friedman [4, 5] have defined certain special classes of differential games, and have shown that under their definitions the games have nonzero value functions.

C. N-Person Differential Games

In a differential game between many players, the state vector $x(t) \in \mathbf{R}^n$ is governed by

$$dx/dt = f(t, x, u_1, \ldots, u_N), \qquad x(\tau) = \xi,$$

where $u_i \in \mathbf{R}^{p_i}$ is chosen at each instant of time by player i. Each u_i is constrained to lie in a fixed closed subset U_i of $\mathbf{R}^{p_i}$. Let S_i be the class of functions $u_i(t,x)$ that are piecewise C^1 in x on $\bar{D}$ and have their range in U_i.

We say that an element $u = (u_1, \ldots, u_N)$ of $S_1 \times \ldots \times S_N$ is **playable** if, for every $(\tau, \xi) \in D$, every solution of the differential equation

$$dx/dt = f(t, x, u_1(t,x), \ldots, u_N(t,x)), \qquad x(\tau) = \xi, \tag{5}$$

stays in D and reaches a terminal manifold F in finite time. Let $\Omega_i \subset S_i$ $(i = 1, \ldots, N)$ be the maximal subclasses such that each element $u = (u_1, \ldots, u_N) \in \Omega_1 \times \ldots \times \Omega_N$ is playable. We call the functions $u_i \in \Omega_i$ $(i = 1, \ldots, N)$ the **strategies**. For each strategy N-tuple we define a **payoff**

$$J_i(\tau, \xi, u_1, \ldots, u_N) = g_i(t_1, x(t_1))$$
$$+ \int_{t_0}^{t_1} h_i(t, x(t), u_1(t, x(t)), \ldots, u_N(t, x(t)))\,dt,$$

where $x(t)$ is a solution of (5) and t_1 is the first time that $(t, x(t))$ reaches the terminal manifold F. Each player i is to choose his strategy $u_i \in \Omega_i$ so as to maximize his own payoff J_i.

There are many definitions of "solution" for games involving more than two players. A strategy N-tuple $u^* = (u_1^*, \ldots, u_N^*)$ is called an **equilibrium point** for the game if the inequalities

$$J_i(u_1^*, \ldots, u_{i-1}^*, u_i, u_{i+1}^*, \ldots)$$
$$\leqslant J_i(u_1^*, \ldots, u_{i-1}^*, u_i^*, u_{i+1}^*, \ldots) \qquad (i = 1, \ldots, N)$$

hold for any $u_1 \in \Omega_1, \ldots, u_N \in \Omega_N$. J. H. Case [9] has shown that the conclusions drawn for zero-sum two-person games also hold for N-person differential games.

References

[1] R. P. Isaacs, Differential games, Wiley, 1965.
[2] W. H. Fleming, The convergence problem for differential games II, Advances in game theory, Ann. Math. Studies 52, Princeton Univ. Press, 1964, 195–210.
[3] L. D. Berkovitz, Necessary conditions for optimal strategies in a class of differential games and control problems, SIAM J. Control, 5 (1967), 1–24.
[4] A. Friedman, On the definition of differential games and the existence of value and of saddle points, J. Differential Equations, 7 (1970), 69–91.
[5] A. Friedman, Differential games, Wiley, 1971.
[6] P. Varaiya and J. Lin, Existence of saddle points in differential games, SIAM J. Control, 7 (1969), 142–157.
[7] H. W. Kuhn and G. P. Szegö (eds.), Differential games and related topics, North-Holland, 1971.
[8] L. S. Pontryagin, On the theory of differential games, Russian Math. Surveys, 21 (1966), 193–246.
[9] J. H. Case, Toward a theory of many player differential games, SIAM J. Control, 7 (1969), 179–197.

109 (VII.1) Differential Geometry

In differential geometry in the classical sense, we use differential calculus to study the prop-

erties of figures such as curves and surfaces in Euclidean planes or spaces. Owing to his studies of how to draw tangents to smooth plane curves, P. †Fermat is regarded as a pioneer in this field. Since his time, differential geometry of plane curves, dealing with curvature, †circles of curvature, †evolutes, †envelopes, etc., has been developed as a part of calculus. Also, the field has been expanded to analogous studies of space curves and surfaces, especially of †asymptotic curves, †lines of curvature, †curvatures and †geodesics on surfaces, and †ruled surfaces. C. F. †Gauss founded the theory of surfaces by introducing concepts of the †geometry on surfaces (*Disquisitiones circa superficie curvas*, 1827). Gauss recognized the importance of the intrinsic geometry of surfaces, and it is generally agreed that differential geometry as it is known today was initiated by him. Thus differential geometry came to occupy a firm position as a branch of mathematics. The influence that differential-geometric investigations of curves and surfaces have exerted upon branches of mathematics, physics, and engineering has been profound. For example, E. Beltrami discovered an intimate relation between the geometry on a †pseudo-sphere and †non-Euclidean geometry. The study of †geodesics is a fertile topic deeply related to dynamics, the calculus of variations, and topology, on which there is excellent work by J. Hadamard, H. †Poincaré, P. Funk, G. D. Birkhoff, M. Morse, R. Bott, W. Klingenberg, and M. Berger, among others. The theory of minimal surfaces initiated by J. L. Lagrange was an application of the calculus of variations. At the early stages of development, G. Monge, J. B. M. C. Meusnier, A. M. Legendre, O. Bonnet, B. Riemann, K. Weierstrass, H. A. Schwarz, Beltrami, and S. Lie contributed to the theory. Weierstrass and Schwarz established its relationship with the theory of functions. J. A. Plateau showed experimentally that †minimal surfaces can be realized as soap films by dipping wire in the form of a closed space curve into a soap solution (1873). The Plateau problem, i.e., the problem of proving mathematically the existence of a minimal surface with prescribed boundary curve, was solved by Tibor Radó in 1930 and independently by J. Douglas in 1931. Although the relationship to function theory is lost for higher-dimensional minimal submanifolds, their study is intimately related to the calculus of variations and topology.

Euclidean geometry is a geometry belonging to F. Klein's Erlangen program (→ 137 Erlangen Program). For other geometries in the sense of F. Klein we may also consider the corresponding differential geometries. For instance, in †projective differential geometry we study by means of differential calculus the properties of curves and surfaces that are invariant under projective transformations. This subject was studied by E. J. Wilczynski, G. Fubini, and others; †affine differential geometry and †conformal differential geometry were studied by W. Blaschke and others (→ 110 Differential Geometry in Specific Spaces).

Influenced by Gauss's geometry on surfaces, in his inaugural address at Göttingen in 1854 Riemann advocated an intrinsic differential geometry completely independent of embeddings (Über die Hypothesen, welche der Geometrie zugrund liegen; *Werke*, 2nd ed. 1892, 272–287) (→ 364 Riemannian Manifolds). Removing the restriction to two dimensions and considering abstract manifolds of dimension n, he introduced what is now known as the Riemannian metric; actually he considered the more general metrics that had formed the subject matter of the dissertation of P. Finsler in 1918 (→ 152 Finsler Spaces). Riemannian geometry includes Euclidean and non-Euclidean geometry as special cases, and is important for the great influence it exerted on geometric ideas of the 20th century. Under the influence of the algebraic theory of invariants, Riemannian geometry was then studied as a theory of invariants of quadratic †covariant tensors by E. B. Christoffel, C. G. Ricci, and others. Riemannian geometry attracted wide attention after A. Einstein applied it to the †general theory of relativity in 1916.

In the same year, T. Levi-Civita introduced the notion of †Levi-Civita parallelism, which contributed greatly to the clarification of geometric properties of Riemannian spaces. Observing parallelism to be an affine-geometric concept, H. Weyl and A. S. Eddington developed a theory of Riemannian spaces "affinely" based on the notion of parallelism without using metrical methods. Such a geometry is called a geometry of an affine connection (→ 80 Connections).

Every straight line in a Euclidean space has the property that all tangents to the line are parallel. In a space with an affine connection, we may define a family of curves called †paths as an analog of straight lines. Such curves are solutions of a system of ordinary differential equations of the second order of a certain type. Coefficients of such differential equations determine a parallelism and hence an affine connection. H. Weyl discovered transformations of coefficients that leave the family of paths invariant as a whole, namely, projective transformations of an affine connection. A geometry that aims to study properties of paths or affine connections that are invariant under these transformations is called a **projective geometry of paths**. Such geometry was

studied by L. P. Eisenhart, O. Veblen, and others. The concept of projective connections was an outcome of such studies. Similarly, the concept of conformal connections was developed from the consideration of †conformal transformations of Riemannian spaces.

These geometries cannot in general be regarded as geometries in the sense of Klein. Actually, any one of these geometries generally has no transformations that correspond to †congruent transformations of geometries in the sense of Klein; even if it has such transformations, they do not act transitively on the space. Thus geometries are naturally divided into two categories, one consisting of geometries in the sense of Klein (based on the group concept) and the other of geometries based on Riemann's idea. Under such circumstances, E. Cartan unified the thoughts of Klein and Riemann from a higher standpoint and constructed his theory of connections in a series of papers published between 1923 and 1925. He developed the theory of affine, projective, and conformal connections from a viewpoint consistent with that of Klein. Just as each tangent space of a Riemannian manifold is viewed as a Euclidean space, an affine connection regards the tangent space at each point as an affine space and develops it onto the tangent space at an infinitesimally nearby point. In discussing projective connections Cartan attached a projective space to each point of a manifold as an infinitesimal approximation, and similarly for conformal connections. More generally, he attached to each point of a manifold a fixed Klein space, i.e., a homogeneous space of a Lie group, called the structure group. Thus Cartan introduced the concept of fiber bundle (→ 147 Fiber Bundles). Then he defined a connection as a development of the fiber, i.e., the generalized tangent space, at each point onto the fiber at an infinitesimally nearby point (→ 80 Connections B). If G is the group of congruent transformations in Euclidean space, a manifold with connection having G as its structural group is called a **manifold with Euclidean connection**. Among manifolds with Euclidean connection, Riemannian manifolds are characterized as those without †torsion. If we take the group of congruent transformations of projective (conformal) geometry as G, we have manifolds with projective (conformal) connection in the sense of Cartan. Among these, there are remarkable ones called manifolds with normal projective (or conformal) connections, which are essentially the same as the ones studied by Veblen and others. Cartan's idea had a profound influence on modern differential geometry. The method of moving frames, created by G. Darboux and extensively used by Cartan in his theory of connections, was a forerunner of the theory of fiber bundles. Combining †Grassmann algebra with differential calculus, Cartan developed a powerful computational tool known as calculus of differential forms (→ 105 Differentiable Manifolds Q). Differential forms have become indispensable in topology, algebraic geometry, and in studies of functions of several complex variables, as well as in differential geometry.

The work of Lie on transformation groups also had a profound influence on Cartan. The latter's work on †Lie groups, particularly on simple Lie groups, and on differential geometry culminated in 1926 in his discovery of †Riemannian symmetric spaces. These spaces are natural generalizations of the spherical surface and the unit disk in the complex plane with Poincaré metric, and play essential roles in unitary representation theory and in other areas of mathematics.

A tangent line to a curve C at a point P of C is the limit line of the line PQ, where Q is a point on C aporoaching P; hence we can define it locally. A concept (or property), such as this, that can be defined in an arbitrary small neighborhood of a point of a given figure or a space is called a **local concept** (or **local property**) or a concept (or property) **in the small**. In the early stages of the development of differential geometry, differential calculus was the main tool of study, so most of the results were local. On the other hand, a concept (or property) that is defined in connection with a whole figure or a whole space is called **global** or **in the large**. In modern differential geometry, the study of relations between local and global properties has attracted the interest of mathematicians. This view was emphasized by Blaschke, who worked on the differential geometry of †ovals and †ovaloids. The study of †rigidity of ovaloids by S. Cohn-Vossen belongs in this category, and many works on geodesics and minimal surfaces were done from this standpoint.

From the viewpoint of modern mathematics, the basic concepts on which we construct Riemannian geometry and geometries of connections are global concepts of †differentiable manifolds. However, in Riemann's time the theory of †Lie groups and topology were not yet developed; consequently, Riemannian geometry remained a local theory. In 1925, H. Hopf began to study the relations between local differential-geometric structures and the topological structures of Riemannian spaces. However, except for the work of Cartan, Hopf, and a few others, differential geometry in the 1920s was still largely concerned with surfaces in the 3-dimensional Euclidean space or local properties of Riemannian manifolds, and with affine, projective, and conformal connections.

Gradually the concept of differentiable manifolds was clarified, the global theory of Lie groups made progress, and topology developed; and the trend toward global differential geometry began slowly in the early 1930s. The dissertation of G.-W. de Rham published in 1931 showed that the cohomology of a manifold can be computed in terms of differential forms (→ 105 Differentiable Manifolds R). His theorem provides the theoretical foundation for expressing cohomological invariants of a manifold in terms of differential geometric invariants. In a series of papers immediately following de Rham's, W. V. D. Hodge established that, on a compact Riemannian manifold, every r-dimensional cohomology class can be uniquely represented by a harmonic form of degree r (→ 194 Harmonic Integrals).

An important class of complex manifolds with compatible Riemannian metric was discovered by J. A. Schouten, D. van Dantzig, and E. Kähler around 1929–1932. This class of manifolds, called Kähler manifolds today, comprises the projective algebraic manifolds. Hodge's theory of harmonic integrals is most effective when applied to compact Kähler manifolds, (→ 232 Kähler Manifolds).

The most celebrated global theorem in classical differential geometry of surfaces is the Gauss-Bonnet formula (1848) (→ 364 Riemannian Manifolds D). The formula was generalized to closed hypersurfaces of Euclidean space by Hopf in 1925, to closed submanifolds of Euclidean space by C. B. Allendoerfer and W. Fenchel in 1940, and finally to arbitrary closed Riemannian manifolds by Allendoerfer and A. Weil in 1943. But the simple proof given by S. S. Chern in 1944 contained the notion of transgression, which has become essential in the theory of characteristic classes (→ 56 Characteristic Classes). The discovery of Pontryagin classes for Riemannian manifolds (1944) and Chern classes for Hermitian manifolds (1946) culminated in the †index theorem and the †Riemann-Roch theorem of F. E. P. Hirzebruch, and finally in the †Atiyah-Singer index theorem.

A simple but fruitful idea of S. Bochner, relating harmonic forms to curvature, established †vanishing theorems for harmonic forms of Riemannian manifolds and for holomorphic forms of Kähler manifolds under suitable positivity conditions for curvature. His idea has led to the vanishing theorems of K. Kodaira and others (→ 232 Kähler Manifolds D).

The work of C. Ehresmann in 1950 on connections in principal fiber bundles established a solid foundation to Cartan's theory of connections. Gauge theory in physics is largely based on the theory of connections in principal fiber bundle.

R. H. Nevanlinna's value distribution theory and its subsequent generalization by Chern and others can be best described in differential-geometric terms.

Differentiable manifolds are currently objects of research in both differential geometry and differential topology. While topology studies manifolds per se, differential geometry may be considered as the study of differentiable manifolds equipped with geometric structures, such as metric tensors, connections, (almost) complex structures, and various other tensors. Through these geometric structures, differential geometry enjoys close contact with many branches of mathematics. From its early days, differential geometry has had close ties to topology (as exemplified by the Gauss-Bonnet formula) and to partial differential equations and analytic functions (through, e.g., the study of minimal surfaces). The bonds with topology were strengthened by Morse theory and, more recently, by the theory of characteristic classes. In the most recent proof of the Atiyah-Singer index theorem, differential geometry is an important intermediary between topology and analysis. Differential geometry and algebraic geometry have enriched each other through Kähler manifolds. The theory of functions of several complex variables also has points of contact with differential geometry, such as value distribution theory and Cauchy-Riemann structures. Contact and symplectic structures are basic to mechanics. Lorentz manifolds and connections in principal bundles are essential mathematical tools in the general theory of relativity and in gauge theory. Topics such as minimal submanifolds, manifolds of positive curvature, and closed geodesics are active and important areas of research belonging to Riemannian geometry proper; at the same time, differential geometry provides a language and methods that are important in wider areas of mathematics.

References

[1] L. P. Eisenhart, A treatise on the differential geometry of curves and surfaces, Ginn, 1909.

[2] G. Darboux, Leçons sur la théorie générale des surfaces, Gauthier-Villars, I, second edition, 1914; II, second edition, 1915; III, 1894; IV, 1896.

[3] W. Blaschke, Vorlesungen über Differentialgeometrie, Springer, I, third edition, 1930; II, 1923; III, 1929 (Chelsea, 1967).

[4] L. P. Eisenhart, Riemannian geometry, Princeton Univ. Press, 1926.

[5] E. Cartan, Leçons sur la géométrie des espaces de Riemann, Gauthier-Villars, second edition, 1946; English translation, Geometry of Riemannian spaces, Math. Sci. Press, 1983.
[6] J. A. Schouten, Ricci-calculus, Springer, second edition, 1954.
[7] O. Veblen and J. H. C. Whitehead, The foundations of differential geometry, Cambridge Tracts, 1932.
[8] S. S. Chern, Topics in differential geometry, Lecture notes, Inst. for Adv. Study, Princeton, 1951.
[9] W. V. D. Hodge, The theory and applications of harmonic integrals, Cambridge Univ. Press, second edition, 1952.
[10] S. Kobayashi and K. Nomizu, Foundations of differential geometry, Interscience, I, 1953; II, 1969.
[11] E. Cartan, Oeuvres complètes I–III, Gauthier-Villars, 1952–1955; second edition, CNRS, 1984.
[12] A. L. Besse, Manifolds all of whose geodesics are closed, Springer, 1978.
[13] M. Berger, P. Gauduchon, and E. Mazet, Le spectre d'une variété Riemanninne, Lecture notes in math. 194, Springer, 1971.
[14] D. Gromoll, W. Klingenberg, and W. Meyer, Riemannsche Geometrie im Groosen, Lecture notes in math. 55, Springer, 1968.
[15] S. Helgason, Differential geometry, Lie groups, and symmetric spaces, Academic Press, 1978.
[16] S. Kobayashi, Transformation groups in differential geometry, Springer, 1972.
[17] T. Levi-Civita, The absolute differential calculus, Blackie, 1926.
[18] J. W. Milnor, Lectures on Morse theory, Ann. Math. Studies 51, Princeton Univ. Press, 1963.
[19] R. Osserman, A survey of minimal surfaces, Van Nostrand, 1969.
[20] G. de Rham, Variétés differentiables, Hermann, 1955.
[21] A. Weil, Introduction à l'étude des variétés kähleriennes, Hermann, 1958.
[22] K. Yano and S. Bochner, Curvature and Betti numbers, Ann. Math. Studies 32, Princeton Univ. Press, 1953.
[23] Y. Matsushima, Differentiable manifolds, Dekker, 1972.

110 (VII.17) Differential Geometry in Specific Spaces

A. The Method of Moving Frames

The main theme of this article is the theory of surfaces (i.e., submanifolds) in a differentiable manifold V on which a †Lie transformation group G acts.

If a †Lie group G of dimension r acts †transitively on a space G_* and the †stability group for any point of G_* consists of the identity element only, then G_* is called the **group manifold** of G, and an element of G_* is called a **frame**. If f_0 is a fixed frame, then the mapping $a \to af_0 \in G_*$ $(a \in G)$ gives a †diffeomorphism of G to G_*. Let $I(G)$ be the set of all †differential forms ω of degree 1 on G_* that are invariant under transformations of G. Then $I(G)$ is a linear space of dimension r and is the †dual space of the Lie algebra $\mathfrak{g}$ of G. A basis $\{\omega_\lambda \mid 1 \leqslant \lambda \leqslant r\}$ of $I(G)$ is called a set of **relative components** of G. The **structure equations** hold:

$$d\omega_\lambda = \frac{1}{2}\sum_{\mu,\nu=1}^{r} c_{\lambda\mu\nu}\omega_\mu \wedge \omega_\nu,$$

where $c_{\lambda\mu\nu}$ are †structure constants of the Lie algebra $\mathfrak{g}$.

Let G be a Lie transformation group of a space E. Then a G-invariant submanifold of E on which G acts transitively is called an **orbit**. Each point $y \in E$ determines an orbit containing y. When there exist parameters k_j $(1 \leqslant j \leqslant t)$ such that any G-invariant on E is a function of $k_1, \ldots, k_t$, then these parameters are called the **fundamental invariants** of E. Let H $(\subset G)$ be the stability group at a point y_0 on an orbit M; then M is identified with the †homogeneous space G/H by the diffeomorphism $\varphi: M \to G/H$, $\varphi(ay_0) = aH$ $(a \in G)$. Furthermore, a †principal fiber bundle (G_*, M, H, τ) is determined by the projection $\tau: G_* \to M$, $\tau(af_0) = ay_0$ $(a \in G)$. The †fiber H_y on a point $y \in M$ is a group manifold of H. H_y is called the **family of frames** on y, and an element of H_y is a frame on y. Local coordinates θ_μ $(1 \leqslant \mu \leqslant s)$ of the group H are called the **secondary parameters** and are used to indicate frames in H_y. When H is not connected, let H^0 be the connected component of the identity of H and $\tilde{M}$ be the †covering manifold G/H^0 of M. An element $\tilde{y} \in \tilde{M}$ over $y \in M$ is called an **oriented element**. Now assume that the group H is connected. Then the family of frames H_y on each $y \in M$ is given as an †integral manifold of a †completely integrable system of total differential equations $\pi_i = 0$ $(1 \leqslant i \leqslant r-s$, $\pi_i \in I(G))$ on the group manifold G_*. Here the π_i are linearly independent and are called the **horizontal components** of M. The π_i are linear combinations of the relative components ω_λ of G, and their coefficients are generally functions of the fundamental invariants k_j. For simplicity, we assume that the relative components $\{\omega_\lambda\}$ are chosen such that the horizontal components π_i and components ω_α $(r-s < \alpha \leqslant r)$ are linearly independent. Then the ω_β $(1 \leqslant \beta \leqslant s)$ are called the **secondary com-**

ponents. The differentials $d\theta_\mu$ of the secondary parameters are linear combinations of the ω_β. Furthermore, let $\{x_\sigma\}$ $(1 \leqslant \sigma \leqslant n)$ be local coordinates of E; then the differentials dx_σ are linear combinations of the differentials dk_j and the horizontal components π_i.

Let G be a Lie transformation group of a space V. We regard two m-dimensional surfaces W_1 and W_2 passing through $x \in V$ as equivalent if they have a †contact of order p at x. Then an equivalence class of submanifolds is called a **contact element** of order p at the point x. Let E_p be the set of all contact elements of order p at x, where x runs over all the points in V. A contact element of order p naturally determines a contact element of order $p-1$, and we denote this correspondence by $\psi: E_p \rightarrow E_{p-1}$. Thus we obtain the series of correspondences

$$V = E_0 \overset{\psi}{\leftarrow} E_1 \leftarrow \ldots \leftarrow E_{p-1} \overset{\psi}{\leftarrow} E_p \leftarrow \ldots,$$

where a contact element of order 0 is identified with a point of V. Since a transformation on the space V induces a transformation on E_p, G is also a Lie transformation group of E_p, and this transformation commutes with the mapping ψ. The fundamental invariants k_j of E_p are said to be of order p. We use similar terminology (such as frames of order p, etc.) throughout this article. The fundamental invariants k_j of order p $(1 \leqslant j \leqslant t_p)$ can be chosen such that they contain the fundamental invariants k_i $(1 \leqslant i \leqslant t_{p-1})$ of order $p-1$. The additional $t_p - t_{p-1}$ invariants k_α $(t_{p-1} < \alpha \leqslant t_p)$ are called the **invariants of order** p. The family H_y^p $(y \in E_p)$ of frames of order p can be chosen such that H_y^p is contained in the family H_z^{p-1} $(z = \psi y \in E_{p-1})$ of frames of order $p-1$. If necessary, the family H_y^p of frames of order p can be made connected by defining an **orientation** of contact elements of order p. Furthermore, the horizontal components π_j $(1 \leqslant j \leqslant r - s_p)$ of order p can be chosen such that they contain the horizontal components π_i $(1 \leqslant i \leqslant r - s_{p-1})$. The additional $s_{p-1} - s_p$ components π_α $(r - s_{p-1} < \alpha \leqslant r - s_p)$ are called the **principal components of order** p.

Let W be an m-dimensional surface of a space V. The contact element of order p $(\geqslant 0)$ is determined at every point of W and expressed by the family of frames of order p and the values of invariants of orders less than or equal to p. Let $\{u_i\}$ $(1 \leqslant i \leqslant m)$ be local coordinates on W. Then the differentials du_i are given as linear combinations of linearly independent differential forms π_i $(1 \leqslant i \leqslant m)$, where the π_i, called the **basic components** of W, are certain linear combinations of the differentials of the fundamental invariants of V and of the horizontal components of orbits of V. Let $F^p(W)$ be the set of all families of frames of order p; then $F^p(W)$ depends on m parameters u_i and s_p secondary parameters θ_μ of order p. On the space $F^p(W)$, the differentials of invariants of orders less than or equal to $p-1$ and the principal components of orders less than or equal to $p-1$ are linear combinations of the basic components, whose coefficients are functions of the invariants of orders less than or equal to p. The differentials of invariants of order p and the principal components of order p are linear combinations of the basic components:

$$dk_\alpha = h_{\alpha 1}\pi_1 + \ldots + h_{\alpha m}\pi_m, \qquad t_{p-1} < \alpha \leqslant t_p,$$

$$\pi_\alpha = b_{\alpha 1}\pi_1 + \ldots + b_{\alpha m}\pi_m, \qquad r - s_{p-1} < \alpha \leqslant r - s_p,$$

where the coefficients $h_{\alpha i}$, $b_{\alpha i}$ are functions of the invariants of orders less than or equal to p and, in general, the secondary parameters θ_μ of order p. These coefficients are called the **coefficients of order** p. Let Γ_p be a subgroup of G preserving a family of frames of order p and D_p be a space whose coordinates are coefficients $(h_{\alpha i}, b_{\alpha i})$ of order p. Then Γ_p acts on D_p as a transformation group. Knowledge of the properties of contact elements of order less than or equal to p can be utilized to obtain information about the invariants of order $p+1$, etc. In fact, if we can choose in the Γ_p-space D_p a subspace C_p that intersects each orbit in D_p at one and only one point, then in general the secondary parameters of order p associated with the points in C_p correspond to the frames of order p, and the parameters associated with the points in C_p are the invariants of order $p+1$. The restrictions of the coefficients of order p to C_p are functions of the invariants of orders less than or equal to $p+1$; they are independent of the secondary parameters of order p. Thus the frames of order $p+1$ and the invariants of orders less than or equal to $p+1$ determine the contact elements of order p of W and their differentials; generally, the latter can be utilized to determine the contact elements of order $p+1$.

This process of obtaining information of "order $p+1$" utilizing a suitable subspace C_p of D_p is the so-called general **method of moving frames**. However, the surface W may contain points for which the general method does not apply. Actually, there are surfaces W for which the method does not apply for any point in W. Thus various methods of moving frames are necessary to cope with different kinds of surfaces. In the actual application of the method of moving frames, we use certain devices that help to simplify the calculations. In fact, an infinitesimal transformation $(\delta h_{\alpha i}, \delta b_{\alpha i})$ of the group Γ_p acting on the space D_p is expressed as a linear combination of the secondary components of order p; this expression is easily obtained by means of the structure equations of

G. The group Γ_{p+1} is a subgroup of Γ_p fixing every point of the subspace C_p, and its infinitesimal transformation is such that $\delta h_{\alpha i}=0$, $\delta b_{\alpha i}=0$. The secondary components of order $p+1$ are immediately obtained from the equations for dk_α and π_α. Furthermore, when $m\geqslant 2$, the condition for the principal components of every order to satisfy the structure equations of G is essential to the problem of the existence of (m-dimensional) surfaces.

As we apply the method of moving frames consecutively to a surface W, we eventually arrive at the order q having the following properties: The families of frames of order $q+1$ coincide with those of order q, and the invariants k_β of order $q+1$ are expressed as functions $\varphi_\beta(k_\alpha)$ of invariants k_α of order less than or equal to q. In this case, the families of frames of orders $q+j$ ($j\geqslant 1$) are all equal, and the invariants of orders $q+j$ are partial ($j-1$)-derivatives of functions $\varphi_\beta(k_\alpha)$. The family of frames of order q is called the **Frenet frame**. The **differential invariants** on a surface are defined to be differential forms generated by the basic components and the invariants of each order.

Specifically, assume that the group G is an analytic transformation group of V, and the m-dimensional surfaces W_1, W_2 are analytic. Then there exists an element g of G such that $gW_1=W_2$ if and only if W_1 and W_2 are of the same kind and have the same relations among the invariants of orders less than or equal to $q+1$. These relations are called the **natural equations** of the surface. The theory of surfaces based on the analysis of the natural equations of surfaces is called **natural geometry**. The **reduction formula** can be obtained by utilizing the Frenet frame; it gives the equation of the surface in the form of power series containing the invariants of each order.

Various results are known concerning the theory of surfaces of the spaces V_1, V_2 sharing the same transformation group G. We also have a theory of **special surfaces** whose invariants satisfy specific functional relations. Furthermore, we have problems concerning the **deformation of a surface** (preserving some differential invariants). Actually, the theory of surfaces of dimension m other than curves and hypersurfaces is in general quite difficult. The methods of tensor calculus can be applied to the study of surfaces. The theory of †connections can be considered to be an outgrowth of the study of surfaces by means of the method of moving frames and tensor calculus.

B. Projective Differential Geometry

The rudiments of differential geometry subordinated to the †projective transformation group, or **projective differential geometry**, can be found in the *Theory of surfaces* by J. G. Darboux. The subject has been systematically studied by H. G. H. Halphen, E. J. Wilczynski, and G. Fubini. The Fubini theory was enriched substantially by E. Cartan, E. Čech, E. Bompiani, and J. Kanitani.

In this section we consider a surface S in a 3-dimensional projective space. Let $A(u^1,u^2)$ (u^1,u^2 are parameters on S) be a point of S, and associate with A all the frames $[A,A_1,A_2,A_3]$ ($|A,A_1,A_2,A_3|=1$), where A_1, A_2, A_3 are points of the tangent plane to S at A. A family of such frames is called the **family of frames of order 1**, and we express its differential by

$$dA_\alpha=\sum_{\beta=0}^{3}\omega_\alpha^\beta A_\beta,\qquad \alpha=0,1,2,3,\qquad A_0=A.$$

The ω_α^β are †Pfaffian forms that depend on two principal parameters determining the origin A and ten secondary parameters determining the frame. We have $\omega_0^0+\omega_1^1+\omega_2^2+\omega_3^3=0$, $\omega_0^3=0$. Furthermore, $\omega^1=\omega_0^1$, $\omega^2=\omega_0^2$ are independent of each other and depend on the principal parameters only. Let z^1, z^2, z^3 be †nonhomogeneous coordinates with respect to a frame of order 1. Then in a neighborhood of the origin, S is expressed by $z^3=\sum_{r=2}^{\infty}f_r$, where the f_r are homogeneous functions of degree r with respect to z^1, z^2.

If we write $f_2=(a_0(z^1)^2+2a_1z^1z^2+a_2(z^2)^2)/2$, then it follows from the structure equations of the projective transformation group that $\omega_1^3=a_0\omega^1+a_1\omega^2$, $\omega_2^3=a_1\omega^1+a_2\omega^2$. If we put $\varphi_2=a_0(\omega^1)^2+2a_1\omega^1\omega^2+a_2(\omega^2)^2$, then a curve on S defined by $\varphi_2=0$ is called the **asymptotic curve** and its tangent the **asymptotic tangent**. At any point of this curve, the plane tangent to S is in contact of order 2 with this curve, and there are in general two asymptotic curves through any point of S. Equations of the asymptotic tangent at A are given by $z^3=0$, $f_2=0$. A point of S at which the asymptotic tangents coincide is called a **parabolic point**. If every point of S is parabolic, then S is a †developable surface, and the general theory is not applicable to such a surface.

Among the family of frames of order 1, a frame satisfying $a_0=a_2=0$, $a_1=1$ is called the **frame of order 2**. For this frame, the straight lines $\overline{AA_1}$, $\overline{AA_2}$ are asymptotic tangents. With respect to this frame, if $f_3=-(b_0(z^1)^3+3b_1(z^1)^2z^2+3b_2z^1(z^2)^2+b_3(z^2)^3)/3$, then $\omega_1^2=b_0\omega^1+b_1\omega^2$, $-\omega_0^0+\omega_1^1+\omega_2^2-\omega_3^3=2(b_1\omega^1+b_2\omega^2)$, $\omega_2^1=b_2\omega^1+b_3\omega^2$, and the quadric surface $z^3=z^1z^2-z^3(b_1z^1+b_2z^2+pz^3)$ (with p arbitrary) is called **Darboux's quadric** at A, an especially interesting one among contact quadrics of S. **Darboux's curve** is a curve on S such

that Darboux's quadric is in contact of order 3 at any point of it. Its tangent is called **Darboux's tangent** and is given by $z^3=0$, $b_0(z^1)^3+b_3(z^2)^3=0$.

We have $b_0 b_3 \neq 0$, except in the case of †ruled surfaces. We take special frames of order 2 determined by $b_1=b_2=0$, $b_0=b_3=1$ and call them the **frames of order 3**. If a frame of order 3 satisfies $f_4=-(c_0(z^1)^4+4c_1(z^1)^3z^2+6(c_2-1)(z^1z^2)^2+4c_3z^1(z^2)^3+c_4(z^2)^4)/12$, then $\omega_0^0-2\omega_1^1+\omega_2^2=c_0\omega^1+c_1\omega^2$, $\omega_3^2-\omega_1^0=c_1\omega^1+c_2\omega^2$, $\omega_3^1-\omega_2^0=c_2\omega^1+c_3\omega^2$, $\omega_0^0+\omega_1^1-2\omega_2^2=c_3\omega^1+c_4\omega^2$. With respect to this family of frames of order 3, $((\omega^1)^3+(\omega^2)^3)/2\omega^1\omega^2$ is an invariant associated with two neighboring points of S, called the **projective line element**. Also with respect to this frame, two straight lines $\overline{AA_3}$, $\overline{A_1A_2}$ are polar with respect to Darboux's quadric.

Among the families of frames of order 3, a frame satisfying $c_1=c_2=c_3=0$ is called a **frame of order 4**. With respect to this frame, there exist λ, μ, ν, ρ such that $\omega_1^0=\lambda\omega^1+\mu\omega^2$, $\omega_2^0=\nu\omega^1+\rho\omega^2$, $\omega_3^0=\rho\omega^1+\lambda\omega^2$. Hence if we put $c_0=-3a$, $c_4=-3b$, it follows that

$$(\omega_\alpha^\beta)=\begin{pmatrix}\tau_0^0 & \omega^1 & \omega^2 & 0\\ \omega_1^0 & \tau_1^1 & \omega^1 & \omega^2\\ \omega_2^0 & \omega^2 & -\tau_1^1 & \omega^1\\ \omega_3^0 & \omega_2^0 & \omega_1^0 & -\tau_0^0\end{pmatrix},$$

$$\tau_0^0=-(3/2)(a\omega^1+b\omega^2),\quad \tau_1^1=(1/2)(a\omega^1+b\omega^2).$$

Thus the frame of order 4 is the Frenet frame and is attached to every point of S. This frame is called the **normal frame**, and the invariants a, b, λ, μ, ν, ρ are called the **fundamental differential invariants**. The straight lines $\overline{AA_3}$ and $\overline{A_1A_2}$ associated with the normal frame are called **directrices of Wilczynski** of the first and second kind, respectively. With respect to the normal frame, S is expressed by

$$z^3=z^1z^2-((z^1)^3+(z^2)^3)/3$$
$$+(a(z^1)^4+b(z^2)^4)/4+(z^1z^2)^2/2+\ldots.$$

A necessary and sufficient condition for two surfaces S, $\bar{S}$ to be projectively equivalent is that there be normal frames having the same ω^1, ω^2 and the same six fundamental differential invariants. For six quantities a, b, λ, μ, ν, ρ to be fundamental differential invariants of a surface, they must satisfy a certain condition of existence [6].

A frame of order 1 such that $\overline{AA_3}$ and $\overline{A_1A_2}$ are polar with respect to Darboux's quadric is called **Darboux's frame**. With respect to this frame also, a theory of surfaces has been established.

Consider a pointwise correspondence between two surfaces S, $\bar{S}$, and denote by $\bar{A}\in\bar{S}$ the point corresponding to $A\in S$. If there exists a projective transformation φ that transforms A into $\bar{A}$ and the image $\varphi(S)$ is in contact of order 2 with $\bar{S}$ at $\bar{A}$, then the pointwise correspondence is called a **projective deformation**. A necessary and sufficient condition for the existence of a projective deformation between two surfaces is that these surfaces have the same projective line element [6]. A ruled surface is projectively deformable only to a ruled surface. Given an arbitrary surface S, it is generally impossible to find a surface that is different from S and projectively deformable to S; some conditions must be satisfied [6, 8].

Let p^{01}, p^{02}, p^{03}, p^{12}, p^{13}, p^{23} be †Plücker coordinate of a straight line in a 3-dimensional projective space P^3. Then we have $p^{01}p^{23}-p^{02}p^{13}+p^{03}p^{12}=0$, and there is a one-to-one correspondence between the ratios of $\{p^{ij}\}$ and straight lines (→ 90 Coordinates B). If the p^{ij} are regarded as homogeneous coordinates of a 5-dimensional projective space P^5, then the previous equation defines a hyperquadric Q in P^5. Thus there is a one-to-one correspondence between points of Q and straight lines in P^3. A curve on Q corresponds to a set of one-parameter families of straight lines, or a ruled surface. Sets of 2-parameter or 3-parameter families of straight lines corresponding to surfaces of 2 or 3 dimensions on Q in P^5 are called **congruences of lines** or **complexes of lines**, respectively. Thus by using a theory of surfaces in P^5, it is possible to establish the theory of congruences and complexes [2, 6, 8], which is an important part of projective differential geometry.

Specifically, if the surface is either a curve or a hypersurface, there are numerous interesting results [2, 4, 6].

C. Affine Differential Geometry

The theme of general affine differential geometry is the study of differential-geometric properties of a point or set of points in a space that are invariant under the action of the †affine transformation group. **Affine differential geometry** is the study of the properties invariant under the action of the †equivalent affine transformation group, i.e., a subgroup of the affine transformation group formed by elements sending (x_i) to $(\bar{x}_i)$ such that

$$\bar{x}_i=a_i+\sum_{j=1}^{n}a_{ij}x_j,\qquad i=1,\ldots,n,\qquad \det(a_{ij})=1.$$

The latter transformation leaves invariant the volume surrounded by an oriented closed hypersurface. The method of moving frames is effective in affine differential geometry.

Let C be a plane curve, and associate with any point $A=A(t)$ of C a family of frames $[A,\mathbf{e}_1,\mathbf{e}_2]$, where the area of the parallelogram

determined by the two vectors $\mathbf{e}_1, \mathbf{e}_2$ is equal to 1. This frame is called the **frame of order 0**. Its differential is expressed by

$$dA=\sum_{s=1}^{2}\omega^s\mathbf{e}_s,\qquad d\mathbf{e}_r=\sum_{s=1}^{2}\omega_r^s\mathbf{e}_s,\qquad r=1,2,$$

$$\omega_1^1+\omega_2^2=0.$$

The frames of order 1, 2, and 3 are characterized by $\omega^2=0$; $\omega^2=0$, $\omega_1^2=\omega^1$; and $\omega^2=0$, $\omega_1^2=\omega^1$, $\omega_1^1=0$, respectively. Then a frame of order 3 can be associated with each point of C and coincides with the Frenet frame. We call $\omega^1=d\sigma$ the **affine arc element**; the **affine curvature** κ is defined by $\omega_2^1=-\kappa\, d\sigma$. Then the Frenet formula is given by

$$dA=d\sigma\mathbf{e}_1,\qquad d\mathbf{e}_1=d\sigma\mathbf{e}_2,\qquad d\mathbf{e}_2=-\kappa\, d\sigma\mathbf{e}_1.$$

With respect to this frame, C is expressed as

$$y=x^2/2+\kappa x^4/8+(d\kappa/d\sigma)x^5/40+\ldots.$$

Further, $d\sigma$ and κ are given analytically by

$$d\sigma=|dA,d^2A|^{1/3},\qquad \kappa=|d^2A/d\sigma^2,d^3A/d\sigma^3|,$$

where $|M,N|=\det(M,N)$. M, N are column vectors with two entries. We call σ the **affine arc length**. The straight line on which $\mathbf{e}_2$ is situated is called the **affine normal**, the diameter of the parabola osculating C at A. If κ is constant, then C is a conic section. Furthermore, C is an ellipse, hyperbola, or parabola according as the constant κ is positive, negative, or zero. In affine geometry, parabolas play a role similar to that played by straight lines in Euclidean geometry.

There are numerous results concerning the theory of skew curves and surfaces [1]. Concerning the theory of skew curves, results on **affine length, affine curvature, affine torsion, affine principal normals**, and **affine binormals** are similar to those in Euclidean geometry. The affine transformation group is situated between the projective transformation group and the congruent transformation group and hence has properties analogous to theirs. The theory of surfaces has a character similar to that of projective differential geometry [1]. We may also consider the variation of the affine area of a surface surrounded by a closed skew curve C. We call the extremal surface the **affine minimal surface**. W. Blaschke and others obtained many results on the global properties of such surfaces.

D. Conformal Differential Geometry

Let S^n be a †conformal space of dimension n, and associate with each point $A_0\in S^n$ a frame $\mathfrak{R}[A_0,A_1,\ldots,A_n,A_\infty]$ of the †$(n+2)$-hyperspherical coordinates with origin A_0 (→ 76 Conformal Geometry). Then denoting by $A\cdot B$ the †inner product of hyperspheres A, B, we obtain

$$A_\alpha\cdot A_\beta=g_{\alpha\beta},\qquad \alpha,\beta=0,1,\ldots,n,\infty,$$

where

$$(g_{\alpha\beta})=\begin{pmatrix}0&0&-1\\0&g_{ij}&0\\-1&0&0\end{pmatrix},$$

$$i,j=1,\ldots,n,\qquad g_{ij}=g_{ji}.$$

Let z^α be homogeneous coordinates with respect to $\mathfrak{R}$. Then the †Möbius transformation $z\to\bar z$ of S^n is characterized by $\bar z^\alpha=c_\beta^\alpha z^\beta$, where $g_{\alpha\beta}c_\sigma^\alpha c_\tau^\beta=g_{\sigma\tau}$, $|c_\beta^\alpha|\neq0$. The differential of the family of the frames is defined by

$$dA_\alpha=\sum_{\beta=0,1}^{n,\infty}\omega_\alpha^\beta A_\beta,\tag{1}$$

where

$$(\omega_\alpha^\beta)=\begin{pmatrix}\omega_0^0&\omega_0^j&0\\ \sum g_{ik}\omega_\infty^k&\omega_i^j&\sum g_{ik}\omega_0^k\\ 0&\omega_\infty&-\omega_0^0\end{pmatrix},$$

$$\sum(g_{ik}\omega_j^k+g_{jk}\omega_i^k)=dg_{ij}.$$

There are $(n+1)(n+2)/2$ linearly independent forms among ω, and this is the number of parameters of the Möbius transformation group. The structure equations of this group are

$$d\omega_\alpha^\beta=\sum\omega_\alpha^\sigma\wedge\omega_\sigma^\beta.\tag{2}$$

The theme of **conformal differential geometry** is the properties of Pfaffian forms ω_α^β satisfying (1) and (2).

Consider a transformation $c:\sum z^\alpha A_\alpha\to\sum z^\alpha(A_\alpha+dA_\alpha)$. (1) If all ω vanish except ω_0^0, then all the circles through A_0, A_∞ are invariant, and any point P is transformed to a neighboring point $\bar P$ on the circle, such that the cross ratio $(P,\bar P;A_0,A_\infty)$ is constant. This transformation is called the **homothety** with centers A_0, A_∞. (2) If all ω vanish except ω_0^j, $\omega_i^\infty=\sum g_{ik}\omega_0^k$, then all the circles tangent to a fixed direction at A_∞ are transformed into themselves, and any hypersphere through A_∞ and orthogonal to those circles is transformed into a hypersphere having the same property. This transformation is called the **elation** with center A_∞. (3) If all ω vanish except $\omega_i^0=\sum g_{ik}\omega_\infty^k$, ω_∞^j, then the transformation is an elation with center A_0. (4) If all ω vanish except ω_i^j, then the transformation is an infinitesimal rotation with center A_0, with A_∞ regarded as a point at infinity. Thus any infinitesimal Möbius transformation is decomposed into the previous four types of transformation.

To study the theory of curves and hypersurfaces in S^n, we again utilize the Frenet frame chosen from a family of frames associated with

A_0. For example, the Frenet formula of a curve in S^3 is given by

$$(\omega_\alpha^\beta)=\begin{pmatrix} 0 & d\sigma & 0 & 0 & 0 \\ \kappa d\sigma & 0 & 0 & 0 & d\sigma \\ -d\sigma & 0 & 0 & \tau d\sigma & 0 \\ 0 & 0 & -\tau d\sigma & 0 & 0 \\ 0 & \kappa d\sigma & -d\sigma & 0 & 0 \end{pmatrix}.$$

We call $d\sigma$, κ, and τ the **conformal arc element**, **conformal curvature**, and **conformal torsion**, respectively. There are many results on conformal deformation [5].

Concerning Laguerre differential geometry, we have results dual to those in conformal differential geometry (a point is replaced by a straight line and an angle by a distance between the points of contact of the common tangents of two oriented circles).

E. Contact Manifolds

Consider a $(2n+1)$-dimensional differentiable manifold M^{2n+1} with a 1-form η such that $\eta\wedge(d\eta)^n\neq 0$, where $d\eta$ is the exterior derivative of η and $\wedge$ denotes exterior multiplication. (Note that this is true for the 1-form in the left-hand side of eq. (2) in 82 Contact Transformations A.) Such a manifold is called a **contact manifold** with **contact form** η. The structure group of the tangent bundle of a contact manifold M^{2n+1} reduces to $U(n)\times 1$, where $U(n)$ is the unitary group; hence every contact manifold is orientable. Simple but typical examples are given by the unit sphere S^{2n+1} in Euclidean space E^{2n+2} and the tangent sphere bundle of an $(n+1)$-dimensional Riemannian manifold M^{n+1}, both with natural contact forms (S. S. Chern [9]). Every 3-dimensional compact orientable differentiable manifold is a contact manifold (J. Martinet [12]).

Now a differentiable manifold M^{2n+1} is said to be an **almost contact manifold** if it admits a tensor field φ of type (1, 1), a vector field ξ, and a 1-form η such that

$$\varphi^2X=-X+\eta(X)\xi, \qquad \eta(\xi)=1, \tag{3}$$

where X is an arbitrary vector field on M^{2n+1}; and the triple (φ,ξ,η) is then called an **almost contact structure**. (3) implies that $\varphi\xi=0$ and $\eta(\varphi X)=0$ (S. Sasaki [14, I]). The structure group of the tangent bundle of an almost contact manifold M^{2n+1} reduces to $U(n)\times 1$. Indeed, J. W. Gray [10] took this property as his definition of almost contact structure. For any pair of vector fields X and Y on M^{2n+1}, let

$$\begin{aligned} N(X,Y)=&[X,Y]+\varphi[\varphi X,Y] \\ &+\varphi[X,\varphi Y]-[\varphi X,\varphi Y] \\ &-\{X\cdot\eta(Y)-Y\cdot\eta(X)\}\xi, \end{aligned}$$

where [,] is the Poisson bracket; then N is a tensor field of type (1, 2) over M^{2n+1}, which we call the **torsion tensor** of the almost contact structure (φ,ξ,η). When N vanishes identically on M^{2n+1}, we say that the almost contact structure is **normal**.

An almost contact structure (φ,ξ,η) on M^{2n+1} induces naturally an almost complex structure J on $M^{2n+1}\times R$ (resp. $M^{2n+1}\times S^1$), which reduces to a complex structure if and only if (φ,ξ,η) is normal. A similar statement is also valid for the product space of two almost contact manifolds (A. Morimoto [13]).

If M^{2n+1} is an almost contact manifold with structure tensor (φ,ξ,η), we can find a positive definite Riemannian metric g so that $g(\varphi X,\varphi Y)=g(X,Y)-\eta(X)\eta(Y)$ for any pair of vector fields X and Y, and the set (φ,ξ,η,g) is then said to be an **almost contact metric structure.**

When M^{2n+1} is a contact manifold with contact form η, there exists a unique vector field ξ which satisfies $d\eta(X,\xi)=0$, $\eta(\xi)=1$ for any vector field X. We can then find a tensor field φ of type (1, 1) and a positive definite metric tensor g so that (i) $d\eta(X,Y)=g(X,\varphi Y)$ is satisfied for any pair of vector fields X and Y, and (ii) (φ,ξ,η,g) is an almost contact metric structure. The almost contact metric structure determined in this way by a contact form η is called a **contact metric structure**. A differentiable manifold with normal contact metric structure is called a **normal contact Riemannian manifold** or a **Sasakian manifold**. Brieskorn manifolds are examples of such manifolds. They include, besides the standard sphere S^{2n+1}, all exotic $(2n+1)$-spheres that bound compact oriented parallelizable manifolds. An almost contact manifold is said to be **regular** or nonregular according as the integral curve of ξ is regular or not as a submanifold. A compact regular contact manifold is a principal circle bundle over a symplectic manifold, and it admits a normal contact metric structure if and only if the base manifold is a Hodge manifold (Boothby and Wang [8], Hatakeyama [11]).

Many research papers on the topology and differential geometry of manifolds with the structures defined above have been published by S. Tanno, S. Tachibana, D. E. Blair, M. Okumura, K. Ogiue, S. I. Goldberg, and others.

References

[1] W. Blaschke, Vorlesungen über Differentialgeometrie, Springer, II, 1923; III, 1929 (Chelsea, 1967).

[2] G. Bol, Projektive Differentialgeometrie I, II, Vandenhoeck & Ruprecht, 1950.
[3] E. Cartan, La théorie des groupes finis et continus, Gauthier-Villars, 1951.
[4] E. Cartan, Leçons sur la théorie des espaces à connexion projective, Gauthier-Villars, 1937.
[5] P. C. Delens, Méthodes et problèmes des géométries différentielles euclidienne et conforme, Gauthier-Villars, 1927.
[6] G. Fubini and E. Čech, Introduction à la géométrie projective différentielle des surfaces, Gauthier-Villars, 1931.
[7] J. Kanitani, Géométrie différentielle projective des hypersurfaces, Mem. Ryojun Coll. of Eng., 1931.
[8] W. M. Boothby and H. C. Wang, On contact manifolds, Ann. Math., (2) 68 (1958), 721–734.
[9] S. S. Chern, Pseudo-groupes continus infinis, Coll. int. du CNRS, Geom. Diff., Strasbourg (1953), 119–135.
[10] J. W. Gray, Some global properties of contact structure, Ann. Math., (2) 69 (1959), 421–450.
[11] Y. Hatakeyama, Some notes on differentiable manifolds with almost contact structure, Tôhoku Math. J., (2) 15 (1963), 176–181.
[12] J. Martinet, Formes de contact sur les variétés de dimension 3, Proc. Liverpool Singularities Symposium II (1971), 142–163.
[13] A. Morimoto, On normal almost contact structures, J. Math. Soc. Japan, 15 (1963), 420–436.
[14] S. Sasaki, On differentiable manifolds with certain structures which are closely related to almost contact structure. I, Tôhoku Math. J., (2) 12 (1960), 459–476. II (with Y. Hatakeyama), ibid., 13 (1961), 281–294.
[15] K. Yano and S. Ishihara, Tangent and cotangent bundles, Dekker, 1973.
[16] D. E. Blair, Contact manifolds in Riemannian geometry, Lecture notes in math. 509, Springer, 1976.

111 (VII.12) Differential Geometry of Curves and Surfaces

A. General Remarks

Let f be an †immersion of an m-dimensional †differentiable manifold M of class C^r into an n-dimensional Euclidean space E^n. More precisely, f is a differentiable mapping of class C^r such that the †differential df_p is injective at every point p of M. The pair (M, f) is called an **immersed submanifold** (or a **surface**) of E^n. When $m=1$, we call it a **curve** of E^n, and when $m=n-1$, a **hypersurface** in E^n. The cases of $n=2$ and $n=3$ have been the main objects of study in differential geometry of curves and surfaces. The differential-geometric properties for the general case of immersion are discussed in 365 Riemannian Submanifolds.

B. Frames in E^n

Every †Euclidean motion in E^n can be expressed as the product of a parallel translation and an †orthogonal transformation that keeps the origin of E^n fixed. The set of all parallel translations is a commutative group that can be identified with $\mathbf{R}^n$. It is a normal subgroup of the group of motions $I(E^n)$ of E^n. So we see that $I(E^n)$ is a †semidirect product of $\mathbf{R}^n$ and the †orthogonal group $O(n)$. The Lie algebra of $I(E^n)$ is the direct sum of $\mathbf{R}^n$ and the Lie algebra $\mathfrak{o}(n)$ of the orthogonal group, where both are regarded as additive groups. Corresponding to this decomposition, we can write the †Maurer-Cartan differential form over $I(E^n)$ as $\omega+\Omega$, where ω belongs to $\mathbf{R}^n$ and Ω to $\mathfrak{o}(n)$. The †structural equation $d(\omega+\Omega)=-(1/2)(\omega+\Omega)\wedge(\omega+\Omega)$ can be divided into the following two parts: $d\omega=\Omega\wedge\omega$; $d\Omega=-(1/2)\Omega\wedge\Omega$. These are known as the **structure equations** of E^n. By an **orthogonal frame** in E^n we mean an ordered set $(x, \mathbf{e}_1, \ldots, \mathbf{e}_n)$ consisting of a point x and a set of †orthonormal vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$. We denote by $\mathcal{O}(n)$ the set of all orthogonal frames in E^n. If we denote the translation identified with $x\in\mathbf{R}^n$ by T_x, then there is a one-to-one correspondence φ: $I(E^n)\to\mathcal{O}(n)$ given by $\varphi(T_xA)=(x, A\mathbf{e}_1, \ldots, A\mathbf{e}_n)$ $(A\in O(n))$. We can make $\mathcal{O}(n)$ into a differentiable manifold so that φ is a †diffeomorphism. We denote the differential forms over $\mathcal{O}(n)$, which are images of ω and Ω under the †dual mapping of φ^{-1}, by the same letters ω and Ω, respectively. For $\mathcal{O}(n)$ as a †principal fiber bundle over $\mathbf{R}^n$ with the projection π: $\pi(x, \mathbf{e}_1, \ldots, \mathbf{e}_n)=x$ and n vector-valued functions φ_i: $\varphi_i(x, \mathbf{e}_1, \ldots, \mathbf{e}_n)=\mathbf{e}_i$ over $\mathcal{O}(n)$, we have

$$\omega=\sum_i \omega^i \mathbf{e}_i, \qquad \Omega=\sum_{i<j}\Omega^{ij}E_{ij},$$

$$\omega^i=(d\pi, \varphi_i), \qquad \Omega^{ij}=(d\varphi_i, \varphi_j), \tag{1}$$

$$d\omega^i=\sum_j \Omega^{ij}\wedge\omega^j, \quad d\Omega^{ij}=\sum_k \Omega^{ik}\wedge\Omega^{kj},$$

where $\{E_{ij}\}$ is a basis of $\mathfrak{o}(n)$ defined by $E_{ij}\mathbf{e}_j=\mathbf{e}_i$, $E_{ij}\mathbf{e}_i=-\mathbf{e}_j$, $E_{ij}\mathbf{e}_k=0$ $(k\neq i,j)$ and (,) is the scalar product of vector-valued forms induced from the scalar product of E^n. Any diffeomorphism of $\mathcal{O}(n)$ onto itself preserving ω and Ω must be a Euclidean motion.

C. Theory of Curves

Let (M,f) be an immersion of a 1-dimensional differentiable manifold M into E^n. We identify the tangent space of E^n at each point with E^n itself. Then df_x maps the origin of the tangent space M_x to $f(x)$, and the image $df_x(M_x)$ of M_x by df_x is a straight line passing through $f(x)$ in E^n, called the **tangent line** of $f(M)$ at $f(x)$. By $\mathcal{O}_f(M)$ we mean the set of all ordered sets $(x, \mathbf{e}_1, \ldots, \mathbf{e}_n)$, where $x \in M$ and $\{\mathbf{e}_i\}$ is an orthonormal basis of E^n such that $\mathbf{e}_1 \in df_x(M_x)$. Then $\mathcal{O}_f(M)$ can be naturally immersed in $\mathcal{O}(n)$ by the mapping $\hat{f}: \hat{f}(x, \mathbf{e}_1, \ldots, \mathbf{e}_n) = (f(x), \mathbf{e}_1, \ldots, \mathbf{e}_n)$. We can pull back the differential forms ω, Ω, ω^i, Ω^{ij} over $\mathcal{O}(n)$ to $\mathcal{O}_f(M)$ $\hat{f}^*$ and denote them by θ, Θ, θ^i, and Θ^{ij}, respectively; then we have $\theta^i = 0$ $(i > 1)$. Let f_1 and f_2 be two immersions of M into E^n. Then in order for there to exist a Euclidean motion α of E^n such that $f_1 = \alpha \circ f_2$, it is necessary and sufficient that there exist a diffeomorphism φ of $\mathcal{O}_{f_1}(M)$ onto $\mathcal{O}_{f_2}(M)$ such that $\theta_{f_1} = \varphi^*(\theta_{f_2})$, $\Theta_{f_1} = \varphi^*(\Theta_{f_2})$. Let π_f be the projection of the fiber bundle $\mathcal{O}_f(M)$, and let φ_i be naturally defined vector-valued functions over $\mathcal{O}_f(M)$. Then we have $d(f \circ \pi_f) = \theta^1 \varphi_1$, $d\varphi_i = \sum_{j=1}^n \Theta^{ij} \varphi_j$. If we put $ds^2(X) = \|df_x(X)\|^2$ $(X \in M_x)$, then we have $(\theta^1)^2 = \pi_f^*(ds^2)$. For each point $x \in M$ there are two possibilities for the choice of $\mathbf{e}_1$ corresponding to two orientations of the curve. But since $(d\varphi_1, d\varphi_1) = \pi_f^*(\rho^2 ds^2)$, ρ^2 depends only on the point x of M. We call ρ $(\geqslant 0)$ the **absolute curvature**. We now choose an orientation of the curve and then $\mathbf{e}_1$ in accordance with the orientation. Thus we get a submanifold of $\mathcal{O}_f(M)$, which we again express by the notation $\mathcal{O}_f(M)$. If we define the form ds by $ds(X) = (df(X), \mathbf{e}_1)$, we have $\theta^1 = \pi_f^*(ds)$, and ds is called the **line element**. Any †local cross section $R: \pi_f \circ R = 1$ of the bundle $\mathcal{O}_f(M)$ is called a **moving frame**. Putting

$$R^*(\theta^1) = ds, \qquad R^*(\Theta^{ij}) = \rho^{ij} ds,$$

we see that the following equation holds over M:

$$d\mathbf{e}_i = \sum_j \rho^{ij} ds\, \mathbf{e}_j. \tag{2}$$

For two immersions (M, f_1) and (M, f_2), we have $f_1 = \alpha \circ f_2$ (α is a Euclidean motion) if and only if they have the same ds and ρ^{ij} for some moving frames.

D. Frenet's Formulas

In order to study local properties of curves it is sufficient to consider them on †Jordan arcs of class C^r. With respect to orthogonal coordinates $(x^1, \ldots, x^n)$ in E^n, such a curve is represented parametrically by $x^i = f^i(t)$ $(t \in [a,b]$, $\sum (dx^i/dt)^2 > 0)$ or by a vector representation $\mathbf{x} = \mathbf{x}(t)$. If φ is a diffeomorphism of a closed interval $[a', b']$ onto $[a,b]$, then $f \circ \varphi$ and f are representations of the same arc in E^n, and φ is called a **transformation of the parameter**. Any curve of class C^1 is †rectifiable, and its **arc length** is given by $s = \int_a^b (\sum_{i=1}^n (dx^i/dt)^2)^{1/2} dt$. We may choose the arc length s measured from a point on the arc as a parameter, called the **canonical parameter** of the arc. Consider an arc C of class C^n given by the vector representation $\mathbf{x} = \mathbf{x}(s)$, $s \in [a,b]$. We assume that its †Wronskian $|\mathbf{x}'(s), \ldots, \mathbf{x}^{(n)}(s)|$ is not identically zero (we denote by $'$ the derivative with respect to the canonical parameter), which means that the arc C is not contained in a hyperplane in E^n. A point at which the Wronskian vanishes is called a **stationary point**, and we assume that there exists no stationary point on C. By the †Gram-Schmidt orthonormalizing process we obtain an orthonormal basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ $(|\mathbf{e}_1, \ldots, \mathbf{e}_n| > 0)$ from n vectors $\mathbf{x}'(s), \ldots, \mathbf{x}^{(n)}(s)$ at each point of C. We call the frame thus determined the **Frenet frame**. With respect to the Frenet frame, (2) is rewritten as

$$\begin{aligned} &\mathbf{e}_i'(s) = -\kappa_{i-1}(s)\mathbf{e}_{i-1}(s) + \kappa_i(s)\mathbf{e}_{i+1}(s), && i = 1, \ldots, n; \\ &\kappa_0(s) = \kappa_n(s) = 0; && \\ &\kappa_j(s) > 0, && j = 1, \ldots, n-2. \end{aligned} \tag{3}$$

These are called **Frenet's formulas** (or the **Frenet-Serret formulas**). We call $\kappa_1, \kappa_2, \ldots, \kappa_{n-2}$ the first, second, ..., $(n-2)$nd **curvature**, respectively, while we call κ_{n-1} the **torsion** for $n \geqslant 3$. For a curve in a lower-dimensional subspace $E^m \subset E^n$, we set $\kappa_i = 0$ $(i > m)$. The curvatures and the torsion of a straight line are zero. To get Frenet's formulas in these special cases, we fix $\mathbf{e}_i$ $(i > m)$ in the subspace orthogonally complementary to E^m in E^n and proceed as in the general case. Suppose that C_1, C_2 are arcs such that both of their Frenet frames are of class C^1. If there exists a diffeomorphism of C_1 to C_2 that preserves arc length and the κ_i $(i = 1, \ldots, n-1)$ are equal at corresponding points, C_1 and C_2 are mapped onto each other by a motion of E^n. This is the **fundamental theorem of the theory of curves**. Given $n-1$ functions of class C^1 $\kappa_1(s) \geqslant 0, \ldots, \kappa_{n-2}(s) \geqslant 0$ (we assume that the equality signs occur at most at a finite number of points) and $\kappa_{n-1}(s)$ for $0 \leqslant s \leqslant L$, there exists an arc that has $\kappa_1, \ldots, \kappa_{n-2}, \kappa_{n-1}$ as its first, ..., $(n-2)$nd curvatures and its torsion, respectively. The equations $\kappa_i = \kappa_i(s)$ are called the **natural equations** of the curve.

E. Plane Curves

Let $\mathbf{x} = \mathbf{x}(s)$ be a curve of class C^2 in E^2, and $(\mathbf{x}(s), \mathbf{e}_1, \mathbf{e}_2)$ its Frenet frame. The tangent and

the normal of this curve at $\mathbf{x}(s)$ have parametric representations $\mathbf{x}(s)+t\mathbf{e}_1$, $\mathbf{x}(s)+t\mathbf{e}_2$, respectively (with parameter t). Frenet's formulas are written as $\mathbf{x}'=\mathbf{e}_1$, $\mathbf{x}''=\mathbf{e}_1'=\kappa\mathbf{e}_2$, and ρ is called the **curvature** of the curve C. The natural equation is given by $\kappa=\kappa(s)$. If $\kappa(s)=$ constant $\neq 0$ along C, C must be a portion of a circle. Another way of defining the curvature is as follows: We take a fixed direction (for example, the positive direction of the x-axis on E^2) and denote by $\theta(s)$ the angle made by the tangent T_s of the curve C at $\mathbf{x}(s)$ with the direction. Then we have $\kappa(s)=d\theta/ds$. If $n=2$, the curvature can take both positive and negative values. Figs. 1 and 2 suggest a geometric meaning of $\kappa>0$ and $\kappa<0$, respectively. The circle with center $\mathbf{x}=\mathbf{x}(s)+(1/\kappa)\mathbf{e}_2$ and radius $1/\kappa$ has a contact of higher order than any other circle in E^2. We call this circle the **osculating circle** (or **circle of curvature**) at the point $\mathbf{x}=\mathbf{x}(s)$, its center **the center of curvature**, and $1/\kappa$ the **radius of curvature**. The locus C' of the center of curvature of a curve C is called an **evolute** of C. Conversely, C is called an **involute** of C', the †envelope of the family of normal lines of C. When a curve is given in terms of its canonical parameter s, the curvature is given by $|\mathbf{x}'(s),\mathbf{x}''(s)|$; when the curve is given by another parameter t as $\mathbf{x}=\mathbf{x}(t)$, the curvature is given by $\kappa(t)=|\mathbf{x}'(t),\mathbf{x}''(t)|/|\mathbf{x}'(t)|^3$, where $'$ means d/dt.

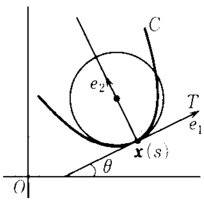


Fig. 1
$\kappa>0$.

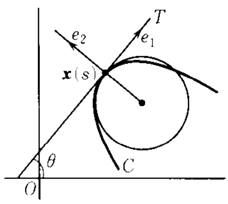


Fig. 2
$\kappa<0$.

The facts we have just stated concern local properties of plane curves. We shall now discuss the global theory of curves, which deals with properties of each curve as a whole. Let $C:\mathbf{x}=\mathbf{x}(s)$, $a\leqslant s\leqslant b$, be a closed curve. Let $\theta(s)$, $0\leqslant\theta(s)<2\pi$, be the angle that $\mathbf{e}(s)$ makes with the x axis. Put $\Theta=\int_a^b\theta'(s)ds$. Intuitively, Θ measures the total rotation of $\mathbf{e}_1(s)$ as we run along the curve C from a to b. Since C is closed, Θ is an integer multiple I of 2π. The integer I is called the **rotation number** of C, and is equal to $(1/2\pi)\int_a^b\kappa(s)ds$. Let $\bar{D}$ be a closed domain consisting of points in the interior and on the boundary of a simple closed curve C. C is called a **closed convex curve** or an **oval** if $\bar{D}$ is †convex in E^2. Among all ovals of given length, the circle has the maximum area. Various generalizations of this theorem have been obtained, and the collection of problems of this kind is called the **isoperimetric problem**. This problem has intimate connections with fields such as integral geometry. The oval has a convenient parameter other than the arc length parameter s. Given a number t, $0\leqslant t\leqslant 2\pi$, there exists a unique point $\mathbf{x}(t)$ in the oval such that $\mathbf{e}_2=(\cos t,\sin t)$ at $\mathbf{x}(t)$. When we describe the oval in terms of the parameter t, the tangent vector at $\mathbf{x}(t)$ is parallel to that at $\mathbf{x}(t+\pi)$, and we can define the width $W(t)$ at $\mathbf{x}(t)$. $W(t)$ is called the **width of the oval**. A curve is called a **curve of constant width** if the curve is an oval whose width $W(t)$ does not depend on t. The circle is a typical example of a curve of constant width. Reuleaux's triangle is another well-known example of a curve of constant width (Fig. 3). For a curve of constant width of width W and length L, we have $L=\pi W$.

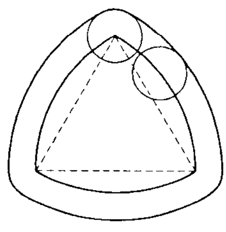


Fig. 3

There are also some results concerning the relations between local properties (for example, curvature) and properties of the whole figure. An example is given by the **four-vertex theorem**. A vertex on a curve C is by definition a point where $d\kappa/ds=0$. Then there are at least four vertices on an oval of class C^3. A simple closed curve with $\kappa\geqslant 0$ ($\leqslant 0$) must be convex (→ 89 Convex Sets).

F. Space Curves

Let $\mathbf{x}=\mathbf{x}(s)$ ($s\in[a,b]$) be a curve C of class C^3 in E^3 defined in terms of the canonical parameter s. Let $(\mathbf{x}(s),\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3)$ be Frenet frames along C. Then we have the Frenet formulas

$$\mathbf{e}_1'=\kappa_1\mathbf{e}_2,\quad \mathbf{e}_2'=-\kappa_1\mathbf{e}_1+\kappa_2\mathbf{e}_3,\quad \mathbf{e}_3'=-\kappa_2\mathbf{e}_2.$$

We call $1/\kappa_1$, $1/\kappa_2$ the **radius of curvature** and the **radius of torsion**, respectively. The line $\mathbf{x}=\mathbf{x}(s_0)+t\mathbf{e}_1$ is the tangent of C at $\mathbf{x}(s_0)$. The two straight lines through the point $\mathbf{x}(s_0)$ defined by $\bar{\mathbf{x}}=\mathbf{x}(s_0)+t\mathbf{e}_2$ and $\bar{\mathbf{x}}=\mathbf{x}(s_0)+t\mathbf{e}_3$ are called the **principal normal** and the **binormal** of C at $\mathbf{x}(s_0)$, respectively. The three planes through $\mathbf{x}(s_0)$ defined by $\bar{\mathbf{x}}=\mathbf{x}(s_0)+t\mathbf{e}_2+\bar{t}\mathbf{e}_3$, $\bar{\mathbf{x}}=\mathbf{x}(s_0)+t\mathbf{e}_3+\bar{t}\mathbf{e}_1$, and $\bar{\mathbf{x}}=\mathbf{x}(s_0)+t\mathbf{e}_1+\bar{t}\mathbf{e}_2$ are called the **normal plane**, the **rectifying plane**, and the **osculating plane**, respectively.

At a point $\mathbf{x}(s_0)$ of a curve $\mathbf{x}=\mathbf{x}(s)$ of class C^ω, we take $\mathbf{e}_1(s_0)$, $\mathbf{e}_2(s_0)$, and $\mathbf{e}_3(s_0)$ as unit vectors of the coordinate axes. Substituting the Frenet formulas into the Taylor expansion of $\mathbf{x}(s)$, we see that the new coordinate $\mathbf{x}_1(s)$, $\mathbf{x}_2(s)$, $\mathbf{x}_3(s)$ of C are given by

$$x_1=(s-s_0)-(\kappa_1(s_0)/6)(s-s_0)^3+\ldots,$$

$$x_2=(\kappa_1(s_0)/2)(s-s_0)^2+(\kappa_1'(s_0)/6)(s-s_0)^3+\ldots,$$

$$x_3=(\kappa_1(s_0)\kappa_2(s_0)/6)(s-s_0)^3+\ldots.$$

These are called **Bouquet's formulas**. Utilizing these formulas we can see the nature of the curve with given κ_1 and κ_2. A curve and its osculating plane at a point on it have contact of order higher than any other plane through that point. The family of osculating planes of C †envelops a †developable surface S and coincides with the locus of tangent lines to C. We call S the **tangent surface** of C, and C the **line of regression** of S. The family of rectifying planes of C also envelops a developable surface called the **rectifying surface**, and C is a †geodesic on this surface. The family of normal planes of C envelops either a cone or a tangent surface of another curve $\bar{C}$. When the natural equation of a space curve has a special form, the shape of the curve is simple. For example, $\kappa_1(s)=$constant, $\kappa_2(s)=$constant represent a curve, called an **ordinary helix**, on a cylinder which cuts all the generators of the cylinder at a constant angle. More generally, it is known that if $\kappa_1/\kappa_2=$constant, the tangent at each point of the curve makes a constant angle with a fixed direction. Such a curve is called a **generalized helix** or a **curve of constant inclination**. Each curve satisfying $a\kappa_1+b\kappa_2=c$ $(ab\neq 0)$ is called a **Bertrand curve**. For a Bertrand curve there exists another curve $\bar{C}$ and a correspondence of C onto $\bar{C}$ such that they have a common principal normal at corresponding points. Conversely, this property is also a sufficient condition for C to be a Bertrand curve. A **Mannheim curve** is defined analogously as a curve having a correspondence with another curve $\bar{C}$ such that the principal normal of C and the binormal of $\bar{C}$ coincide at corresponding points. When a correspondence of C and $\bar{C}$ has the property that tangents at corresponding points are parallel, then the correspondence is called a **correspondence of Combescure**.

We have stated mainly local properties of space curves. There are also several results about global properties of curves in E^3 analogous to the case of plane curves. For a simple closed curve C of length L, we call $K=\int_0^l \kappa_1(s)ds$ the **total curvature** of C. Generally we have $K\leqslant 2\pi$, while $K=2\pi$ if and only if C is a closed convex curve lying in a plane (W. Fenchel) [5, 6]. The total curvature is deeply related to the properties of †knots. If a simple closed curve in E^3 is knotted, then the total curvature is at least 4π [7, 8]. We fix an origin O in E^3 and draw a unit tangent vector with initial point O parallel to the unit tangent vector at each point of a space curve C; then the endpoint of this vector traces a curve $\bar{C}$ on the unit sphere with the center O. We call $\bar{C}$ the **spherical indicatrix** of C and the correspondence of C to $\bar{C}$ a **spherical representation**. The total curvature K of a curve C is equal to the length of $\bar{C}$. Consequently, we have $K=\oint_C d\theta$, where θ is the angular deflection of the tangent line along the closed curve C.

G. Theory of Hypersurfaces

Let (M,f) be an immersion of an $(n-1)$-dimensional differentiable manifold M of class C^r into E^n. Then we can define on the hypersurface M a positive definite differential form g of degree 2 induced from the inner product of E^n: $g_x(X,X)=(df_x(X),df_x(X))$, $X\in M_x$. Then M becomes a †Riemannian manifold with †Riemannian metric g. We call g the **first fundamental form** of (M,f). The Riemannian geometry on a surface with its first fundamental form as Riemannian metric is called **geometry on a surface** (→ 364 Riemannian Manifolds).

By $\mathcal{O}_f(M)$ we mean the set of all the ordered sets $(x,\mathbf{e}_1,\ldots,\mathbf{e}_n)\in\mathcal{O}(n)$, where $x\in M$ and $\{\mathbf{e}_i\}$ $(i=1,2,\ldots,n)$ is an orthonormal system of E^n such that $\mathbf{e}_i\in df_x(M_x)$ $(i=1,\ldots,n-1)$. Then $\mathcal{O}_f(M)$ with natural projection π_f and natural differentiable structure is a †principal fiber bundle over M and has a natural immersion $\hat{f}:\hat{f}(x,\mathbf{e}_1,\ldots,\mathbf{e}_n)=(f(x),\mathbf{e}_1,\ldots,\mathbf{e}_n)$ in the principal fiber bundle $\mathcal{O}(n)$. We can †pull back the forms on $\mathcal{O}(n)$ to $\mathcal{O}_f(M)$ by $\hat{f}^*$ and put $\theta=\hat{f}^*(\omega)$, $\Theta=\hat{f}^*(\Omega)$; then the structural equations of E^n are transformed to $d\theta=\Theta\wedge\theta$, $d\Theta=(-1/2)\Theta\wedge\Theta$. Furthermore, if we put $\theta^i=\hat{f}^*(\omega^i)$, $\Theta^{ij}=\hat{f}^*(\Omega^{ij})$, then $\theta^n=0$ and θ^i, $\Theta^{ij}(i,j<n)$ depend only on the first fundamental form of (M,f). Let f_1 and f_2 be two immersions of M into E^n. Then in order that there exist a Euclidean motion α of E^n such that $f_1=\alpha\circ f_2$, it is necessary and sufficient that there exist a diffeomorphism φ of $\mathcal{O}_{f_1}(M)$ onto $\mathcal{O}_{f_2}(M)$ such that $\theta_{f_1}=\varphi^*(\theta_{f_2})$ and $\Theta_{f_1}=\varphi^*(\Theta_{f_2})$. Suppose that M is †orientable and oriented. Then the unit vector field normal to $df_x(M_x)$ at every point $x\in M$ in E^n defines a mapping of M into the unit sphere in E^n called the **spherical representation** of M or the **Gauss mapping (Gauss map)**. Regarding the unit normal vector field ξ of M as a vector-valued function over M, we can define a symmetric product of df and $d\xi$ by $-(df,d\xi)(X,Y)=$

$(1/2)[(df(X), d\xi(Y)) + (df(Y), d\xi(X))]$, called the **second fundamental form** of (M, f).

Two immersions f_1 and f_2 of M that induce the same first and second fundamental forms have a Euclidean motion α such that $f_1 = \alpha \circ f_2$; and the converse is also true. This fact is called the **fundamental theorem of the theory of surfaces**.

H. Theory of Surfaces in E^3 (→ 365 Riemannian Submanifolds; Appendix A, Table 4.I)

A surface in E^n is locally expressed by parametric equations $x_i = x_i(u_\alpha)$ $(i = 1, \ldots, n;\ \alpha = 1, \ldots, m)$ or by a single vector equation $\mathbf{x} = \mathbf{x}(u_\alpha)$. We are mainly concerned with the case $n = 3$, $m = 2$, and we express the surface by a vector representation $\mathbf{x} = \mathbf{x}(u, v)$. The first and second fundamental forms are written as

$$E\,du^2 + 2F\,du\,dv + G\,dv^2,$$

$$L\,du^2 + 2M\,du\,dv + N\,dv^2.$$

If we use the usual notation of †tensor analysis, then the first and second fundamental forms are also denoted by $g_{\alpha\beta}\,du^\alpha du^\beta$ and $H_{\alpha\beta}\,du^\alpha du^\beta$, respectively, where u^α $(\alpha = 1, 2)$ are parameters (with Σ omitted by †Einstein's convention). We call $\{g_{\alpha\beta}\}$, $\{H_{\alpha\beta}\}$ the **first** and **second fundamental quantities**, respectively (→ 365 Riemannian Submanifolds). At the point $p_0 = x(u_0, v_0)$ on a surface S that corresponds to parameter values (u_0, v_0), the curves expressed by $v = v_0$ and $u = u_0$ are called a **u-curve** and a **v-curve** through p_0, respectively. Let $\mathbf{x}_u$, $\mathbf{x}_v$ denote the tangent vectors $\partial\mathbf{x}/\partial u$, $\partial\mathbf{x}/\partial v$ at p_0 to the u-curve and v-curve, respectively, through the given point p_0 and ξ denote the unit vector orthogonal to $\mathbf{x}_u$ and $\mathbf{x}_v$. Then ξ is called the **normal vector** of S at p_0 and $(\mathbf{x}_u, \mathbf{x}_v, \xi)$ the **Gaussian frame** of S at p_0. Although a Gaussian frame is not in general an orthogonal frame, it is intimately related to local parameters. The plane that passes through the point p_0 and is spanned by $\mathbf{x}_u$, $\mathbf{x}_v$ is called the **tangent plane** to S at p_0. The coefficients of the second fundamental form $L(u, v)$, $M(u, v)$, $N(u, v)$ are expressed by the inner products $L = (-\mathbf{x}_u, \xi_u)$, $M = (-\mathbf{x}_u, \xi_v)$, $N = (-\mathbf{x}_v, \xi_v)$ $(\xi_u = \partial\xi/\partial u,\ \xi_v = \partial\xi/\partial v)$.

Let (X, Y) be the coordinates of a point on the tangent space at p_0 with respect to the Gaussian frame. We call the curve of the second order defined by $LX^2 + 2MXY + NY^2 = \varepsilon$ (ε is a suitable constant) the **Dupin indicatrix**. The point p_0 is called an **elliptic point** or a **hyperbolic point** on S according as the Dupin indicatrix at the point is an ellipse or a hyperbola. If p_0 is an elliptic point, then points near p_0 on the surface lie on one side of the tangent plane at p_0, whereas if p_0 is a hyperbolic point, points near p_0 on the surface lie on both sides of the tangent plane at p_0 (Figs. 4, 5). A hyperbolic point is also called a **saddle point**, since in a neighborhood of the point the surface looks like a saddle. A point that is neither elliptic nor hyperbolic is called a **parabolic point**; at a parabolic point we have $LN - M^2 = 0$. If at least one of L, M, N does not vanish at p_0, then there is a neighborhood of p_0 of the surface that lies on one side of the tangent plane at p_0 (Fig. 6). If a vector (X, Y) on the tangent plane of the surface at p_0 satisfies the equation $LX^2 + 2MXY + NY^2 = 0$, then the direction of the vector is called an **asymptotic direction**. If the point p_0 is elliptic, such a direction does not exist; if p_0 is hyperbolic, the direction is an †asymptotic direction of the Dupin indicatrix on the tangent plane at p_0. A curve C on a surface such that the tangent line at each point of the curve coincides with an asymptotic direction of the surface at the point is called an **asymptotic curve**.

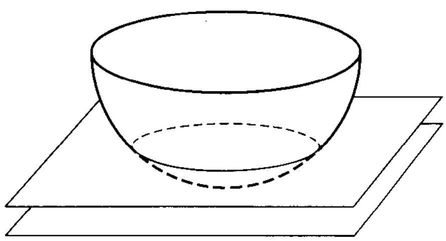


Fig. 4
Elliptic point.

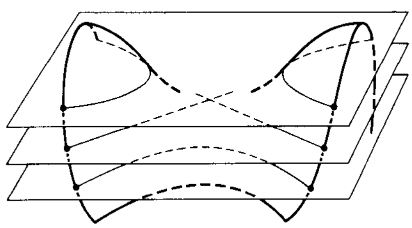


Fig. 5
Hyperbolic point.

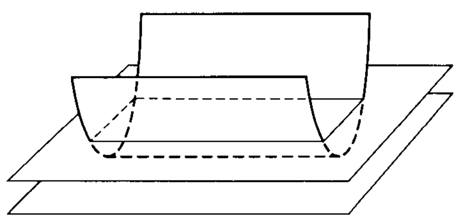


Fig. 6
Parabolic point.

Let $C: \mathbf{x} = \mathbf{x}(u(t), v(t))$ be a curve through p_0 on the surface $\mathbf{x} = \mathbf{x}(u, v)$. Then the curvature κ of C as a space curve is given by

$$\kappa\cos\theta = \frac{L\,du^2 + 2M\,du\,dv + N\,dv^2}{E\,du^2 + 2F\,du\,dv + G\,dv^2},$$

where $du\,dv$ is the direction of C on the surface at p_0 and θ is the angle between the normal of the surface at p_0 and the principal normal of C at p_0. The center of curvature at a point p_0 of a curve C of class C^2 on a surface of class C^2 is the projection on its osculating plane of the center of curvature of the section C^* (of the surface) cut by the plane determined by the tangent to the curve at p_0 and the normal of the surface at the point (**Meusnier's theorem**). The curvature of the curve C^* at p_0 is called the **normal curvature** of the surface at the point for the tangent direction. Since the normal curvature for a direction at a point is a continuous function of this direction that can be represented as a point on a unit circle, there exist two directions that realize the maximum and minimum of the normal curvature. These directions are given by the equation

$$\begin{vmatrix} E\,du+F\,dv & F\,du+G\,dv \\ L\,du+M\,dv & M\,du+N\,dv \end{vmatrix}=0.$$

When this quadratic equation in du/dv has nonzero discriminant, it determines two directions defined by its two roots. These directions are called **principal directions** at the point. A curve C on a surface such that the tangent line at each point of the curve coincides with a principal direction at the point is called a **line of curvature**. When all lines of curvature of a surface are circles, the surface is called a **cyclide of Dupin**. The two normal curvatures corresponding to two principal directions are given by $1/R$, satisfying the following second-order equation:

$$\left(\frac{1}{R}\right)^2-\frac{EN+GL-2FM}{EG-F^2}\frac{1}{R}+\frac{LR-M^2}{EG-F^2}=0.$$

They are called **principal curvatures**, and each of their inverses is called a **radius of principal curvature**. The mean value $H=(\kappa_1+\kappa_2)/2$ of two principal curvatures $\kappa_i=1/R_i$ $(i=1,2)$ is called the **mean curvature** (or **Germain's curvature**), and the product $K=\kappa_1\kappa_2$ is called the **total curvature** (or **Gaussian curvature**). These are given by

$$H=\frac{1}{2}\frac{EN+GL-2FM}{EG-F^2},\qquad K=\frac{LN-M^2}{EG-F^2}.$$

A point on a surface is elliptic, hyperbolic, or parabolic according as $K>0$, $K<0$, or $K=0$ at the point. A point where the second fundamental form is proportional to the first fundamental form is called an **umbilical point**, and a point where the second fundamental form vanishes is called a **flat point** or a **geodesic point**. If a surface consists of umbilical points only, the ratio $(L(du)^2+2M\,du\,dv+N(dv)^2)/(E(du)^2+2F\,du\,dv+G(du)^2)$ is a constant, and the surface is either a sphere or a portion of it. If a surface consists of flat points only, the surface must be either a plane or a portion of it. The mean curvature and the Gaussian curvature of a sphere are constant, and those of a plane are both equal to zero. If we use the spherical representation of a surface stated in Section G, we can give to the Gaussian curvature the following geometric meaning: Let A be the area of the domain enclosed by a closed curve C around a point p_0 on a surface, and let A^* be the area of the domain on the unit sphere enclosed by the curve that is the image of C under the spherical representation of the surface. Then the limit of A^*/A as the closed curve C tends to the point p_0 is equal to K at p_0.

Let us denote by $(g^{\alpha\beta})$ the inverse matrix of the matrix $(g_{\alpha\beta})$ whose elements are coefficients of the first fundamental form $g_{\alpha\beta}du^\alpha du^\beta$. We easily see that $g^{11}=G/(EG-F^2)$, $g^{12}=g^{21}=-F/(EG-F^2)$, $g^{22}=E/(EG-F^2)$. We introduce the symbols

$$[\beta\gamma,\alpha]=\frac{1}{2}\left(\frac{\partial g_{\alpha\beta}}{\partial u^\gamma}+\frac{\partial g_{\gamma\alpha}}{\partial u^\beta}-\frac{\partial g_{\beta\gamma}}{\partial u^\alpha}\right),$$

$$\begin{Bmatrix}\alpha\\ \beta\gamma\end{Bmatrix}=g^{\alpha\delta}[\beta\gamma,\delta],$$

which are called the **Christoffel symbols** of the first and second kinds, respectively. Suppose that a surface is given by the vector representation $\mathbf{x}=\mathbf{x}(u_1,u_2)$, and put $\mathbf{x}_\alpha=\partial\mathbf{x}/\partial u^\alpha$, $\mathbf{x}_{\alpha\beta}=\partial\mathbf{x}_\alpha/\partial u^\beta$. Then for the derivatives of the Gaussian frame, we obtain

$$\mathbf{x}_{\alpha\beta}=\begin{Bmatrix}\gamma\\ \alpha\beta\end{Bmatrix}\mathbf{x}_\gamma+H_{\alpha\beta}\xi,\qquad \xi_\alpha=-g^{\gamma\beta}H_{\beta\alpha}\mathbf{x}_\gamma.$$

We call the former **Gauss's formula** and the latter **Weingarten's formula**. The integrability conditions of these partial differential equations are

$$R^\alpha{}_{\beta\gamma\delta}=H_{\delta\beta}H_\gamma{}^\alpha-H_{\gamma\beta}H_\delta{}^\alpha,\qquad H_\gamma{}^\alpha=g^{\alpha\sigma}H_{\gamma\sigma},$$

$$\frac{\partial H_{\alpha\beta}}{\partial u^\gamma}-\frac{\partial H_{\alpha\gamma}}{\partial u^\beta}+\begin{Bmatrix}\sigma\\ \alpha\beta\end{Bmatrix}H_{\sigma\gamma}-\begin{Bmatrix}\sigma\\ \alpha\gamma\end{Bmatrix}H_{\sigma\beta}=0,$$

where

$$R^\alpha{}_{\beta\gamma\delta}=\frac{\partial}{\partial u^\gamma}\begin{Bmatrix}\alpha\\ \delta\beta\end{Bmatrix}-\frac{\partial}{\partial u^\delta}\begin{Bmatrix}\alpha\\ \gamma\beta\end{Bmatrix}+\begin{Bmatrix}\sigma\\ \delta\beta\end{Bmatrix}\begin{Bmatrix}\alpha\\ \gamma\sigma\end{Bmatrix}-\begin{Bmatrix}\sigma\\ \gamma\beta\end{Bmatrix}\begin{Bmatrix}\alpha\\ \delta\sigma\end{Bmatrix}$$

are components of the curvature tensor. The former are called the **Gauss equations**, and the latter the **Codazzi-Mainardi equations**. In connection with these equations, **Bonnet's fundamental theorem** states the following: Suppose that a positive definite symmetric matrix $(g_{\alpha\beta})$ and a symmetric matrix $(H_{\alpha\beta})$ are given that are functions of class C^2 and C^1, respectively, defined over a †simply connected domain D in $\mathbf{R}^2$. If they satisfy the Gauss

equations and the Codazzi-Mainardi equations, then there exists a surface $\mathbf{x}=\mathbf{x}(u_1, u_2)$ with the given $(g_{\alpha\beta})$ and $(H_{\alpha\beta})$ as coefficients of its first and second fundamental forms, respectively. Such a surface is determined uniquely if, for an arbitrary fixed point (u_1^0, u_2^0) of D, we assign an arbitrary point p_0 and a frame $(\mathbf{x}_1^0, \mathbf{x}_2^0, \xi^0)$ at p_0 so that $\mathbf{x}_1^0, \mathbf{x}_2^0$ are orthogonal to the unit vector ξ^0 and $(\mathbf{x}_\alpha^0, \mathbf{x}_\beta^0) = g_{\alpha\beta}(u_1^0, u_2^0)$ as the Gaussian frame at p_0. On the other hand, we can investigate surfaces as Riemannian manifolds defined by the first fundamental form.

A †diffeomorphism between two surfaces preserving arc length is called an **isometric mapping**. The condition of preserving arc length is equivalent to the condition that the first fundamental quantities of the surfaces coincide at each pair of corresponding points, provided that we have introduced parameters on the two surfaces so that corresponding points have the same parameter values. In such a case two surfaces are said to be **isometric**. From the Gauss equation we can see that the total curvature depends only on the first fundamental quantities. So K is a quantity that is preserved under isometric mappings (**Gauss's theorema egregium**).

A vector field $\lambda^\alpha(t)\mathbf{x}_\alpha$ defined along a curve $u^\alpha = u^\alpha(t)$ on a surface is said to be **parallel in the sense of Levi-Civita** along the curve if its †covariant derivative along the curve vanishes, i.e., if

$$\delta\lambda^\alpha/dt = d\lambda^\alpha/dt + \left\{ \begin{matrix} \alpha \\ \beta\gamma \end{matrix} \right\} \lambda^\beta du^\gamma/dt = 0.$$

The length of a vector belonging to a vector field that is parallel along a curve C is constant along C. The angle of two vectors both belonging to vector fields that are parallel along C is also constant along C. Choose two vector fields $\lambda^\alpha_{(a)}$ parallel along a curve C on a surface that satisfy $g_{\alpha\beta}\lambda^\alpha_{(a)}\lambda^\beta_{(b)} = \delta_{ab}$. Then the tangent vector to C is expressed by $du^\alpha/dt = \lambda^\alpha_{(a)}v^a(t)$. Take a 2-plane and fix an orthogonal coordinate system on it; then the integral curve C of a set of ordinary differential equations $dx^\alpha/dt = C^\alpha_{(a)}v^a(t)$ $(C^\alpha_a = \lambda^\alpha_{(a)}(P_0))$ is called the **development** of C (→ 80 Connections). We denote by κ the curvature of a curve C of class C^2 on a surface S of class C^2 and by σ the angle between the binormal of C and the normal of S at the same point. Then the quantity $\kappa_g = \kappa\cos\sigma$, belonging to the geometry on the surface, is called the **geodesic curvature** of the curve at the point. A curve with vanishing geodesic curvature is called a **geodesic**. It satisfies the differential equations

$$\frac{d^2u^\alpha}{ds^2} + \left\{ \begin{matrix} \alpha \\ \beta\gamma \end{matrix} \right\} \frac{du^\beta}{ds}\frac{du^\gamma}{ds} = 0.$$

The development of a geodesic is a straight line (→ 364 Riemannian Manifolds).

Let us consider a simply connected, orientable bounded domain D on a surface such that the boundary of D is a simple closed curve C that consists of a finite number of arcs of class C^2. If we denote by α_i $(i=1,2,\ldots,m)$ the external angles at vertices of the curvilinear polygon C (Fig. 7), we have

$$\int_C \kappa_g ds + \sum_{i=1}^m \alpha_i + \iint_D K\, d\sigma = 2\pi.$$

This is called the **Gauss-Bonnet formula**. In particular, if all the arcs of C are geodesics, we have

$$\sum_{i=1}^m \alpha_i + \iint_D K\, d\sigma = 2\pi.$$

This formula implies as special cases the following well-known theorems in Euclidean geometry and spherical trigonometry: (i) The sum of interior angles of a triangle is equal to π. (ii) The area of a spherical triangle is proportional to its spherical excess. The formula also implies the following theorem: On any closed orientable surface we have $\iint K\, d\sigma = 2\pi\chi$, where χ is the †Euler characteristic of the surface. We call $\iint K\, d\sigma$ the **integral curvature** (or **total Gaussian curvature**).

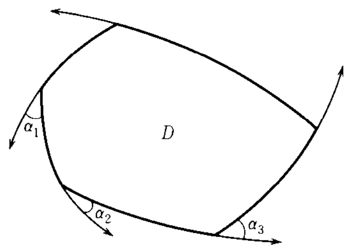


Fig. 7

I. Special Surfaces in E^3

A surface is called a **surface of revolution** if it is generated by a curve C on a plane π when π is rotated around a straight line l in π. Then l and C are called an **axis of rotation** and a **generating curve**, respectively. A surface of revolution having the x_3-axis as the axis of rotation is given by the equations $x_1 = r\cos\theta$, $x_2 = r\sin\theta$, $x_3 = \varphi(r)$; its first fundamental form is $(1+\varphi'^2)dr^2 + r^2 d\theta^2$. The section of a surface of revolution by a half-plane through its axis of rotation is called a **meridian**. According as the meridian is a straight line parallel to the axis of rotation or a straight line intersecting the axis nonorthogonally, the surface of revolution is called a **circular cylinder** or a **circular cone**, respectively. If the

meridian is a circle that does not intersect the axis of rotation, it is called a **torus**.

A surface of class C^2 whose mean curvature H vanishes everywhere is called a **minimal surface** (→ 275 Minimal Submanifolds). A surface of class C^1 realizing a relative minimum of areas among all surfaces of class C^1 with a given closed curve as their boundaries is an †analytic surface such that $H=0$. Conversely, a surface of class C^2 with vanishing mean curvature is an analytic surface. The equation of a surface of revolution with a catenary as its generating line is given by $x_1^2+x_2^2=a(e^{x_3/a}+e^{-x_3/a})/2$. This surface is called a **catenoid** and is a minimal surface. Conversely, a minimal surface of revolution is necessarily a catenoid. For a surface obtained by rotating a †Delaunay curve around its base line, the mean curvature H is equal to a constant ($\neq 0$). Conversely, a surface of revolution with nonzero constant mean curvature must be such a surface. A surface with constant Gaussian curvature is called a **surface of constant curvature**, and is a 2-dimensional Riemannian space of constant curvature (→ 364 Riemannian Manifolds). A non-Euclidean plane can be represented locally as a surface of constant curvature (→ 285 Non-Euclidean Geometry). Two surfaces of the same constant curvature are locally isometric to each other.

Surfaces of revolution of constant curvature are classified. The simplest surface of constant negative curvature is a **pseudosphere**, which is a surface of revolution obtained by rotating a †tractrix $x_1=a\cos\varphi$, $x_3=a\log\tan((\varphi/2)+(\pi/4))-a\sin\varphi$ $(-\pi/2<\varphi<\pi/2)$ around the x_3-axis. A surface generated by a 1-parameter family of straight lines is called a **ruled surface**; a hyperboloid of one sheet, a hyperbolic paraboloid, a circular cylinder, and a circular cone are examples. The first two can be regarded as ruled surfaces in two ways. Each of the straight lines that generate a ruled surface is called a **generating line**. A surface consisting of straight lines parallel to a fixed line and passing through each point of a space curve C is called a **cylindrical surface** with the director curve C. A surface generated by a straight line that connects a certain point o with each point of a curve C is called a **conical surface**. Both a cylindrical surface and a conical surface are ruled surfaces such that $K=0$ everywhere. For ruled surfaces we have $K\leqslant 0$. In particular, a surface such that $H\neq 0$ and $K=0$ everywhere is called a **developable surface**. A developable surface must be either a cylindrical surface, a conical surface, or a tangent surface of a space curve. There exist ruled surfaces that are not developable, for example, hyperboloids of one sheet and hyperbolic paraboloids. A nondevelopable ruled surface is called a **skew surface**. A ruled surface generated by a straight line that moves under a certain rule intersecting a fixed straight line l orthogonally is called a **right conoid**. If we take l as the x_3-axis, the surface is given by the equations $x_1=u\cos v$, $x_2=u\sin v$, $x_3=f(v)$. A surface generated by a curve C (C may be chosen as a plane curve) that moves in the direction of a fixed line l with constant velocity and turns around l with certain constant angular velocity is called a **helicoidal surface**. If we take l as the x_3-axis, the surface is given by the equations $x_1=u\cos v$, $x_2=u\sin v$, $x_3=f(u)+kv$, where k is a constant and $x_3=f(x_1)$ is the equation of C. In particular, if C is a straight line that intersects l orthogonally, then $f(u)=0$, and the surface is called a **right helicoid** (or **ordinary helicoid**). A right conoid is both a ruled surface and a minimal surface. Conversely, a ruled surface that is also a minimal surface is necessarily a right conoid. A helicoidal surface with a tractrix as the curve C is called a **Dini surface** and is a surface of constant negative curvature. On the normal of a surface S two points q_i $(i=1,2)$ are centers of principal curvature at p. The locus of each of these points is a surface called a **center surface** of S. When S is a sphere, two center surfaces degenerate to a point; if S is a surface of revolution, one of the center surfaces degenerates to the axis of revolution and the other is a certain surface of revolution. If S is general, each of the center surfaces is the locus of an edge of regression of the developable surface generated by normals of S along a line of curvature.

When a 1-parameter family of surfaces S_t is given by the equation $F(x_1,x_2,x_3,t)=0$, a surface E that does not belong to this family is called an **enveloping surface** of the family of surfaces $\{S_t\}$ if E is tangent to some S_t at each point of E, that is, if E and S_t have the same tangent plane. The equation of E is obtained by eliminating t from $F(x_1,x_2,x_3,t)=0$ and $(\partial F/\partial t)(x_1,x_2,x_3,t)=0$. In general, if we denote by $\varphi(x_1,x_2,x_3)=0$ the equation obtained by eliminating t from $F=0$ and $\partial F/\partial t=0$, then the surface defined by $\varphi=0$ is either the enveloping surface of $\{S_t\}$ or the locus of singular points of S_t. The intersection C_{t_0} of the enveloping surface E of $\{S_t\}$ and S_{t_0} is a curve defined by $F(x_1,x_2,x_3,t_0)=0$, $(\partial F/\partial t)(x_1,x_2,x_3,t_0)=0$. We call C_{t_0} a **characteristic curve** of $\{S_t\}$. Since $\{C_t\}$ is a family of curves on the enveloping surface E, there may exist an envelope F on E. In such a case, F is called the **line of regression** of $\{S_t\}$. The equation of F is obtained by eliminating t from $F=0$, $\partial F/\partial t=0$, and $\partial^2F/\partial t^2=0$. In particular, the enveloping surface of a family of planes is a developable surface, and their characteristic curves are straight lines. Moreover, the line of

regression coincides with the line of regression of the tangent surface.

If there exists a diffeomorphism between two surfaces such that first fundamental forms at each pair of corresponding points are proportional, then the surfaces are said to be in a **conformal correspondence**. In particular, when the proportionality factor is a constant, they are said to be in a **similar** (or **homothetic**) **correspondence**. There exists a local conformal correspondence between any analytic surface and a plane. Namely, if we choose suitable parameters, we can reduce the first fundamental form of any analytic surface to the form $A(\xi,\eta)(d\xi^2+d\eta^2)$. Such parameters are called **isothermal parameters**. From the existence of isothermal parameters we can see that there exists a local conformal correspondence between any two analytic surfaces. The assumption of analyticity in these theorems is not necessary [10]. If there exists a diffeomorphism between two surfaces under which geodesics are mapped to geodesics, then the surfaces are said to be in **geodesic correspondence**. A surface has a locally geodesic correspondence with a plane if and only if it is a surface of constant curvature. If two surfaces are in geodesic correspondence, then with respect to parameters with the same values at corresponding points, we have the relation

$$\overline{\left\{\begin{matrix}\alpha\\ \beta\gamma\end{matrix}\right\}}=\left\{\begin{matrix}\alpha\\ \beta\gamma\end{matrix}\right\}+\delta^\alpha_\beta A_\gamma+\delta^\alpha_\gamma A_\beta$$

for coefficients of connections of the two surfaces.

By the †Alexander-Pontryagin duality theorem, a submanifold M in E^3 that is homeomorphic to S^2 divides E^3 into two domains, and two points belonging to different domains cannot be connected by a broken segment unless the segment meets the surface. Such a manifold M is called a **closed surface**. One of the two domains consists of those points with bounded distance from a point belonging to the domain. Such a domain is called the interior of the closed surface M. If the set M^* consisting of M and its interior is convex in E^3, the surface M is called a **closed convex surface** (or **ovaloid**).

The Gaussian curvature of an ovaloid cannot be negative at any point. A closed surface with $K>0$ must be an ovaloid. Moreover, it is known that on any closed surface there exists at least one point where $K>0$ (J. Hadamard). If there exists no umbilical point and K is strictly positive in a domain on a surface, then the two principal curvatures regarded as continuous functions on the domain cannot take their local maximum and local minimum values at the same point (D. Hilbert). A compact, connected surface of class C^4 with constant Gaussian curvature is a sphere. A closed surface with $K>0$ and $H=$constant is a sphere (H. Liebmann).

A problem proposed by H. Hopf asks whether an orientable compact surface with constant mean curvature is a sphere. In connection with this problem, Hopf showed that a closed orientable surface of class C^3 of genus zero with constant mean curvature is a sphere. If there is a certain relation $W(k_1,k_2)=0$ between the two principal curvatures k_1, k_2 ($k_1\geqslant k_2$) of a surface, the surface is called a **Weingarten surface** (or **W-surface**). There are many interesting results for W-surfaces. As an extension of the convex surface, tight immersions have been studied (→ 365 Riemannian Submanifolds).

The Gaussian curvature K is invariant under isometries. Hence a sphere is transformed to a sphere by each isometry. This fact is sometimes described as the **rigidity** of a sphere. More generally, if two ovaloids are isometric, then they are congruent (**Cohn-Vossen's theorem**). It is known that if we remove a small circular disk from a sphere, then the remaining portion of the sphere is isometrically deformable. On the existence of closed geodesics on ovaloids, G. D. Birkhoff proved the following theorem: There exist at least three closed geodesics on any ovaloid of class C^3. It is also known that there exist surfaces of revolution that are not spheres but whose geodesics are all closed (→ 178 Geodesics). On a hyperbolic non-Euclidean compact †space form of genus p ($p\geqslant 2$) there exists a geodesic whose points are everywhere dense in it (E. Hopf) (for the ergodicity of flows along geodesics on this surface → 136 Ergodic Theory; also 126 Dynamical Systems).

J. Singular Points of a Surface

Suppose that a neighborhood of a point p_0 of a surface S in E^3 is given by a certain vector-valued function f of class C^r as $\mathfrak{r}=\mathfrak{f}(u,v)$. Then a point p_0 where two vectors $(\partial\mathfrak{f}/\partial u)_{p_0}$, $(\partial\mathfrak{f}/\partial v)_{p_0}$ are linearly independent is called a **regular point**. A point on S that is not regular is called a **singular point**. If for suitable parameters we have $(\partial\mathfrak{f}/\partial u)_{p_0}=0$ but $(\partial\mathfrak{f}/\partial v)_{p_0}$, $(\partial^2\mathfrak{f}/\partial u^2)_{p_0}$, $(\partial^2\mathfrak{f}/\partial u\,\partial v)_{p_0}$ are linearly independent, then such a singular point is called a **semiregular point**. In general, shapes of neighborhoods of singular points are extremely complicated. However, we note the following: (i) by a small deformation of the function f (and its derivatives of orders at most r) we can reduce p_0 to a regular or semiregular point of the deformed surface; (ii) if p_0 is semiregular, we can choose

suitable parameters and curvilinear coordinates of class C^r in E^3 near p_0 so that the surface S in the neighborhood of the origin p_0 is expressed by the equations $x_1 = u^2$, $x_2 = v$, $x_3 = uv$ (H. Whitney's theorem [15]). (The higher-dimensional case has also been considered (Whitney [16]).)

References

[1] W. Blaschke, Vorlesungen über Differentialgeometrie I, Springer, 1924 (Chelsea, 1967).
[2] A. Duschek and W. Mayer, Lehrebuch der Differentialgeometrie I, Teubner, 1930.
[3] L. P. Eisenhart, An introduction to differential geometry, Princeton Univ. Press, 1940; revised edition, 1947.
[4] W. Klingenberg, Eine vorlesung über differentialgeometrie, Springer, 1973.
[5] W. Fenchel, On the differential geometry of closed space curves, Bull. Amer. Math. Soc., 57 (1951), 44–54.
[6] S. S. Chern, Curves and surfaces in Euclidean space, Studies in global geometry and analysis, Studies in math. 4, Math. Assoc. Amer.
[7] I. Fary, Sur la courbure totale d'une courbe gauche fisant un noeud, Bull. Soc. Math. France, 77 (1949), 128–138.
[8] J. W. Milnor, On the total curvature of knots, Ann. Math., (2) 52 (1950), 248–257.
[9] E. Cartan, La théorie des groupes finis et continus et la géométrie différentielle traitees par la méthode du repère mobile, Gauthier-Villars, 1937.
[10] L. Bers, Riemann surfaces, Courant Inst. Math. Sci., 1957–1958.
[11] S. Sternberg, Lectures on differential geometry, Prentice-Hall, 1964.
[12] A. D. Aleksandrov, Intrinsic geometry of convex surfaces (in Russian), Gostekhizdat, 1948; German translation, Die innere Geometrie der konvexen Flächen, Akademie-Verlag, 1955.
[13] S. S. Chern, Topics in differential geometry, Lecture notes, Institute for Advanced Study, Princeton, 1951.
[14] H. Hopf, Zur Differential Geometrie geschlossener Flächen in Euklidischen Raum, Convegne Internazionale di Geometria Differenziale, Italy, 1953, 45–54 (Cremona, 1954).
[15] H. Whitney, The singularities of a smooth n-manifold in $(2n-1)$-space, Ann. Math., (2) 45 (1944), 247–293.
[16] H. Whitney, Singularities of mappings of Euclidean spaces, Symposium Internacional de Topologia Algebraica, Universidad Nacional Autonoma de Mexico and UNESCO, 1958, 285–301.
[17] M. P. do Carmo, Differential geometry of curves and surfaces, Prentice-Hall, 1976.
[18] T. J. Willmore, An introduction to differential geometry, Clarendon Press, 1959.
[19] N. J. Hicks, Notes on differential geometry, Van Nostrand, 1964.
[20] D. Laugwits, Differential and Riemannian geometry, Academic Press, 1965.
[21] B. O'Neill, Elementary differential geometry, Academic Press, 1966.
[22] J. J. Stoker, Differential geometry, Wiley, 1969.

112 (XII.15) Differential Operators

A. Definition

A mapping (or an operator) A of a function space $\mathbf{F}_1$ to a function space $\mathbf{F}_2$ is said to be a **differential operator** if the value $f(x)$ of the image $f = Au$ ($u \in \mathbf{F}_1, f \in \mathbf{F}_2$) at each point x is determined by the values at x of u and a finite number of its derivatives. If u and f are †distributions, the definition applies with the derivative interpreted in the sense of distributions (→ 125 Distributions and Hyperfunctions). In this article we restrict ourselves to the case of linear differential operators and consider only those of the form

$$P(x, D) = \sum_{|\alpha| \leqslant m} a_\alpha(x) D^\alpha, \qquad (1)$$

where α denotes n-tuples $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ of nonnegative integers, called **multi-indices**; $|\alpha|$ the **length** of α: $|\alpha| = \alpha_1 + \alpha_2 + \ldots + \alpha_n$; and D^α the differential operator $D^\alpha = D_1^{\alpha_1} D_2^{\alpha_2} \ldots D_n^{\alpha_n}$, with $D_j = (-i)\partial/\partial x_j$. The coefficient $(-i)$ is sometimes omitted. The coefficients $a_\alpha(x)$ are functions defined on an open set Ω in n-dimensional spacc. We call $P(x, D)$ an **ordinary differential operator** if the dimension n of Ω is 1 and a **partial differential operator** if $n \geqslant 2$. Ordinary differential operators and partial differential operators behave quite differently in many respects.

We set

$$P(x, \xi) = \sum_{|\alpha| \leqslant m} a_\alpha(x) \xi^\alpha, \qquad \xi^\alpha = \xi_1^{\alpha_1} \ldots \xi_n^{\alpha_n},$$

where $\xi = (\xi_1, \xi_2, \ldots, \xi_n) \in \mathbf{R}^n$ or $\mathbf{C}^n$. The **order** of $P(x, D)$ is the greatest integer $|\alpha|$ for which $a_\alpha(x) \not\equiv 0$. In expression (1) m is assumed to be equal to the order, and in that case

$$P_m(x, D) = \sum_{|\alpha| = m} a_\alpha(x) D^\alpha$$

is called the **principal part** of $P(x, D)$, and the corresponding polynomial $P_m(x, \xi)$ the **characteristic polynomial**.

Differential operators have been investigated for a long time in connection with the linear differential equations

$$P(x,D)u(x)=f(x), \qquad x\in\Omega. \tag{2}$$

Except for ordinary differential operators, however, it is rather recently that the properties of such operators have been studied from the general viewpoint.

We denote a differential operator with constant coefficients by $P(D)$. In general, $P(x,D)$ is assumed to be a linear differential operator with coefficients that are C^∞-functions. However, many of the results for C^∞-coefficients also hold when the coefficients are sufficiently differentiable. (For †function spaces $\mathscr{D}(\Omega)$, $\mathscr{D}'(\Omega)$, $\mathscr{E}(\Omega)$, $\mathscr{E}'(\Omega)$, $C(\Omega)$, $C^\infty(\Omega)$, $C_0^\infty(\Omega)$, $L_p(\Omega)$, $\mathscr{A}(\Omega)$, $\mathscr{B}(\Omega)$, etc., → 125 Distributions and Hyperfunctions, 168 Function Spaces).

Differential operators are classified according to their properties. The most important are the elliptic, hyperbolic, and parabolic types. A differential operator $P(x,D)$ is said to be an **elliptic** operator if the characteristic polynomial $P_m(x,\xi)$ has no real zero except for $\xi=0$ for each $x\in\Omega$. Typical examples are the Laplacian $\Delta=-(D_1^2+\ldots+D_n^2)$ and the Cauchy-Riemann operator $\partial/\partial\bar{z}=(1/2)(\partial/\partial x+i\partial/\partial y)$.

A differential operator is said to be **hyperbolic** if the associated Cauchy problem is well posed (→ 325 Partial Differential Equations of Hyperbolic Type). The d'Alembertian $D_1^2-(D_2^2+\ldots+D_n^2)$ is an example. A differential operator of the form $\partial/\partial t+P(t,x,D_x)$ is called **parabolic** if $P(t,x,D_x)$ is strongly elliptic (→ Section G) in x. The heat operator $iD_{n+1}-\Delta$ is typical.

These three types of operators appear most often in applications, and if $n=2$, then any operator of order 2 with real coefficients in $(\partial/\partial x_j)^\alpha$ belongs to one of them at a generic point. In other cases, however, there are differential operators that do not belong to any of them.

B. Fundamental Solutions

If a differential operator $P(x,D)$ with $\mathscr{D}(\Omega)$ as its domain has a left inverse F that is expressed as an †integral operator with †kernel distribution (in $\mathscr{D}'_{x,y}$; → 125 Distributions and Hyperfunctions F), then the kernel is said to be a **fundamental solution** (or **elementary solution**). F is usually a right inverse of the weak extension (→ Section F) of $P(x,D)$ and maps $\mathscr{D}(\Omega)$ into $\mathscr{E}(\Omega)$. The image is mapped to the original function by $P(x,D)$. Nevertheless, F is not a genuine right inverse, and hence the fundamental solution is not unique if it exists.

When $P(D)$ is a differential operator with constant coefficients, we call a distribution $E(x)$ a **fundamental solution** if it satisfies

$$P(D)E(x)=\delta(x), \tag{3}$$

where $\delta(x)$ is †Dirac's distribution (δ function). If $E(x)$ is a fundamental solution in this sense, then $F(x,y)=E(x-y)$ is the kernel of a left inverse of $P(D)$ and is a fundamental solution in the sense of the preceding paragraph.

Every differential operator $P(D)$ with constant coefficients has a fundamental solution in the sense of (3) (**Ehrenpreis-Malgrange theorem**; see L. Hörmander [4] for a proof).

General operators $P(x,D)$ with variable coefficients do not necessarily have fundamental solutions. However, if $P(x,D)$ belongs to one of the classical types of operators (elliptic, hyperbolic, or parabolic), then it has a fundamental solution at least locally. (See F. John [8] for elliptic operators, J. Leray [11] for strongly hyperbolic operators, and S. Mizohata [12] and S. D. Eĭdel'man [9] for parabolic operators.) Leray has generalized John's method to strongly hyperbolic operators in an enormous work [13].

C. Ranges of Differential Operators

Let $P(D)$ be a differential operator with constant coefficients. Then it follows from the Ehrenpreis-Malgrange theorem that $P(D)\mathscr{D}'(\Omega)\supset\mathscr{D}(\Omega)$ holds for any open set Ω. However, there are differential operators $P(x,D)$ with variable coefficients such that for any Ω, $P(x,D)\mathscr{D}'(\Omega)\not\supset\mathscr{D}(\Omega)$. H. Lewy first devised such an example:

$$P(x,D)=-iD_1+D_2-2(x_1+ix_2)D_3.$$

Let $C_{2m-1}(x,D)$ be the homogeneous part of order $2m-1$ of the commutator

$$P(x,D)\overline{P(x,D)}-\overline{P(x,D)}P(x,D).$$

Then in order that $P(x,D)\mathscr{D}'(\Omega)\supset\mathscr{D}(\Omega)$, it is necessary that

$$P_m(x,\xi)=0 \quad \text{imply} \quad C_{2m-1}(x,\xi)=0$$

for all $x\in\Omega$, $\xi\in\mathbf{R}^n$ (**Hörmander's theorem** [4]). When $P(x,D)$ is a differential operator that does not satisfy this condition (e.g., Lewy's operator), choose an $f(x)\in\mathscr{D}(\Omega)$ that is not in $P(x,D)\mathscr{D}'(\Omega)$. Then the differential equation (2) has no distribution solutions at all. P. Schapira extended this result to the case of †hyperfunctions (also → 274 Microlocal Analysis).

Concerning the ranges of differential operators $P(D)$ with constant coefficients, we have the following detailed results due to L. Ehrenpreis [28], B. Malgrange [14], and Hörmander [4].

An open set Ω is said to be **P-convex** for a differential operator $P(D)$ if for each compact set $K \subset \Omega$ there exists a compact set $K' \subset \Omega$ such that $\varphi \in C_0^\infty(\Omega)$ and $\operatorname{supp} P(-D)\varphi \subset K$ imply $\operatorname{supp} \varphi \subset K'$. Convex sets are P-convex for any $P(D)$. All open sets are P-convex if and only if $P(D)$ is an elliptic operator.

Theorem: The following conditions are equivalent: (i) Ω is P-convex; (ii) $P(D)\mathscr{D}'(\Omega) \supset \mathscr{E}(\Omega)$; (iii) $P(D)\mathscr{E}(\Omega) = \mathscr{E}(\Omega)$. Property (iii), the Mittag-Leffler theorem, and the solvability of †Cousin's first problem for the solutions of $P(D)u = 0$ are equivalent.

An open set Ω is said to be **strongly P-convex** if for each compact set $K \subset \Omega$ there exists a compact set K' such that $\mu \in \mathscr{E}'(\Omega)$ and $\operatorname{supp} P(-D)\mu \subset K$ imply $\operatorname{supp} \mu \subset K'$; and $\mu \in \mathscr{E}'(\Omega)$ and sing supp $P(-D)\mu \subset K$ imply sing supp $\mu \subset K'$; where the **singular support** of μ is the closure of the set of all points at which μ is not a C^∞-function. Convex sets are strongly P-convex, and strongly P-convex sets are P-convex.

Theorem: Ω is strongly P-convex if and only if $P(D)\mathscr{D}'(\Omega) = \mathscr{D}'(\Omega)$.

R. Harvey has shown that every domain Ω is P-convex in the sense of hyperfunctions, i.e., the equation $P(D)u = f$ always has a hyperfunction solution u on Ω for any hyperfunction f on Ω. For the †real analytic functions $\mathscr{A}(\Omega)$, however, $P(D)\mathscr{A}(\Omega) = \mathscr{A}(\Omega)$ does not hold for convex open set Ω in general. Hörmander [15] gave a necessary and sufficient condition for $P(D)$ and Ω in order that this hold.

D. Hypoellipticity

A differential operator $P(x, D)$ is called **hypoelliptic** in Ω if for any distribution $u(x) \in \mathscr{D}'(\Omega)$, $Pu \in C^\infty(\Omega)$ implies $u \in C^\infty(\Omega)$. Further, a differential operator with real analytic coefficients $P(x, D)$ is called **analytically hypoelliptic** in Ω if P is hypoelliptic and if for any distribution $u(x) \in \mathscr{D}'(\Omega)$, $Pu \in \mathscr{A}(\Omega)$ implies $u(x) \in \mathscr{A}(\Omega)$.

There are two fundamental facts about such operators. Let P have constant coefficients; then $P(D)$ is hypoelliptic if and only if $P(\xi + i\eta) = 0$ and $|\xi + i\eta| \to \infty$ imply that $|\eta| \to \infty$ (**Hörmander's theorem** [4]). Furthermore, $P(D)$ is analytic hypoelliptic if and only if P is elliptic (**Petrovskiĭ's theorem** [16]). The heat operator is not elliptic, but hypoelliptic. If $P(D)$ is elliptic, then actually any hyperfunction $u(x) \in \mathscr{B}(\Omega)$ such that $Pu \in \mathscr{A}(\Omega)$ is real analytic (R. Harvey, G. Bengel). On the other hand, if $P(D)$ is not elliptic, there is a hyperfunction solution u of $Pu = 0$ that is not a distribution.

Strictly speaking, the notion of the hypoellipticity for general differential operators was first formulated explicitly by L. Schwartz [27]. Before that time, D. Hilbert, E. E. Levi and K. O. Friedrichs, and others investigated this problem for some elliptic operators, and the hypoellipticity was called **Weyl's lemma** (→ 323 Partial Differential Equations of Elliptic Type).

Similarly to the constant coefficient case the following two theorems are fundamental: Elliptic and parabolic operators $P(x, D)$ with C^∞ coefficients are hypoelliptic (Schwartz [27], Mizohata [12]). Elliptic operators $P(x, D)$ with real analytic coefficients are analytic hypoelliptic (I. G. Petrovskiĭ [16], C. B. Morrey and L. Nirenberg). The latter holds also for hyperfunction solutions (M. Sato and Schapira). Moreover, the following result is known: If for each compact set K in an open set Ω, there exists a constant C such that $\|P^k u\|_K \leqslant C^{k+1}(mk)!$, then $u \in \mathscr{A}(\Omega)$, where $\|\cdot\|_K$ denotes the L_p-norm on K (H. Komatsu, T. Kotake, and M. S. Narasimhan).

However, as seen by the example $D_1 + ix_1^k D_2$ in $\mathbf{R}^2$ (k is even), the analytic hypoellipticity also holds for nonelliptic operators. Such an operator is called a subelliptic operator; subelliptic operators have been investigated by Hörmander, Yu. V. Egorov, F. Treves, and others [19].

Hörmander has obtained a fairly complete result on the hypoellipticity of the operators of the form

$$L = \sum_{i=1}^{r} X_i^2 + X_0 + c(x),$$

where $X_0, \ldots, X_r$ are homogeneous first-order differential operators with real coefficients ([17]; O. A. Oleĭnik and E. V. Radkevich [18]). Hypoellipticity was investigated extensively after the introduction of pseudodifferential operators and Fourier integral operators (→ 345 Pseudodifferential Operators).

E. Differential Operators in Banach Spaces

We consider differential operators $P(x, D)$ defined on a domain Ω as operators in the function spaces $C(\Omega)$ or $L_p(\Omega)$. Differential operators of order $m \geqslant 1$ are always unbounded operators in the Banach space $X = C(\Omega)$ or $L_p(\Omega)$. Moreover, their domains of definition as operators in X are not generally determined uniquely by the expressions $P(x, D)$ as differential operators.

$P(x, D)$ is a linear operator that maps $C_0^\infty(\Omega)$ into X. This operator has a closed extension. The minimal closed extension P_0 is called the **minimal operator** of $P(x, D)$ in X. We have $u \in \mathscr{D}(P_0)$ and $P_0 u = f$ if and only if there exists a sequence $\varphi_n \in C_0^\infty(\Omega)$ such that $\varphi_n \to u$, $P(x, D)\varphi_n \to f$. On the other hand, the

closed linear operator P_1 whose domain is the set of all $u \in X$ such that $P(x,D)u \in X$ in the sense of distribution is called the **maximal operator** (or **weak extension**) of $P(x,D)$. We have $u \in \mathscr{D}(P_1)$ and $P_1 u = f$ if and only if $\langle u, {}^tP(x,D)\varphi \rangle = \langle f, \varphi \rangle$ for any $\varphi \in C_0^\infty(\Omega)$, where ${}^tP(x,D)$ is the **transposed operator**

$$
{}^tP(x,D)\varphi(x) = \sum_{|\alpha| \leqslant m} (-D)^\alpha (a_\alpha(x)\varphi(x)).
$$

Integration by parts shows that P_1 is an extension of P_0 and that when X is the †dual space of a space Y, the weak extension P_1 in X is the †dual of the minimal operator of ${}^tP(x,D)$ in Y. Let $X = L_p(\Omega)$ $(1 < p < \infty)$, Ω be a bounded open set with smooth boundary, and $P(D)$ have constant coefficients. Then P_1 coincides with the smallest closed extension of the operator $P(D)$ having as its domain the set of all $u \in C^\infty(\Omega) \cap X$ such that $P(D)u \in X$. The latter closed extension is called the **strong extension**. The difference between the weak and the strong extension is not obvious in the variable coefficient case.

P_0 coincides with P_1 when Ω is the entire space and $P(x,D)$ is an elliptic operator whose coefficients are constants or close to constants (J. Peetre, *Medd. Lunds Univ. Mat. Sem.*, 16 (1959); T. Ikebe and T. Kato, *Arch. Rational Mech. Anal.*, 9 (1962)). In general, we have $P_0 \neq P_1$. Let $P(x,D)$ be an ordinary differential operator with bounded coefficients such that $|a_m(x)| \geqslant \delta > 0$ and Ω be the bounded interval (a,b). Then the domain of P_1 coincides with the set of all $(m-1)$-times continuously differentiable functions u such that the $(m-1)$st derivative is absolutely continuous and $P(x,D)u \in X$, while the domain of P_0 is the set of all functions u which satisfy in addition the boundary conditions

$$
\begin{aligned}
u(a) &= u'(a) = \ldots = u^{(m-1)}(a) \\
&= u(b) = \ldots = u^{(m-1)}(b) = 0.
\end{aligned}
$$

(Moreover, $u^{(m)}(a) = u^{(m)}(b) = 0$ when $X = C(a,b)$.)

Let $G(P_0)$, $G(P_1)(\subset X \times X)$ be the †graphs of P_0, P_1. Then the quotient space $\mathscr{B} = G(P_1)/G(P_0)$ is called the **boundary space**, and an element of the dual $\mathscr{B}'$ of $\mathscr{B}$, i.e., a continuous linear functional on $G(P_1)$ which vanishes on $G(P_0)$, is called a **boundary value** relative to $P(x,D)$. For the ordinary differential operators discussed above, the boundary space is the set of all linear combinations of $u^{(i)}(a)$ and $u^{(j)}(b)$. When $P(x,D)$ is an ordinary differential operator, we can explicitly determine the boundary values also in the case where the interval (a,b) is infinite, the coefficients $a_\alpha(x)$ are not bounded, or $a_m(x) \to 0$ $(x \to a, b)$; and we can show that $\mathscr{B}$ is finite-dimensional. When $P(x,D)$ is a partial differential operator, $\mathscr{B}$ is generally of infinite dimension, and the concrete forms of the elements of $\mathscr{B}$ and $\mathscr{B}'$ are not known. However, we have some information by M. I. Vishik (*Amer. Math. Soc. Transl.*, (2) 24 (1963)) about the boundary values of elliptic operators of the second order. Combining this with the results by J.-L. Lions and E. Magenes (*J. Anal. Math.*, 11 (1963)), we can obtain information for elliptic operators of higher order.

F. Differential Operators with Boundary Conditions

A closed operator between the minimal operator P_0 and the maximal operator P_1 is determined by designating a closed subspace B of the boundary space $\mathscr{B}$. This operator is called the **operator with the boundary condition** B. Particularly important are boundary conditions expressed in the form

$$
Q_i(x,D)u(x) = 0, \qquad x \in \partial\Omega, \qquad i = 1, \ldots, k, \tag{4}
$$

with differential operators $Q_i(x,D)$ $(i = 1, \ldots, k)$ defined on the boundary $\partial\Omega$ of Ω.

When $P(x,D)$ is an ordinary differential operator defined on a finite interval and the orders of Q_i are at most $m-1$ (or m), (4) always has a definite meaning. However, for partial differential operators, we need an interpretation of (4), i.e., (4) does not necessarily determine the subspace B of $\mathscr{B}$ uniquely.

Let P_s be the smallest closed extension of $P(x,D)$ with $\{u \in C^\infty(\bar{\Omega}) \cap X \mid Q_i(x,D)u(x) = 0, x \in \partial\Omega;\ P(x,D)u \in X\}$ as its domain. P_s is called the **strong extension** of the differential operator $P(x,D)$ with boundary condition (4).

On the other hand, when Ω, P, and Q_i satisfy suitable conditions, we can define the transposed differential operator ${}^tP(x,D)$ with the transposed boundary operators $R_j(x,D)$ $(j = 1, \ldots, l)$. Namely, there are differential operators $R_j(x,D)$ on the boundary such that a necessary and sufficient condition for $u \in C^\infty(\bar{\Omega})$ to satisfy (4) and $Pu(x) = f(x)$ is

$$
\int_\Omega f(x)v(x)\,dx = \int_\Omega u(x)\,{}^tPv(x)\,dx \tag{5}
$$

for all $v(x) \in C^\infty(\bar{\Omega})$ with the boundary conditions $R_j v = 0$, $x \in \partial\Omega$. Then the operator P_w defined by $P_w u(x) = f(x)$ for the pairs $u(x)$, $f(x) \in X$ satisfying (5) is called the **weak extension** of the differential operator $P(x,D)$ with boundary condition (4). As in the case of operators without boundary conditions, the weak extension is an extension of the strong extension, and generally is the dual of the strong extension in the dual space X' of the transposed differential operator with the transposed boundary condition.

Regularity up to the boundary. The fundamental problems for the differential operator $P(x, D)$ with the homogeneous boundary condition (4) are to determine, for both the strong and the weak extensions, the spaces of solutions of the homogeneous equations $Pu=0$ and their ranges. The problems mostly reduce to determining when the strong and the weak extensions coincide and, including this, also to the problem of regularity on the closed domain $\bar{\Omega}$ containing the boundary of the solutions u of the equation $P_w u = f$.

This problem was solved by Nirenberg for strongly elliptic operators in $L^2(\Omega)$ with the Dirichlet boundary condition

$$\partial^{j-1}u(x)/\partial n^{j-1}=0, \quad j=1,2,\ldots,m/2,$$

and generalized later by F. Browder, M. Schechter [20], S. Agmon, Lions, and others for elliptic operators in $L_p(\Omega)$ with a kind of coercive boundary condition (→ Section H).

Consequently, when Ω is bounded and smooth, P_w is equal to P_s for those operators. Write P for P_w. Then the space $N(P)$ of the solutions of $Pu=0$ is a subspace of finite dimension, and the range $R(P)$ is a closed subspace of finite codimension. In particular, it follows that the †index of P, $\dim N(P) - \operatorname{codim} R(P)$, is finite.

G. Strongly Elliptic Operators

A differential operator $P(x, D)$ is said to be **strongly elliptic** if its characteristic polynomial satisfies

$$\operatorname{Re} P_m(x, \xi) \geqslant C|\xi|^m > 0, \qquad \xi \neq 0.$$

Many of the elliptic operators, such as the Laplacian, that appear in applications are strongly elliptic. L. Gårding treated the boundary value problem with the Dirichlet condition (in the generalized sense) for strongly elliptic operators. His work initiated the general study of differential operators (→ 323 Partial Differential Equations of Elliptic Type). His theory is based on the following inequality, called **Gårding's inequality**:

$$\|\nabla^{m/2}u\|_{L_2}^2 \leqslant C\left(\operatorname{Re}\int Pu\cdot\bar{u}\,dx + \|u\|_{L_2}^2\right), \qquad u \in C_0^\infty(\Omega).$$

H. Coercive Boundary Conditions

The boundary condition (4) is said to be **coercive** if

$$\|\nabla^m u\| \leqslant C(\|Pu\| + \|u\|) \tag{6}$$

holds for any $u \in C^\infty(\bar{\Omega})$ that satisfies (4). In order that a differential operator P have a coercive boundary condition, it is necessary that it be a special type of elliptic operator. In this case, Agmon, N. Aronszajn, Schechter, and others found conditions under which (4) is coercive. Agmon, A. Douglis, and Nirenberg show that the inequality (6) holds in $L_p(\Omega)$ and in the normed spaces of Hölder continuous functions under a suitable condition. The classical boundary conditions $au + b\partial u/\partial n = 0$ ($a \geqslant 0, b \geqslant 0, a+b=1$) are coercive for elliptic operators of the second order (→ 323 Partial Differential Equations of Elliptic Type). However, problems remain when the coefficients of $Q_j(x, D)$ are discontinuous. In order to have coincidence of the strong and weak extensions or regularity up to the boundary, it is necessary neither that P be elliptic nor that the boundary condition be coercive. But it is not known to what extent these conditions can be weakened. At present, major contributions are Hörmander's work (*Acta Math.*, 99 (1958)) dealing with operators with constant coefficients and flat boundaries, and works by J. J. Kohn [21], Nirenberg, and Hörmander concerning noncoercive boundary conditions. The latter works are connected with the theory of several complex variables, and have attracted much attention.

I. Self-Adjoint Extension

One of the fundamental problems in the case $X = L_2(\Omega)$ is whether the minimal operator P_0 has a self-adjoint extension. P_0 is symmetric if and only if $P(x, D)$ is **formally self-adjoint**: $P(x, D) = {}^t\overline{P(x, D)}$. Under this condition the boundary space $\mathscr{B}$ turns out to be the direct sum of two subspaces $\mathscr{B}_\pm = \{(x, P_1x) \mid x \in \mathscr{D}(P_1), P_1x = \pm ix\} + G(P_0)$. The numbers $n_\pm = \dim \mathscr{B}_\pm$ are called the **deficiency indices** of P_0, and P_0 has a self-adjoint extension if and only if $n_+ = n_-$.

H. Weyl gave a method for computing $n_\pm$ for the **Sturm-Liouville operators**:

$$P(x, D) = -\left(\frac{d}{dx}p(x)\frac{d}{dx}\cdot\right) + q(x), \qquad x \in (a, b).$$

We say that a (resp. b) is of **limit circle type** if the solutions $u(x)$ of $P(x, D)u(x) + lu(x) = 0$ ($l \in \mathbf{C}$) always belong to L_2 in a neighborhood of a (resp. b) and of **limit point type** if a solution does not belong to L_2. This classification does not depend on the choice of $l \in \mathbf{C}$. (1) If both a and b are of limit point type, then $n_+ = n_- = 0$ and hence P_0 is self-adjoint. (2) If a is of limit circle type and b is of limit point type, then $n_+ = n_- = 1$, and the self-adjoint extensions of P_0 are the operators P_α that are obtained from P_1 by assigning the boundary

condition

$p(a)u'(a)\cos\alpha+u(a)\sin\alpha=0.$

The same is true when a and b are interchanged. (3) If both a and b are of limit circle type, then $n_+=n_-=2$, and we can impose two boundary conditions to obtain the self-adjoint extensions.

These results have been extended by K. Kodaira [23] and N. Dunford and J. Schwartz [3] to the case of ordinary differential operators of order m. There are formally self-adjoint operators that have no self-adjoint extensions. For example, the operator $-id/dx$ in $L_2(0,\infty)$ has deficiency indices $n_+=1\neq n_-=0$.

For partial differential operators, it is difficult to determine explicitly all self-adjoint extensions of a given formally self-adjoint operator because the boundary space is complicated. If the boundary condition (4) is formally self-adjoint and coercive, then it follows from the results of Schechter and others that the differential operator with boundary condition (4) is self-adjoint. Furthermore, conditions under which P_0 is self-adjoint or has a self-adjoint extension are known. The following theorem is often used as a condition of the latter type. If a †symmetric operator defined on a dense subspace of a Hilbert space X is positive definite:

$$(Tx,x)\geqslant 0,\qquad x\in\mathscr{D}(T),$$

then there is a positive definite self-adjoint extension $\tilde{T}$ (Friedrichs's theorem). The self-adjoint extension obtained by this theorem is called the **Friedrichs extension**.

J. Generators of Semigroups

From the point of view of probability theory, W. Feller investigated the extensions of the Laplacian d^2/dx^2 and similar operators that are the generators of order-preserving semigroups. Recently various attempts have been made to generalize his results to the multidimensional case (→ 115 Diffusion Processes; 378 Semigroups of Operators and Evolution Equations).

P. D. Lax and A. N. Milgram proved that if $P(x,D)$ is a strongly elliptic operator, then $-P(x,D)$ with the Dirichlet condition in $L_2(\Omega)$ is the generator of a semigroup [1].

K. Boundary Value Problems

There are two methods of solving the inhomogeneous boundary value problem

$$P(x,D)u(x)=f(x),\qquad x\in\Omega,$$

$$Q_i(x,D)u(x)=g_i(x),\qquad x\in\partial\Omega,\qquad i=1,\ldots,k.$$

In the first method, we take a function $v(x)$ that satisfies $Q_i(x,D)v(x)=g_i(x)$ and reduce the problem to the homogeneous one for $u_0=u-v$. In the second method, we consider the pair $P(x,D)$ and $Q_i(x,D)$ as an operator that maps a function u to the pair of functions (Pu,Q_iu) and investigate it directly. The latter method was adopted by Peetre and Hörmander [4].

L. Estimates in Weighted Spaces

J. F. Treves, Hörmander [4], and H. Kumanogo obtained estimates similar to (6) in L_p-spaces relative to the weighted measure $w_t(x)\,dx$ instead of the usual L_p-spaces, and applied them to the proof of the uniqueness of Cauchy problems for differential equations with variable coefficients (→ 321 Partial Differential Equations (Initial Value Problems)). Hörmander [22] applied similar estimates to the proof of the †fundamental theorems of Stein manifolds.

M. Eigenfunction Expansions

When a self-adjoint operator P in the Hilbert space $L_2(\Omega)$ is a self-adjoint extension of a differential operator $P(x,D)$, the †spectral decomposition of P is concretely expressed by the expansion of functions $u\in L_2(\Omega)$ into eigenfunctions of $P(x,D)$.

If Ω is bounded, $P(x,D)$ is an elliptic operator defined on a neighborhood of $\bar{\Omega}$, and the boundary condition is coercive, then the †spectrum of P is composed solely of eigenvalues, and the eigenvectors of P are eigenfunctions of $P(x,D)$ in the classical sense and are of class C^∞ up to the boundary.

N. Asymptotic Distribution of Eigenvalues

If $P=-\Delta$, the number $\nu(\lambda)$ of eigenvalues less than λ satisfies the asymptotic relation

$$\nu(\lambda)\sim\frac{\lambda^{n/2}A}{2^{n-1}\pi^{n/2}n\Gamma(n/2)},\quad\text{as}\quad\lambda\to\infty$$

regardless of the shape of the domain and the boundary condition, where n is the dimension and A is the volume of Ω [2]. This was first proved by Weyl and extended by T. Carleman, Gårding, and others to the case of operators of higher order with variable coefficients (→ 323 Partial Differential Equations of Elliptic Type).

O. Weyl-Stone-Titchmarsh-Kodaira Theory

When Ω is unbounded or the coefficients of $P(x, D)$ have singularities near the boundary of Ω, the spectrum of P may have a continuous part.

Let $P(x, D)$ be an ordinary differential operator on an interval (a, b). Then for each $\lambda \in \mathbf{C}$ the equation $(P(x, D)-\lambda)\varphi(x)=0$ has m linearly independent solutions $\varphi_k(x, \lambda)$ $(k=1, \ldots, m)$, and any solution is represented as a linear combination of them. Weyl and M. H. Stone obtained the spectral decomposition of P (of the second order) in the form

$$u(x)=\sum_{j,k=1}^{m}\int_{-\infty}^{\infty}\varphi_j(x,\lambda)\,d\rho_{jk}(\lambda)\times\int_a^b\overline{\varphi_k(y,\lambda)}u(y)\,dy,$$

$$Pu(x)=\sum_{j,k=1}^{m}\int_{-\infty}^{\infty}\lambda\varphi_j(x,\lambda)\,d\rho_{jk}(\lambda)\times\int_a^b\overline{\varphi_k(y,\lambda)}u(y)\,dy,$$

where the $\rho_{jk}(\lambda)$ are functions of bounded variation and their variations represent the spectral measure. This formula shows that linear combinations of the $\varphi_j(x, \lambda)$ form generalized eigenfunctions even when λ belongs to the continuous spectrum. Later, E. C. Titchmarsh and Kodaira gave a formula to obtain the density matrix $\rho_{jk}(\lambda)$ and completed the theory (Titchmarsh [6], Kodaira [23], Dunford and Schwartz [3]). This expansion theorem makes it possible to deduce in a unified manner expansion theorems for classical special functions, such as the †Fourier series expansion theorem, the expansions by †Hermite polynomials and †Laguerre polynomials, the †Fourier integral theorem, and various expansions in terms of †Bessel functions [6, 7].

The relation between the coefficients of the differential operator and the spectral distribution of P is important in applications and is the subject of many papers [3, 5, 6].

For example, let $P(x, D)=-d^2/dx^2+q(x)$ and $\Omega=(-\infty, \infty)$. If $q(x)\rightarrow\infty$ as $|x|\rightarrow\infty$, then $P(x, D)$ is essentially self-adjoint, the spectrum is entirely composed of the point spectrum, and a detailed estimate of the jth eigenvalue λ_j is also known [6]. If $q(x)$ converges rapidly to 0 as $|x|\rightarrow\infty$, then $P(x, D)$ is essentially self-adjoint, and there is only a continuous spectrum for $\lambda>0$ and eigenvalues for $\lambda<0$ with at most 0 as accumulation point. If $q(x)$ is a periodic function of x, then $P(x, D)$ is again essentially self-adjoint, and the spectrum consists of a continuous spectrum decomposed in a sequence of nonoverlapping intervals. The converse problem of determining $q(x)$ when the spectral measure is given has been studied by I. M. Gel'fand and B. M. Levitan (*Amer. Math. Soc. Transl.*, (2) 1 (1955); original in Russian, 1951).

P. Eigenfunction Expansion for Partial Differential Operators

The theory of eigenfunction expansion for partial differential operators with continuous spectra is not complete, as it is for ordinary differential operators. What causes difficulties is that the solutions of $(P(x, D)-\lambda)u=0$, which should be the generalized eigenfunctions, form an infinite-dimensional space, and except for special cases it is impossible to introduce convenient parameters in it. Many proofs are known for a general result, saying that any self-adjoint elliptic operator has an eigenfunction expansion into generalized eigenfunctions of the form

$$u(x)=\int_{-\infty}^{\infty}\sum_{j=1}^{k(\lambda)}\varphi_j(x,\lambda)\,d\rho_j(\lambda)\int_\Omega\overline{\varphi_j(y,\lambda)}u(y)\,dy,$$

[10]. There are few operators, however, for which we know how to construct $\varphi_j(x, \lambda)$ and the measure $d\rho_j(\lambda)$. The Fourier transform gives the expansion for operators with constant coefficients defined on the whole space. By means of a generalized form of the Fourier transform, Ikebe gave an eigenfunction expansion for the Schrödinger operator $-\Delta+q(x)$ in $\mathbf{R}^3$ under the condition that $q(x)\in L_2$ and $O(|x|^{-2-\varepsilon})$ as $|x|\rightarrow\infty$ [24]. Y. Shizuta, Mizohata, Lax and R. S. Phillips, N. A. Shenk, Ikebe, D. K. Fadeev, and others proved the same results for similar operators defined on an exterior domain with a bounded set deleted from a higher-dimensional Euclidean space. These theories are closely related to the †scattering theory of the Schrödinger equation $(-id/dt-P)u=0$ or the wave equation $(d^2/dt^2+P)u=0$ associated with those operators. Lax and Phillips have developed the latter scattering theory [26] (→ 375 Scattering Theory).

Many of the problems in †quantum mechanics reduce to finding the spectral distribution of self-adjoint partial differential operators.

Q. Expansion Theorems for Non-Self-Adjoint Operators

A kind of eigenfunction expansion theorem may hold for non-self-adjoint differential operators or for non-Hilbert spaces. (See papers by the Russian school for ordinary differential operators and those by Browder,

Agmon, and others for partial differential operators.)

R. Systems of Differential Operators

We have so far dealt with single differential operators that map functions u to functions f. A linear differential operator that maps a p-tuple $(u_1, \ldots, u_p)$ of functions to a q-tuple $(f_1, \ldots, f_q)$ of functions can be written as

$$f_i(x) = \sum_{j=1}^{p} P_{ij}(x, D) u_j(x), \qquad i = 1, \ldots, q,$$

where $P(x, D) = (P_{ij}(x, D))$ is a matrix of single differential operators. Such a matrix is called a **system of differential operators**. A system $P(x, D)$ is said to be **underdetermined** if $p > q$, **determined** if $p = q$, and **overdetermined** if $p < q$. Many propositions that hold for single operators hold for determined systems under appropriate conditions.

However, there is a fundamental difference between overdetermined (underdetermined) systems and determined systems, as is seen from the theory of several complex variables, which is the theory of a typical overdetermined system $\partial/\partial\bar{z} = (\partial/\partial\bar{z}_i)$. The theory is much more difficult for overdetermined (underdetermined) systems. The general theory of overdetermined and underdetermined systems with constant coefficients has been constructed by Ehrenpreis [28], Malgrange, V. Palamodov, and Hörmander [22] for C^∞-functions and distributions. It has been extended by H. Komatsu to the case of hyperfunctions. T. Miwa has discussed the same problem for real analytic functions.

S. Symmetric Systems of the First Order

Determined systems of the first order are important in applications. Many problems in mathematical physics are formulated in terms of them. Also, single equations of higher order can be reduced to determined systems of the first order by regarding the derivatives as unknown functions. In some cases determined systems of the first order are easier to handle than single operators of higher order. In particular, a system of differential operators

$$P(x, D) = \sum A_i(x) \partial/\partial x_i + B(x)$$

is said to be a **symmetric positive system** if the matrices $A_i(x)$ and $B(x)$ satisfy the following conditions: $A_i^* = A_i$; $B + B^* + \sum \partial A_i/\partial x_i$ is positive semidefinite. Symmetric positive systems have been studied in detail by Friedrichs [25], Phillips, C. S. Morawetz, Lax, and others.

References

[1] K. Yosida, Functional analysis, Springer, 1965.
[2] R. Courant and D. Hilbert, Methods of mathematical physics, Interscience, I, 1953; II, 1962.
[3] N. Dunford and J. T. Schwartz, Linear operators II, Interscience, 1963.
[4] L. Hörmander, Linear partial differential operators, Springer, 1963.
[5] M. A. Naimark, Linear differential operators I, II, Ungar, 1967. (Original in Russian, 1954.)
[6] E. C. Titchmarsh, Eigenfunction expansions associated with second-order differential equations, Claredon Press, I, revised edition, 1962; II, 1958.
[7] K. Yosida, Lectures on differential and integral equations, Interscience, 1955. (Original in Japanese, 1950.)
[8] F. John, Plane waves and spherical means applied to partial differential equations, Interscience, 1955.
[9] S. D. Èĭdel'man, Parabolic systems, North-Holland, 1969. (Original in Russian, 1964.)
[10] I. G. Gel'fand and N. Ya. Vilenkin, Generalized functions IV, Applications of harmonic analysis, Academic Press, 1964.
[11] J. Leray, Hyperbolic differential equations, Lecture notes, Institute for Advanced Study, Princeton Univ. Press, 1953.
[12] S. Mizohata, Hypoellipticité des équations paraboliques, Bull. Soc. Math. France, 85 (1957), 15–50.
[13] J. Leray, Problème de Cauchy I–V, Bull. Soc. Math. France, 85 (1957), 389–429; 86 (1958), 75–96; 87 (1959), 81–180; 90 (1962), 39–156; 92 (1964), 263–361.
[14] B. Malgrange, Existence et approximation des solutions des équations aux dérivées partielles et des équations de convolution, Ann. Inst. Fourier, 6 (1955–1956), 271–355.
[15] L. Hörmander, On the existence of real analytic solutions of partial differential equations with constant coefficients, Inventiones Math., 21 (1973), 151–182.
[16] I. G. Petrovskiĭ, Sur l'analyticité des solutions des systèmes d'équations différentielles, Mat. Sb., 5 (47) (1939), 3–70.
[17] L. Hörmander, Hypoelliptic second-order differential equations, Acta Math., 119 (1967), 147–171.
[18] O. A. Oleĭnik and E. V. Radkevich, Second-order equations with nonnegative characteristic form, Amer. Math. Soc., 1973. (Original in Russian, 1971.)
[19] L. Hörmander, Seminar on singularities of solutions of linear partial differential equations, Ann. Math. Studies 91, Princeton Univ. Press, 1979.

[20] M. Schechter, On L^p estimates and regularity I, II, Amer. J. Math., 85 (1963), 1–13; Math. Scand., 13 (1963), 47–69.
[21] J. J. Kohn, A priori estimates in several complex variables, Bull. Amer. Math. Soc., 70 (1964), 739–745.
[22] L. Hörmander, An introduction to complex analysis in several variables, Van Nostrand, 1966.
[23] K. Kodaira, On ordinary differential equations of any even order and the corresponding eigenfunction expansions, Amer. J. Math., 72 (1950), 502–544.
[24] T. Ikebe, Eigenfunction expansions associated with the Schrödinger operators and their applications to scattering theory, Arch. Rational Mech. Anal., 5 (1960), 1–34.
[25] K. O. Friedrichs, Symmetric positive linear differential equations, Comm. Pure Appl. Math., 11 (1958), 333–418.
[26] P. D. Lax and R. S. Phillips, Scattering theory, Academic Press, 1967.
[27] L. Schwartz, Théorie des distributions, Hermann, 1966.
[28] L. Ehrenpreis, Fourier analysis in several complex variables, Wiley-Interscience, 1970.
[29] J.-E. Björk, Rings of differential operators, North-Holland, 1979.
[30] L. Hörmander, The analysis of linear partial differential operators, I–IV, Springer, 1983–1985.

113 (III.17) Differential Rings

Let R be a †commutative ring with a unity element 1. If a map δ of R into R is such that for any pair x, y of elements of R, (i) $\delta(x+y) = \delta x + \delta y$, and (ii) $\delta(xy) = \delta x \cdot y + x \cdot \delta y$, then δ is called a **derivation** (or **differentiation**) in R. A ring R provided with a finite number of mutually commutative differentiations in R is called a **differential ring**. In this article we consider only the case where R contains a subfield that has the unity element in common with R. In particular, if R is a field, we call it a **differential field**.

In the above definition of differential ring, it is not necessary to mention the †characteristic of the subfield contained in R. However, to make it more effectively applicable in the case of nonzero characteristics, we may define differential rings using *higher differentiation* in place of the differentiation defined above. If a sequence $\delta = \{\delta_\nu\}$ of maps $\delta_0, \delta_1, \delta_2, \ldots$ of R into R satisfies the following conditions (i)–(iv) for any pair x, y of elements of R and any pair λ, μ of nonnegative integers, then δ is called a **higher differentiation** in R: (i) $\delta_\lambda(x+y) = \delta_\lambda x + \delta_\lambda y$; (ii) $\delta_\lambda(xy) = \sum \delta_\alpha x \cdot \delta_\beta y$ (the addition is performed over all pairs α, β of nonnegative integers that satisfy $\alpha + \beta = \lambda$); (iii) $\delta_\lambda(\delta_\mu x) = \binom{\lambda+\mu}{\lambda}\delta_{\lambda+\mu}x$; (iv) $\delta_0 x = x$. Two higher differentiations $\delta = \{\delta_\nu\}$ and $\delta' = \{\delta'_\nu\}$ are said to be commutative if and only if δ_λ and δ'_μ commute for all pairs λ, μ of nonnegative integers. Higher differentiations were introduced by H. Hasse (1935) for the study of the field of †algebraic functions of one variable in the case of nonzero characteristics.

These two definitions of differential rings coincide if the characteristic of R is zero. For simplicity, we shall restrict ourselves to that case.

Let $\delta_1, \ldots, \delta_m$ be the differentiations of the differential ring R. If x is an element of R, $\delta_1^{s_1}\delta_2^{s_2}\ldots\delta_m^{s_m}x (s_1, s_2, \ldots, s_m$ are nonnegative integers) is called a **derivative** of x. We call x constant if and only if $\delta_1 x = \ldots = \delta_m x = 0$. An †ideal $\mathfrak{a}$ of R with $\delta_i \mathfrak{a} \subset \mathfrak{a}$ $(i = 1, 2, \ldots, m)$ is called a **differential ideal** of R. If it is a †prime ideal (**semiprime ideal** (i.e., an ideal containing all those elements x that satisfy $x^g \in \mathfrak{a}$ for some natural number g)), then $\mathfrak{a}$ is called a **prime differential ideal (semiprime differential ideal)**. A subring S of R with $\delta_i S \subset S$ can be regarded as a differential ring with respect to the differentiations $\delta_1, \ldots, \delta_m$. We call S a **differential subring** of R and R a **differential extension ring** of S.

Let $X_1, \ldots, X_n$ be elements of a differential extension field of a differential field K with the differentiations $\delta_1, \ldots, \delta_m$, and let $\delta_1^{s_1}\delta_2^{s_2}\ldots\delta_m^{s_m}X_i$ $(s_1 \geq 0, \ldots, s_m \geq 0,\ 1 \leq i \leq n)$ be †algebraically independent over K. The totality of their polynomials over K, which forms a differential ring, is called the **ring of differential polynomials** of the **differential variables** $X_1, \ldots, X_n$ over K, and is denoted by $K\{X_1, \ldots, X_n\}$. Its elements are called **differential polynomials**. For this ring of differential polynomials we have an analog of †Hilbert's basis theorem in the ring of ordinary polynomials, **Ritt's basis theorem**: If we are given any set $\mathfrak{M}$ of differential polynomials of $X_1, \ldots, X_n$ over K, we can choose a finite number of differential polynomials $P_1, \ldots, P_r$ from $\mathfrak{M}$ such that each element Q of $\mathfrak{M}$ has an integral power Q^g equal to a linear combination of $P_1, \ldots, P_r$ and their derivatives, where the coefficients of the linear combination are elements of $K\{X_1, \ldots, X_n\}$. This theorem implies that in the ring of differential polynomials, every semiprime differential ideal can be expressed as the intersection of a finite number of prime differential ideals; if this expression is †irredundant, it is unique (→ 67 Commutative Rings).

The equation obtained by equating a dif-

ferential polynomial to zero is called an **algebraic differential equation**. Concerning these equations, we are able to use methods similar to those used in †algebraic geometry in studying the usual algebraic equations. J. Ritt made interesting studies on solutions of algebraic differential equations by such methods, principally in the case when the ground field K consists of †meromorphic functions.

Since that time, basic study of differential rings and fields has been fairly well organized and has developed into theories such as the following two:

(1) **Picard-Vessiot theory**. This is a classical theory of †linear homogeneous differential equations originated by E. Picard and E. Vessiot; it resembles the †Galois theory concerning algebraic equations. The Galois group in this case is a linear group, and its structure characterizes the solution of the differential equation. E. Kolchin introduced the general concept of the **Picard-Vessiot extension field** of a differential field and studied in detail the group of **differential automorphisms** (i.e., the group of all those automorphisms that commute with the differentiations and fix elements of the ground field), thus making the classical theory more precise and more general.

(2) **Galois theory of differential fields**. Generalizing the concept of the Picard-Vessiot extension, Kolchin introduced the notion of the **strongly normal extension field** and established the Galois theory for such extensions. In this theory, we see that the Galois group is an †algebraic group relative to a †universal domain over the field of constants of the ground field. Conversely, every algebraic group is the Galois group of a strongly normal extension field. We also see that a strongly normal extension can be decomposed, in a certain sense, into a Picard-Vessiot extension and an Abelian extension (i.e., an extension whose Galois group is an †Abelian variety) (→ Kolchin [2–5], Okugawa [6]).

References

[1] I. Kaplansky, An introduction to differential algebra, Actualités Sci. Ind., Hermann, 1957.
[2] E. R. Kolchin, Algebraic matric groups and the Picard-Vessiot theory of homogeneous ordinary linear differential equations, Ann. Math., (2) 49 (1948), 1–42.
[3] E. R. Kolchin, Galois theory of differential fields, Amer. J. Math., 75 (1953), 753–824.
[4] E. R. Kolchin, On the Galois theory of differential fields, Amer. J. Math., 77 (1955), 868–894.
[5] E. R. Kolchin, Abelian extensions of differential fields, Amer. J. Math., 82 (1960), 779–790.
[6] K. Okugawa, Basic properties of differential fields of an arbitrary characteristic and the Picard-Vessiot theory, J. Math. Kyoto Univ., 2 (1963), 295–322.
[7] J. F. Ritt, Differential algebra, Amer. Math. Soc. Colloq. Publ., 1950.

114 (IX.18) Differential Topology

A. General Remarks

Differential topology can be defined as the study of those properties of †differentiable manifolds that are invariant under †diffeomorphisms. The basic objects studied in this field are the topological, combinatorial, and differentiable structures of manifolds and the relationships among them. Some of the remarkable contributions, such as H. Whitney's embedding theorem [1], the triangulation theorems of J. H. C. Whitehead and S. S. Cairnes [2, 3], and Morse theory [4], were made in the 1930s. In the late 1950s, outstanding results were obtained by R. Thom, J. W. Milnor, S. Smale, M. Kervaire, and F. Hirzebruch, among others. Differential topology thus became a new, fascinating branch of mathematics.

B. Differentiable Structures

We assume that all manifolds are †paracompact. Let M be a †topological manifold. A C^r-equivalence class of †atlases of class C^r $(1 \leq r \leq \infty)$ on M is called a **C^r-structure** on M. Any C^r-structure on M contains an atlas of class C^∞ on M, and its C^∞-equivalence class is uniquely determined (Whitney [1]). This C^∞-structure is called a **differentiable structure** on M **compatible** with the given C^r-structure. Moreover, any C^∞-structure admits a †real analytic structure compatible with it in this sense (Whitney [1]). A C^∞-manifold is also called a **smooth manifold**, and a differentiable structure a **smooth structure**. Let D_0, D_1 be differentiable structures on M. If two C^∞-manifolds (M, D_0) and (M, D_1) are not †diffeomorphic, we say that the differentiable structures D_0, D_1 are **distinct** (→ 105 Differentiable Manifolds). Milnor defined a certain invariant of differentiable structure using the Hirzebruch index theorem (→ 56 Characteristic Classes) and proved that there are several distinct differentiable structures on the 7-dimensional sphere S^7 (Milnor [5]). Milnor's

example of such structures was: Let $f_k: S^3 \to SO(4)$ be the mapping defined by $f_k(\sigma)\tau = \sigma^h \tau \sigma^j$, where k is an odd integer, h, j are integers determined by $h+j=1$, $h-j=k$, and σ, τ are †quaternions of norm 1. Let M_k^7 be the total space of the S^3-bundle over S^4 corresponding to the mapping f_k. This is an oriented closed manifold with the naturally defined differentiable structure. Moreover, for each k, M_k^7 is homeomorphic to the 7-dimensional sphere S^7. But if k, l are odd and $k^2 \not\equiv l^2 \pmod 7$, M_k^7 is not diffeomorphic to M_l^7. Differentiable manifolds (such as M_k^7) that are homeomorphic, but not diffeomorphic, to the natural sphere are called **exotic spheres**. After the discovery made by Milnor, many topological manifolds other than the 7-dimensional sphere and possessing several distinct differentiable structures have been found (N. Shimada, I. Tamura). Moreover, topological manifolds admitting no differentiable structures have been constructed (M. Kervaire [6], Smale, Tamura [7], J. Eells and N. Kuiper, C. T. C. Wall [8]). On the other hand, manifolds of dimension $\leqslant 3$ admit unique differentiable structures.

We can introduce a †Riemannian metric on a C^∞-manifold. Let M_1^k and M_2^n be C^∞-manifolds (k-dimensional and n-dimensional, respectively) and $f: M_1^k \to M_2^n$ be an †immersion. We fix a Riemannian metric on M_2^n. For each point p of M_1^k, let $N_p(f)$ be the linear space of all tangent vectors of M_2^n at $f(p)$ that are orthogonal to the †tangent space of $f(U(p))$, where $U(p)$ is a small open neighborhood of p in M_1^k. Then $\{N_p(f) \mid p \in M_1^k\}$ forms an $(n-k)$-dimensional †vector bundle ν_f over M_1^k, called the **normal bundle** of the immersion f. The equivalence class of ν_f is independent of the choice of Riemannian metric on M_2^n. When f is an †embedding, we denote the submanifold $f(M_1^k)$ by L^k. Let $N_\varepsilon(L^k)$ be the set of all points whose distances (with respect to the Riemannian metric) from L^k are $\leqslant \varepsilon$. If L^k is compact and ε is sufficiently small, $N_\varepsilon(L^k)$ is an n-dimensional submanifold of M_2^n with boundary and is uniquely determined independently of the choice of ε up to diffeomorphism. This submanifold is called a **tubular neighborhood** of L^k in M_2^n and is denoted by $N(L^k)$. The interior of $N(L^k)$ is called an **open tubular neighborhood**. The total space $E(\nu_f)$ of the normal bundle ν_f of the embedding f is diffeomorphic to the open tubular neighborhood of L^k in M_2^n.

C. C^r-Triangulation and the Smoothing Problem

Let (K, f) be a †triangulation of the C^r-manifold M^n $(1 \leqslant r \leqslant \infty)$ $(f: |K| \to M^n)$ satisfying the following conditions: (i) For any closed n-simplex σ of K, $f \mid \sigma: \sigma \to M^n$ is a C^r-mapping; (ii) for any point p of σ, the †Jacobian matrix of $f \mid \sigma$ has rank n at p. Then we say that (K, f) is a **C^r-triangulation** of the C^r-manifold M^n, and the C^r-structure of M^n is **compatible** with the triangulation (K, f). Concerning C^r-triangulations of C^r-manifolds, we have the following results (S. Cairns [2], J. H. C. Whitehead [3]): (i) any C^r-manifold $(1 \leqslant r \leqslant \infty)$ has a C^r-triangulation, and any C^r-triangulation of the boundary can be extended to the whole manifold; (ii) for a C^r-triangulation (K, f) of a C^r-manifold M^n, the triangulated space $(K, f; M^n)$ is a †combinatorial manifold; (iii) for two C^r-triangulations (K_1, f_1) and (K_2, f_2) of a C^r-manifold M^n, the triangulated spaces $(K_1, f_1; M^n)$ and $(K_2, f_2; M^n)$ are combinatorially equivalent.

Conversely, for a combinatorial manifold $(K, f; M^n)$, a differentiable structure on M that is compatible with the triangulation (K, f) is called a **smoothing** of M. The question of whether a smoothing of a combinatorial manifold exists is called the **smoothing problem**. We have some criteria for its solvability. One is expressed in terms of transverse fields, and another in terms of †microbundles (→ 147 Fiber Bundles P). We also have the following method using the theory of †obstruction (J. Munkres [9]; M. W. Hirsch and B. Mazur [10]). Let Γ_k be the group of oriented differentiable structures on the †combinatorial k-sphere (an exact definition of this group is given in Section I). Let K^i be the i-†skeleton of a combinatorial n-manifold $(K, f; M^n)$ and α a smoothing of a neighborhood U of $f(K^i)$ in M^n. Then there exists an obstruction cocycle $C_\alpha \in Z^{i+1}(K, \Gamma_i)$ satisfying the following conditions: (i) if $C_\alpha(\sigma^{i+1}) = 0$ for an $(i+1)$-simplex σ^{i+1} of K, the smoothing α can be extended over a neighborhood of $f(K^i \cup \sigma^{i+1})$, and vice versa; (ii) if there exists another smoothing α' of a neighborhood of $f(K^i)$ that coincides with the smoothing α of a neighborhood of $f(K^{i-1})$, then $C_{\alpha'} - C_\alpha$ is a †coboundary; conversely, for any $(i+1)$-coboundary b with coefficients in Γ_i, there exists a smoothing α' of a neighborhood of $f(K^i)$ such that $b = C_{\alpha'} - C_\alpha$. Using this method we can prove that for $n \leqslant 7$ all combinatorial manifolds of dimension n are smoothable.

D. Embedding and Immersion Theorems

In the following, M^n and X^m are C^∞-manifolds of dimensions n and m, respectively. Two †immersions $f_0, f_1: M^n \to X^m$ are said to be **regularly homotopic** if there exists a †homotopy $f_t: M^n \to X^m$, $0 \leqslant t \leqslant 1$, such that f_t is an immer-

sion for each t and the induced mapping (†differential of f_t) $df_t: T(M^n) \to T(X^m)$ on the tangent spaces naturally gives rise to a continuous mapping over $T(M^n) \times I$. Two †embeddings are **isotopic** if they are regularly homotopic and the homotopy f_t is an embedding for each t. Given M^n and X^m, a fundamental problem of embedding (immersion) theory is to classify the embeddings (immersions) of M^n in X^m according to their isotopy (regular homotopy) classes. Whitney proved that a continuous mapping $f: M^n \to X^m$ can be approximated by an immersion if $m \geqslant 2n$ and by an embedding if $m \geqslant 2n+1$ [1]. The following results are also due to Whitney [1]: Any two immersions of M^n in X^m that are homotopic are regularly homotopic if $m \geqslant 2n+2$, and any two embeddings of M^n in X^m that are homotopic are isotopic if $m \geqslant 2n+3$. The range $m \geqslant 2n+3$ is called the **stable range** of embeddings.

In [11, 12] (1944) Whitney improved his classical theorems and showed that M^n can always be immersed in $\mathbf{R}^{2n-1}$ for $n > 1$ and M^n can always be embedded in $\mathbf{R}^{2n}$. The methods used in his proof have played an important role in the subsequent development. Classification of immersions of the n-sphere S^n in $\mathbf{R}^m$ was determined by Smale [13]. Let $p: T(M^n) \to M^n$ and $p': T(X^m) \to X^m$ be the projections of the tangent bundles of M^n and X^m, respectively. A mapping $\varphi: T(M^n) \to T(X^m)$ is a **linear fiber mapping (linear fiber map)** if, for each $x \in M^n$, $\varphi(p^{-1}(x))$ is contained in a fiber of $T(X^m)$ and $\varphi \mid p^{-1}(x)$ is a linear mapping of rank n. A **linear homotopy** $\varphi_t: T(M^n) \to T(X^m)$ $(0 \leqslant t \leqslant 1)$ is a homotopy such that each φ_t is a linear fiber mapping. The following theorem of Hirsch [14] is fundamental to immersion theory: Assume $n < m$. If $\varphi: T(M^n) \to T(X^m)$ is a linear fiber mapping, the mapping $\bar{\varphi}: M^n \to X^m$ induced by φ can be approximated by an immersion $f: M^n \to X^m$ such that df and φ are linearly homotopic. Two immersions $f, g: M^n \to X^m$ are regularly homotopic if and only if df and dg are linearly homotopic. If M^n is immersible in $\mathbf{R}^{m+r}$, where $m > n$ with r linearly independent fields of †normal vectors, then M^n is immersible in $\mathbf{R}^m$. In particular, if M^n is a π-manifold (→ Section I), M^n is immersible in $\mathbf{R}^{n+1}$.

Let $\mathscr{E}(M^n)$ be the set of isomorphism classes of real †vector bundles over M^n, and consider $\theta: \mathscr{E}(M^n) \to KO(M^n)$ (→ 237 K-Theory). An element $\xi \in KO(M^n)$ is said to be **positive** if ξ is in the image of θ. If $\xi_0 \in \widetilde{KO}(M^n)$, the **geometric dimension** of ξ_0, written $g(\xi_0)$, is the least integer k such that $\xi_0 + k$ is positive. Then Hirsch's theorem [14] can be expressed as follows: M^n is immersible in $\mathbf{R}^{n+k}$ $(k > 0)$ if and only if $g(n - \tau(M^n)) \leqslant k$, where $\tau(M^n)$ is the tangent bundle of M^n.

A. Haefliger [15] obtained the following important result: Let M^n be compact and †$(k-1)$-connected, and let X^m be k-connected. Then any continuous mapping of M^n in X^m is homotopic to an embedding if $2k < n$, $m \geqslant 2n - k + 1$, and any two homotopic embeddings of M^n in X^m are isotopic if $2k < n+1$, $m \geqslant 2n - k + 2$. Thus if $m > 3(n+1)/2$, any two embeddings of S^n in $\mathbf{R}^m$ are isotopic. The range $m > 3(n+1)/2$ is called the **metastable range**. Haefliger further classified the embeddings of S^{4n-1} in $\mathbf{R}^{6n}$ and showed the existence of embeddings of S^{4n-1} in $\mathbf{R}^{6n}$ that are not isotopic to the natural one. More complete results for the classification of embeddings of †homotopy n-spheres in S^m were obtained by J. Levine [16]. Furthermore, Levine proved the following unknotting theorem in higher-dimensional knot theory [16]: Let $f: S^{n-2} \to S^n$ be a C^∞-embedding for $n \geqslant 6$; then $f(S^{n-2})$ is unknotted if and only if $S^n - f(S^{n-2})$ is homotopy equivalent to S^1 (→ 235 Knot Theory G; also 65 Combinatorial Manifolds D).

We list some results about embeddings and immersions. If M^n is noncompact, M^n can always be embedded in $\mathbf{R}^{2n-1}$; if M^n is a noncompact π-manifold, M^n can always be immersed in $\mathbf{R}^n$; if M^n is compact and orientable and $n > 4$, M^n can always be embedded in $\mathbf{R}^{2n-1}$.

E. Nonembedding and Nonimmersion Theorems

We denote the †total Stiefel-Whitney class of M^n by $w(M^n)$ and the †total Pontryagin class of M^n by $p(M^n)$ (→ 56 Characteristic Classes). Then $(w(M^n))^{-1}$ $(\in H^*(M^n; \mathbf{Z}_2))$ and $(p(M^n))^{-1}$ $(\in H^*(M^n; \mathbf{Z}))$ can be written as $\bar{w}(M^n) = \sum \bar{w}_i(M^n)$ $(\bar{w}_i \in H^i(M^n; \mathbf{Z}_2))$ and $\bar{p}(M^n) = \sum \bar{p}_i(M^n)$ $(\bar{p}_i \in H^{4i}(M^n; \mathbf{Z}))$. Then the property of characteristic classes for the †Whitney sum implies the following theorem: If M^n can be immersed in $\mathbf{R}^{n+k}$, then $\bar{w}_i(M^n) = 0$ for $i > k$ and $\bar{p}_i(M^n) = 0$ for $i > [k/2]$. Furthermore, if M^n can be embedded in $\mathbf{R}^{n+k}$, then $\bar{w}_k(M^n) = 0$. As an application, these results yield the nonembedding (nonimmersion) theorem for projective spaces (→ Appendix A, Table 6.VII). Sharper theorems were obtained subsequently. In particular, Atiyah proved the following: Let λ^i $(i = 0, 1, \ldots)$ be †exterior power operations (→ 237 K-Theory), and let γ^i be the operations defined by the formal power series $\sum_{i=0}^{\infty} \gamma^i t^i = (\sum_{i=0}^{\infty} \lambda^i t^i)(1-t)^{-1}$. Then $\gamma^i(n - \tau(M^n)) = 0$ for $i > k$ $(i \geqslant k)$ if M^n can be immersed (embedded) in $\mathbf{R}^{n+k}$. Furthermore, we have an interesting result for the differentiable case. For any positive integer q, there exists a differentiable

manifold M^n such that M^n is immersible in $\mathbf{R}^k$ but not embeddable in $\mathbf{R}^{k+q}$.

F. Handlebodies

Let W be a C^∞-manifold with boundary ∂W. Then the boundary ∂W has a neighborhood U that is homeomorphic to $\partial W \times \mathbf{R}_+$, where $\mathbf{R}_+ = [0, \infty)$. Let W_1, W_2 be C^∞-manifolds with boundaries and $f: \partial W_1 \rightarrow \partial W_2$ be a diffeomorphism. Then the quotient space M of $W_1 \cup W_2$ obtained by identifying points corresponding under f has a natural differentiable structure. This construction of the C^∞-manifold M from the C^∞-manifolds W_1, W_2 is called **pasting together the boundaries**.

Now we consider the †topological product $W_1 \times W_2$ of manifolds W_1, W_2 with boundaries ∂W_1, ∂W_2. $W_1 \times W_2 - \partial W_1 \times \partial W_2$ has a natural atlas of class C^∞. By introducing a suitable atlas of class C^∞ in a neighborhood of $\partial W_1 \times \partial W_2$ in $W_1 \times W_2$, we obtain a C^∞-manifold with boundary homeomorphic to $W_1 \times W_2$. More generally, let M be an n-manifold with boundary, let N be a finite union of submanifolds of dimension $\leqslant n-2$ in ∂M, and suppose that M has a corner along N. Extension of the C^∞-structure of $M - N$ over M as in this paragraph is called **straightening the angle**.

Let D^n be the oriented unit †n-disk in the n-dimensional Euclidean space $\mathbf{R}^n$, M_1^n and M_2^n be oriented compact C^∞-manifolds, and $f_i: D^n \rightarrow M_i^n$ $(i=1,2)$ be †C^∞-embeddings, f_1 †orientation-preserving and f_2 †orientation-reversing. Then, pasting together the boundaries of $M_1^n - \operatorname{Int} f_1(D^n)$ and $M_2^n - \operatorname{Int} f_2(D^n)$ by the mapping $f_2 \circ f_1^{-1}$, we obtain an oriented C^∞-manifold, called the **connected sum** of M_1^n and M_2^n and denoted by $M_1^n \# M_2^n$. The connected sum $M_1^n \# M_2^n$ has the orientation induced from those of M_1^n and M_2^n, and its differentiable structure is uniquely determined independently of the mappings f_i. Let S^n be the natural n-dimensional sphere. Then we have $M^n \# S^n \approx M^n$ (here $\approx$ means diffeomorphic), $(M_1 \# M_2) \# M_3 \approx M_1 \# (M_2 \# M_3)$, $M_1 \# M_2 \approx M_2 \# M_1$.

Let M^n be a manifold with boundary and $f: (\partial D^s) \times D^{n-s} \rightarrow \partial M^n$ be a C^∞-embedding. Then we call the quotient space $X(M^n; f; s)$ of $M^n \cup (D^s \times D^{n-s})$ obtained by the identification of corresponding points under f the **manifold with a handle attached by** f. Also, we call the construction of $X(M^n; f; s)$ from M^n **attaching a handle** and call $D^s \times D^{n-s}$ an **s-handle**. After straightening the angle, $X(M^n; f; s)$ is considered naturally a C^∞-manifold with boundary. Let $f_i: \partial D_i^s \times D_i^{n-s} \rightarrow \partial M^n$ $(i=1, \ldots, k)$ be embeddings whose images are mutually disjoint. Then similarly, using embeddings $f_1, \ldots, f_k$, we can define a C^∞-manifold with handles $X(M^n; f_1, \ldots, f_k; s)$. In particular, $X(X(\ldots(X(X(D^n; f_1; s_1); f_2; s_2) \ldots); f_j; s_j))$ is called a **handlebody**.

Let M^{n-1} be an oriented $(n-1)$-manifold and $f: \partial D^s \times D^{n-s} \rightarrow (M^{n-1} - \partial M^{n-1})$ an orientation-preserving C^∞-embedding. Then, by straightening the angle, the quotient space of $(M^{n-1} - \operatorname{Int} f(\partial D^s \times D^{n-s})) \cup (D^s \times \partial D^{n-s})$ obtained by the identification of points corresponding under $f \mid \partial D^s \times \partial D^{n-s}$ is a C^∞-manifold $\chi(M^{n-1}; f)$. We say this manifold is obtained by a **spherical modification** (or **surgery**) of type $(s, n-s)$ from M^{n-1}. The manifold $\chi(M^{n-1}; f)$ has a natural orientation, and $\partial \chi(M^{n-1}; f) = \partial M^{n-1}$. The process of spherical modification has the following relation to that of attaching a handle: Let W be an n-dimensional manifold and $f: \partial D^s \times D^{n-s} \rightarrow \partial W$ an embedding. Then $\partial X(W; f; s) = \chi(\partial W; f)$. When $W = M^{n-1} \times [0,1]$ and $f: \partial D^s \times D^{n-s} \rightarrow M \times \{1\}$, $\partial X(W; f; s) = \chi(M \times \{1\}; f) \cup M \times \{0\}$, and therefore M is cobordant (→ Section H) to $\chi(M; f)$. Conversely, let M_1 be cobordant to M_2. Then we can obtain M_2 from M_1 by a finite sequence of spherical modifications (A. Wallace [17], Milnor). Let M^n be an n-dimensional C^∞-manifold and $f: M^n \rightarrow \mathbf{R}^1$ be a C^∞-function. If f satisfies the following conditions, then it is called a **nice function**: (i) All †critical points of f are †nondegenerate; (ii) for any critical point p of f, the †index (→ 279 Morse Theory) of f at p is equal to $f(p)$. We have the following theorems.

1. Let M be a compact C^∞-manifold. Then there exists a nice function on M (M. Morse, Smale [18], Wallace [17]).

2. Let M be a compact C^∞-manifold and $f: M \rightarrow \mathbf{R}^1$ a C^∞-function all of whose critical points are nondegenerate. Suppose that the number of critical points on $f^{-1}[-\varepsilon, \varepsilon]$ is k and that they are all contained in $f^{-1}(0)$. Suppose further that the indices of these critical points are all equal to s. Then $f^{-1}(-\infty, \varepsilon]$ is diffeomorphic to the manifold with handles $X(f^{-1}(-\infty, -\varepsilon]; f_1, \ldots, f_k; s)$ (Morse, Thom, Smale [18]).

3. †Generalized Poincaré conjecture. Let M^n be an n-dimensional †homotopy sphere of class C^∞ $(n \geqslant 5)$. Then M^n can be obtained by pasting together the boundaries of two n-disks. Consequently, M^n is homeomorphic to the n-dimensional sphere S^n (Smale [18], H. Yamasuge [19]).

4. Let M^n be a †contractible compact n-dimensional manifold $(n > 5)$, with ∂M connected and simply connected. Then M^n is diffeomorphic to the n-disk D^n (Smale [18]).

5. Let M_1^n, M_2^n be oriented, compact, simply connected n-dimensional manifolds $(n > 4)$. If M_1 is h-cobordant (→ Section I) to M_2, then M_1 is diffeomorphic to M_2 (**h-cobordism theorem**; Smale [20]).

6. If $M_1^n \times \mathbf{R}^k$ is diffeomorphic to $M_2^n \times \mathbf{R}^k$, we say that M_1^n is **k-equivalent** to M_2^n. Let M_1, M_2 be compact n-dimensional manifolds of the same †homotopy type. Then M_1 is k-equivalent to M_2 if and only if for a †homotopy equivalence $f: M_1 \rightarrow M_2$, $\tau(M_1) \oplus \varepsilon_k$ is equivalent to $f^*\tau(M_2) \oplus \varepsilon_k$, where $\tau(M_i)$ is the †tangent bundle of M_i and ε_k the †trivial vector bundle of dimension k (Mazur).

7. In some special cases the classification of manifolds by diffeomorphism is completely determined. (i) The classification of simply connected 5-dimensional manifolds M with vanishing second †Stiefel-Whitney class $w_2(M)$ is determined by $H_2(M)$. (ii) A 2-connected compact 6-dimensional manifold is diffeomorphic to either S^6 or the connected sum of a finite number of copies of $S^3 \times S^3$ (Smale [20]). Besides these, the classifications of $(n-1)$-connected $2n$-dimensional manifolds (Wall [8]) and $(n-1)$-connected $(2n+1)$-dimensional manifolds (Tamura [21], Wall [22]) have been obtained.

G. Thom Complexes

Let ξ be a real n-dimensional vector bundle over a paracompact space X, A_ξ be the total space of the associated bundle of ξ with fiber the closed n-disk D^n, and $\dot{A}_\xi$ be the total space of the associated bundle of ξ with fiber the $(n-1)$-sphere ∂D^n. Then the quotient space $X_\xi = A_\xi/\dot{A}_\xi$, obtained from A_ξ by contracting $\dot{A}_\xi$ to a point, is called the **Thom space** of the vector bundle ξ. If X is a †CW-complex, then the Thom space X_ξ of ξ has the homotopy type of a CW-complex and is called a **Thom complex**. The Thom space X_ξ has the canonical base point p_ξ corresponding to $\dot{A}_\xi$.

For a coefficient group G, denote by G_ξ the †local system of coefficient groups with stalk G associated with the †orientation sheaf of an $\mathbf{R}^n$-bundle ξ. Then we have the **Thom-Gysin isomorphism**:

$$H^q(X; G_\xi) \cong H^{n+q}(X_\xi, p_\xi; G),$$

$$H_q(X; G_\xi) \cong H_{n+q}(X_\xi, p_\xi; G).$$

Let G be a closed subgroup of the orthogonal group $O(n)$ and BG be the base space of the universal n-dimensional vector bundle γ_G with structure group G. Then we can take a connected CW-complex as BG. We denote the Thom complex of the vector bundle γ_G by MG and call it the **Thom complex associated with** (G, n). The Thom complex associated with (G, n) is $(n-1)$-†connected. If G is connected, then we have $\pi_n(MG) \cong \mathbf{Z}$, $\pi_n(MO(n)) \cong \mathbf{Z}_2$, $H^n(MG; \mathbf{Z}) \cong \mathbf{Z}$, and $H^n(MO(n); \mathbf{Z}_2) \cong \mathbf{Z}_2$ ($\mathbf{Z}_2$ stands for the quotient group $\mathbf{Z}/2\mathbf{Z}$). The generator U of $H^n(MG; \mathbf{Z})$ is called the **fundamental (cohomology) class of the Thom complex** MG. For a general Thom space X_ξ, we can also define the fundamental class of X_ξ using the Thom-Gysin isomorphism.

A C^∞-submanifold W^p of a compact manifold V^n is a †support of a †singular cycle and represents a homology class $z \in H_p(V^n; G)$ ($G = \mathbf{Z}_2$ or $\mathbf{Z}$). In this case we say that the homology class z of V^n is **realizable** by a submanifold. A homology class $z \in H_{n-k}(V^n; G)$ is realizable by a submanifold if and only if there exists a mapping $f: V^n \rightarrow MO(k)$ (or $MSO(k)$) such that $f^*(U)$ is the dual cohomology class of z (Thom [23]).

The homotopy group $\pi_{n+k}(MO(n))$ (resp. $\pi_{n+k}(MSO(n))$ $(n > k)$ is determined only by k for any n up to isomorphism. It is called the k-dimensional **stable homotopy group of the Thom spectrum** $MO = \{MO(n)\}$ ($MSO = \{MSO(n)\}$) and is denoted by $\pi_k(MO)$ ($\pi_k(MSO)$). For the †unitary group $U(n)$ and the †symplectic group $Sp(n)$, we define Thom complexes $MU(n)$ and $MSp(n)$ as the Thom complexes associated with $(U(n), 2n)$ and $(Sp(n), 4n)$ by the canonical inclusions $U(n) \subset O(2n)$ and $Sp(n) \subset O(4n)$, respectively. The k-dimensional stable homotopy groups $\pi_k(MU) = \lim \pi_{2n+k}(MU(n))$, $\pi_k(MSp) = \lim \pi_{4n+k}(MSp(n))$ of Thom spectra $MU = \{\ldots, MU(n), SMU(n), MU(n+1), \ldots\}$, $MSp = \{\ldots, MSp(n), SMSp(n), S^2MSp(n), S^3MSp(n), MSp(n+1), \ldots\}$ can be defined similarly to the case of $\pi_k(MO)$, where $SMU(n)$ denotes the †reduced suspension of $MU(n)$. The stable homotopy groups of Thom spectra are calculated in connection with the cohomology groups utilizing the Thom-Gysin isomorphism (→ 202 Homotopy Theory T) (Thom [23], Milnor [24]).

H. Cobordism

Cobordism theory is a theory of classification of differentiable manifolds initiated by L. S. Pontryagin and V. A. Rokhlin, who called it **intrinsic homology**. The theory was brought to maturity by Thom [23]. Its fundamental problem is to determine whether a given compact manifold is the boundary of another manifold. Corresponding cobordism theories for combinatorial and topological manifolds are being developed (Wall, R. Williamson).

We consider only compact C^∞-manifolds that are not necessarily connected. Let $\mathfrak{D}$ be the set of all diffeomorphism classes of C^∞-manifolds, and let $\mathfrak{D}_0$ be the set of all orientation-preserving diffeomorphism classes of oriented C^∞-manifolds. For an oriented manifold V, we write $-V$ for the manifold with reversed orientation.

For two compact k-manifolds V, $W \in \mathfrak{D}_0$, we say that V is **cobordant** to W if there exists a compact $(k+1)$-manifold $X \in \mathfrak{D}_0$ such that $\partial X = V \cup (-W)$. In this definition, considering $\mathfrak{D}$ instead of $\mathfrak{D}_0$, we say that V is **cobordant** to W **mod** 2. The equivalence class of V^k with respect to the cobordism relation (the cobordism relation mod 2) is called an **oriented (unoriented) cobordism class** and is denoted by $[V^k]$ ($[V^k]_2$). The set of all oriented (unoriented) cobordism classes $[V^k]$ ($[V^k]_2$) of k-manifolds forms an Abelian group Ω_k ($\mathfrak{N}_k$) by the natural addition $[V^k]+[W^k]=[V^k \cup W^k]$ ($[V^k]_2+[W^k]_2=[V^k \cup W^k]_2$). This is called the k-dimensional **oriented (unoriented) cobordism group**. Define the product $[V^k]\times[W^l]=[V^k \times W^l]$ ($[V^k]_2 \times [W^l]_2 = [V^k \times W^l]_2$). Then the direct sum $\Omega = \sum \Omega_k$ ($\mathfrak{N} = \sum \mathfrak{N}_k$) forms an anticommutative (commutative) †graded algebra, which is called the **cobordism ring** or **Thom algebra**. We have the following theorems.

1. Ω_k, $\mathfrak{N}_k$ are isomorphic to the stable homotopy groups $\pi_k(MSO)$, $\pi_k(MO)$ of the Thom spectra MSO, MO, respectively (**Thom's fundamental theorem**) [23].

2. For a natural number k not of the form 2^i-1, there exists a compact manifold $P(k)$, and $\mathfrak{N}$ is a polynomial ring over $\mathbf{Z}_2$ with generators $\{[P(k)]_2 \mid k \neq 2^i-1\}$. $\Omega \otimes \mathbf{Q}$ ($\mathbf{Q}$ is the field of rational numbers) is a polynomial ring over $\mathbf{Q}$ with generators $\{[PC^{2m}] \mid m \geqslant 1\}$, where PC^{2m} is the complex $2m$-dimensional projective space (Thom [23]). Moreover, Milnor [24] proved that the p-component of Ω_k is zero for an odd prime p. Wall proved that the 2-component of Ω_k contains no element of order 4, using a certain exact sequence which contains the natural homomorphism $\Omega_k \rightarrow \mathfrak{N}_k$ [25].

3. Let T be the ideal of Ω consisting of all torsion elements in Ω. Then Ω/T is a polynomial ring over $\mathbf{Z}$ with generators $\{[Y_{4k}] \mid k \geqslant 1\}$, where we can take for Y_{4k} a complex $2k$-dimensional nonsingular algebraic variety (Milnor).

†Characteristic numbers are invariants of cobordism classes. Combining the results stated in 2, we have the following theorem:

4. Let V, W be manifolds. V is cobordant to W mod 2 if and only if they have the same corresponding †Stiefel-Whitney numbers. Let V, W be oriented manifolds. V is cobordant to W if and only if they have the same corresponding Stiefel-Whitney numbers and †Pontryagin numbers. The †index of an oriented $4k$-dimensional compact manifold is a cobordism class invariant and can be expressed as a linear combination with rational coefficients of †Pontryagin numbers (→ 56 Characteristic Classes).

A manifold whose †stable tangent bundle has a complex structure is called a **stably** (or **weakly**) **almost complex manifold**. Let V, W be 2n-dimensional compact stably almost complex manifolds ($n \geqslant 1$). We say that V is **C-equivalent** to W if they have the same †Chern numbers. The set of all C-equivalence classes $[V]_C$ forms the **complex cobordism group** $\mathfrak{U}_n$ as in the (real) case of cobordism groups (the existence of the inverse of an element is not trivial). Introducing multiplication into the direct sum $\mathfrak{U}_* = \sum \mathfrak{U}_n$ by the product of manifolds, we obtain the **complex cobordism ring**. We have $\mathfrak{U}_n \cong \pi_{2n}(MU)$ and $\mathfrak{U}_*$ is a polynomial ring over $\mathbf{Z}$ with generators $\{[Y'_{2k}]_C \mid k \geqslant 1\}$, where we can take for Y'_{2k} a complex k-dimensional nonsingular algebraic variety (Milnor [24]).

I. h-Cobordism Groups of Homotopy Spheres

A manifold M is said to be **parallelizable** if its tangent bundle $\tau(M)$ is trivial, and **almost parallelizable** if there exists a finite number of points x_i in M such that $M - \bigcup\{x_i\}$ is parallelizable. A manifold M is called **stably parallelizable** (or **s-parallelizable**) if the †Whitney sum $\tau \oplus \varepsilon_1$ of its tangent bundle $\tau(M)$ and the trivial line bundle ε_1 is trivial. A manifold M^n is called a **π-manifold** if M^n has a trivial normal bundle when it is embedded into a Euclidean space $\mathbf{R}^N$ ($N > 2n$). A manifold M^n is a π-manifold if and only if M^n is s-parallelizable. The concepts defined in this paragraph are related by inclusions as follows: parallelizable $\subsetneqq$ s-parallelizable $\subsetneqq$ almost parallelizable. For a connected manifold with boundary, these three concepts are equivalent. †Group manifolds are parallelizable. An n-dimensional manifold homeomorphic to the n-dimensional sphere is parallelizable if and only if $n=1, 3, 7$ (Milnor [26]). Suppose that we are given an $(n-1)$-dimensional sphere S^{n-1} (n even). We can consider the problem of determining the maximal number r such that there exists a tangent r-frame field over S^{n-1}. J. F. Adams solved this problem using †K-theory (→ 237 K-Theory) as follows: Let $n=(2a+1)2^b$, $b=c+4d$, where a, b, c, d are integers and $0 \leqslant c \leqslant 3$. Put $\rho(n)=2^c+8d$. Then $r=\rho(n)-1$. On the other hand, homotopy spheres are π-manifolds [27].

Let V_1, V_2 be oriented compact manifolds. Then we say that V_1 is **h-cobordant** to V_2 if there exists an oriented manifold W with boundary $\partial W = V_1 \cup (-V_2)$ such that V_i $(i=1,2)$ is a deformation retract of W. The set of all h-cobordism classes of oriented homotopy n-spheres forms an Abelian group θ_n with connected sum as addition. This is called the **(h-cobordism) group of homotopy n-spheres**. We

denote by $\theta_n(\partial\pi)$ the subgroup of homotopy n-spheres that are boundaries of π-manifolds. Kervaire and Milnor gave certain exact sequences that contain the groups θ_n, the stable homotopy groups G_n of spheres, and the stable homotopy groups $\pi_n(SO)$. They clarified relations among these groups and proved that $\theta_n(\partial\pi)=0$ for n even, $\theta_n(\partial\pi)=$ finite group for n odd $\neq 3$, $\theta_n/\theta_n(\partial\pi)\subset \mathrm{Coker}\, J_n$, etc. [27], where $J_n:\pi_n(SO)\rightarrow G_n$ is the †J-homomorphism. From these results it follows that the θ_n ($n\neq 3$) are finite Abelian groups, $\theta_n=0$ for $n<7$, $\neq 3$, $\theta_7\cong \mathbf{Z}_{28}$, etc. (→ Appendix A, Table 6.I).

By pasting together the boundaries of two n-disks D_1^n and D_2^n, we obtain an oriented manifold which is considered as a smoothing of the combinatorial n-sphere. The set of all orientation-preserving diffeomorphism classes of smoothings of the combinatorial n-sphere forms an Abelian group Γ_n with connected sum as addition. This is called the **group of oriented differentiable structures on the combinatorial sphere**. By the generalized Poincaré conjecture and the h-cobordism theorem, we obtain $\Gamma_n=\theta_n$ ($n\neq 3,4$). Furthermore, $\Gamma_3=0$ and $\Gamma_4=0$ (J. Cerf [28]). The group Γ_n can also be defined as follows: Let $\mathrm{Diff}^+ D^n$, $\mathrm{Diff}^+ S^{n-1}$ be the **groups of orientation-preserving diffeomorphisms** of D^n, S^{n-1}, respectively, where multiplication is defined by composition. Let $r:\mathrm{Diff}^+ D^+\rightarrow \mathrm{Diff}^+ S^{n-1}$ be the homomorphism induced by the restriction $D^n\rightarrow S^{n-1}$. Then the image of r is a normal subgroup, and $\Gamma_n=\mathrm{Diff}^+ S^{n-1}/r\,(\mathrm{Diff}^+ D^n)$.

J. Surgery Theory

A process of modifying a manifold into another by a sequence of spherical modifications is called **surgery** on the manifold (→ Section F). The technique of surgery proved to be a powerful tool for the development of differential topology in the 1960s. Kervaire and Milnor exploited this technique in their study of homotopy spheres [27]. W. Browder [29] and S. P. Novikov [30] used surgery to construct differentiable manifolds with the same homotopy type as that of a given Poincaré complex in dimension greater than 4.

To explain the main points of surgery theory, we introduce some terminology. A pair of finite †CW complexes (X,Y) is called a **Poincaré pair of formal dimension** n if there exists a class $\mu\in H_n(X,Y;\mathbf{Z})$ called the **fundamental class** such that the †cap product $\mu\frown$: $H^q(X;\mathbf{Z})\rightarrow H_{n-q}(X,Y;\mathbf{Z})$ is an isomorphism for each q. When $Y=\varnothing$, X is called a **Poincaré complex**. Let M be an n-dimensional smooth manifold. Consider an embedding g of M into a Euclidean space $\mathbf{R}^{n+k}$, where k is large enough (i.e., $k\geqslant n+2$). Then the isomorphism class of the †normal bundle of g is independent of the choice of embedding, and it depends only on M. Any representative of the isomorphism class is called the **normal k-vector bundle** of M and is denoted by ν_M^k. Let (X,Y) be a Poincaré pair of formal dimension n, and let ξ be a real k-vector bundle over X. A **normal mapping (normal map)** (f,b) consists of a degree 1 mapping $f:(M,\partial M)\rightarrow(X,Y)$ and a †bundle mapping $b:\nu_M^k\rightarrow\xi$ which covers f.

A normal mapping $(f,b):(M,\partial M)\rightarrow(X,Y)$ is **normally cobordant** to a normal mapping $(f',b'):(M',\partial M')\rightarrow(X,Y)$ if $\partial M=\partial M'$ and there exist a smooth $(n+1)$-manifold W and a mapping $F:W\rightarrow X$ such that $\partial W=M\cup(-M')\cup\partial M\times I$, F is covered by a bundle mapping $B:\nu_W^k\rightarrow\xi$, and $(F,B)|(M,\partial M)=(f,b)$, $(F,B)|(M',\partial M')=(f',b')$.

The fundamental theorem of surgery theory is formulated as follows: Let (X,Y) be a Poincaré pair of formal dimension n. Suppose that X is simply connected and $n\geqslant 5$. Let (f,b): $(M,\partial M)\rightarrow(X,Y)$ be a normal mapping that restricts to a homotopy equivalence on the boundary $f|\partial M:\partial M\rightarrow Y$. Then (f,b) is normally cobordant to a normal mapping (f',b'): $(M',\partial M')\rightarrow(X,Y)$ with $f':M'\rightarrow X$ a homotopy equivalence if and only if a well-defined obstruction $\sigma(f,b)$ vanishes. $\sigma(f,b)$ is called the **surgery obstruction**. When n is odd, $\sigma(f,b)$ always vanishes. When $n\equiv 0 \bmod 4$, $\sigma(f,b)$ is an integer. If $Y=\varnothing$, it is given by $(I(M)-I(X))/8$, where I( ) denotes the †index of the manifold or of the Poincaré complex. If $n\equiv 2 \bmod 4$, $\sigma(f,b)$ is an integer mod 2, called the **Arf-Kervaire invariant**. For a thorough development of simply connected surgery → [31].

In the PL or even in the topological categories, surgery theory can be developed similarly (Browder and Hirsch, R. C. Kirby and L. C. Siebenmann [32]).

In his study of Hauptvermutung for simply connected manifolds, D. Sullivan reformulated surgery in terms of the "surgery exact sequences" involving the classifying spaces G/PL or G/O. (These spaces are "homotopy theoretic fibers" of $B\mathrm{PL}\rightarrow BG$ or $BO\rightarrow BG$ respectively.) In the special case where X is a closed simply connected PL n-manifold, the surgery exact sequence can be formulated as follows ($n\geqslant 5$):

$$0\rightarrow hT(X)\xrightarrow{\eta}[X,G/\mathrm{PL}]\xrightarrow{\theta}\begin{cases}\mathbf{Z} & (n\equiv 0 \bmod 4),\\ \mathbf{Z}_2 & (n\equiv 2 \bmod 4),\\ 0 & (n \text{ odd}).\end{cases}$$

The set $hT(X)$ is the totality of equivalence classes of pairs (M,f), where M is a closed PL n-manifold and f is a homotopy equivalence $M\rightarrow X$. Two such pairs (M',f') and (M'',f'') are defined to be equivalent if there exists a PL

homeomorphism $h: M' \to M''$ such that $f'' \circ h$ is homotopic to f'. The set of homotopy classes $[X, G/\mathrm{PL}]$ of mappings $X \to G/\mathrm{PL}$ is appropriately identified with the set of all normal cobordism classes of normal mappings with target X or with the set of all PL reductions of the **Spivak normal fiber space** of X as a Poincaré complex [31]. The mapping η can be defined naturally, and the mapping θ corresponds to the assignment of the surgery obstruction. The image $\eta(M, f) \in [X, G/\mathrm{PL}]$ is sometimes called the **normal invariant** of (M, f). Sullivan reduced the classification problem of manifolds within a given homotopy type to a homotopy theoretic problem of the classifying spaces G/PL, G/O [33] (also → [31, 34]).

In the latter half of the 1960s, surgery theory was extended by Wall to cover all compact non-simply connected manifolds which are not necessarily orientable. He introduced a certain Abelian group $L_n(\pi; w)$, now called the **Wall group**, which functorially depends on the fundamental group $\pi = \pi_1(X)$, with the orientation character $w: \pi_1(X) \to \mathbf{Z}_2$, and which is of period 4 with respect to the formal dimension n of X. In Wall's theory, the surgery obstruction $\sigma(f, b)$ takes its value in this group. The group structure of Wall groups has been calculated in many cases. Sullivan's exact sequences are extended to the non-simply connected cases [34]. Making use of the extended exact sequences, Wall and W. C. Hsiang and J. L. Shaneson classifield homotopy tori. The result played an important role in the work of Kirby and Siebenmann on the †annulus conjecture and stable homeomorphisms which led to the solution of †Hauptvermutung and triangulation problems on topological manifolds in 1969 (→ 65 Combinatorial Manifolds). Surgery theory has many applications to other geometric problems. Among them are finding missing boundaries for open manifolds, equivariant surgery, homology surgery [35], and surgery on codimension two submanifolds [36, 37].

K. 4-Dimensional Manifolds

The results in differential topology discussed above are mainly concerned with manifolds of dimension $\geqslant 5$. For 4-manifolds, however, because of their peculiar nature which is not observed in other dimensions, many fundamental problems had remained unsolved until M. H. Freedman's epoch-making paper [54] appeared in 1982. His paper, together with S. K. Donaldson's theorem [55] which was published a little later, was a breakthrough in the theory of 4-manifolds.

It was Rokhlin who first discovered a strange property of 4-manifolds [38]. **Rokhlin's theorem** states: Let M be a closed oriented smooth 4-manifold. If M is almost parallelizable (or equivalently, if M is a spin 4-manifold), then the index of M is divisible by 16. Milnor and Kervaire gave an alternative proof of this theorem from the differential-topological point of view [39]. Freedman and Kirby and Y. Matsumoto gave geometric or elementary proofs.

Until 1981, Rokhlin's theorem had been a constant source of many curious phenomena in 4-dimensional topology. The list contains, for example: (1) the class $(4m+2, 4n+2) \in H_2(S^2 \times S^2) \cong \mathbf{Z} \oplus \mathbf{Z}$ cannot be represented by a smoothly embedded 2-sphere in $S^2 \times S^2$ (Kervaire and Milnor). This result was improved by W.-C. Hsiang and R. H. Szczarba, A. G. Tristram, and Rokhlin. (2) The h-cobordism theorem fails to hold in 4 or 3 dimensions (T. Matumoto, Siebenmann [40]). (3) There exists a closed smooth 4-manifold M that is homotopy equivalent to the real projective space $\mathbf{R}P^4$, but $M \not\cong \mathbf{R}P^4$ (Cappell and Shaneson). (4) There exists an open 4-manifold W^4 properly homotopy equivalent to $S^3 \times \mathbf{R}$ but distinct (Freedman).

Closed, connected, simply connected 4-manifolds M and N are homotopy equivalent if and only if the intersection forms defined on the 2-dimensional (co)homology groups are equivalent as inner product spaces over $\mathbf{Z}$ [41]. Wall [42] proved that if closed simply connected smooth 4-manifolds M and N are homotopy equivalent, then they are h-cobordant. Moreover, if M and N are h-cobordant, then there exists an integer $k \geqslant 0$ such that $M \# k(S^2 \times S^2)$ is diffeomorphic to $N \# k(S^2 \times S^2)$.

About 1973, using a certain infinite repetition process, A. Casson constructed a family of noncompact smooth 4-manifolds which are properly homotopy equivalent to the open 2-handle $D^2 \times \mathbf{R}^2$. A 4-manifold belonging to the family is called a **Casson handle**. He observed that if all the Casson handles are diffeomorphic to $D^2 \times \mathbf{R}^2$, then theories analogous to surgery and the h-cobordism theorem in higher dimensions can also be developed in dimension 4.

In his 1982 paper [54], Freedman proved that each Casson handle is homeomorphic to $D^2 \times \mathbf{R}^2$. (It was proved later that Casson handles are not, in general, diffeomorphic to $D^2 \times \mathbf{R}^2$.) This result and the proper h-cobordism theorem, also due to Freedman, proved many fundamental results on the topological structure of 4-manifolds: (1) If closed simply connected smooth 4-manifolds M, N are h-cobordant, they are homeomorphic. In

particular, a 4-dimensional homotopy 4-sphere is homeomorphic to the 4-sphere S^4. (Proof of the 4-dimensional Poincaré conjecture.) (2) A topological 4-manifold properly homotopy equivalent to $\mathbf{R}^4$ is homeomorphic to $\mathbf{R}^4$. (3) Given a nonsingular symmetric bilinear form ω over $\mathbf{Z}$, there exists a closed, connected, simply connected topological 4-manifold whose intersection form is equivalent to ω. (Therefore Rokhlin's theorem cannot be extended to topological 4-manifolds. From this, it follows that there exist simply connected topological 4-manifolds which are nonsmoothable or even nontriangulable.) (4) The homeomorphism class of a closed connected simply connected 4-manifold is determined by the intersection form and the Kirby-Siebenmann class. (This statement is an improved one by F. Quinn [56].)

Donaldson [55] revealed a sharp contrast existing between smooth and topological 4-manifolds. Donaldson's theorem states: If the intersection form of a closed, connected, simply connected smooth 4-manifold is positive definite, then the form is equivalent to the standard one $\langle 1\rangle \oplus \langle 1\rangle \oplus \ldots \oplus \langle 1\rangle$. This theorem, together with Casson's theory and Freedman's result (2) stated above, implies that there exists an exotic differential structure on a 4-dimensional Euclidean space $\mathbf{R}^4$. Moreover, as an application of Donaldson's theorem, the problem of representing a 2-dimensional homology class of $S^2 \times S^2$ by a smoothly embedded 2-sphere was completely solved (K. Kuga [57]): the class $(p, q) \in H_2(S^2 \times S^2)$ is represented by a smoothly embedded S^2 if and only if $|p| \leqslant 1$ or $|q| \leqslant 1$.

Many problems concerning differential structures on 4-manifolds remain unsolved. It is not known whether an exotic smooth 4-sphere exists.

Among other results, the proof of the 4-dimensional annulus conjecture by Quinn [56] is remarkable.

L. Miscellaneous Results

The Browder-Livesay invariant. Let Σ^n be a homotopy n-sphere, T an **involution** on Σ^n, that is, a smooth mapping $T: \Sigma^n \to \Sigma^n$ with $T \circ T = \mathrm{id}_{\Sigma^n}$. Assume that T is free from fixed points. The involution T is said to **desuspend** if there exists a homotopy $(n-1)$-sphere Σ' smoothly embedded in Σ^n which is invariant under the action of T. Browder and G. R. Livesay defined an obstruction $\sigma(T)$ in the group 0 (n: even), $\mathbf{Z}$ ($n \equiv 3 \bmod 4$), $\mathbf{Z}_2$ ($n \equiv 1 \bmod 4$) such that $\sigma(T)=0$ if and only if T desuspends (provided that $n \geqslant 6$) [43]. The invariant $\sigma(T)$ is now called the **Browder-Livesay invariant**. The technique used there was equivariant surgery under the action of T.

De Rham homotopy theory. A new approach to topology of manifolds was invented by Sullivan [44], who considered the exterior algebra of differential forms (with polynomial coefficients) on a simplicial complex. Through the construction of the **minimal model** for the algebra of differential forms, he recovers the rational homotopy type of the complex under some reasonable condition on the fundamental group. Following upon the classical de Rham theorem which recovers the (real) cohomology algebra from the algebra of differential forms, this approach is called the **de Rham homotopy theory**. This method has proved to be useful for the explicit calculation of the algebraic topological invariants of †Kähler manifolds, †loop spaces, cross-section spaces, †algebraic varieties, etc.

Kirby calculus. Kirby [45] initiated a link-theoretic approach to 3- and 4-manifolds. Let L be a link in S^3. Suppose that each component of L is given a **framing**, which means a trivialization of a tubular neighborhood. Such a link is called a **framed link**. A framing of a component is determined by the †linking number between the component and its replacement along the framing. By attaching 2-handles to the 4-disk D^4 along a framed link in $S^3 = \partial D^4$, we obtain a handlebody; here we denote it by W_L. It is known that each closed connected oriented 3-manifold is obtained as the boundary of such a handlebody (W. B. R. Lickorish, Wallace). Kirby proved that for framed links L and L', the boundaries ∂W_L and $\partial W_{L'}$ belong to the same orientation-preserving diffeomorphism class if and only if L is transformed into L' by a sequence of the following two kinds of elementary operations: (i) adding or deleting a trivial knot (separated from L) with framing ± 1, (ii) band summation of components corresponding to "handle sliding." Kirby's approach is called **Kirby calculus** on framed links. For applications → [45, 53].

References

[1] H. Whitney, Differentiable manifolds, Ann. Math., (2) 37 (1936), 645–680.
[2] S. S. Cairns, Triangulation of the manifold of class one, Bull. Amer. Math. Soc., 41 (1935), 549–552.
[3] J. H. C. Whitehead, On C^1-complexes, Ann. Math., (2) 41 (1940), 809–824.
[4] M. Morse, The calculus of variations in the large, Amer. Math. Soc. Colloq. Publ., 1934.
[5] J. W. Milnor, On manifolds homeomor-

phic to the 7-sphere, Ann. Math., (2) 64 (1956), 399–405.
[6] M. A. Kervaire, A manifold which does not admit any differentiable structure, Comment. Math. Helv., 34 (1960), 257–270.
[7] I. Tamura, 8-manifolds admitting no differentiable structure, J. Math. Soc. Japan, 13 (1961), 377–382.
[8] C. T. C. Wall, Classification of $(n-1)$-connected $2n$-manifolds, Ann. Math., (2) 75 (1962), 163–189.
[9] J. Munkres, Obstructions to the smoothing of piecewise differentiable homeomorphisms, Ann. Math., (2) 72 (1960), 521–554.
[10] M. W. Hirsch and B. Mazur, Smoothings of piecewise linear manifolds, Ann. Math. Studies, Princeton Univ. Press, 1974.
[11] H. Whitney, The self-intersections of a smooth n-manifold in $2n$-space, Ann. Math., (2) 45 (1944), 220–246.
[12] H. Whitney, The singularities of a smooth n-manifold in $(2n-1)$-space, Ann. Math., (2) 45 (1944), 247–293.
[13] S. Smale, The classification of immersions of spheres in Euclidean spaces, Ann. Math., (2) 69 (1959), 327–344.
[14] M. W. Hirsch, Immersions of manifolds, Trans. Amer. Math. Soc., 93 (1959), 242–276.
[15] A. Haefliger, Plongements differéntiables de variétés dans variétés, Comment. Math. Helv., 36 (1961), 47–82.
[16] J. Levine, A classification of differentiable knots, Ann. Math., (2) 82 (1965), 15–50.
[17] A. H. Wallace, Modifications and cobounding manifolds, Canad. J. Math., 12 (1960), 503–528.
[18] S. Smale, Generalized Poincaré's conjecture in dimensions greater than four, Ann. Math., (2) 74 (1961), 391–406.
[19] H. Yamasuge, On the Poincaré conjecture for M^5, J. Math. Osaka City Univ., 12 (1961), 1–17.
[20] S. Smale, On the structure of manifolds, Amer. J. Math., 84 (1962), 387–399.
[21] I. Tamura, Classification des variétés différentiables, $(n-1)$-connexes, sans torsion, de dimension $2n+1$, Sém. H. Cartan 16–19, 1962–1963, Inst. H. Poincaré, Univ. Paris, 1964.
[22] C. T. C. Wall, Classification problems in differential topology, I, II, Topology, 2 (1963), 253–261, 263–272.
[23] R. Thom, Quelques propriétés globales des variétés différentiables, Comment. Math. Helv., 28 (1954), 17–86.
[24] J. W. Milnor, On the cobordism ring Ω^* and a complex analogue I, Amer. J. Math., 82 (1960), 505–521.
[25] C. T. C. Wall, Determination of the cobordism ring, Ann. Math., (2) 72 (1960), 292–311.
[26] J. W. Milnor, Some consequences of a theorem of Bott, Ann. Math., (2) 68 (1958), 444–449.
[27] M. A. Kervaire and J. W. Milnor, Groups of homotopy spheres I, Ann. Math., (2) 77 (1963), 504–537.
[28] J. Cerf, Sur les difféomorphismes de la sphère de dimension trois $(\Gamma_4=0)$, Lecture notes in math. 53, Springer, 1968.
[29] W. Browder, Homotopy type of differentiable manifolds, Colloquium on Algebraic Topology, Aarhus, 1962, 42–46.
[30] S. P. Novikov, Diffeomorphisms of simply connected manifolds (in Russian), Dokl. Akad. Nauk. SSSR, 143 (1962), 1046–1049.
[31] W. Browder, Surgery on simply connected manifolds, Erg. Math., 65 (1972).
[32] R. C. Kirby and L. C. Siebenmann, Foundational essays on topological manifolds, smoothings, and triangulations, Ann. Math. Studies 88, Princeton Univ. Press, 1977.
[33] D. Sullivan, Triangulating and smoothing homotopy equivalences and homeomorphisms, Geometric Topology Seminar Notes, Princeton Univ. Press, 1967.
[34] C. T. C. Wall, Surgery on compact manifolds, Academic Press, 1970.
[35] S. E. Cappell and J. L. Shaneson, The codimension two placement problem and homology equivalent manifolds, Ann. Math., (2) 99 (1974), 277–348.
[36] Y. Matsumoto, Knot cobordism groups and surgery in codimension two, J. Fac. Sci. Univ. Tokyo, 20 (1973), 253–317.
[37] M. H. Freedman, Surgery on codimension 2 submanifolds, Mem. Amer. Math. Soc., vol. 12, no. 191, 1977.
[38] V. A. Rokhlin (Rohlin), New results in the theory of 4-dimensional manifolds (in Russian), Dokl. Akad. Nauk SSSR, 84 (1952), 221–224.
[39] J. W. Milnor and M. A. Kervaire, Bernoulli numbers, homotopy groups, and a theorem of Rohlin, Proc. Int. Congress Math., 1958, Cambridge Univ. Press, 1960, 454–458.
[40] L. C. Siebenmann, Disruption of low-dimensional handle body theory by Rohlin's theorem, Topology of Manifolds, Markham, 1970, 57–76.
[41] J. W. Milnor, On simply connected 4-manifolds, Symp. Int. de Top. Alg. (Mexico 1956), Mexico, 1958, 122–128.
[42] C. T. C. Wall, On simply connected 4-manifolds, J. London Math. Soc., 39 (1964), 141–149.
[43] S. López de Medrano, Involutions on manifolds, Erg. Math., 59, 1971.
[44] D. Sullivan, Infinitesimal computation in topology, Publ. Math. Inst. HES, 47 (1977), 269–331.
[45] R. C. Kirby, A calculus for framed links in S^3, Inventiones Math., 45 (1978), 35–56.

[46] J. Munkres, Elementary differential topology, Ann. Math. Studies, Princeton Univ. Press, 1961.
[47] S. Smale, A survey of some recent developments in differential topology, Bull. Amer. Math. Soc., 69 (1963), 131–145.
[48] C. T. C. Wall, Topology of smooth manifolds, J. London Math. Soc., 40 (1965), 1–20.
[49] J. W. Milnor, Lectures on the h-cobordism theorem, Princeton Univ. Press, 1965.
[50] R. Stong, Notes on cobordism theory, Princeton Univ. Press, 1968.
[51] J. W. Milnor, Topology from the differentiable viewpoint, Univ. Press of Virginia, 1969.
[52] M. W. Hirsch, Differential topology, Springer, 1976.
[53] R. Mandelbaum, Four-dimensional topology: An introduction, Bull. Amer. Math. Soc., (New Series) 2 (1980), 1–159.
[54] M. H. Freedman, The topology of four-dimensional manifolds, J. Differential Geometry, 17 (1982), 357–453.
[55] S. K. Donaldson, An application of gauge theory to four-dimensional topology, J. Differential Geometry, 18 (1983), 279–315.
[56] F. Quinn, Ends of maps III: dimension 4 and 5, J. Differential Geometry, 17 (1982), 503–521.
[57] K. Kuga, Representing homology classes of $S^2 \times S^2$, Topology, 23 (1984), 133–137.

115 (XVII.8) Diffusion Processes

A. General Remarks

Let $(\Omega, \mathfrak{B}, P)$ be a †probability space. A †Markov process $\{X_t\}_{0 \leqslant t < \infty}$ on a †topological space S with †continuous time parameter t is called a **diffusion process** if the †sample function $X_t(\omega)$ is continuous in t with probability 1 until a random time $\zeta(\omega)$, called the †terminal time. After the terminal time, $X_t(\omega)$ stays at the terminal point ∂. Such a process is said to be **1-dimensional** or **multidimensional** according as S is an interval or a manifold (possibly with boundary) with dimension $\geqslant 2$. Brownian motion is a typical diffusion process (→ 45 Brownian Motion).

To give conditions for a Markov process to be a diffusion process, let us assume that S is a †complete metric space with metric ρ. Let $\{X_t\}$ be a Markov process on S with time parameter t ranging over a finite interval $[t_1, t_2]$ and satisfying

$$\sup P(s, x, t, S - U_\varepsilon(x)) = O(h) \quad (h \downarrow 0) \tag{1}$$

for every $\varepsilon > 0$, where $U_\varepsilon(x)$ is the ε-neighborhood of x and sup is taken over all $x \in S$, and $s, t \in [t_1, t_2]$ such that $0 < t - s < h$. Then $\{X_t\}$ is a diffusion process if and only if

$$\int_{t_1}^{t_2} P(\rho(X_t, X_{t+h}) > \varepsilon)\, dt = o(h) \quad (h \downarrow 0)$$

for every $\varepsilon > 0$ [5]. This includes the following result due to E. B. Dynkin (1952) and J. R. Kinney (1953) as a special case: If in (1) we can replace $O(h)$ with $o(h)$, then $\{X_t\}$ will be a diffusion process. Specifically, when $\{X_t\}$ is a 1-dimensional †strong Markov process, the latter condition is also necessary for the process to be a diffusion process.

Diffusion processes are intimately related to a certain class of †partial differential equations. Let S be the real line. Assume that the †transition probability $P(s, x, t, E)$ $(s < t)$ of the process $\{X_t\}_{0 \leqslant t < \infty}$ satisfies

$$1 - P(s, x, s+h, (x-\varepsilon, x+\varepsilon)) = o(h) \quad (h \downarrow 0) \tag{2a}$$

for every $\varepsilon > 0$ and that the following limits exist:

$$\lim_{h \to 0+} \frac{1}{h} \int_{x-\varepsilon}^{x+\varepsilon} (y-x)^2 P(s, x, s+h, dy) = 2a(s, x) > 0, \tag{2b}$$

$$\lim_{h \to 0+} \frac{1}{h} \int_{x-\varepsilon}^{x+\varepsilon} (y-x) P(s, x, s+h, dy) = b(s, x), \tag{2c}$$

$$\lim_{h \to 0+} \frac{1}{h} (P(s, x, s+h, S) - 1) = c(s, x) \leqslant 0. \tag{2d}$$

Assume further that the transition probability is †absolutely continuous with respect to Lebesgue measure: $P(s, x, t, dy) = p(s, x, t, y)\, dy$. Then under some suitable additional conditions, $p(s, x, t, y)$ satisfies

$$\frac{\partial p}{\partial s} = -a(s, x) \frac{\partial^2 p}{\partial x^2} - b(s, x) \frac{\partial p}{\partial x} - c(s, x) p,$$

$$p(t-0, x, t, y) = \delta(x-y), \tag{3}$$

and

$$\frac{\partial p}{\partial t} = \frac{\partial^2}{\partial y^2}(a(t, y) p) - \frac{\partial}{\partial y}(b(t, y) p) + c(t, y) p,$$

$$p(s, x, s+0, y) = \delta(y-x), \tag{4}$$

where δ is the †Dirac delta function. The coefficient c vanishes if $\{X_t\}$ is †conservative. Equations (3) and (4) are called **Kolmogorov's backward equation** and **forward equation**, respectively. They are also called the **Fokker-Planck partial differential equations**.

A. N. Kolmogorov [1] derived equations (3) and (4) in 1931, and W. Feller proved that (3) (or (4)) has a unique solution under certain regularity conditions on the coefficients a, b, and c, and that the solution $p(s, x, t, y)$ is non-

negative, has an integral with respect to y that does not exceed 1, and satisfies the †Chapman-Kolmogorov equation

$$\int p(s,x,t,y)p(t,y,u,z)\,dy = p(s,x,u,z)$$

for every $0<s<t<u$. Hence $p(s,x,t,y)$ determines a †Markov process analytically. However, rigorous proof establishing the sufficiency of condition (2) for the Markov process $\{X_t\}$ to be a diffusion process did not appear until the 1950s.

In the †temporally homogeneous case, $p(s,x,s+h,y)$ does not depend on s and can be written as $p(h,x,y)$. Then $a(s,x)$, $b(s,x)$, and $c(s,x)$ are also independent of s, and $p(t,x,y)$ satisfies

$$\frac{\partial p}{\partial t} = a(x)\frac{\partial^2 p}{\partial x^2} + b(x)\frac{\partial p}{\partial x} + c(x)p,$$

$$p(0+,x,y) = \delta(x-y). \tag{5}$$

Feller made an intensive study of this case and completely solved the problem of existence and uniqueness of the solution of (5) assuming that $p(t,x,y)$ is nonnegative and that its integral with respect to y does not exceed 1 [7]. In particular, when t varies in $[0,+\infty)$ and S is an interval $[r_1,r_2]$, Feller used the Hille-Yosida theory of †semigroups of operators to determine the conditions that should be satisfied by r_1 and r_2 in order that the differential equation (5) (with the initial condition and his additional assumptions) yield one and only one solution. Feller also introduced the notion of generalized differential operators, which expresses the differential operator in the right-hand side of (5) in the most general form [8]. The probabilistic meaning of his results was clarified by Dynkin, H. P. McKean, K. Itô, D. B. Ray, and others, and all 1-dimensional diffusion processes with the †strong Markov property have now been completely determined.

Since not much research on temporally nonhomogeneous diffusion processes has been done so far, we restrict our explanation to temporally homogeneous ones. Let $\mathfrak{M} = (X_t, W, P_x \mid x \in S)$ be a Markov process, where S is a †state space, W is the †path space consisting of all paths $w:[0,+\infty] \to S \cup \{\partial\}$ which are continuous in t for $0 \leqslant t < \zeta(w)$ ($w(t) = \partial$ for $t \geqslant \zeta(w)$ while $w(t) \in S$ for $0 \leqslant t < \zeta(w)$), and P_x is a probability measure on W under the condition that the process starts from x at $t=0$ (→ 261 Markov Processes). We can actually identify W with the †basic space Ω and set $X_t(w) = w(t)$ for $w \in \Omega$. Assume that $\mathfrak{M}$ has the †strong Markov property. It follows from †Dynkin's formula for †infinitesimal generators (→ 261 Markov Processes) that the infinitesimal generator $\mathfrak{G}$ of a diffusion process has the local property stating that if u and v belong to the domain of $\mathfrak{G}$ and coincide in a neighborhood of x_0, then $\mathfrak{G}u(x_0) = \mathfrak{G}v(x_0)$.

B. 1-Dimensional Diffusion Processes

Let S be a straight line. A point $x \in S$ is called a **right singular point** if $X_t(w) \geqslant x$ for all $t \in [0, \zeta(w))$ with P_x-probability 1. A **left singular point** is defined analogously, with $\geqslant$ replaced by $\leqslant$. A right and left singular point is called a **trap**, while a right singular point which is not left singular is called a **right shunt** (a **left shunt** is defined analogously). A point is called a **regular point** if it is neither right nor left singular.

The set of all regular points is open. Let (r_1, r_2) be a connected component of this open set. One of the most important results concerning this situation is the proof of the existence of a strictly increasing function $s(x)$ defined on (r_1, r_2) and two measures m and k on (r_1, r_2) such that the infinitesimal generator $\mathfrak{G}$ of $\mathfrak{M}$ is represented as

$$\mathfrak{G}u(x) = \frac{u^+(dx) - u(x)k(dx)}{m(dx)}, \tag{6}$$

where $u^+(dx)$ is the measure $du^+(x)$ induced by the †right derivative $u^+(x)$ of $u(x)$ with respect to $s(x)$ (i.e., $u^+(x) = \lim_{\Delta x \to +0}\{u(x+\Delta x) - u(x)\}/\{s(x+\Delta x) - s(x)\}$). Equation (6) gives a generalization of second-order †differential operator $au'' + bu' + cu$ ($a>0, c \leqslant 0$) [12]. Here m is positive for nonempty open sets, both m and k are finite for compact sets in (r_1, r_2), and s, m, and k are unique in the following sense: If there are two sets of values of s_i, m_i, and k_i ($i = 1, 2$), then $s_2(x) = cs_1(x) + \text{constant}$, $m_2(dx) = c^{-1}m_1(dx)$, and $k_2(dx) = c^{-1}k_1(dx)$ for some positive constant c. We call s, m, and k, respectively, the **canonical scale**, **canonical measure** (or **speed measure**), and **killing measure** for $\mathfrak{M}$. They determine the behavior of $X_t(w)$ belonging to $\mathfrak{M}$ inside the interval (r_1, r_2). Conversely, given any such set of s, m, and k, we can find a 1-dimensional diffusion process $\mathfrak{M}$ such that s, m, and k are, respectively, the canonical scale, canonical measure, and killing measure of $\mathfrak{M}$. If $X_t(w)$ is nonvanishing in (r_1, r_2) with probability 1, the killing measure k is identically zero, and the canonical scale s satisfies the equation

$$P_x(\sigma_{x_2} < \sigma_{x_1}) = \frac{s(x) - s(x_1)}{s(x_2) - s(x_1)}$$

for $x_1 < x < x_2$, where σ_y is the †hitting time of the point y.

The motion $X_t(w)$ belonging to the process $\mathfrak{M}$ and contained in (r_1, r_2) can be constructed

from the Brownian motion by means of a topological transformation of the state space (interval) based on s, a †time change based on m, and a †killing based on k. More precisely, we first transform the interval (r_1, r_2) by $x \rightarrow s(x)$ into the interval $(s(r_1+0), s(r_2-0))$ so that the diffusion process on this new interval has a canonical scale coincident with x. The speed and killing measures are transformed accordingly. We can, therefore, assume that the canonical scale is x. Let us consider the case $(r_1, r_2) = (-\infty, +\infty)$ for simplicity. Let $\mathfrak{t}(t, x)$ be the †local time of Brownian motion at x (→ 45 Brownian Motion). Next, we apply the †time change to the Brownian motion by means of the †additive functional

$$\varphi(t) = \int_{-\infty}^{+\infty} \mathfrak{t}(t, x) m(dx),$$

and finally †kill the latter process by means of the †multiplicative functional

$$\alpha(t) = \exp\left(-\int_{-\infty}^{+\infty} \mathfrak{t}(\varphi^{-1}(t), x) k(dx)\right).$$

Thus we obtain the process $\mathfrak{M}$ in $(-\infty, \infty)$. In particular, if $m(dx) = a(x)^{-1} dx$ and $k(dx) = |c(x)| dx$, we have

$$\varphi(t) = \int_0^t a(X_\tau)^{-1} d\tau,$$

$$\alpha(t) = \exp\left(\int_0^t c(X_{\varphi^{-1}(\tau)}) d\tau\right),$$

and $\mathfrak{G}u = au'' + cu$.

At a shunt the infinitesimal generator $\mathfrak{G}$ has a form that is a generalization of the first-order differential operator $bu' + cu$ $(c \leqslant 0)$, with $b > 0$ or $b < 0$ according as it is a right shunt or a left shunt. At a trap we have $\mathfrak{G}u(x) = -u(x)/E_x(\zeta)$.

When S is an interval with endpoints r_1 and r_2 and all interior points are regular, the left endpoint r_1 is classified into the following 4 types, according to the behavior of $\mathfrak{M}$ near r_1: Take an arbitrary fixed point $r \in (r_1, r_2)$, and set $n(dx) = m(dx) + k(dx)$ and

$$\alpha = \int_{(r_1, r)} (s(r) - s(x)) n(dx),$$

$$\beta = \int_{(r_1, r)} n((x, r)) s(dx).$$

Then r_1 is a **regular boundary** if $\alpha < \infty$, $\beta < \infty$; an **entrance boundary** if $\alpha < \infty$, $\beta = \infty$; an **exit boundary** if $\alpha = \infty$, $\beta < \infty$; and a **natural boundary** if $\alpha = \infty$, $\beta = \infty$. This classification is independent of the choice of r. A similar classification of r_2 can be established. $X_t(w)$ approaches r_1 in finite time with positive or null probability according as β is finite or infinite. If $\alpha = \infty$, it never happens that $X_t(w)$ starts from r_1 and reaches the interior of the interval S even if $r_1 \in S$, whereas if $\alpha < \infty$ we can construct (adjoining r_1 to S if necessary) a diffusion process that enters the interior from r_1 and whose motion in the interior coincides with that of $X_t(w)$.

If r_1 is a regular boundary for $\mathfrak{M}$ and $r_1 \in S$, then there are various possibilities for the behavior of $X_t(w)$ at r_1. They are expressed by the boundary conditions satisfied by the functions u belonging to the domain of the infinitesimal generator $\mathfrak{G}$. The condition is in general of the form

$$\gamma u(r_1) + \delta \mathfrak{G} u(r_1) + \mu u^+(r_1) = 0,$$

where γ, δ, and μ are constants, $\gamma, \delta \leqslant 0$, $\mu \geqslant 0$, and $|\delta| + \mu > 0$. If $\gamma = \delta = 0$, then r_1 is said to be a **reflecting barrier**. If r_1 is regular for $\mathfrak{M}$ and does not belong to S, then $X_t(w)$ vanishes exactly as $X_t(w)$ reaches r_1, and r_1 is called an **absorbing barrier**. This case corresponds to the boundary condition $u(r_1) = 0$. Whatever the boundary condition may be, $\mathfrak{M}$ is constructed from the Brownian motion with reflecting barrier by topological transformation of the state space, time change, and killing. Here if $\gamma \neq 0$, then killing may occur at r_1; if $\delta \neq 0$, the set of visiting times of r_1 has positive Lebesgue measure; and if $\mu \neq 0$, the trace of the motion may go beyond the point r_1 and reach the interior points of S [2, 7–9].

If we weaken the assumption of continuity of paths and admit jumps from r_1, the general boundary condition becomes

$$\gamma u(r_1) + \delta \mathfrak{G} u(r_1) + \mu u^+(r_1) + \int_{(r_1, r_2]} (u(x) - u(r_1)) \nu(dx) = 0,$$

where ν is a measure with respect to which $\min(1, s(x) - s(r_1+0))$ is integrable.

When $S = (r_1, r_2)$, the transition probability is absolutely continuous with respect to the canonical measure, the density $p(t, x, y)$ has an †eigenfunction expansion

$$p(t, x, y) = \int_{-\infty}^{0} e^{\lambda t} \mathfrak{e}(d\lambda; x, y),$$

and $p(t, x, y)$ is positive, jointly continuous in 3 variables, and symmetric in x and y. A similar result is also known when S is half-open or closed [9].

If $\mathfrak{M}$ is †recurrent, i.e., $P_x(\sigma_y < +\infty) = 1$ for every x and y in S, then there exists a unique (up to a multiplicative constant) †invariant measure for S that is finite for all closed intervals in the interior of S. If $\mathfrak{M}$ is conservative and all the interior points of S are regular, then the canonical measure is an invariant measure, provided that the endpoints are either entrance, natural, or regular reflecting.

C. Multidimensional Diffusion Processes

Let the state space S be a domain or the closure of a domain in the n-dimensional Euclidean space $\mathbf{R}^n$. Consider a temporally homogeneous diffusion process $\{X_t\}_{0 \leqslant t < \infty}$ on S. Under suitable regularity conditions the infinitesimal generator $\mathfrak{G}$ coincides, for sufficiently smooth functions u in its domain, with the following †elliptic partial differential operator A:

$$A = \sum_{i,j=1}^{n} a^{ij}(x)\frac{\partial^2}{\partial x^i \partial x^j} + \sum_{i=1}^{n} b^i(x)\frac{\partial}{\partial x^i} + c(x), \qquad c \leqslant 0; \qquad (7)$$

where S has a boundary, u satisfies a boundary condition of the form

$$\sum_{i,j=1}^{n} \alpha^{ij}(x)\frac{\partial^2 u(x)}{\partial x^i \partial x^j} + \sum_{i=1}^{n-1} \beta^i(x)\frac{\partial u(x)}{\partial x^i}$$

$$+\gamma(x)u(x) + \delta(x)Au(x) + \mu(x)\frac{\partial u(x)}{\partial n} = 0, \qquad (8)$$

where, for simplicity, we assume that S is the closed half-space defined by $x^n \geqslant 0$, (α^{ij}) is a symmetric nonnegative definite matrix, $\gamma \leqslant 0$, $\delta \leqslant 0$, $\mu \geqslant 0$, and $\partial/\partial n$ is the inward-directed †conormal derivative associated with a^{ij}. This boundary condition was discovered by A. D. Venttsel' [13]. Conversely, given an operator A such as (7) and a boundary condition (8) (if S has a boundary), the existence and uniqueness of the corresponding diffusion process are known for several special cases. If $S = \mathbf{R}^n$ and A has continuous coefficients, there exists at least one diffusion process corresponding to A [14–16].

A probabilistic approach to constructing diffusion processes corresponding to the operator A with $c = 0$ is given by Itô's method of †stochastic differential equations (→ 406 Stochastic Differential Equations E). A somewhat different approach was introduced by D. W. Stroock and S. R. S. Varadhan [20] under the name of **martingale problems** (→ 261 Markov Processes C). Let $W^n = C([0, \infty) \to \mathbf{R}^n)$ be the space of all continuous functions w: $[0, \infty) \to \mathbf{R}^n$ endowed with compact uniform topology and $\mathfrak{B}(W^n)$ be the topological σ-field. Given $x \in \mathbf{R}^n$, a **solution to the martingale problem** for the operator A with $c = 0$ starting from x is a probability measure P_x on $(W^n, \mathfrak{B}(W^n))$ satisfying $P_x(w(0) = x) = 1$ such that $f(w(t)) - \int_0^t Af(w(s))\,ds$ is a P_x-martingale for all $f \in C_0^\infty(\mathbf{R}^n)$, where $C_0^\infty(\mathbf{R}^n)$ denotes the set of all C^∞-functions on $\mathbf{R}^n$ having compact support. If $a = (a^{ij})$ is uniformly positive definite, bounded, and continuous and if $b = (b^i)$ is bounded and Borel measurable, the martingale problem for the operator A with $c = 0$ is **well posed**, i.e., for each $x \in \mathbf{R}^n$, there is exactly one solution starting from x. (For further information related to the theory of martingale problems → [15, 20].)

Suppose next that S is the closure of a bounded domain with a sufficiently smooth boundary and A is given with sufficiently smooth coefficients. If the boundary condition is $\gamma u + \mu \partial u/\partial n = 0$, $\mu \neq 0$, then there exists a unique diffusion process on S corresponding to this situation. Moreover, if $\gamma = 0$, then the process is said to have a **reflecting barrier**. S. Watanabe [22] gave a probabilistic condition which characterizes the reflecting diffusion processes in the normal direction among all reflecting diffusion processes in oblique directions. Suppose that a general boundary condition (8) is given. We write the left-hand side of (8) as Lu. Write $u = Hf$ for the solution of $Au = 0$ with boundary value f. Under some natural additional conditions, T. Ueno proved that if LH †generates a Markov process on the boundary, then there exists a diffusion process for A with boundary condition (8) [21]. If $\{X_t\}$ has a reflecting barrier, the Markov process on the boundary is, conversely, obtained from $\{X_t\}$ through time change by a nonnegative continuous †additive functional which increases only when the value of X_t is on the boundary. Stochastic differential equations are also used in constructing diffusion processes with boundary condition (8) [19] (→ 406 Stochastic Differential Equations). An elegant method for constructing diffusion processes with boundary condition (8) has been introduced by Watanabe [22]. It consists of piecing together excursions from the boundary to the boundary. In this construction, the stochastic integrals for †Poisson point processes play an important role. There are various other results on multidimensional diffusion processes with general boundary conditions which cannot be covered by the method mentioned above (→ [23] and M. Motoo, *Proc. Intern. Symp. Stochastic Differential Equations, Kyoto*, 1976).

When we have a diffusion process on S with infinitesimal generator of the form A, we can obtain a probabilistic expression for the solutions of various partial differential equations involving A. Let σ be the hitting time of the boundary of S. Then $Hf(x)$ can be expressed as $E_x(f(X_{\sigma-0}))$. The solution of $Au = -f$ with boundary value 0 is given by $u(x) = E_x(\int_0^\sigma f(X_t)\,dt)$, while the solution $u(t, x)$ of $\partial u/\partial t = Au$ with boundary value 0 and initial value $u(0, x) = f$ is $E_x(f(X_t); t < \sigma)$. The first case gives the solution of a †Dirichlet problem; and in this case, the condition for a boundary point to be †regular relative to the Dirichlet problem can also be expressed probabilistically (→ 45 Brownian Motion, 261 Markov

Processes). Furthermore, if $f(X_t)$ is replaced by $f(X_t)\exp(\int_0^t k(X_s)ds)$ in the expressions for $u(x)$ and $u(t,x)$, A is replaced by $A+k$ (M. Kac [24]). When $k\leqslant 0$, this replacement gives rise to a killing of the process (→ 261 Markov Processes E).

By using the theory of †Dirichlet forms we can investigate a general class of †symmetric multidimensional diffusion processes which are not in the framework of the classical diffusion processes, i.e., diffusion processes whose infinitesimal generators are not necessarily differential operators (M. Fukushima, *Dirichlet Forms and Markov Processes*, 1980)).

D. Diffusion Processes on Manifolds

Let M be a †connected †oriented †σ-compact C^∞ †manifold of dimension n. Let $\hat{M}$ be M or $M\cup\{\partial\}$ (= the one-point †compactification of M) according as M is †compact or noncompact. Suppose that we are given a system of C^∞-†vector fields $A_0, A_1, \dots, A_r$ on M. We consider the following stochastic differential equation on M:

$$dX_t=\sum_{k=1}^r A_k(X_t)\circ dw^k(t)+A_0(X_t)dt \tag{9}$$

(→ 406 Stochastic Differential Equations), where $w(t)=(w^1(t), w^2(t), \dots, w^r(t))$ denotes an r-dimensional Brownian motion and the first term of the right-hand side is understood in the sense of the †Stratonovich stochastic differential. Let $X(t,x,w)$ be the solution of (9) with $X_0=x\in M$ defined on the r-dimensional Wiener space (W_0^r, P^W) (→ 406 Stochastic Differential Equations) and P_x be the probability law on $\hat{W}(M)$ of $(X(t,x,w))_{t\geqslant 0}$, where $\hat{W}(M)$ is the space of all continuous mappings $w:[0,\infty)\to\hat{M}$ with ∂ as a trap. Then $\{P_x \mid x\in M\}$ defines a diffusion process $\mathfrak{M}$ on M which is generated by the second-order differential operator $\sum_{k=1}^r A_k^2/2+A_0$, i.e., a diffusion process $\mathfrak{M}$ with the infinitesimal generator $\mathfrak{G}$ such that

$$\mathfrak{G}f(x)=\frac{1}{2}\sum_{k=1}^r A_k^2 f(x)+A_0 f(x) \quad \text{for } f\in C_0^\infty(M),$$

where $C_0^\infty(M)$ denotes the space of all C^∞-functions on M with compact support. By appealing to the analytical theory of partial differential equations we can discuss regularity properties of the transition probability of $\mathfrak{M}$ ([15]; L. Hörmander, *Acta Math.*, 119 (1967)). Recently, P. Malliavin [25] also suggested a probabilistic method for proving elliptic regularity results (see also [19]).

We now assume that all linear sums of $A_0, A_1, \dots, A_r$ are †complete. Then the terminal time of $\mathfrak{M}$ is infinite a.s. Let $W(M)$ be the space of all continuous mappings $w:[0,\infty)\to M$ endowed with compact uniform topology, and set $W_x(M)=\{w \mid w\in W(M), w(0)=x\}$. We describe the †topological support $\mathscr{S}(P_x)$ of the probability P_x on $W_x(M)$, i.e., the smallest closed subset of $W_x(M)$ that carries the measure P_x. Let $\mathscr{S}$ be the set of all piecewise constant mappings $u:[0,\infty)\to\mathbf{R}^r$. For a given $u=(u^1(t), u^2(t), \dots, u^r(t))$ of $\mathscr{S}$, we consider a system of ordinary differential equations

$$\frac{d}{dt}\varphi(t)=A_0(\varphi(t))+\sum_{k=1}^r A_k(\varphi(t))u^k(t), \tag{10}$$

i.e., for every C^∞-function f on M with compact support,

$$\frac{d}{dt}f(\varphi(t))=A_0 f(\varphi(t))+\sum_{k=1}^r (A_k f)(\varphi(t))u^k(t).$$

Then for every $u\in\mathscr{S}$ and $x\in M$, we obtain a curve $\varphi=\varphi(x,u)=(\varphi_t(x,u))$ on M by solving (10) with $\varphi(0)=x$. Set $\mathscr{S}^x=\{\varphi(x,u) \mid u\in\mathscr{S}\}$. Then we have

$$\mathscr{S}(P_x)=\overline{\mathscr{S}^x} \quad \text{for every } x\in M.$$

Let $\mathscr{L}$ be the †Lie algebra generated by $A_1, A_2, \dots, A_r$, and set $\mathscr{L}(x)=\{V_x \mid V\in\mathfrak{L}\}$. If $\dim\mathscr{L}(x)=n$ for every $x\in M$, then $\mathscr{S}(P_x)=W_x(M)$ for every $x\in M$. For further information → Stroock and Varadhan (*Proc. 6th Berkeley Symp. Math. Statist. Prob.* III, 1972) and H. Kunita (*Proc. Int. Symp. Stochastic Differential Equations, Kyoto*, 1976).

Let A be a smooth †nondegenerate second-order †elliptic differential operator on M which is expressed in local coordinates as

$$A=\frac{1}{2}\sum_{i,j=1}^n a^{ij}(x)\frac{\partial^2}{\partial x^i\partial x^j}+\sum_{i=1}^n \bar{b}^i(x)\frac{\partial}{\partial x^i},$$

where $(a^{ij}(x))$ is symmetric and †strictly positive definite. Then there exist a †Riemannian metric g and a C^∞-†vector field b on M such that

$$A=\frac{1}{2}\Delta_M+b, \tag{11}$$

where Δ_M is the †Laplace-Beltrami operator on the †Riemannian manifold (M,g). We now construct the diffusion process generated by the operator A, introducing a stochastic differential equation on the †bundle $O(M)$ of the †orthonormal frames. There exists an †affine connection ∇ †compatible with the Riemannian metric g such that for every C^∞-function f on M,

$$\frac{1}{2}\sum_{k=1}^n L_k^2(f\circ\pi)(r)=\left(\frac{1}{2}\Delta_M+b\right)(f)(\pi(r)), \quad r\in O(M),$$

where $(L_1, L_2, \ldots, L_n)$ is the system of †canonical horizontal vector fields (†basic vector fields) on $O(M)$ corresponding to the affine connection ∇ and $\pi: O(M) \to M$ is the natural projection [19]. We now consider the following stochastic differential equation on $O(M)$:

$$dr_t = \sum_{k=1}^{n} L_k(r_t) \circ dw^k(t). \tag{12}$$

Let $r(t, r, w)$ be the solution of (12) with $r_0 = r \in O(M)$ defined on the n-dimensional Wiener space (W_0^n, P^W). Now a stochastic curve $X(t, r, w)$ on M is defined by $X(t, r, w) = \pi(r(t, r, w))$. Set $X(t, r, w) = \partial$ for $t \geqslant \zeta$, where ζ is the †explosion time of $r(t, r, w)$. Then the probability law of $(X(t, r, w))_{t \geqslant 0}$ on $\hat{W}(M)$ depends only on $x = \pi(r)$, and it defines a diffusion process $\mathfrak{M}$ on M which is generated by the operator A of (11). (For details → [19, 26].) When $A = \Delta_M/2$, i.e., $b = 0$ in (11), the diffusion process $\mathfrak{M}$ is called the †Brownian motion on the Riemannian manifold M (→ 45 Brownian Motion).

Next consider the case when $(L_1, L_2, \ldots, L_n)$ in (12) is a system of canonical horizontal vector fields corresponding to the †Riemannian connection, and let b be a C^∞-vector field on M. Let $\tilde{b}$ be the scalalization of b, i.e., $\tilde{b} = (\tilde{b}^1, \tilde{b}^2, \ldots, \tilde{b}^n)$, $\tilde{b}^i(r) = \sum_{j=1}^n b^j(x) f_j^i$, and $(f_j^i) = (e_j^i)^{-1}$ for $r = (x, e_1, e_2, \ldots, e_n) \in O(M)$, where $b = \sum_{i=1}^n b^i \partial/\partial x^i$ and $e_i = \sum_{j=1}^n e_i^j \partial/\partial x^j$ in local coordinates. Suppose that $\tilde{b}^i(r)$ is bounded on $O(M)$ for $i = 1, 2, \ldots, n$. Then

$$M(t) = \exp\left[\sum_{i=1}^n \int_0^t \tilde{b}^i(r_s) dw^i(s) - \frac{1}{2}\sum_{i=1}^n \int_0^t (\tilde{b}^i(r_s))^2 ds\right]$$

is an †exponential martingale. The diffusion process generated by the operator $\Delta_M/2 + b$ is obtained from the Brownian motion on M constructed above by the †transformation of drift, i.e., the transformation by the †multiplicative functional $M(t)$ (→ 261 Markov Processes). (G. Maruyama (*Nat. Sci. Rep. Ochanomizu Univ.*, 15 (1954)) studied this for 1-dimensional processes, I. V. Girsanov [27] and M. Motoo (*Ann. Inst. Statist. Math.*, 12 (1960–1961)) for multidimensional cases.)

Consider a diffusion process $\mathfrak{M} = \{P_x \mid x \in M\}$ on M generated by the operator A of (11). Let ω_b be the †differential 1-form defined from the vector field b by

$$\omega_b(V)_x = g_x(b_x, V_x)$$

for every $x \in M$ and every C^∞-vector field V on M. Then $\mathfrak{M}$ is symmetrizable if and only if ω_b is †exact, i.e., if there exists a function F on M such that $\omega_b = dF$. The †invariant measures are then of the form const. $\times \exp[2F(x)] m(dx)$, where $m(dx)$ is the †Riemannian volume (E. Nelson, *Duke Math. J.*, 25 (1958); Kolmogorov, *Math. Ann.*, 113 (1937)). The diffusion process $\mathfrak{M}$ is said to be **locally symmetrizable** if for every †simply connected domain $D \subset M$ there exists a Borel measure $v^D(dx)$ on D such that

$$\int_D T_t^D f(x) g(x) v^D(dx) = \int_D f(x) T_t^D g(x) v^D(dx)$$

for all bounded continuous functions f and g, where

$$T_t^D f(x) = \int_{\hat{W}(M)} f(w(t)) I_{\{w \mid \tau_D(w) > t\}} P_x(dw),$$

$$\tau_D(w) = \inf\{t \mid w(t) \notin D\}.$$

Then $\mathfrak{M}$ is locally symmetrizable if and only if ω_b is †closed, i.e., $d\omega_b = 0$. R. Z. Khas'minskiĭ [28] proved a pair of useful tests for explosions of diffusion processes on $M = \mathbf{R}^n$, similar to Feller's test for $n = 1$ mentioned in Section B (→ [18]). In case of a general manifold we can investigate the possibility of explosions for $\mathfrak{M}$ by appealing to the comparison theorems for †curvatures in the theory of differential geometry [19, 29]; S. T. Yau, *J. Math. Pures Appl.*, 57 (1978)). The diffusion process $\mathfrak{M}$ on M is said to be †recurrent if $P_x[X_t \in U$ for some $t > 0] = 1$ for any open subset U of M; otherwise it is called †transient. It is well known that an n-dimensional Brownian motion on $\mathbf{R}^n$ is recurrent if $n \leqslant 2$ and transient if $n \geqslant 3$ [9]. There also are some results for the criterion which determines whether $\mathfrak{M}$ is recurrent or transient ([19, 28]; A. Friedman, *Stochastic Differential Equations and Applications* I, II, 1975, and K. Ichihara, *Publ. Res. Inst. Math. Sci.*, 14 (1978)). Explicit formulas for the transition probability of the Brownian motion on a hyperbolic space are given in [29]. For information related to the asymptotic behavior of diffusion processes as $t \downarrow 0$ we refer the reader to Varadhan (*Comm. Pure Appl. Math.*, 20 (1967)) and S. A. Molchanov (*Russian Math. Surveys*, 30 (1975)).

Diffusion approximations for suitably normalized random sequences have been studied extensively beginning with the work of A. J. Khinchin [30]. The theory of †limiting distributions of sums of †independent (or weakly dependent) random variables is among the best-known examples (→ 250 Limit Theorems in Probability Theory). For information related to the theory of diffusion approximations, see Yu. V. Prokhorov (*Theory of Prob. Appl.*, 1 (1956)), A. V. Skorokhod [16], and Stroock, Varadhan, and G. Papanicolaou (*Proc. 1976 Duke Turbulence Conference*). Recently, many interesting examples of multidimensional diffusion processes have also been

introduced to describe probabilistic models in physics, biology, etc. (e.g., R. L. Stratonovich, *Topics in the Theory of Random Noise* I, II, 1963; J. F. Crow and M. Kimura, *An Introduction to Population Genetics Theory*, 1970; and K. Sato, *Proc. Int. Symp. Stochastic Differential Equations, Kyoto*, 1976).

In general, if a diffusion process $\{X_t\}$ is given on a noncompact space S and X_t has no limit points in S as $t\uparrow\zeta$, then some natural compactification should be induced by $\{X_t\}$. The notion of a †Martin boundary for Markov processes is introduced in this connection.

References

[1] A. N. Kolmogorov, Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung, Math. Ann., 104 (1931), 451–458.
[2] E. B. Dynkin, Markov processes I, II, Springer, 1965. (Original in Russian, 1963.)
[3] J. L. Doob, Stochastic processes, Wiley, 1953.
[4] K. Itô, Lectures on stochastic processes, Tata Inst., 1960.
[5] L. V. Seregin, Continuity conditions for stochastic processes, Theory of Prob. Appl., 6 (1961), 1–26. (Original in Russian, 1961.)
[6] W. Feller, Zur Theorie der Stochastischen Prozesse (Existenz- und Eindeutigkeits-sätz), Math. Ann., 113 (1936), 113–160.
[7] W. Feller, The parabolic differential equations and the associated semi-groups of transformations, Ann. Math., (2) 55 (1952), 468–519.
[8] W. Feller, On second order differential operators, Ann. Math., (2), 61 (1955), 90–105.
[9] K. Itô and H. P. McKean, Jr., Diffusion processes and their sample paths, Springer, 1965.
[10] P. Lévy, Processus stochastiques et mouvement brownien, Gauthier-Villars, 1948.
[11] E. B. Dynkin, One-dimensional continuous strong Markov processes, Theory of Prob. Appl., 4 (1959), 1–52. (Original in Russian, 1959.)
[12] I. S. Kats (Kac) and M. G. Kreĭn, On the spectral functions of the string, Amer. Math. Soc. Transl., (2) 103 (1974). (Original in Russian, supplement of the Russian translation of the book by F. V. Atkinson, Discrete and continuous boundary problems.)
[13] A. D. Venttsel' (Ventcel'), On boundary conditions for multidimensional diffusion processes, Theory of Prob. Appl., 4 (1959), 164–177. (Original in Russian, 1959.)
[14] N. V. Krylov, On the selection of a Markov process from a system of processes and the construction of quasidiffusion processes, Math. USSR-Izv., 7 (1973), 691–709. (Original in Russian, 1973.)
[15] D. W. Stroock and S. R. S. Varadhan, Multidimensional diffusion processes, Springer, 1979.
[16] A. V. Skorokhod, Studies in the theory of random processes, Addison-Wesley, 1965. (Original in Russian, 1961.)
[17] K. Itô, On stochastic differential equations, Mem. Amer. Math. Soc., 1951.
[18] H. P. McKean, Jr., Stochastic integrals, Academic Press, 1969.
[19] N. Ikeda and S. Watanabe, Stochastic differential equations and diffusion processes, North-Holland and Kodansha, 1981.
[20] D. W. Stroock and S. R. S. Varadhan, Diffusion processes with continuous coefficients I, II, Comm. Pure. Appl. Math., 22 (1969), 345–400, 479–530.
[21] K. Sato and T. Ueno, Multidimensional diffusions and the Markov process on the boundary, J. Math. Kyoto Univ., 4 (1965), 529–605.
[22] S. Watanabe, Poisson point process of Brownian excursions and its applications to diffusion processes, Amer. Math. Soc. Proc. Symp. Pure Math., 31 (1977), 153–164.
[23] M. Motoo, Application of additive functionals to the boundary problem of Markov processes, Proc. 5th Berkeley Symp. Math. Stat. Prob. II, pt. 2, Univ. of California Press, 1967, 75–110.
[24] M. Kac, On some connections between probability theory and differential and integral equations, Proc. 2nd Berkeley Symp. Math. Stat. Prob., Univ. of California Press, 1951, 189–215.
[25] P. Malliavin, C^k-hypoelliptic with degeneracy, Stochastic Analysis, A. Friedman and M. Pinsky (eds.), Academic Press, 1978, 199–214, 327–340.
[26] P. Malliavin, Formule de la moyenne, calcul de perturbations et théorèmes d'annulation pour les formes harmoniques, J. Functional Anal., 17 (1974), 274–291.
[27] I. V. Girsanov, On transforming a certain class of stochastic processes by absolutely continuous substitution of measures, Theory of Prob. Appl., 5 (1960), 285–301. (Original in Russian, 1960.)
[28] R. Z. Khas'minskiĭ, Ergodic properties of recurrent diffusion processes and stabilization of the solution of the Cauchy problem for parabolic equations, Theory of Prob. Appl., (1960), 179–196. (Original in Russian, 1960.)
[29] A. Debiard, B. Gaveau and E. Mazet, Théorèmes de comparison en géométrie riemannienne, Publ. Res. Inst. Math. Sci., 12 (1976), 391–425.
[30] A. Ya. Khinchin, Asymptotische Gesetze der Wahrscheinlichkeitsrechnung, Springer, 1933.

116 (XX.2) Dimensional Analysis

The system of units for physical quantities is derived from a certain set of **fundamental units**. If the fundamental units are denoted by θ, φ, ψ, etc., any other unit α (called a **derived unit**) can always be expressed in the form $\alpha = c\theta^l\varphi^m\psi^n \ldots$ ($c, l, m, n, \ldots$ are constants), by definition or by physical laws. The exponents l, m, n, ... are called the **dimensions** of α, and the content of the previous statement is expressed as $[\alpha] = [\theta^l\varphi^m\psi^n \ldots]$, which is called the **dimensional formula**. The usual practice is to take as fundamental units length, time, mass, temperature, and energy, which are denoted by L, T, M, θ, and H, respectively. **Dimensional analysis** investigates the relation between physical quantities by use of the π theorem and the law of similitude given below.

A. The π Theorem

If a relationship $f(\alpha, \beta, \ldots) = 0$ holds among n physical quantities α, β, ... independently of the choice of fundamental units, the equation $f(\alpha, \beta, \ldots) = 0$ can always be transformed into $F(\pi_1, \pi_2, \ldots) = 0$, where the π_i are $n-m$ dimensionless quantities (m is the number of fundamental units) of the form $\pi_i = \alpha^{a_i}\beta^{b_i} \ldots$. If we choose the π_i so that $\pi_1 = \alpha\beta^{-r_1}\gamma^{-z_1}\ldots$ and π_2, π_3, etc. do not contain α, then $f = 0$ implies $\alpha = \beta^{r_1}\gamma^{z_1}\ldots\Phi(\pi_2, \pi_3, \ldots)$, which clearly shows the manner in which the quantity α is related to other quantities $\beta, \gamma, \ldots$.

B. The Law of Similitude

In general, if two physical systems of the same kind have the same values of the π_i, then the physical states of the systems are similar. If we are given a family of mutually similar systems, it is sufficient to observe a particular one among them (a "model") in order to estimate physical values attached to any one of the given systems.

Consider, for example, the case of the drag D acting on geometrically similar bodies placed in the flow of a viscous incompressible fluid. If v is the velocity, l the representative length of the body, ρ the density of the fluid, and μ the viscosity (which has the dimensional formula $ML^{-1}T^{-1}$), then the π theorem gives $D/\rho v^2 l^2 = f(\rho vl/\mu)$. Hence the drag coefficient as given by the left-hand side can be obtained by the experiments performed on a geometrically similar model. The dimensionless quantity $R = vl/\nu$ ($\nu = \mu/\rho$) is called the **Reynolds number**. If the wave resistance due to gravity as well as the effect of compressibility are taken into account, gravitational acceleration g and sound velocity a must be included, so that we have $D/\rho v^2 l^2 = f(vl/\nu, v^2/lg, v/a, C_1, C_2, \ldots)$, where C_1, C_2, ... are other dimensionless quantities depending on the physical properties of the fluid. $Fr = v^2/lg$ is called the **Froude number**, and $M = v/a$ the **Mach number**.

Next, consider the case of heat transfer between a solid surface and a flowing fluid. Let the area of the solid surface be denoted by S, the heat transferred per unit time by Q (HT^{-1}), the thermal conductivity of the fluid by k ($HL^{-1}T^{-1}\theta^{-1}$), the specific heat by C ($HM^{-1}\theta^{-1}$), the two representative temperatures by T_0 and T_1, and the representative length by l, where expressions in parentheses represent dimensional formulas. Then we have, as dimensionless quantities, the **Nusselt number** $Nu = Q/(kS(T_1 - T_0)/l)$, the **Prandtl number** $Pr = \nu/\kappa$ ($\kappa = k/\rho c$), the **Grashoff number** $Gr = d^3 g(T_1 - T_0)/\nu^2 T_0$, and R, so that from the π theorem we have the relation $Nu = f(R, Pr, Gr, C_1, C_2, \ldots)$. Furthermore, $Pe = vl/\kappa = Pr\,R$ is called the **Péclet number**.

References

[1] A. W. Porter, The method of dimensions, Methuen, third edition, 1946.
[2] D. C. Ipsen, Units, dimensions, and dimensionless numbers, McGraw-Hill, 1960.
[3] H. L. Langhaar, Dimensional analysis and theory of models, Wiley, 1951.
[4] H. E. Huntley, Dimensional analysis, Macdonald, 1952.
[5] C. M. Focken, Dimensional methods and their applications, Arnold, 1953.
[6] R. Kurth, Dimensional analysis and group theory in astrophysics, Pergamon, 1972.
[7] L. I. Sedov, Similarity and dimensional methods in mechanics, Academic Press, 1959. (Original in Russian, fourth edition, 1957.)
Also → 414 Systems of Units.

117 (II.21) Dimension Theory

A. Introduction

Toward the end of the 19th century, G. Cantor discovered that there exists a one-to-one correspondence between the set of points on a line segment and the set of points on a square; and also, G. Peano discovered the existence of a †continuous mapping from the segment onto the square. Soon, the progress of the theory of

point-set topology led to the consideration of sets which are more complicated than familiar sets, such as polygons and polyhedra. Thus it became necessary to give a precise definition to dimension, a concept which had previously been used only vaguely. In 1913, L. E. J. Brouwer [9] gave a definition of dimension based on an idea of H. Poincaré. In 1922, the foundations of dimension theory for separable metric spaces were established by K. Menger [11] and P. Uryson [10]. Subsequently, P. S. Aleksandrov and W. Hurewicz contributed much to the development of the theory. The foundations of dimension theory for general metric spaces were established independently by M. Katětov amd K. Morita. More general theory for †normal spaces has also been investigated; the same results as in metric spaces, however, do not always hold.

B. Definition of Dimension

Let X be a normal space. If any finite open †covering of X has an open covering of †order $\leqslant n+1$ as its refinement (→ 425 Topological Spaces R) (i.e., if for any open sets G_i ($i=1, \ldots, s$) such that $X=G_1 \cup \ldots \cup G_s$, there exist open sets H_i ($i=1, \ldots, s$) such that $H_i \subset G_i$, $X=H_1 \cup \ldots \cup H_s$ and any $n+2$ of the H_i have no point in common), then we write $\dim X \leqslant n$. If $\dim X \leqslant n$ but $\dim X \leqslant n-1$ does not hold, then we define X to be **n-dimensional** and write $\dim X=n$. We call $\dim X$ the **covering dimension**, (or **Lebesgue dimension**) of X. The idea behind this definition is due to H. Lebesgue.

There are other definitions of *dimension* that are given inductively. Let us define $\operatorname{Ind} X=-1$ if X is empty. If for any pair consisting of a closed set F and an open set G with $F \subset G$ in X there exists an open set V such that $F \subset V \subset G$ and $\operatorname{Ind}(\bar{V}-V) \leqslant n-1$, then we define $\operatorname{Ind} X \leqslant n$. Next, we define $\operatorname{ind} \varnothing=-1$. For any point p of X and any neighborhood G of p, suppose that there exists an open neighborhood V of p such that $V \subset G$ and $\operatorname{ind}(\bar{V}-V)=n-1$. Then we define $\operatorname{ind} X \leqslant n$. As before, we set $\operatorname{Ind} X=n$ ($\operatorname{ind} X=n$) if $\operatorname{Ind} X \leqslant n$ ($\operatorname{ind} X \leqslant n$) but $\operatorname{Ind} X \leqslant n-1$ ($\operatorname{ind} X \leqslant n-1$) does not hold. (The definition of $\operatorname{ind} X$ is due to Menger.) We call $\operatorname{Ind} X$ ($\operatorname{ind} X$) the **large inductive dimension** of X (the **small inductive dimension** of X).

If $\dim X \leqslant n$ does not hold for any n, then X is called **infinite-dimensional**, written $\dim X=\infty$; we define $\operatorname{Ind} X=\infty$ and $\operatorname{ind} X=\infty$ similarly. These dimensions are invariant under †homeomorphisms.

The set of irrational points in a Euclidean space, the †Cantor discontinuum, and †Baire zero-dimensional spaces are all 0-dimensional. The set of rational points in a separable †Hilbert space is 1-dimensional.

C. Dimension of Metric Spaces

The following theorems hold for the dimension of metric spaces (M. Katětov, *Czechoslovak Math. J.*, 2 (1952); K. Morita, *Math. Ann.*, 128 (1954)). Let X and Y be metric spaces. The equality $\dim X=\operatorname{Ind} X$ holds. If $Y \subset X$, then $\dim Y \leqslant \dim X$. If X is a union of a countable number of closed sets F_i ($i=1,2, \ldots$), then $\dim X=\max(\dim F_i)$ (**sum theorem for dimension**). The inequality $\dim(X \cup Y) \leqslant \dim X+\dim Y+1$ holds. If $\dim X=n$, then X is a union of $n+1$ 0-dimensional subsets (**decomposition theorem for dimension**). We have $\dim(X \times Y) \leqslant \dim X+\dim Y$, where $X \neq \varnothing$ (**product theorem for dimension**).

Each of the following is a necessary and sufficient condition for $\dim X \leqslant n$: (i) There exists a subspace A of a Baire zero-dimensional space $B(\tau)$ and a continuous closed mapping f of A onto X such that $f^{-1}(x)$ consists of at most $n+1$ points for each point x of X (K. Morita, *Sci. Rep. Tokyo Kyoiku Daigaku*, 5 (1955)); (ii) there exists a metric of X which gives the same topology on X such that for any positive number ε, any point x of X, and any $n+2$ points x_i ($i=1, \ldots, n+2$) at a distance less than ε from the ($\varepsilon/2$)-neighborhood of x, there are at least two points x_i and x_j ($i \neq j$) with distance $<\varepsilon$ (J. Nagata, *Fund. Math.*, 45 (1958)).

Hurewicz's problem asked whether the equality $\dim X=n+m$ ($m>0$) implies the existence of an m-dimensional space A and a mapping f of A onto X having property (i). It was solved affirmatively for separable metric spaces by J. H. Roberts and for general metric spaces by K. Nagami (*Japan. J. Math.*, 30 (1960)).

If X is the union of a countable number of closed †strongly paracompact subspaces, in particular if X is separable, then $\operatorname{Ind} X=\operatorname{ind} X$ [1, 2]. However, it was shown by P. Roy (*Bull. Amer. Math. Soc.*, 68 (1962)) that this equality does not hold in general.

D. Euclidean Spaces and Dimension

The n-dimensional †Euclidean space $\mathbf{R}^n$ is exactly n-dimensional in the sense mentioned above; thus this concept of dimension agrees with our intuition. The proof of $\dim \mathbf{R}^n \geqslant n$ comes from **Lebesgue's theorem**: If each member of a finite closed covering of an n-cube has

sufficiently small diameter, then the order of the covering is not less than $n+1$. (The proof of $\dim \mathbf{R}^n \leqslant n$ is easy.) Let X be a subset of $\mathbf{R}^n$ and f a homeomorphism from X onto a subset $f(X)$ of $\mathbf{R}^n$. If x is an interior point of X, then $f(x)$ is an interior point of $f(X)$. Also, if an open set A of $\mathbf{R}^n$ is homeomorphic to a subset B of $\mathbf{R}^n$, then B is open in $\mathbf{R}^n$ (**Brouwer's theorem on the invariance of domain** [8]). This theorem holds for any manifold but not for general separable metric spaces. By the theorem of invariance of domain it can be shown that $\mathbf{R}^m$ and $\mathbf{R}^n$, $m \neq n$, are not homeomorphic (**theorem on invariance of dimension of Euclidean spaces**). Any n-dimensional separable metric space is embedded in a Euclidean space $\mathbf{R}^{2n+1}$, or, more precisely, in the subset of $\mathbf{R}^{2n+1}$ consisting of all points x of which at most n coordinates are rational (**Menger-Nöbeling embedding theorem**, G. Nöbeling, *Math. Ann.*, 104 (1930)). Thus, from the topological point of view, any finite-dimensional separable metric space can be identified with a subset of a Euclidean space. Moreover, it is known that any n-dimensional separable metric space is homeomorphic to a subset of some n-dimensional compact metric space.

If F is a bounded closed subset of $\mathbf{R}^m$, then $\dim F \leqslant n$ if and only if for any positive number ε, there exists a continuous mapping f from F into an n-dimensional polyhedron in $\mathbf{R}^m$ such that the distance between x and $f(x)$ is less than ε for each point x of F.

E. Dimension of Normal Spaces

Let X be a normal space. Then $\operatorname{Ind} X \geqslant \dim X$ and $\operatorname{Ind} X \geqslant \operatorname{ind} X$, but the equalities do not necessarily hold here. The following theorems were obtained by E. Čech, Aleksandrov, C. H. Dowker, E. Hemmingsen, and Morita [1]. If $\dim X \leqslant n$, then any locally finite open covering of X has an open covering of order $\leqslant n+1$ as its refinement; if A is an †F_σ subset of X or A is strongly paracompact, then $\dim A \leqslant \dim X$; if X has a †σ-locally finite closed covering $\{F_\alpha\}$, then $\dim X = \max(\dim F_\alpha)$.

In order that $\dim X \leqslant n$, it is necessary and sufficient that any continuous mapping from a closed subset of X into an n-sphere S^n can be extended continuously to X. If X and Y are †paracompact and X is †locally compact, or if $X \times Y$ is strongly paracompact, then $\dim(X \times Y) \leqslant \dim X + \dim Y$, where $X \neq \varnothing$; if X is a †CW complex, then the equality holds [14]. M. L. Wage (*Proc. Nat. Acad. Sci. US*, 75 (1978)) proved under the continuum hypothesis (CH) that $\dim(X \times Y) \leqslant \dim X + \dim Y$ does not hold in general even if $X \times Y$ is locally compact and normal, and $\dim X = \dim Y = 0$; T. Przymusiński (*Proc. Amer. Math. Soc.*, 76 (1979)) noted that CH can be avoided by a modification of Wage's construction. Katětov (*Časopis Pěst. Mat. Fys.*, 75 (1950)) proved that $\dim X$ is determined by the ring $C^*(X)$ of bounded real-valued continuous functions on X.

F. Homological Dimension

Aleksandrov contributed much to the development of dimension theory in introducing the concept of homological dimension (*Math. Ann.*, 106 (1932)). The **homological dimension** of a compact Hausdorff space X with respect to an Abelian group G is the largest integer n such that the n-dimensional †Čech homology group $\check{H}_n(X, A; G)$ is nonzero for some closed subset A of X. The **cohomological dimension** $D(X; G)$ is defined similarly by using the †Čech cohomology group $\check{H}^n(X, A; G)$. If $\dim X < \infty$, then $\dim X = D(X; \mathbf{Z})$ ($\mathbf{Z}$ is the additive group of integers). The cohomological dimension of X with respect to an arbitrary Abelian group is determined by the cohomological dimension with respect to some specified groups, and the cohomological dimension of the product space $X \times Y$ is expressed in terms of those of X and Y (M. F. Bokshteĭn, [5]). A compact Hausdorff space X has the property that $\dim(X \times Y) = \dim X + \dim Y$ for any compact Hausdorff space Y if and only if $\dim X = D(X; \mathbf{Q}(p))$, where $\mathbf{Q}(p)$ is the additive group of rationals mod 1 of the form m/p^s for any prime number p (V. Boltyanskiĭ, [5]); this result holds also when X is paracompact (Y. Kodama, *J. Math. Soc. Japan*, 18 (1966)).

G. Dimension and Measure

Let X be a separable metric space. Then $\dim X \leqslant n$ if and only if X is homeomorphic to a subset of a Euclidean space $\mathbf{R}^{2n+1}$ whose $(n+1)$-dimensional †Hausdorff measure is 0 (E. Szpilrajn. *Fund. Math.*, 28 (1937); also → [1, 2]). The infimum of $\alpha \geqslant 0$ such that the Hausdorff measure $\Lambda_\alpha(X)$ of dimension α vanishes is called the **Hausdorff dimension** of X.

H. Dimension Type (Fréchet's Definition)

In analogy to the theory of †cardinal numbers in set theory, M. Fréchet (1909) defined the dimension type of topological spaces as fol-

lows: Two spaces X and Y are said to have the same **dimension type** if X is homeomorphic to a subset of Y and Y is homeomorphic to a subset of X.

I. Infinite-Dimensional Spaces

If X is a metric space with $0<\dim X<\infty$, then for each positive integer m with $m \leqslant \dim X$, X contains a (closed) subset S with $\dim S=m$. Tumarkin asked the following question: For an infinite-dimensional compact metric space X and for each positive integer m, does X contain a closed subset S with $\dim S=m$? D. Henderson (*Amer. J. Math.*, 89 (1967)) answered this question in the negative. Furthermore, J. Walsh (*Bull. Amer. Math. Soc.*, 84 (1978)) constructed an infinite-dimensional compact metric space X such that if S is an arbitrary subset of X with $\dim S>0$ then $\dim S=\infty$.

References

[1] K. Morita, Dimension theory (in Japanese), Iwanami, 1950.
[2] W. Hurewicz and H. Wallman, Dimension theory, Princeton Univ. Press, 1941.
[3] J. Nagata, Modern dimension theory, Noordhoff, second edition, 1965.
[4] K. Nagami, Dimension theory, Academic Press, 1970.
[5] P. S. Aleksandrov, The present status of the theory of dimension, Amer. Math. Soc. Transl., (2) 1 (1955), 1–26. (Original in Russian, 1951.)
[6] L. E. J. Brouwer, Beweis der Invarianz der Dimensionenzahl, Math. Ann., 70 (1911), 161–165.
[7] H. Lebesgue, Sur la non-applicabilité de deux domaines appartenant respectivement à des espaces à n et $n+p$ dimensions, Math. Ann., 70 (1911), 166–168.
[8] L. E. J. Brouwer, Beweis der Invarianz des n-dimensionalen Gebiets, Math. Ann., 71 (1912), 305–313.
[9] L. E. J. Brouwer, Über den natürlichen Dimensionsbegriff, J. Reine Angew. Math., 142 (1913), 146–152.
[10] P. Uryson, Les multiplicités cantoriennes, C. R. Acad. Sci. Paris, 175 (1922), 440–442.
[11] K. Menger, Dimensionstheorie, Teubner, 1928.
[12] A. Pears, Dimension theory for general spaces, Cambridge Univ. Press, 1975.
[13] R. Engelking, Dimension theory, North-Holland, 1978.
[14] K. Morita, Dimension of general topological spaces, Surveys in General Topology, G. M. Reed (ed.), Academic Press, 1980.

118 (V.9) Diophantine Equations

A. General Remarks

A **Diophantine equation** is an †algebraic equation whose coefficients lie in the ring **Z** of rational integers and whose solutions are sought in that ring. The name comes from Diophantus, an Alexandrian mathematician of the third century A.D., who proposed many Diophantine problems; but such equations have a very long history, extending back to ancient Egypt, Babylonia, and Greece. As early as the sixth century B.C., Pythagoras is said to have partially solved the equation $x^2+y^2=z^2$ by $x=2n+1$, $y=2n^2+2n$, $z=y+1$. A general solution is given by the Pythagorean numbers $x=m^2-n^2$, $y=2mn$, $z=m^2+n^2$. †Fermat's problem also concerns a Diophantine equation.

Systematic studies of Diophantine equations over **Z** have been made for the linear equation $\sum_{i=1}^{n} a_i x_i = a$ $(a_i, a \in \mathbf{Z})$ and for the quadratic equation $ax^2+bxy+cy^2=k$ $(a,b,c,k \in \mathbf{Z})$ in two unknowns. The latter forms a principal topic of C. F. Gauss's *Disquisitiones arithmeticae* and can be regarded as a starting point of modern algebraic number theory. The special quadratic equation $t^2-Du^2=\pm 4$ $(D \in \mathbf{Z})$ is called **Pell's equation**. If $D<0$, then Pell's equation has only a finite number of solutions. If $D>0$, then all solutions t_n, u_n of Pell's equation are given by $\pm((t_1+u_1\sqrt{D})/2)^n=(t_n+u_n\sqrt{D})/2$, provided that the pair t_1, u_1 is a solution with the smallest $t_1+u_1\sqrt{D}>1$ [15]. Using continued fractions (→ 83 Continued Fractions), we can determine t_1, u_1 explicitly. A general quadratic Diophantine equation $ax^2+bxy+cy^2=k$ with two unknowns can be solved completely if we use solutions of Pell's equation; this is an application of the arithmetic of quadratic fields (→ 347 Quadratic Fields) [1]. On quadratic Diophantine equations of several unknowns, there are deep studies by C. L. Siegel (→ 348 Quadratic Forms).

Diophantine problems consist of giving criteria for the existence of solutions of algebraic equations in rings and fields and eventually determining the number of such solutions. The fundamental ring of interest is **Z** and the fundamental field of interest is **Q**. One discovers rapidly, however, that to have all the technical maneuverability necessary for handling general problems, one must consider rings and fields of finite type over **Z** and **Q**. Furthermore, one is led to consider finite fields and local fields when one deals with a localization of the problems under consideration.

Techniques from various fields of mathematics, e.g., algebraic number theory, algebraic geometry, analysis, Diophantine approximation, etc., have been successfully applied to solve Diophantine problems. However, much remains unsolved today. Yu. V. Matiyasevich (1970) showed that Hilbert's tenth problem is unsolvable; there is no general method of telling whether a Diophantine equation has a solution. This theorem in a sense indicates the complexity of Diophantine problems. For many centuries, no other topic has engaged the attentions of so many mathematicians, both professional and amateur, or resulted in so many published papers. For these miscellaneous results → Dickson [1] and Mordell [2].

B. Equations over Finite Fields

Let k be a finite field of characteristic p consisting of q $(=p^f)$ elements. **Chevalley's theorem**: Let f be a †form of degree d in n variables with coefficients in k such that $d<n$. Then $f=0$ has a nontrivial solution in k. A generalization is **Warning's theorem**: Let $f_1, \ldots, f_t$ be polynomials with coefficients in k in n variables of degrees $d_1, \ldots, d_t$, respectively, and suppose that $d=d_1+\ldots+d_t<n$. Then the number N of common zeros of $f_1, \ldots, f_t$ satisfies $N \equiv 0 \pmod p$. **Warning's second theorem** asserts that if $N>0$ then $N \geqslant q^{n-d}$. Warning's theorem was also improved by J. Ax (1964) to the effect that $N \equiv 0 \pmod{q^b}$ for any integer $b<n/d$ [3]. For equations over finite fields, counting the number of solutions is important. Let $f(x, y)$ be an †absolutely irreducible polynomial in x and y over k. Let N denote the number of zeros in k of $f(x, y)$. A. Weil proved $|N-q| \leqslant 2g\sqrt{q}+c(d)$, where g is the genus of the curve $f(x, y)=0$ and $c(d)$ is a constant depending on d. Weil's proof requires the use of deep results from algebraic geometry. This theorem is equivalent to the †Riemann hypothesis for algebraic curves over finite fields [4]. Later, using Stepanov's method, W. M. Schmidt and E. Bombieri (1973) independently gave new proofs which do not depend on algebraic geometry [3]. P. Deligne [5] proved a far-reaching generalization of Weil's theorem to †nonsingular absolutely irreducible equations in n variables. He showed $|N-q^{n-1}| = O(q^{(n-1)/2})$. This is a part of the †Weil conjecture for zeta functions of algebraic varieties over finite fields (→ Section E; 450 Zeta Functions Q). Schmidt obtained in an elementary manner a weaker estimate $|N-q^{n-1}|=O(q^{n-3/2})$ but without the assumption of nonsingularity (→ Section F).

C. Equations over Local Fields

A method of solving problems in number theory by use of embeddings of the ground field into its †completions is called a local method. Such methods have important consequences when applied to Diophantine equations. Let f be a polynomial in n variables with rational integer coefficients. The congruence $f \equiv 0 \pmod{p^k}$ is solvable for all $k \geqslant 1$ if and only if $f=0$ is solvable in †p-adic integers. This is an easy consequence of the compactness of the ring of p-adic integers. We can solve $f=0$ in p-adic integers provided that we can solve an infinite sequence of congruences. It is generally difficult to tell when we may limit our consideration to only a finite number of these. In this respect, the following lemma is most useful. **Hensel's lemma**: Let $f(x_1, \ldots, x_n)$ be a polynomial whose coefficients are p-adic integers. Let $\gamma_1, \ldots, \gamma_n$ be p-adic integers such that for some i $(1 \leqslant i \leqslant n)$ and an integer $\delta \geq 0$, we have $f(\gamma_1, \ldots, \gamma_n) \equiv 0 \pmod{p^{2\delta+1}}$ and $\partial f/\partial x_i(\gamma_1, \ldots, \gamma_n) \equiv 0 \pmod{p^\delta}$, $\not\equiv 0 \pmod{p^{\delta+1}}$. Then there exist p-adic integers $\theta_1, \ldots, \theta_n$, such that $f(\theta_1, \ldots, \theta_n)=0$ and $\theta_i \equiv \gamma_i \pmod{p^{\delta+1}}$ for $i=1, \ldots, n$. The case $\delta=0$ is often useful; it implies that a nonsingular solution mod p can be extended to a p-adic solution. Generalization to simultaneous equations is also known [6]. Skolem's method is sometimes useful when we investigate certain types of equations over †local fields. This method is based on some simple properties of local analytic manifolds over local fields [7]. If a quadratic form has zeros in each local field, then it has a rational zero (†Minkowski-Hasse theorem). When a theorem of this type holds, we say that the †Hasse principle holds (→ 348 Quadratic Forms). For forms of higher degree, the Hasse principle no longer holds even if the forms are absolutely irreducible and nonsingular. Counterexamples were first found for cubic (E. S. Selmer, 1951) and quintic (M. Fujiwara, 1972) forms. Asymptotic formulas in Waring's problem (→ 4 Additive Number Theory E) can be regarded as an analytic form of the Hasse principle. As to the quantitative formulation of the Hasse principle, there are deep results of Siegel for quadratic forms and their generalization by T. Tamagawa. R. Brauer (1945) showed that forms in sufficiently many variables represent zero in all p-adic fields. Forms of odd degree represent zero in the field of rational numbers if the number of variables is sufficiently large compared with the degree (B. J. Birch, 1957). Let f be a polynomial with p-adic integer coefficients and c_m $(m \geqslant 0)$ be the number of solutions to the congruence $f \equiv 0 \pmod{p^m}$. The series $\varphi(t)= \sum_{m=0}^{\infty} c_m t^m$ is called the Poincaré series of f.

J. Igusa (1975) proved, by using his theory of asymptotic expansions together with Hironaka's †resolution theorem, that $\varphi(t)$ is a rational function of t [8].

D. Integral Solutions of Some Diophantine Equations

In this section we are concerned with those equations for which some "theory" exists. For isolated results → [1,2].

(1) Binary Forms. Thue's theorem (1908): If $f(x)=\sum_{v=0}^{n}a_v x^v (a_v \in \mathbf{Z}, n>2)$ has distinct roots, then the number of rational integral solutions of $\sum_{v=0}^{n}a_v x^v y^{n-v}=a$ $(\mathbf{Z}\ni a\neq 0)$ is finite. This theorem is a direct consequence of Thue's theorem on Diophantine approximation, which says that there are only a finite number of rational numbers p/q $(p, q\in\mathbf{Z}, q>0)$ with $|\alpha-p/q|<1/q^{(n/2)+1}$ for a given algebraic number α of degree n $(n>2)$ [9, p. 122]. K. F. Roth proved that $(n/2)+1$ in this formula can be replaced by $2+\varepsilon$ (ε is an arbitrary positive number independent of n) (*Mathematika*, 2 (1955), 1–20). **Roth's theorem** was generalized to some cases of number fields and function fields (→ 182 Geometry of Numbers) and is applied to Diophantine equations [9, 10]. A. Baker (1968), using a completely different method, has given explicit upper bounds for the solutions of Thue's equations, thus enabling one to compute effectively all the solutions. More precisely, if f in Thue's equation is irreducible over the rationals, then every integer solution (x, y) of the equation satisfies $\max(|x|,|y|)<\exp((nH)^{(10n)^5}+(\log a)^{2n+2})$, where H is the †height of f. The proof of this remarkable theorem is based on his deep result concerning the lower bound for the linear forms in the logarithm of algebraic numbers (*Mathematica*, 13 (1966); 14 (1967)). Baker's method has been applied to elliptic, hyperelliptic, and other curves (Baker, H. Stark, J. Coates, V. G. Sprindzhuk, etc.).

(2) Higher-Degree Forms. A natural generalization of binary forms is a **norm form**. Let K be an algebraic number field of degree $t\geqslant 3$ and $\alpha_1, \ldots, \alpha_n$ be elements of K. Then the norm $N(\alpha_1x_1+\ldots+\alpha_nx_n)=\prod_{i=1}^{t}(\alpha_1^{(i)}x_1+\ldots+\alpha_n^{(i)}x_n)$, where $\alpha^{(i)}$ denotes a conjugate of α, is a form of degree t with rational coefficients. It is easy to see that every form which has rational coefficients and is irreducible over $\mathbf{Q}$ but which is a product of linear forms with algebraic coefficients is a constant multiple of a norm form [7]. A module M in an algebraic number field K is called **degenerate** if M has a submodule N such that, for some $\alpha\in K$, αN is a full module in some subfield K' of K, where K' is neither $\mathbf{Q}$ nor an imaginary quadratic field. The most basic in the theory of norm forms is (W. M.) **Schmidt's theorem**: Let $\alpha_1, \ldots, \alpha_n$ be linearly independent over $\mathbf{Q}$ and suppose that the module generated by $\alpha_1, \ldots, \alpha_n$ is nondegenerate; then $N(\alpha_1x_1+\ldots+\alpha_nx_n)=c$ $(c\in\mathbf{Q})$ has only finitely many solutions in integers $x_1, \ldots, x_n$. (*Math. Ann.*, 191 (1971)). The proof is based on his remarkable result on †simultaneous approximation which generalizes Roth's theorem (→ 182 Geometry of Numbers G). There are investigations on special norm forms (T. Skolem, N. I. Fel'dman, K. Ramanathan, K. Györy, M. Fujiwara, etc.). For general forms of higher degree, not much is yet known except for the additive forms (→ 4 Additive Number Theory E). H. Davenport [11] proved that if $f(x)$ is a cubic form with rational integer coefficients in n variables, then $f(x)=0$ has a nontrivial integral solution, provided that $n\geqslant 16$. This theorem was proved by means of an exquisite application of the †circle method together with some geometry of numbers. A well-known conjecture is that $n\geqslant 10$ instead of $n\geqslant 16$. It is known that over local fields, any cubic form in 10 variables has a nontrivial zero. There are various results of this type for simultaneous additive, quadratic, and cubic forms (Davenport, D. Lewis, R. Cook, Schmidt, etc.) [17]. A satisfactory theory of forms of higher degree, like that of quadratic forms, is not yet known but is quite desirable. In this vein, Igusa has obtained some new results of considerable interest, e.g., a †Poisson summation formula for higher-degree forms, using his theory of asymptotic expansions [8].

(3) Algebraic Curves. The fundamental result is **Siegel's theorem** (1929): Assume that the equations $f_i(X_1, \ldots, X_n)=0$ $(1\leqslant i\leqslant m)$ determine an algebraic curve with a positive †genus in an †affine space of dimension n. Then the number of rational integral solutions of $f_i(X_1, \ldots, X_n)=0$ $(1\leqslant i\leqslant m)$ is finite. This theorem was generalized by S. Lang in the following form: Let K be a finitely generated field over $\mathbf{Q}$ and I a subring of K that is finitely generated over $\mathbf{Z}$. Furthermore, let C be a nonsingular projective algebraic curve with a positive genus defined over K, and let φ be a rational function on C defined over K. Then there are only a finite number of points P on C with $\varphi(P)\in I$ [10]. The proof of this theorem is based on a generalization of Roth's theorem in the above sense and on the weak Mordell-Weil theorem (→ Section C). A. Robinson and P. Roquette gave another approach to Siegel's theorem from the standpoint of nonstandard arithmetic (*J. Number Theory* 7 (1975)). On the other hand, a necessary condition for the exis-

tence of infinitely many solutions of $f(X, Y)=0$ with rational integral coefficients was given by C. Runge (*J. Reine Angew. Math.*, 100 (1887)).

(4) Elliptic Curves. An elliptic curve E is an †Abelian variety of dimension 1, or what is the same, an irreducible nonsingular †projective algebraic curve of †genus 1 furnished with a point O as origin. The †Riemann-Roch theorem defines a group law on the set of †divisor classes of E. Actually, if P, P' are points of E, then there exists a unique point P'' such that $(P)+(P')\sim(P'')+(O)$, where $\sim$ means linear equivalence, i.e., the left-hand side minus the right-hand side is the divisor of a rational function on the curve. The group law on E is then defined by $P+P'=P''$. If the characteristic $\neq 2$ or 3, using the Riemann-Roch theorem one finds that the curve E can be defined by a Weierstrass equation $y^2=x^3+ax+b$ with a, b in the ground field over which the curve is defined. Conversely, any homogeneous nonsingular cubic equation has genus 1 and defines an elliptic curve in the projective plane once the origin has been selected. If both the curve and the origin are defined over a field k, then the group law is also defined over k, and it becomes a 1-dimensional Abelian variety defined over k. If the ground field k is the field of complex numbers, the group law is the same as that given by the †addition theorem of the †Weierstrass $\wp$-function with invariants $g_2=-4a$ and $g_3=-4b$ through the parametrization $x=\wp(u)$, $y=\frac{1}{2}\wp'(u)$. Much of the Diophantine theorems on elliptic curves are generalized to Abelian varieties. Here we shall deal mainly with elliptic curves defined by Weierstrass equations over $\mathbf{Q}$. Extension to algebraic number fields usually causes no trouble. For more general elliptic curves → [19]. The Lutz-Mattuck theorem (→ Section E) obviously implies that the points of finite order in E_k $(k=\mathbf{Q}_p)$ form a finite group. This torsion group is computable. In case a and b are in $\mathbf{Z}$ then any point of finite order in $E_{\mathbf{Q}}$ has coordinates (x, y) in $\mathbf{Z}$ and, if $y\neq 0$, $y^2 \mid 4a^3+27b^2$ (Lutz-Nagell). The **WC group** (Weil-Châtelet group) of E over k is the birational class of principal homogeneous spaces over k. The extent of validity of the Hasse principle for elliptic curves can be measured by the **Tate-Shafarevich group,** which is defined as the set of elements of the WC group that are everywhere locally trivial. This group is conjectured to be a finite group. For other results and interesting conjectures → [12–14]. The number of integral points on elliptic curves is finite according to Siegel's theorem on algebraic curves. Explicit bounds for the size of these points have been given for several types of elliptic curves by using Baker's method. For example, if $f(x, y)$ is an absolutely irreducible polynomial with coefficients in $\mathbf{Z}$ such that the curve $f=0$ has genus 1, then $\max(|x|, |y|)<\exp\exp\exp((2H)^m)$, where $m=10^{d^{10}}$, $d=\deg f$, and H is the height of f (Baker and Coates, 1970. The method of proof was to reduce it to the Weierstrass equation case, which had been treated earlier by Baker, with a better bound.) By the †Mordell-Weil theorem (→ Section E), $A_{\mathbf{Q}}\cong \mathbf{Z}^r\times$ finite torsion group. Here r is called the **rank** of E over $\mathbf{Q}$. There is a rather doubtful conjecture to the effect that the rank r is bounded. The rank r is conjectured to be equal to the order of the zero of $L(s, E)$ at $s=1$ (**Birch-Swinnerton-Dyer conjecture**). Much numerical and theoretical evidence supports this famous conjecture [13].

E. Rational Points of Algebraic Varieties

Let V be an †abstract algebraic variety defined over a field k, and let P be a point of V. Then P is called a **rational point** over k of V if the coordinates of the †representative P_α contained in an †affine open set V_α of V are in k (→ 16 Algebraic Varieties D). This definition is independent of the choice of the representative P_α. In particular, if V is a †projective variety, the point P given by the †homogeneous coordinates $(x_0, x_1, \dots, x_n)$ is rational if and only if $x_i/x_p \in k$ $(0\leqslant i\leqslant n,\ x_p\neq 0)$. In the following we state main results concerning rational points of algebraic varieties, especially results concerning †Abelian varieties, restricting k to be either an †algebraic number field of finite degree, a †p-adic number field, or a †finite field.

Mordell-Weil Theorem. Let A be an Abelian variety of dimension n defined over an algebraic number field k of finite degree. Then the group A_k of all k-rational points on A is finitely generated. This theorem was proved by L. J. Mordell (1922) for the case of $n=1$ and by Weil (1928) for the general case [10]. The assertion that A_k/mA_k is a finite group for any rational integer m is called the **weak Mordell-Weil theorem**; this theorem is basic in the proof of the Mordell-Weil theorem and is used in the proof of Siegel's theorem, too. A generalization of the Mordell-Weil theorem is obtained when k is a field (of arbitrary characteristic) finitely generated over the †prime field [10].

If A is defined over a finite algebraic number field k, we have the following conjectures of Birch, Swinnerton-Dyer, and Tate on the rank of A_k. Let $\mathfrak{p}$ be a prime ideal of k at which A has a good †reduction, and denote by $A_{\mathfrak{p}}$ the reduced variety. Let $\pi_{\mathfrak{p}}^{(1)}, \dots, \pi_{\mathfrak{p}}^{(2n)}$ be the eigen-

values of the $N(\mathfrak{p})$th power endomorphism of $A_\mathfrak{p}$ with respect to an l-adic representation, where $N(\mathfrak{p})$ denotes the norm of $\mathfrak{p}$ (→ 3 Abelian Varieties E, N), and put $L_\mathfrak{p}(s, A) = \prod_{i=1}^{2n} (1 - \pi_\mathfrak{p}^{(i)} N(\mathfrak{p})^{-s})^{-1}$. The L-function of A defined by $L(s, A) = \prod' L_\mathfrak{p}(s, A)$, where the product ranges over all good primes, is the principal part of the zeta function of A (→ 450 Zeta Functions S). Birch and Swinnerton-Dyer conjectured that if $k = \mathbf{Q}$ and A is of dimension 1, then there exists a constant $C \neq 0$ such that $L(s, A) \sim C(s-1)^g$ as $s \to 1$. Tate generalized this conjecture to any A and k. Moreover, the constant C, appropriately modified by factors corresponding to the bad primes and the infinite primes, is thought to be expressible in terms of certain arithmetic invariants of A [13].

Lutz-Mattuck Theorem. The group of rational points of an Abelian variety A of dimension n over a †$\mathfrak{p}$-adic number field k contains a subgroup of finite index isomorphic to the direct sum of n copies of the †ring $\mathfrak{o}$ of $\mathfrak{p}$-adic integers in k (E. Lutz, *J. Reine Angew. Math.*, 177 (1937); A. Mattuck, *Ann. Math*, 62 (1955)).

Mordell's Conjecture. In his 1922 paper, in which the above theorem on the set of rational fields on 1-dimensional Abelian varieties (i.e., on elliptic curves) was established, Mordell stated the conjecture: Any algebraic curve of genus $g \geqslant 2$ defined over $\mathbf{Q}$ has only a finite number of rational points. The same can be conjectured for such curves defined over any algebraic number field k of finite degree. This had remained as a conjecture until 1983. In 1961 I. R. Shafarevich conjectured: Let k be any algebraic number field of finite degree, S a finite set of finite prime spots of k, and g any natural number $\geqslant 2$. Then there are, up to k-isomorphism, only a finite number of nonsingular algebraic curves of genus g defined over k having good reduction at every finite prime spot outside S.

In 1973, A. N. Parshin showed that Mordell's conjecture followed from this conjecture, which was finally proved in 1983 by G. Fallings [7]. Mordell's conjecture was thus settled in the affirmative. Analogs of these conjectures on algebraic function fields over finite fields are easier than the original ones and had been proved in the 1960s for Mordell's conjecture by Yu. I. Manin, H. Grauert, and M. Miwa, and for Shafarevich's conjecture by S. Alakelov, A. N. Parshin, and L. Szpiro.

F. C_i-Fields

Let F be a field, and let $i \geqslant 0$, $d \geqslant 1$ be integers. Let f be a homogeneous polynomial of n variables of degree d with coefficients in F. If the equation $f=0$ has a solution $(x_1, \ldots, x_n) \neq (0, \ldots, 0)$ in F for any f with $n > d^i$, then F is called a **$C_i(d)$-field**. If F is a $C_i(d)$-field for any $d \geqslant 1$, then F is called a **C_i-field**. In order for F to be a C_0-field, it is necessary and sufficient that F be an †algebraically closed field. A C_1-field is sometimes called a **quasi-algebraically closed field**. There exists no noncommutative algebra over a C_1-field F. A finite field is C_1 (C. Chevalley (1936)). If F_0 is algebraically closed, then $F = F_0(X)$ (rational function field of one variable) is a C_1-field **(Tsen's theorem)**. A homogeneous polynomial f of $n = d^i$ variables of degree d with coefficients in F such that $f=0$ has no solution in F except $(0, \ldots, 0)$ is called a **normic form** of order i in F. If a C_i-field F_0 has at least one normic form of order i, then (i) $F_0(X_1, \ldots, X_k)$ is a C_{i+k}-field; and (ii) an extension of F_0 of finite degree is a C_i-field. A complete field F with respect to an †exponential valuation is a C_1-field whenever its residue field F_0 is algebraically closed. The field F of power series of one variable over a finite field F_0 is a C_2-field (Lang). E. Artin conjectured that a p-adic field $\mathbf{Q}_p$ is a C_2-field. It was proved by H. Hasse (1923) that $\mathbf{Q}_p$ is a $C_2(2)$-field and by D. Lewis (1952) that $\mathbf{Q}_p$ is a $C_2(3)$-field. However, G. Terjanian (1966) [17] gave a counterexample to Artin's conjecture; that is, he gave a quartic form of 18 variables with coefficients in $\mathbf{Q}_2$ having only trivial zero in $\mathbf{Q}_2$. Ax and S. Kochen (1965) [18] proved that for any integer $d \geqslant 1$ there exists an integer $p_0(d)$ such that $\mathbf{Q}_p$ is a $C_2(d)$-field for $p > p_0(d)$ (→ 276 Model Theory E).

References

[1] L. Dickson, History of the theory of numbers I–III, Chelsea, 1952.
[2] L. J. Mordell, Diophantine equations, Academic Press, 1969.
[3] W. M. Schmidt, Equations over finite fields, Lecture notes in math. 536, Springer, 1976.
[4] A. Weil, Sur les courbes algébriques et les variétés qui s'en déduisent, Actualites Sci. Ind., 1041 (1948).
[5] G. Faltings, Endlichkeitssätze für abelische Varietäten über Zahlkörpern, Inventiones Math., 73 (1983), 349–366.
[6] M. Greenberg, Lectures on forms in many variables, Benjamin, 1969.
[7] Z. I. Borevich and I. R. Shafarevich, Number theory, Academic Press, 1966. (Original in Russian, 1964.)
[8] J. Igusa, Lectures on forms of higher degree, Tata Inst., 1978.
[9] W. J. LeVeque, Topics in number theory, Addison-Wesley, 1956.

[10] S. Lang, Diophantine geometry, Interscience, 1962.
[11] H. Davenport, Analytic methods for Diophantine equations and Diophantine inequalities, Univ. of Michigan, 1962.
[12] J. W. S. Cassels, Diophantine equations with special reference to elliptic curves, J. London Math. Soc., 41 (1966), 193–291.
[13] H. P. F. Swinnerton-Dyer, The conjectures of Birch and Swinnerton-Dyer and of Tate, Proc. Conf. on Local Fields, Driebergen, Springer, 1967.
[14] S. Lang, Elliptic curves: Diophantine analysis, Springer, 1978.
[15] I. Niven and H. Zuckerman, An introduction to the theory of numbers, Wiley, 1960.
[16] J. Joly, Equations et variétés algébriques sur un corps fini, Enseignement Math., (2) 19 (1973), 1–117.
[17] D. Lewis, Diophantine equations: p-adic method, Studies in Math., vol. 6, Math. Assoc. Amer., 1969.
[18] J. Ax and S. Kochen, Diophantine problems over local fields I, II, Amer. J. Math., 87 (1965), 605–630, 631–648.
[19] J. Tate, The arithmetic of elliptic curves, Inventiones Math., 23 (1974), 170–206.

119 (XXI.20) Dirichlet, Peter Gustav Lejeune

Peter Gustav Lejeune Dirichlet (February 13, 1805–May 5, 1859) was born of a French family in Düren, Germany. From 1822 to 1827 he was in Paris, where he became a friend of J.-B. †Fourier. In 1827, he was appointed lecturer at the University of Breslau; in 1829, lecturer at the University of Berlin; and in 1839, professor at the University of Berlin. In 1855, he was invited to succeed C. F. †Gauss at the University of Göttingen, where he spent his last four years as a professor.

His works cover many aspects of mathematics; however, those on number theory, analysis, and potential theory are the most famous. He greatly admired Gauss and is said to have kept Gauss's *Disquisitiones arithmeticae* at his side even when traveling.

In number theory, he created the †Dirichlet series and proved that a sequence in arithmetic progression contains infinitely many prime numbers, provided that the first term and the common difference are relatively prime. Also, using his "drawer principle," which states that if there are $n+1$ objects in n drawers then at least one drawer contains at least 2 objects, he clarified the structure of †unit groups of †algebraic number fields. In †potential theory he dealt with the †Dirichlet problem concerning the existence of †harmonic functions. He also gave †Dirichlet's condition for the convergence of trigonometric series.

References

[1] P. G. L. Dirichlet's Werke I, II, Reimer, 1889, 1897 (Chelsea, 1969).
[2] P. G. L. Dirichlet, Vorlesungen über Zahlentheorie, with supplements by R. Dedekind, Vieweg, fourth edition, 1894 (Chelsea, 1968).
[3] F. Klein, Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert I, Springer, 1926 (Chelsea, 1956).

120 (X.30) Dirichlet Problem

A. The Classical Dirichlet Problem

Let D be a bounded or unbounded †domain in $\mathbf{R}^n$ $(n \geqslant 2)$ with compact boundary S. The classical **Dirichlet problem** is the problem of finding a †harmonic function in D that assumes the values of a prescribed continuous function on S. This problem is also called the †first boundary value problem (→ 193 Harmonic Functions and Subharmonic Functions). In this article f always stands for a boundary function given on S. The problem is called an **interior problem** if D is bounded and an **exterior problem** if D is unbounded. In an exterior problem, it is further required that when an †inversion with center at an exterior point P_0 of D is performed on D and a †Kelvin transformation is performed on the solution in D (when the solution exists), the function thus obtained on the inverted image of D be harmonic at P_0 $(n \geqslant 3)$. When $n=2$, the solution in D, which is already regarded as a function on the inverted image of D, is required to be harmonic at P_0. Thus an interior problem can be transformed to an exterior problem, and vice versa. We now explain the history of the classical problem.

Let D be a bounded domain with boundary S in $\mathbf{R}^3$. G. Green (1828) asserted that if S is sufficiently smooth,

$$u(P) = -\frac{1}{4\pi}\int_S f(Q)\frac{\partial G(P,Q)}{\partial n_Q}\,d\sigma(Q) \qquad (1)$$

is the solution for the Dirichlet problem, where f is prescribed on S, $G(P,Q)$ is †Green's function with the pole at Q in D, n_Q is the outward-drawn normal to S at Q, and $d\sigma$ is the †sur-

face element on S. He took for granted the existence of Green's function from physical consideration of the problem. Thus his discussion was not quite rigorous. This defect was corrected by A. M. Lyapunov (1898) under a certain condition on S. Denote by u_m the Newtonian potential of a measure with density $m \geqslant 0$ on S. Assume that a continuous function f on S and a positive constant a are given. In 1840, C. F. Gauss investigated the existence of a density $m_f \geqslant 0$ on S of total mass a which satisfies $\int_S (u_{m_f} - 2f) m_f d\sigma = \min_m \int_S (u_m - 2f) m \, d\sigma$, where the total mass of m is equal to a. He asserted also that $u_{m_f} - f$ is equal to a constant b on S. If $f \equiv 0$, then u_{m_0} must be equal to a positive constant c on S, and hence $u_{m_f} - bc^{-1} u_{m_0}$ must be a solution of the exterior problem for the boundary function f on S. However, his discussion was incomplete because we cannot always ensure the existence of a density that gives a measure minimizing the integral. Moreover, even in the case where D is a ball, there exists a continuous function $f \geqslant 0$ on S such that there is no Newtonian potential that is equal to f on S up to a constant (M. Ohtsuka, 1961). Gauss (1840), W. Thomson (Lord Kelvin) (1847), and G. L. Dirichlet solved the Dirichlet problem by making use of the so-called **Dirichlet principle**, which is explained in detail in Section F. After K. Weierstrass (1870) pointed out that there is a case where no minimizing function exists, D. Hilbert (1899) gave a rigorous proof of the Dirichlet principle under a certain condition. Meanwhile, C. G. Neumann (1870) solved the Dirichlet problem rigorously for the first time, although he assumed that D is a convex domain with a smooth boundary. First, he considered the potential $W_1 = (1/2\pi) \cdot \int_S f(\partial r^{-1}/\partial n) d\sigma$ of a double layer in D; then he formed the potential $W_2 = (1/2\pi) \cdot \int_S f_1 (\partial r^{-1}/\partial n) d\sigma$ of a double layer with the values f_1 of W_1 on S and defined $W_3, W_4, \ldots$ similarly. The series $W_1 - W_2 + W_3 - W_4 + \ldots$ plus a suitable constant gives a solution to the exterior Dirichlet problem for the boundary function f. In 1887, H. Poincaré also used (1) to solve the Dirichlet problem. He obtained Green's function with the pole at O in a bounded domain D in the following manner: Let D' be the image of D by an inversion with center O, S_0 be a spherical surface surrounding the boundary $\partial D'$ of D', and μ be a uniform measure on S_0 such that the potential of μ is equal to 1 inside S_0. By †sweeping out μ to $\partial D'$, the solution in D' of the exterior problem for the boundary function 1 is obtained. A †Kelvin transformation of this solution yields the solution h in D of the interior problem for the boundary function $1/\overline{OP}$. Now $1/\overline{OP} - h(P)$ is Green's function in D. In 1899 Poincaré used another method (without utilizing (1)) to solve the Dirichlet problem [8]. He observed that it is sufficient to consider the case where f is equal to the restriction to S of a polynomial g and that g is expressed in D as the sum of the Newtonian potential of a signed measure τ with density $-\Delta g/(4\pi)$ and a function that is harmonic in D and continuous on $D \cup S$. If it is possible to sweep out τ to ∂D, then the solution is obtained. He showed that this is in fact the case if at every point P of S there exists a cone that is disjoint from D and has its vertex at P. This condition is called **Poincaré's condition.** In 1900, I. Fredholm discussed the Dirichlet problem by reducing it to a problem of †integral equations. A domain D is called a **Dirichlet domain** if the (classical) Dirichlet problem is always solvable in D. H. Lebesgue (1912) showed that a solution is obtained by the method of iterative averaging in every Dirichlet domain.

B. The Dirichlet Problem in a General Domain

It has been believed that the classical Dirichlet problem is always solvable in every domain until S. Zaremba observed in 1909 that the problem is not always solvable for a punctured ball. In 1913, Lebesgue gave a decisive example in which the domain is homeomorphic to a ball and bounded by a surface sufficiently smooth except at one point. Thus the central interest shifted to finding a harmonic function in D that depends only on a continuous function f given on ∂D and coincides with the classical solution when D is a Dirichlet domain. Extend f to a continuous function in the whole space, and denote it by f_0. Approximate D by an increasing sequence $\{D_n\}$ of Dirichlet domains, and denote by u_n the solution in D_n of the Dirichlet problem for the boundary function f_0. N. Wiener proved in 1924 that u_n coverges to a harmonic function that is independent of the choice of the extension of f and $\{D_n\}$. The problem of deciding where on ∂D the general solution assumes the given boundary values is treated in Section D. O. D. Kellogg (1928) found a general method that includes Poincaré's method of sweeping out, Schwarz's alternating method, and the result of Wiener. Both Poincaré's method of sweeping out and Lebesgue's method of iterative averaging yield Wiener's general solution.

C. Perron's Method

We explain O. Perron's method (1923) by considering the improved method of M. Brelot [1]. For simplicity we assume that the domain

D is bounded in $\mathbf{R}^3$. Let U be the family of subharmonic functions u bounded above and satisfying $\limsup_{P \to M} u(P) \leqslant f(M)$ for any boundary point M. Define $\underline{H}_f(P)$ as $\sup_u u(P)$, where u runs through U, if this family is not empty; otherwise, set $\underline{H}_f \equiv -\infty$. Call $\underline{H}_f$ a **hypofunction**. Define $\bar{H}_f$ by $-\underline{H}_{-f}$ and call it a **hyperfunction**. If $\underline{H}_f = \bar{H}_f$, the common function is denoted by H_f; if $H_f(P) < \infty$, then H_f is harmonic. This is called a **Perron-Brelot solution (Perron-Wiener-Brelot solution** or simply PWB solution). The method of defining H_f is called **Perron's method** (the Perron-Brelot method or the Perron-Wiener-Brelot method).

Wiener showed in 1923 that the †Daniell-Stone integral can be regarded as a general solution if a †Daniell-Stone integrable function f is given on the boundary of a Dirichlet domain; in 1925, he showed that the same is true for a general domain (not necessarily Dirichlet). He showed also that his solution coincides with the Perron-Brelot solution H_f if f is continuous. Unfortunately, however, from a wrong example he concluded that $\underline{H}_f \neq \bar{H}_f$ can hold even for a simple discontinuous f, and so he lost interest in Perron's method. Brelot (1939) corrected Wiener's erroneous conclusion and proved that the Daniell upper and lower integrals are equal to $\bar{H}_f$ and $\underline{H}_f$, respectively. To any continuous f there corresponds an H_f, and there exists a †Radon measure μ_P satisfying $H_f(P) = \int f d\mu_P$. This measure is called a **harmonic measure** or harmonic measure function. Brelot showed that $\underline{H}_f = \bar{H}_f$ if and only if f is μ_P-integrable for one (or every) P. In particular, If D is a Dirichlet domain and E is a closed set on the boundary ∂D, then the harmonic measure function $\mu_P(E)$ takes the value 1 at an inner point (in the space ∂D) of E and vanishes on $\partial D - E$. We note that μ_P is equal to the measure obtained by sweeping out the unit mass at P to ∂D.

D. Regular Boundary Points

If $H_f(P) \to f(M)$ as $P \to M \in \partial D$ for any continuous function f on ∂D, then M is called **regular**. The regularity of a point is a local property. A boundary point that is not regular is called **irregular**. The regularity of M is equivalent to the convergence of μ_P as $P \to M$ to the unit mass at M with respect to the †vague topology. There are many sufficient conditions and necessary conditions for a boundary point to be regular. The existence of a **barrier** is a qualitative condition that is necessary and sufficient for a boundary point to be regular. It was used by Poincaré and so named and used effectively by Lebesgue. A barrier is a continuous superharmonic function in D that assumes the boundary value 0 at M and has a positive lower bound outside every ball with center at M. A positive superharmonic function defined in the intersection of D and a neighborhood of M and taking the boundary value 0 can be used as a barrier. A necessary and sufficient condition for a boundary point M to be regular is the existence of a Green's function in D assuming the value 0 at M. This condition was given by G. Bouligand (1925), and it follows from the existence of a barrier. Another necessary and sufficient condition of a quantitative nature was obtained by Wiener. It is equivalent to the requirement that the complement of D is not †thin at M (→ 338 Potential Theory G). Kellogg conjectured that the set of irregular boundary points is of capacity zero and verified this in $\mathbf{R}^2$ (1928). The conjecture was proved first by G. C. Evans (1933) in $\mathbf{R}^3$, and different proofs were given by F. Vasilesco (1935) and O. Frostman (1935). The conjecture is also true in $\mathbf{R}^n$ for $n \geqslant 4$.

E. The More General Dirichlet Problem

So far, we have been concerned with $\mathbf{R}^n$. More generally, Brelot and G. Choquet [3] obtained the following result in a Green space $\mathscr{E}$ (→ 193 Harmonic Functions and Subharmonic Functions): Consider a metric space that contains $\mathscr{E}$ and in which $\mathscr{E}$ is everywhere dense, and denote by Δ the complement of $\mathscr{E}$ with respect to the space. Let $\{F\}$ be a family of †filters on $\mathscr{E}$ such that each F converges to a certain point of Δ. Suppose that $u \leqslant 0$ whenever u is subharmonic and bounded above on $\mathscr{E}$ and $\limsup u \leqslant 0$ along every F. Assume the existence of a barrier v in a neighborhood in $\mathscr{E}$ of the limit point Q of every F; that is, v is to be positive superharmonic, to tend to 0 along F, and to have a positive lower bound outside every neighborhood of Q. Under these assumptions, we obtain the PWB solution on $\mathscr{E}$ as in $\mathbf{R}^3$. There are various examples of Δ and F that satisfy these conditions. In particular, L. Naïm [6] investigated in detail the case where Δ is the †Martin boundary. More generally, it is possible to treat the Dirichlet problem axiomatically (→ 193 Harmonic Functions and Subharmonic Functions).

F. The Dirichlet Principle

Let D be a bounded domain with a sufficiently smooth boundary in $\mathbf{R}^n$ and f be a piecewise C^1-function on D with finite **Dirichlet integral** $\|f\|^2 = \int_D |\text{grad} f|^2 d\tau$, where $d\tau$ is the volume element. Suppose that f has a continuous boundary value φ on ∂D. The classical Dirich-

let principle asserts that the solution of the Dirichlet problem for φ has the smallest Dirichlet integral among the functions that are piecewise of class C^1 in D and assume the boundary value φ. In a general domain, H_φ minimizes $\|u-f\|$ among harmonic functions u in D. Brelot [2] discussed the principle for a family of competing functions that are defined in a domain in $\mathscr{E}$ and whose boundary values cannot be defined in the classical manner.

References

[1] M. Brelot, Familles de Perron et problème de Dirichlet, Acta Sci. Math. Szeged., 9 (1939), 133–153.
[2] M. Brelot, Etude et extensions du principe de Dirichlet, Ann. Inst. Fourier, 5 (1955), 371–419.
[3] M. Brelot and G. Choquet, Espaces et lignes de Green, Ann. Inst. Fourier, 3 (1952), 199–263.
[4] R. Courant, Dirichlet's principle, conformal mapping, and minimal surfaces, Interscience, 1950.
[5] O. D. Kellogg, Foundations of potential theory, Springer, 1929.
[6] L. Naïm, Sur le rôle de la frontière de R. S. Martin dans la théorie du potentiel, Ann. Inst. Fourier, 7 (1957), 183–281.
[7] M. Ohtsuka, Dirichlet problem, extremal length and prime ends, Lecture notes, Washington Univ., St. Louis, 1962–1963.
[8] H. Poincaré, Théorie du potentiel Newtonien, Leçons professées à la Sorbonne, Gauthier-Villars, 1899.
[9] L. L. Helms, Introduction to potential theory, Interscience, 1969.

121 (XI.3) Dirichlet Series

A. Dirichlet Series

For $z=x+iy$, $\lambda_n>0$, and $\lambda_n\uparrow+\infty$, the series of the form

$$f(z)=\sum_{n=1}^{\infty} a_n\exp(-\lambda_n z) \tag{1}$$

is called a **Dirichlet series** (more precisely, a **Dirichlet series of the type** $\{\lambda_n\}$). If $\lambda_n=n$, then (1) is a power series with respect to e^{-z}. If $\lambda_n=\log n$, the series (1) becomes

$$\sum_{n=1}^{\infty} a_n/n^z, \tag{2}$$

which is called an **ordinary Dirichlet series**. If $a_n=1$, then (2) is the †Riemann zeta function.

Series of the form (2) were introduced by P. G. L. Dirichlet in 1839 and utilized in an investigation of the problems of †analytic number theory. Later J. Jensen (1884) and E. Cohen (1894) extended the variable z to complex numbers. The Dirichlet series is not only a useful tool in analytic number theory, but is also investigated as a generalization of power series. The †Laplace transform is the generalization of the Dirichlet series to the integral, and similar formulas often hold for both cases.

B. Convergence Regions

If the series (1) converges at $z=z_0$, then it converges in the half-plane $\operatorname{Re}z>\operatorname{Re}z_0$. Therefore there is a uniquely determined real number S such that (1) converges in $\operatorname{Re}z>S$ and diverges in $\operatorname{Re}z<S$. If (1) always converges (diverges), we put $S=-\infty$ $(+\infty)$. We call S the **abscissa of convergence** (or **abscissa of simple convergence**). Similarly, there is a uniquely determined real number A such that (1) converges absolutely in $\operatorname{Re}z>A$ and is not absolutely convergent in $\operatorname{Re}z<A$. We call A the **abscissa of absolute convergence**. Furthermore, there is a uniquely determined real number U such that (1) converges uniformly in $\operatorname{Re}z>U'$ for every $U'>U$ and does not converge uniformly in $\operatorname{Re}z>U''$ for every $U''<U$. The number U is called the **abscissa of uniform convergence**. Among these abscissas we always have the relations

$$-\infty\leqslant S\leqslant U\leqslant A+\infty,$$

$$A-S\leqslant\limsup_{n\to\infty}\frac{\log n}{\lambda_n}.$$

The latter was proved by Cohen (1894). The numbers S, A, U are determined from a_n, λ_n by means of the formulas

$$S=\limsup_{x\to\infty}\frac{1}{x}\log\left|\sum_{[x]\leqslant\lambda_n<x} a_n\right|. \tag{3}$$

$$A=\limsup_{x\to\infty}\frac{1}{x}\log\left(\sum_{[x]\leqslant\lambda_n<x}|a_n|\right), \tag{4}$$

$$U=\limsup_{x\to\infty}\frac{1}{x}\log T_x,$$

$$T_x=\sup_{-\infty<y<+\infty}\left|\sum_{[x]\leqslant\lambda_n<x} a_n\exp(-i\lambda_n y)\right|, \tag{5}$$

where [ ] is the †Gauss symbol. Formulas (3) and (4) were proved by T. Kojima (1914) and (5) by M. Kunieda (1916). In particular, when $\lim_{n\to\infty}(\log n)/\lambda_n=0$, we have

$$S=U=A=\limsup_{n\to\infty}\frac{\log|a_n|}{\lambda_n} \tag{6}$$

(O. Szász, 1922; G. Valiron, 1924).

The series (1) converges uniformly in the angular domain $\{z \,|\, |\arg(z-z_0)|<\alpha<\pi/2\}$, where the vertex z_0 lies on the line $\operatorname{Re} z_0=S$. Hence it represents a holomorphic function in the domain $\operatorname{Re} z>S$, but it is possible that there is no singularity on the line $\operatorname{Re} z=S$. For example, if $a_n=(-1)^n$, then the series (2) has $S=0$, but the sum is an †entire function $(2^{1-z}-1)\,\zeta(z)$. Taking the analytic continuation $f(z)$ of the series (1), the infimum R of ρ such that $f(z)$ is holomorphic in $\operatorname{Re} z>\rho$ is called the **abscissa of regularity**. It is still possible that there is no singularity on the line $\operatorname{Re} z=R$. We always have $R\leqslant S$, and R is given by

$$R=\sup_{-\infty<y<\infty}\lim_{x\to-\infty}(\log\log^+|\varphi(x+iy)|+x),$$

$$\varphi(z)=\sum_{n=1}^{\infty}\frac{a_n\exp(-\lambda_n z)}{\Gamma(1+\lambda_n)}, \tag{7}$$

where $\log^+ a=\max(\log a,0)$ (C. Tanaka, 1951). The infimum B of ρ such that $f(z)$ is bounded in $\operatorname{Re} z>\rho$ is called the **abscissa of boundedness**. We always have $R\leqslant B\leqslant A$. H. Bohr proved the following three theorems concerning these values: (1) If $\{\lambda_n\}$ are linearly independent over the ring of integers, then $A=B$ (1911). (2) If $(\lambda_{n+1}-\lambda_n)^{-1}=O(\exp e^{\lambda_n\varepsilon})$ for every $\varepsilon>0$, then $U=B$ (1913). (3) If $\limsup(\log n)/\lambda_n=0$, then $S=U=A=B$ (1913). In the final case, the values are given by (6).

C. Properties of Functions Given by Dirichlet Series

The coefficients a_n in (1) are given in terms of the function $f(z)$ by

$$\sum_{\nu=1}^{n}a_\nu=\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}f(z)\frac{e^{\omega z}}{z}dz, \tag{8}$$

where $c>\max(S,0)$, $\lambda_n<\omega<\lambda_{n+1}$, and the integration contour does not pass through $\{\lambda_n\}$. If $\omega=\lambda_n$, then the term a_n in the sum of the left-hand side of (8) is replaced by $a_n/2$ (O. Perron, 1908). Furthermore, if $S<x$, then we have

$$a_n=\lim_{T\to\infty}\frac{1}{T}\int_{\alpha_0}^{\alpha_0+T}f(x+iy)\exp[\lambda_n(x+iy)]\,dy \tag{9}$$

(J. Hadamard, 1908; C. Tanaka, 1952).

If $x=\operatorname{Re} z>S$, then $f(z)=o(|y|)$ $(|y|\to\infty)$. In order to investigate its behavior more precisely, Bohr introduced

$$\mu(x)=\limsup_{|y|\to+\infty}\frac{\log|f(x+iy)|}{\log|y|}$$

in his thesis (1910) and called it the **order** over $\operatorname{Re} z=x$. The function $\mu(x)$ is nonnegative, monotone decreasing, convex, and continuous with respect to x. Bohr later found that there is a kind of periodicity for the values of $f(z)$ over $\operatorname{Re} z=x$; this was the origin of the theory of †almost periodic functions.

As for the zeros of the function $f(z)$, the following theorems are known: If $f(z)$ is not identically zero, it has only a finite number of zeros in $x\geqslant S+\varepsilon$, $e^{-Mx}\leqslant y\leqslant e^{Mx}$ for arbitrary positive numbers ε, M (Perron, 1908). If we denote by $N(T)$ the number of zeros in $x>S+\varepsilon$, $T<y<T+2\delta\log T$, then $\limsup_{T\to\infty}N(T)/(\log T)^2\leqslant\delta/\varepsilon$ (E. Landau, 1927).

There have been many investigations into the connection between the singularities of $f(z)$ and the coefficients a_n. If the a_n are real and positive, the point $z=S$ is always a singularity of $f(z)$. Moreover, if $S=0$, $\operatorname{Re} a_n\geqslant 0$, and $\lim_{n\to\infty}(\cos(\arg a_n))^{1/\lambda_n}=1$, then $z=0$ is a singularity of $f(z)$ (C. Biggeri, 1939). Furthermore, if $\lambda_n/n\to\infty$, $\liminf_{n\to\infty}(\lambda_{n+1}-\lambda_n)>0$, then the line $\operatorname{Re} z=S$ is the †natural boundary of $f(z)$ (F. Carleson and Landau, 1921; A. Ostrowski, 1923). If $S=0$, $\liminf_{n\to\infty}(\lambda_{n+1}-\lambda_n)=q>0$, then there always exist singularities on every interval on the imaginary axis with the length $2\pi/q$ (G. Pólya, 1923). S. Mandelbrojt (1954, 1963) gave some interesting results concerning the relations between the singularities of (1) and the Fourier transform of an entire function.

If $U=-\infty$, the function $f(z)$ is an entire function. Its †order (in the sense of entire function) ρ is given by

$$\rho=\limsup_{x\to-\infty}\frac{\log^+\log^+M(x)}{|x|},$$

$$M(x)=\sup_{-\infty<y<\infty}|f(x+iy)|.$$

There have been many investigations into the †Julia direction of $f(z)$ and related topics by Mandelbrojt, Valiron, and Tanaka.

D. Tauberian Theorems

As in the case of power series, if the series $\sum a_n$ converges to s, then $f(+0)=s$ (**Abel's continuity theorem**). The converse is not necessarily true. The converse theorems, with additional conditions on a_n and λ_n, are called **Tauberian theorems**, as in the case of power series. Many theorems are known about this field. The most famous additional conditions are $\lim_{n\to\infty}\lambda_n a_n/(\lambda_n-\lambda_{n-1})=0$ (Landau, 1926) and $a_n=O((\lambda_n-\lambda_{n-1})/\lambda_n)$ (K. Ananda-Rau, 1928).

For the summation of Dirichlet series (especially †Riesz's method of summation) → 379 Series R. For Tauberian theorems (especially the Wiener-Ikehara-Landau theorem) of the ordinary Dirichlet series → 123 Distribution of Prime Numbers B.

E. Series Related to Dirichlet Series

A series of the form

$$\sum_{n=1}^{\infty}\frac{n!a_n}{z(z+1)(z+2)\ldots(z+n)},\quad z\neq 0,\ -1,\ -2,\ldots$$

is called a **factorial series** with the coefficients $\{a_n\}$. It converges or diverges simultaneously with the ordinary Dirichlet series $\sum a_n/n^z$ except at $z=0$ and negative integers. The series

$$\sum_{n=1}^{\infty}\frac{(z-1)(z-2)\ldots(z-n)}{n!}a_n=\sum_{n=1}^{\infty}a_n\binom{z-1}{n}$$

is called a **binomial coefficient series.** It converges or diverges simultaneously with the ordinary Dirichlet series $\sum(-1)^n a_n/n^z$ except at $z=0$ and positive integers.

References

[1] V. Bernshteĭn, Leçons sur les progrès récents de la théorie des séries de Dirichlet, Gauthier-Villars, 1933.
[2] K. Chandrasekharan and S. Minakshisundaram, Typical means, Oxford Univ. Press, 1952.
[3] G. H. Hardy and M. Riesz, The general theory of Dirichlet's series, Cambridge, 1915.
[4] S. Mandelbrojt, Séries lacunaires, Actualités Sci. Ind., Hermann, 1936.
[5] S. Mandelbrojt, Séries adhérentes, régularisation des suites, applications, Gauthier-Villars, 1952.
[6] S. Mandelbrojt, Dirichlet series, Reidel, 1972.
[7] G. Valiron, Théorie générale des séries de Dirichlet, Mém. Sci. Math., Gauthier-Villars, 1926.

122 (IV.14) Discontinuous Groups

A. Definitions [1–4]

Suppose that a group Γ acts continuously on a †Hausdorff space X, that is, for every $\gamma\in\Gamma$ and $x\in X$, an element γx of X is assigned in such a way that the mapping $x\rightarrow\gamma x$ is a homeomorphism of X onto itself and that we have $\gamma_1(\gamma_2 x)=(\gamma_1\gamma_2)x$, $1x=x$, where 1 is the identity element of Γ. Two points x, $y\in X$ are said to be **Γ-equivalent** if there exists a $\gamma\in\Gamma$ such that $y=\gamma x$. (Γ-equivalence for subsets of X is defined similarly.)

We consider the following conditions of discontinuity of Γ. (i) For every $x\in X$ and any infinite sequence $\{\gamma_i\}$ consisting of distinct elements of Γ, the sequence $\{\gamma_i x\}$ has no †cluster point in X. (ii) For every $x\in X$, there exists a neighborhood U_x such that $\gamma U_x\cap U_x=\varnothing$ for all but finitely many $\gamma\in\Gamma$. (ii′) If x, $y\in X$ are not Γ-equivalent, there exist neighborhoods U_x, U_y of x, y, respectively, such that $\gamma U_x\cap U_y=\varnothing$ for all $\gamma\in\Gamma$. (iii) For any compact subset M of X, $\gamma M\cap M=\varnothing$ for all but finitely many $\gamma\in\Gamma$.

It is easy to see that (ii)$\Rightarrow$(i), (ii)+(ii′)$\Rightarrow$(iii); if, moreover, X is †locally compact, we also have (iii)$\Rightarrow$(ii), (ii′). When (i) holds, Γ is called a **discontinuous transformation group** of X, and when (ii) holds, Γ is called a **properly discontinuous transformation group.** In particular, when X can be identified with a †homogeneous space G/K of a locally compact group G by a compact subgroup K, the conditions (i), (ii), and (iii) for a subgroup Γ of G are all equivalent, and they are also equivalent to the condition that Γ is a †discrete subgroup of G.

For a discontinuous group Γ acting on X, the †stabilizer $\Gamma_x=\{\gamma\in\Gamma\mid\gamma x\times x\}$ of $x\in X$ is always a finite subgroup. When $\Gamma_x=\{1\}$ for all $x\in X$, Γ is said to be **free** (or to **act freely** on X). If $\Gamma_X=\bigcap_{x\in X}\Gamma_x=\{1\}$, Γ is said to act †effectively on X. A point $x\in X$ is called a **fixed point** of Γ if $\Gamma_x\neq\Gamma_X$. In the following, we assume for simplicity that Γ acts effectively on X, unless otherwise specified.

Since Γ-equivalence is clearly an †equivalence relation, we can decompose X into Γ-equivalence classes, or Γ-†orbits. The space of all Γ-orbits, called the **quotient space** of X by Γ, is denoted by $\Gamma\backslash X$. When Γ satisfies the conditions (ii) and (ii′), the space $\Gamma\backslash X$ becomes a †Hausdorff space with respect to the topology of the quotient space. If, moreover, Γ is free, X is an (unramified) †covering space of $\Gamma\backslash X$ with the †covering transformation group Γ. (Conversely, a covering transformation group is always a free, properly discontinuous transformation group.) In general, X may be viewed as a covering space of $\Gamma\backslash X$ with ramifications, and the ramifying points (in X) are nothing but the fixed points of Γ.

B. Fundamental Regions

A complete set of representatives F of $\Gamma\backslash X$ in X (that is, a subset F of X such that $\Gamma F=X$, $\gamma F\cap F=\varnothing$ for $\gamma\in\Gamma$, $\gamma\neq 1$) is called a **fundamental region** of Γ in X if it further satisfies suitable topological or geometrical requirements. Here we assume that $\bar{F}$, the closure of F, is the closure of its interior F^i. (In this case, $\bar{F}$ or F^i is sometimes called the fundamental region of Γ instead of F itself.) Such a fundamental region exists if Γ satisfies the

conditions (ii), (ii′), and the set of fixed points is †nowhere dense in X (R. Baer, F. W. Levi, 1931). A fundamental region F is called **normal** if the set $\{\gamma F\}$ $(\gamma \in \Gamma)$ is locally finite, that is, if, for every $x \in X$, there exists a neighborhood U_x such that $\gamma F \cap U_x = \varnothing$ for all but a finite number of $\gamma \in \Gamma$. If X is †connected and F is normal, then Γ is generated by the set of $\gamma \in \Gamma$ such that $\gamma \bar{F} \cap \bar{F} \neq \varnothing$. Thus it is useful to have a fundamental region in order to find a set of generators of Γ and a set of †fundamental relations for them. When X has a Γ-invariant †Borel measure μ and Γ is countable, then the measure $\mu(F)$ of F is independent of the choice of F. Hence it is legitimate to put $\mu(\Gamma \backslash X) = \mu(F)$; Γ is called a **discontinuous group of the first kind** (C. L. Siegel [2]) if Γ is a discontinuous transformation group which has a normal fundamental region F such that $\{\gamma \mid \gamma \bar{F} \cap \bar{F} = \varnothing\}$ is finite and $\mu(F) < \infty$. For instance, if X is locally compact and $\bar{F}$ is compact ($\Leftrightarrow \Gamma \backslash X$: compact), then Γ is of the first kind.

When we are concerned only with the qualitative properties of Γ, it is sometimes convenient to relax the conditions for a fundamental region, replacing it by a **fundamental (open) set** Ω of Γ, that is, an (open) subset Ω of X such that $\Gamma\Omega = X$ and $\gamma\Omega \cap \Omega = \varnothing$ for all but a finite number of $\gamma \in \Gamma$ [5–9].

C. The Case of a Riemann Surface

Let Γ be a discontinuous group of analytic automorphisms of a †Riemann surface X. In virtue of †uniformization theory, it is enough, in principle, to consider the case where X is †simply connected. Thus we have the following three cases:

(1) $X = \mathbf{C} \cup \{\infty\}$ (†Riemann sphere). Γ is a finite group. Since Γ can also be considered as a group of motions of the sphere, it is either a cyclic, †dihedral, or †regular polyhedral group [10].

(2) $X = \mathbf{C}$ (complex plane). Γ is contained in the group of motions of the plane. The subgroup consisting of all parallel translations contained in Γ is a †free Abelian group of rank $v \leqslant 2$. If $v = 0$, then Γ is a finite cyclic group. When $v > 0$, Γ consists of the transformations of the following form:

When $v = 1$, $\quad z \rightarrow \varepsilon^k z + m\omega \qquad (k, m \in \mathbf{Z})$,

When $v = 2$, $\quad z \rightarrow \varepsilon^k z + m_1\omega_1 + m_2\omega_2 \qquad (k, m_i \in \mathbf{Z})$,

where ω, ω_1, ω_2 are nonzero complex numbers with $\operatorname{Im}(\omega_2/\omega_1) > 0$, and $\varepsilon = \pm 1$ in general, except in special cases when $v = 2$ and $\omega_2/\omega_1 = \zeta_4$ (resp. ζ_3 or ζ_6, where $\zeta_l = \exp(2\pi i/l)$), in which cases we may put $\varepsilon = \zeta_4$ (resp. ζ_3 or ζ_6). For the fundamental regions corresponding to these values of ε, see Fig. 1. In the cases $v = 1$ and 2, the †automorphic functions with respect to Γ are essentially given by exponential functions and elliptic functions, respectively (→ 134 Elliptic Functions).

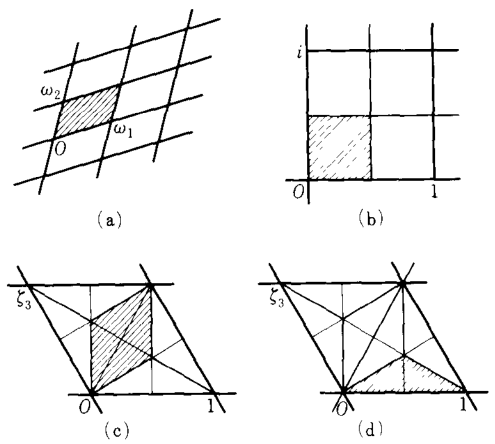


Fig. 1
(a) $v = 2$, $\varepsilon = 1$. (b) $\omega_1 = 1$, $\omega_2 = i$, $\varepsilon = i$. (c) $\omega_1 = 1$, $\omega_2 = \zeta_3$, $\varepsilon = \zeta_3$. (d) $\omega_1 = 1$, $\omega_2 = \zeta_3$, $\varepsilon = \zeta_6$.

(3) $X = \{|z| < 1\}$ (unit disk) [3, 10, 11]. By a †Cayley transformation, the unit disk can be transformed to the upper half-plane $\mathfrak{H} = \{z = x + iy \mid y > 0\}$. Any analytic automorphism of $\mathfrak{H}$ is given by a real †linear fractional transformation (Möbius transformation) $z \rightarrow (az + b)(cz + d)^{-1}$ $(a, b, c, d \in \mathbf{R}, ad - bc = 1)$. The totality of real linear fractional transformations acts transitively on $\mathfrak{H}$. Hence $\mathfrak{H}$ can be identified with the †homogeneous space G/K of $G = SL(2, \mathbf{R})$ by $K = SO(2)$ (which is the stabilizer of the point $\sqrt{-1}$). Hence discontinuous groups Γ of analytic automorphisms of $\mathfrak{H}$ are obtained as discrete subgroups of G. Actually, every element of G defines an analytic automorphism of the whole Riemann sphere, which leaves the real axis $\mathbf{R} \cup \{\infty\}$ invariant. For any $z \in \mathbf{C} \cup \{\infty\}$ and a sequence $\{\gamma_i\}$ consisting of distinct elements of Γ, a cluster point of the sequence $\{\gamma_i z\}$ in $\mathbf{C} \cup \{\infty\}$ is called a **limit point** of Γ. When only one or two limit points exist, Γ can easily be transformed to one of the groups given in (2). Otherwise, the set $\mathfrak{L}$ of all limit points of Γ is infinite, and either $\mathfrak{L} = \mathbf{R} \cup \{\infty\}$ or $\mathfrak{L}$ is a †perfect, †nowhere dense subset of $\mathbf{R} \cup \{\infty\}$. When $\mathfrak{L}$ is infinite, Γ is called a **Fuchsoid group**.

Since $\mathfrak{H}$ has a G-invariant †Riemannian metric $ds^2 = y^{-2}(dx^2 + dy^2)$ (called the †Poincaré metric), by which $\mathfrak{H}$ becomes a **hyperbolic plane** (†non-Euclidean plane with negative curvature), we can construct a fundamental region F of Γ which is a normal polygon bounded by geodesics, that is, the arcs of cir-

cles orthogonal to the real axis. A set of generators of Γ and the fundamental relations for them are easily obtained by observing the correspondence of the equivalent sides of the fundamental polygon. (Conversely, starting from a normal polygon satisfying a suitable condition, one can construct a discontinuous group Γ having F as a fundamental region. In this manner, we generally obtain a (nontrivial) continuous family of discrete subgroups of G.)

A Fuchsoid group Γ is finitely generated if and only if the fundamental polygon F has a finite number of sides, and in that case Γ is called a **Fuchsian group**. (More generally, a finitely generated discontinuous group of linear fractional transformations acting on a domain in the complex plane is called a **Kleinian group**.) A Fuchsian group Γ is **of the first kind** if and only if $\mathfrak{L}=\mathbf{R}\cup\{\infty\}$; otherwise, it is **of the second kind**. It is also known that a discontinuous group Γ is a Fuchsian group of the first kind if and only if $\mu(\Gamma\backslash\mathfrak{H})<\infty$ [12]. For a real point $x\in\mathbf{R}\cup\{\infty\}$, we also denote by Γ_x the stabilizer of x (in Γ). The point x is called a **(parabolic) cusp** of Γ if Γ_x is a free cyclic group generated by a †parabolic transformation ($\neq\pm1$). Cusps of Γ are represented by vertices of the fundamental polygon on the real axis. On the other hand, if a fixed point z of Γ lies in $\mathfrak{H}$, then the stabilizer Γ_z is always a finite cyclic group generated by an †elliptic transformation. Hence such a point z is also called an **elliptic point** of Γ. For a Fuchsian group Γ of the first kind, let $\{z_1,\dots,z_s\}$ be a complete set of representatives of the Γ-equivalence classes of the elliptic points of Γ (which can also be chosen from among the vertices of the fundamental polygon), and let e_i be the order of Γ_{z_i}; furthermore, let t be the number of the Γ-equivalence classes of parabolic cusps of Γ. Then the quotient space $\Gamma\backslash\mathfrak{H}$ can be compactified by adjoining t points at infinity, and the resulting space becomes a compact Riemann surface $\mathfrak{R}_\Gamma$ if we define an analytic structure on it in a suitable manner. The area $\mu(\mathfrak{R}_\Gamma)$ measured by the Poincaré metric is given by the †Gauss-Bonnet formula:

$$\mu(\mathfrak{R}_\Gamma)=\int_F\frac{dx\,dy}{y^2}$$
$$=2\pi\left[2g-2+\sum_{i=1}^{s}\left(1-\frac{1}{e_i}\right)+t\right],$$

where g is the †genus of the Riemann surface $\mathfrak{R}_\Gamma$. It is known that there exists a lower bound ($=\pi/21$) for $\mu(\mathfrak{R}_\Gamma)$ [12]. Automorphic functions (or Fuchsian functions) with respect to a Fuchsian group Γ, which are essentially the same thing as algebraic functions on the Riemann surface $\mathfrak{R}_\Gamma$, have been objects of extensive study since Poincaré (1882).

D. Modular Groups [10, 13]

The group

$$\Gamma=SL(2,\mathbf{Z})$$
$$=\left\{\begin{pmatrix}a&b\\c&d\end{pmatrix}\middle|\,a,b,c,d\in\mathbf{Z},ad-bc=1\right\}$$

(or the corresponding group of linear fractional transformations) is called the **(elliptic) modular group**. The modular group Γ is a Fuchsian group of the first kind acting on $\mathfrak{H}$, and its fundamental region together with the correspondence of the equivalent sides is shown in Fig. 2. Fig. 3 illustrates the transformations under Γ of the fundamental triangle, where Γ is regarded as acting on the unit disk. From Fig. 2 we obtain the generators of Γ (mod$\{\pm I_2\}$) ($I_2=\begin{pmatrix}1&0\\0&1\end{pmatrix}$) and the fundamental relations:

$$\sigma_1=\begin{pmatrix}1&1\\0&1\end{pmatrix},\quad\sigma_2=\begin{pmatrix}0&-1\\1&0\end{pmatrix},$$
$$\sigma_2^2=(\sigma_2\sigma_1)^3=-I_2.$$

There are two Γ-equivalence classes of elliptic points of Γ, which are represented by $\zeta_4=i$ and ζ_3, with $[\Gamma_i:\{\pm I_2\}]=2$, $[\Gamma_{\zeta_3}:\{\pm I_2\}]=3$; and only one Γ-equivalence class of parabolic cusps, which coincides with $\mathbf{Q}\cup\{\infty\}$. The corresponding Riemann surface $\mathfrak{R}_\Gamma$ is analytically equivalent to the Riemann sphere.

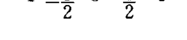


Fig. 2

Fig. 3

For a positive integer N, the totality $\Gamma(N)$ of elements in Γ satisfying the condition $\begin{pmatrix}a&b\\c&d\end{pmatrix}\equiv\begin{pmatrix}1&0\\0&1\end{pmatrix}$ (mod N) forms a normal subgroup of Γ, called a **principal congruence subgroup of level** N. (For the case $N=2$, see Fig. 4.) In general, a subgroup Γ' of Γ containing $\Gamma(N)$ for some N is called a **congruence subgroup** of Γ. (It is known that there actually exists a subgroup Γ' of Γ with a finite index, which is not a congruence subgroup.) For $N\geqslant3$, $-I_2\notin\Gamma(N)$, so that $\Gamma(N)$ is effective. (For $N=1,2$, we have $\Gamma(N)_{\mathfrak{H}}=\{\pm I_2\}$.) If $N\geqslant2$, $\Gamma(N)$ has no elliptic point. The number $t(N)$ of the equivalence classes of cusps of $\Gamma(N)$ and the genus $g(N)$ of the corresponding Rie-

mann surface $\mathfrak{R}_{\Gamma(N)}$ are given as follows:

$t(1)=1, \quad t(2)=3,$

$t(N)=(1/2N)[\Gamma:\Gamma(N)] \quad (N\geqslant 3),$

$g(1)=g(2)=0,$

$g(N)=1+((N-6)/24N)[\Gamma:\Gamma(N)] \quad (N\geqslant 3),$

where $[\Gamma:\Gamma(N)]=N^3\prod_{p|N}(1-1/p^2)$. Automorphic functions with respect to $\Gamma(N)$ are called †modular functions of level N.

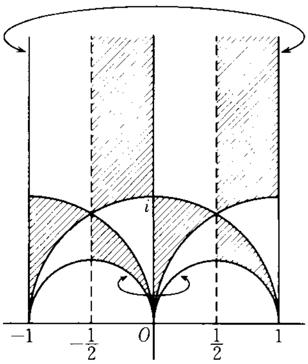


Fig. 4
A fundamental region of $\Gamma(2)$ that consists of six fundamental regions of $\Gamma(1)$.

E. The Case of Many Variables

Up to the present time, discontinuous groups Γ and the corresponding automorphic functions have been studied only in the following cases: (2′) $X=\mathbf{C}^n$, $\Gamma\cong\mathbf{Z}^{2n}$ (the free Abelian group of rank $2n$, consisting of parallel translations) [14] (→ 3 Abelian Varieties); (3′) X is a bounded domain in $\mathbf{C}^n$ and Γ is a discontinuous group of analytic automorphisms of X. (In this case, conditions (i), (ii), (iii) are equivalent.)

In the case (3′), the group $\mathscr{A}(X)$ of all (complex) analytic automorphisms of X, endowed with its natural (†compact-open) topology, becomes a †Lie group, which has Γ as a discrete subgroup. When $\Gamma\backslash X$ is compact, it is known by the theory of automorphic functions (or by a theorem of Kodaira) that $\Gamma\backslash X$ becomes a †projective variety, which is a †minimal model [3, 14]. In particular, when X is a †symmetric bounded domain, i.e., when X becomes a †symmetric Riemannian space with respect to its †Bergman metric, the connected component G of the identity element of $G'=\mathscr{A}(X)$ (which incidentally coincides with that of the group $I(X)$ of all †isometries of X) is a †semisimple Lie group of noncompact type (i.e., without compact simple factors), and X can be identified with the homogeneous space of G by a maximal compact subgroup K of G. The theory of discontinuous groups of this type, initiated by Siegel (especially in the case where $X=\mathfrak{H}_n=Sp(n,\mathbf{R})/K$, †Siegel's upper half-space; $\Gamma=Sp(n,\mathbf{Z})$, †Siegel's modular group of degree n), O. Blumenthal, H. Braun, and L.-K. Hua, and continued by those in the German school such as M. Koecher, H. Maass, and others, has undergone substantial development in recent years under the influence of the theory of algebraic groups [5, 8, 14–17] (→ 32 Automorphic Functions).

On the other hand, for a symmetric Riemannian space X of negative curvature, the group of isometries $G=I(X)$ is a †semisimple Lie group of noncompact type with a finite number of connected components and with a finite center, and X can be identified with the homogeneous space of G by a maximal compact subgroup. Therefore the study of discontinuous groups of isometries of X can be reduced to that of discrete subgroups of a Lie group G of this type. A typical example is the case where X is the space of all real †positive definite symmetric matrices of degree n with determinant 1; this space can be identified with the quotient space $SL(n,\mathbf{R})/SO(n)$ ($A\in SL(n,\mathbf{R})$ acts on X by $X\ni S\rightarrow {}^tASA$). The unimodular group $\Gamma=SL(n,\mathbf{Z})$ is a discontinuous group of the first kind acting on this space X, and a method of constructing a fundamental region of Γ in X is provided by the **Minkowski reduction theory** [6, 7].

F. Discrete Subgroups of a Semisimple Lie Group

Two subgroups Γ, Γ' of a group G are called **commensurable** if $\Gamma\cap\Gamma'$ is of finite index in both Γ and Γ'. For a real †linear algebraic group $G\subset GL(n,\mathbf{R})$ defined over $\mathbf{Q}$, a subgroup Γ commensurable with $G_{\mathbf{Z}}=G\cap GL(n,\mathbf{Z})$ is called an **arithmetic subgroup** (in the original sense) of G (examples: $SL(n,\mathbf{Z}), Sp(n,\mathbf{Z})$). An arithmetic subgroup Γ is always discrete, and when G is semisimple, the quotient space $\Gamma\backslash G$ is of finite volume ($\mu(\Gamma\backslash G)<\infty$) with respect to an invariant measure μ. Moreover, $\Gamma\backslash G$ is compact if and only if G is †$\mathbf{Q}$-compact (or †$\mathbf{Q}$-anisotropic), that is, if $G_{\mathbf{Q}}$ or $G_{\mathbf{Z}}$ consists of only †semisimple elements (A. Borel, Harish-Chandra, G. D. Mostow, and T. Tamagawa [6, 7]); the same results remain true if G is †Zariski connected and has no †character defined over $\mathbf{Q}$. The proofs of these facts (and the compactification of the quotient space $\Gamma\backslash X$ for the noncompact case) depend on a construction of fundamental open sets that generalizes the reduction theory of Minkowski and Siegel [5, 8, 15–18].

For a connected semisimple Lie group G of noncompact type and a discrete subgroup Γ with $\mu(\Gamma\backslash G)<\infty$, the following **density**

theorem holds (Borel, *Ann Math.*, (2) 72 (1960)): (i) For any linear representation ρ of G, the linear closure of $\rho(\Gamma)$ coincides with that of $\rho(G)$; (ii) if G is algebraic, Γ is †Zariski dense in G. Furthermore, suppose that G is a direct product of simple groups G_i and the center of G is finite; Γ is called **irreducible** if its projection on any (proper) partial product of $\{G_i\}$ is not dicrete. For instance, if G is a †**Q**-simple algebraic group then $\Gamma = G_{\mathbf{Z}}$ is irreducible. In general, there exists a partition of the set of indices $\{i\}$ such that Γ is commensurable with a direct product of irreducible discrete subgroups of the partial products corresponding to this partition, and these irreducible components are unique up to commensurability.

The method of constructing a discrete subgroup Γ of $G = SL(2, \mathbf{R})$ in a geometric manner using the upper half-plane can be generalized to some extent to the construction of discrete subgroups of certain groups using hyperbolic spaces of low dimensions (E. B. Vinberg). Except for these few cases, today it is known that a discrete subgroup Γ of a semisimple Lie group G (of **R**-rank $\geqslant 2$) with $\mu(\Gamma \backslash G) < \infty$ is arithmetic in a certain sense (→ Section G). This implies there are only very few discrete subgroups for a semisimple Lie group of higher rank. Actually a number of facts suggesting this result were already known in the 1960s. First, the only subgroups of $SL(n, \mathbf{Z})$ $(n \geqslant 3)$, $Sp(n, \mathbf{Z})$ $(n \geqslant 2)$ with finite index are congruence subgroups (H. Bass, M. Lazard, and J.-P. Serre; this result has been generalized to the case of an arbitrary †Chevalley group over an algebraic number field by C. C. Moore and H. Matsumoto). Second, $G_{\mathbf{Z}} = SL(n, \mathbf{Z})$, $Sp(n, \mathbf{Z})$ are maximal in G [8]. Finally, it is known (rigidity theorem) that if a connected semisimple Lie group G with a finite center does not contain a simple factor which is locally isomorphic to $SL(2, \mathbf{R})$, then any discrete subgroup Γ of G with compact quotient $\Gamma \backslash G$ has no nontrivial †deformation (i.e., all deformations are obtained from inner automorphisms of G) (A. Selberg, E. Calabi and E. Vesentini, and A. Weil [19]). This last result amounts to the vanishing of the cohomology group $H^1(\Gamma, X, \mathrm{ad})$, and in this connection an extensive study has been made by Y. Matsushima, S. Murakami, G. Shimura, and K. G. Raguanathan to determine the †Betti numbers of $\Gamma \backslash X$, and more generally the cohomology groups of the type $H(\Gamma, X, \rho)$ with an arbitrary representation ρ of G. These cohomology groups are closely related to automorphic forms with respect to Γ [8, 20].) For a †nilpotent or †solvable Lie group, a general method of constructing discrete subgroups is known; see, for example, M. Saito, *Amer. J. Math.*, 83 (1961).)

G. Rigidity and Arithmeticity

A discrete subgroup Γ of a Lie group G with $\mu(\Gamma \backslash G) < \infty$ is usually called a **lattice** of G. A lattice Γ of G is said to be **uniform** if $\Gamma \backslash G$ is compact. By a theorem of Borel and Harish-Chandra [7], an arithmetic subgroup of a real linear group (defined over **Q**) is a lattice. Using this result, Borel showed further, by a constructive method, that any semisimple Lie group has a lattice, especially a uniform one (Borel, *Topology*, 2 (1963)).

For a long time there were no known examples of nonarithmetic irreducible lattices in semisimple Lie groups other than those locally isomorphic to $SL_2(\mathbf{R})$. This naturally led to Selberg's conjecture that any irreducible (nonuniform) lattice in a semisimple Lie group G not locally isomorphic to $SL_2(\mathbf{R})$ is arithmetic (Selberg, International Colloquium on Function Theory, Bombay, 1960). The conjecture seemed to be well-grounded by the rigidity theorem of Weil and Selberg [19].

However, in 1966–1967, V. S. Makarov and Vinberg constructed nonarithmetic nonuniform lattices in $SO(n,1)$ $(n = 3, 4, 5)$ by a geometric method; the lattices are generated by †reflections [21]. Thus rigidity and arithmeticity do not necessarily coincide, and the conjecture should be considered under stronger conditions.

As for rigidity of uniform lattices, Mostow established in 1970 the following **strong rigidity theorem** [22]: If G, G' are semisimple Lie groups with trivial center and without compact factors, and are not locally isomorphic to $SL_2(\mathbf{R})$, and if Γ, Γ' are irreducible uniform lattices, then any isomorphism $\theta: \Gamma \to \Gamma'$ extends to an analytic isomorphism $\tilde{\theta}: G \to G'$ (namely, $\tilde{\theta} | \Gamma = \theta$). The previous rigidity theorem is now implied by Mostow's. If X is a simply connected symmetric space, then a †locally symmetric space M covered by X is expressed as a quotient of X by a fixed-point-free properly discontinuous group Γ in the group $I(X)$ of total isometries (when the Lie algebra of $I(X)$ does not have a compact factor, this condition for Γ is equivalent to saying that it is a torsion-free discrete subgroup of the semisimple Lie group $I(X)$), and the fundamental group $\pi_1(M)$ of $M = \Gamma \backslash X$ is isomorphic to Γ. The strong rigidity theorem implies that compact locally symmetric spaces M_1 and M_2 (of higher dimensions) are isomorphic as Riemannian spaces if and only if $\pi_1(M_1)$ and $\pi_1(M_2)$ are isomorphic as abstract groups.

On the other hand, in 1973, G. A. Margulis proved the arithmeticity of irreducible nonuniform lattices in semisimple real linear groups of **R**-rank greater than 1 (*Russian Math. Surveys*, 29 (1974) (original in Russian,

1974); *Functional Anal. Appl.* 9 (1975) (original in Russian, 1975). M. S. Raghunathan also proved independently the same fact under a slightly stronger condition. Their results, together with the rigidity of nonuniform lattices in (higher-dimensional) semisimple Lie groups of **R**-rank 1 established by H. Garland, Raghunathan, and G. Prasad (*Inventiones Math.*, 21 (1973)), imply that the strong rigidity theorem holds similarly for nonuniform lattices.

The results of Margulis and Raghunathan show that the Selberg conjecture in the original sense is affirmative for the case of **R**-rank greater than 1. But neither argument is applicable to uniform lattices, for they depend deeply on the fact (proved by D. A. Kazdan and Margulis) that a nonuniform lattice contains a nonidentity unipotent element. Previously, in a lecture at the international congress of mathematicians at Moscow, 1966, I. I. Pyatetskiĭ-Shapiro generalized the definition of arithmeticity and suggested that arithmeticity of lattices should be investigated without the distinction of whether they are uniform or nonuniform. His definition is equivalent to the following [9, 24]: For a connected semisimple algebraic group **G** defined over **R**, a subgroup $\Gamma \subset G = G_{\mathbf{R}}$ is an **arithmetic subgroup** (of **G**) if there is an algebraic group **H** defined over **Q** and a surjective homomorphism $\varphi: \mathbf{H} \rightarrow \mathrm{Ad}\,\mathbf{G}$ defined over **R** such that the Lie group $(\mathrm{Ker}\,\varphi)_{\mathbf{R}}$ is compact and that $\varphi(H_{\mathbf{Z}})$ and $\mathrm{Ad}\,\Gamma$ are commensurable. The uniform lattice in G that is constructed by the method of Borel is arithmetic in this sense. In 1974, Margulis finally established the following arithmeticity theorem [23, 24]: If the **R**-rank of **G** is not less than 2, an irreducible lattice Γ in G is an arithmetic subgroup of **G** (even if it is uniform). In the same lecture, Pyatetskiĭ-Shapiro also extended the Selberg conjecture to such "semisimple Lie groups" as those containing p-adic Lie groups as factors. Margulis proved this Pyatetskiĭ-Shapiro conjecture affirmatively by showing that an analog of the strong rigidity theorem holds for such groups.

As for semisimple Lie groups of **R**-rank 1, besides the lattices constructed by Makarov and Vinberg there are only the few examples of nonarithmetic lattices in $SU(2,1)$ presented by Mostow (*Proc. Nat. Acad. Sci. US*, 75 (1978); *Pacific J. Math.*, 86 (1980)). The problem of arithmeticity still remains open for groups locally isomorphic to $SO(n,1)$ $(n \geqslant 6)$, $SU(n,1)$ $(n \geqslant 3)$, $Sp(n,1)$, or F_4.

H. Geometric Discontinuous Groups [1, 25]

The study of discontinuous groups Γ acting on a Euclidean or projective space X as a transformation group of a given structure is a classical problem. All possibilities for such Γ have been enumerated in low-dimensional cases. For instance, there are 230 kinds of discontinuous groups of †Euclidean motions acting on 3-dimensional Euclidean space without fixed subspaces, which are classified into 32 †crystal classes (A. Schönflies and E. S. Fedorov, 1891–1892; → 92 Crystallographic Groups). All discontinuous groups of a Euclidean space generated by †reflections have also been enumerated (H. S. M. Coxeter, 1934 [25, 26]).

I. Kleinian Groups

The last decade has seen considerable research on (finitely generated) Kleinian groups. This research is closely related to the theory of †quasiconformal mappings and the †moduli of Riemann surfaces.

Making use of Eichler cohomology and †potentials, L. V. Ahlfors established his finiteness theorem and L. Bers his area theorem. Bers and B. Maskit investigated the boundaries of †Teichmüller spaces and discovered Kleinian groups with the property that the complement of the set $\mathfrak{L}$ of limit points is connected and simply connected.

Numerous mathematicians have subsequently discussed the classification, deformation, and stability properties of the set $\mathfrak{L}$, uniformization and deformation of Riemann surfaces with or without nodes, and other geometric properties. In their discussions, the theory of quasiconformal mappings has played an important role. The discontinuous groups of motions of hyperbolic 3-space have also been studied.

References

[1] B. L. van der Waerden, Gruppen von linearen Transformationen, Erg. Math., Springer, 1935 (Chelsea, 1948).
[2] C. L. Siegel, Discontinuous groups, Ann. Math., (2) 44 (1943), 674–689 (Gesammelte Abhandlungen, Springer 1966, vol. 2, 390–405).
[3] Sém. H. Cartan 6, Fonctions automorphes et espaces analytiques, Ecole Norm. Sup., 1953–1954.
[4] G. Shimura, Introduction to the arithmetic theory of automorphic functions, Publ. Math. Soc. Japan, 11 (1971).
[5] Sém. H. Cartan 10, Fonctions automorphes, Ecole Norm. Sup., 1957–1958.
[6] A. Weil, Discontinuous subgroups of classical groups, Lecture notes, Univ. of Chicago, 1958.
[7] A. Borel and Harish-Chandra, Arithmetic subgroups of algebraic groups, Ann. Math., (2) 75 (1962), 485–535.

[8] Algebraic groups and discontinuous subgroups, Amer. Math. Soc. Proc. Symp. Pure Math., 1966, vol. 9.
[9] M. S. Raghunathan, Discrete subgroups of Lie groups, Erg. Math., Springer, 1972.
[10] R. Fricke and F. Klein, Vorlesungen über die Theorie der automorphen Funktionen I, Teubner, 1897, second edition, 1926.
[11] H. Poincaré, Théorie des groupes fuchsiens, Acta Math., 1 (1882), 1–62 (Oeuvres, Gauthier-Villars, 1916, vol. 2, 108–168).
[12] C. L. Siegel, Some remarks on discontinuous groups, Ann. Math., (2) 46 (1945), 708–718 (Gesammelte Abhandlungen, Springer, 1966, vol. 3, 67–77).
[13] F. Klein and R. Fricke, Vorlesungen über die Theorie der elliptischen Modulfunktionen I, Teubner, 1890.
[14] C. L. Siegel, Analytic functions of several complex variables, Lecture notes, Institute for Advanced Study, Princeton, 1948–1949.
[15] C. L. Siegel, Symplectic geometry, Amer. J. Math., 65 (1943), 1–86 (Gesammelte Abhandlungen, Springer, 1966, vol. 2, 274–359; Academic Press, 1964).
[16] I. Satake, On compactifications of the quotient spaces for arithmetically defined discontinuous groups, Ann. Math., (2) 72 (1960), 555–580.
[17] W. L. Baily and A. Borel, Compactifications of arithmetic quotients of bounded symmetric domains, Ann. Math., (2) 84 (1966), 442–528.
[18] H. Minkowski, Diskontinuitätsbereich für arithmetische Äquivalenz, J. Reine Angew. Math., 129 (1905), 220–274 (Gesammelte Abhandlungen, Teubner, 1911, vol. 2, 53–100; Chelsea, 1967).
[19] A. Weil, On discrete subgroups of Lie groups, Ann. Math., (2) 72 (1960), 369–384; II, ibid., 75 (1962), 578–602.
[20] Y. Matsushima and S. Murakami, On vector bundle valued harmonic forms and automorphic forms on symmetric Riemannian manifolds, Ann. Math., (2) 78 (1963), 365–416.
[21] E. B. Vinberg, Discrete groups generated by reflections in Lobacevskii spaces, Math. USSR-Sb., 1 (1967), 429–444. (Original in Russian, 1967.)
[22] G. D. Mostow, Strong rigidity of locally symmetric spaces, Ann. Math. Studies 78, Princeton Univ. Press, 1973.
[23] G. A. Margulis, Discrete groups of motions of manifolds of nonpositive curvature, Amer. Math. Soc. Transl., (2) 109 (1977), 33–45. (Original in Russian, 1974.)
[24] G. A. Margulis, Arithmetic irreducible lattices in semisimple groups of rank greater than 1 (in Russian), Mir, 1977. (Appendix to the Russian translation of [9].)
[25] H. S. M. Coxeter and W. O. J. Moser, Generators and relations for discrete groups, Erg. Math., Springer, 1957.
[26] N. Bourbaki, Eléments de mathématique, Groupes et algèbres de Lie, ch. 4, 5, 6, Actualités Sci. Ind., 1377, Hermann, 1968.
[27] L. Bers and I. Kra, A crash course on Kleinian groups, Lecture notes in math. 400, Springer, 1974.

123 (V.7) Distribution of Prime Numbers

A. General Remarks

Given a real number x, we denote by $\pi(x)$ the number of primes not exceeding x. A. M. Legendre (1808) obtained empirically the formula $\pi(x) \fallingdotseq x/(\log x - B)$ for some constant B, and C. F. Gauss (1849) obtained the formula

$$\pi(x) \fallingdotseq \int_2^x \frac{du}{\log u},$$

assuming the average density of primes to be $1/\log x$. The **Bertrand conjecture**, which asserts the existence of at least one prime between x and $2x$, was proved by P. L. Chebyshev (1848), who introduced the functions

$$\theta(x) = \sum_{m=1}^{\infty} \sum_{p \leqslant x} \log p$$

and

$$\psi(x) = \sum_{p^m \leqslant x} \log p$$
$$= \theta(x) + \theta(\sqrt{x}) + \theta(\sqrt[3]{x}) + \ldots .$$

(In this section, p represents a prime number.) He thereby proved

$$Ax + O(\sqrt{x}) < \theta(x) < \psi(x) < (6/5)Ax + O(\sqrt{x}),$$

where $A = \log(2^{1/2} 3^{1/3} 5^{1/5} 30^{-1/30})$. G. F. B. Riemann (1858) considered the function $\zeta(s)$ (where $s = \sigma + it$ is a complex variable), expressed by the †Dirichlet series $\sum_{n=1}^{\infty} n^{-s}$, which is convergent for $\sigma > 1$. He found relations between the zeros of $\zeta(s)$ (→ 450 Zeta Functions) and $\pi(x)$. F. Mertens (1874) obtained the formulas

$$\sum_{p \leqslant x} \frac{\log p}{p} = \log x + O(1),$$

$$\sum_{p \leqslant x} \frac{1}{p} = \log\log x + B + O\left(\frac{1}{\log x}\right),$$

$$\prod_{p \leqslant x} \left(1 - \frac{1}{p}\right) = \frac{e^{-c}}{\log x}\left(1 + O\left(\frac{1}{\log x}\right)\right),$$

where c is the Euler constant and B is some constant.

B. Prime Number Theorem

The **prime number theorem**

$$\lim_{x\to\infty}\frac{\pi(x)\log x}{x}=1, \quad \text{or} \quad \pi(x)\sim\frac{x}{\log x},$$

was proved almost simultaneously (1896) by J. Hadamard and C. J. de La Vallée-Poussin. Without using the theory of †entire functions, E. Landau (1908) established the formula

$$\pi(x)=\mathrm{Li}\,x+O(xe^{-c\sqrt{\log x}}),$$

where

$$\mathrm{Li}\,x=\lim_{\delta\to+0}\left(\int_0^{1-\delta}+\int_{1+\delta}^{x}\right)\frac{du}{\log u}$$

is the †logarithmic integral. It can be shown by integration by parts that

$$\mathrm{Li}\,x=\frac{x}{\log x}+\frac{1!x}{\log^2 x}+\dots$$
$$+\frac{(k-1)!x}{\log^k x}+O\left(\frac{x}{\log^{k+1}x}\right).$$

For example, by taking $x=10^7$, we get $\pi(x)=664{,}579$, $\mathrm{Li}\,x\fallingdotseq 664{,}918$, and $x/\log x\fallingdotseq 620{,}417$.

If the Dirichlet series $f(s)=\sum_{n=1}^{\infty}a_n n^{-s}$ satisfies the condition $\sum_{n\leqslant x}a_n\sim cx$, then its abscissa of convergence is 1, and we have $\lim_{s\to 1+0}(s-1)f(s)=c$. The converse is known as the †Tauberian theorem. If $F(s)=\sum_{n=1}^{\infty}a_n n^{-s}$ $(a_n\geqslant 0)$ converges absolutely for $\sigma>1$ and $F(s)-c/(s-1)$ is analytic for $\sigma\geqslant 1$, then we obtain $\sum_{n\leqslant x}a_n\sim cx$ (**Wiener-Ikehara-Landau theorem**, 1932). Specifically, if we put $-\zeta'(s)/\zeta(s)=\sum_{n=1}^{\infty}\Lambda(n)n^{-s}$, then the conditions of the theorem are satisfied, and we obtain $\sum_{n\leqslant x}\Lambda(n)=\psi(x)\sim x$.

It is easily seen that the prime number theorem is equivalent to $\psi(x)\sim x$ or $\theta(x)\sim x$. The †number-theoretic function $\Lambda(n)$ (**Mangoldt's function**) introduced above satisfies $\sum_{d|n}\Lambda(d)=\log n$. It follows from the †Möbius inversion formula that $\Lambda(n)=\sum_{d|n}\mu(d)\log(n/d)$. Hence $\Lambda(n)=\log p$ $(n=p^m)$ and $=0$ otherwise. Thus we obtain $\psi(x)=\sum_{n\leqslant x}\Lambda(n)=O(x)$. When $f(x)=\theta(x)$ or $\psi(x)$, it is easy to show that $\int_2^x(f(t)/t^2)dt=\log x+O(1)$ and $\liminf_{x\to\infty}f(x)/x\leqslant 1\leqslant\limsup_{x\to\infty}f(x)/x$. However, it is not easy to prove $f(x)\sim x$. To do so, introduce the number-theoretic function $M(n)$, which satisfies $\sum_{d|n}M(d)=\log^2 n$. As before, we have $M(n)=\sum_{d|n}\mu(d)\log^2(n/d)$; hence

$$M(n)=\begin{cases}(2l-1)\log p & (n=p^l,\ l\geqslant 1),\\ 2\log p\log q & (n=p^l q^m,\ l\geqslant 1,\ m\geqslant 1),\\ 0 & (\text{otherwise}).\end{cases}$$

Thus we obtain $\sum_{n\leqslant x}M(n)=2x\log x+O(x)$. This leads to A. Selberg's well-known formula (1949):

$$\theta(x)\log x+\sum_{p\leqslant x}\theta(x/p)\log p=2x\log x+O(x),$$

which enabled him to prove $\theta(x)\sim x$. Thus he obtained for the first time a proof of the prime number theorem that does not use complex analytic methods. The simple formulas $\sum_{n=1}^{\infty}\mu(n)/n=0$ and $\sum_{n\leqslant x}\mu(n)=o(x)$, obtained by H. von Mangoldt (1897), were revealed by Landau to have a deep meaning concerning the prime number theorem. Let $\pi_r(x)$ denote the number of integers not exceeding x that can be expressed as the product of r distinct primes. In generalizing the prime number theorem, Landau (1911) proved that

$$\pi_r(x)\sim\frac{1}{(r-1)!}\frac{x(\log\log x)^{r-1}}{\log x}.$$

Let us write $\vartheta(x)=\sum_{n=1}^{\infty}e^{-\pi x n^2}$. Riemann proved that

$$\Gamma\left(\frac{s}{2}\right)\pi^{-s/2}\zeta(s)$$
$$=\frac{1}{s(s-1)}+\int_1^{\infty}\vartheta(x)(x^{s/2-1}+x^{-1/2-s/2})dx$$

and obtained the well-known functional equation for the zeta function (→ 450 Zeta Functions B)

$$\pi^{-s/2}\Gamma(s/2)\zeta(s)=\pi^{-1/2+s/2}\Gamma(1/2-s/2)\zeta(1-s).$$

This enables us to extend $\zeta(s)$ as a meromorphic function to the whole complex plane. Utilizing this extended $\zeta(s)$ and the following result of O. Perron on Dirichlet series, we can estimate $\psi(x)$. Let σ_0 $(\neq\infty)$ be the abscissa of convergence of $F(s)=\sum_{n=1}^{\infty}f(n)n^{-s}$, and let $a>0$, $a>\sigma_0$, and $x>0$. If

$$\lim_{T\to\infty}\frac{1}{2\pi i}\int_{a-iT}^{a+iT}F(s)\frac{x^s}{s}ds$$

exists, then the limit is equal to $\sum'_{n\leqslant x}f(n)$, where $\sum'$ means that in the summation the last term $f(x)$ is replaced by $f(x)/2$ if x is an integer. In many cases, $F(s)$ has a pole at $s=1$, and the principal part of the sum is obtained from the residue at $s=1$, whereas the residual part is given by a certain contour integral. To estimate $\psi(x)$, we use $-\zeta'(s)/\zeta(s)$ as $F(s)$; hence the problem arises of determining the zeros of $\zeta(s)$. Riemann conjectured that all the zeros of $\zeta(s)$ in the strip $0\leqslant\sigma\leqslant 1$ must be situated on the vertical line $\sigma=1/2$. If this so-called †Riemann hypothesis (→ 450 Zeta Functions) is true, then it follows that $\pi(x)=\mathrm{Li}\,x+O(\sqrt{x}\log x)$. The ultimate validity of Riemann's hypothesis remains in doubt. Concerning this, the most recent major result is the following formula, obtained by I. M. Vinogradov (1958): $\pi(x)=\mathrm{Li}\,x+$

$O(x\exp(-c\log^{3/5}x/\log\log^{1/5}x))$. Without using Riemann's hypothesis, J. E. Littlewood (1918) proved that

$$\limsup_{x\to\infty}\frac{\pi(x)-\mathrm{Li}\,x}{(\sqrt{x}/\log x)\log\log\log x}>0,$$

$$\liminf_{x\to\infty}\frac{\pi(x)-\mathrm{Li}\,x}{(\sqrt{x}/\log x)\log\log\log x}<0.$$

If we denote by $N(T)$ the number of zeros of $\zeta(s)$ in the domain $0<\sigma<1$, $0<t<T$, then we have

$$N(T)=\frac{1}{2\pi}T\log T-\frac{1+\log 2\pi}{2\pi}T+O(\log T).$$

Let $N_0(T)$ denote the number of zeros of $\zeta(s)$ on the interval $\sigma=1/2$, $0<t<T$. Selberg (1942) obtained the impressive result

$$N_0(T)>cT\log T.$$

E. C. Titchmarsh (1936) showed that there exist 1041 zeros of $\zeta(s)$ in the domain $0<\sigma<1$, $0<t<1468$ and that all lie on the line $\sigma=1/2$. Computers have provided further results that seem to justify the Riemann hypothesis. For example, it has been calculated that there are 75,000,000 zeros of $\zeta(s)$ in the domain $0<\sigma<1$, $0<t<32{,}585{,}736.4$ and that all are simple zeros and lie on the line $\sigma=1/2$ (R. P. Brent, *Math. Comp.*, 33 (1979)). N. Levinson proved in 1974 by another method that at least one-third of the zeros of the Riemann zeta function are on the line $\sigma=1/2$. The minimum of the modulus of the imaginary part of the zeros with $\sigma=1/2$ is $t=14.13\ldots$.

C. Twin Primes

Let p_n be the nth prime. We know from the prime number theorem that $p_n\sim n\log n$; more precisely, $p_n=n\log n+n\log\log n+O(n)$. A pair of primes differing only by 2 are called **twin primes**. It is still unknown whether there exist infinitely many twin primes. There exist infinitely many n satisfying $p_{n+1}-p_n<c\log p_n$ (P. Erdös, 1940). Suppose that $\zeta(1/2+it)=O(|t|^c)$. A. E. Ingham (1937) proved that

$$p_{n+1}-p_n<p_n^{\Theta},\quad \Theta=(1+4c)/(2+4c)+\varepsilon,$$

by using the following density theorem related to the zeros of $\zeta(s)$: If we denote by $N(\alpha,T)$ the number of zeros of $\zeta(s)$ in the domain $\alpha\leqslant\sigma\leqslant 1$, $0\leqslant t\leqslant T$ $(1/2\leqslant\alpha\leqslant 1)$, then there is a positive constant c such that $N(\alpha,T)=O(T^{2(1+2c)(1-\alpha)}\log^5 T)$. The **Lindelöf hypothesis** asserts that the constant c can be made arbitrarily small. If the Riemann hypothesis holds, then the Lindelöf hypothesis also holds. It is clear that we can substitute $1/6+\varepsilon$ $(\varepsilon>0)$ for c, ε being arbitrarily small. W. Haneke (1963) showed that c can be replaced by $6/37+\varepsilon$, and consequently that $\Theta\leqslant 61/98+\varepsilon$. D. R. Heath-Brown and H. Iwaniec (1979) proved that $\Theta\leqslant 11/20+\varepsilon$ by using the †sieve method and the zero density theorem of L-functions (Section E). R. A. Rankin (1935) proved that

$$p_{n+1}-p_n>c\log p_n\log\log p_n\log\log\log\log p_n \times(\log\log\log p_n)^{-2}$$

holds for infinitely many n. If we denote by $\pi_2(x)$ the number of primes $p\leqslant x$ such that $p+2$ is also a prime, then it has been conjectured by Hardy and Littlewood (*Acta Math.*, 44 (1922)) that

$$\pi_2(x)\sim C\int_2^x\frac{du}{\log^2 u}$$

as $x\to\infty$, where

$$C=2\prod_{p>2}\left\{1-\frac{1}{(p-1)^2}\right\}=1.32032\ldots.$$

The numerical evidence provided by the computation $\pi_2(10^9)=3424506$ (Brent, *Math. Comp.*, 28 (1974)) tends to indicate the truth of this conjecture. At present, $76\cdot 3^{139}\pm 1$ seems to be the largest known pair of twin prime numbers (H. C. Williams and C. R. Zarnke, *Math. Comp.*, 26 (1972)).

D. Prime Numbers in Arithmetic Progressions

Let k be a positive integer, $\chi(n)$ be a residue character modulo k (→ 295 Number-Theoretic Functions D), and $L(s,\chi)=\sum_{n=1}^{\infty}\chi(n)n^{-s}$ $(\sigma>1)$ be the †Dirichlet L-function. The function $L(s,\chi)$ of s defined by this series can be extended to an analytic function in the whole complex plane in the same way as the Riemann zeta function. In particular, when χ is the principal character, then $L(s,\chi)$, thus extended, is a meromorphic function whose only pole is situated at $s=1$ and is simple; otherwise the function $L(s,\chi)$ is holomorphic on C. Using this function $L(s,\chi)$ and in connection with his research concerning the †class numbers of quadratic forms, P. G. L. Dirichlet (1837) proved that there exist infinitely many primes in the arithmetical progression $l, l+k, l+2k, \ldots$, where l is the initial term and k a common difference relatively prime to l. This result is called the **Dirichlet theorem** (or **prime number theorem for arithmetic progressions**).

Suppose that $\mathfrak{a}$ runs over all integral ideals in a †quadratic number field K of discriminant d. Then the †Dedekind zeta function $\zeta_K(s)$ of K is defined by $\sum(N\mathfrak{a})^{-s}$ for $\sigma>1$. By virtue of the decomposition law of prime ideals (→ 347 Quadratic Fields C), we have $\zeta_K(s)=\zeta(s)L(s,\chi)$, where $\chi(n)=(d/n)$ is the †Kronecker symbol.

Utilizing $\zeta_K(s)$, we obtain formulas concerning the †class number $h(d)$ of the field K. If $d>0$, then $h(d)=(\sqrt{d}/2\log\varepsilon)L(1,\chi)$, where ε is the †fundamental unit of K. On the other hand, if $d<0$, then $h(d)=(w\sqrt{-d}/2\pi)L(1,\chi)$, where w denotes the number of the roots of unity contained in K. It follows that $L(1,\chi)\geqslant 2\log((1+\sqrt{5})/2)/\sqrt{|d|}$. Let χ be a character modulo k induced by a primitive character χ^0. Since we have $L(s,\chi)=L(s,\chi^0)\prod_{p|k}(1-\chi^0(p)p^{-s})$, it can be shown that $L(1,\chi)\neq 0$ for a real character χ. It is easy to prove that $L(1,\chi)\neq 0$ for a complex character χ. These statements then lead to the Dirichlet theorem. The proof was simplified by H. N. Shapiro (1951). Besides Landau's three proofs for $L(1,\chi)\neq 0$ for real character χ (1908), there are elegant proofs by T. Estermann (1952), Selberg (1949), and others. For a character $\chi \bmod k$, we always have $L(1,\chi)=O(\log k)$, while $L(1,\chi)^{-1}=O(\log k)$ with one possible exception, which may occur only if χ is a real character. Even in this case, we have $L(1,\chi)^{-1}=O(k^{\varepsilon})$ (where $\varepsilon>0$ is arbitrary, but O depends on ε). This result was obtained by K. L. Siegel (1934) from his study concerning class numbers of imaginary quadratic number fields. His proof was simplified by Estermann (1948) and S. D. Chowla (1950). The importance of the prime number theorem for arithmetic progressions was revealed when it was applied to the Goldbach problem (→ 4 Additive Number Theory C). Concerning this problem, the manner in which the remainder term depends on the modulus k became an object of investigation. The Page-Siegel-Walfisz theorem is convenient to use: Denote by $\pi(x;k,l)$ the number of primes not exceeding x and of the form $ky+l$, where $(k,l)=1$. If $x\geqslant\exp(k^{\varepsilon})$ (where $\varepsilon>0$ is arbitrary), then we have

$$\pi(x;k,l)=\frac{\operatorname{Li}x}{\varphi(k)}+O\left(\frac{xe^{-c(\varepsilon)\sqrt{\log x}}}{\varphi(k)}\right).$$

Further research on the distribution of zeros of $L(s,\chi)$ is necessary for the study of $\pi(x;k,l)$ when x takes smaller values. If χ is a nonprincipal real character, then $L(s,\chi)$ may have at most one real zero β_1 around 1; this is called **Siegel's zero**. Because of this fact, when x is small we are unable to obtain any formula to indicate the uniform distribution of primes. However, we have the following deep result, obtained by E. Fogels (1962). For a given positive ε, there exist $c_0(\varepsilon)$ and $c(\varepsilon)$ such that $\pi(x;k,l)>c(\varepsilon)x/\varphi(k)k^{\varepsilon}\log x$, provided that $x\geqslant k^{c_0(\varepsilon)}$. On the other hand, Titchmarsh (1930), using the †sieve method, obtained $\pi(x;k,l)=O(x/\varphi(k)\log x)$ for $x\geqslant k^{c_0}$. A theorem of this type is called the **Burn-Titchmarsh theorem** (→ Section E). Fogels's theorem is based on the following theorem by Yu. V. Linnik (1947) and K. Prachar (1957), which is an extension of Page's theorem (1935): We let δ be the function defined by $\delta=1-\beta_1$ if $L(s,\chi)$ has an exceptional real zero β_1, and $\delta=c_1/\log(k(|t|+2))$ otherwise (where c_1 is a suitably small number). If we denote by β the real part of any zero ($\neq\beta_1$) of $L(s,\chi)$, then the theorem states that

$$\beta\leqslant 1-\frac{c_1}{\log(k(|t|+2))}\log\left(\frac{c_1 e}{\delta\log(k(|t|+2))}\right),$$

provided that $\delta\log(k(|t|+2))\leqslant c_1$. Linnik called this result the **Deuring-Heilbronn phenomenon**. Using this theorem and the zero density theorem, Linnik (1944) proved that $p(k,l)\ll k^L$, where $p(k,l)$ is the least prime in the arithmetic progression $l, l+k, l+2k, \ldots$ and L is a constant. We call L **Linnik's constant**. M. Jutila (1977) and Chen Jing-Run (1979) proved that $L\leqslant 80$ and $L\leqslant 17$, respectively.

Let s be positive, b_j, z_j $(1\leqslant j\leqslant s)$ be complex, and l, m be real numbers satisfying $\max|z_j|\geqslant 1$, $l\geqslant s$, and $m\geqslant 0$. Under these conditions P. Turán (1953) obtained the following fundamental theorem, which is called the **power sum theorem**:

$$\max_{m\leqslant r\leqslant l+m}|b_1z_1^r+\ldots+b_sz_s^r|$$

$$\geqslant\left(\frac{l}{8e(l+m)}\right)^l\min_{1\leqslant j\leqslant s}|b_1+\ldots+b_j|.$$

This theorem is effective in research on the distribution of zeros of zeta functions. Based on this new method, Turán (1961), S. Knapowski (1962), and Fogels (1965) reached the results cited above.

Another method of research, considered to be a new sieve method, on the distribution of primes was introduced by Selberg and A. I. Vinogradov. This method was followed by W. B. Jurkat and H. E. Richert (1965). Linnik, A. Rényi (1950), and E. Bombieri (1965) founded still another method, called the **large sieve method**, by which P. T. Batemann, Chowla, and P. Erdös studied the value of $L(1,\chi)$.

E. Sieve Method

Let A be a set of integers, and P a set of primes. For each $p\in P$, let Ω_p be a set of residues mod p, and $\omega(p)$ the number of residues belonging to Ω_p. The sieve method is a device for estimating (from above or below) the number of integers n belonging to the set $S(A,P,\Omega_p)=\{n\mid n\in A, n \bmod p\notin\Omega_p \text{ for } p\in P\}$. The combinatorial methods of the Brun, Bustab, and Richert sieves are interesting and efficient but quite complicated. Here, we shall briefly describe the **Selberg sieve**. As an example, denote by $S(x;q,l)$ the number of n satisfying $n\equiv l \bmod q$, $n\leqslant x$, $(n,D)=1$, where q

is a prime not exceeding $x, z \leqslant x$, $D=\prod_{p\leqslant z} p$, and $(l,q)=1$; then

$$S(x;q,l)=\sum_{\substack{n\leqslant x\\ n\equiv l \bmod q}}\sum_{d|(n,D)}\mu(d)$$

$$\leqslant \sum_{\substack{n\leqslant x\\ n\equiv l \bmod q}}\left(\sum_{d|(n,D)}\lambda_d\right)^2,$$

where $\lambda_1=1$ and λ_d is arbitrary for $d>1$. Thus the problem reduces to the optimization of λ_d. Proceeding in this manner, C. Hooley (1967) and Y. Motohashi (1975), using analytic methods, obtained certain deep results relating to the Brun-Titchmarsh theorem. Using the large sieve, H. L. Montgomery and R. C. Vaughan (1973) expressed this theorem in a precise form: $\pi(x,q,l)<2x/\varphi(q)\log(x/q)$ for all $q<x$.

Let $n_1, n_2, \ldots, n_Z$ be Z natural numbers not exceeding N, and $Z(p,a)$ the number of n_j's such that $n_j\equiv a \bmod p$. Rényi (1950) proved that

$$\sum_{p\leqslant\sqrt[3]{N/12}} p\sum_{a=1}^{p-1}\left\{Z(p,a)-\frac{Z}{p}\right\}^2\leqslant 2NZ.$$

Set $a_n=1$ for $n=n_j$ and $a_n=0$ otherwise, and set $S(\alpha)=\sum_{n\leqslant N}a_n\exp(2\pi i n\alpha)$; then

$$p\sum_{a=1}^{p-1}\left\{Z(p,a)-\frac{Z}{p}\right\}^2=\sum_{a=1}^{p-1}\left|S\left(\frac{a}{p}\right)\right|^2.$$

In view of this simple fact, Bombieri (1965) and P. X. Gallagher (1968) extended the problem and proved that, in general,

$$\sum_{q\leqslant Q}\sum_{\substack{a=1\\(a,q)=1}}^{q}\left|\sum_{M<n\leqslant M+N}a_n\exp\left(2\pi i\frac{a}{q}n\right)\right|^2$$

$$\leqslant (N+2Q^2)\sum_{M<n\leqslant M+N}|a_n|^2.$$

Similar results can be obtained for character sums $\sum_{M<n\leqslant M+N}a_n\chi(n)$. In this connection Montgomery proved that

$$S(\{n; M<n\leqslant M+N\}; P,\Omega_p)$$

$$\leqslant (N+2Q^2)\left\{\sum_{\substack{q\leqslant Q\\ q|P(z)}}\prod_{p|q}\frac{\omega(p)}{p-\omega(p)}\right\}^{-1},$$

where

$$P(z)=\prod_{\substack{p\in P\\ p\leqslant z}}p.$$

Using these methods, the following estimate was obtained by Bombieri:

$$\sum_{q\leqslant x^{1/2}(\log x)^{-B}}\max_{y\leqslant x}\max_{\substack{l\\(q,l)=1}}\left|\pi(y,q,l)-\frac{1}{\varphi(q)}\int_2^y\frac{du}{\log u}\right|\ll x(\log x)^{-A},$$

where A is arbitrary and B is a certain function of A (Vaughan). These methods, known collectively as the **large sieve**, were first directed toward proving the **Rényi theorem** stating that every sufficiently large even integer can be represented as the sum of a prime and an almost prime integer (M. B. Barban). Afterward, combining Richert's sieve with this large sieve, Chen Jing-Run (1973) improved this result to a remarkable degree (→ 4 Additive Number Theory C). Several applications of Bombieri's theorem have been demonstrated by P. D. T. A. Elliot and H. Halberstam (1966): e.g., the estimation of the number of representations of n as $p+x^2+y^2$ (Hooley, Linnik) and the estimation of $\sum_{p\leqslant n}d(n-p)$ (Linnik, B. M. Bredihin).

Let $N(\alpha,T,\chi)$ denote the number of zeros of $L(s,\chi)$ in the rectangle $\alpha\leqslant\sigma\leqslant 1$, $|t|\leqslant T$. Combining the large sieve with new Fourier integral techniques and the Turán †power-sum method, Gallagher (1970) proved that there exists a positive constant c satisfying

$$\sum_{\chi \bmod Q}N(\alpha,T,\chi)\ll(QT)^{c(1-\alpha)},$$

$$\sum_{q\leqslant Q}\sum_{\chi \bmod q}N(\alpha,T,\chi)\ll(QT)^{c(1-\alpha)}.$$

Similar results were also obtained by G. Halasz and Montgomery (1969) using another method, which was further exploited by Jutila, M. H. Huxley, and Iwaniec. In particular, Heath-Brown and Iwaniec (1979) deduced that $\pi(x+y)-\pi(x)\geqq cy(\log x)^{-1}$ if $y\geqq x^{11/20+\varepsilon}$. Combined with the Deuring-Heilbronn phenomenon, the zero density theorem above not only establishes the Linnik theory, but also yields the following result, due to K. A. Rodoskiĭ, T. Tatuzawa, and A. I. Vinogradov:

$$\pi(x;q,l)=\frac{1}{\varphi(q)}\int_2^x\frac{du}{\log u}$$

$$-E\frac{\chi_1(l)}{\varphi(q)}\int_2^x\frac{u^{\beta-1}}{\log u}du+O\left(\frac{1}{\varphi(q)}x^{1-\frac{c}{\Delta}}\right)$$

if $x\geqslant\exp(\log q\log\log q)$, where $E=1$ or 0 according as Siegel's zero β exists or not, and where

$$\Delta=\mathrm{Max}(\log q,(\log x)^{2/5}(\log\log x)^{1/5}).$$

Throughout these researches, estimates of the type

$$\sum_{\chi \bmod q}\left|L\left(\frac{1}{2}+it,\chi\right)\right|^4$$

$$\ll\varphi(q)(|t|+2)\log^c\{q(|t|+2)\}$$

and

$$\sum_{\chi \bmod q}\int_{-T}^{T}\left|L\left(\frac{1}{2}+it,\chi\right)\right|^4dt$$

$$\ll\varphi(q)(T+2)\log^c q(T+2)$$

are of great importance and have been studied by A. F. Lavrik (1968), Linnik (1961), Huxley (1972), and K. Ramachandra (1975).

F. The Prime Ideal Theorem in Algebraic Number Fields

In an algebraic number field of finite degree, the prime number theorem is replaced by the **prime ideal theorem** (T. Mitsui, 1956, Fogels, 1962), which is based on the theory of the Hecke †*L*-function (E. Hecke, 1917; Landau, 1918). Let K be a finite Galois extension over an algebraic number field k of finite degree. Suppose that $\mathfrak{p}$ is a prime ideal of k and is not ramified in K. The †Frobenius automorphism of a prime divisor of $\mathfrak{p}$ in K determines a conjugate class C of the Galois group of K/k. Let $\pi(x; C)$ denote the number of prime ideals in k associated with the class C in the above sense and whose norm does not exceed x. Then we have

$$\pi(x, C) = \frac{h(C)}{[K:k]} \operatorname{Li} x + O(xe^{-c\sqrt{\log x}}),$$

where $h(C)$ is the number of elements contained in C and c is a positive constant depending on K/k. This is an extension of Chebotarev's theorem (E. Artin, 1923).

References

[1] R. Ayoub, An introduction to the analytic theory of numbers, Amer. Math. Soc. Math. Surveys, 1963.
[2] H. Bohr and H. Cramér, Die neuere Entwicklung der analytischen Zahlentheorie, Enzykl. Math., II C (8), 1922.
[3] K. Chandrasekharan, Introduction to analytic number theory, Springer, 1968.
[4] P. Erdös, Some recent advances and current problems in number theory, Lectures on Modern Mathematics III, T. L. Saaty (ed.), Wiley, 1965, 196–244.
[5] E. Hecke, Vorlesungen über die Theorie der algebraischen Zahlen, Akademische Verlag, Leipzig, 1923 (Chelsea, 1970).
[6] E. Landau, Handbuch der Lehre von der Verteilung der Primzahlen I, II, Teubner, 1909 (Chelsea, 1969).
[7] E. Landau, Vorlesungen über Zahlentheorie I, II, Hirzel, 1927 (Chelsea, 1946).
[8] G. H. Hardy, Divergent series, Clarendon Press, 1949.
[9] H. M. Edwards, Riemann's zeta function, Academic Press, 1974.
[10] A. E. Ingham, The distribution of prime numbers, Cambridge Univ. Press, 1932.
[11] D. N. Lehmer, List of prime numbers from 1 to 10,006,721, Carnegie Institution of Washington, 1914.
[12] N. Levinson, More than one third of zeros of Riemann's zeta-function on $\sigma = 1/2$, Advances in Math., 13 (1974), 383–436.
[13] K. Prachar, Primzahlverteilung, Springer, 1957.
[14] A. Selberg, An elementary proof of the prime number theorem for arithmetic progressions, Canad. J. Math., 2 (1950), 66–78.
[15] E. C. Titchmarsh, The theory of the Riemann zeta-function, Clarendon Press, 1951.
[16] I. M. Vinogradov, A new estimate of the function $\zeta(1+it)$ (in Russian), Izv. Akad. Nauk SSSR, 22 (1958), 161–164.
[17] M. N. Huxley, The distribution of prime numbers, Oxford Univ. Press, 1972.
[18] P. Turán, Eine neue Methode in der Analysis und deren Anwendungen, Akadémiai Kiadó, Budapest, 1953.
[19] H. L. Montgomery, Topics in multiplicative number theory, Lecture notes in math., 227, Springer, 1971.
[20] E. Bombieri, Le grand crible dans la théorie analytique des nombres, Société Mathématique de France, 1974.
[21] H. Halberstam and H. E. Richert, Sieve methods, Academic Press, 1974.
[22] C. Hooley, Applications of sieve methods to the theory of numbers, Cambridge Univ. Press, 1976.
[23] H. E. Richert, Lectures on sieve methods, Tata Inst., 1976.

124 (XI.8) Distribution of Values of Functions of a Complex Variable

A. General Remarks

Suppose that we are given a function $f: A \to B$ and that the variables z, w take on values in A, B, respectively. A **value distribution** of $f(z)$ is a set of points z where $f(z)$ takes on a certain value w (called w-points of $f(z)$). Value distribution theory is usually concerned with the study of value distributions of complex †analytic functions. Value distribution theory has been developed extensively and deeply for the case where A is the finite plane $|z| < \infty$ or the unit disk $|z| < 1$ and B is the extended complex plane $|w| \leqslant \infty$, and there are many interesting results in this case (→ 272 Meromorphic Functions). Value distributions for analytic functions on general domains or on Riemann

surfaces or of several complex variables have also been studied.

B. The Case of $|z|<R\leqslant\infty$

For a †transcendental entire function $f(z)$, every value (including ∞) is a value of the †cluster set of $f(z)$ at the point at infinity (†Weierstrass's theorem). This theorem was improved in the following way by E. Picard in 1879: A transcendental entire function $f(z)$ has an infinite number of w-points for any finite value w except for at most one finite value (†Picard's theorem). A value w for which the w-points are at most finite is said to be a Picard's exceptional value. E. Borel gave a precise form of this theorem, taking into consideration the order of a function, and G. Julia proved the existence of Julia's directions (→ 272 Meromorphic Functions; 429 Transcendental Entire Functions). After other results in value distribution theory had been obtained by J. Hadamard, G. Valiron, and others, R. Nevanlinna published an important work in 1925 in which he established the so-called **Nevanlinna theory** of meromorphic functions in $|z|<R\leqslant\infty$, unifying results obtained up until that time, and which became the starting point of the subsequent value distribution theory (→ 272 Meromorphic Functions). T. Shimizu and L. V. Ahlfors gave a geometric meaning to the Nevanlinna †characteristic function $T(r)$. Ahlfors established the theory of covering surfaces by metricotopological methods in 1935, and applied it to obtain the Nevanlinna theory and many other results on meromorphic functions. This theory revealed that the topological meaning of the number 2 of Picard's exceptional values is closely related to the †Euler characteristic 2 of the sphere. H. Selberg established the value distribution theory of †algebroidal functions and gave a precise form of G. Rémoundos's theorem, which corresponds to Picard's theorem for algebroidal functions.

Moreover, as an extension of the value distribution theory of meromorphic functions, there is the theory of holomorphic curves or of systems of entire functions. Let $f_0, f_1, \dots f_n$ $(n\geqslant 1)$ be entire functions without common zeros for all and for which $f_0:f_1:\dots:f_n$ is not constant. Put $f=(f_0,f_1,\dots,f_n)$. This is a reduced representation of a nonconstant holomorphic curve $f: C\to P^n(C)$. For $\alpha=(\alpha_0,\alpha_1,\dots,\alpha_n)\in C^{n+1}-\{0\}$, considering the zeros of

$$(\alpha,f)=\alpha_0 f_0+\alpha_1 f_1+\dots+\alpha_n f_n \qquad (\neq 0),$$

we can extend Picard's theorem and the Nevanlinna theory for meromorphic functions to holomorphic curves.

We define the characteristic function $T(r)\ (=T(r,f))$ as

$$T(r)=\frac{1}{2\pi}\int_0^{2\pi}\log|f|\,d\theta\,\bigg|_{r_0}^{r},$$

$$|f|=(|f_0|^2+|f_1|^2+\dots+|f_n|^2)^{1/2},$$

where r_0 is a fixed positive number and $A(r)|_{r_0}^{r}=A(r)-A(r_0)$. Using a bivector $[ff']=(f_if_j'-f_jf_i')$ of f and $f'=(f_0',f_1',\dots,f_n')$, we have another representation of $T(r)$:

$$T(r)=\frac{1}{\pi}\int_{r_0}^{r}\frac{ds}{s}\int_0^{s}\int_0^{2\pi}\frac{|[ff']|^2}{|f|^2}t\,dt\,d\theta.$$

f is transcendental when $\lim T(r)/\log r=\infty$, and the number

$$\lambda=\dim\{(c_0,c_1,\dots,c_n)\in C^{n+1}\mid c_0f_0+c_1f_1+\dots+c_nf_n=0\}$$

is the degeneracy index of f. It holds that $0\leqslant\lambda\leqslant n-1$. As an extension of Picard's theorem, J. Dufresnoy stated that, for transcendental holomorphic curves f and among α in general position, the zeros of (α,f) are infinite except for at most $n+\lambda+1$. X $(\subset C^{n+1}-\{0\})$ is in general position when any p $(\leqslant n+1)$ vectors in X are linearly independent. In connection with this result, there are many Picard-type theorems, and when $\lambda>0$, there are some particular results for holomorphic curves. H. Cartan, H. and J. Weyl, and Ahlfors extended the Nevanlinna theory to holomorphic curves as follows. For $\alpha=(\alpha_0,\alpha_1,\dots,\alpha_n)\in C^{n+1}-\{0\}$, we put

$$|\alpha|=\text{the length of }\alpha,$$

$$\|\alpha f\|=|(\alpha,f)|/|\alpha||f|,\qquad (\alpha,f)\neq 0,$$

and define the proximity function

$$m(r,\alpha)=\frac{1}{2\pi}\int_0^{2\pi}\log\frac{1}{\|\alpha f\|}d\theta$$

and the counting function

$$N(r,\alpha)=\frac{1}{2\pi}\int_0^{2\pi}\log|(\alpha,f)|\,d\theta]\bigg|_{r_0}^{r}.$$

Then we have the first main theorem: For any α for which $(\alpha,f)\not\equiv 0$,

$$T(r)+m(r_0,\alpha)=N(r,\alpha)+m(r,\alpha);$$

and the second main theorem: When $\lambda=0$, for any $\alpha_1,\alpha_2,\dots,\alpha_q$ in general position,

$$(q-n-1)T(r)<\sum_{j=1}^{q}N(r,\alpha_j)+S(r),$$

where $S(r)=O(\log rT(r))$ except for r in a set e of finite logarithmic measure, (i.e., $\int_e d\log r<\infty$). For $\lambda>0$, it is conjectured that "$q-n-1$" can be changed into "$q-n-\lambda-1$." This is unsolved except for some special cases.

Some results show relations between exceptional values and the order as in the case of meromorphic functions. Nevanlinna theory has been extended to the associated curves f^p ($p=1,2,\ldots,n; f^1=f$) for $\lambda=0$, and there have been attempts to extend this theory of holomorphic curves even further by generalizing domains or ranges.

C. The Case of General Domains

The value distributions of meromorphic functions defined in a general domain or an open Riemann surface depend on the function-theoretic "size" of the set of singularities (→ 169 Function-Theoretic Null Sets) or the type of the Riemann surface (→ 367 Riemann Surfaces). For instance, we have the following theorem of Picard type: A single-valued meromorphic function with a set of singularities of †logarithmic capacity zero takes on every value infinitely often in any neighborhood of each singularity except for at most an F_σ-set of values of logarithmic capacity zero (**Hällström-Kametani theorem**). For the study of value distribution at general singularities, it is useful to investigate cluster sets (→ 62 Cluster Sets). In order to generalize the Nevanlinna theory to the case of general domains or Riemann surfaces, we take their exhaustions depending on a real parameter r and define the †counting function and so on (→ 272 Meromorphic Functions). G. J. Hällström established the value distribution theory of meromorphic functions defined in the complementary domain D of a compact set E of logarithmic capacity zero by taking $D_r=\{z \mid v(z)<r\}$ as the exhaustion of D, where $v(z)$ denotes the Evans potential for E, i.e., $v(z)$ is the potential corresponding to a positive mass distribution μ on E of total mass 1 which tends to $+\infty$ as z tends to any point of E. Thus the number of Picard's exceptional values is not greater than $2+\xi$, where $\xi=\limsup_{r\to\infty} F(r)/T(r)$ with $-n(r)=$ the Euler characteristic of D_r and $F(r)=\int_{r_0}^{r} n(r)\,dr$ (Hällström-Tsuji theorem). J. Tamura, L. Sario, and others studied the value distributions of meromorphic functions defined on Riemann surfaces. Sario succeeded in extending the Nevanlinna theory to analytic mappings of a Riemann surface $\mathfrak{R}$ into another Riemann surface $\mathfrak{S}$ by introducing a suitable metric in $\mathfrak{S}$ to define the †proximity function.

In the Nevanlinna theory on general domains, we must sometimes impose conditions that the functions must satisfy in order to obtain a concrete conclusion, for instance, the condition that ξ be finite in the Hällström-Tsuji theorem. But it is also important to determine the domains where some result can be obtained without imposing any additional conditions on the functions. The Hällström-Kametani theorem is an example. Although the set of exceptional values in this theorem cannot be replaced generally by a smaller set than an F_σ-set of logarithmic capacity zero, we have the following theorem: Let E be a †Cantor set with successive ratios $\xi_n=2l_n/l_{n-1}$, where l_n denotes the length of the segments that remain after repeating n times the process of deleting an open segment from the middle of another segment. Then any single-valued meromorphic function with E as the set of singularities has at most 3 Picard's exceptional values if $\lim_{n\to\infty}\xi_n=0$ and at most 2 if $\xi_{n+1}=o(\xi_n^2)$ (L. Carleson, K. Matsumoto). By weakening the conditions on E, one can get several improvements of this result.

References

[1] K. Noshiro, Modern theory of functions (in Japanese), Iwanami, 1954.
[2] A. Dinghas, Vorlesgungen über Funktionentheorie, Springer, 1961.
[3] W. K. Hayman, Meromorphic functions, Clarendon Press, 1964.
[4] H. Cartan, Sur les zéros des combinaisons linéaires de p fonctions holomorphes données, Mathematica, 7 (1933), 5–31.
[5] L. V. Ahlfors, The theory of meromorphic curves, Acta Soc. Sci. Fenn., ser. A, 3 (1941), 3–31.
[6] H. and J. Weyl, Meromorphic functions and analytic curves, Ann. Math. Studies 12, Princeton Univ. Press, 1943.
[7] M. Cowen and P. Griffiths, Holomorphic curves and metrics of negative curvature, J. Analyse Math., 29 (1976), 93–153.
[8] L. Sario and K. Noshiro, Value distribution theory, Van Nostrand, 1966.
[9] K. Matsumoto, Existence of perfect Picard sets, Nagoya Math. J., 27 (1966), 213–222.

125 (XII.7) Distributions and Hyperfunctions

A. History

The advancement of analysis, particularly in the field of partial differential equations and harmonic analysis, necessitated the generalization of the notion of functions and derivatives. For instance, "functions" such as Dirac's †delta function and †Heaviside's function were used by physicists and engineering scientists

even though the former is not a function and the latter is not a differentiable function in the classical sense. The finite parts of divergent integrals, used by J. Hadamard to investigate the fundamental solutions of wave equations (1932), and the Riemann-Liouville integrals due to M. Riesz (1938) were the notions that eventually led to the theory of generalized functions. The rudiments of the idea of distribution, however, can also be found in other earlier works. S. Bochner (1932) and T. Carleman (1944) discussed the Fourier transforms of locally integrable functions on the reals with growth as large as a polynomial. S. L. Sobolev introduced the notion of generalized derivative and also of generalized solution of differential equations by means of integration by parts in studying the Cauchy problem for hyperbolic equations (1936); J. Leray (1934), K. O. Friedrichs (1939), and C. B. Morrey, Jr. (1940) also discussed generalized derivatives. On the other hand, L. Fantappié (1943) investigated analytic functionals that are elements of the dual of the space of analytic functions and applied them to the theory of partial differential equations. Based on a systematic generalization of these investigations, L. Schwartz [1] established the theory of distributions (1945), which not only provided a mathematical foundation for a number of formal methods that had been used in mathematical physics but also gave new and powerful tools for the theories of differential equations (L. Hörmander [2]) and †Fourier transforms. Furthermore, it has been applied to †representation theory of locally compact groups, the theory of probability, and the theory of manifolds (G. de Rham [3]). As will be seen in Section B, distributions are defined as continuous †functionals on a certain function space, and it is essential to select a function space appropriate to the problems concerned. For this reason, I. M. Gel'fand and G. E. Shilov defined various classes of generalized functions [4] as a natural extension of Schwartz's theory. In this direction there are also various classes of ultradistributions introduced by C. Roumieu [5] and A. Beurling.

Another but completely different approach was given by M. Sato [6, 15] in the form of the theory of hyperfunctions (1958). Intuitively a hyperfunction is the sum of (ideal) boundary values of holomorphic functions at the real axis. We obtain in this way an ultimate class of localizable generalized functions. Sato employed the relative (or local) cohomology theory to define the "boundary values" and to prove their localizing property. Such an algebraic approach to generalized functions led naturally to an algebraic treatment of systems of linear partial differential operators [7], called **algebraic analysis** by comparison with algebraic geometry. Independently, Hörmander established a similar theory for distributions using Fourier analysis (→ 274 Microlocal Analysis).

B. Definition of Distributions

Let $\varphi(x)$ be a complex-valued function of $x=(x_1, \dots, x_n)$ defined on an open set Ω in the n-dimensional Euclidean space $\mathbf{R}^n$. By the **support** (or **carrier**) of φ, denoted by $\operatorname{supp}\varphi$, we mean the †closure of $\{x \mid \varphi(x) \neq 0\}$ in Ω. For multi-indices p, i.e., n-tuples $p=(p_1, \dots, p_n)$ of nonnegative integers, we set $|p|=p_1+\dots+p_n$. For a function $\varphi(x)$ of class $C^{|p|}$, we write

$$D^p\varphi=\frac{\partial^{|p|}\varphi(x)}{\partial x_1^{p_1}\dots\partial x_n^{p_n}}. \qquad (1)$$

In particular, $D^{(0,\dots,0)}\varphi=\varphi$. To indicate the variables x we shall adopt the notation D_x^p.

$\mathscr{D}(\Omega)$ denotes the set of all complex-valued functions of class C^∞ defined on Ω with †compact support, which is a †linear space under the usual addition and scalar multiplication in function spaces. A sequence $\{\varphi_m\}$ in $\mathscr{D}(\Omega)$ is said to converge to 0 (the function identically equal to zero) as $m\to\infty$, denoted by $\varphi_m \Rightarrow 0$, if there exists a compact set E in Ω such that E contains $\operatorname{supp}\varphi_m$ for every m, and for every p, $\{D^p\varphi_m\}$ converges uniformly to 0 as $m\to\infty$.

We sometimes abbreviate $\mathscr{D}(\Omega)$ either as $\mathscr{D}$ or, when we want to indicate the variables x, as $\mathscr{D}_x$.

A complex-valued †linear functional T defined on $\mathscr{D}(\Omega)$ is called a **distribution on** Ω if it is continuous on $\mathscr{D}(\Omega)$, i.e., $\varphi_m \Rightarrow 0$ implies $T(\varphi_m)\to 0$. The set of all distributions on Ω is denoted by $\mathscr{D}'(\Omega)$ (or $\mathscr{D}'$). For distributions S and T, the sum $S+T$ and scalar multiple αT are defined by $(S+T)(\varphi)=S(\varphi)+T(\varphi)$ and $(\alpha T)(\varphi)=\alpha T(\varphi)$, respectively, which are also distributions. Hence $\mathscr{D}'(\Omega)$ is a †linear space.

C. Examples of Distributions

(1) Let $f(x)$ be a †measurable and †locally integrable function on Ω. Then a distribution T_f is defined by $T_f(\varphi)=\int\varphi(x)f(x)dx$. Here dx is the †Lebesgue measure and the domain of integration is Ω (in fact, $\operatorname{supp}\varphi$). If $T_f=T_g$, then $f(x)=g(x)$ †almost everywhere. Thus we can identify the distribution T_f with the corresponding function f, and sometimes T_f will be denoted simply by f.

(2) Let μ be a †Radon measure on Ω, i.e., a complex-valued †regular completely additive set function on the †Borel sets in Ω. Then a distribution T_μ is defined by $T_\mu(\varphi)=\int\varphi(x)\mu(dx)$. Example (1) is a special case where

$\mu(dx)=f(x)dx$. Two Radon measures μ and ν coincide if $T_\mu=T_\nu$. If the measure is concentrated at the origin, then $T_\mu(\varphi)=c\varphi(0)$, denoted by $c\delta$, and δ is called **Dirac's distribution**. Sometimes δ is denoted by δ_x, $\delta_{(x)}$, or $\delta(x)$ to indicate that it operates on functions of x. A distribution $T\in\mathcal{D}'(\Omega)$ is a measure if and only if $T(\varphi_m)\to 0$ whenever the supports $\operatorname{supp}\varphi_m$ of a sequence $\varphi_m\in\mathcal{D}(\Omega)$ are in a fixed compact set in Ω and φ_m converges uniformly to zero. A distribution T is called a **positive distribution** if $T(\varphi)\geqslant 0$ for any $\varphi\in\mathcal{D}(\Omega)$ that has nonnegative values at all points x in Ω. Every positive distribution is equal to a T_μ corresponding to a positive measure μ.

(3) For given p the distribution $\delta^{(p)}$ is defined by $\delta^{(p)}(\varphi)=(-1)^{|p|}D^p\varphi(0)\cdot\delta^{(0,\ldots,0)}=\delta$.

(4) Let $g(x)$ be a function defined but not integrable on an interval (a,b), and assume that for any positive number ε it is integrable on $(a+\varepsilon,b)$. Moreover, assume that

$$g(x)=\sum_{\nu=1}^{n}A_\nu(x-a)^{-\lambda_\nu}+h(x),$$

where $\operatorname{Re}\lambda_\nu>1$, λ_ν is not an integer, and $h(x)$ is integrable on (a,b). Then

$$\int_{a+\varepsilon}^{b}g(x)dx-\sum A_\nu(\lambda_\nu-1)^{-1}\varepsilon^{1-\lambda_\nu}=F(\varepsilon)$$

tends to a finite value as $\varepsilon\to 0$. This limit is called the **finite part** (in French *partie finie*) of the integral $\int_a^b g(x)dx$, denoted by $\mathrm{Pf}\int_a^b g(x)dx$:

$$\mathrm{Pf}\int_a^b g(x)dx$$

$$=-\sum_\nu\frac{A_\nu}{\lambda_\nu-1}(b-a)^{1-\lambda_\nu}+\int_a^b h(x)dx.$$

In the same way, for every $\varphi\in\mathcal{D}(\mathbf{R}^1)$, $T(\varphi)=\mathrm{Pf}\int_a^b g(x)\varphi(x)dx$ is defined, and T is a distribution which will be denoted by $\mathrm{Pf}\,g$, and which is frequently called a **pseudofunction**. This notion is extended to the n-dimensional case and used to express the fundamental solutions (→ Section EE; Appendix A, Table 15.V) of †hyperbolic partial differential equations.

D. Localization of Distributions

Let Ω' be an open subset of Ω. Every function $\varphi\in\mathcal{D}(\Omega')$ can be extended to a function $\bar\varphi\in\mathcal{D}(\Omega)$ by setting $\bar\varphi(x)=0$ for $x\notin\Omega'$. Thus if $T\in\mathcal{D}'(\Omega)$, then a distribution $S\in\mathcal{D}'(\Omega')$ is defined by $S(\varphi)=T(\bar\varphi)$ for every $\varphi\in\mathcal{D}(\Omega')$, and S is called the **restriction** of T to Ω'. Two distributions T and S are said to be equal in Ω' if their restrictions to Ω' are equal. If every point in Ω has a neighborhood where $T=S$, then $T=S$. In this sense a distribution is determined completely by its local data, although the notion of pointwise value, having exact meaning for functions, has no meaning for distributions. Moreover, we can construct a distribution with given local data in the following way: Suppose that an †open covering $\{\Omega_j\}$ of Ω and a set of distributions $T_j\in\mathcal{D}'(\Omega_j)$ are given and that T_j and T_k are equal in $\Omega_j\cap\Omega_k$ for any j and k. Then there exists a unique distribution $T\in\mathcal{D}'(\Omega)$ such that $T=T_j$ in Ω_j for each j. (Define $T(\varphi)=\sum T_j(\alpha_j\varphi)$ for a †partition of unity $\{\alpha_j\}$ subordinate to $\{\Omega_j\}$). Namely, the distributions $\mathcal{D}'$ form a †sheaf of linear spaces over $\mathbf{R}^n$. It is not †flabby but is soft, i.e., for any distribution T defined on a neighborhood of a closed set F in $\mathbf{R}^n$, there is a distribution on $\mathbf{R}^n$ that coincides with T on a neighborhood of F.

For each distribution $T\in\mathcal{D}'(\Omega)$ there is a largest open subset Ω' of Ω on which T vanishes. Its complement is called the **support** (or **carrier**) of T and is denoted by $\operatorname{supp}T$. The support of the distribution given in example (1) coincides with that of the function f. The support of $\delta^{(p)}$ is the origin.

E. Derivatives of Distributions

In example (1) in Section C, if $f(x)$ is a function of class C^k, then by integration by parts $T_{D^pf}(\varphi)=(-1)^{|p|}T_f(D^p\varphi)$ for $|p|\leqslant k$. The right-hand side defines a distribution even if f is not differentiable. In view of this example, we define **derivatives** D^pT of any distribution T by

$$(D^pT)(\varphi)=(-1)^{|p|}T(D^p\varphi),\qquad\varphi\in\mathcal{D}_x.\tag{2}$$

Any distribution is infinitely differentiable. Any locally integrable function is infinitely differentiable in the sense of distributions, and its derivatives D^pT_f are called **distribution derivatives** (or **generalized derivatives** or **weak derivatives**).

For example: (1) $D^p\delta=\delta^{(p)}$; (2) (1-dimensional case) $dx_+/dx=\mathbf{1}$, $d\mathbf{1}/dx=\delta$, where $x_+=\max(x,0)$ and $\mathbf{1}(x)$ is **Heaviside's function**, which is equal to 1 for $x\geqslant 0$ and to 0 for $x<0$.

F. The Operation of Linear Differential Operators on Distributions

Let T be a distribution and $\alpha(x)$ a C^∞-function on Ω. Then we can define the product αT by $(\alpha T)(\varphi)=T(\alpha\varphi)$, where the right-hand side is a continuous linear functional on $\mathcal{D}(\Omega)$ since the multiplication $\varphi\mapsto\alpha\varphi$ is continuous in $\mathcal{D}(\Omega)$. We define the **dual operator** (or **conjugate operator**) $P'(x,D)$ of a linear differential operator $P(x,D)=\sum a_p(x)D^p$ by

$$P'(x,D)\varphi=\sum(-1)^{|p|}D^p(a_p(x)\varphi(x)),\tag{3}$$

where the $a_p(x)$ are C^∞-functions on Ω. Com-

bining differentiation of distributions, multiplication by functions, and addition, we can apply partial differential operators to distributions, and we have $(P(x,D)T)(\varphi)=T(P'(x,D)\varphi)$.

Linear differential operators commute with restriction mappings. In particular, we have $\operatorname{supp} P(x,D)T\subset \operatorname{supp} T$. In other words, linear differential operators are †sheaf homomorphisms on the sheaf $\mathscr{D}'$ of distributions over Ω. On the other hand, every continuous sheaf homomorphism on $\mathscr{D}'$ is a linear differential operator whose order is finite on each compact set in Ω. Even when no continuity is assumed, a sheaf homomorphism on $\mathscr{D}'$ is a linear differential operator except on a discrete subset of Ω (J. Peetre).

G. The Topology of $\mathscr{D}$ and $\mathscr{D}'$

Let $\varnothing=K_1\Subset K_2\Subset\ldots$ be a sequence of compact sets in Ω that exhausts Ω. For every nondecreasing sequence of positive numbers $\{a\}=\{a_0,a_2,\ldots\}$ and nondecreasing sequence of nonnegative integers $\{k\}=\{k_0,k_1,\ldots\}$ we set

$$\rho_{\{a\},\{k\}}(\varphi)=\sup_{j\geqslant 0}\sup_{|p|\leqslant k_j}\sup_{x\notin K_j} a_j|D^p\varphi(x)| \qquad (4)$$

for $\varphi\in\mathscr{D}(\Omega)$. We define a †locally convex topology of the space $\mathscr{D}(\Omega)$ by taking the totality of the †seminorms $\rho_{\{a\},\{k\}}$ as a fundamental system of continuous seminorms. Then $\mathscr{D}(\Omega)$ is a †nuclear †(LF)-space, and the convergence $\varphi_m\Rightarrow 0$ of a sequence φ_m is identical to the convergence $\varphi_m\to 0$ in this topology. A subset $B\subset\mathscr{D}(\Omega)$ is †bounded in the topology if and only if there exist a compact set $E\subset\Omega$ such that $\operatorname{supp}\varphi\subset E$ for every $\varphi\in B$, and positive numbers M_p for each p such that $\sup|D^p\varphi(x)|\leqslant M_p$ for every $\varphi\in B$.

The topology of the space $\mathscr{D}'(\Omega)$ is the †strong topology of the †dual of $\mathscr{D}(\Omega)$: the †topology of uniform convergence on every bounded set in $\mathscr{D}(\Omega)$. Under this topology $\mathscr{D}(\Omega)$ is a †nuclear †reflexive linear topological space. With respect to this topology, the linear differential operators of Section F are continuous in $\mathscr{D}'$.

By virtue of the following convergence theorems, various limiting processes concerning distributions can be treated easily. If the limit $\lim T_j(\varphi)=T(\varphi)$ exists for every $\varphi\in\mathscr{D}$, where $\{T_j\}$ is a sequence of distributions, then $T\in\mathscr{D}'$ and $\{T_j\}$ is convergent to the distribution T (**convergence theorem**). Moreover, for any p, D^pT_j is convergent to D^pT (**theorem of termwise differentiation**). Any bounded set in $\mathscr{D}$ or $\mathscr{D}'$ is totally bounded. Thus weak convergence of a sequence $\{T_j\}$ implies strong convergence (convergence in the topology of the space $\mathscr{D}'$) (**strong convergence theorem**).

H. Distributions Depending on a Parameter

Consider distributions T_λ depending on a parameter λ, where λ ranges over the real line, the complex plane, or more generally an open set in Euclidean space. The convergence theorem and the strong convergence theorem also hold in the case of a continuous parameter λ. Thus T_λ is **continuous (differentiable) with respect to the parameter** λ if $T_\lambda(\varphi)$ is continuous (differentiable) with respect to λ for any $\varphi\in\mathscr{D}$. If T_λ is continuous with respect to λ, then $D_x^pT_\lambda$ is also continuous with respect to λ. If T_λ is differentiable with respect to a real variable λ, then $D_x^pT_\lambda$ is also differentiable, and $\partial D_x^pT_\lambda/\partial\lambda=D_x^p(\partial T_\lambda/\partial\lambda)$. The same facts hold for the case of several real variables. For a complex parameter λ, T_λ is **holomorphic with respect to** λ if $T_\lambda(\varphi)$ is holomorphic for any $\varphi\in\mathscr{D}$. The fundamental properties of holomorphic functions also hold in this case.

If T_λ is defined and continuous in an interval $[a,b]$, then the **integral** $T=\int T_\lambda d\lambda$ **with respect to** λ exists in $\mathscr{D}'$ and we have

$$T(\varphi)=\int_a^b T_\lambda(\varphi)\,d\lambda, \qquad \varphi\in\mathscr{D}.$$

I. Distributions with Compact Support

We denote by $\mathscr{E}(\Omega)$ (abbreviated to $\mathscr{E}$) the space of all complex valued C^∞-functions on Ω. $\mathscr{E}(\Omega)$ is a nuclear †Fréchet space with the locally convex topology defined by the seminorms

$$\rho_{p,K}(\varphi)=\sup_{x\in K}|D^p\varphi(x)| \qquad (5)$$

as p ranges over all multi-indices and K ranges over all compact sets in Ω.

We denote by $\mathscr{E}'(\Omega)$ (abbreviated to $\mathscr{E}'$) the †strong dual of $\mathscr{E}(\Omega)$, i.e., the set of all continuous linear functionals on $\mathscr{E}(\Omega)$ equipped with the topology of uniform convergence on bounded sets in $\mathscr{E}(\Omega)$. $\mathscr{E}'(\Omega)$ is a nuclear †(DF)-space. If $T\in\mathscr{E}'(\Omega)$, then its restriction to $\mathscr{D}(\Omega)$ is a distribution with compact support. Conversely if T is a distribution with compact support in Ω, then choosing $\alpha\in\mathscr{D}(\Omega)$ which is identically equal to 1 on a neighborhood of the support of T, we define a linear functional S on $\mathscr{E}(\Omega)$ by $S(\varphi)=T(\alpha\varphi)$. Then $S\in\mathscr{E}'(\Omega)$, and S is independent of the choice of α. In this sense, we can consider that S is the extension of T to $\mathscr{E}(\Omega)$ and identify S and T. thus $\mathscr{E}'(\Omega)$ coincides with the set of all distributions with compact support in Ω.

J. Structure Theorems for Distributions

A distribution $T\in\mathscr{D}'(\Omega)$ is said to be of order at most k if $|T(\varphi)|\leqslant\rho_{\{a\},\{k\}}(\varphi)$ for $\{k\}=\{k,k,k,\ldots\}$

and for some $\{a\}$. A distribution of order 0 is a measure. Every **distribution of finite order** can be represented as a finite linear combination of derivatives (in the sense of distributions) of measures or locally integrable functions. The restriction of any distribution to a relatively compact open subset Ω' can be represented in this way because it is of finite order. Therefore the distributions form the smallest class of generalized functions that contains all locally integrable functions, is stable under differentiation, and forms a sheaf. Furthermore, for any distribution T there exists a sequence $\{\varphi_j\}\subset\mathscr{D}$ that converges to T in the distribution sense.

A closed set F is said to be **regular** if for every point a in F we have a neighborhood U of a and constants ω and $1\geqslant\alpha>0$ such that every pair of points x and y in $F\cap U$ is connected by a curve contained in F with length less than or equal to $\omega|x-y|^{\alpha}$. If the support of a distribution T is contained in a regular closed set F, then

$$T=\sum D^{p_j}T_{\mu_j}$$

(the expression is not necessarily unique, and the sum is locally finite), where the μ_j are complex-valued measures with support contained in F. In particular, if the support of a distribution T contains only one point a, then T can be represented uniquely as a finite linear combination

$$T=\sum_{|p|\leqslant m}\alpha_p D^p\delta_{(x-a)}$$

of derivatives of the distribution $\delta_{(x-a)}$ defined by $\delta_{(x-a)}(\varphi)=\varphi(a)$.

K. Tensor Products of Distributions

Let $\Omega_x\subset\mathbf{R}^n$ and $\Omega_y\subset\mathbf{R}^m$ be open sets. If $T\in\mathscr{D}'(\Omega_x)$ and $S\in\mathscr{D}'(\Omega_y)$, then there is a unique distribution $T\otimes S\in\mathscr{D}'(\Omega_x\times\Omega_y)$, called the **tensor product** (or **direct product**), such that $T\otimes S(\varphi(x)\psi(y))=T(\varphi)S(\psi)$ for any $\varphi\in\mathscr{D}(\Omega_x)$ and $\psi\in\mathscr{D}(\Omega_y)$. More directly we have Fubini's theorem: $T\otimes S(\varphi(x,y))=T_x(S_y(\varphi(x,y)))=S_y(T_x(\varphi(x,y)))$ for any $\varphi\in\mathscr{D}(\Omega_x\times\Omega_y)$. If F and G are linear spaces of distributions on Ω_x and Ω_y, respectively, then the †tensor product $F\otimes G$ is identified with the linear combinations of tensor products $T\otimes S$ of $T\in F$ and $S\in G$, which form a linear space of distributions on $\Omega_x\times\Omega_y$. When F and G have †locally convex topologies, the †completed tensor products $F\hat{\otimes} G$ and $F\hat{\hat{\otimes}} G$ are also usually identified with subspaces of $\mathscr{D}'(\Omega_x\times\Omega_y)$. For example, we have $\mathscr{D}'(\Omega_x)\hat{\otimes}\mathscr{D}'(\Omega_y)=\mathscr{D}'(\Omega_x\times\Omega_y)$, $\mathscr{E}'(\Omega_x)\hat{\otimes}\mathscr{E}'(\Omega_y)=\mathscr{E}'(\Omega_x\times\Omega_y)$ and $\mathscr{E}(\Omega_x)\hat{\otimes}\mathscr{E}(\Omega_y)=\mathscr{E}(\Omega_x\times\Omega_y)$ (including the topologies). Similarly, we have $\mathscr{D}(\Omega_x)\hat{\otimes}\mathscr{D}(\Omega_y)=\mathscr{D}(\Omega_x\times\Omega_y)$, but the left-hand side has a strictly weaker topology than the right-hand side. Spaces of distributions with parameters (→ Section H) are often identified with completed tensor products of spaces of distributions [8–11].

L. The Kernel Theorem

Let Ω_x and Ω_y be as above. Every distribution K on $\Omega_x\times\Omega_y$ induces the continuous linear mapping $L_K:\mathscr{D}(\Omega_x)\to\mathscr{D}'(\Omega_y)$ defined by $(L_K(\varphi(x)))(\psi(y))=K(\varphi(x)\psi(y))$, $\psi\in\mathscr{D}(\Omega_y)$. Conversely every continuous linear mapping $L:\mathscr{D}(\Omega_x)\to\mathscr{D}'(\Omega_y)$ is equal to L_K for a $K\in\mathscr{D}'(\Omega_x\times\Omega_y)$ and the correspondence $K\mapsto L_K$ is a topological isomorphism of the locally convex space $\mathscr{D}'(\Omega_x\times\Omega_y)$ onto $L(\mathscr{D}(\Omega_x),\mathscr{D}'(\Omega_y))$ equipped with the topology of uniform convergence on the bounded sets (L. Schwartz's kernel theorem, [8–11]). K is called the **kernel (distribution)** of the mapping L_K.

L_K is a continuous linear mapping $\mathscr{D}(\Omega_x)\to\mathscr{E}(\Omega_y)$ (resp. $\mathscr{D}(\Omega_x)\to\mathscr{E}'(\Omega_y)$) if and only if $K\in\mathscr{D}'(\Omega_x)\hat{\otimes}\mathscr{E}(\Omega_y)$ (resp. $\mathscr{D}'(\Omega_x)\hat{\otimes}\mathscr{E}'(\Omega_y)$), in which case K is said to be **regular** (resp. **compact**) **in** y. L_K can be extended to a continuous linear mapping $\mathscr{E}'(\Omega_x)\to\mathscr{D}'(\Omega_y)$ (resp. $\mathscr{E}(\Omega_x)\to\mathscr{D}'(\Omega_y)$) if and only if $K\in\mathscr{E}(\Omega_x)\hat{\otimes}\mathscr{D}'(\Omega_y)$ (resp. $\mathscr{E}'(\Omega_x)\hat{\otimes}\mathscr{D}'(\Omega_y)$), in which case K is said to be **regular** (resp. **compact**) **in** x. K is said to be **regular** (resp. **compact**) if it is regular (resp. compact) both in x and y. L_K can be extended to a continuous linear mapping $\mathscr{E}'(\Omega_x)\to\mathscr{E}(\Omega_y)$ (resp. $\mathscr{E}(\Omega_x)\to\mathscr{E}'(\Omega_y)$) if and only if $K\in\mathscr{E}(\Omega_x\times\Omega_y)$ (resp. $\mathscr{E}'(\Omega_x\times\Omega_y)$). Then K is said to be **regularizing** (resp. **compactifying**) [8, 10].

M. Convolution

For distributions S and T on $\mathbf{R}^n$, assume that either S or T has a compact support or more generally that $\operatorname{supp}T\cap(\{x\}-\operatorname{supp}S)$ is locally uniformly bounded with respect to x. Then the correspondence

$$\varphi\in\mathscr{D}\to S_{(x)}T_{(y)}(\varphi(x+y))$$

defines a distribution called the **convolution** of S and T and is denoted by $S*T$. In particular, $T_f*T_g=T_{f*g}$, where $f*g$ is the †convolution of functions f and g, and T_f, T_g, and T_{f*g} denote the distributions corresponding to f, g, and $f*g$, respectively. For example: $T*\delta=\delta*T=T$, $D^pT*S=T*(D^pS)=D^p(T*S)$.

Thus a solution of the partial differential equation $P(D)T=S$ with constant coefficients is given by the convolution $S*E$ of S with a †fundamental solution E of $P(D)$, whenever the convolution exists.

If $T\in\mathscr{D}'$ and $\varphi\in\mathscr{D}$, the convolution $T*\varphi$ is equal to a function f of class C^∞ (or the

distribution corresponding to f), and $f(y)=T_{(x)}(\varphi(y-x))$. f is called the **regularization** of T. For distributions S and T, assume that $|f(x)g(y-x)|$ is integrable on $\mathbf{R}^n$ for any regularization $f=S*\varphi$ and $g=T*\psi$. Then there exists a unique distribution V such that

$$V*(\varphi*\psi)=\int f(x)g(y-x)\,dx.$$

This distribution V is called the **generalized convolution** and is denoted by $S*T$ (C. Chevalley).

N. Tempered Distributions $\mathscr{S}'$

We denote by $\mathscr{S}(\mathbf{R}^n)$ (abbreviated to $\mathscr{S}$) the space of all †rapidly decreasing functions of class C^∞ on $\mathbf{R}^n$. $\mathscr{S}$ is a nuclear Fréchet space with the locally convex topology defined by the seminorms

$$p_{p,q}(\varphi)=\sup_x |x^p D^q\varphi(x)|, \tag{6}$$

as p and q range over all multi-indices, where $x^p=x_1^{p_1}\dots x_n^{p_n}$. We denote by $\mathscr{S}'(\mathbf{R}^n)$ (abbreviated to $\mathscr{S}'$) the strong dual of $\mathscr{S}(\mathbf{R}^n)$. Its elements are called **tempered distributions** (or **slowly increasing distributions**) and are identified with their restrictions to $\mathscr{D}(\mathbf{R}^n)$. A distribution $T\in\mathscr{D}'$ can be continuously extended to $\mathscr{S}$ if and only if any regularization $f=T*\varphi$ of T is a slowly increasing continuous function (i.e., there is a polynomial $P(x)$ such that $|f(x)|\leqslant|P(x)|$). $\mathscr{S}'(\mathbf{R}^n)$ is a nuclear (DF)-space. Let $1<p<\infty$, and let q be the †conjugate exponent. The strong dual of the †function space $\mathscr{D}_{L_q}(\mathbf{R}^n)$ is denoted by $\mathscr{D}'_{L_p}(\mathbf{R}^n)$. Since $\mathscr{S}$ is dense in $\mathscr{D}_{L_q}$, $\mathscr{D}'_{L_p}$ is identified with a linear subspace of $\mathscr{S}'$. The strong dual of $\mathscr{D}'_{L_p}(\mathbf{R}^n)$ is equal to $\mathscr{D}_{L_q}(\mathbf{R}^n)$. Let $\dot{\mathscr{B}}(\mathbf{R}^n)$ be the closed linear subspace of $\mathscr{B}(\mathbf{R}^n)$ consisting of all functions φ such that $D^p\varphi\in C_\infty(\mathbf{R}^n)$ for all p (→ 168 Function Spaces). It is also the closure of $\mathscr{S}$ in $\mathscr{B}$. The strong dual of $\dot{\mathscr{B}}$ is denoted by $\mathscr{D}'_{L_1}(\mathbf{R}^n)$. Its elements are called **integrable distributions** because the strong dual of $\mathscr{D}'_{L_1}(\mathbf{R}^n)$ is equal to $\mathscr{B}(\mathbf{R}^n)$ so that the integral $T(1)$ has a meaning. Here 1 is the function identically equal to 1.

†Sobolev spaces $W_p^l(\mathbf{R}^n)$ and †Besov spaces $B_{p,q}^s(\mathbf{R}^n)$ are also linear subspaces of $\mathscr{S}'(\mathbf{R}^n)$.

O. Fourier Transforms

The Fourier transform

$$\mathscr{F}\varphi(x)=(\sqrt{2\pi})^{-n}\int\varphi(\xi)\exp(-ix\xi)\,d\xi \tag{7}$$

is an isomorphism of $\mathscr{S}_\xi$ onto $\mathscr{S}_x$, where $x\xi=x_1\xi_1+\dots+x_n\xi_n$. The Fourier transform $\mathscr{F}T\in\mathscr{S}'_\xi$ of $T\in\mathscr{S}'_x$ is defined by

$$(\mathscr{F}T)(\varphi)=T(\mathscr{F}\varphi),\qquad \varphi\in\mathscr{S}. \tag{8}$$

The **inverse Fourier transform** is defined similarly, except that $-i$ is replaced by i. For example: (1) $\mathscr{F}1=(\sqrt{2\pi})^n\delta$, where 1 is the distribution corresponding to the function 1. (2) $\mathscr{F}(D^pT)=i^{|p|}x^p\mathscr{F}T$.

A function φ of class C^∞ is called a **slowly increasing C^∞-function** if φ and its derivatives of any order are slowly increasing continuous functions. The space $\mathscr{O}_M$ is the set of all such functions. Its image under the Fourier transform $\mathscr{F}(\mathscr{O}_M)$, denoted by $\mathscr{O}'_C$, coincides with the set of all distributions T such that any regularization $f=T*\varphi$ is a rapidly decreasing C^∞-function. A member of the space $\mathscr{O}'_C$ is called a **rapidly decreasing distribution.** If $\alpha\in\mathscr{O}_M$ ($\beta\in\mathscr{O}'_C$) and $T\in\mathscr{S}'$, then $\alpha T\in\mathscr{S}'$ and $\mathscr{F}(\alpha T)=(\sqrt{2\pi})^{-n}\mathscr{F}\alpha*\mathscr{F}T$ ($\beta*T\in\mathscr{S}'$ and $\mathscr{F}(\beta*T)=(\sqrt{2\pi})^n\mathscr{F}\beta\mathscr{F}T$). If T is a distribution with compact support in $\mathbf{R}^n$, then its Fourier transform is the †entire function $f(\zeta)=(\sqrt{2\pi})^{-n}T(\exp(-ix\zeta))$, where $\zeta=\xi+i\eta\in\mathbf{C}^n$. For a convex compact set K in $\mathbf{R}^n$, define its **supporting function** H_K by $H_K(\eta)=\sup_{x\in K}x\eta$.

Paley-Wiener theorem. An entire function $f(\zeta)$ on $\mathbf{C}^n$ is the Fourier transform of a distribution (resp. a function of class C^∞) with support in K if and only if (i) for any $\varepsilon>0$ there is a constant C_ε such that $|f(\zeta)|\leqslant C_\varepsilon\exp(H_K(\eta)+\varepsilon|\zeta|)$ on $\mathbf{C}^n$, and (ii) there are constants C and N such that $|f(\xi)|\leqslant C(1+|\xi|)^N$ (resp. for any N there is a constant C such that $|f(\xi)|\leqslant C(1+|\xi|)^{-N}$) on $\mathbf{R}^n$.

P. Fourier Series and Distributions on Tori

The n-dimensional †torus $\mathbf{T}^n$ is the quotient space of $\mathbf{R}^n$ with respect to the equivalence relation $x_j\equiv y_j \bmod \mathbf{Z}$ ($j=1,\dots,n$). The space $\mathscr{D}'(\mathbf{T}^n)$ of distributions on $\mathbf{T}^n$ is defined to be the strong dual of $\mathscr{D}(\mathbf{T}^n)=\mathscr{E}(\mathbf{T}^n)$. The volume element dx of $\mathbf{T}^n$ is defined from that of $\mathbf{R}^n$. Thus, for an integrable function f, we can define a distribution T_f by

$$T_f(\varphi)=\int_{\mathbf{T}^n}f(x)\varphi(x)\,dx,\qquad \varphi\in\mathscr{D}(\mathbf{T}^n).$$

Consider the family of functions $f_p(x)=\exp(2\pi ipx)$, where p ranges over all n-tuples of integers. Then any distribution T on $\mathbf{T}^n$ has the **Fourier series** expansion $T=\sum c_pf_p(x)$, where the Fourier coefficients $c_p=T(f_{-p})$ are slowly increasing, i.e., $|c_p|\leqslant C(1+|p|^2)^k$ for some k and C. Conversely if $\{c_p\}$ is a slowly increasing sequence, then $\sum c_pf_p$ converges to a distribution T on $\mathbf{T}^n$.

Q. Substitution

Let $f=(f_1,\ldots,f_m)$ be a C^∞-mapping from an open set Ω in $\mathbf{R}^n$ into an open set $\tilde{\Omega}$ in $\mathbf{R}^m$, and assume that the †rank of the Jacobian matrix $(\partial f_i/\partial x_j)$ $(i=1,\ldots,m; j=1,\ldots,n)$ is equal to m in a neighborhood of the †inverse image E of the support of a distribution $S=S_{(y)}\in\mathcal{D}'(\tilde{\Omega})$. In a neighborhood U of E, we can choose $u_1=g_1(x),\ldots,u_{n-m}=g_{n-m}(x)$ so that the transformation $(y,u)=(f(x),g(x))$ has the inverse transformation $x=\psi(y,u)$ of class C^∞. If the support of $\varphi\in\mathcal{D}(\Omega)$ is contained in U, then defining $\tilde{\varphi}$ by

$$\tilde{\varphi}(y)=\int_{\mathbf{R}^{n-m}}\varphi(\psi(y,u))J(y,u)\,du,$$

we have $\tilde{\varphi}\in\mathcal{D}(\tilde{\Omega})$, where $J(y,u)$ is the absolute value of the †Jacobian of $\psi(y,u)$, and $\tilde{\varphi}$ is independent of the choice of $g_1\ldots,g_{n-m}$. Now we define $T(\varphi)=S(\tilde{\varphi})$ if $\operatorname{supp}\varphi\subset U$, and $T(\varphi)=0$ if $\operatorname{supp}\varphi$ does not intersect E. Since a distribution can be determined by its local behavior, we have a distribution $T\in\mathcal{D}'(\Omega)$ with support E, which is denoted by $T=S\circ f=S(f)=S_{f(x)}$ and is called the **substituted distribution** of $S_{(y)}$ by $y=f(x)$. It is also called the **pullback** of S by f and is denoted by f^*S. The chain rule for derivatives of composites

$$\frac{\partial}{\partial x_j}(S\circ f)=\sum_{i=1}^{m}\frac{\partial f_i}{\partial x_j}\left(\frac{\partial S}{\partial y_i}\circ f\right) \qquad (9)$$

also holds.

For example: (1) If $S=\delta_{(y)}^{(q)}$ $(y\in\mathbf{R}^m)$, then E is the surface $f_1(x)=f_2(x)=\ldots=f_m(x)=0$. Assuming that $(f_1,\ldots,f_m)$ satisfies the condition in the previous paragraph, we obtain the distribution $\delta^{(q)}(f_1,\ldots,f_m)$ with support contained in a surface. From the fact $\delta^{(q)}=D_y^q\delta$, we can write

$$\delta^{(q)}(f_1,\ldots,f_m)=\frac{\partial^{|q|}\delta(f_1,\ldots,f_m)}{(\partial f_1)^{q_1}\ldots(\partial f_m)^{q_m}}$$

(Gel'fand-Shilov notation). (2) For the mapping $f(x)=Ax+b$, where A is an $n\times n$ regular matrix and b is an n-vector, we can define the substituted distribution of S by $f=f(x)$ for any $S\in\mathcal{D}'(\mathbf{R}^n)$. For instance, $\delta_{(x-b)}(\varphi)=\varphi(b)$. $\delta_{(x^2-c^2)}=(2c)^{-1}(\delta_{(x-c)}+\delta_{(x+c)})$ $(c>0, x\in\mathbf{R}^1)$.

R. Distributions and Currents on a Differentiable Manifold

Let M be an n-dimensional †differentiable manifold of class C^∞. Let $\{(U_\alpha,\psi_\alpha)\}$ be an †atlas. Then a **distribution** T on M is defined to be a collection of distributions T_α on $\psi_\alpha(U_\alpha)\subset\mathbf{R}^n$ such that T_β on $\psi_\beta(U_\alpha\cap U_\beta)$ is equal to $T_\alpha\circ(\psi_\alpha\circ\psi_\beta^{-1})$ for any α and β. Since the distributions obey the same transformation law as functions under coordinate transformation, this definition has an invariant meaning, and every locally integrable function on M is regarded as a distribution. More generally, a distribution †cross section of a †vector bundle of class C^∞ is defined in the same way. The most important is the case of the †p-fold exterior power of the †cotangent bundle. Its distribution cross sections, i.e., †exterior differential forms of order p with distribution coefficients, are called **currents** of degree p. If M is †oriented, then the space of currents of degree p on M is identified with the dual space of the locally convex space $\mathcal{D}^{(n-p)}(M)$ of all differential forms of degree $n-p$, of class C^∞ and with compact support in M. For example, if C is an $(n-p)$-dimensional †singular chain, a current T_C of degree p is defined by $T_C(\alpha)=\int_C\alpha$. If M is not †orientable, we have to consider either the double covering $\tilde{M}$ of M or the cross sections of the tensor product of the bundle of $(n-p)$-covectors and the orientation bundle [1, 3]. Currents were introduced by de Rham to prove his celebrated isomorphism theorem of the †de Rham cohomology groups defined by differential forms and the †singular cohomology groups defined by singular cochains (→ 105 Differentiable Manifolds).

S. Gel'fand-Shilov Generalized Functions

Let E and F be †locally convex spaces for which a continuous linear injection $i:E\to F$ with dense range is defined. Then the dual i': $F'\to E'$ is a continuous linear injection with weak*-dense range. Therefore if we shrink the function space E, we obtain a larger space E' of continuous linear functionals. Gel'fand and Shilov [4] introduced many function spaces E, called **fundamental spaces** or **test function spaces**, and defined corresponding spaces E' of **generalized functions**. Their motivations were in applications to the theory of partial differential equations, where Fourier transforms play an essential role and Schwartz's framework of tempered distributions is too restrictive. They considered a pair of function spaces E and $\tilde{E}$ such that the Fourier transformation $\mathcal{F}:E\to\tilde{E}$ is an isomorphism of locally convex spaces. Then the Fourier transformation defined to be the dual of $\mathcal{F}$ gives an isomorphism $\tilde{E}'\to E'$ of spaces of generalized functions. The function spaces E and $\tilde{E}$ are often spaces of entire functions, so that the Gel'fand-Shilov generalized functions are not necessarily localizable in $\mathbf{R}^n$. As typical examples of spaces E and $\tilde{E}$ we shall mention only spaces of type S in the next section.

L. Ehrenpreis's **analytically uniform spaces**

[12] are also a class of generalized function spaces introduced from a similar point of view.

T. Spaces of Type S

Let α, β, A, B, $\bar{A}$, and $\bar{B}$ be n-dimensional vectors. By $A > B$ we mean $A_j > B_j$ ($j = 1, \ldots, n$). We use notations $A^p = A_1^{p_1} \ldots A_n^{p_n}$, $p^{p\alpha} = p_1^{p_1\alpha_1} \ldots p_n^{p_n\alpha_n}$. (i) For $\alpha \geqslant 0$, $A > 0$ the space $S_{\alpha,A}$ consists of all C^∞-functions φ such that seminorms

$$\rho_{q,\bar{A}}(\varphi) = \sup_x \sup_p \frac{|x^p D^q \varphi(x)|}{\bar{A}^p p^{p\alpha}} \tag{10}$$

are finite for all $\bar{A} > A$ and q. For example, $S_{0,A}$ is the space of all functions of class C^∞ with support in $\{x \mid |x_j| \leqslant A_j, j = 1, \ldots, n\}$. (ii) For $\beta \geqslant 0$, $B > 0$ the space $S^{\beta,B}$ consists of all C^∞-functions φ such that the seminorms

$$\rho_{p,\bar{B}}(\varphi) = \sup_x \sup_q \frac{|x^p D^q \varphi(x)|}{\bar{B}^q q^{q\beta}} \tag{11}$$

are finite for all $\bar{B} > B$ and q. (iii) For α, $\beta \geqslant 0$ and A, $B > 0$, the space $S_{\alpha,A}^{\beta,B}$ consists of all C^∞-functions such that the seminorms

$$\rho_{\bar{A},\bar{B}}(\varphi) = \sup_x \sup_p \sup_q \frac{|x^p D^q \varphi(x)|}{\bar{A}^p \bar{B}^q p^{p\alpha} q^{q\beta}} \tag{12}$$

are finite for all $\bar{A} > A$ and $\bar{B} > B$. The topologies for these spaces are given in terms of the seminorms (10), (11), and (12), respectively. These spaces are generically called **spaces of type** S.

$\mathscr{F}(S_{\alpha,A}) = S^{\alpha,A}$ $\quad (\alpha > 0)$,

$\mathscr{F}(S^{\beta,B}) = S_{\beta,B}$ $\quad (\beta > 0)$,

$\mathscr{F}(S_{\alpha,A}^{\beta,B}) = S_{\beta,B}^{\alpha,A}$ $\quad (\alpha + \beta > 1)$.

$\mathscr{F}(S_{0,A}) = S^{0,A'}$,

$\mathscr{F}(S^{0,B}) = S_{0,B'}$,

$\mathscr{F}(S_{\alpha,A}^{\beta,B}) = S_{\beta,B'}^{\alpha,A'}$ $\quad (\alpha + \beta = 1)$,

where $A' = A\exp(1/A)$ and $B' = B\exp(1/B)$.

We set $S_\alpha = \bigcup S_{\alpha,A}$, $S^\beta = \bigcup S^{\beta,B}$, and $S_\alpha^\beta = \bigcup S_{\alpha,A}^{\beta,B}$, where A and B range over all positive n-dimensional vectors. $S_\alpha^\beta \neq \{0\}$ if and only if one of the following conditions is satisfied: (i) $\alpha + \beta \geqslant 1$, $\alpha > 0$, $\beta > 0$; (ii) $\alpha > 1$, $\beta = 0$; (iii) $\alpha = 0$, $\beta > 1$. In such a case the space $S_\alpha^{\beta\prime}$ of generalized functions contains the tempered distributions $\mathscr{S}'(\mathbf{R}^n)$.

U. Ultradistributions

The localizing property of distributions is proved mainly from the existence of †partitions of unity by functions in $\mathscr{D}$. Therefore if a test function space admits partitions of unity, the corresponding class of generalized functions forms a sheaf. Roumieu [5] took the space $\mathscr{D}_{\{M_p\}}(\Omega)$ of †ultradifferentiable functions of class $\{M_p\}$ with compact support. We can also use the space $\mathscr{D}_{(M_p)}(\Omega)$ of class (M_p). The corresponding generalized functions are called **ultradistributions of class** $\{M_p\}$ and **of class** (M_p), respectively. Here M_p is a sequence of positive numbers satisfying the logarithmic convexity $M_p^2 \leqslant M_{p-1} M_{p+1}$ and the Denjoy-Carleman condition $\sum M_p / M_{p+1} < \infty$. $\mathscr{D}_{\{M_p\}}(\Omega)$ (resp. $\mathscr{D}_{(M_p)}(\Omega)$) is the space of all functions $\varphi \in \mathscr{D}(\Omega)$ such that there are constants k and C (resp. for any $k > 0$ there is a C) for which we have $|D^\alpha \varphi(x)| \leqslant C k^{|\alpha|} M_{|\alpha|}$. In particular, the **Gevrey classes** $\{s\} = \{p!^s\}$ and $(s) = (p!^s)$ defined for $s > 1$ are important, and they appear often in theory of differential equations. Almost all results for distributions have been extended to ultradistributions under appropriate conditions on M_p, which are satisfied by Gevrey sequences $p!^s$ (→ [14] in particular). Closely connected are A. Beurling's test function spaces $\mathscr{D}_\omega(\Omega)$ (→ [13]), where ω is a function on $\mathbf{R}^n$ continuous at the origin, satisfying $0 = \omega(0) \leqslant \omega(\xi + \eta) \leqslant \omega(\xi) + \omega(\eta)$ for all ξ and η in $\mathbf{R}^n$ and such that $\int \omega(\xi)(1 + |\xi|)^{-n-1} d\xi < \infty$. $\mathscr{D}_\omega(\Omega)$ is defined to be the space of all $\varphi \in \mathscr{D}(\Omega)$ such that the Fourier transform $\hat{\varphi}$ satisfies $\int |\hat{\varphi}(\xi)| \exp(\lambda \omega(\xi)) d\xi < \infty$ for any $\lambda > 0$. If $\omega(\xi) = \log(1 + |\xi|)$, then $\mathscr{D}_\omega = \mathscr{D}$; if $\omega(\xi) = |\xi|^{1/s}$, then $\mathscr{D}_\omega = \mathscr{D}_{(s)}$. Let $\check{\omega}(\xi) = \omega(-\xi)$. The space $\mathscr{D}'_\omega(\Omega)$ of **Beurling's generalized distributions** is defined to be the strong dual of $\mathscr{D}_{\check{\omega}}(\Omega)$.

V. Hyperfunctions

In this and the following sections we mean by a cone $\Gamma \subset \mathbf{R}^n$ a convex open cone with vertex at 0. For two cones Γ, Δ, we write $\Delta \Subset \Gamma$ if $\Delta \cap \mathbf{S}^{n-1}$ is relatively compact in $\Gamma \cap \mathbf{S}^{n-1}$, where $\mathbf{S}^{n-1} = \{|x| = 1\}$. By a **wedge** we mean an open subset of $\mathbf{C}^n$ of the form $\Omega + i\Gamma$, where $\Omega \subset \mathbf{R}^n$ is an open subset and $\Gamma \subset \mathbf{R}^n$ is a cone. Γ is called the opening of the wedge and Ω its edge. By an **infinitesimal wedge** (0-wedge for short) or a **tuboid** of opening Γ and edge Ω, we mean a complex open set U such that $U \subset \Omega + i\Gamma$ and that for any $\Delta \Subset \Gamma$, U contains the part of $\Omega + i\Delta$ which is contained in some complex neighborhood of the edge Ω. The symbol $\Omega + i\Gamma 0$ will represent any one of such open sets, and $\mathscr{O}(\Omega + i\Gamma 0)$ (the inductive limit of) the totality of functions holomorphic on some of them.

A **hyperfunction** $f(x)$ on an open set $\Omega \subset \mathbf{R}^n$ is an equivalence class of formal expressions of the form

$$f(x) = \sum_{j=1}^{N} F_j(x + i\Gamma_j 0), \tag{13}$$

where $F_j(z) \in \mathscr{O}(\Omega + i\Gamma_j 0)$. $\{F_j(z)\}$ is called a set

of **defining functions** of $f(x)$. Here the equivalence relation is given as

$$F_j(x+i\Gamma_j0)+F_k(x+i\Gamma_k0)=(F_j+F_k)(x+i\Gamma_j\cap\Gamma_k0) \tag{14}$$

if $\Gamma_j\cap\Gamma_k\neq\varnothing$, that is, we can contract two terms into a single term as above and, conversely, can decompose, if possible, a term in the inverse way. These are considered to be modifications of the expression of the same hyperfunction. The totality of hyperfunctions on Ω is denoted by $\mathscr{B}(\Omega)$. It is a linear space by virtue of the linear structure naturally induced from that of holomorphic functions, combined with the above equivalence relation.

The symbol $F_j(x+i\Gamma_j0)$, which represents by itself a hyperfunction, is called the **boundary value** of $F_j(z)$ to the real axis. It is merely formal and does not imply any topological limit, though there is some justification for the terminology as will be seen in Section Z.

In the case of one variable, we have only two kinds of wedges $\Omega+i\mathbf{R}^{\pm}$, hence a hyperfunction can be expressed by two terms:

$$F_+(x+i0)-F_-(x-i0). \tag{15}$$

Some examples of hyperfunctions of one variable are $\delta(x)=-(2\pi i)^{-1}((x+i0)^{-1}-(x-i0)^{-1})$; $\mathbf{1}(x)=-(2\pi i)^{-1}(\log(-x-i0)-\log(-x+i0))$; $\mathrm{Pf}\,x^{-1}=((x+i0)^{-1}+(x-i0)^{-1})/2$.

W. Localization of Hyperfunctions

If $\Omega'\subset\Omega$ is an open subset, the restriction mapping $\mathscr{B}(\Omega)\to\mathscr{B}(\Omega')$ is induced from that for holomorphic functions. With this structure the correspondence $\Omega\mapsto\mathscr{B}(\Omega)$ becomes a †presheaf. It is in fact a sheaf, because it can be expressed by the terminology of relative (or local) cohomology as follows: Let $H^k_\Omega(\mathbf{C}^n,\mathscr{F})$ denote the kth **relative cohomology group** of the pair $(\mathbf{C}^n,\mathbf{C}^n\setminus\Omega)$ (also called the kth **local cohomology group** with support in Ω) with coefficients in a sheaf $\mathscr{F}$ on $\mathbf{C}^n$. (It is by definition the kth †derived functor of $\mathscr{F}\mapsto\Gamma_\Omega(\mathbf{C}^n,\mathscr{F})=$ the totality of sections of $\mathscr{F}$ defined on a neighborhood of Ω and with support in Ω and is calculated as the kth cohomology group of the †complex $\Gamma_\Omega(\mathbf{C}^n,\mathscr{L})$, where $\mathscr{L}$ denotes any **flabby resolution** (i.e., †resolution by †flabby sheaves) of $\mathscr{F}$.) Let $\mathscr{H}^k_{\mathbf{R}^n}(\mathscr{F})$ denote the kth **derived sheaf** of $\mathscr{F}$ to $\mathbf{R}^n$. (It is by definition the sheaf on $\mathbf{R}^n$ associated with the presheaf $\Omega\mapsto H^k_\Omega(\mathbf{C}^n,\mathscr{F})$.) Then the cohomological definition of the sheaf of hyperfunctions is $\mathscr{B}=\mathscr{H}^n_{\mathbf{R}^n}(\mathscr{O})$ (the orientation being neglected). A fundamental theorem by Sato says that $\mathbf{R}^n\subset\mathbf{C}^n$ is **purely** n**-codimensional** with respect to $\mathscr{O}$ (i.e., $\mathscr{H}^k_{\mathbf{R}^n}(\mathscr{O})=0$ for $k\neq n$), and moreover $H^k_\Omega(\mathbf{C}^n,\mathscr{O})=0$ for $k\neq n$ for any open set $\Omega\subset\mathbf{R}^n$. Then by the general theory it can be shown that the remaining $H^n_\Omega(\mathbf{C}^n,\mathscr{O})$ agrees with the section module $\mathscr{B}(\Omega)$ of $\mathscr{H}^n_{\mathbf{R}^n}(\mathscr{O})$. Since $H^n(V,\mathscr{O})=0$ for any open set $V\subset\mathbf{C}^n$ (B. Malgrange), it follows further that the sheaf $\mathscr{B}$ is flabby, i.e., its sections on any open set can always be extended to the whole space.

If U is a †Stein neighborhood of Ω, then $H^n_\Omega(\mathbf{C}^n,\mathscr{O})$ can be expressed using the covering cohomology as the quotient space

$$\mathscr{O}(U\#\Omega)\Big/\sum_{j=1}^{n}\mathscr{O}(U\#_j\Omega), \tag{16}$$

where

$$U\#\Omega=\{z\in U\mid \mathrm{pr}_k(z)\notin\mathrm{pr}_k(\Omega)\quad\text{for all }k\},$$

$$U\#_j\Omega=\{z\in U\mid \mathrm{pr}_k(z)\notin\mathrm{pr}_k(\Omega)\quad\text{for }k\neq j\}, \tag{17}$$

and pr_k is the projection from $\mathbf{C}^n$ to the kth coordinate. Then the isomorphism $H^n_\Omega(\mathbf{C}^n,\mathscr{O})=\mathscr{B}(\Omega)$ is induced by the correspondence

$$\mathscr{O}(U\#\Omega)\ni F(z)\mapsto[F(z)]$$
$$=\sum_\sigma \mathrm{sgn}\,\sigma F(x+i\Gamma_\sigma0)\in\mathscr{B}(\Omega), \tag{18}$$

where Γ_σ is the σ-orthant $\{y\in\mathbf{R}^n\mid\sigma_jy_j>0, j=1,\dots,n\}$ and $\mathrm{sgn}\,\sigma=\sigma_1\dots\sigma_n$. $F(z)$ is called a defining function of the corresponding hyperfunction. For one variable, any complex neighborhood $U\supset\Omega$ is Stein, and the above isomorphism reads $\mathscr{O}(U\setminus\Omega)/\mathscr{O}(U)=\mathscr{B}(\Omega)$, from which the naturality of the sign in (15) follows.

Thus the notion of support is also legitimate for hyperfunctions. The sheaf of hyperfunctions $\mathscr{B}$ does not admit partitions of unity as for $\mathscr{D}'$. It is, however, flabby. Consequently, given a decomposition of a closed set into locally finite closed subsets $E=\bigcup_{\lambda\in\Lambda}E_\lambda$ and a hyperfunction f with support in E, we can always find hyperfunctions f_λ with support in E_λ such that $f=\sum_{\lambda\in\Lambda}f_\lambda$. For distributions this property holds only under some regularity assumption for the decomposition.

There are several practical criteria to determine whether or not a hyperfunction is zero in some open set Ω. These are called the **edge of the wedge theorem**. A hyperfunction $F(x+i\Gamma0)$ with single expression is zero if and only if $F(z)$ itself is zero. $F_1(x+i\Gamma_10)=F_2(x+i\Gamma_20)$ if and only if they stick together to a function in $\mathscr{O}(\Omega+i(\Gamma_1+\Gamma_2)0)$ (Epstein type). (Note that $\Gamma_1+\Gamma_2$ is equal to the convex hull of $\Gamma_1\cup\Gamma_2$, e.g., $\Gamma+(-\Gamma)=\mathbf{R}^n$ (Bogolyubov type).) $\sum_{j=1}^N F_j(x+i\Gamma_j0)=0$ if and only if there exist $G_{jk}(z)\in\mathscr{O}(\Omega+i(\Gamma_j+\Gamma_k)0)$, $j,k=1,\dots,N$, such that $G_{jk}(z)=G_{kj}(z)$ and $F_j=\sum_{k=1}^N G_{jk}$, $j=1,\dots,N$ (A. Martineau [19]). These are interpretations of cohomology in terms of coverings and have global variants concerning the envelope of holomorphy.

The real analytic functions $\varphi\in\mathscr{A}(\Omega)$ on Ω are naturally included in $\mathscr{B}(\Omega)$ via the ex-

pression $\varphi(x+i\Gamma 0)$ for any Γ or with $\Gamma=\mathbf{R}^n$. They form a subsheaf. Let f be a hyperfunction on Ω. The complement in Ω of the largest open subset $\Omega'\subset\Omega$, where $f(x)$ is real analytic, is called the **singular support** of $f(x)$ and is denoted by $\operatorname{sing\,supp} f$. If Ω is bounded and $\operatorname{sing\,supp} f\subset K$, then we can choose an expression of the form (13), where $F_j(z)$ can be continued analytically to $\Omega\setminus K$. If $\operatorname{supp} f\subset K$, then these $F_j(z)$ satisfy $\sum F_j(x)=0$ on $\Omega\setminus K$ in the usual sense.

X. Operations on Hyperfunctions

The **derivatives** of a hyperfunction $f(x)$ with the expression (13) are defined through the defining functions as $D^p f(x)=\sum(D_z^p F_j)(x+i\Gamma_j 0)$. The **product** by a real analytic function $\psi(x)$ is defined by $\psi(x)f(x)=\sum(\psi F_j)(x+i\Gamma_j 0)$. Combining these, we have the operation of a linear partial differential operator with real analytic coefficients $P(x,D)$ on hyperfunctions. It is a †sheaf homomorphism.

Let $f\in\mathscr{B}(\Omega)$ and $D\subset\Omega$ be a compact set with piecewise smooth boundary. If $\operatorname{sing\,supp} f\cap\partial D=\varnothing$, then by means of the special expression (13) mentioned at the end of Section W the **definite integral** is defined as

$$\int_D f(x)\,dx=\sum_{j=1}^{N}\int_{D_j}F_j(z)\,dz,$$

where D_j is a path deformed from D in such a way that $\partial D_j=\partial D$ and $F_j(z)$ is holomorphic on D_j. The result is independent of the choice of deformations or of the boundary value expression employed.

If $\operatorname{supp} f\Subset D$, the result is also independent of D; hence it can be written as $\int_{\mathbf{R}^n}f(x)\,dx$. If $f(x,t)$ is a hyperfunction of the two groups of variables $(x,t)\in\Omega\times V$ such that $\operatorname{sing\,supp} f\cap\partial D\times V=\varnothing$, then the integral $\int_D f(x,t)\,dx\in\mathscr{B}_t(V)$ is defined by the same method. It commutes with differentiation or integration with respect to t.

If $f(x)=\sum F_j(x+i\Gamma_j 0)$, $g(x)=\sum G_k(x+i\Delta_k 0)$, then we can define the **product** by $f(x)g(x)=\sum(F_jG_k)(x+i\Gamma_j\cap\Delta_k 0)$ under the assumption that $\Gamma_j\cap\Delta_k\neq\varnothing$ for every pair j,k. Especially, the product $f(x)g(t)$ is always legitimate for two hyperfunctions depending on different groups of variables. (It can also be interpreted as the tensor product.)

A real analytic coordinate transformation $x=\Phi(\tilde{x}):\tilde{\Omega}\to\Omega$ extends naturally to a holomorphic coordinate transformation $z=\Phi(\tilde{z})$ on a complex neighborhood. A 0-wedge $\Omega+i\Gamma 0$ is transformed by Φ^{-1} to a twisted wedge containing, for each $x\in\Omega$ and $\Delta\Subset\Gamma$, a 0-wedge $\tilde{\Omega}_{\tilde{x}}+i(D\Phi^{-1})_x\Delta 0$ with some (real) neighborhood $\tilde{\Omega}_{\tilde{x}}$ of $\tilde{x}=\Phi^{-1}(x)$, where $(D\Phi^{-1})_x$ is the derivative or the Jacobian matrix of the mapping Φ^{-1} at x. Thus the pullback $\Phi^*f(\tilde{x})=f(\Phi(\tilde{x}))\in\mathscr{B}(\tilde{\Omega})$ of $f(x)\in\mathscr{B}(\Omega)$ by the transformation $x=\Phi(\tilde{x})$ can be defined by substituting $z=\Phi(\tilde{z})$ into the defining functions. It is consistent with the definition of the coordinate transformation for real analytic functions. Also, the law of the change of variables for the definite integral is the same as that for real analytic functions. This can be extended to the general operation of **substitution** as mentioned in Section Q for distributions or even to much more general operations (→ Section CC).

The **convolution** $(f*g)(x)=\int_{\mathbf{R}^n}f(x-t)g(t)\,dt$ is defined under the same assumption on support as in the case of distributions. It can be defined either literally as the composition of the above operations, or directly by way of the defining functions: For example, if g has compact support $\Subset D$ and $\sum G_k(x+i\Delta_k 0)$ is an expression as mentioned at the end of Section W, then by choosing a suitable deformation D_k of D, we have

$$(f*g)(x)=\sum_{j,k}\left[\int_{D_k}F_j(z-\tau)G_k(\tau)\,d\tau\right]_{z\to x+i(\Gamma_j+\Delta_k)0}.$$

Y. Hyperfunctions and Analytic Functionals

The totality $\mathscr{B}[K]$ of hyperfunctions with support in a compact set K becomes a nuclear Fréchet space. It is the dual of the nuclear (DF)-space $\mathscr{A}(K)$ of real analytic functions defined on a neighborhood of K. Thus $\mathscr{B}[K]$ is the space of analytic functionals with †porter in the real compact set K. The duality is given by the definite integral $\langle f,\varphi\rangle=\int_{\mathbf{R}^n}f(x)\varphi(x)\,dx$ for $f\in\mathscr{B}[K]$ and $\varphi\in\mathscr{A}(K)$. A sequence $\{f_k(x)\}$ in $\mathscr{B}[K]$ converges if and only if it admits an expression (13) for a common fixed set of 0-wedges $\Omega+i\Gamma_j 0$, $j=1,\dots,N$, such that the defining functions $F_{kj}\in\mathscr{O}(\Omega+i\Gamma_j 0)$ are also holomorphic on a fixed complex neighborhood U of $\Omega\setminus K$ and converges locally uniformly in $\Omega+i\Gamma_j 0\cup U$, where Ω is a (real) neighborhood of K.

$\mathscr{B}[K]$ is isomorphic to $H_K^n(\mathbf{C}^n,\mathscr{O})$, and the above duality is a special case of the **Martineau-Harvey duality** $H_K^n(\mathbf{C}^n,\mathscr{O})=\mathscr{O}(K)'$ for a Stein compact set $K\subset\mathbf{C}^n$. In the 1-dimensional case this is due to G. Köthe. If $K=[a_1,b_1]\times\ldots\times[a_n,b_n]$, then $H_K^n(\mathbf{C}^n,\mathscr{O})$ can be represented, including the topology, by the quotient space

$$\mathscr{O}(U\#K)\Big/\sum_{j=1}^{n}\mathscr{O}(U\#_jK),$$

where U is a Stein neighborhood of K and $U\#K$, $U\#_jK$ are defined in the same way as (17). If we choose $U=U_1\times\ldots\times U_n$, then by

way of a defining function $F(z) \in \mathcal{O}(U \# K)$ for $f \in \mathscr{B}[K]$ the inner product is given by the contour integral

$$\int_{\mathbf{R}^n} f(x)\varphi(x)\,dx = (-1)^n \oint_{\gamma_1} \ldots \oint_{\gamma_n} F(z)\varphi(z)\,dz_1 \ldots dz_n,$$

where $\gamma_j \subset U_j$ is a closed path surrounding $[a_j, b_j]$ once in the positive sense. Similar integral formulas are known for some special K of various types.

Starting from analytic functionals we can reconstruct the sheaf of hyperfunctions. For example, we can put $\mathscr{B}(\Omega)$ = the totality of locally finite sums of analytic functionals with porter in Ω modulo the rearrangement of supports by decomposition (Martineau [17]). If Ω is bounded, we can also put $\mathscr{B}(\Omega) = \mathscr{B}[\bar{\Omega}]/\mathscr{B}[\partial\Omega]$ (Schapira [18]). The proof of localizability and/or flabbiness is based on the decomposability of support (which is the dual of the exact sequence $0 \to \mathscr{A}(K \cup L) \to \mathscr{A}(K) \oplus \mathscr{A}(L) \to \mathscr{A}(K \cap L) \to 0$) and the denseness of $\mathscr{B}[K] \subset \mathscr{B}[L]$ (which is the dual of the unique continuation property $0 \to \mathscr{A}(L) \to \mathscr{A}(K)$) for a pair $K \subset L$ with the same family of connected components. Note that in no way is the topology of $\mathscr{B}[K]$ localizable, or equivalently, $\mathscr{B}(\Omega)$ does not admit a reasonable topology.

Z. Embedding of Distributions

As the dual of the natural mapping $\mathscr{A}(K) \to \mathscr{E}(K)$ we have the topological embedding $\mathscr{E}'(K) \hookrightarrow \mathscr{A}(K)' = \mathscr{B}[K]$. This embedding conserves the support and hence gives rise to an embedding of sheaf $\mathscr{D}' \hookrightarrow \mathscr{B}$ (R. Harvey).

For a distribution T with compact support a set of its defining functions as a hyperfunction is given by $F_\sigma(z) = T_x(W(z-x, \Gamma_\sigma))$, σ being the multisignature, where $W(z, \Gamma_\sigma) = \int_{S^{n-1} \cap \Gamma_\sigma} W(z, \omega)\,d\omega$ and $W(z, \omega)$ is the component of a Radon decomposition of $\delta(x)$ (→ Section CC). If $\operatorname{supp} T \subset K = [a_1, b_1] \times \ldots \times [a_n, b_n]$, then as a hyperfunction it is represented by the following element of $\mathcal{O}(\mathbf{C}^n \# K)$:

$$F(z) = T_x\left(\frac{1}{(2\pi i)^n (x_1 - z_1) \ldots (x_n - z_n)}\right). \qquad (19)$$

It is in fact in $\mathcal{O}((\mathbf{P}^1)^n \# K)$ and vanishes at infinity, where $\mathbf{P}^1 = \mathbf{C}^1 \cup \{\infty\}$ is the Riemann sphere. These formulas are valid also for hyperfunctions, and they give defining functions of some canonical types. Especially, the one given by (19) is called the **standard defining function** and is characterized by the foregoing properties.

For the above-mentioned defining functions for a distribution T, we have the following convergence in the sense of distributions on $\mathbf{R}^n$:

$$T_x = \sum_\sigma \operatorname{sgn}\sigma \lim_{\varepsilon \downarrow 0} F_\sigma(x + i\varepsilon y_\sigma) \quad \text{for} \quad y_\sigma \in \Gamma_\sigma.$$

More generally if the limit

$$\sum \lim_{\varepsilon \downarrow 0} F_j(x + i\varepsilon y_j) \quad \text{for} \quad y_j \in \Gamma_j \qquad (20)$$

exists in $\mathscr{D}'(\Omega)$, then the distribution defined as this limit admits $\{F_j(z)\}$ as a set of defining functions when it is considered as a hyperfunction. However, for an arbitrary set of defining functions for a distribution, (20) does not necessarily converge in $\mathscr{D}'(\Omega)$. (This is because we can add terms with any bad behavior which cancel each other formally.) If a distribution admist the boundary value expression with only one term, or if the dimension is one, then the convergence of (20) in $\mathscr{D}'(\Omega)$ necessarily holds.

The convergence of (20) in $\mathscr{D}'(\Omega)$ is equivalent to the locally uniform estimate for the defining functions of the type $F_j(z) = O(|y|^{-M})$ for some $M > 0$. These assertions can be generalized to ultradistributions. For $\mathscr{D}'_{\{M_p\}}(\Omega)$ (resp. $\mathscr{D}'_{(M_p)}(\Omega)$) the last growth condition for the defining functions reads as follows: $F_j(z) = O(\exp M^*(L/|y|))$ for some L (resp. every $L > 0$), where $M^*(\rho) = \sup_p \log(\rho^p p! M_0 / M_p)$; especially for $M_p = p!^s$ we have $M^*(\rho) \sim \rho^{1/(s-1)}$.

AA. Hyperfunctions on a Real Analytic Manifold

Sticking the hyperfunctions on coordinate patches by the transformation law mentioned in Section X, we can define the sheaf of hyperfunctions on a real analytic manifold. More generally, for a real analytic vector bundle over a real analytic manifold, we can consider the sheaf of its hyperfunction local cross sections, which is also flabby. Thus, especially on a real analytic manifold M, we can obtain a concrete flabby resolution of the constant sheaf $\mathbf{C}_M$ of length $\dim M$ by the sheaves of differential forms with hyperfunction coefficients: $0 \to \mathbf{C}_M \to \mathscr{B}_M^{(0)} \xrightarrow{d} \mathscr{B}_M^{(1)} \xrightarrow{d} \ldots \to \mathscr{B}_M^{(\dim M)} \to 0$. With this sequence we can calculate the relative cohomology groups of open pairs with coefficients in $\mathbf{C}$ by an analytic method. This is an extension of the de Rham theory for distributions [16].

If M is a compact manifold equipped with a nowhere vanishing real analytic density globally defined on M, then we have the topological duality $\mathscr{B}(M) = \mathscr{A}(M)'$. The inner product is given by the definite integral with respect to the density.

The Fourier series is an example of hyperfunctions on real analytic manifolds. The series $\sum c_p \exp(2\pi i p x)$ converges in $\mathscr{B}(\mathbf{T}^n)$ and defines a periodic hyperfunction if and only if c_p is of **infra-exponential growth**, i.e., $c_p = O(e^{\varepsilon|p|})$ for any $\varepsilon > 0$. $\mathbf{T}^n$ has the global complex neighborhood $(\mathbf{P}^1)^n$, and $f(x) \in \mathscr{B}(\mathbf{T}^n)$ has the corresponding boundary value expression

$$F_\sigma(z) \in \mathscr{O}(\{|z_1|^{\sigma_1} < 1, \ldots, |z_n|^{\sigma_n} < 1\}),$$

which represents the terms in the Fourier series such that $\sigma_j p_j > 0$.

BB. Fourier Hyperfunctions

In place of $\mathscr{S}$ we take as the basic space the space $\mathscr{P}_*$ of **exponentially decreasing real analytic functions** in the sense of M. Sato [15]: $f(x) \in \mathscr{P}_*$ if and only if there exist $\delta > 0$ and $\varepsilon > 0$ such that for any $\delta' < \delta$ and $\varepsilon' < \varepsilon$, $f(x)$ extends holomorphically to the neighborhood $\{|\mathrm{Im}\, z| < \delta'\}$ of the real axis and has order $O(e^{-\varepsilon'|\mathrm{Re}\, z|})$ there. $\mathscr{P}_*$ is endowed with the structure of nuclear (DF)-space via the inductive limit for $\delta > 0$, $\varepsilon > 0$. The classical Fourier transform $\mathscr{F}$ acts isomorphically on $\mathscr{P}_*$. (In fact, δ and ε change their roles under $\mathscr{F}$.) The strong dual of $\mathscr{P}_*$ is called the space of **Fourier hyperfunctions** and is denoted by $\mathscr{Q}$. It is a nuclear Fréchet space. It contains $\mathscr{S}'$ as a dense subspace in view of the continuous and dense inclusion $\mathscr{P}_* \hookrightarrow \mathscr{S}$. It also contains classical locally integrable functions of **infra-exponential growth**, i.e., of order $e^{\varepsilon|x|}$ for any $\varepsilon > 0$. Thus by the duality we obtain a wider extension of Fourier transformation on $\mathscr{Q}$.

In the following a 0-wedge of the form $\mathbf{R}^n + i\Gamma 0$ will be called a 0-wedge of the form $\mathbf{D}^n + i\Gamma 0$ at the same time if it is a tubular domain (i.e., with fixed imaginary part $\Gamma 0$). Then an element $f(x) \in \mathscr{Q}$ can be expressed in the form (13), where each $F_j(z)$ is holomorphic in a 0-wedge $\mathbf{D}^n + i\Gamma_j 0$ and is of infra-exponential growth there locally uniformly in $\mathrm{Im}\, z$. The inner product of such $f(x)$ with $\varphi \in \mathscr{P}_*$ is given by the definite integral

$$\int f(x)\varphi(x)\,dx = \sum_{j=1}^{N} \int_{\mathrm{Im}\, z = y_j} F_j(z)\varphi(z)\,dz,$$

where the $y_j \in \Gamma_j 0$ are fixed. Given a cone Δ we define its **dual cone** by $\Delta^\circ = \{\eta \in \mathbf{R}^n \mid \langle \eta, y \rangle \geq 0$ for all $y \in \Delta\}$. If $F_j(z)$ are all of exponential decrease in $\mathrm{Re}\, z$ locally uniformly with respect to $\mathrm{Im}\, z \in \Gamma_j 0$ and $\mathrm{Re}\, z / |\mathrm{Re}\, z| \notin \Delta^\circ$, then the definite integral

$$G(\zeta) = (\sqrt{2\pi})^{-n} \int e^{-ix\zeta} f(x)\,dx$$
$$= \sum_{j=1}^{N} (\sqrt{2\pi})^{-n} \int_{\mathrm{Im}\, z = y_j} e^{-iz\zeta} F_j(z)\,dz \qquad (21)$$

converges locally uniformly for ζ in some 0-wedge $\mathbf{D}^n - i\Delta 0$, and defines there a holomorphic function of infra-exponential growth. Thus we obtain a Fourier hyperfunction $G(\xi - i\Delta 0)$ that agrees with $\mathscr{F}f$ calculated by the duality. For a general $f(x) \in \mathscr{Q}$, the Fourier transform in the manner of Sato is calculated as follows: First we decompose $f(x)$ into the sum $\sum f_k(x)$ for which the defining functions of $f_k(x)$ decrease exponentially outside Δ_k°. Then we calculate $G_k(\zeta)$ by (21) and put $\mathscr{F}f = \sum G_k(\xi - i\Delta_k 0)$. An example of such decomposition is given by multiplication by $\chi_\sigma(x) = \prod_{j=1}^n 1/(1 + \exp \sigma_j x_j)$, which decreases exponentially outside $\overline{\Gamma}_\sigma = \{\sigma_j x_j \geq 0\}$.

The relation between $\mathscr{Q}$ and $\mathscr{B}$ is more complicated than the relation $\mathscr{S}' \hookrightarrow \mathscr{D}'(\mathbf{R}^n)$. The growth condition for $\mathscr{Q}$ is interpreted as a condition concerning germs at infinity. Thus $\mathscr{Q}$ can be considered to be a sheaf on the directional compactification $\mathbf{D}^n = \mathbf{R}^n \cup \mathbf{S}_\infty^{n-1}$ such that $\mathscr{Q}|_{\mathbf{R}^n} = \mathscr{B}$. Just as the sheaf $\mathscr{B}$ is obtained from $\mathscr{O}$, the sheaf $\mathscr{Q}$ is obtained as the nth derived sheaf $\mathscr{H}^n_{\mathbf{D}^n}(\tilde{\mathscr{O}})$ from the sheaf $\tilde{\mathscr{O}}$ on $\mathbf{D}^n + i\mathbf{R}^n$ consisting of germs of holomorphic functions of infra-exponential growth with respect to $\mathrm{Re}\, z$. We have $H^k_\Omega(\mathbf{D}^n + i\mathbf{R}^n, \tilde{\mathscr{O}}) = 0$ for $k \neq n$ for any open set $\Omega \subset \mathbf{D}^n$. Especially, $\mathscr{Q}$ is flabby, and the decomposition of support is available to calculate the Fourier transform. The symbol $\mathscr{Q}$ employed at the beginning to express the global Fourier hyperfunctions corresponds to $\mathscr{Q}(\mathbf{D}^n)$, and $\mathscr{Q}(\mathbf{R}^n) = \mathscr{B}(\mathbf{R}^n)$ by definition. The canonical restriction mapping $\mathscr{Q}(\mathbf{D}^n) \to \mathscr{B}(\mathbf{R}^n)$ is surjective but not injective.

As for tempered distributions $\mathscr{S}'$, we can introduce various subclasses of Fourier hyperfunctions, e.g., **exponentially decreasing Fourier hyperfunctions** $\bigcup_{\delta > 0} \exp(-\delta\sqrt{x^2+1})\mathscr{Q}$, real analytic functions of infra-exponential growth $\tilde{\mathscr{O}}(\mathbf{D}^n)$, etc. We can also consider operations such as convolution and multiplication between adequate pairs, and apply differential operators with suitable coefficients. Concerning these we can avail ourselves of the same formulas as given in Section O.

A hyperfunction with compact support is naturally considered as a Fourier hyperfunction, and its Fourier transform agrees with the inner product $\langle f(x), (\sqrt{2\pi})^{-n} \exp(-ix\xi) \rangle$, which gives an entire function of exponential type.

Paley-Wiener theorem. An entire function $f(\zeta)$ is the Fourier transform of a hyperfunction with support in a compact convex set K in $\mathbf{R}^n$ if and only if it satisfies condition (i) of Section O.

The theory of Fourier hyperfunctions described above is mainly due to Sato and T. Kawai [20]. They are not the largest class of generalized functions stable under the Fourier

transformation. Following a suggestion of Sato, S. Nagamachi and N. Mugibayashi, Y. Saburi, and Y. Ito have extended them to the **modified Fourier hyperfunctions** in which the radial compactification $\mathbf{D}^{2n}$ of $\mathbf{C}^n$ is employed instead of the horizontal compactification $\mathbf{D}^n+i\mathbf{R}^n$ in the above theory. If we discard the localizing property, they can be extended further to the **Fourier ultrahyperfunctions** or **ultradistributions** of J. Sebastiaõ e Silva (1958), M. Hasumi (1961), and M. Morimoto (1973).

CC. Micro-Analyticity of Hyperfunctions

The boundary value expression (13) for a hyperfunction $f(x)$ can be interpreted conversely as the description of the state of analytic continuation of $f(x)$ to the complex domain. Thus we say that f is **micro-analytic** at (x_0, ξ_0) if on a neighborhood of x_0 it admits the analytic continuation into the half-space $\langle \operatorname{Im} z, \xi_0\rangle<0$ in the sense that it admits an expression (13) satisfying $\Gamma_j \cap \{\langle \operatorname{Im} z, \xi_0\rangle<0\} \neq \varnothing$ for every j. The set of points $(x_0, \xi_0) \in \Omega \times S^{n-1}$, where $f \in \mathscr{B}(\Omega)$ is not micro-analytic, is called the **singularity spectrum** or **singular spectrum** of f and is denoted by S.S.f. We have by definition $\mathrm{S.S.}F(x+i\Gamma 0) \subset \Omega \times (\Gamma^\circ \cap S^{n-1})$. Micro-analyticity can be characterized by the Fourier transformation as follows: f is micro-analytic at (x_0, ξ_0) if and only if there exists a Fourier hyperfunction $g(x)$ such that the Fourier transform $\hat{g}(\xi)$ decreases exponentially on a †conical neighborhood of ξ_0 and that $f-g$ is real analytic on a neighborhood of x_0. A hyperfunction $f(x)$ is real analytic in a neighborhood of x_0 if and only if it is micro-analytic at (x_0, ξ) for any $\xi \in S^{n-1}$.

With this notion we can clarify the operations on hyperfunctions. (In the following S^{n-1} is identified with $(\mathbf{R}^n \setminus \{0\})/\mathbf{R}^+$ and $+$ denotes the sum in the latter.) We have

$$\mathrm{S.S.}(fg) \subset \{(x, \lambda\xi+(1-\lambda)\eta) \mid (x,\xi) \in \mathrm{S.S.}f,\ (x,\eta) \in \mathrm{S.S.}g, 0 \leqslant \lambda \leqslant 1\} \cup \mathrm{S.S.}f \cup \mathrm{S.S.}g,$$

$$\mathrm{S.S.}f(\Phi(\tilde{x})) \subset \{(\tilde{x}, {}^t D\Phi\xi) \mid (\Phi(\tilde{x}), \xi) \in \mathrm{S.S.}f\},$$

and these operations are legitimate if and only if 0 does not appear in the direction component of the right-hand side. We also have

$$\mathrm{S.S.}\int f(x,t)\,dx \subset \{(x,\xi) \mid (x,t,\xi,0) \in \mathrm{S.S.}f \text{ for a } t\},$$

$$\mathrm{S.S.}(f*g) \subset \{(x+y,\xi) \mid (x,\xi) \in \mathrm{S.S.}f, (y,\xi) \in \mathrm{S.S.}g\}$$

under suitable conditions for support.

The classical **decomposition formula of Radon** (or the **plane wave decompostion** of δ)

$$\delta(x)=\frac{(n-1)!}{(-2\pi i)^n}\int_{S^{n-1}} \frac{d\omega}{(x\omega+i0)^n} \tag{22}$$

can be interpreted as the decomposition of δ into hyperfunctions with S.S. in the single direction ω. By the convolution, this formula supplies similar decomposition for a general hyperfunction (called the singular spectrum decomposition). (22) can be generalized to

$$\delta(x)=\frac{(n-1)!}{(-2\pi i)^n}\int_{S^{n-1}} \frac{\det(\operatorname{grad}_\omega \Psi(x,\omega))}{(\Phi(x,\omega)+i0)^n}\,d\omega, \tag{23}$$

where the twisted phase $\Phi(x,\omega)$ satisfies (i) $\Phi(x,\omega)$ is a real analytic function of positive type (i.e., $\operatorname{Re}\Phi(x,\omega)=0$ implies $\operatorname{Im}\Phi(x,\omega) \geqslant 0$) and (ii) $\Phi(x,\omega)$ is homogeneous of order 1 in ω and $\Phi(0,\omega)=0$, $\operatorname{grad}_x \Phi(0,\omega)=\omega$; and the vector $\Psi(x,\omega)$ is such that $\langle \Psi(x,\omega), x\rangle = \Phi(x,\omega)$. If $\Phi(x,\omega)$ further satisfies $\Phi(x,\omega) \neq 0$ for $x \neq 0$, then the component becomes a hyperfunction (even a distribution) of x whose S.S. is precisely one point $(0,\omega)$, and this fact is useful in theoretical applications [7, 21] (→ 274 Microlocal Analysis).

DD. Structure Theorems of Hyperfunctions

A hyperfunction whose support is concentrated at the origin is expressed as the infinite derivative $J(D)\delta(x)=\sum a_p D^p \delta(x)$ of the Dirac measure, with the coefficients satisfying $\lim(|a_p|p!)^{1/|p|}=0$. Such an operator $J(D)$ is called a **local operator** with constant coefficients and acts on $\mathscr{B}$ as a sheaf homomorphism. Its total symbol $J(\zeta)$ is an entire function of infra-exponential growth or of type $[1,0]$. By way of such an operator, a general (Fourier) hyperfunction can be written in the form $J(D)g(x)$ with a continuous function on $\mathbf{R}^n$ (of infra-exponential growth).

If $0 \in \operatorname{supp} f \subset \{\langle v, x\rangle \geqslant 0\}$, then $\mathrm{SS}\, f \ni (0, \pm v)$ (**Holmgren type theorem** of Kashiwara and Kawai), and furthermore, the direction component of S.S.f at 0 has the form of $\{\pm v\} \cup \rho^{-1}(E)$ with some $E \subset S^{n-2}$, where $\rho: S^{n-1} \setminus \{\pm v\} \to S^{n-2}$ is the projection to the equator (the **watermelon-slicing theorem** of Morimoto, Kashiwara, and K. Kataoka). E is called the reduced S.S. of f at 0. These theorems have many applications in the theory of linear partial differential equations and also in physics.

A hyperfunction $f(x)$ with support in the hyperplane $x_n=0$ has several further remarkable properties in x_n. It admits a formal expansion of the form $\sum_{k=0}^{\infty} f_k(x')\delta^{(k)}(x_n)$, where $x'=(x_1, \ldots, x_{n-1})$ and $f_k(x')=\int_{-\infty}^{\infty} f(x)x_n^k/k!\,dx_n$. The sum converges in the sense of the topology if f has compact support. It reduces locally to a finite sum if f is a distribution, and the kth term represents the k-ple layer distribution of mass, electric charge, etc. For a general f the coefficients $\{f_k(x')\}$ do not neces-

sarily determine f, though they are determined by f.

EE. Complex Powers of Polynomials

Among examples of special generalized functions the most important are those of the form f_+^λ, where $f_+ = \max\{f(x), 0\}$ (or more generally it can be replaced by zero on some connected components of $\{f(x)>0\}$), and $\lambda \in \mathbf{C}$ denotes a holomorphic parameter. (The discussion is the same for $f_- = \max\{-f(x), 0\}$.) The simplest example, x_+^λ, is defined as the analytic continuation of the locally integrable function x_+^λ for $\mathrm{Re}\,\lambda > -1$ by repeated use of the formula $x_+^\lambda = (\lambda+1)^{-1} D_x x_+^{\lambda+1}$, and becomes meromorphic in λ with simple poles at $\lambda = -1, -2, \ldots$. As a hyperfunction we have $x_+^\lambda = \{(-x+i0)^\lambda - (-x-i0)^\lambda\}/2i\sin\pi\lambda$. At a negative integer $\lambda = -n$, x_+^λ has residue $(-1)^{n-1}\delta^{(n-1)}(x)/(n-1)!$ and finite part $[-(2\pi i)^{-1} z^{-n}\log(-z)]$. The latter is often denoted by x_+^{-n}. In general, for a germ of a real-valued real analytic function $f(x)$ we can find a differential operator $P(\lambda, x, D_x)$ with polynomial coefficients in λ and a monic polynomial $b(\lambda)$ of minimum degree such that

$$P(\lambda, x, D_x) f_+^{\lambda+1} = b(\lambda) f_+^\lambda \tag{24}$$

(Sato, I. N. Bernshteĭn, Kashiwara, J.-E. Björk [22]). This formula gives the analytic continuation of f_+^λ just as for x_+^λ. The polynomial $b(\lambda)$ is called the **b-function** or the **Sato-Bernshteĭn polynomial** and contains valuable information regarding the singularity of f. It has only negative rational roots (Kashiwara).

We have $f \cdot f_\pm^\lambda = f_\pm^{\lambda+1}$; hence $f_+^{-1} - f_-^{-1}$ (suitably interpreted as above) gives a solution of the division problem $u \cdot f = 1$. Thus if f is a polynomial, its inverse Fourier transform gives a tempered fundamental solution of $f(-iD)$. Furthermore, when f is the relative invariant of a †prehomogeneous vector space, we can calculate $b(\lambda)$ explicitly by way of the holonomy diagram. Also, the Fourier transform of f_+^λ can be calculated explicitly by way of the real holonomy diagram as a linear combination of the corresponding objects for the dual prehomogeneous vector space with coefficients similar to the †Maslov index. The simplest example is

$$\begin{aligned}\mathscr{F} x_\pm^\lambda &= (\sqrt{2\pi})^{-1}(\mp i) e^{-\pi i\lambda/2}\Gamma(\lambda+1)(\xi \mp i0)^{-\lambda-1}\\ &= (\sqrt{2\pi})^{-1}(\mp i)\Gamma(\lambda+1)\{e^{\mp\pi i\lambda/2}\xi_-^{-\lambda-1}\\ &\qquad - e^{\pm\pi i\lambda/2}\xi_-^{-\lambda-1}\}.\end{aligned}$$

Among practical examples are the following classical formulas: Let $P(x)$ be a nondegenerate real quadratic form and $Q(\xi)$ its †dual form, and let q denote the number of negative eigenvalues of $P(x)$. Then

$$\begin{aligned}\mathscr{F} P_+^\lambda = &-(\sqrt{2\pi})^{-n} 2^{2\lambda+n}\pi^{n/2-1}\Gamma(\lambda+1)\Gamma(\lambda+n/2)\\ &\times |\det P|^{-1/2}\{\sin\pi(\lambda+q/2) Q_+^{-\lambda-n/2}\\ &-\sin\pi((n-q)/2) Q_-^{-\lambda-n/2}\}.\end{aligned}$$

Here the arguments in the Γ-factor $(\lambda+1)(\lambda+n/2)$ give the b-function of $P(x)$. If $q = n-1$, we further have, letting $P_{+\pm}^\lambda = P_+^\lambda 1(\pm\langle x, v\rangle)$ for an eigenvector v corresponding to the unique positive eigenvalue,

$$\begin{aligned}\mathscr{F} P_{+\pm}^\lambda = &(2\pi)^{-n/2} 2^{2\lambda+n-1}\pi^{n/2-1}\Gamma(\lambda+1)\\ &\times\Gamma(\lambda+n/2)|\det P|^{-1/2}\{e^{\mp\pi i(\lambda+n/2)} Q_{++}^{-\lambda-n/2}\\ &+ e^{\pm\pi i(\lambda+n/2)} Q_{+-}^{-\lambda-n/2} + Q_-^{-\lambda-n/2}\}.\end{aligned}$$

From these formulas (taking the finite part if necessary) we obtain the fundamental solution of the wave equation, the Laplacian, and their iterations. These are exactly the distributions introduced by Hadamard, M. Riesz, and others, as mentioned in Section A (→ Appendix A, Table 15.V).

References

[1] L. Schwartz, Théorie des distributions, Hermann, revised edition, 1966.
[2] L. Hörmander, Linear partial differential operators, Springer, 1963.
[3] G. de Rham, Variétés différentiables, Hermann, 1955.
[4] I. M. Gel'fand and G. E. Shilov, Generalized functions. I, Properties and operations; II, Spaces of fundamental and generalized functions; III, Theory of differential equations; IV (I. M. Gel'fand and N. Ya. Vilenkin), Applications of harmonic analysis; V (I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin), Integral geometry and representation theory, Academic Press, 1964, 1968, 1967, 1964, 1966; VI (I. M. Gel'fand, M. I. Graev, and I. I. Pyatetskiĭ-Shapiro), Representation theory and automorphic functions, Saunders, 1969. (Originals in Russian.)
[5] C. Roumieu, Sur quelques extensions de la notion de distribution, Ann. Sci. Ecole Norm. Sup. Paris, 77 (1960), 41–121.
[6] M. Sato, Theory of hyperfunctions I, II, J. Fac. Sci. Univ. Tokyo, sec. I, 8 (1959), 139–193, 387–437.
[7] M. Sato, T. Kawai, and M. Kashiwara, Microfunctions and pseudo-differential equations, in [16, pp. 265–529].
[8] L. Schwartz, Espaces de fonctions différentiables à valeurs vectorielles, J. Analyse Math., 4 (1954–1955), 88–148.
[9] A. Grothendieck, Produits tensoriels topologiques et espaces nucléaires, Mem. Amer. Math. Soc., 16 (1955).

[10] L. Schwartz, Théorie des distributions à valeurs vectorielles, Ann. Inst. Fourier, 7 (1957), 1–141; 8 (1958), 1–209.
[11] F. Treves, Topological vector spaces, distributions and kernels, Academic Press, 1967.
[12] L. Ehrenpreis, Fourier analysis in several complex variables, Wiley-Interscience, 1970.
[13] G. Björck, Linear partial differential operators and generalized distributions, Ark. Mat., 6 (1966), 351–407.
[14] H. Komatsu, Ultradistributions. I, Structure theorems and a characterization; II, The kernel theorem and ultradistributions with support in a submanifold; III, Vector-valued ultradistributions and the theory of kernels, J. Fac. Sci. Univ. Tokyo, sec. IA, 20 (1973), 25–105; 24 (1977), 607–628; 29 (1982), 653–718.
[15] M. Sato, Theory of hyperfunctions (in Japanese), Sugaku, 10 (1958), 1–27.
[16] H. Komatsu (ed.), Hyperfunctions and pseudo-differential equations, Lecture notes in math. 287, Springer, 1973.
[17] A. Martineau, Les hyperfonctions de M. Sato, Sém. Bourbaki, 13 (1960–1961), no. 214.
[18] P. Schapira, Théorie des hyperfonctions, Lecture notes in math. 126, Springer, 1970.
[19] A. Martineau, Le "edge of the wedge theorem" en théorie des hyperfonctions de Sato, Proc. Intern. Conf. Functional Analysis and Related topics, Tokyo, 1969, Univ. Tokyo Press, 1970, 95–106.
[20] T. Kawai, On the theory of Fourier hyperfunctions and its applications to partial differential equations with constant coefficients, J. Fac. Sci. Univ. Tokyo, sec. IA, 17 (1970), 467–517.
[21] K. Kataoka, On the theory of Radon transformations of hyperfunctions, J. Fac. Sci. Univ. Tokyo, sec. IA, 28 (1981), 331–413.
[22] J.-E. Björk, Rings of differential operators, North-Holland, 1979.

126 (IX.22) Dynamical Systems

A. History

The theory of dynamical systems began with the investigation of the motion of planets in ancient astronomy. Qualitative investigation of mechanics in antiquity and the Middle Ages culminated in the work of J. Kepler and G. Galilei in 17th century. At the end of that century, I. Newton founded his celebrated Newtonian mechanics, by means of which Kepler's law on the motion of planets and Galileo's observations of movement can be explained theoretically. Following this, L. Euler, J. L. Lagrange, P. S. Laplace, W. R. Hamilton, C. G. J. Jacobi, and others developed the theory using analytical methods, and founded analytical dynamics. From the end of the 18th century through the 19th century, the †three-body problem attracted the attention of many mathematicians. At the end of the 19th century, H. Bruns and H. Poincaré found that general solutions of the three-body problem could not be obtained by †quadrature, and this gave rise to a crisis of analytical dynamics. But this was resolved by Poincaré himself. He pointed out the importance of the qualitative theory based on topological methods, and obtained many fundamental results. A. M. Lyapunov with his theory of stability and G. D. Birkhoff with his many important concepts of topological dynamics established foundations of the new qualitative theory.

In 1937 A. A. Andronov and L. S. Pontryagin introduced the concept of structural stability, which attracted the attention of S. Lefschetz. Lefschetz's school investigated structural stability and †nonlinear oscillations, and obtained many important results (H. F. de Baggis, L. Markus, M. M. Peixoto, and others). In about 1960, S. Smale initiated study of differentiable dynamical systems under the influence of Lefschetz's school. Smale and his school founded a new theory of differentiable dynamical systems using †differential topology. D. V. Anosov generalized the work of E. Hopf and G. A. Hedlund on †geodesic flows of closed surfaces of †constant negative curvature and established the concept of Anosov systems, which played an important role in Smale's theory. The work of Hopf, Hedlund, and Anosov is closely related to †ergodic theory. Ya. G. Sinaĭ and R. Bowen obtained important results in ergodic theory. The concept of structural stability and its generalization are essential in the †catastrophe theory of R. Thom (→ 51 Catastrophe Theory); the theory of bifurcation of dynamical systems is another essential part of catastrophe theory. D. Ruelle and F. Takens proposed a new mathematical mechanism for the generation of turbulence using Smale's theory and Hopf bifurcation. The new theory of dynamical systems developed by Smale and others is now applied to the mathematical explanation of chaotic phenomena in many branches of science. Finally, we mention that in the 1960s A. N. Kolmogorov, V. I. Arnold, and J. Moser obtained remarkable results on the existence of quasi-periodic solutions for the n-body problem, which turned out to solve the long-standing problem of the stability of the solar system.

B. Definitions of Dynamical Systems

In the study of the evolution of physical, biological, and other systems, we construct mathematical models of the systems. Usually, the state of a given system is completely described by a collection of continuous parameters, which may be related in some cases. Thus the space X of all possible states of the system can be regarded as a Euclidean space or a subset of a Euclidean space defined by some equations. In general, we assume that the space X of all possible states of the system forms a †topological space, and we call it a **state space** or a **phase space**. Second, we assume that the law of evolution of states in time is given, by which we can tell the state x_1 at any time r_1 if we know the state x_0 at time t_0. Assigning x_1 to x_0, we have a mapping $\pi(t_1, t_0): X \to X$ for any times t_0 and t_1, which satisfies the following conditions: (i) $\pi(t_2, t_1) \circ \pi(t_1, t_0) = \pi(t_2, t_0)$; (ii) $\pi(t_0, t_0) = 1_X$, the identity mapping of X. Finally, we assume that the mapping $\pi(t_1, t_0)$ depends only on $t = t_1 - t_0$. Writing $\pi_t = \pi(t_1, t_0)$ if $t = t_1 - t_0$, we have the following conditions from (i) and (ii) above: (i') $\pi_s \circ \pi_t = \pi_{s+t}$, $s, t \in \mathbf{R}$; (ii') $\pi_0 = 1_X$.

In general, the theory of topological dynamics can be regarded as the study of topological transformation groups (→ 431 Transformation Groups) originating in the topological investigations of problems arising from classical mechanics. Here, we restrict our attention to some important special cases.

(1) Let X be a topological space and $\mathbf{R}$ the additive topological group of real numbers. Let $\varphi: X \times \mathbf{R} \to X$ be a continuous mapping. For each $t \in \mathbf{R}$, we define a mapping $\varphi_t: X \to X$ by $\varphi_t(x) = \varphi(x, t)$, $x \in X$. If the family of mappings $\{\varphi_t\}_{t \in \mathbf{R}}$ satisfies the following conditions, we say that (X, φ) is a **(continuous) R-action**, a **(continuous) flow**, or a **(continuous) dynamical system** on X, and that X is the **phase space**: (i) $\varphi_s \circ \varphi_t = \varphi_{s+t}$ for all $s, t \in \mathbf{R}$; (ii) $\varphi_0 = 1_X$.

Let (X, φ) be a flow. Then $\varphi_t: X \to X$ is a †homeomorphism with $(\varphi_t)^{-1} = \varphi_{-t}$ for each $t \in \mathbf{R}$. For each $x \in X$, define a mapping $\varphi^x: \mathbf{R} \to X$ by $\varphi^x(t) = \varphi(x, t)$, $t \in \mathbf{R}$. The mapping φ^x is called a **motion** through x, and its image $C(x) = \{\varphi^x(t) \mid t \in \mathbf{R}\}$ is called the **orbit** or the **trajectory** through x. An orbit is nonempty, and two orbits are either identical or mutually disjoint. The family of orbits fills up the phase space X and is called the **phase portrait**.

Let $\mathbf{R}_+$ be the additive †semigroup of all nonnegative real numbers. If we replace $\mathbf{R}$ by $\mathbf{R}_+$ in the definition of a (continuous) flow, we obtain a definition of a **(continuous) semiflow**. For a semiflow (X, φ), the mapping $\varphi_t: X \to X$, $t \in \mathbf{R}_+$ is in general not a homeomorphism but a continuous mapping.

Let (X, φ) and (Y, ψ) be flows. A homeomorphism $h: X \to Y$ is called a **topological equivalence** from (X, φ) to (Y, ψ) if it maps each orbit of φ onto an orbit of ψ preserving orientations of orbits (i.e., there exists an increasing homeomorphism $\alpha_x: \mathbf{R} \to \mathbf{R}$ for each $x \in X$ such that $h\varphi(x, t) = \psi(h(x), \alpha_x(t))$ for all $t \in \mathbf{R}$). Two flows are **topologically equivalent** if there exists a topological equivalence from one to the other. If two flows are topologically equivalent, their phase portraits have the same topological structure. Two flows (X, φ) and (Y, ψ) are **flow equivalent** if there exist a $c > 0$ and a homeomorphism $h: X \to Y$ such that $h\varphi(x, t) = \psi(h(x), ct)$ for all $t \in \mathbf{R}$. Such an h is a topological equivalence from (X, φ) to (Y, ψ).

(2) Let X be a topological space and $\mathbf{Z}$ the additive group of integers. If we replace $\mathbf{R}$ by $\mathbf{Z}$ in the definition of a continuous flow, we obtain a definition of a **(continuous) Z-action**, a **discrete flow**, or a **discrete dynamical system** on X. If (X, φ) is a discrete flow, then $f = \varphi_1: X \to X$ is a homeomorphism and $\varphi_n = f^n$ for all $n \in \mathbf{Z}$. Conversely, for a given homeomorphism $f: X \to X$, define a mapping $\varphi: X \times \mathbf{Z} \to X$ by $\varphi(x, n) = f^n(x)$, $x \in X$ and $n \in \mathbf{Z}$. Then (X, φ) is a discrete flow such that $\varphi_n = f^n$ for $n \in \mathbf{Z}$. So we identify a homeomorphism with a discrete flow. Thus the orbit of a homeomorphism $f: X \to X$ through x is $C(x) = \{f^n(x) \mid n \in \mathbf{Z}\}$.

Let $\mathbf{Z}_+$ be the additive semigroup of all nonnegative integers. If we replace $\mathbf{R}$ by $\mathbf{Z}_+$ in the definition of a continuous flow, we obtain a definition of a **discrete semiflow** on X. For a discrete semiflow (X, φ), the mapping $\varphi_n: X \to X$, $n \in \mathbf{Z}_+$ is in general not a homeomorphism but a continuous mapping. We can identify a continuous mapping $f: X \to X$ with a discrete semiflow (X, φ) in a natural way as above.

Let $f: X \to X$ and $g: Y \to Y$ be two homeomorphisms (continuous mappings). A homeomorphism $h: X \to Y$ such that $h \circ f = g \circ h$ is called a **topological conjugacy** from f to g. And f and g are called **topologically conjugate** if there exists a topological conjugacy from f to g. Topologically conjugate homeomorphisms have the same phase portrait in a topological sense.

(3) Let M be a †differentiable manifold of class C^r ($1 \leqslant r \leqslant \infty$ or $r = \omega$). A continuous flow (M, φ) is a **flow of class C^r**, a **C^r-flow**, a **differentiable dynamical system of class C^r**, or a **one-parameter group of transformations of class C^r**, if $\varphi: M \times \mathbf{R} \to M$ is of class C^r. A **semiflow of class C^r** is defined similarly.

Let (M, φ) and (N, ψ) be C^r-flows. A topological equivalence $h: M \to N$ from (M, φ) to (N, ψ) is called a **C^r-equivalence** if it is a †C^r-diffeomorphism. Two flows are **C^r-equivalent** if there is a C^r-equivalence from one to the other. Classification of C^r-flows by C^r-equivalence is

dificult and sometimes too unwieldy to work with. On the other hand, there are many problems which can be solved by the knowledge of the topological structure of the phase portrait of C^r-flows.

(4) Let (M, π) be a discrete flow on a differentiable manifold M of class C^r. If $\pi: M \times \mathbf{Z} \to M$ is of class C^r, then (M, π) is called a **Z-action of class C^r**, a **discrete flow of class C^r**, a **discrete C^r-flow**, or a **discrete dynamical system of class C^r**. If (M, π) is a discrete C^r-flow, then $f = \pi_1: M \to M$ is a C^r-diffeomorphism. Conversely, a C^r-diffeomorphism $f: M \to M$ defines a discrete C^r-flow in a natural way on M, and we identify a C^r-diffeomorphism $f: M \to M$ with the discrete C^r-flow on M defined by f. A **discrete semiflow of class C^r** is defined similarly, and a †C^r-mapping $f: M \to M$ is identified with a discrete semiflow of class C^r on M defined by f.

Let $f: M \to M$ and $g: N \to N$ be C^r-diffeomorphisms (C^r-mappings) of differentiable manifolds M and N of class C^r. A topological conjugacy from f to g is called a **C^r-conjugacy** if it is a C^r-diffeomorphism. f and g are **C^r-conjugate** if there is a C^r-conjugacy from f to g.

C. Examples and Remarks

(1) Let D be an open set of $\mathbf{R}^n$ and $f: D \to \mathbf{R}^n$ a continuous mapping. Consider the †autonomous system of ordinary differential equations

$$dx/dt = f(x), \qquad x \in D. \tag{1}$$

We assume that for each $x \in D$ there exists a unique †nonextendable solution $\varphi(x, t)$ with the initial condition $\varphi(x, 0) = x$ defined on a maximal interval (a_x, b_x), $-\infty \leqslant a_x < 0 < b_x \leqslant \infty$ (→ 316 Ordinary Differential Equations (Initial Value Problems)). The set $\{\varphi(x, t) \mid a_x < t < b_x\}$ is called the **trajectory** through x. By the uniqueness assumption, we have $\varphi(\varphi(x, t), s) = \varphi(x, t + s)$ whenever both sides of the equality are defined. When $(a_x, b_x) = \mathbf{R}$ for all $x \in D$, then equation (1) is called **complete**. If (1) is complete, the mapping $\varphi: D \times \mathbf{R} \to D$ defined by the solutions $\varphi(x, t)$ determines a continuous flow (D, φ). Furthermore, if f is of class C^r, then (D, φ) is of class C^r. If (1) is not complete, then there exists a continuous positive scalar function $\alpha: D \to \mathbf{R}$ such that

$$dx/dt = \alpha(x) f(x), \qquad x \in D, \tag{2}$$

is complete. The trajectories of (1) and (2) through x coincide for all $x \in D$, and thus the phase portraits of (1) and (2) are the same.

(2) Let M be a differentiable manifold of class C^∞. A †vector field of class C^r on M $(1 \leqslant r \leqslant \infty)$ gives rise to an autonomous system of ordinary differential equations in a coordinate neighborhood of each point of M, and it generates a †local one-parameter group of local transformations of class C^r. If M is †compact, this local one-parameter group of local transformations is uniquely extended to a one-parameter group of transformations of class C^r (→ 105 Differentiable Manifolds). Thus a vector field of class C^r on M generates a unique C^r-flow on M if M is compact. Therefore, sometimes we identify a vector field of class C^r with the C^r-flow generated by it.

(3) Let (M, φ) be a C^r-flow. Then $\varphi_1: M \to M$ is a C^r-diffeomorphism, which we call the **time-one mapping (time-one map)** of (M, φ). Thus every C^r-flow (M, φ) induces a C^r-diffeomorphism as a time-one mapping. But the set of C^1-diffeomorphisms that are time-one mappings of C^1-flows is of the †first category in the space of all C^1-diffeomorphisms with C^1 topology (J. Palis). Thus most C^1-diffeomorphisms are not expressed as time-one mappings of C^1-flows.

(4) Let M be a compact differentiable manifold of class C^∞ with dimension m and (M, φ) a C^r-flow $(1 \leqslant r \leqslant \infty)$. An $(m-1)$-dimensional closed submanifold Σ of M is called a **cross section** of (M, φ) if the following conditions are satisfied: (i) For any $x \in X$, there exist $t_1 > 0$ and $t_2 < 0$ such that $\varphi_{t_1}(x)$, $\varphi_{t_2}(x) \in \Sigma$; (ii) Every orbit intersects Σ †transversally whenever it meets Σ. Let Σ be a cross-section of (M, φ) and $x \in \Sigma$. Let t_1 be the least positive number with $\varphi_{t_1}(x) \in \Sigma$. Such a t_1 exists for every $x \in \Sigma$, and $f: \Sigma \to \Sigma$ defined by $f(x) = \varphi_{t_1}(x)$, $x \in \Sigma$ is a C^r-diffeomorphism. We call this diffeomorphism f the **first-return mapping (first-return map)** or the **Poincare mapping (Poincaré map)** for Σ.

(5) Let N be a compact differentiable manifold of class C^∞ and $f: N \to N$ a C^r-diffeomorphism $(1 \leqslant r \leqslant \infty)$. Define an equivalence relation $\sim$ on $N \times \mathbf{R}$ generated by $(x, t+1) \sim (f(x), t)$ for $x \in X$, $t \in \mathbf{R}$. Then the quotient space $N(f) = N \times \mathbf{R}/\sim$ has a natural †differentiable structure of class C^r, and a C^r-flow $(N(f), \psi)$ is defined by $\psi([x, t], s) = [x, t+s]$ for $x \in N$, $t, s \in \mathbf{R}$, where $[x, t] \in N(f)$ is the equivalence class of (x, t). The flow $(N(f), \psi)$ thus obtained is called the **suspension** of f. The suspension $(N(f), \psi)$ has a cross section $\Sigma = \{[x, 0] \mid x \in N\}$, and the Poincaré mapping for Σ is C^r-conjugate to f. Conversely, if (M, φ) has a cross section Σ and the Poincaré mapping for Σ is $f: \Sigma \to \Sigma$, then the suspension $(\Sigma(f), \psi)$ is C^r-equivalent to (M, φ).

(6) Let U be an open set of $\mathbf{R}^n$ and $f: U \to \mathbf{R}^n$ a continuous mapping. Consider the †difference equation

$$x_{m+1} = f(x_m), \qquad x_m \in U. \tag{3}$$

For each $x \in U$, let $\varphi(x, m)$ be the solution of (3)

with $\varphi(x,0)=x$. If $f(U)\subset U$, then $\varphi(x,m)$ exists for all $m\in\mathbf{Z}_+$ and $x\in U$, and φ defines a discrete semiflow on U. If $f: U\to U$ is a homeomorphism, then $\varphi(x,m)$ exists for all $m\in\mathbf{Z}$ and $x\in U$, and φ defines a discrete flow on U (→ 104 Difference Equations).

D. Basic Concepts

For simplicity we assume that the phase spaces of dynamical systems are metric spaces, and we denote their metrics by d.

(1) Let (X,φ) be a continuous flow on a metric space X and $x\in X$. A subset A of X is called **invariant** if $\varphi_t(A)\subset A$ for all $t\in\mathbf{R}$. A subset of X is invariant if and only if it is a union of orbits. A subset A of X is **positively** (resp. **negatively**) **invariant** if $\varphi_t(A)\subset A$ for all $t\geqslant 0$ (resp. $t\leqslant 0$). The subset $C_+(x)=\{\varphi^x(t)\mid t\geqslant 0\}$ (resp. $C_-(x)=\{\varphi^x(t)\mid t\leqslant 0\}$) is called the **positive** (resp. **negative**) **semiorbit** or the **positive** (resp. **negative**) **half-trajectory** starting from x. A subset is positively (resp. negatively) invariant if and only if it is a union of positive (resp. negative) semiorbits. A subset is invariant if and only if it is both positively and negatively invariant.

The union and the intersection of invariant sets are invariant. If A is invariant, then its closure $\bar{A}$, its interior $\operatorname{Int} A$, its boundary ∂A, and its complement $A^c=X-A$ are invariant. If A is invariant, then $\varphi(A\times\mathbf{R})\subset A$ and the restriction mapping $\varphi\mid A\times\mathbf{R}$ defines a continuous flow $(A,\varphi\mid A\times\mathbf{R})$ on A. The flow thus obtained is called the **restriction** of (X,φ) on A.

(2) A point $x\in X$ is a **singular point**, an **equilibrium point**, a **critical point**, a **rest point**, or a **fixed point** if $C(x)=\{x\}$ (→ 290 Nonlinear Oscillation). A point is **regular** or **nonsingular** if it is not a singular point. The set of all singular points is a closed invariant set, and the set of all nonsingular points is an open invariant set. If A is a positively invariant set which is homeomorphic to the closed unit ball in $\mathbf{R}^n$, then there exists a singular point in A (†Brouwer fixed-point theorem).

A point $x\in X$ is **periodic** if there exists a $T\neq 0$ such that

$$\varphi(x,t+T)=\varphi(x,t) \tag{4}$$

holds for all $t\in\mathbf{R}$. If x is periodic, the motion φ^x and the orbit $C(x)$ are said to be **periodic**. A point x is periodic if and only if there exists a $T\neq 0$ with $\varphi(x,T)=x$. A singular point is periodic. For a periodic point x, a number T satisfying (4) is called a **period** of x. If x is nonsingular and periodic, then there exists a smallest positive period T_0 of x, and any period is an integral multiple of T_0. An orbit of a nonsingular periodic point is called a **closed orbit**. If C is a closed orbit, then all points of C are nonsingular periodic points, and their smallest positive periods coincide. A closed orbit is compact.

(3) Let x be a point of X. A point $y\in X$ is called an **ω-limit** (resp. **α-limit**) **point** or a **positive** (resp. **negative**) **limit point** of x if there exists a sequence $\{t_n\}$ of real numbers such that (i) $t_n\to\infty$ (resp. $t_n\to-\infty$) as $n\to\infty$ and (ii) $\varphi(x,t_n)\to y$ as $n\to\infty$. The set of all ω-limit (resp. α-limit) points of x is denoted by $\omega(x)$ (resp. $\alpha(x)$) and is called the **ω-limit** (resp. **α-limit**) **set** of x. For each $x\in X$, $\omega(x)$ and $\alpha(x)$ are closed invariant sets, and the following equalities hold: $\overline{C_+(x)}=C_+(x)\cup\omega(x)$, $\overline{C_-(x)}=C_-(x)\cup\alpha(x)$, and $\overline{C(x)}=C(x)\cup\alpha(x)\cup\omega(x)$. If x is a periodic point, $\overline{C(x)}=C(x)=\alpha(x)=\omega(x)$. If C is a closed orbit, then $\bar{C}=C=C(x)=\alpha(x)=\omega(x)$ for all $x\in C$. If A is a compact invariant set and $x\in A$, then $\alpha(x)$ and $\omega(x)$ are nonempty.

Assume that X is †locally compact and $x\in X$. Then $\omega(x)$ is connected if it is compact, and none of the connected components of $\omega(x)$ is compact if $\omega(x)$ is not compact.

(4) Let x be a point of X. Let $J_+(x)$ (resp. $J_-(x)$) be the set of all points y satisfying the following condition: There exist a sequence $\{t_n\}$ of numbers and a sequence $\{x_n\}$ of points in X such that (i) $t_n\to\infty$ (resp. $t_n\to-\infty$) as $n\to\infty$, (ii) $x_n\to x$ as $n\to\infty$, and (iii) $\varphi(x_n,t_n)\to y$ as $n\to\infty$. The set $J_+(x)$ (resp. $J_-(x)$) is a closed invariant set containing $\omega(x)$ (resp. $\alpha(x)$), called the **first positive** (resp. **negative**) **prolongational limit set** of x. If X is locally compact, then $J_+(x)$ is connected if it is compact, and none of the connected components of $J_+(x)$ is compact when $J_+(x)$ is not compact.

Notions of higher prolongations have been defined and investigated by T. Ura, J. Auslander, and P. Seibert.

(5) For a discrete flow, we can similarly define basic notions such as an invariant set, fixed point, periodic point, and so on defined in Sections D(1)–D(4) by replacing $\mathbf{R}$ by $\mathbf{Z}$. The propositions and theorems stated above hold for discrete flows, except those concerning connectedness.

E. Recursive Concepts and Dispersive Concepts

(1) Let (X,φ) be a flow on a metric space X. Let $x\in X$ be a point such that there exist a neighborhood U of x and a $T>0$ satisfying the condition $U\cap\varphi_t(U)=\varnothing$ for $t\geqslant T$. Then x is called **wandering**. The set of all wandering points is an open invariant set. A point is **nonwandering** if it is not wandering. The set Ω of all nonwandering points is a closed invariant set and is called the **nonwandering set**.

The following conditions are equivalent: (i) x is nonwandering, (ii) $x \in J_+(x)$, (iii) $x \in J_-(x)$. The nonwandering set Ω contains all singular points, closed orbits, and $\omega(x)$ and $\alpha(x)$ for all $x \in X$.

(2) A point $x \in X$ is **positively** (resp. **negatively**) **Poisson stable** if $x \in \omega(x)$ (resp. $x \in \alpha(x)$). It is **Poisson stable** if it is both positively and negatively Poisson stable. A positively Poisson stable point and a negatively Poisson stable point are nonwandering. The following conditions are equivalent: (i) x is positively Poisson stable, (ii) $\overline{C_+(x)} = \omega(x)$, (iii) $C(x) \subset \omega(x)$, (iv) for any neighborhood U of x and $T > 0$, there exists a $t > T$ with $\varphi(x, t) \in U$. If X is †complete and $x \in X$ is positively Poisson stable but not periodic, then $\omega(x) - C(x)$ is †dense in $\omega(x)$. This implies that if X is complete, then $C(x)$ is periodic if and only if $C(x) = \omega(x)$.

(3) If all the points of the phase space X are wandering, then (X, φ) is called **completely unstable**. If all the points of X are nonwandering, then (X, φ) is called **regionally recurrent**. If A is an invariant set such that the restriction of (X, φ) on A is regionally recurrent, then (X, φ) is said to be **regionally recurrent on** A. If (X, φ) is regionally recurrent and X is locally compact, then the set of all Poisson stable points is dense in X.

For a given flow (X, φ), we obtain a sequence of invariant sets $\{\Omega_n\}$ and a sequence of restriction flows $\{(\Omega_n, \varphi_n)\}$ such that (i) $(\Omega_0, \varphi_0) = (X, \varphi)$; (ii) Ω_{n+1} is the nonwandering set of (Ω_n, φ_n), $n \geq 0$; and (iii) $(\Omega_{n+1}, \varphi_{n+1})$ is the restriction of (X, φ) on Ω_{n+1}. Then $X = \Omega_0 \supset \Omega_1 \supset \ldots \supset \Omega_n \supset \ldots$. Put $\Omega_\omega = \bigcap_n \Omega_n$. Then Ω_ω is an invariant set of (X, φ), and we denote the restriction of (X, φ) on Ω_ω by $(\Omega_\omega, \varphi_\omega)$. Starting from $(\Omega_\omega, \varphi_\omega)$, we obtain similarly a sequence of invariant sets $\{\Omega_{\omega+n}\}$ and a sequence of flows $\{(\Omega_{\omega+n}, \varphi_{\omega+n})\}$. If we obtain an ordinal number γ such that $\Omega_\gamma = \Omega_{\gamma+1} \neq \varnothing$ by continuing this process, then we call Ω_γ the set of **central motions**. In this case, the flow $(\Omega_\gamma, \varphi_\gamma)$ is regionally recurrent, and every invariant subset of X on which (X, φ) is regionally recurrent is contained in Ω_γ. When X is †separable and Ω_1 is compact and nonempty, then the minimum of such ordinal γ is an †ordinal of at most the second number class.

(4) Let $\mathscr{R}$ be the set of all $x \in X$ satisfying the following condition: For each ε, $T > 0$, there exist a sequence $\{x_0 = x, x_1, \ldots, x_k = x\}$ of points in X and a sequence $\{t_0, t_1, \ldots, t_{k-1}\}$ of numbers such that $t_i \geq T$ and $d(\varphi_{t_i}(x_i), x_{i+1}) < \varepsilon$ for $0 \leq i \leq k-1$. The set $\mathscr{R}$ is a closed invariant set containing the nonwandering set Ω and is called the **chain recurrent set**. If $X = \mathscr{R}$, then (X, φ) is called **chain recurrent**. The restriction $(\mathscr{R}, \varphi \mid \mathscr{R} \times \mathbf{R})$ of (X, φ) on $\mathscr{R}$ is chain recurrent (C. C. Conley).

(5) A nonempty closed invariant set is called a **minimal set** if none of its nonempty proper closed subsets is invariant. For a nonempty compact subset A of X, the following conditions are equivalent: (i) A is minimal, (ii) $\overline{C(x)} = A$ for all $x \in A$, (iii) $\overline{C_+(x)} = A$ for all $x \in A$, (iv) $\overline{C_-(x)} = A$ for all $x \in A$, (v) $\omega(x) = A$ for all $x \in A$, (vi) $\alpha(x) = A$ for all $x \in A$. A point $x \in X$ is **positively** (resp. **negatively**) **Lagrange stable** if $\overline{C_+(x)}$ (resp. $\overline{C_-(x)}$) is compact. If $\overline{C(x)}$ is compact, then x is called **Lagrange stable**. Every nonempty compact invariant set contains a compact minimal set. If $x \in X$ is positively (resp. negatively) Lagrange stable, then $\omega(x)$ (resp. $\alpha(x)$) contains a compact minimal set.

A closed invariant set A is called a **saddle set** if there exists a neighborhood U of A such that every neighborhood of A contains a point x such that $C_+(x) \not\subset U$ and $C_-(x) \not\subset U$. Otherwise A is called a **nonsaddle set**. For a point x of X, let $E(x)$ be the subset of X consisting of the points y satisfying the following condition: There exist a sequence $\{x_n\}$ of points in X and two sequences $\{t_n\}$, $\{s_n\}$ of numbers such that (i) $x_n \to x$, $t_n \to \infty$, $s_n \to -\infty$ as $n \to \infty$ and (ii) $\varphi(x_n, t_n) \to y$, $\varphi(x_n, s_n) \to y$ as $n \to \infty$. For a subset B of X, put $E(B) = \bigcup_{x \in B} E(x)$. Let $\{S_\alpha\}$ be the family of all saddle minimal sets and $\{F_\beta\}$ the family of all nonsaddle minimal sets. If the phase space X is compact, then the nonwandering set $\Omega = (\bigcup_\beta F_\beta) \cup (\bigcup_\alpha E(S_\alpha)) = (\bigcup_\beta F_\beta) \cup E(\bigcup_\alpha S_\alpha)$ (T. Saito).

(6) A point $x \in X$ is said to be **recurrent** if for any $\varepsilon > 0$ there exists a $T > 0$ such that the †ε-neighborhood U of $\varphi^x([t, t+T])$ contains $C(x)$ for all $t \in \mathbf{R}$. If x is recurrent, the motion φ^x and the orbit $C(x)$ are said to be **recurrent**. Every orbit of a compact minimal set is recurrent, and if the phase space is complete, then the closure of a recurrent orbit is a compact minimal set (Birkhoff).

A set D of real numbers is called **relatively dense** if there exists a $T > 0$ such that $D \cap (t, t+T) \neq \varnothing$ for all $t \in \mathbf{R}$. Assume that $x \in X$ is Lagrange stable, then x is recurrent if and only if for every $\varepsilon > 0$ the set $\{t \mid d(x, \varphi(x, t)) < \varepsilon\}$ is relatively dense.

(7) A flow (X, φ) is **dispersive** if for any x, $y \in X$ there exist neighborhoods U_x of x and U_y of y and a $T > 0$ such that $U_x \cap \varphi_t(U_y) = \varnothing$ for all $t \geq T$. The following conditions are equivalent: (i) (X, φ) is dispersive; (ii) For any x, $y \in X$, there exist neighborhoods U_x of x and U_y of y and a $T > 0$ such that $U_x \cap \varphi_t(U_y) = \varnothing$ if $|t| \geq T$; (iii) $J_+(x) = \varnothing$ for all $x \in X$.

A flow (X, φ) is **parallelizable** if there exist a subset S of X and a homeomorphism h: $X \to S \times \mathbf{R}$ such that (i) $\varphi(S \times \mathbf{R}) = X$ and (ii) $h \circ \varphi(x, t) = (x, t)$ for all $x \in S$ and $t \in \mathbf{R}$. A flow (X, φ) is parallelizable if and only if there is a subset S of X satisfying the following con-

ditions: (i) For each $x \in X$ there exists a unique $\tau(x) \in \mathbf{R}$ with $\varphi(x, \tau(x)) \in S$; (ii) $\tau: X \to \mathbf{R}$ is continuous. A parallelizable flow is dispersive. A flow on a locally compact separable metric space is parallelizable if and only if it is dispersive (V. V. Nemytskiĭ, V. V. Stepanov).

An open set U of X is called a **tube** if there exist a $T > 0$ and a subset S of U satisfying the following conditions: (i) $\varphi(S \times (-T, T)) = U$, (ii) For each $x \in U$ there is a unique $\tau(x) \in \mathbf{R}$ with $|\tau(x)| < T$ such that $\varphi(x, \tau(x)) \in S$. The set S in the above definition is called a **local section**. If $x \in X$ is a regular point, then there exists a tube containing x (M. Bebutov, H. Whitney). The notion of a local section is a generalization of Poincare's "surface sans contact" [1] or Birkhoff's "surface of section" [4], and it is related to the notion of the cross section.

(8) For discrete flows (homeomorphisms), basic notions, such as a nonwandering set, Poisson stability, regional recurrence, central motion, and minimal set, are defined similarly, and many of the propositions and theorems in Sections E(1)–E(5) hold similarly for discrete flows (homeomorphisms).

F. Stability

(1) Let (X, φ) be a continuous flow on a metric space X. A point $x \in X$ is called **orbitally stable** if for any $\varepsilon > 0$ there exists a $\delta > 0$ such that $C_+(y)$ is contained in the ε-neighborhood of $C_+(x)$ for all y with $d(x, y) < \delta$ (→ 394 Stability). A nonempty set A is called **stable** if every neighborhood of A contains a positively invariant neighborhood of A. If A is compact (in particular, if A is a periodic orbit), then orbital stability and stability for A are equivalent. A nonempty set A is called **asymptotically stable** if A is stable and there exists a neighborhood V of A such that $\omega(x) \subset A$ for any $x \in V$. If A is stable and $\omega(x) \subset A$ for all $x \in X$, then A is called **globally asymptotically stable**. A point $x \in X$ is **Lyapunov stable** if for any $\varepsilon > 0$ there exists a $\delta > 0$ such that $d(\varphi_t(x), \varphi_t(y)) < \varepsilon$ for all $t \geqslant 0$ and y with $d(x, y) < \delta$. Lyapunov stability implies orbital stability. A point x is **uniformly Lyapunov stable** if for any $\varepsilon > 0$ there exists a $\delta > 0$ such that for $z \in C(x)$ and y with $d(y, z) < \delta$ we have $d(\varphi_t(z), \varphi_t(y)) < \varepsilon$ for all $t \geqslant 0$. Uniform Lyapunov stability implies Lyapunov stability. For a singular point, the notions of uniform Lyapunov stability, Lyapunov stability, orbital stability, and stability are equivalent. Assume that the phase space X is locally compact and A is a nonempty compact subset of X. Then A is asymptotically stable if and only if there exist a neighborhood N of A and a continuous real-valued function L on N such that (i) $L(x) = 0$ if $x \in A$ and $L(x) > 0$ if $x \notin A$; (ii) $L(\varphi(x, t)) < L(x)$ if $x \notin A$, $t > 0$, and $\varphi(\{x\} \times [0, t]) \subset N$. Such a function L is called a **Lyapunov function** for A (→ 394 Stability). Assume that X is locally compact and A is nonempty, stable, and invariant. Then A is asymptotically stable if and only if there exists a neighborhood U of A such that any invariant set in U is necessarily contained in A (Ura). Let A be a nonempty closed invariant set. A is called an **attractor** if it has an open neighborhood U satisfying the following conditions: (i) U is positively invariant; (ii) For each open neighborhood V of A, there is a $T > 0$ such that $\varphi_t(U) \subset V$ for all $t \geqslant T$. Condition (ii) implies that $\bigcap_{t \geqslant 0} \varphi_t(U) = A$ and $\omega(x) \subset A$ for all $x \in U$. Thus an attractor is asymptotically stable. If A is an attractor, the **basin** of A is the union of all open neighborhoods of A satisfying (i) and (ii).

(2) Assume that the phase space X is complete. A motion φ^x $(x \in X)$ is called **almost periodic** if for any $\varepsilon > 0$ there exists a relatively dense subset $\{\tau_n\}$ in $\mathbf{R}$ such that $d(\varphi^x(t), \varphi^x(t + \tau_n)) < \varepsilon$ for all $t \in \mathbf{R}$ and τ_n (→ 18 Almost Periodic Functions). If φ^x is almost periodic, then φ^y is almost periodic for all $y \in C(x)$. If A is a compact minimal set and if φ^x is almost periodic for some x in A, then every motion φ^y, $y \in A$, is almost periodic. An almost periodic motion is recurrent. The converse is not true. But if x is recurrent and Lyapunov stable in $C(x)$ (i.e., in the restriction flow $(C(x), \varphi \mid C(x) \times \mathbf{R})$), then φ^x is almost periodic. If x is uniformly Lyapunov stable in $C(x)$ and negatively Lagrange stable, then φ^x is almost periodic (A. A. Markov).

G. Singular Points and Closed Orbits

In this section we assume that the phase space is a †paracompact differentiable manifold of class C^∞ with metric d.

(1) Let E be a finite-dimensional real vector space and $L: E \to E$ a linear automorphism. L is called **hyperbolic** if it has no †eigenvalues of absolute value 1. If $L: E \to E$ is hyperbolic, there are unique vector subspaces E^s and E^u satisfying the following conditions: (i) $E = E^s \oplus E^u$, (ii) $L(E^s) = E^s$ and $L(E^u) = E^u$, (iii) if $\|\cdot\|$ is a †norm on E, then there exist constants $c > 0$ and $0 < \lambda < 1$ such that, for any positive integer m, $\|L^m(v)\| \leqslant c\lambda^m \|v\|$ when $v \in E^s$ and $\|L^{-m}(v)\| \leqslant c\lambda^m \|v\|$ when $v \in E^u$. The zero 0 of E is the only fixed point of a hyperbolic linear automorphism. Put $s = \dim E^s$ and $u = \dim E^u$. Then $s + u = \dim E$, and s (resp. u) is the number of eigenvalues of absolute value < 1 (resp. > 1) counted with multiplicity. A topological conjugacy class of a hyperbolic linear automorphism $L: E \to E$ is determined by s, u, and the signs of $\det(L \mid E^s)$ and $\det(L \mid E^u)$, where $\det(L \mid E^\sigma)$ is the †determinant of the restriction

$L|E^\sigma: E^\sigma \to E^\sigma$ of L on E^σ ($\sigma = s, u$). Further investigations of topological classification of linear automorphisms have been carried out by N. H. Kuiper and J. W. Robbin.

(2) Let $f: \mathbf{R}^n \to \mathbf{R}^n$ be a C^r-diffeomorphism ($1 \leqslant r \leqslant \infty$). Assume that the origin $0 \in \mathbf{R}^n$ is a fixed point of f. Then the †differential $df_0: \mathbf{R}^n \to \mathbf{R}^n$ of f at 0 is a linear automorphism. It is given by $df_0(x) = J_0(f)x$, $x \in \mathbf{R}^n$, where $J_0(f)$ is the †Jacobian matrix of f at 0 and x is expressed as a column vector. If df_0 is hyperbolic, then f is topologically conjugate to df_0 in a sufficiently small neighborhood of 0 (P. Hartman). Assume that f is of class C^∞, and let $\lambda_1, \ldots, \lambda_n$ (possibly repeated) be the eigenvalues of df_0. If $\lambda_i \neq \lambda_1^{m_1} \ldots \lambda_n^{m_n}$ for all $1 \leqslant i \leqslant n$ and for all nonnegative integers $m_1, \ldots, m_n$ with $2 \leqslant m_1 + \ldots + m_n$, then in a sufficiently small neighborhood of 0, f is C^∞-conjugate to df_0 (S. Sternberg).

(3) Let $f: M \to M$ be a C^r-diffeomorphism ($1 \leqslant r \leqslant \infty$) of a differentiable manifold M of class C^∞. Let $p \in M$ be a fixed point of f. Then the †differential df_p of f at p is a linear automorphism of the †tangent space $T_p(M)$ of M at p. If df_p is hyperbolic, then p is called a **hyperbolic fixed point** of f. Let p be a hyperbolic fixed point of f and $T_p(M) = E^s \oplus E^u$ be the direct sum decomposition with respect to df_p. Put $W^s(p) = \{x \in M \mid f^n(x) \to p \text{ as } n \to \infty\}$ and $W^u(p) = \{x \in M \mid f^{-n}(x) \to p \text{ as } n \to \infty\}$. $W^s(p)$ and $W^u(p)$ are invariant sets and images of †injective immersions of class C^r of vector spaces E^s and E^u, respectively. At the point p, $W^s(p)$ and $W^u(p)$ are tangent to E^s and E^u, respectively [21]. We call $W^s(p)$ (resp. $W^u(p)$) the **stable** (resp. **unstable**) **manifold** of f at p. In a sufficiently small neighborhood of a hyperbolic fixed point p of f, f is topologically conjugate to its differential df_p. Therefore a hyperbolic fixed point is an **isolated fixed point** (i.e., there are no fixed points in its sufficiently small neighborhood except itself). A hyperbolic fixed point p is called a **source** if $\dim W^s(p) = 0$ and a **sink** if $\dim W^u(p) = 0$. Otherwise it is a **saddle point**. The number of topological conjugacy classes of hyperbolic fixed points on an n-dimensional manifold is $4n$.

Let $p \in M$ be a periodic point of f and m the smallest positive period of p. Replacing f by f^m, we obtain notions of hyperbolicity, stable manifold, and so on for the periodic point p. We also obtain propositions and theorems similar to those stated above for periodic points.

(4) Let A be a real $n \times n$ matrix. Consider the system of linear differential equations

$$dx/dt = Ax, \qquad x \in \mathbf{R}^n. \tag{5}$$

This equation generates a C^∞-flow $(\mathbf{R}^n, \varphi)$, and $\varphi_t: \mathbf{R}^n \to \mathbf{R}^n$ is given by $\varphi_t(x) = e^{tA}x$, where e^{tA} is the †exponential of the matrix tA. Thus φ_t is a linear automorphism for all $t \in \mathbf{R}$. The origin $0 \in \mathbf{R}^n$ is a singular point of $(\mathbf{R}^n, \varphi)$. If none of the real parts of the eigenvalues of A are zero, we call $0 \in \mathbf{R}^n$ a **hyperbolic singular point** of (5). If $0 \in \mathbf{R}^n$ is a hyperbolic singular point of (5), then there exist two vector subspaces E^s and E^u of $\mathbf{R}^n$ satisfying the following conditions: (i) $R^n = E^s \oplus E^u$; (ii) $A(E^s) = E^s$ and $A(E^u) = E^u$; (iii) E^s and E^u are invariant sets of $(\mathbf{R}^n, \varphi)$; (iv) there exist positive constants c and λ such that for all $t \geqslant 0$, $\|\varphi_t(x)\| \leqslant ce^{-\lambda t}\|x\|$ when $x \in E^s$ and $\|\varphi_{-t}(x)\| \leqslant ce^{-\lambda t}\|x\|$ when $x \in E^u$, where $\|\cdot\|$ is the norm on $\mathbf{R}^n$. The origin 0 is a hyperbolic singular point of (5) if and only if the time-one mapping $\varphi_1 = e^A: \mathbf{R}^n \to \mathbf{R}^n$ of φ is a hyperbolic linear automorphism. If 0 is a hyperbolic singular point of (5), the above direct sum decomposition $\mathbf{R}^n = E^s \oplus E^u$ coincides with the one with respect to φ_1. Put $s = \dim E^s$ and $u = \dim E^u$. Then $s + u = n$, and s (resp. u) is the number of all eigenvalues of A with negative (resp. positive) real parts counted with multiplicity. Also we obtain the following properties: (i) $x \in E^s$ (resp. E^u) $\Leftrightarrow \varphi_t(x) \to 0$ as $t \to \infty$ (resp. $t \to -\infty$), (ii) $x \notin E^s$ (resp. E^u) $\Rightarrow \|\varphi_t(x)\| \to \infty$ as $t \to \infty$ (resp. $t \to -\infty$), (iii) the origin 0 is the only singular point.

Let A and B be $n \times n$ real matrices. Let $(\mathbf{R}^n, \varphi_A)$ and $(\mathbf{R}^n, \varphi_B)$ be the flows determined by the equations $dx/dt = Ax$ and $dx/dt = Bx$, respectively. Assume that $0 \in \mathbf{R}^n$ is a hyperbolic singular point for both of the equations. Let s and u (resp. s' and u') be the integers defined for φ_A (resp. φ_B) in the above paragraph. Then the following conditions are equivalent: (i) $(\mathbf{R}^n, \varphi_A)$ and $(\mathbf{R}^n, \varphi_B)$ are flow equivalent, (ii) $(\mathbf{R}^n, \varphi_A)$ and $(\mathbf{R}^n, \varphi_B)$ are topologically equivalent, (iii) $s = s'$, (iv) $u = u'$, (v) $(\varphi_A)_1$ and $(\varphi_B)_1$ are topologically conjugate. Further investigations of the phase portrait of the equation $dx/dt = Ax$ without the assumption of hyperbolicity were done by Kuiper.

(5) Let D be an open set of $\mathbf{R}^n$ and $f: D \to \mathbf{R}^n$ a continuous mapping. Consider the system of differential equations (1), which we write down again here:

$$dx/dt = f(x), \quad x \in D. \tag{6}$$

A point $p \in D$ is a **singular point** of (6) if the trajectory through p consists of a single point p. If (6) generates a flow (D, φ), then p is a singular point of (6) if and only if p is a singular point of (D, φ). A point p is a singular point of (6) if and only if $f(p) = 0$.

Suppose further that f is a C^r-mapping ($1 \leqslant r \leqslant \infty$). A singular point p of (6) is called **hyperbolic** if none of the real parts of the eigenvalues of the Jacobian matrix $J_p(f)$ of f at p is zero. Let p be a singular point of (6). We assume for simplicity that $p = 0$ and (6) generates

a flow (D, φ). Denote the Jacobian matrix of f at 0 by A, and let $(\mathbf{R}^n, \varphi_A)$ be the flow generated by the equation $dx/dt = Ax$. If 0 is a hyperbolic singular point, then in a sufficiently small neighborhood of 0, the flows (D, φ) and $(\mathbf{R}^n, \varphi_A)$ are flow equivalent and hence topologically equivalent (Hartman, D. M. Grobman). Let $\lambda_1, \ldots, \lambda_n$ (possibly repeated) be the eigenvalues of A. If f is of class C^∞ and $\lambda_i \neq m_1\lambda_1 + \ldots + m_n\lambda_n$ for all $1 \leqslant i \leqslant n$ and for all nonnegative integers $m_1, \ldots, m_n$ with $2 \leqslant m_1 + \ldots + m_n$, then in a sufficiently small neighborhood of 0, two flows (D, φ) and $(\mathbf{R}^n, \varphi_A)$ are C^∞-equivalent (Sternberg).

(6) Let M be a paracompact differentiable manifold of class C^∞ and V a vector field of class C^r $(1 \leqslant r \leqslant \infty)$ on M. For simplicity we assume that V generates a C^r-flow (M, φ). For each point p of M, take a coordinate neighborhood (U, α) of class C^∞ around p, where U is an open neighborhood of p in M and $\alpha: U \to \mathbf{R}^n$ is a homeomorphism onto an open set D in $\mathbf{R}^n$. Using (U, α), we can express the vector field V as a system of ordinary differential equations of the form (6) in D. The eigenvalues $\lambda_1, \ldots, \lambda_n$ (possibly repeated) of the Jacobian matrix of f at $\alpha(p)$ are independent of the choice of local coordinates (U, α) around p. Thus $\lambda_1, \ldots, \lambda_n$ are invariants of the C^r-equivalence, but they are not invariants of the topological equivalence.

A singular point p of a flow (M, φ) (or a vector field V) is called **hyperbolic** if $\alpha(p)$ is a hyperbolic singular point of the corresponding equation (6) (i.e., none of the real parts of the above eigenvalues $\lambda_1, \ldots, \lambda_n$ are zero). A singular point of (M, φ) is hyperbolic if and only if it is a hyperbolic fixed point of the time-one mapping φ_1 of φ. A hyperbolic singular point p is an isolated singular point.

Let p be a hyperbolic singular point of (M, φ) and $T_p(M) = E^s \oplus E^u$ the direct sum decomposition with respect to φ_1. Put $s = \dim E^s$ and $u = \dim E^u$. Then s (resp. u) is the number of all eigenvalues λ_i of negative (resp. positive) real parts. Put $W^s(p) = \{x \in M \mid \varphi_t(x) \to p \text{ as } t \to \infty\}$ and $W^u(p) = \{x \in M \mid \varphi_{-t}(x) \to p \text{ as } t \to \infty\}$. $W^s(p)$ and $W^u(p)$ are invariant sets and images of injective immersions of class C^r of vector spaces E^s and E^u, respectively. At the point p, $W^s(p)$ and $W^u(p)$ are tangent to E^s and E^u, respectively. We call $W^s(p)$ (resp. $W^u(p)$) the **stable** (resp. **unstable**) manifold of f at p. The stable manifold and the unstable manifold of (M, φ) at a hyperbolic singular point p coincide with those of the time-one mapping φ_1 at p. A hyperbolic singular point p is called a **source** if $\dim W^s(p) = 0$ and a **sink** if $\dim W^u(p) = 0$. Otherwise it is called a **saddle point**. If p is a source (resp. a sink), then $W^u(p)$ (resp. $W^s(p)$) is a neighborhood of p. A singular point p is a source (resp. a sink) if and only if all real parts of the corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$ are positive (resp. negative).

(7) Let V be a vector field of class C^r on a C^∞-differentiable manifold M of dimension $n \geqslant 2$ and (M, φ) the C^r-flow generated by V. Let C be a closed orbit of (M, φ), p a point of C, and $T > 0$ its smallest positive period. Then p is a fixed point of the C^r-diffeomorphism $\varphi_T: M \to M$. Let $V(p) \in T_p(M)$ be the value of V at p and $(d\varphi_T)_p: T_p(M) \to T_p(M)$ the differential of φ_T at p. Then $(d\varphi_T)_p(V(p)) = V(p)$, and $V(p) \neq 0$. Therefore 1 is an eigenvalue of $(d\varphi_T)_p$. The other eigenvalues $\lambda_1, \ldots, \lambda_{n-1}$ (possibly repeated) of $(d\varphi_T)_p$ do not depend on a choice of $p \in C$ and are called the **characteristic multipliers** of C. If none of the characteristic multipliers of C is of absolute value 1, we call C a **hyperbolic closed orbit**. If C is a hyperbolic closed orbit, then there exist vector subspaces E_p^s and E_p^u for each $p \in C$ satisfying the following conditions: (i) $T_p(M) = L(V(p)) \oplus E_p^s \oplus E_p^u$, where $L(V(p))$ is the 1-dimensional subspace generated by $V(p)$; (ii) $(d\varphi_t)_p(E_p^\sigma) = E_{\varphi_t(p)}^\sigma$ $(\sigma = s, u)$ for all $t \in \mathbf{R}$, where $(d\varphi_t)_p: T_p(M) \to T_{\varphi_t(p)}(M)$. In particular, $(d\varphi_T)_p(E_p^\sigma) = E_p^\sigma$ $(\sigma = s, u)$; (iii) $\dim E_p^s$ (resp. $\dim E_p^u$) is independent of $p \in C$ and is equal to the number of λ_i of absolute value < 1 (resp. > 1). Put $W^s(C) = \{x \in M \mid d(\varphi_t(x), C) \to 0 \text{ as } t \to \infty\}$ and $W^u(C) = \{x \in M \mid d(\varphi_{-t}(x), C) \to 0 \text{ as } t \to \infty\}$. Then $W^s(C)$ (resp. $W^u(C)$) is an invariant set and is an injectively immersed C^r-submanifold of M which is tangent to $L(V(p)) \oplus E_p^s$ (resp. $L(V(p)) \oplus E_p^u$) for each $p \in C$. $W^s(C)$ (resp. $W^u(C)$) is called the **stable** (resp. **unstable**) **manifold** for C.

Let p be a point of a closed orbit C. An embedded $(n-1)$-dimensional disk D of class C^r in M containing p is called a **cross section for a closed orbit** C if V is transverse to D (i.e., $T_x(M) = L(V(x)) \oplus T_x(D)$ for each $x \in D$) and $D \cap C = \{p\}$. For a given cross section D for C, there exists a neighborhood U of p in D such that for any $x \in U$ there exists a $\tau(x) > 0$ such that $\varphi(x, \tau(x)) \in D$ and $\varphi(x, t) \notin D$ for $0 < t < \tau(x)$. A mapping $f: U \to D$, defined by $f(x) = \varphi(x, \tau(x))$, $x \in U$, is called a **Poincaré mapping** for D. The C^r-conjugacy class of the †germ of a Poincaré mapping $f: V \to D$ at p is independent of the choice of $p \in C$ and the cross section D for the closed orbit C. The point p is a fixed point of f, and the eigenvalues of df_p coincide with the characteristic multipliers of C including multiplicity. Therefore C is hyperbolic if and only if $p \in C$ is a hyperbolic fixed point of f. In a sufficiently small neighborhood of C, the flow is C^r-equivalent to the suspension of f. The topological equivalence of the flow in a sufficiently small neighborhood of C is determined by the topological conjugacy class of f. The number of topological equivalence classes of hyperbolic closed orbits which appear in a

flow on an n-dimensional manifolds is $4(n-1)$ (M. C. Irwin).

(8) Let V be a vector field of class C^r ($1 \leqslant r \leqslant \infty$) on a compact C^∞-manifold M. If M has a nonempty boundary, we assume that V points outward at all boundary points. Let p be an isolated singular point of V and denote by $i(p)$ the †index of the vector field V at the singular point P. If all singular points are isolated, then the sum $\sum_p i(p)$ is independent of V and is equal to the †Euler-Poincaré characteristic $\chi(M)$ of M (Poincaré, H. Hopf) (→ 153 Fixed-Point Theorems). A vector field (or a flow) is called **nonsingular** if it has no singular points. If $\chi(M) \neq 0$, then any vector field (and hence any flow) on M has a singular point. If $\chi(M)=0$, then there exists a nonsingular vector field on M. If M is 2-dimensional and without boundary, then M admits a nonsingular vector field if and only if M is a †torus or a †Klein bottle.

There are many directions in which generalizations of the Poincaré-Hopf theorem can be made. For example, an index of an isolated closed orbit of a flow has been defined and investigated by F. B. Fuller.

The phase portrait near a singular point or a closed orbit which is not hyperbolic is complicated. Many results in general situations have been obtained for planer flows, but there are only a few for higher-dimensional flows.

H. Generic Properties and Structural Stability

(1) Let M be a compact C^∞-manifold without boundary. The set $\mathscr{X}^r(M)$ of all vector fields of class C^r ($1 \leqslant r \leqslant \infty$) on M forms a real vector space in a natural way. We can give a norm $\|V\|_r$ (called a **C^r-norm**) for $V \in \mathscr{X}^r(M)$ using its expressions in a given suitable system of local coordinates of class C^∞ on M and their partial derivatives up to order r. By virtue of this norm, $\mathscr{X}^r(M)$ is a Banach space with the topology of uniform C^r-convergence (→ 168 Function Spaces). A subset of a topological space is called a **residual set** or a **Baire set** if it is the intersection of a countable number of dense open sets. A residual set in $\mathscr{X}^r(M)$ is a dense set. A proposition P concerning a vector field of class C^r is called **generic** if the set $\{V \in \mathscr{X}^r(M) \mid P(V)\}$ contains a residual set for any M. The following properties are generic properties: (i) All singular points and all closed orbits are hyperbolic [21, 22]; (ii) any stable manifold (of a hyperbolic singular point or a hyperbolic closed orbit) meets transversally with any unstable manifold (of a hyperbolic singular point or a hyperbolic closed orbit) at any point of their intersection [21, 22]; (iii) the set of all periodic points (i.e., the union of all singular points and all closed orbits) is dense in the nonwandering set Ω for $V \in \mathscr{X}^1(M)$ [23]; (iv) there is no regular first integral for $v \in \mathscr{X}^1(M)$, where a **regular first integral** of a vector field V on M is a C^1-function $f: M \to \mathbf{R}$ such that f is not constant on any open set of M and is constant along any orbit of (the flow generated by) V (R. C. Robinson).

Let V (resp. V') be an element of $\mathscr{X}^r(M)$ (resp. $\mathscr{X}^r(M')$), (M, φ) (resp. (M', φ')) the flow generated by V (resp. V'), and Ω (resp. Ω') the nonwandering set of (M, φ) (resp. (M', φ')). V and V' are called **topologically equivalent** (resp. **Ω-equivalent**) if (M, φ) and (M', φ') (resp. the restrictions $(\Omega, \varphi \mid \Omega \times \mathbf{R})$ and $(\Omega', \varphi' \mid \Omega' \times \mathbf{R})$) are topologically equivalent. A vector field $V \in \mathscr{X}^r(M)$ (or the flow generated by V) is called **C^r-structurally stable** (resp. **C^r-Ω-stable**) if there exists a neighborhood $\mathscr{N}$ of V in $\mathscr{X}^r(M)$ such that any V' in $\mathscr{N}$ is topologically equivalent (resp. Ω-equivalent) to V. If $V \in \mathscr{X}^r(M)$ is structurally stable (resp. Ω-stable), then the topological structure of the phase portrait of (the flow generated by) V in the whole space (resp. the nonwandering set) remains invariant under a sufficiently small C^r-perturbation of V. The generic properties (i)–(iv) above hold for C^1-structurally stable vector fields (Markus, Thom, Peixoto, J. Arraut). The generic properties (i) and (iii) above hold for C^1-Ω-stable vector fields.

(2) Let M be a compact C^∞-manifold without boundary. Let $F^r(M)$ be the set of all C^r-mappings of M into itself with the topology of uniform C^r-convergence. Let $\mathrm{Diff}^r(M)$ be the subset of $F^r(M)$ consisting of all C^r-diffeomorphisms of M onto itself. Then $F^r(M)$ is a complete metric space, and $\mathrm{Diff}^r(M)$ is open in $F^r(M)$. Thus $\mathrm{Diff}^r(M)$ is a †Baire space and also a topological group with respect to this topology. A proposition P concerning $f \in \mathrm{Diff}^r(M)$ is called **generic** if the set $\{f \in \mathrm{Diff}^r(M) \mid P(f)\}$ contains a residual set. The following properties are generic: (i) Every periodic point is hyperbolic [21]; (ii) for each pair of hyperbolic periodic points p and q, the stable manifold $W^s(p)$ intersects the unstable manifold $W^u(q)$ transversally [21]; (iii) for a C^1-diffeomorphism $f \in \mathrm{Diff}^1(M)$ the set of all periodic points is dense in the nonwandering set ([23], Robinson).

Let $f: M \to M$ and $g: N \to N$ be C^r-diffeomorphisms and $\Omega(f)$ and $\Omega(g)$ their nonwandering sets. f and g are called **Ω-conjugate** if $f \mid \Omega(f): \Omega(f) \to \Omega(f)$ and $g \mid \Omega(g): \Omega(g) \to \Omega(g)$ are topologically conjugate. A diffeomorphism $f \in \mathrm{Diff}^r(M)$ is called **C^r-structurally stable** (resp. **C^r-Ω-stable**) if there exists a neighborhood $\mathscr{N}$ of f in $\mathrm{Diff}^r(M)$ such that any g in $\mathscr{N}$ is topologically conjugate (resp. Ω-conjugate) to f. Generic properties (i)–

(iii) above hold for C^1-structurally stable diffeomorphisms, and (i) and (iii) above hold for C^1-Ω-stable diffeomorphisms.

I. Low-Dimensional Systems

(1) Let $S^1=\{z\in\mathbf{C}\mid |z|=1\}$ be the unit circle in the complex plane $\mathbf{C}$ and $p:\mathbf{R}\to S^1$ the †covering projection defined by $p(x)=e^{2\pi ix}$, $x\in\mathbf{R}$. Let $f:S^1\to S^1$ be an †orientation-preserving homeomorphism. Then f can be lifted to a mapping $F:\mathbf{R}\to\mathbf{R}$ satisfying the following conditions: (i) $p\circ F=f\circ p$; (ii) F is monotone increasing; (iii) $F-1_{\mathbf{R}}$ is a periodic function of period 1, where $1_{\mathbf{R}}$ is the identity mapping of $\mathbf{R}$. The limit $\rho(f)=\lim_{n\to\infty}F^n(x)/n$ exists for all $x\in\mathbf{R}$, and its †residue class modulo $\mathbf{Z}$ is independent of F and x. We call $\rho(f)$ the **rotation number** of f. Let $f, g:S^1\to S^1$ be orientation-preserving homeomorphisms. If f and g are topologically conjugate by an orientation-preserving (resp. †reversing) homeomorphism, then $\rho(f)\equiv\rho(g)$ (resp. $\rho(f)\equiv-\rho(g)$) modulo $\mathbf{Z}$. An orientation preserving homeomorphism $f:S^1\to S^2$ has a periodic point of the smallest positive period s if and only if $\rho(f)=r/s$, where r and s are relatively prime integers. In this case all periodic points are of the smallest positive period s. If $\rho(f)$ is irrational, then the ω-limit set $\omega(x)$ of $x\in S^1$ is independent of x, and $E=\omega(x)$ is either †perfect and †nowhere dense or the whole space S^1. If $\rho(f)$ is irrational and $E=\omega(x)=S^1$, then f is called **transitive**. If f is transitive, then it is topologically conjugate to the rotation $r_{\rho(f)}:S^1\to S^1$ defined by $r_{\rho(f)}(e^{2\pi ix})=e^{2\pi i(x+\rho(f))}$, $x\in\mathbf{R}$. Let $f:S^1\to S^1$ be of class C^1 with $\rho(f)$ irrational. If its derivative f' is of †bounded variation, then f is transitive (A. Denjoy). In particular, if f is of class C^2 with $\rho(f)$ irrational, then f is topologically conjugate to the rotation $r_{\rho(f)}$. However, there are C^1-diffeomorphisms of S^1 onto itself whose nonwandering sets are not the whole space. Those C^1-diffeomorphisms are never topologically conjugate to C^2-diffeomorphisms. M. R. Herman gave a sufficient condition for a diffeomorphism of S^1 onto itself to be differentiably conjugate to a rotation.

A C^r-diffeomorphism $f:S^1\to S^1$ is structurally stable if and only if $\Omega(f)$ is finite (hence $\Omega(f)$ consists of a finite number of periodic points), and all periodic points are hyperbolic. The set of all structurally stable C^r-diffeomorphisms of S^1 onto itself is a dense open set in $\mathrm{Diff}^r(S')$ (Peixoto).

(2) Let D be an open set in $\mathbf{R}^2$ and $f:D\to\mathbf{R}^2$ a continuous mapping. We assume that for each $x\in D$ there exists a unique nonextendable solution $\varphi(x,t)$ with the initial condition $\varphi(x,0)=x$ defined on (a_x,b_x) for the equation $dx/dt=f(x)$, $x\in D$. Suppose that for a given point $x\in D$ there exists a compact set K in D containing the positive semiorbit $C_+(x)=\{\varphi(x,t)\mid 0\leqslant t<b_x\}$ of x. Then $b_x=\infty$. Further, we assume that the ω-limit set $\omega(x)$ of x contains no singular points. Then we have either (i) $C_+(x)=\omega(x)$ and it is a closed orbit, or (ii) $C_+(x)\neq\omega(x)$ and $\omega(x)$ is a closed orbit. In case (ii), $\omega(x)$ is a †simple closed curve and $C_+(x)$ is a spiral which tends the closed orbit $\omega(x)$ (Poincaré, I. Bendixson). We call such a closed orbit $\omega(x)$ a **limit cycle**. Let f_1 and f_2 be polynomials of two variables and m the maximum of degrees of f_1 and f_2. Let $f=(f_1,f_2):\mathbf{R}^2\to\mathbf{R}^2$ be the mapping defined by f_1 and f_2. The equation $dx/dt=f(x)$, $x\in\mathbf{R}^2$, defined by such an f is called a polynomial system of degree m. The following is Hilbert's 16th problem: Is there a number $N(m)$ depending only on m such that the number of limit cycles for any polynomial system of degree m is bounded by $N(m)$?

Let M be a closed (i.e., compact, without boundary) C^∞-manifold of dimension 2 and (M,φ) a C^2-flow on M. Then a minimal set of (M,φ) is either (i) a singular point, (ii) a closed orbit, or (iii) the whole space M. For case (iii), M is a torus (Poincaré, Denjoy, C. L. Siegel, A. J. Schwartz). Let T be a 2-dimensional torus and (T,φ) a C^r-flow. Suppose that (T,φ) has a cross section Σ which is C^1-diffeomorphic to S^1. Let $f:\Sigma\to\Sigma$ be the Poincaré mapping for Σ. Then (T,φ) has a closed orbit if and only if the rotation number $\rho(f)$ of f is rational. If $\rho(f)$ is irrational and (T,φ) is of class C^2, then T is a minimal set.

Let M be an orientable 2-dimensional closed C^∞-manifold and $V\in\mathscr{X}^1(M)$. Then V is structurally stable if and only if the following conditions are satisfied: (i) There are only a finite number of singular points, all hyperbolic; (ii) There are only a finite number of closed orbits, all hyperbolic; (iii) There are no orbits which connect saddle points; (iv) The $\alpha(x)$ and $\omega(x)$ for any $x\in M$ are singular points or closed orbits (Peixoto). The above theorem was first proved by Andronov and Pontryagin for analytic vector fields on a 2-dimensional disk which are transverse to the boundary of the disk at any boundary point. The set of all structurally stable vector fields of class C^1 on M is a dense open set in $\mathscr{X}^1(M)$ (Peixoto).

J. Axiom A Systems

In this section we assume that phase spaces are closed C^∞-manifolds with metric d and $1\leqslant r\leqslant\infty$ unless stated otherwise.

(1) A vector field $V \in \mathscr{X}^r(M)$ (resp. a C^r-flow (M, φ)) is called a **Morse-Smale vector field** (resp. a **Morse-Smale flow**) if the following conditions are satisfied: (i) The nonwandering set is the union of a finite number of singular points and a finite number of closed orbits; (ii) The singular points and closed orbits are all hyperbolic; (iii) The stable manifolds and the unstable manifolds of the singular points and closed orbits intersect each other transversely. If $\dim M = 2$, the Morse-Smale vector fields on M are exactly the structurally stable ones discussed in Section I(2). A Morse-Smale vector field is structurally stable [25]. The set of all Morse-Smale vector fields on M is open but not dense in $\mathscr{X}^r(M)$ if $\dim M > 2$ (Palis, Smale). However, it contains a dense open subset of the set of all †gradient vector fields with respect to a given †Riemannian metric (Smale). In particular, any closed manifold admits Morse-Smale vector fields and hence structurally stable vector fields. For a Morse-Smale vector field, the †Morse inequalities hold as they do in the †calculus of variations in the large [24].

(2) A vector field $V \in \mathscr{X}^r(M)$ (or the C^r-flow (M, φ) generated by V) is called an **Anosov vector field** (or an **Anosov flow**) if the following conditions are satisfied: (i) There is a direct sum decomposition $T_x(M) = L(V(x)) \oplus E_x^s \oplus E_x^u$ of the tangent space $T_x(M)$ for each $x \in M$ which depends continuously on $x \in M$; (ii) $(d\varphi_t)_x(E_x^s) = E_{\varphi_t(x)}^s$ and $(d\varphi_t)_x(E_x^u) = E_{\varphi_t(x)}^u$ for all $x \in M$ and $t \in \mathbf{R}$; (iii) There are a Riemannian metric on M and constants $c, \lambda > 0$ such that, for all $t > 0$ and $x \in M$, $\|(d\varphi_t)_x(w)\| \leqslant ce^{-\lambda t}\|w\|$ when $w \in E_x^s$, and $\|(d\varphi_{-t})_x(w)\| \leqslant ce^{-\lambda t}\|w\|$ when $w \in E_x^u$, where $\|\cdot\|$ is the norm induced by the Riemannian metric. The suspensions of Anosov diffeomorphisms and the †geodesic flows on Riemannian manifolds of negative curvature are important examples of Anosov flows (J. Hadamard [8]). There are examples of Anosov flows other than the ones stated above (M. Handel and W. P. Thurston, J. Franks and R. F. Williams). The following have been proved by Anosov [8]: (i) An Anosov flow is structurally stable; (ii) there are countably many closed orbits for an Anosov flow; (iii) if there exists a **smooth invariant measure** (i.e., an †invariant measure which has a †smooth density with respect to the measure associated with the Riemannian metric), then the set of all closed orbits is dense in M. If we assume further that the flow is of class C^2, then it is †ergodic; (iv) the set of all Anosov vector fields is open in $\mathscr{X}^r(M)$; (v) $\{E_x^s\}$, $\{E_x^u\}$ $(x \in M)$ define †foliations on M, which are called **Anosov foliations**.

(3) A vector field $V \in \mathscr{X}^r(M)$ (or the C^r-flow (M, φ) generated by V) is called an **Axiom A vector field** (or an **Axiom A flow**) if the following conditions are satisfied: A(a) The nonwandering set consists of a finite set F of singular points, all hyperbolic, and the closure Λ of the union of closed orbits, and $F \cap \Lambda = \varnothing$; A(b) The conditions (i)–(iii) in the definition of Anosov flow in which we replace the terms "$x \in M$" by "$x \in \Lambda$." For each $x \in M$, put $W^s(x) = \{y \in M \mid d(\varphi_t(x), \varphi_t(y)) \to 0 \text{ as } t \to \infty\}$ and $W^u(x) = \{y \in M \mid d(\varphi_{-t}(x), \varphi_{-t}(y)) \to 0 \text{ as } t \to \infty\}$. We call $W^s(x)$ (resp. $W^u(x)$) the **stable** (resp. **unstable**) **manifold** of V at x. For a subset A of M, put $W^\sigma(A) = \bigcup_{x \in A} W^\sigma(x)$ $(\sigma = s, u)$. For $x \in M$, put $W^{w\sigma}(x) = W^\sigma(C(x))$ $(\sigma = s, u)$, where $C(x)$ is the orbit of x. If the flow satisfies Axiom A, then $W^s(x)$, $W^u(x)$, $W^{ws}(x)$, and $W^{wu}(x)$ are injectively immersed submanifolds for all $x \in M$ (M. W. Hirsch and C. C. Pugh). If $x \in M$ is a hyperbolic singular point, then $W^s(x)$ and $W^u(x)$ defined above coincide with those defined before. If $C(x)$ is a hyperbolic closed orbit, then $W^s(C(x))$ and $W^u(C(x))$ defined above coincide with those defined before.

For an Axiom A flow, there is a decomposition of the nonwandering set $\Omega = \Omega_1 \cup \ldots \cup \Omega_m$ (disjoint union), where each Ω_i is closed, invariant, and **transitive** (i.e., has a dense orbit), and $M = \bigcup_{i=1}^m W^s(\Omega_i) = \bigcup_{i=1}^m W^u(\Omega_i)$ (disjoint union) [7]. This decomposition is called the **spectral decomposition** of Ω, and each Ω_i is called a **basic set**. Let $\Omega = \Omega_1 \cup \ldots \cup \Omega_m$ be the spectral decomposition of the nonwandering set for an Axiom A flow. Denote $\Omega_i \leqslant \Omega_j$ if $W^s(\Omega_i) \cap W^u(\Omega_j) \neq \varnothing$. A sequence of basic sets $\Omega_{i_0}, \Omega_{i_1}, \ldots, \Omega_{i_k}$ $(k > 1)$ is called a **cycle** if $\Omega_{i_0} \leqslant \Omega_{i_1} \leqslant \ldots \leqslant \Omega_{i_k}, \Omega_{i_0} = \Omega_{i_k}$, and otherwise $\Omega_{i_p} \neq \Omega_{i_q}$ for $p \neq q$. An Axiom A flow which has no cycles in the above sense is said to satisfy the **no cycle condition**. The **Ω-stability theorem**: An Axiom A flow with the no cycle condition is Ω-stable [27]. **Ω-explosion**: If an Axiom A flow has a cycle, then it is not Ω-stable (Palis). An Axiom A flow is said to satisfy the **strong transversality condition** if $W^{ws}(x)$ and $W^{wu}(x)$ intersect transversely at any $x \in M$. An Axiom A flow with the strong transversality condition satisfies the no cycle condition [7]. The **structural stability theorem**: An Axiom A flow of class C^1 with the strong transversality condition is structurally stable [29, 31]. Morse-Smale flows and Anosov flows are Axiom A flows with the strong transversality condition. There are Axiom A flows other than Morse-Smale flows and Anosov flows that satisfy the strong transversality condition [7]. **Stability conjecture**: A C^r-flow is structurally stable (resp. Ω-stable) if and only if it is an Axiom A flow with the strong transversality condition (resp. the no cycle condition). S. Newhouse, V. A. Pliss, Robinson, R. Mañé, S. D. Liao,

and A. Sannami have made important contributions to the study of the stability conjecture.

Neither the set of all Axiom A flows nor the set of all Ω-stable flows nor the set of structurally stable flows is dense in $\mathfrak{X}^r(M)$ if $\dim M > 2$ (R. Abraham and Smale, Newhouse). However, the set of all structurally stable flows is dense in $\mathfrak{X}^r(M)$ in the C^0 topology (M. Shub, M. M. C. de Oliverira).

(4) A diffeomorphism $f \in \mathrm{Diff}^r(M)$ is called a **Morse-Smale diffeomorphism** if the following conditions are satisfied: (i) The nonwandering set Ω is a finite set, and hence it consists of periodic points; (ii) all periodic points are hyperbolic; (iii) for each pair $p, q \in \Omega$, $W^s(p)$ intersects $W^u(q)$ transversally. The Morse-Smale diffeomorphisms on the unit circle S^1 are exactly the structurally stable ones described in Section I(1). A Morse-Smale diffeomorphism is structurally stable [25]. The set of all Morse-Smale diffeomorphisms on M is open but not dense in $\mathrm{Diff}^r(M)$ if $\dim M \geqslant 2$ (Palis, Smale). However, it contains a dense open subset of the set of all time-one mappings of the flows generated by gradient vector fields. In particular, any closed manifold admits Morse-Smale diffeomorphisms and hence structurally stable diffeomorphisms. For a Morse-Smale diffeomorphism, the Morse inequalities hold [24].

Let Λ be a closed invariant set of $f \in \mathrm{Diff}^r(M)$. Λ is called **hyperbolic** if the following conditions are satisfied: (i) There is a splitting $T_x(M) = E^s_x \oplus E^u_x$ of the tangent space $T_x(M)$ for each $x \in \Lambda$, which depends continuously on $x \in \Lambda$; (ii) $df_x(E^s_x) = E^s_{f(x)}$ and $df_x(E^u_x) = E^u_{f(x)}$ for all $x \in \Lambda$; (iii) There is a Riemannian metric on M and constants $c > 0$, $0 < \lambda < 1$ such that, for any integer $m > 0$ and $x \in \Lambda$, $\|df^m_x(w)\| \leqslant c\lambda^m \|w\|$ when $w \in E^s_x$ and $\|df^{-m}_x(w)\| \leqslant c\lambda^m \|w\|$ when $w \in E^u_x$. A diffeomorphism $f \in \mathrm{Diff}^r(M)$ is called an **Anosov diffeomorphism** if M itself is hyperbolic for f. Manifolds which admit Anosov diffeomorphisms are restricted (Hirsch, K. Shiraiwa). Examples of Anosov diffeomorphisms are given as follows: Let $L: \mathbf{R}^n \to \mathbf{R}^n$ be a hyperbolic linear automorphism with $L(\mathbf{Z}^n) = \mathbf{Z}^n$, where $\mathbf{Z}^n$ is the discrete subgroup of $\mathbf{R}^n$ consisting of all elements with integral coordinates. Then L induces an automorphism $f: T^n \to T^n$ of the n-dimensional torus $T^n = \mathbf{R}^n/\mathbf{Z}^n$, which is an Anosov diffeomorphism. There are similar constructions using hyperbolic automorphisms of simply connected †nilpotent Lie groups and their uniform discrete subgroups (Smale). A homeomorphism $h: X \to X$ of a metric space X is called **expansive** if there is a constant $\varepsilon > 0$ such that $x, y \in X$ and $d(h^n(x), h^n(y)) < \varepsilon$ for all $n \in \mathbf{Z}$ imply $x = y$. An Anosov diffeomorphism is expansive. For Anosov diffeomorphisms, theorems similar to those stated in Section J(2) hold. Franks, Newhouse, A. Manning, and J. N. Mather obtained important results concerning Anosov diffeomorphisms.

A diffeomorphism $f \in \mathrm{Diff}^r(M)$ is called an **Axiom A diffeomorphism** if the following conditions are satisfied: A(a) The nonwandering set Ω is hyperbolic; A(b) the set of all periodic points is dense in Ω. There are examples that satisfy Axiom A(a) but not Axiom A(b) (A. Dankner, M. Kurata). For $x \in M$, put $W^s(x) = \{y \in M \mid d(f^n(x), f^n(y)) \to 0 \text{ as } n \to \infty\}$ and $W^u(x) = \{y \in M \mid d(f^{-n}(x), f^{-n}(y)) \to 0 \text{ as } n \to \infty\}$. We call $W^s(x)$ (resp. $W^u(x)$) the **stable** (resp. **unstable**) **manifold** of f at x. For a subset A of M, put $W^\sigma(A) = \bigcup_{x \in A} W^\sigma(x)$ $(\sigma = s, u)$. If f is an Axiom A diffeomorphism, $W^s(x)$ and $W^u(x)$ are injectively immersed submanifolds of M. If x is a hyperbolic periodic point, $W^s(x)$ and $W^u(x)$ defined here coincide with those defined before. For Axiom A diffeomorphisms, notions such as spectral decomposition, basic sets, the strong transversality condition, and the no cycle condition are defined similarly, and theorems similar to those stated in Section J(3) hold.

Let $h: X \to X$ be a homeomorphism of a metric space X and $\alpha, \beta > 0$. A sequence $\{x_i\}_{i \in \mathbf{Z}}$ of points in X is an **α-pseudo-orbit** if $d(h(x_i), x_{i+1}) < \alpha$ for all $i \in \mathbf{Z}$. We say that $\{x_i\}_{i \in \mathbf{Z}}$ is **β-shadowed** (or **β-traced**) by a point $x \in X$ if $d(h^i(x), x_i) \leqslant \beta$ for all $i \in \mathbf{Z}$. We say that h has the **pseudo-orbit tracing property** if for any $\beta > 0$ there exists an $\alpha > 0$ so that any α-pseudo-orbit is β-shadowed by some point. If f is an Axiom A diffeomorphism, then $f \mid \Omega(f): \Omega(f) \to \Omega(f)$ has the pseudo-orbit tracing property (Bowen). Axiom A diffeomorphisms with the strong transversality condition (in particular, Anosov diffeomorphisms) have the pseudo-orbit tracing property (Bowen, K. Sawada, K. Kato, A. Morimoto).

(5) Let S be a discrete topological space with n elements $(n \geqslant 2)$ and $\Sigma(S) = \prod_{i \in \mathbf{Z}} S_i$ the doubly infinite product of copies S_i of S with the product topology. A point $x = (x_i)_{i \in \mathbf{Z}}$ of $\Sigma(S)$ is a doubly infinite sequence of points in S. $\Sigma(S)$ is a †totally disconnected, perfect, compact, †metrizable space (i.e., a †Cantor set). Let $\sigma: \Sigma(S) \to \Sigma(S)$ be a mapping defined by $\sigma(x) = y$, $x = (x_i)$, $y = (y_i)$, $y_i = x_{i+1}$ $(i \in \mathbf{Z})$. Then σ is a homeomorphism of $\Sigma(S)$ and is called the **shift automorphism** with n symbols.

Define a diffeomorphism $f: S^2 \to S^2$ of the 2-dimensional sphere S^2 as follows. We consider S^2 as $\mathbf{R}^2 \cup \{\infty\}$. Let U be a region of $\mathbf{R}^2$ consisting of the rectangle R $(= ABCD)$ and an upper and a lower cap as shown in Fig. 1. f maps U into itself as in Fig. 1 ($f(A) = A'$, $f(B) = B'$, $f(C) = C'$, $f(D) = D'$, and so on). Here, $R \cap f(R) = P \cup Q$ is the union of two rectangles,

$P'=f^{-1}(P)$ and $Q'=f^{-1}(Q)$ are rectangles, and f is †affine on each of P' and Q' (stretching in the vertical direction and contracting in the horizontal direction). The lower cap is contracted into itself, and every point in it tends to a sink x_0. The upper cap is mapped into the lower one and stays there. The point ∞ is the only source, and $\lim_{n\to\infty} f^{-n}(x)=\infty$ for all $x\in\mathbf{R}^2-U$. Thus the nonwandering set of f consists of a sink x_0, a source ∞, and $\Lambda=\bigcap_{n\in\mathbf{Z}} f^n(R)$. The mapping f constructed here is called the **horseshoe diffeomorphism**. It is an Axiom A diffeomorphism with the strong transversality condition (and hence structurally stable). The restriction $f|\Lambda$ of f on a basic set Λ is topologically conjugate to the shift automorphism with two symbols. It is neither a Morse-Smale diffeomorphism nor an Anosov diffeomorphism [7].

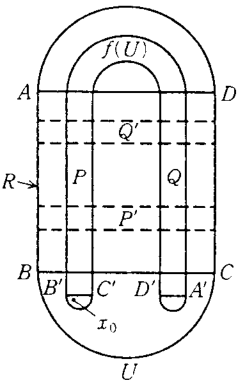


Fig. 1

A restriction of a shift automorphism on a closed invariant set is called a **subshift**. Let $A=(A_{ij})$ be an $n\times n$ matrix with (i,j)-component $A_{ij}=0$ or 1 for all $i,j=1,\ldots,n$. It is **irreducible** if for each i,j there is a positive integer m such that A^m has a nonzero (i,j)-component. Assume that A is irreducible and $S=\{1,\ldots,n\}$. Put $\Sigma_A=\{x=(x_i)\in\Sigma(S)\,|\,A_{x_i x_{i+1}}=1$ for all $i\in\mathbf{Z}\}$. Then Σ_A is a closed invariant set of $\Sigma(S)$. The restriction $\sigma_A=\sigma|\Sigma_A:\Sigma_A\to\Sigma_A$ is called a **subshift of finite type** or a **Markov subshift**, and A is called the **transition matrix**. A topological classification of the subshifts of finite type has been investigated by Williams.

Let $f\in\mathrm{Diff}^r(M)$ be an Axiom A diffeomorphism and Λ its basic set. Bowen constructed a †Markov partition for $f|\Lambda$ by generalizing the Sinaĭ construction for Anosov diffeomorphisms. The Markov partition connects $f|\Lambda$ with a suitable subshift of finite type and is applied to the study of Axiom A diffeomorphisms, especially to the ergodic theory of Axiom A diffeomorphisms (Bowen, Ruelle). A similar theory for flows was developed by Bowen, Ruelle, and M. E. Ratner.

(6) Let $x\in M$ be a hyperbolic fixed point of $f\in\mathrm{Diff}^r(M)$. A point $p\in W^s(x)\cap W^u(x)$ $(p\neq x)$ is called a **homoclinic point**. If $W^s(x)$ and $W^u(x)$ intersect transversally at a homoclinic point p, then p is called a **transversal homoclinic point**. In a neighborhood of a transversal homoclinic point, there is a closed invariant set Λ of f^m for some positive integer m such that $f^m|\Lambda$ is topologically conjugate to the shift automorphism with two symbols (Smale). There are generalizations of this theorem for semiflows by F. R. Marotto, Shiraiwa, and Kurata.

K. Topological Entropy and Zeta Functions

(1) The notion of topological entropy was first defined by R. L. Adler, A. G. Konheim, and M. H. McAndrew as an analog to measure-theoretic †entropy. Let X be a compact topological space and α an †open covering of X. Let $N(\alpha)$ be the minimum number of members of a †subcovering of α. Let $f:X\to X$ be a continuous mapping and α an open covering of X. Then $\lim_{n\to\infty}(1/n)\log N(\alpha\vee f^{-1}\alpha\vee\ldots\vee f^{-n+1}\alpha)$ exists, where $\alpha\vee f^{-1}\alpha\vee\ldots\vee f^{-n+1}\alpha$ is the open covering $\{A_0\cap f^{-1}(A_1)\cap\ldots\cap f^{-n+1}(A_{n-1})\,|\,A_0, A_1,\ldots,A_{n-1}\in\alpha\}$. We denote the above limit by $h(f,\alpha)$ and call it the **topological entropy of** f **with respect to** α. The **topological entropy** $h(f)$ of f is defined as the $\sup h(f,\alpha)$, where the supremum is taken over all open coverings α of X. Now assume that X is a compact metric space. For an open covering α of X, put $\operatorname{diam}\alpha=\sup\{\operatorname{diam}A\,|\,A\in\alpha\}$, where $\operatorname{diam}A$ is the †diameter of A. If $\{\alpha_n\}_{n\geq 1}$ is a sequence of open coverings of X such that $\operatorname{diam}\alpha_n\to 0$ as $n\to\infty$, then $h(f,\alpha_n)\to h(f)$ as $n\to\infty$. The topological entropy of the shift automorphism with n symbols is equal to $\log n$. Let $L:\mathbf{R}^n\to\mathbf{R}^n$ be a linear mapping with $L(\mathbf{Z}^n)\subset L(\mathbf{Z}^n)$ and $\lambda_1,\ldots,\lambda_n$ its eigenvalues. Let $f:T^n\to T^n$ be the induced endomorphism of the n-dimensional torus $T^n=\mathbf{R}^n/\mathbf{Z}^n$. Then the topological entropy is given by $h(f)=\sum_{|\lambda_i|>1}\log|\lambda_i|$.

Let $f:X\to X$ be a continuous mapping and $M(f)$ the set of all †f-invariant probability measures on the †Borel sets of X. For each $\mu\in M(f)$, denote by $h_\mu(f)$ the (measure-theoretic) entropy of f with respect to μ. Then $h(f)=\sup\{h_\mu(f)\,|\,\mu\in M(f)\}$ (E. I. Dinaburg, T. N. T. Goodman, L. W. Goodwyn). The following properties hold: (i) If $f:X\to X$ is a homeomorphism, then $h(f^n)=|n|h(f)$ for all $n\in\mathbf{Z}$. (ii) If (X,φ) is a continuous flow, then $h(\varphi_t)=|t|h(\varphi_1)$ for all $t\in\mathbf{R}$. (iii) Let $f_i:X_i\to X_i$ $(i=1,2)$ be continuous. Then $h(f_1\times f_2)=h(f_1)+h(f_2)$. (iv) Let $f_i:X_i\to X_i$ $(i=1,2)$ be continuous. If there is a continuous mapping $g:X_1\to X_2$ such that $g(X_1)=X_2$ and $g\circ f_1=f_2\circ g$, then $h(f_1)\geq h(f_2)$. In particular, if f_1 and f_2 are topologically conjugate, then $h(f_1)=$

$h(f_2)$. (v) If $f: X \to X$ is a homeomorphism and Ω is the nonwandering set of f, then $h(f) = h(f|\Omega)$. In particular, Ω-conjugate homeomorphisms have the same topological entropy (Bowen). (vi) If $f: M \to M$ is an Axiom A diffeo morphism of a closed C^∞-manifold M, then $h(f) = \limsup_{n \to \infty} (1/n) \log N_n(f)$, where $N_n(f)$ is the number of all periodic points of period n. And $h(f) = 0$ if and only if the nonwandering set is finite (Bowen).

Bowen gave alternative definitions of the topological entropy and generalized it for uniformly continuous mappings of (not necessarily compact) metric spaces. If $f: M \to M$ is a C^1-mapping of an n-dimensional Riemannian manifold M, then $h(f) \leqslant \max\{0, n \log \sup\{\|df_x\| \mid x \in M\}\}$, where $\|df_x\|$ is the norm of $df_x: T_x(M) \to T_{f(x)}(M)$, and hence $h(f)$ is finite (Bowen, S. Ito).

(2) Let $f: M \to M$ be a continuous mapping of a closed C^∞-manifold M and $f_*: H_*(M) \to H_*(M)$ (resp. $f_{*1}: H_1(M) \to H_1(M)$) the †induced homomorphism if f on the †homology group $H_*(M)$ (resp. the †first homology group $H_1(M)$) with coefficients in $\mathbf{R}$. The **spectral radius** $s(L)$ of a linear mapping $L: E \to E$ of a real vector space E is the maximum of the absolute values of the eigenvalues of L. The following **entropy conjecture** is still open: If $f: M \to M$ is a C^1-diffeomorphism (C^1-mapping), then $h(f) \geqslant \log s(f_*)$. Concerning the entropy conjecture, the following are known: (i) The conjecture holds for a dense open set of $\mathrm{Diff}^r(M)$ in the C^0-topology (Shub); (ii) for any continuous mapping f, $h(f) \geqslant \log s(f_{*1})$ (Manning); (iii) the conjecture fails for a homeomorphism (Pugh); (iv) the conjecture holds for an Axiom A diffeomorphism with the no cycle condition (Shub and Williams); (v) the conjecture holds for any continuous mapping of the n-dimensional torus (M. Misiurewicz, F. Przytycki); (vi) for a C^1-mapping $f: M \to M$, $h(f) \geqslant \log |\deg f|$, where $\deg f$ is the †mapping degree of f (Misiurewicz, Przytycki).

(3) M. Artin and B. Mazur first defined the zeta function of a diffeomorphism by analogy to †Weil's zeta function. Let $f: X \to X$ be a homeomorphism of a compact metric space X. Assume that the number $N_m = N_m(f)$ of all periodic points of period m is finite for all m. Put $\zeta(t) = \exp(\sum_{m=1}^\infty N_m t^m / m)$ and call it the **zeta function** of f. For any closed C^∞-manifold M, there is a dense set E of $\mathrm{Diff}^r(M)$ such that $\tilde{N}_m(f) \leqslant ck^m$ for $f \in E$, where $\tilde{N}_m(f)$ is the number of isolated periodic points of period m and c and k are positive constants depending only on f (Artin, Mazur). Hence, for such $f \in E$, the series $\exp(\sum_{m=1}^\infty \tilde{N}_m(f) t^m / m)$ has a positive radius of convergence. Originally, Artin and Mazur called this the zeta function of f.

Let $\sigma_A: \Sigma_A \to \Sigma_A$ be a subshift of finite type with transition matrix A; then $\zeta(t) = 1/\det(E - tA)$, where E is the unit matrix (Bowen and O. E. Lanford III). The zeta function of an Axiom A diffeomorphism has a positive radius of convergence and is a rational function (K. Meyer, J. Guckenheimer, Manning).

Let (M, φ) be a nonsingular C^r-flow on a closed C^∞-manifold M. Let Γ be the set of all closed orbits and $l(\gamma)$ the smallest positive period of $\gamma \in \Gamma$. Smale defined the **zeta function** of (M, φ) by $Z(s) = \prod_{\gamma \in \Gamma} \prod_{k=0}^\infty (1 - [\exp l(\gamma)]^{-s-k})$. If (M, φ) is the geodesic flow on a surface of constant negative curvature, $Z(s)$ reduces to the †Selberg zeta function.

There are generalizations and modifications of the notion of zeta function by Ruelle and Franks.

L. Classical Dynamical Systems

(1) Let M be a C^∞-manifold without boundary. A C^r-flow (M, φ) or a C^r-diffeomorphism $f: M \to M$ with a smooth invariant measure is called a **classical dynamical system**. Most important classical dynamical systems are Hamiltonian or Lagrangian systems. In the modern formulation, a **Hamiltonian system** consists of M, a **symplectic form** ω on M (i.e., a †nondegenerate †closed 2-form), and the vector field X_H on M defined by a C^{r+1}-function $H: M \to \mathbf{R}$. We call X_H the **Hamiltonian vector field** with **energy function** H. Let (M, ω, X_H) be a Hamiltonian system. Then the following hold: (i) M is of even dimension and there is a system of local coordinates $(q^1, \ldots, q^n, p_1, \ldots, p_n)$ such that $\omega = \sum_{i=1}^n dq^i \wedge dp_i$ (J. G. Darboux). In these coordinates X_H is expressed by †Hamilton's equations $dq^i/dt = \partial H / \partial p_i$, $dp_i/dt = -\partial H / \partial q^i$, $i = 1, \ldots, n$; (ii) the smooth measure defined by the †volume element $\Omega = ((-1)^{[n/2]}/n!)\omega^n$ is an invariant measure for the flow generated by X_H (J. Liouville); (iii) the energy function H is constant along any trajectory of X_H. Especially, $H^{-1}(e)$ ($e \in \mathbf{R}$) is an invariant set for each e and is called an **energy surface**. Energy surfaces are submanifolds of codimension one for almost all $e \in \mathbf{R}$.

Important examples of Hamiltonian systems are given as follows. Let Q be an n-dimensional manifold and $T^*(Q)$ the †cotangent bundle of Q. Then $T^*(Q)$ has a canonical symplectic form ω_0. We call Q a **configuration space** and $T^*(Q)$ a **momentum phase space**. For any C^{r+1}-function H on $T^*(Q)$, we have a Hamiltonian system $(T^*(Q), \omega_0, X_H)$ in which X_H is of class C^r. The Hamiltonian formalism is translated into Lagrangian formalism by using the †tangent bundle $T(Q)$ instead of $T^*(Q)$. $T(Q)$ is called a **velocity phase space**.

For a given differentiable function L on $T(Q)$, the **energy function** E on $T(Q)$ and the **Lagrangian vector field** X_E on $T(Q)$ are constructed. If L satisfies a certain condition, there is a system of local coordinates $(q^1, \ldots, q^n, \dot{q}^1, \ldots, \dot{q}^n)$ such that X_E is expressed in these coordinates by the †Euler-Lagrange equations $dq^i/dt = \dot{q}^i$, $d/dt(\partial L/\partial \dot{q}^i) = \partial L/\partial q^i$, $i = 1, \ldots, n$. The projection of an integral curve (i.e., an orbit) of X_E into Q is called a base integral curve. Under a suitable condition, the base integral curve of a Lagrangian (or a Hamiltonian) system is the geodesic of the Jacobi metric on Q up to a reparametrization, and the restriction of the flow generated by X_X on an energy surface is the geodesic flow described below.

(2) Let M be a †complete Riemannian manifold of class C^∞ and $S(M) = \{v \in T(M) \mid \|v\| = 1\}$. Let $\pi: S(M) \to M$ be the projection defined by $\pi(v) = x$ for $v \in T_x(M)$. Then $S(M)$ is a †sphere bundle over M, which is called the **unit tangent sphere bundle** over M. Each $v \in S(M)$ determines a unique †geodesic $C_v: \mathbf{R} \to M$ such that $C_v(0) = \pi(v)$ and the †tangent vector $C_v'(0)$ to C_v at $t = 0$ is equal to v. Let v_t be the tangent vector $C_v'(t)$ to C_v at t for $t \in \mathbf{R}$. Then $\|v_t\| = 1$ and $\pi(v_t) = C_v(t)$ for all $t \in \mathbf{R}$. Define a mapping $\varphi: S(M) \times \mathbf{R} \to S(M)$ by $\varphi(v, t) = v_t$. Then $(S(M), \varphi)$ is a C^∞-flow, which is called the **geodesic flow** on M. By the classical Liouville theorem, a geodesic flow has a smooth invariant measure.

(3) Let $T^n = \mathbf{R}^n/\mathbf{Z}^n$ be the n-dimensional torus and $\omega_1, \ldots, \omega_n$ real constants. Define a mapping $\varphi: T^n \times \mathbf{R} \to T^n$ by $\varphi([x_1, \ldots, x_n], t) = [x_1 + \omega_1 t, \ldots, x_n + \omega_n t]$, where $[x_1, \ldots, x_n] \in \mathbf{R}^n/\mathbf{Z}^n = T^n$ is the residue class of $(x_1, \ldots, x_n) \in \mathbf{R}^n$ modulo $\mathbf{Z}^n$. Then (T^n, φ) is a C^ω-flow. We call it the **translation flow** with **frequencies** $\omega_1, \ldots, \omega_n$. If $\omega_1, \ldots, \omega_n$ are linearly independent over $\mathbf{Z}$, they are called **independent**. Every orbit of the translation flow is dense if and only if its frequencies are independent (Poincaré, H. Weyl). A translational flow with independent frequencies is called **quasiperiodic**.

Let (M, ω, X_H) be a Hamiltonian system on a $2n$-dimensional manifold M. Under a certain condition, there exist an open set U of M and a diffeomorphism $f: U \to T^n \times \mathbf{R}^n$ such that the following holds: Identify U with $T^n \times \mathbf{R}^n$ by f, and let $(q^1, \ldots, q^n, p_1, \ldots, p_n)$ be the coordinates of $T^n \times \mathbf{R}^n$. Then the energy function H is independent of $q = (q^1, \ldots, q^n)$ so that Hamilton's equations becomes $dq^i/dt = \partial H/\partial p_i$, $dp_i/dt = -\partial H/\partial q^i = 0$ for $i = 1, \ldots, n$. Therefore the solutions are given by $q^i(t) = (\partial H/\partial p_i(c))t + q^i(0)$ modulo $\mathbf{Z}$, $p_i(t) = p_i(0) = c_i$, $i = 1, \ldots, n$, where $c = (c_1, \ldots, c_n)$. Therefore $N_c = f^{-1}(T^n \times \{c\})$ is an **invariant torus** (i.e., an invariant set diffeomorphic to T^n) of (the flow generated by) X_H for all $c \in \mathbf{R}^n$, and the restriction flow on N_c is the translational flow with frequencies $\omega_1 = \partial H/\partial p_1(c), \ldots, \omega_n = \partial H/\partial p_n(c)$ (Arnold). Assume further that the Hessian $\det(\partial^2 H/\partial p_i \partial p_j)$ does not vanish at c and $\omega_1, \ldots, \omega_n$ are independent. Let $\tilde{H}$ be an energy function obtained by adding a sufficiently small perturbation to H. Then for almost all c' near c, there exists an invariant torus $\tilde{N}$ of $X_{\tilde{H}}$ near $N_{c'}$ such that the restriction flow on $\tilde{N}$ is differentiably equivalent to the translational flow with the same frequencies (Kolmogrov, Arnold, Moser).

(4) Generic properties for Hamiltonian systems were investigated by M. Buchner, Markus, Meyer, Pugh, Robinson, Takens, and Newhouse.

M. Bifurcation

(1) Consider a differential equation with a parameter. For example, let X be a domain of $\mathbf{R}^n$, $J = (-1, 1)$, and $f: J \times X \to \mathbf{R}^n$ a C^r-mapping. For each $\mu \in J$, define $f_\mu: X \to \mathbf{R}^n$ by $f_\mu(x) = f(\mu, x)$, $x \in X$. Consider the differential equation

$$dx/dt = f_\mu(x), \qquad x \in X \text{ and } \mu \in J. \tag{7}$$

As μ varies, the topological structure of the phase portrait of (7) may change. Suppose that there exists $\mu_0 \in J$ such that the topological structure of the phase portrait of (7) changes at $\mu = \mu_0$ but remains the same when $\mu_0 - \varepsilon < \mu < \mu_0$ or $\mu_0 < \mu < \mu_0 + \varepsilon$ for some $\varepsilon > 0$. Then μ_0 is called a **bifurcation point** of (7).

Hopf bifurcation: Assume that $X = \mathbf{R}^2$ and the origin $0 \in \mathbf{R}^2$ is a singular point of (7) for all $\mu \in J$. Assume further that the Jacobian matrix of f_μ at $0 \in \mathbf{R}^2$ has two distinct complex conjugate eigenvalues $\lambda(\mu)$ and $\overline{\lambda(\mu)}$ such that the real part $\operatorname{Re}\lambda(\mu)$ of $\lambda(\mu)$ is positive when $\mu > 0$, zero when $\mu = 0$, and negative when $\mu < 0$. Then $0 \in \mathbf{R}^2$ is a sink for $\mu < 0$ and a source for $\mu > 0$. Now assume further that $d/d\mu(\operatorname{Re}\lambda(\mu))|_{\mu=0}$ is positive and $0 \in \mathbf{R}^2$ is a "vague attractor." Then $0 \in J$ is a bifurcation point, and there exists an asymptotically stable closed orbit for (7) near and around $0 \in \mathbf{R}^2$ which depends continuously on μ for $\mu > 0$ [37]. Thus a sink of (7) ($\mu < 0$) changes to a source and an asymptotically stable closed orbit ($\mu > 0$) when μ changes its sign. The Hopf bifurcation theorem can be generalized to a higher-dimensional case, and there is a diffeomorphism version of the theorem.

(2) Bifurcations in more general settings have been investigated by many mathematicians, including Thom, Arnold, R. J. Sacker, G. R. Sell, D. H. Sattinger, G. Iooss, Ruelle, and Takens. Generic bifurcations of dynamical systems have been investigated by J. Soto-

mayer, Meyer, P. Brunovský, and others; bifurcations of Morse-Smale systems by Newhouse, Palis, Peixoto, and S. Matsumoto; and bifurcations of Axiom A diffeomorphisms by Newhouse and Palis.

N. Miscellaneous Topics

(1) Let S^3 be the 3-dimensional sphere and $V \in \mathscr{X}^1(S^3)$. H. Seifert proved that if V is sufficiently close (in C^0 topology) to a nonsingular vector field tangent to the fibers of the †Hopf fibration $S^3 \rightarrow S^2$, then V has a closed orbit. He conjectured that every nonsingular vector field $V \in \mathscr{X}^1(S^3)$ had a closed orbit. (**Seifert conjecture**; → 154 Foliations D). If a nonsingular vector field $V \in \mathscr{X}^1(S^3)$ is transverse to a codimension 1 foliation of class C^2, then it has a closed orbit (S. P. Novikov). Let M be a 3-dimensional C^∞-manifold. Then there exists a nonsingular vector field $V \in \mathscr{X}^1(M)$ with no closed orbit in any †homotopy class of a nonsingular vector field on M (P. A. Schweitzer). Thus the Seifert conjecture fails for a vector field of class C^1, but the conjecture for a vector field of class C^r $(r \geqslant 2)$ is an open problem. Related work has been done by Fuller, H. Chu, and A. Weinstein.

(2) Let M be a closed connected C^∞-manifold. A C^r-flow (M, φ) (resp. $f \in \mathrm{Diff}^r(M)$) is a **minimal flow** (resp. a **minimal diffeomorphism**) if M itself is a minimal set. If M admits a †locally free S^1-action of class C^∞, then it admits a minimal C^∞-diffeomorphism, and if M admits a locally free special (in particular, a †free) T^2-action of class C^∞, then it admits a minimal C^∞-flow (A. Fathi, Herman, A. B. Katok). Open problems: What are the topological properties of the manifolds admitting minimal flows? Does S^3 admit a minimal flow?

(3) E. N. Lorenz studied numerical solutions of the following nonlinear equations in $\mathbf{R}^3$ which arose from the convection equation: $dx/dt = -\sigma x + \sigma y$, $dy/dt = -xz + rx - y$, $dz/dt = xy/bz$. When $\sigma = 10$, $r = 28$, and $b = 8/3$, he found irregular behavior in this dynamical system. R. M. May studied numerical solutions of the following difference equation in connection with the growth of biological populations with nonoverlapping generations: $x_{n+1} = ax_n(1 - x_n)$, $x_n \in [0, 1]$ $(1 < a < 4)$. He found that the dynamical structure of the above difference equation was delicate and complicated. Y. Ueda and H. Kawakami also found similar phenomena in their numerical study of Duffing's equation of the type $d^2x/dt^2 + k\,dx/dt + x^3 = B\cos t$. The phenomena observed in these investigations were called **chaos**, which exhibits **strange attractors**.

These investigations have attracted the attention of many mathematicians and scientists. For 1-dimensional semidynamical systems such as May's equation, T. Y. Li, J. A. Yorke, A. N. Sharkovskiĭ, J. W. Milnor, Thurston, and many others have obtained notable results, while for Lorenz's equation we have results by Ruelle, Guckenheimer, Williams, Sinaĭ, and many others. Chaos arising from discretization of differential equations has been studied by M. Yamaguti, S. Ushiki, and others (→ 433 Turbulence and Chaos).

References

[1] H. Poincaré, Mémoire sur les courbes définies par une équation différentielle, J. Math. Pures Appl., (3) 7 (1881), 375–422; 8 (1882), 251–296. (Oeuvres I, Gauthier-Villars, 1928, 3–84.)
[2] H. Poincaré, Sur les courbes définies par les équations différentielles, J. Math. Pures Appl., (4) 1 (1885), 167–244; 2 (1886), 151–217. (Oeuvres I, 90–158, 167–222.)
[3] H. Poincaré, Les méthodes nouvelles de la mécanique céleste. I, Solutions périodiques, Non-existence des intégrals uniforms, Solutions asymptotiques. II, Méthodes de MM. Newcomb, Gyldén, Lindstedt et Bohlin. III, Invariants intégraux, Solutions périodiques du deuxième genre, Solutions doublement asymptotiques. Gauthier-Villars, I, 1982; II, 1893; III, 1899; English translation, New methods of celestial mechanics I–III, Clearinghouse for Federal Scientific and Technical Information, Springfield, 1967.
[4] G. D. Birkhoff, Dynamical systems, Amer. Math. Soc. Colloq. Publ., 1927.
[5] A. A. Andronov and L. S. Pontryagin, Systèmes grossiers, Dokl. Akad. Nauk SSSR, 14 (1937), 247–250.
[6] S. Lefschetz, Differential equations: geometric theory, Interscience, 1957.
[7] S. Smale, Differentiable dynamical systems, Bull. Amer. Math. Soc., 73 (1967), 747–817.
[8] D. V. Anosov, Geodesic flows on closed Riemannian manifolds with negative curvature, Proc. Steklov Inst. Math., 90 (1967), 1–235.
[9] R. Bowen, Equilibrium states and the ergodic theory of Anosov diffeomorphisms, Lecture notes in math. 470, Springer, 1975.
[10] D. Ruelle and F. Takens, On the nature of turbulence, Comm. Math. Phys., 20 (1971), 167–192.
[11] J. Moser, Stable and random motions in dynamical systems, Ann. Math. Studies 77, Princeton Univ. Press, 1973.
[12] L. H. Loomis and S. Sternberg, Advanced calculus, Addison-Wesley, 1968.

[13] V. V. Nemytskiĭ and V. V. Stepanov, Qualitative theory of differential equations, Princeton Univ. Press, 1960. (Original in Russian, 1947.)
[14] N. P. Bhatia and G. P. Szegö, Stability theory of dynamical systems, Springer, 1970.
[15] W. H. Gottschalk and G. A. Hedlund, Topological dynamics, Amer. Math. Soc. Colloq. Publ., 1955.
[16] R. Abraham and J. W. Robbin, Transversal mappings and flows, Benjamin, 1967.
[17] L. Markus, Lectures in differentiable dynamics, revised edition, Regional Conf. Series in Math. 3, Amer. Math. Soc., 1980.
[18] Z. Nitecki, Differentiable dynamics, MIT Press, 1971.
[19] M. Shub, Stabilité globale des systèmes dynamiques, Astérisque, Soc. Math. France, 56 (1978).
[20] M. C. Irwin, Smooth dynamical systems, Academic Press, 1980.
[21] S. Smale, Stable manifolds for differential equations and diffeomorphisms, Ann. Scuola Norm. Sup. Pisa, (3) 17 (1963), 97–116.
[22] I. Kupka, Contribution à la théorie des champs génériques, Contributions to Differential Equations, 2 (1963), 457–482.
[23] C. Pugh, An improved closing lemma and a general density theorem, Amer. J. Math., 89 (1967), 1010–1021.
[24] S. Smale, Morse inequalities for a dynamical system, Bull. Amer. Math. Soc., 66 (1960), 43–49.
[25] J. Palis and S. Smale, Structural stability theorems, Amer. Math. Soc. Proc. Symp. Pure Math., 14 (1970), 223–231.
[26] M. W. Hirsch and C. C. Pugh, Stable manifolds and hyperbolic sets, Amer. Math. Soc. Proc. Symp. Pure Math., 14 (1970), 133–163.
[27] S. Smale, The Ω-stability theorem, Amer. Math. Soc. Proc. Symp. Pure Math., 14 (1970), 289–297.
[28] S. Smale, Notes on differentiable dynamical systems, Amer. Math. Soc. Proc. Symp. Pure Math., 14 (1970), 277–287.
[29] J. W. Robbin, A structural stability theorem, Ann. Math., (2) 94 (1971), 447–493.
[30] J. W. Robbin, Topological conjugacy and structural stability for discrete dynamical systems, Bull. Amer. Math. Soc., 78 (1972), 923–952.
[31] R. C. Robinson, Structural stability of C^1 flows, Lecture notes in math. 468, Springer, 1975, 262–277.
[32] M. Shub, Dynamical systems, filtrations and entropy, Bull. Amer. Math. Soc., 80 (1974), 27–41.
[33] M. W. Hirsch, C. C. Pugh, and M. Shub, Invariant manifolds, Lecture notes in math. 583, Springer, 1977.
[34] R, Bowen, On Axiom A diffeomorphisms, Regional Conf. Series in Math. 35, Amer. Math. Soc., 1978.
[35] R. Abraham and J. E. Marsden, Foundations of mechanics, second edition, Benjamin, 1978.
[36] V. I. Arnold and A. Avez, Problèmes ergodiques de la mécanique classique, Gauthier-Villars, 1967; English translation, Ergodic problems of classical mechanics, Benjamin, 1968.
[37] J. E. Marsden and M. McCracken, The Hopf bifurcation and its applications, Applied math. sci. 19, Springer, 1976.

127 (XIX.8) Dynamic Programming

A. General Remarks

There are two types of multistage †decision processes. In one of them, an outcome of the whole process is determined at the final stage without any consideration of the outcome for each intermediate stage. The †extensive form of a †game is of this type. In the other type, an outcome is assigned at each stage of a multistage decision process. The theory of **dynamic programming**, dealing with this latter type, has been developed by R. Bellman and others since 1950 and is now one of the fundamental branches of mathematical programming, along with the theories of linear and nonlinear programming. The following examples illustrate some features of multistage decision processes.

Multistage allocation process. We are given a quantity $x>0$ that can be divided into two parts y and $(x-y)$. From y we obtain a return $g(y)$, and from $(x-y)$ a return $h(x-y)$. In so doing, we expend a certain amount of our original resources and are left with a new quantity, $ay+b(x-y)$, $0<a, b<1$, with which the process is continued. How do we proceed so as to maximize the total return obtained in a finite or unbounded number of stages?

Multistage choice process. Suppose that we possess two gold mines A and B, the first of which contains an amount x of gold, while the second contains an amount y. In addition, we have a single gold mining machine with the property that if used to mine gold in the mine A, there is a probability P_1 that it will mine a fraction r_1 of the gold there and remain in working order, and a probability $1-P_1$ that it will mine no gold and be damaged beyond repair. Similarly, the mine B has associated

with it the corresponding probabilities P_2 and $1-P_2$ and fraction r_2. How do we proceed in order to maximize the total amount of gold before the machine is defunct?

These two processes have the following features in common: (1) In each case we have a physical system characterized in any state by a small set of parameters, the **state variables**. (2) In each state of either process we have a choice of a number of **decisions**. (3) The effect of a decision is a transformation of the state variables. (4) The past history of the system is of no importance in determining future actions (†Markov property). (5) The purpose of the process is to maximize some function of the state variables.

A **policy** is a rule for making decisions that yields an allowable sequence of decisions; an **optimal policy** is a policy that maximizes a preassigned function of the final state variables. A convenient term for this preassigned function of the final state variables is **criterion function**. One of the characteristic features of Bellman's methodology of dynamic programming is the appeal to the **principle of optimality**: An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision.

In the multistage allocation process the state variables are x (the quantity of resources) and z (the return obtained up to the current stage). The decision at any stage consists of an allocation of a quantity $0 \leqslant y \leqslant x$. This decision has the effect of transforming x into $ay+b(x-y)$ and z into $z+g(y)+h(x-y)$. The purpose of the process is to maximize the final value of z. Denote by $f_N(x)$ the N-stage return obtained starting from an initial state x and using an optimal policy. Then we have

$$f_1(x)=\max_{0\leqslant y\leqslant x}[g(y)+h[x-y)],$$

$$f_n(x)=\max_{0\leqslant y\leqslant x}[g(y)+h(x-y)+f_{n-1}(ay+b(x-y))], \qquad n\geqslant 2.$$

This recurrence relation yields a method for obtaining the sequence $\{f_n(x)\}$ inductively.

In the stochastic gold-mining process, the state variables are x and y (the present level of the two mines) and z (the amount of gold mined to date). The decision at any stage consists of a choice of A and B. If A is chosen, (x, y) goes into $((1-r_1)x, y)$ and z into $z+r_1x$, and if B is chosen, (x, y) goes into $(x, (1-r_2)y)$ and z into $z+r_2y$. The purpose of the process is to maximize the expected value of z obtained before the machine becomes defunct.

Denote by $f(x, y)$ the expected amount of gold obtained using an optimal sequence of choice. Then we have

$$f(x,y)=\max\begin{bmatrix}p_1\{r_1x+f((1-r_1)x,y)\}\\ p_2\{r_2y+f(x,(1-r_2)y)\}\end{bmatrix}.$$

The optimal policy can be described in the following way. We choose A or B according as $p_1r_1x/(1-p_1)$ is greater or less than $p_2r_2y/(1-p_2)$. We can choose either A or B if equality holds. After an operation according to such a choice, the machine may become defunct and terminate the process. If the machine is usable, then we can apply our policy to a new combination of the amounts of gold in A and B.

B. Discrete Deterministic Processes

By a deterministic process we mean a process in which the outcome of a decision is uniquely determined by the decision. We assume that the state of the system, apart from time dependence, is described in any stage by an M-dimensional vector p constrained to lie within some region D. Let $T=\{T_q\}$ (where q runs over a set S) be a set of transformations with the property that $p\in D$ implies that $T_q(p)\in D$ for all $q\in S$, i.e., any transformation T_q carries D into itself. The term "discrete" signifies here that we have a process consisting of a finite or denumerably infinite number of stages. A policy, for the finite process which we consider first, consists of a selection of N transformations in order, $P=(T_1, T_2, \ldots, T_N)$, yielding successively the sequence of states $p_i=T_i(p_{i-1})$ $(i=2,3,\ldots,N)$ with $p_1=T_1(p)$. These transformations are to be chosen to maximize a given function R of the final state p_N. Observe that the maximum value of $R(p_N)$, as determined by an optimal policy, will be a function of the initial vector p and the number N of stages only. Let us then define our basic auxiliary functions $f_N(p)=\max R(p_N)=$ the N-stage return obtained starting from an initial state p and using an optimal policy. This sequence is defined for $N=1,2,\ldots,$ and $p\in D$. The essential uses of the principle of optimality can be observed from the following two features. The first is the use of the **embedding principle**. The original process is embedded in a family of similar processes. In place of attempting to determine the characteristics of an optimal policy for an isolated process, we attempt to deduce the common properties of the set of optimal policies possessed by the members of the family. The second feature is the derivation of recurrence relations by which the functional equations connecting the members of the sequence $\{f_k(p)\}$ are established. Assume that we choose some transformation T_q as a result of our first decision, obtaining in this way a new state vector $T_q(p)$. The **maximum return**

from the following $k-1$ stages is, by definition, $f_{k-1}(T_q(p))$. It follows that, if we wish to maximize the total k-stage return, q must now be chosen to maximize this $(k-1)$-stage return. The result is the basic recurrence relation $f_k(p)=\max_{q\in S}f_{k-1}(T_q(p))$, for $k\geqslant 2$, with $f_1(p)=\max_{q\in S}R(T_q(p))$. For the case of an unbounded process, the sequence $\{f_k(p)\}$ is replaced by a single function $f(p)$, the total return obtained by using an optimal policy starting from state p, and the recurrence relation is replaced by the functional equation $f(p)=\max_q f(T_q(p))$.

C. Discrete Stochastic Processes

We again consider a discrete process, but one in which the transformations are stochastic rather than deterministic. The initial vector p is transformed into a stochastic vector z with an associated distribution function $dG_q(p,z)$ dependent on p and the choice of q. We assume that z is known after the decision has been made and before the next decision is to be made. We agree to measure the value of a policy in terms of some average value of the function of the final state. Let us call this expected value the **return**. Beginning with the case of a finite process, we define $f_k(p)$ as before. The expected return as a result of the initial choice of T_q is therefore

$$\int_{z\in D} f_{k-1}(z)\,dG_q(p,z).$$

Consequently, the recurrence relation for the sequence $\{f_k(p)\}$ is

$$f_k(p)=\max_{q\in S}\int_{z\in D} f_{k-1}(z)\,dG_q(p,z),\qquad k\geqslant 2,$$

with $f_1(p)=\max_{q\in S}\int_{z\in D}R(z)\,dG_q(p,z)$. Considering the unbounded process, we obtain the functional relation

$$f(p)=\max_{q\in S}\int_{z\in D} f(z)\,dG_q(p,z).$$

D. Continuous Deterministic Processes

A number of interesting processes require that decisions be made at each point of a continuum, such as a time interval. The simplest examples of processes of this character are provided by the †calculus of variations. Let us denote by $f(p;T)$ the return obtained over a time interval $[0,T]$ starting from the initial state p and employing an optimal policy. Although we consider the process as one consisting of choices made at each point t on $[0,T]$, it is better to begin with the concept of choosing policies (functions) over intervals, and then pass to the limit as these intervals shrink to points. Let d be an allowable decision made over the interval $[0,S]$, and let p_d be the state at S starting from the initial state p and employing d. The application of the principle of optimality suggests that

$$f(p;S+T)=\sup_D f(p_d,T),\tag{1}$$

where the supremum is taken over the set D of all allowable decisions d.

The limiting form of (1) as $S\to 0$ is a nonlinear partial differential equation (Bellman partial differential equation). This expression is important for use in actual analysis. For numerical purposes, S is kept nonzero but small. R. Bellman showed that it is possible to avoid many of the quite difficult rigorous details involved in this limiting procedure if we are interested only in the computational solution of variational processes.

E. Markovian Decision Processes

Consider a physical system which at any of the times $t=0,\Delta,2\Delta,\ldots$ must lie in one of the states $S_1,S_2,\ldots,S_N$. Let $y_i(n)$ be the probability that the system is in S_i at times $n\Delta$, and let P_{ij} be the probability that the system is in state S_j at $t+\Delta$ if it is in state S_i at time t. We suppose that the transition probabilities P_{ij} are independent of t. We assume that the P_{ij} depend on a parameter q, which may be a vector, and that at each stage of the process q is to be chosen so as to maximize the probability that the system is in the state S_1. We obtain the nonlinear system

$$y_1(n+1)=\max_q\sum_{j=1}^N p_{1j}(q)y_j(n)$$

$$=\sum_{i=1}^N y_i(n)p_{i1}(q^*);$$

$$y_i(n+1)=\sum_{j=1}^N p_{ij}(q^*)y_j(n),\qquad i=2,3,\ldots,N,$$

where $q^*=q^*(n)$ in the remaining $N-1$ equations is one of the values of q that maximize $y_1(n+1)$. There are similar processes that can be considered as continuous analogs of this type of decision process. These are called **Markovian decision processes** and were discussed by Bellman. There is, however, another type of Markovian decision process in which a reward is given at each stage. For each state S_i of the system there are k alternatives $1,2,3,\ldots,k$. If we choose the alternative h among these k alternatives, then the transition probabilities p_{ij}^h $(j=1,2,\ldots,n)$ are determined, and a reward r_{ij}^h is associated with each state S_j. Let us denote by $v_i(n)$ the total expected return obtained at the nth stage by appealing to an optimal policy when the initial state is S_i. Then the principle of optimality in the theory

of dynamic programming yields

$$v_i(n+1)=\max_h\left(\sum_{j=1}^{N} p_{ij}^h(r_{ij}^h+v_j(n))\right).$$

A policy-iteration method involving a value-determination operation with a policy-improvement routine was given by R. A. Howard [4].

F. Dynamic Programming and the Calculus of Variations

Problems in the calculus of variations can be viewed as multistage decision problems of a continuous type. It can be shown that the dynamic-programming approach yields formal derivations of classical necessary conditions for the calculus of variations. Let us consider the problem of minimizing the functional

$$J(y)=\int_a^b f(x,y(x),y'(x))\,dx,$$

where the function y is subject to $y(a)=c$. We embed this problem within the family of problems generated by allowing a and c to be parameters with the ranges of variation $-\infty<a<b$, $-\infty<c<\infty$. Now we define the optimal value function $S(a,c)=\min_y J(y)$. Then the principle of optimality yields the functional equation

$$S(a,c)$$
$$=\min\left(\int_a^{a+\Delta} f(x,y,y')\,dx+S(a+\Delta,c(y))\right),$$

where the minimization is taken over all functions defined over $[a,a+\Delta]$ with $y(a)=c$ and $c(y)=y(a+\Delta)$. Then, writing $v=y'(a)$, we get

$$-\frac{\partial S}{\partial a}=\min_v\left(f(a,c,v)+v\frac{\partial S}{\partial c}\right).$$

This yields the Euler equation, the Legendre condition, the Weierstrass condition, and the Erdman corner conditions. Furthermore, it can be shown that the functional equation characterization yields the Hamilton-Jacobi partial differential equation of classical mechanics.

The dynamic-programming approach can be applied to more general problems in the calculus of variations.

G. Dynamic Programming and the Maximum Principle

In general, the method of dynamic programming carries a more universal character than the maximum principle of optimal control theory. However, in contrast to the latter, this method does not have a rigorous logical foundation. V. G. Boltyanskiĭ [6] has presented a justification of the dynamic programming method.

Let $f_i(x,u)$ $(i=0,1,\ldots,n)$ be defined for $x\in V\subset\mathbf{R}^n$ and $u\in U\subset\mathbf{R}^r$, where V is an open set, and continuously differentiable on $V\times U$. Suppose that two points x^0 and x^1 are given in V. Among all the piecewise continuous controls $u(t)=(u_1(t),\ldots,u_r(t))\in U$ which transfer the phase point moving in accordance with

$$\frac{dx_i}{dt}=f_i(x,u(t)),\qquad i=1,\ldots,n,$$

from $x^0=x(t_0)$ to $x^1=x(t_1)$, find the control $u(t)$ for which the functional

$$J=\int_{t_0}^{t_1} f_0(x(t),u(t))\,dt$$

takes the smallest value.

A continuous function $\omega(x)=\omega(x_1,\ldots,x_n)$ is called a **Bellman function** relative to a point $a\in V$ if it possesses the following properties: (1) $\omega(a)=0$; (2) there exists a set M (the singular set of $\omega(x)$), which is closed in V and does not contain interior points, such that the function $\omega(x)$ is continuously differentiable on the set $V-M$ and satisfies the condition

$$\sup_{u\in U}\left(\sum_{i=1}^{n}\frac{\partial\omega}{\partial x_i}f_i(x,u)-f_0(x,u)\right)=0,\quad x\in V-M.$$

The following theorem gives a sufficient optimality condition.

Theorem: Assume that for $dx/dt=f(x,u(t))$ given in a region $V\subset\mathbf{R}^n$ there exists a Bellman function $\omega(x)$ relative to the point $a\in V$ with a piecewise smooth singular set. Assume, furthermore, that for any point $x^0\in V$ there exists a control $u(t)$ which transfers the phase point from $x^0=x(t_0)$ to $a=x(t_1)$ and satisfies the relation

$$\int_{t_0}^{t_1} f_0(x(t),u(t))\,dt=-\omega(x^0).$$

Then any such control $u(t)$ is optimal in V.

Recently, Vinter and Levis [7] obtained the following general result in this connection. The sufficient condition is given in terms of a solution to the Bellman partial differential equation. It is shown that if this equation is modified so that it is actually an inequality, and if this inequality is required to be satisfied in a limiting sense only, then the condition is also necessary for optimality.

H. Characteristic Features of Dynamic Programming

The characteristic features of the dynamic-programming approach can be summarized in

the following five points: (1) the advantage of lower dimensionality in comparison with the enumeration approach; (2) the possibility of finding maxima and/or minima of functions defined over restricted domains for which differential calculus may not work well; (3) the availability of numerical solutions in recursive form; (4) the possibility of formulating certain problems to which classical methods do not apply; and (5) the applicability of the method to most types of problems in mathematical programming, such as †inventory and production control, optimal searching, and some optimal and adaptive control processes.

References

[1] R. E. Bellman, Dynamic programming, Princeton Univ. Press, 1957.
[2] R. E. Bellman, Applied dynamic programming, Princeton Univ. Press, 1962.
[3] R. E. Bellman, Adaptive control processes; a guided tour, Princeton Univ. Press, 1961.
[4] R. A. Howard, Dynamic programming and Markov processes, Technology Press and Wiley, 1960.
[5] S. E. Dreyfus, Dynamic programming and calculus of variations, Academic Press, 1965.
[6] V. G. Boltyanskiĭ, Sufficient conditions for optimality and the justification of the dynamic programming method, SIAM J. Control, 4 (1966), 326–361.
[7] R. B. Vinter and R. M. Lewis, A necessary and sufficient condition for optimality of dynamic programming type, making no a priori assumption on the control, SIAM J. Control and Optimization, 16 (1978), 571–583.

E

128 (XVIII.15) Econometrics

A. General Remarks

The term **econometrics** can be interpreted in various ways. In its widest sense, it means the application of mathematical methods to economic problems and includes mathematical economics, †mathematical programming, etc. However, here we use it to mean statistical methods applied to economic analysis.

The object of econometrics is to provide methods to analyze relationships between economic variables. We classify these methods into four categories according to the types of relationships involved: (1) Analysis of causal relations: If a set of variables $X_1, \ldots, X_n$ affects an economic variable Y, we can estimate the direction and extent of those effects on Y. (2) Analysis of equilibrium: When a set of economic variables $Y_1, \ldots, Y_m$ is determined through a market equilibrium mechanism, we can analyze the structure of relationships that determines the equilibrium. (3) Analysis of correlation: When a set of economic variables is affected simultaneously by some (unknown) common factors, we can analyze the correlation structure of the variables. (4) Analysis of time interdependence: A process of development in time of a set of economic variables can be analyzed.

There are two types of economic data: (a) **macroeconomic data**, representing quantities and variables related to a national economy as a whole, usually based on national census; and (b) **microeconomic data**, representing information about the economic behavior of individual persons, households, and firms. Macroeconomic data are usually given as a set of time series (→ 421 Time Series Analysis A), while microeconomic data are obtained mainly through statistical surveys and are given as **cross-sectional data**. These two types of data, related to macroeconomic theory and microeconomic theory, respectively, require different approaches; and sometimes information obtained from both types of data has to be combined; obtaining macroeconomic information from microeconomic data is called aggregation.

B. Regression Analysis

The most common technique for the first category of problems is †regression analysis, which is applied to both microeconomic and macroeconomic analysis. However, there are problems peculiar to economic analysis, where a factor can seldom be controlled; and usually there are too many highly related independent variables. In such cases, if all possible independent variables are taken into the model, the accuracy of the †estimators of the coefficients becomes extremely poor. Such a phenomenon, called **multicollinearity**, brings up the problem of selection of independent variables (→ 403 Statistical Models), to which no satisfactory solution has been given. Also, assumptions about the error terms may be dubious, the error terms may be correlated, or the variances may be different. If the †variance-covariance matrix of the errors is given, the †generalized least squares method can be applied, but usually such a matrix is not available.

C. Systems of Simultaneous Equations

The second category of problems is peculiar to economic analysis and applies mainly to macroeconomic data. Suppose that $\mathbf{Y}=(Y_1, \ldots, Y_G)'$ is a vector consisting of G economic variables, among which there exist G relationships that determine the equilibrium levels of the variables. We also suppose that there exist K variables $\mathbf{Z}=(Z_1, \ldots, Z_K)'$ that are independent of the economic relations but affect the equilibrium. The variables $\mathbf{Y}$ are called **endogenous variables**, and the $\mathbf{Z}$ are called **exogenous variables**. If we assume linear relationships among them, we have an expression such as

$$\mathbf{Y}=B\mathbf{Y}+\Gamma\mathbf{Z}+\mathbf{u}, \tag{1}$$

where B and Γ are matrices with constant coefficients and $\mathbf{u}$ is a vector of **disturbances** or errors. (1) is called the **linear structural equation system** and is a system of simultaneous equations. By solving the equations formally, we get the so-called **reduced form**

$$\mathbf{Y}=\Pi\mathbf{Z}+\mathbf{v}, \tag{2}$$

where $\Pi=(I-B)^{-1}\Gamma$, $\mathbf{v}=(I-B)^{-1}\mathbf{u}$. The relation of $\mathbf{Y}$ to $\mathbf{Z}$ is determined through the reduced form (2), and if we have enough data on $\mathbf{Y}$ and $\mathbf{Z}$ we can estimate Π. The problem of **identification** is to decide whether we can determine the unknown parameters in B and Γ uniquely from the parameters in the reduced form. A necessary condition for the parameters in one of the equations in (1) to be identifiable is that the number of unknown parameters (or, since known constants in the system are usually set equal to zero, the number of variables appearing in the equation) not be greater than $K+1$. If it is exactly equal to $K+1$, the equation is said to be **just identified**,

and if it is less than $K+1$, the equation is said to be **overidentified**.

If all the equations in the system are just identified, for arbitrary Π there exist unique B and Γ that satisfy $\Pi=(I-B)^{-1}\Gamma$. Therefore, if we denote the †least squares estimator of Π by $\hat{\Pi}$, we can estimate B and Γ from the equation $(I-\hat{B})\hat{\Pi}=\hat{\Gamma}$. This procedure is called the **indirect least squares method** and is equivalent to the †maximum likelihood method if we assume normality for $\mathbf{u}$.

When some of the equations are over-identified, the estimation problem becomes complicated. Three kinds of procedures have been proposed: (1) full system methods, (2) single equation methods, and (3) subsystem methods. In full system methods all the parameters are considered simultaneously, and if normality is assumed, the maximum likelihood estimator can be obtained by minimizing $|(\mathbf{Y}-\Pi\mathbf{Z})(\mathbf{Y}-\Pi\mathbf{Z})'|$. Since it is usually difficult to compute the maximum likelihood estimator, a simpler, but asymptotically equivalent, **three-stage least squares method** has been proposed. The single equation methods and the subsystem methods take into consideration only the information about the parameters in one equation or in a subset of the equations, and estimate the parameters in each equation separately. There is a single equation method, called the **limited information maximum likelihood method**, based on the maximum likelihood approach, and also a **two-stage least squares method**, which estimates Π first by least squares, computes $\hat{\mathbf{Y}}=\hat{\Pi}\mathbf{Z}$, and then applies the least squares method to the model

$$\mathbf{Y}=B\hat{\mathbf{Y}}+\Gamma\mathbf{Z}+\tilde{\mathbf{u}}.$$

These two and also some others are asymptotically equivalent. Among asymptotically equivalent classes of estimators corresponding to different information structures it has been established that the maximum likelihood estimators have asymptotically higher-order efficiency [5] (→ 399 Statistical Estimation) than other estimators, and Monte Carlo and numerical studies show that they are in most cases better than others if properly adjusted for the biases.

In many simultaneous equation models which have been applied to actual macroeconomic data, the values of endogenous variables obtained in the past appear on the right-hand sides of equations (1). Such variables are called **lagged variables**, and they can be treated, at least in the asymptotic theory of inference, as though they were exogenous. Hence exogenous variables and lagged endogenous variables are jointly called **predetermined variables**. When many lagged variables appear over many time periods and when some structure among the coefficients of those lagged variables can be assumed, such a model is called a distributed lag model.

Sometimes it is necessary to include some nonlinear equations in the simultaneous equation model. Such nonlinear simultaneous equation models are difficult to deal with, partly because the solution of the equation may not be unique, and in practical applications ad hoc procedures are applied to obtain estimates of the parameters.

D. Multivariate and Time Series Analysis

Problems in the third category can be approached by †multivariate analysis techniques. Sometimes †principal component analysis and †canonical correlation analysis have been applied to analyze the variations of a large amount of data. However, the practical meaning of the results obtained is often dubious.

The fourth category is the problem of time series analysis. Sophisticated theories of stochastic processes have little relevance for economic time series, because usually the time series do not satisfy such conditions as being stationary or having the †Markov property, etc. Recently, however, autoregressive moving average (†ARMA) and multivariate ARMA models [4] have been applied to macroeconomic data, especially for the purpose of prediction and for determining the direction of causal relations (→ 421 Time Series Analysis). Traditionally, fluctuations of economic time series have been thought to consist of trend, cyclic variation, seasonal variation, and error. Various ad hoc techniques have been used to separate or eliminate such components, but the theoretical treatment of such problems is far from satisfactory.

References

[1] W. C. Hood and T. C. Koopmans (eds.), Studies in econometric method, Cowles Commission Monograph 14, Wiley, 1953.
[2] H. Theil, Economic forecasts and policy, North-Holland, 1958.
[3] J. Johnston, Econometric methods, McGraw-Hill, 1963.
[4] G. E. P. Box and G. M. Jenkins, Time series analysis: Forecasting and control, Holden-Day, revised edition, 1976.
[5] M. Akahira and K. Takeuchi, Asymptotic efficiency of statistical estimators: Concepts and higher order efficiency, Lecture notes in statistics 7, Springer, 1981.

129 (XXI.21) Einstein, Albert

Albert Einstein (March 14, 1879–April 18, 1955) was born of Jewish parents in the city of Ulm in southern Germany. He became a Swiss citizen soon after graduating from the Eidgenössische Technische Hochschule of Zürich in 1900. Afterward, he obtained a position as examiner of patents at Bern, and while at that post, he published his theories on light quanta, †special relativity, and †Brownian motion. After briefly holding professorships at the University of Zürich and the University of Prague, he became a professor at the University of Berlin in 1913. His general theory of relativity was announced in 1916, and in 1921 he won the Nobel Prize in physics for his contributions to theoretical physics. To escape Nazi persecution, he fled to the United States in 1933, and until his retirement in 1945 he was a professor at the Institute for Advanced Study at Princeton. He advised President Roosevelt of the feasibility of constructing the atomic bomb, but after World War II, along with others who had been connected with the bomb, he was active in promoting the nuclear disarmament movement and the establishment of a world government.

The theory of relativity raises fundamental epistemological problems concerning time, space, and matter. The results of the general theory were verified in 1919 by observations of the solar eclipse.

Through his latter years, Einstein continued to work on †unified field theory and on the generalization of relativity theory.

References

[1] P. Frank, Einstein, his life and time, Knopf, 1947.
[2] A. Einstein, The meaning of relativity, Princeton Univ. Press, fifth edition, 1956.
[3] C. Seelig, Albert Einstein, eine dokumentarische Biographie, Europa Verlag, 1954.

130 (XX.16) Electromagnetism

A. Maxwell's Equations

Mathematical formulation of electromagnetics leads to †initial value and †boundary value problems for Maxwell's equations according to the geometric nature of the medium. **Maxwell's equations** for a vacuum are written in the form

$$\varepsilon_0\partial\mathbf{E}/\partial t=\operatorname{rot}\mathbf{H}-\mathbf{J}_e,\qquad \varepsilon_0\operatorname{div}\mathbf{E}=\rho_e,$$

$$\mu_0\partial\mathbf{H}/\partial t=-\operatorname{rot}\mathbf{E}-\mathbf{J}_m,\qquad \mu_0\operatorname{div}\mathbf{H}=\rho_m,\tag{1}$$

where $\mathbf{E}$ and $\mathbf{H}$ are the **electric** and **magnetic field** vectors, ρ_e and ρ_m the electric and magnetic charge densities, $\mathbf{J}_e$ and $\mathbf{J}_m$ the electric and magnetic current densities, ε_0 and μ_0 constants, and the quantity $1/\sqrt{\varepsilon_0\mu_0}=c$ the speed of light in vacuum (2.99797×10^8 m/s). Charge and current densities must satisfy the **equations of continuity**

$$\partial\rho_e/\partial t+\operatorname{div}\mathbf{J}_e=0,$$

$$\partial\rho_m/\partial t+\operatorname{div}\mathbf{J}_m=0.\tag{2}$$

Following upon the observation that, apparently, $\rho_m=0$ and $\mathbf{J}_m=0$ in nature, we henceforth set them equal to zero. This causes an asymmetry between the electric and magnetic quantities. On the other hand, the proposition "$\rho_m=0$ and $\mathbf{J}_m=0$" cannot be deduced from the classical theory itself.

In the presence of matter, additional charge and current appear due to the **electric** and **magnetic polarization P** and **M** of the material. Therefore, in this case, it is necessary to make the following substitutions in (1):

$$\rho_e\to\rho=\rho_e-\operatorname{div}\mathbf{P},$$

$$\mathbf{J}_e\to\mathbf{J}=\mathbf{J}_e+\partial\mathbf{P}/\partial t+\operatorname{rot}\mathbf{M},$$

$$\mathbf{H}\to\mathbf{H}+\mathbf{M}.\tag{3}$$

Moreover, if we define the **electric flux density** (or **electric displacement**) **D** and the **magnetic flux density** (or **magnetic induction**) **B** by

$$\mathbf{D}=\varepsilon_0\mathbf{E}+\mathbf{P},\qquad \mathbf{B}=\mu_0(\mathbf{H}+\mathbf{M}),\tag{4}$$

then Maxwell's equations (1) are transformed into

$$\partial\mathbf{D}/\partial t=\operatorname{rot}\mathbf{H}-\mathbf{J}_e,\qquad \operatorname{div}\mathbf{D}=\rho_e,$$

$$\partial\mathbf{B}/\partial t=-\operatorname{rot}\mathbf{E},\qquad \operatorname{div}\mathbf{B}=0.\tag{5}$$

In the electromagnetic field in a vacuum there is energy with a density

$$u=(\varepsilon_0/2)\mathbf{E}^2+(\mu_0/2)\mathbf{B}^2,\tag{6}$$

and energy flux with a density expressed by the **Poynting vector**

$$\mathbf{S}=\mathbf{E}\times\mathbf{H}.\tag{7}$$

Between these quantities the following relation holds:

$$\partial u/\partial t+\operatorname{div}\mathbf{S}=0.\tag{8}$$

An electric charge q moving with velocity $\mathbf{v}$ in an electromagnetic field is subject to the **Lorentz force**

$$\mathbf{F}=q\mathbf{E}+q\mathbf{v}\times\mathbf{B}.\tag{9}$$

This force can be interpreted as being caused by the **Maxwell stress tensor**

$$T_{ik}=(\varepsilon_0/2)(-E_iE_k+2\delta_{ik}\mathbf{E}^2)$$
$$+(\mu_0/2)(-H_iH_k+2\delta_{ik}\mathbf{H}^2). \qquad (10)$$

By introducing the **scalar potential** V and the **vector potential A**, we can express the field vectors as

$$\mathbf{B}=\operatorname{rot}\mathbf{A}, \quad \mathbf{E}=-\operatorname{grad}V-\partial\mathbf{A}/\partial t. \qquad (11)$$

Furthermore, if we impose an auxiliary condition (**Lorentz condition**)

$$\operatorname{div}\mathbf{A}+\varepsilon_0\mu_0\partial V/\partial t=0, \qquad (12)$$

then we obtain from (1) the wave equations

$$\Box V=-\rho_e/\varepsilon_0, \quad \Box\mathbf{A}=-\mu_0\mathbf{J}_e, \qquad (13)$$

where $\Box\equiv\Delta-\varepsilon_0\mu_0\partial^2/\partial t^2$ is called the **d'Alembertian** and is sometimes written as $\Diamond^2$. From (13) we conclude that the electromagnetic field can propagate in a vacuum as a wave with speed $c=1/\sqrt{\varepsilon_0\mu_0}$. In †quantum theory the potentials V and $\mathbf{A}$ are regarded as being more fundamental than $\mathbf{E}$ and $\mathbf{H}$ themselves. However, they are not uniquely determined, in the sense that the **gauge transformation**

$$V\to V+\partial\psi/\partial t, \quad \mathbf{A}\to\mathbf{A}-\operatorname{grad}\psi \qquad (14)$$

with an arbitrary function ψ of the space and time variables does not affect the fields.

Maxwell's equations are invariant under the †Lorentz transformation. Therefore they can be written in 4-dimensional tensor form (→ 359 Relativity C).

Maxwell's equations can be regarded as the †wave equations for †bosons with spin 1 (†photons). The equations of quantum electrodynamics are obtained if we regard the field quantities as quantum-mechanical variables (**q-numbers**) and then perform †second quantization.

B. Concrete Problems

In solving Maxwell's equations concretely, we usually make additional assumptions for polarizations, electric current, and field vectors

$$\mathbf{P}=\chi_e\mathbf{E}, \quad \mathbf{M}=\chi_m\mathbf{H}, \quad \mathbf{J}_e=\sigma\mathbf{E}, \qquad (15)$$

where χ_e, χ_m, and σ are called the **electric susceptibility, magnetic susceptibility**, and **conductivity**, respectively. Then we have

$$\mathbf{D}=\varepsilon\mathbf{E}, \quad \mathbf{B}=\mu\mathbf{H}, \qquad (16)$$

where $\varepsilon=\varepsilon_0+\chi_e$ is the **dielectric constant** and $\mu=\mu_0(1+\chi_m)$ is the **magnetic permeability**. Therefore the equations (5) become identical to (1) if ε_0 and μ_0 are replaced by ε and μ, respectively. (ρ_m and $\mathbf{J}_m$ are set equal to zero.)

Some cases of practical importance are given below.

(1) **Electrostatics**. If the fields are time-independent and there is no electric current, then $\mathbf{E}$ and $\mathbf{H}$ are mutually independent. The static electric field is calculated from the solution of the boundary value problem of the †Poisson equation $\Delta V=-\rho_e/\varepsilon$. Specifically, V takes a constant value in each conductor.

(2) **Magnetostatics**. For zero electric current, the problem of magnetostatics is solved in the same way as in electrostatics. For the case of nonvanishing stationary electric current, the problem is reduced to that of solving

$$\Delta\mathbf{A}=-\mu\mathbf{J}_e, \quad \operatorname{div}\mathbf{A}=0. \qquad (17)$$

(3) **Electric current in a conductor**. A stationary electric current in a conductor is governed by the equation of continuity $\operatorname{div}\mathbf{J}=0$, **Ohm's law** $\mathbf{J}=\sigma\mathbf{E}$, and a special case of Maxwell's equation $\operatorname{rot}\mathbf{E}=0$. The electric current produces heat (**Joule heat**) proportional to $\mathbf{J}\cdot\mathbf{E}$ per unit volume per unit time (**Joule's law**). For certain substances, the **specific resistance** σ^{-1} suddenly becomes negligibly small below a critical temperature. This is called **superconductivity**.

(4) **Quasistationary electric circuit**. The problem appearing most often in electrical engineering is that of a quasistationary circuit. Its characteristic feature is that the electric currents exist only in the circuit elements (inductors, capacitors, and resistors) and in the lines connecting them. The current (both $\mathbf{J}_e$ and $\partial\mathbf{D}/\partial t$) can be neglected in all other parts of the system. (This could be compared to the situation in dynamics where we consider systems of material points or of rigid bodies having a finite number of degrees of freedom, although every material body is essentially a continuum.) The system network is constructed as a †linear graph with the circuit elements as its branches. Topological †network theory deals with the relation between the structure of the linear graph and the electrical characteristics of the network, whereas function-theoretic network theory deals with the relation between current and voltage at each part of the network. In the latter theory, current and voltage are considered as functions of the frequency of the sinusoidal alternating voltage applied to some point of the network. Together these constitute a unique theoretical system in engineering mathematics for designing system networks (→ 282 Networks).

(5) **Theory of electromagnetic waves**. The theory of electromagnetic waves deals with the case where the changes of all field quantities are proportional to $e^{i\omega t}$, and in addition the frequency ω is so large that the term $\partial\mathbf{D}/\partial t$ in

(5) is of the same order as or larger than $\mathbf{J}_e$. In such a situation the electromagnetic field behaves like a wave.

Problems of various types arise depending on the geometry of the conducting and dielectric substances, on the type of the energy source, etc. Important problems are: (i) radiation of a wave from a point source into free space; (ii) scattering of a plane wave by small bodies or cylinders; (iii) diffraction of a wave through holes in a conducting plate; (iv) reflection and refraction of a wave at the boundary between different media; (v) wave propagation along a conducting tube (wave guide); and (vi) resonance of the electromagnetic field in a cavity surrounded by a conducting substance. Theoretical treatment similar to that for ordinary networks is possible for microwave circuits consisting of wave guides, cavities, etc.

(6) **Wave guides.** For electromagnetic waves with a harmonic time dependence $e^{-i\omega t}$ propagating inside a hollow tube of uniform cross section extending in the z-direction with perfectly conducting walls (called a **wave guide**), the Maxwell equations become

$$\operatorname{rot}\mathbf{E}=i\omega\mathbf{B},\qquad \operatorname{div}\mathbf{B}=0,$$

$$\operatorname{rot}\mathbf{B}=-i\mu\varepsilon\omega\mathbf{E},\quad \operatorname{div}\mathbf{E}=0,\tag{18}$$

with the boundary condition

$$\mathbf{n}\times\mathbf{E}=0,\qquad \mathbf{n}\cdot\mathbf{B}=0,$$

where $\mathbf{n}$ is a unit normal at the boundary surface. A further reduction is gained by Fourier analysis in the z-variable. For harmonic z-dependence e^{ikz}, the transverse components $\mathbf{E}_t=(\mathbf{e}_z\times\mathbf{E})\times\mathbf{e}_z$ and $\mathbf{B}_t=(\mathbf{e}_z\times\mathbf{B})\times\mathbf{e}_z$ are determined from the z-components $E_z=\mathbf{e}_z\cdot\mathbf{E}$ and $H_z=\mathbf{e}_z\cdot\mathbf{H}$ by the following part of the Maxwell equations:

$$ik\mathbf{E}_t+i\omega\mathbf{e}_z\times\mathbf{B}_t=\operatorname{grad}_t E_z,$$

$$ik\mathbf{B}_t-i\mu\varepsilon\omega\mathbf{e}_z\times\mathbf{E}_t=\operatorname{grad}_t B_z\tag{19}$$

(where grad_t is the transverse components of the gradient), up to the solutions for $E_z=B_z=0$, called **transverse electromagnetic (TEM) waves**, for which $k=\omega\sqrt{\mu\varepsilon}$, $\mathbf{B}_t=\pm\sqrt{\mu\varepsilon}\,\mathbf{e}_z\times\mathbf{E}_t$, and $\mathbf{E}_t$ is a solution of the electrostatic problem in two dimensions $\operatorname{rot}\mathbf{E}_t=0$, $\operatorname{div}\mathbf{E}_t=0$. To have $\mathbf{E}_t\neq 0$ for the TEM solution, it is necessary to have two or more surfaces, such as a coaxial cable (region between two concentric cylinders) or a parallel-wire transmission line. Nonzero longitudinal components E_z and B_z are determined from the 2-dimensional equations

$$\left[\left(\frac{\partial}{\partial x}\right)^2+\left(\frac{\partial}{\partial y}\right)^2+(\mu\varepsilon\omega^2-k^2)\right]\begin{Bmatrix}E_z\\B_z\end{Bmatrix}=0\tag{20}$$

with boundary conditions $E_z=0$ and $\partial B_z/\partial n=0$ on the wall, where $(\partial/\partial n)$ denotes the partial derivative in the normal direction. The solutions with $B_z=0$ are called **transverse magnetic (TM) waves** (or **electric (E) waves**) and those with $E_z=0$ **transverse electric (TE) waves** (or **magnetic (M) waves**); for each of these cases the equations determine an eigen-wave number k for a given angular frequency ω (typically in a waveguide situation) or eigenfrequency $\nu=\omega/(2\pi)$ for a given wavelength $\lambda=2\pi/k$ (typically in a resonant cavity situation).

References

[1] J. D. Jackson, Classical electrodynamics, Wiley, second edition, 1975.
[2] W. K. H. Panofsky and M. Phillips, Classical electricity and magnetism, Addison-Wesley, second edition, 1961.
[3] A. Sommerfeld, Electrodynamics, trans. by E. G. Ramberg, Academic Press, 1952. (Original in German, 1948.)
[4] P. Moon and D. E. Spencer, Foundations of electrodynamics, Van Nostrand, 1960.
[5] J. A. Stratton, Electromagnetic theory, McGraw-Hill, 1941.
[6] B. I. Bleaney, Electricity and magnetism, Clarendon Press, 1957.
[7] R. Fano, L. J. Chu, and R. B. Adler, Electromagnetic fields, energy and forces, Wiley, 1960.
[8] E. G. Hallen, Electromagnetic theory, Chapman & Hall, 1962.

131 (X.8) Elementary Functions

A. Definition

A function of a finite number of real or complex variables that is †algebraic, exponential, logarithmic, trigonometric, or inverse trigonometric, or the composite of a finite number of these, is called an **elementary function**. Elementary functions comprise the most common type of function in elementary calculus.

J. Liouville [1] defined the elementary functions as follows: An algebraic function of a finite number of complex variables is called an elementary function of class 0. Then e^z and $\log z$ are called elementary functions of class 1. Inductively, we define elementary functions of class n under the assumption that elementary functions of class at most $n-1$ have already

been defined. Let $g(t)$ and $g_j(w_1, \ldots, w_n)$ ($1 \leqslant j \leqslant m$) be elementary functions of class at most 1 and $f(z_1, \ldots, z_m)$ be an elementary function of class at most $n-1$. Then the composite functions $g(f(z_1, \ldots, z_m))$ and $f(g_1(w_1, \ldots, w_n), \ldots, g_m(w_1, \ldots, w_n))$ (and only such functions) are called the elementary functions of class at most n. An elementary function of class at most n and not of class at most $n-1$ is called an **elementary function of class** n. A function that is an elementary function of class n for some integer n is called an **elementary function**. In this article, we explain the properties of the most common elementary functions.

B. Exponential and Logarithmic Functions of a Real Variable

Let $a>0$, $a \neq 1$. A function $f(x)$ of a real variable satisfying the functional relation

$$f(x+y)=f(x)f(y), \qquad f(1)=a, \tag{1}$$

satisfies $f(n)=a^n$ for positive integers n and $f(-n)=1/a^n$ for negative integers $-n$. In general, $f(n/m)=\sqrt[m]{a^n}$ for every rational number $r=n/m$. If we assume that $f(x)$ is continuous, then there is a unique strictly monotone function $f(x)$ defined in $(-\infty, \infty)$ whose range is $(0, \infty)$. The function $f(x)$ is called the **exponential function with the base** a and is denoted by a^x, read "a to the power x" and also called a **power** of a with **exponent** x. Its inverse function is called the **logarithmic function to the base** a, and is denoted by $\log_a x$. The specific value $\log_a x$ is called the **logarithm** of x to the **base** a. If $g(x)=\log_a x$, we have

$$g(xy)=g(x)+g(y), \qquad g(a)=1. \tag{2}$$

Hence we have $xy=f(g(x)+g(y))$. Therefore we can reduce multiplication to addition by using a numerical table for the logarithmic function.

C. Logarithmic Computation

The logarithm to the base 10 is called the **common logarithm**. If two numbers x, y expressed in the decimal system differ only in the position of the decimal point (i.e., $y=x \cdot 10^n$ for an integer n), they share the same fractional parts in their common logarithms. The integral part of the common logarithm is called the **characteristic**, and the fractional part is called the **mantissa**. (We note that the word "mantissa" is also frequently used for the fractional part a in the †floating point representation $x=a \cdot 10^n$, $10^{-1} \leqslant a<1$, or $1 \leqslant a<10$.) The common logarithms of integers have been computed and published in tables.

D. Derivatives of Exponential and Logarithmic Functions

The function $f(x)=a^x$ is differentiable, and $f'(x)=k_a f(x)$, where k_a is a constant determined by the base a. If we take the base a to be

$$e=\lim_{\nu \to \infty}\left(1+\frac{1}{\nu}\right)^\nu=\sum_{\nu=0}^{\infty}\frac{1}{\nu!}=2.71828\ldots,$$

then we have $k_e=1$. The number $\sum(1/\nu!)$ is usually called **Napier's number** and is denoted by e after L. Euler (his letter to C. Goldbach of 1731; → Appendix B, Table 6). In 1873, C. Hermite proved that e is a transcendental number. We sometimes denote e^x by $\mathbf{exp}\,x$; the term **exponential function** usually means the function $\exp x$. The function e^x is invariant under differentiation, and conversely, a function invariant under differentiation necessarily has the form Ce^x. The logarithm to the base e is called the **Napierian logarithm** (or **natural logarithm**), and we usually denote it by $\mathbf{log}\,x$ without explicitly naming the base e (sometimes it is denoted by $\ln x$). The derivative of $\log x$ is $1/x$, hence we have the integral representation

$$\log x=\int_1^x \frac{dx}{x}. \tag{3}$$

The constant factor k_a in the derivative of a^x is equal to $\log a$. The graphs of $y=e^x$ and $y=\log x$ are shown in Fig. 1. The functions e^x and $\log(1+x)$ are expanded in the following Taylor series at $x=0$:

$$e^x=\sum_{\nu=0}^{\infty}\frac{x^\nu}{\nu!}, \tag{4}$$

$$\log(1+x)=\sum_{\nu=1}^{\infty}(-1)^{\nu+1}\frac{x^\nu}{\nu}. \tag{5}$$

The power series in the right-hand sides of (4) and (5) are called the **exponential series** and the **logarithmic series**, respectively. The radii of convergence of (4) and (5) are ∞ and 1, respectively.

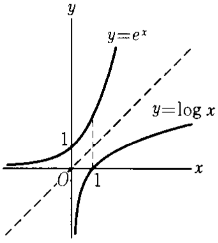


Fig. 1

E. Trigonometric and Inverse Trigonometric Functions of a Real Variable

The trigonometric functions of a real variable x are the functions **sin** x, **cos** x (→ 432 Trigonometry) and the functions **tan** $x = \sin x/\cos x$, **cot** $x = \cos x/\sin x$, **sec** $x = 1/\cos x$, and **cosec** $x = 1/\sin x$ derived from $\sin x$ and $\cos x$. The derivatives of $\sin x$ and $\cos x$ are $\cos x$ and $-\sin x$, respectively. They have the following Taylor expansions at $x=0$:

$$\sin x = \sum_{\nu=0}^{\infty} \frac{(-1)^{\nu}}{(2\nu+1)!} x^{2\nu+1}, \tag{6}$$

$$\cos x = \sum_{\nu=0}^{\infty} \frac{(-1)^{\nu}}{(2\nu)!} x^{2\nu}. \tag{7}$$

The radii of convergence of (6) and (7) are both ∞.

The inverse functions of $\sin x$, $\cos x$, and $\tan x$ are the **inverse trigonometric functions** and are denoted by **arc sin** x, **arc cos** x, and **arc tan** x, respectively. (Instead of this notation, $\mathbf{sin}^{-1} x$, $\mathbf{cos}^{-1} x$, and $\mathbf{tan}^{-1} x$ are also used). These functions are infintely multiple-valued, as shown in Fig. 2. But if we restrict their ranges within the part shown by solid lines in Fig. 2, they are considered single-valued functions. To be more precise, we restrict the range as follows: $-\pi/2 \leqslant \arcsin x \leqslant \pi/2$, $0 \leqslant \arccos x \leqslant \pi$, $-\pi/2 < \arctan x < \pi/2$.

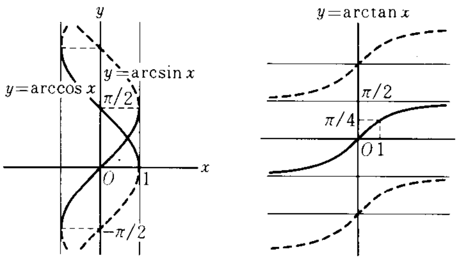


Fig. 2

The functions having these ranges are called the **principal values** and are sometimes denoted by **Arc sin** x, **Arc cos** x, and **Arc tan** x, respectively. The derivatives of these functions are $(1-x^2)^{-1/2}$, $-(1-x^2)^{-1/2}$, $(1+x^2)^{-1}$, respectively (→ Appendix A, Table 9.I; for the Taylor or Laurent expansions of $\tan x$, $\cot x$, $\sec x$, $\operatorname{cosec} x$, $\arcsin x$, $\arccos x$, $\arctan x$, etc., → Appendix A, Table 10.IV).

F. Hyperbolic Functions

Let P be a point on the branch of the hyperbola $x^2 - y^2 = 1$, $x > 0$, and let O be the origin and A the vertex $(1,0)$ of the hyperbola. Denote by $\theta/2$ the area of the domain surrounded by the line segments OA, OP, and the arc $\widehat{AP}$ of the hyperbola. Then we define the coordinates of P to be $(\cosh\theta, \sinh\theta)$ as functions of θ. We have

$$\mathbf{cosh}\, x = (e^x + e^{-x})/2,$$

$$\mathbf{sinh}\, x = (e^x - e^{-x})/2, \tag{8}$$

called the **hyperbolic cosine** and **hyperbolic sine**, respectively. As in the case of trigonometric functions, we define the **hyperbolic tangent** by **tanh** $x = \sinh x/\cosh x$, the **hyperbolic cotangent** by **coth** $x = \cosh x/\sinh x$, the **hyperbolic secant** by **sech** $x = 1/\cosh x$, and the **hyperbolic cosecant** by **cosech** $x = 1/\sinh x$. They are called the **hyperbolic functions**. The graphs of $\sinh x$ and $\cosh x$ are shown in Fig. 3. The trigonometric functions are sometimes called **circular functions**.

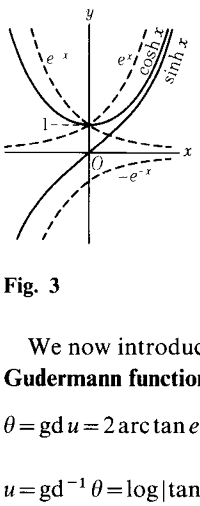


Fig. 3

We now introduce the **Gudermannian** (or **Gudermann function**):

$$\theta = \operatorname{gd} u = 2 \arctan e^u - \pi/2,$$

$$u = \operatorname{gd}^{-1}\theta = \log|\tan\theta + \sec\theta| = \frac{1}{2}\log\frac{1+\sin\theta}{1-\sin\theta}.$$

Then the hyperbolic functions can be expressed in terms of the trigonometric functions. For example,

$$\sinh u = \tan\theta, \qquad \cosh u = \sec\theta, \qquad \tanh u = \sin\theta.$$

G. Elementary Functions of a Complex Variable (→ Appendix A, Table 10)

(1) **Exponential function**. The power series (4) converges for all finite values if we replace x by the complex variable z and gives an †entire function of z with an †essential singularity at the point at infinity. This is the exponential function e^z of a complex variable z. It satisfies the **addition formula** (1), $e^{z_1+z_2} = e^{z_1}e^{z_2}$, and it is also the †analytic continuation of the exponential function of a real variable. For a purely imaginary number $z = iy$, we have the **Euler formula**

$$e^{iy} = \cos y + i \sin y. \tag{9}$$

The function $w = e^z$ gives a †conformal mapping from the z-plane to the w-plane, as shown in Fig. 4, which maps the imaginary axis of

the z-plane onto the unit circle of the w-plane ($w=u+iv$). For $z=x+iy$ (x, y are real numbers), we have $e^z=e^x e^{iy}$; hence e^z is a †simply periodic function with fundamental period $2\pi i$.

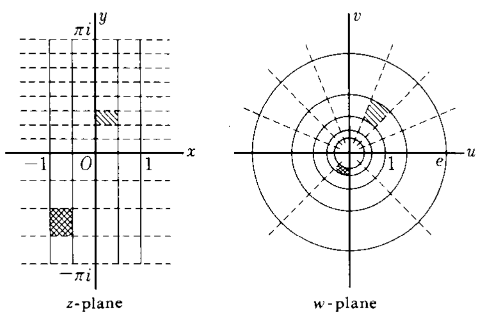


Fig. 4

(2) **Logarithmic function**. The logarithmic function $\log z$ of a complex variable z is the inverse function of e^z. It is an infinitely multiple-valued analytic function that has †logarithmic singularities at $z=0$ and $z=\infty$. All possible values are expressed by $\log z + 2n\pi i$ (n is an arbitrary integer), where we select a suitable value $\log z$. The **principal value** of $\log z$ is usually taken as $\log r + i\theta$, where $z = re^{i\theta}$ ($r=|z|$, θ is the argument of z) and $0 \leqslant \theta < 2\pi$. (Sometimes the range of the argument is taken as $-\pi < \theta \leqslant \pi$.) The principal value of $\log z$ is sometimes denoted by $\mathbf{Log}\, z$. The power series (5) gives one of its †functional elements. The integral representation (3) holds for a complex variable z. The multivalency of $\log z$ results from the selection of a contour of integration; the integral of $1/z$ around the origin is $2\pi i$, which is the increment of $\log z$.

(3) **Power**. The exponential function a^z for an arbitrary complex number a is defined to be $\exp(z \log a)$. Similarly, z^a is defined to be $\exp(a \log z)$. The function z^a is an algebraic function if and only if a is rational. In other cases, the function z^a is an elementary function of class 2.

(4) **Trigonometric functions, inverse trigonometric functions, hyperbolic functions.** The trigonometric, inverse trigonometric, and hyperbolic functions of a complex variable are defined by the analytic continuations of the corresponding functions of a real variable. For example, $\sin z$ and $\cos z$ are defined by the power series (6) and (7), respectively. They are entire functions whose zero points are $n\pi$ and $(n-\frac{1}{2})\pi$ (n is an integer), respectively. They are also represented by †Weierstrass's infinite product (→ Appendix A, Table 10.VI).

The functions $\tan z$, $\cot z$, $\sec z$, and $\operatorname{cosec} z$ are †meromorphic functions of z on the complex z-plane, and they are expressed by †Mittag-Leffler partial fractions (→ Appendix A, Tables 10.IV, 10.V). As can be shown from (8) and (9), we have

$$\cos z = (e^{iz} + e^{-iz})/2, \qquad \sin z = (e^{iz} - e^{-iz})/2i,$$

$$\cosh z = \cos iz, \qquad \sinh z = (\sin iz)/i. \tag{10}$$

Each of these formulas (10) is called an **Euler formula**. For a complex variable, the trigonometric and hyperbolic functions are composites of exponential functions, the inverse trigonometric functions are composites of logarithmic functions, and all of them are elementary functions of class 1. The definition of elementary functions by Liouville described in Section A refers, of course, to the functions of a complex variable. We remark that the inverse function of an elementary function is not necessarily an elementary function. For example, the inverse function of $y = x - a \sin x$ is not an elementary function (→ 309 Orbit Determination B).

The derivative of an elementary function is also an elementary function. However, the †primitive function of an elementary function is not necessarily an elementary function. The primitive function of a rational function or an algebraic function of †genus 0 is again an elementary function. Similar properties hold for rational functions of trigonometric functions. Liouville [1] carried through a deep investigation of the situation where the integral of an elementary function is also an elementary function.

References

[1] J. Liouville, Mémoire sur la classification des transcendantes et sur l'impossibilité d'exprimer les racines de certaines équations en fonction finie explicite des coefficients, J. Math. Pures Appl., 2 (1837), 56–104; 3 (1838), 523–546.

[2] E. G. H. Landau, Einführung in die Differentialrechung und Intergralrechnung, Noordhoff, 1934; English translation: Differential and integral calculus, Chelsea, 1965.

[3] N. Bourbaki, Eléments de mathématique, Fonctions d'une variable réele, ch. 3, Fonctions élémentaires, Actualités Sci. Ind., 1074b, Hermann, second edition, 1958.

Also → references to 106 Differential Calculus.

132 (XX.31) Elementary Particles

A. Introduction

The word "atom" is derived from the Greek word for indivisible. It turns out that an atom

is divisible into its constituent nucleus and electrons. The nucleus, in turn, consists of protons and neutrons (together called nucleons). Photons (quanta of electromagnetic waves), electrons, protons, and neutrons (denoted by γ, e, p, and n, respectively), along with many other subsequently discovered subnuclear particles, are called **elementary particles**, while nuclei, atoms, and molecules are the **composite particles** composed of these elementary particles.

States of a particle form an irreducible (unitary) representation of the †proper inhomogeneous Lorentz group with positive energy (or a finite direct sum of such representations). Thus a **mass** ($m \geqslant 0$) and a **spin** ($j = 0, \frac{1}{2}, 1, \ldots$) are assigned to each particle (→ 258 Lorentz Group). For example, e, p, and n have spin $\frac{1}{2}$ and nonzero masses, while γ has spin 1 and zero mass.

Many elementary particles are unstable, decaying into other particles. The average **lifetime** is denoted by τ and its inverse is called the **half-width**. The time in which half of many samples of the same particle decays is called the **half-life**, given by $\tau \log 2$. For example, electrons, protons, and photons are supposed to be stable (or at least to have very long lifetimes), while a neutron decays into a proton, an electron, and a neutrino ν (β-decay) with a lifetime of about 15 minutes.

From the study of the relativistic equations (Dirac equations) for wave functions of an electron, P. A. M. Dirac predicted the existence of particles with the same mass as the electron, but of opposite electric charge (Dirac's hole theory, 1930). These were discovered in 1932 and called positrons. Every elementary particle is now believed to be associated with an **antiparticle** characterized by the opposite sign of the particle's additive quantum numbers, the two being connected by the †PCT theorem. Hence the positron is the antiparticle of the electron. The antiproton, theoretically expected for a long time and experimentally found in 1955, is the antiparticle of the proton. Antiprotons, antineutrons, and positrons are constituents of antimatter. The particles whose additive quantum numbers are all invariant under change of sign, such as photons (γ), neutral pions (π^0), etc., seem to be antiparticles of themselves and are said to be self-conjugate.

Elementary particles have four distinct types of interaction: gravitational, weak, electromagnetic, and strong, in increasing order of interaction strength. Gravitational and electromagnetic interactions were recognized in earlier centuries because these interactions are of long range. A. Einstein put forward the idea of a light quantum or photon as a lump of electromagnetic energy behaving like a particle. Whether a corresponding quantum (graviton) exists for gravitational interactions is a question related to the existence of gravitational waves themselves, and is not yet settled.

The nuclear force is an example of a strong interaction and is studied to elucidate nuclear structure and to derive the †cross sections of various collision processes involving nuclei. H. Yukawa predicted in 1935 the existence of a particle associated with the nuclear force, just as photons are associated with the electromagnetic interaction. Its mass was predicted, from the range of the nuclear force, to be about 200 times the electron mass, which is intermediate between the masses of electrons and nucleons, and hence the particle was named a mesotron or a meson. These were found in cosmic rays in 1947 and, in fact, it was found that the mesons relevant to the nuclear force (now called pions and denoted by π^+, π^-, π^0 according to their electric charge) decay into other kinds of particles called muons (denoted by μ^+, μ^-) with a charged-pion lifetime of 2.6×10^{-8} sec ($\pi^\pm \rightarrow \mu^\pm + \nu(\bar{\nu})$; the neutral pion π^0 decays, over a shorter lifetime, into photons). The muons then decay into electrons and neutrinos ($\mu^\pm \rightarrow e^\pm + \nu + \bar{\nu}$, $\bar{\nu}$ indicating antineutrinos) with muon lifetimes of about 2.2×10^{-6} sec. Weak interactions are relevant to these decays as well as to the β-decay of the nucleus and the neutron. Since 1962 electron neutrinos ν_e and muon neutrinos ν_μ have been distinguished in these decays, so that electron and muon numbers may be conserved.

Since 1949, many new particles (unstable under weak interactions) have been gradually found in cosmic rays; these are called strange particles. Some of the early ones are hyperons (Λ, Σ, Ξ), of spin $\frac{1}{2}$, and kaons ($\mathrm{K}, \bar{\mathrm{K}}$), of spin 0. For such strange particles, the strangeness quantum number, which is preserved in the strong interaction, has been introduced, and the **Nakano–Nishijima–Gell-Mann formula** concerning this number is known to hold. This says that $Q = I_3 + \frac{1}{2}(B + S)$ for each elementary particle, where Q is the electric charge in units of that of the positron, I_3 is the third component of the isospin, B is the baryon number, and S is the strangeness.

Quantum-mechanical wave functions for a system of identical particles seem to be either totally symmetric (Bose statistics) or totally antisymmetric (Fermi statistics) under permutations of particles. Accordingly, particles are either **bosons** (or Bose particles) or **fermions** (or Fermi particles). All bosons seems to have integer spins and all fermions half-odd-integer spins. This is the **connection of spin and statistics** and follows from certain axioms in quantum field theory (→ 150 Field Theory). There

has been some discussion on intermediate statistics (parabosons and parafermions).

B. Families of Elementary Particles

Elementary particles are classified into families of leptons, photons, and hadrons.

The family of leptons now consists of electrons (e), muons (μ), and tau-leptons (τ), their accompanying neutrinos (ν_e, ν_μ, ν_τ), and their antiparticles; τ was discovered in the late 1970s. Experimentally, ν_τ is not well established nor has the possibility $\nu_\tau = \nu_e$ yet been excluded. Leptons have spin $\frac{1}{2}$. They are characterized by having no strong interactions.

The family of photons consists of photons and the recently discovered intermediary weak vector bosons. Gluons (→ Section C (4)), if they exist, also belong to this family.

The family of hadrons has a large number of members which are either **mesons** or **baryons**, the former being bosons and the latter fermions. We now have, in addition to pions and kaons, many resonant mesons (unstable under strong interaction) such as ρ-mesons and ω-mesons. At present we have, in all, more than 20 species of mesons. This number does not count spin, charge, and antiparticle degrees of freedom. The baryon subfamily includes nucleons, hyperons, and excited states, now consisting of more than 30 species, again not counting spin, charge, and antiparticle degrees of freedom. We have nucleonic resonances with spin as high as 11/2.

Hadrons are now well understood as composites of subhadronic constituents called quarks (and antiquarks), although free quarks have not been observed. (Hence the problem of quark confinement has been discussed extensively.) Mesons are systems made up of a quark and an antiquark. Baryons are systems made up of three quarks. Therefore, at our present level of knowledge, the elementary particles might be leptons, photons, and quarks (and the corresponding antiparticles).

C. Methods in the Theory of Elementary Particles

(1) †Quantum Field Theory. Application of the ideas of quantum mechanics to electromagnetic fields and their interaction with electrons resulted in the formulation of **quantum electrodynamics** (and more generally quantum field theory). Application of quantum-mechanical perturbation theory to quantum electrodynamics with the fine-structure constant $\alpha = e^2/(\hbar c)$ (about 1/137) as an expansion parameter resulted in divergent expressions—the so-called divergence difficulties. The **ultraviolet divergence** comes from integration over high momenta (of virtual particles), and the **infrared divergence** is due to the zero mass of photons. It was later found that the ultraviolet divergence can be combined with a small number of parameters of the theory (i.e., electron mass, (possibly nonzero) photon mass, and electromagnetic coupling constant e) into a revised set of constants (called the renormalized masses and coupling constant), which are then equated to the observed finite values of these constants—a procedure called renormalization. Physically, this renormalization is pictured to be effected by virtual photons (and electron-positron pairs) surrounding (bare) electrons and photons. The infrared divergence is supposed to be a reflection of the fact that electrons can be accompanied by infinitely many photons with negligibly small total energy (a situation made possible by the zero mass of photons); this cannot be experimentally analyzed (and is indistinguishable from a single electron).

The relativistically covariant formulation of the renormalized perturbation theory of quantum electrodynamics proposed by S. Tomonaga, J. S. Schwinger, and R. P. Feynman (independently and in different forms), and in particular the Feynman rules and Feynman diagrams that lead to the Feynman integrals (→ 146 Feynman Integrals), made possible detailed theoretical computations, and the computed values (such as the Lamb shift of hydrogen and the anomalous magnetic moment of an electron) fit marvelously well with observed values—an achievement considered a great success of quantum electrodynamics. F. J. Dyson more or less showed that the renormalization procedure really absorbs all the divergences in all orders of the perturbation expansion in terms of renormalized constants, though there were later refinements and elaborations of the proof. This work also leads to the division of quantum field theories into two classes: **renormalizable** theories, where (infinitely many) ultraviolet divergences can be absorbed into a finite number of constants by renormalized perturbation theory, and **unrenormalizable** theories. The question of whether perturbation series converge in some sense is an unsolved question of quantum electrodynamics.

In quantum field theory, the central role is played by quantum fields, which are operator-valued generalized functions of a space-time point. Particle interpretations of any state at infinite past and infinite future are obtained in the theory from the asymptotic behavior of the fields (†asymptotic fields) at time $-\infty$ and time $+\infty$, and from their relation, expressed by the

†S-matrix, describing how particles scatter by collision. Thus any model of quantum fields makes, in principle, a prediction about what particles appear and how they behave (asymptotically) in mutual collisions.

After the success of quantum electrodynamics, the perturbation theory of quantum fields was applied to systems of pions and nucleons; this proved to be unsuccessful, possibly due to the lack of an appropriate expansion parameter. This has led some people to study the mathematically rigorous consequences of quantum field theories; these mathematical consequences do not require any perturbation calculation, simply following from a small number of mathematically formulated axioms believed to be satisfied by a large class of quantum field theories. This approach, referred to as †axiomatic quantum field theory, has yielded a few physically meaningful consequences of general nature: analyticity of some S-matrix elements, the †PCT theorem, and the connection of spin and statistics (→ Section A).

While the axiomatic approach provides a general framework, the †constructive field theory developed later provides examples of quantum field theories that fit into such a framework. Because of its concrete nature, it can make statements about detailed properties of the model, such as the (non-)existence of composite particles, the establishment of perturbation theory as an asymptotic expansion, and phase-transitions phenomena and the related broken symmetry as the coupling constant varies. It has, however, been successful for only 2 and 3 space-time dimensions.

(2) Analytic S-Matrix Approach. Due to the failure of the perturbation approach in quantum field theory, a new approach was developed based on assumptions about the analyticity properties of the S-matrix elements. The assumed analyticity properties were surmised from examination of Feynman integrals and of nonrelativistic potential scattering, and partially follow from axiomatic field theory. In this approach, the information that scattering amplitudes, or S-matrix elements, possess certain analytic properties with respect to energies, scattering angles, and so on is expressed by means of integral representations. For example, the forward two-body scattering amplitude $f(s)$ as a function of $s=(\text{energy in the center of mass system})^2$ is written as

$$f(s)=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\operatorname{Im} f(s')}{s'-s}ds'. \tag{1}$$

$\operatorname{Im} f(s)$ is related to the total cross section by the optical theorem, which is a statement of the unitarity of the S-matrix. Hence equation (1), called a **dispersion relation**, gives a relation among observable quantities. An integral representation for the general two-body scattering amplitude $f(s,t,u)$ (2 incoming and 2 outgoing particles) has been proposed by S. Mandelstam and is called the **Mandelstam representation**, where $s=(p_1+p_2)^2$, $t=(p_1+p_3)^2$, and $u=(p_1+p_4)^2$ (squares in a Minkowski metric) and p_1, p_2, p_3, p_4 are the 4-momenta of the incoming and outgoing particles, with sign reversed for the latter. (Some relations hold among variables: $p_i^2=m_i^2$ with m_i the mass, $\sum p_i=0$, $s+t+u=\sum m_i^2$.) Under the interchange of incoming and outgoing particles, the corresponding f's are related by analytic continuation. This is called the **crossing symmetry**. The study of the S-matrix directly from its analyticity and unitarity is called the **S-matrix approach** (→ 386 S-Matrices).

The analyticity of the two-body scattering amplitude f as a function of the angular momentum l with a fixed s was investigated by T. Regge for nonrelativistic potential scattering, and later the idea was applied to the S-matrix approach. The poles $l=l(s)$ of f (f is considered to be a function of l for each fixed s) are called **Regge poles**. Regge trajectories $l=l(s)$ for variable $s\leqslant 0$ have been shown to play important roles in the high-energy behavior of scattering amplitudes for small values of the variable t (four-momentum-transfer squared).

An approximate expression of S-matrix elements, with Regge poles and satisfying the crossing symmetry, was introduced by G. Veneziano and is called the **Veneziano model**. It has developed into the so-called **dual resonance model** (dual in the sense that s-channel poles are dual to t-channel poles) and has subsequently evolved into the **string model** of hadrons, according to which hadrons are viewed as systems composed of strings joining quarks (and antiquarks).

(3) Group-Theoretical Approach. In connection with the symmetry properties shown by the spectra and reaction patterns of hadrons, a group-theoretical approach has been developed. For example, the similarity of the behavior of neutrons and protons in nuclei, apart from the difference in their electromagnetic properties, was formulated as isospin invariance (under the group $SU(2)$). The symmetry properties are sometimes understood in terms of new additive quantum numbers that are conserved or nearly conserved, and these properties are made more concrete in the final stage by the introduction of fundamental constituents carrying these quantum numbers.

Based on the canonical formalism of quantum field theory, (Noether) currents, as

quantum-mechanical generators of the symmetry, are introduced in association with conserved quantum numbers. The commutation relation of these currents, referred to as a **current algebra**, shows the structure of a Lie algebra expressing the symmetry of the Lagrangian of the system of hadrons; for example, $SU(3) \times SU(3)$ for three species (flavors) of massless quarks (the first factor for vector currents and the second for axial-vector currents). The approach showed remarkable success when combined with the hypothesis of the **partially conserved axial-vector currents** (PCAC), which requires the divergence of the axial-vector currents to be proportional to the pseudoscalar meson fields.

Even if a theory has a symmetry under a group G in its formulation, a vacuum state of the theory might not have a symmetry under G. If that occurs, we speak of **spontaneously broken symmetry**. Under some assumptions, a particle of zero mass (which is connected to the vacuum by the current for the spontaneously broken symmetry) is associated with the spontaneously broken symmetry. This statement is called **Goldstone's theorem**, and the relevant particle of zero mass is called the **Nambu-Goldstone boson**. The PCAC hypothesis is believed to be connected with the spontaneous breakdown of the axial $SU(3)$ symmetry with pions, kaons, and eta-mesons as Nambu-Goldstone bosons. This approach produced the **Adler-Weisberger sum rule**, which relates the weak axial-vector coupling constants to pion-nucleon scattering cross sections.

A group-theoretical attempt to treat fermions and bosons on an equal footing resulted in the introduction of super Lie algebras (Z_2-graded Lie algebras). The basic new ingredient in this approach is a special class of generators roughly interpreted as the square root of the four-momentum, whose anticommutators (instead of commutators) are linear combinations of ordinary generators. A supermultiplet, which is an irreducible representation of a super Lie algebra, consists of both fermions and bosons. Extensions to local super Lie algebras (and also incorporation of gravitons into the framework) have been tried, with no realistic model emerging so far.

(4) Non-Abelian Gauge Field Theory. Recently, quantum field theory has been revived as non-Abelian †gauge theory, resulting in what are considered to be successful qualitative and semiquantitative predictions. The quantization of the theory was at first carried out in terms of the †Feynman path integral, where fictitious particles, called **Faddeev-Popov ghosts**, appear through the precise definition of the functional measure of the path integral. The canonical quantization has also been formulated with the explicit introduction of Faddeev-Popov ghosts from the beginning. Non-Abelian gauge field theory exhibits the very important property of asymptotic freedom, which states that the interaction at asymptotically high energies, or at very short distances, approaches that of free (noninteracting) theory. This property is required for the description of hadronic systems made of quarks in view of the experimental observation of the scaling behavior of deep inelastic inclusive structure functions. Thus gauge theory is believed to describe the dynamics of systems of quarks in hadrons. It is called **quantum chromodynamics**, and the quantum of the gauge field is called the gluon. The recent development of gauge field theory has been accompanied by many technical refinements of quantum field theory, which include the methods of dimensional regularization and renormalization groups.

Dimensional regularization starts with Feynman integrals defined for n-dimensional momenta with $n \neq 4$, in order to give meaning to integrals divergent for $n=4$. Then Feynman integrals have poles at $n=4$, which are absorbed into the unrenormalized constants by means of the renormalization procedure. This method is particularly suited for non-Abelian gauge theory, since it is the regularization method that keeps gauge invariance at each step of the calculation.

The †renormalization-group equation has been known for a long time. It results from the requirement that physical quantities should be dimensionally covariant under renormalization of constants in the theory. The renormalization-group equation is usually written as a differential equation for a Green's function, expressing the fact that a change of scale of momenta is balanced by changing coupling constants and masses. Further refinements are due to C. G. Callan, K. Symanzik, S. Weinberg, and G. 't Hooft.

D. Models of Elementary Particles

Hadrons are now considered to be made of more fundamental constituents; at present, at least 15 species of subhadronic, fermionic constituents (quarks) have been proposed. Attempts to understand subhadronic structure have a long history, the main landmarks of which include the Fermi-Yang model (where π-mesons are supposed to be made of protons and neutrons), the **Sakata model** (where all hadrons are supposed to be made of protons, neutrons, and Λ-hyperons), and a variety of quark models developed from the **eightfold way** of M. Gell-Mann and Y. Ne'eman (new

assignments of representations of $SU(3)$ to particles somewhat different from the Sakata model; for example, the octet representation for mesons and low-lying baryons, and the decuplet for excited baryons). Originally, quarks were supposed to come in 3 species (u-quarks, d-quarks, s-quarks), each carrying its own quantum number, now called the flavor quantum number. Then it was suggested that each flavored quark has three additional degrees of freedom, i.e., three new quantum numbers (called color quantum numbers) in order to account for experimental data: (1) the spin-statistics problem of baryonic ground state wave functions, (2) the decay rate of $\pi \to 2\gamma$, (3) the Drell ratio (= total cross section for $\{e^- + e^+ \to \text{anything}\}$, divided by the cross section for $e^- + e^+ \to \mu^- + \mu^+$). Recently, two new flavor degrees of freedom besides the old 3 flavors (u, d, s) have been discovered, the carriers of which are c-quarks and b-quarks, c being the constituents of J/ψ-particles, b of Υ-particles. The combination of 5 flavors and 3 colors results in 15 quarks, as stated earlier.

The so-called standard model is a quantum field theory based on a local, non-Abelian gauge group $SU(3) \times SU(2) \times U(1)$. The group $SU(3)$ is called the color $SU(3)$ group, which is supposed to be strictly unbroken and expresses the invariance of the theory under the local $SU(3)$ transformation of the three color degrees of freedom. Quarks having spin $\frac{1}{2}$ transform as its 3-dimensional fundamental representation, and the vector gauge bosons transforming as its 8-dimensional regular representation are **gluons**. This part of the theory is **quantum chromodynamics** (QCD). The remaining part of the theory, based on the local gauge group $SU(2) \times U(1)$, is called the **Glashow-Weinberg-Salam model** or its hadronic extension **quantum flavor dynamics** (QFD), and unifies the electromagnetic and the weak interactions. This gauge group is supposed to be spontaneously broken with the only unbroken subgroup $U(1)'$ corresponding to the electromagnetic gauge transformation. The underlying mechanism for the spontaneous breakdown of the gauge group $SU(2) \times U(1)$ is not well understood, but conventionally it is assumed to occur through the so-called **Higgs mechanism**. This is a mechanism proposed by P. W. Higgs, whereby the Goldstone boson acquires a nonzero mass if the broken symmetry occurs in the presence of an associated massless vector field (called a gauge vector field), which also becomes associated with massive bosons. The gauge vector bosons are identified with photons (γ) (unbroken symmetry and hence massless) and weak intermediary bosons (broken symmetry and hence massive), usually denoted by W and Z. Quarks and leptons transform under the group $SU(2)$ as doublet or singlet representations. The renormalizability of the spontaneously broken gauge field theory has been established by 't Hooft. Major predictions of the Glashow-Weinberg-Salam model, including the existence of the gauge bosons W and Z, have been borne out experimentally.

Grand unified models attempt to unify QCD and QFD, employing a larger Lie group containing $SU(3) \times SU(2) \times U(1)$ as a subgroup. The most popular ones are those based on the groups $SU(5)$, $SO(10)$ and on some of the exceptional groups. Super grand unified models attempt to unify QCD, QFD, and the gravitational interaction, with super Lie groups as a possible basis. Recently, there have been attempts to search for subquark structures.

References

[1] J. D. Bjorken and S. D. Drell, Relativistic quantum mechanics, McGraw-Hill, 1964.
[2] J. D. Bjorken and S. D. Drell, Relativistic quantum fields, McGraw-Hill, 1965.
[3] G. F. Chew, The analytic S-matrix, Benjamin, 1966.
[4] S. L. Adler and R. F. Dashen, Current algebras, Benjamin, 1967.
[5] V. de Alfaro, S. Fubini, G. Furlan, and C. Rossetti, Currents in hadron physics, North-Holland, 1973.
[6] M. Gell-Mann and Y. Ne'eman, The eightfold way, Benjamin, 1964.
[7] R. E. Behrends, J. Dreitlein, C. Fronsdal, and W. Lee, Simple groups and strong interaction symmetries, Rev. Mod. Phys., 34 (1962), 1–40.
[8] E. Fermi and C. N. Yang, Are mesons elementary particles? Phys. Rev., 76 (1949), 1739–1743.
[9] S. Sakata, On a composite model for the new particles, Prog. Theoret. Phys., 16 (1956), 686–688.
[10] L. Corwin, Y. Ne'eman, and S. Sternberg, Graded Lie algebra in mathematics and physics, Rev. Mod. Phys., 47 (1975), 573–603.
[11] M. Jacob (ed.), Dual theory, Physics Reports Reprint Book Series, vol. 1, North-Holland, 1974.
[12] J. Scherk, An introduction to the theory of dual models and strings, Rev. Mod. Phys., 47 (1975), 123–164.
[13] H. Harari, Quarks and leptons, Phys. Rep., 42 (1978), 235–309.
[14] E. S. Abers and B. W. Lee, Gauge theories, Phys. Rep., 9 (1973), 1–141.
[15] S. Weinberg, Recent progress in gauge theories of the weak, electromagnetic and

strong interactions, Rev. Mod. Phys., 46 (1974), 255–277.
[16] J. C. Taylor, Gauge theories of weak interactions, Cambridge Univ. Press, 1976.

133 (XIV.9) Ellipsoidal Harmonics

A. Ellipsoidal Coordinates

If $a>b>c$, then for any given $(x,y,z)\in\mathbf{R}^3$, the three roots of the cubic equation in θ

$$F(\theta)=\frac{x^2}{a^2+\theta}+\frac{y^2}{b^2+\theta}+\frac{z^2}{c^2+\theta}-1=0$$

are real and lie in the intervals $\theta>-c^2$, $-c^2>\theta>-b^2$, and $-b^2>\theta>-a^2$. Denoting these three roots by λ, μ, and ν (they are labeled so as to satisfy the inequalities $\lambda>-c^2>\mu>-b^2>\nu>-a^2$), $F(\lambda)=0$, $F(\mu)=0$, and $F(\nu)=0$ represent an ellipsoid, a hyperboloid of one sheet, and a hyperboloid of two sheets, respectively. They are confocal with the ellipsoid

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}-1=0,$$

pass through the point (x,y,z), and mutually intersect orthogonally.

The quantities λ, μ, ν are called the **ellipsoidal coordinates** of the point (x,y,z). Rectangular coordinates (x,y,z) are expressed in terms of ellipsoidal coordinates (λ,μ,ν) by the formula

$$x^2=\frac{(a^2+\lambda)(a^2+\mu)(a^2+\nu)}{(a^2-b^2)(a^2-c^2)},\tag{1}$$

and two others obtained from (1) by cyclic permutations of (a,b,c) and (x,y,z).

B. Ellipsoidal Harmonics

When a †harmonic function ψ of three real variables is constant on the surface $\lambda=$constant, $\mu=$constant, or $\nu=$constant in ellipsoidal coordinates, the function ψ is called an **ellipsoidal harmonic**. A solution of Laplace's equation $\Delta\psi=0$ in the form $\psi=\Lambda(\lambda)M(\mu)N(\nu)$ can be obtained by the method of separation of variables. The equation $\Delta\psi=0$ is written in the form

$$\sum(\mu-\nu)\Delta_\lambda\frac{\partial}{\partial\lambda}\left(\Delta_\lambda\frac{\partial\psi}{\partial\lambda}\right)=0,$$

where the summation is taken over the even permutations of (λ,μ,ν), and

$$\Delta_\lambda=\sqrt{(a^2+\lambda)(b^2+\lambda)(c^2+\lambda)}.$$

The ordinary differential equation

$$4\Delta_\lambda\frac{d}{d\lambda}\left(\Delta_\lambda\frac{d\Lambda}{d\lambda}\right)=(K\lambda+C)\Lambda\tag{2}$$

is satisfied by Λ and also by M and N if we replace λ by μ and ν, respectively. Equation (2) is called **Lamé's differential equation**, with K and C the separation constants.

Let $K=n(n+1)$ for $n=0,1,2,\ldots$. Then equation (2), for a suitable value (the eigenvalue) of C, has a solution that is a polynomial in λ or a polynomial multiplied by one, two, or three of $\sqrt{a^2+\lambda}$, $\sqrt{b^2+\lambda}$, and $\sqrt{c^2+\lambda}$.

Among these solutions $2n+1$ are linearly independent. We denote these solutions by $\Lambda=f_n^m(\lambda)$ $(m=1,2,\ldots,2n+1)$. They are essentially equivalent to the Lamé functions, to be defined at the end of this section. To be precise, by setting

$$\lambda+(a^2+b^2+c^2)/3=\xi,$$

$$C=B+n(n+1)(a^2+b^2+c^2)/3,$$

$$e_1=(b^2+c^2-2a^2)/3,\ldots;\qquad e_1+e_2+e_3=0,$$

we have

$$\frac{d^2\Lambda}{d\xi^2}+\left(\frac{1/2}{\xi-e_1}+\frac{1/2}{\xi-e_2}+\frac{1/2}{\xi-e_3}\right)\frac{d\Lambda}{d\xi}$$

$$=\frac{n(n+1)\xi+B}{4(\xi-e_1)(\xi-e_2)(\xi-e_3)}\Lambda.\tag{3}$$

This can also be written in the form

$$\frac{d^2\Lambda}{du^2}=(n(n+1)\wp(u)+B)\Lambda\tag{4}$$

by the change of variable $\xi=\wp(u)$ with the Weierstrass †$\wp$-function.

The differential equation (3) has $\xi=e_1,e_2,e_3,\infty$ as †regular singular points. A solution of (3) that is a polynomial in ξ or a polynomial multiplied by one, two, or three of $\sqrt{\xi-e_1}$, $\sqrt{\xi-e_2}$, and $\sqrt{\xi-e_3}$ is called a **Lamé function of the first kind**.

C. Classification of Lamé Functions

The $2n+1$ linearly independent solutions $f_n^m(\lambda)$ of (2) are classified into the following four families. If n is an even number $2p$, then $p+1$ solutions $f_n^m(\lambda)$ among the $2n+1$ solutions are polynomials in λ of degree p, and the other $3p$ functions are polynomials in λ of degree $p-1$ multiplied by

$$\sqrt{(b^2+\lambda)(c^2+\lambda)},\qquad\sqrt{(c^2+\lambda)(a^2+\lambda)},$$

or

$$\sqrt{(a^2+\lambda)(b^2+\lambda)}.$$

Since all these polynomials are products with real factors of degree 1, the solutions belonging

to the first family are of the type

$$f_n^m(\lambda)=(\lambda-\theta_1)(\lambda-\theta_2)\dots(\lambda-\theta_{n/2}), \tag{5}$$

while solutions of the latter kinds are of the type

$$f_n^m(\lambda)=\begin{Bmatrix}\sqrt{(b^2+\lambda)(c^2+\lambda)}\\ \sqrt{(c^2+\lambda)(a^2+\lambda)}\\ \sqrt{(a^2+\lambda)(b^2+\lambda)}\end{Bmatrix} \times(\lambda-\theta_1)(\lambda-\theta_2)\dots(\lambda-\theta_{n/2-1}). \tag{6}$$

The functions (5) and (6) are called **Lamé functions of the first species** and **of the third species**, respectively. On the other hand, if n is an odd number $2p+1$, then $3(p+1)$ solutions among the $2n+1$ functions $f_n^m(\lambda)$ are of the type

$$f_n^m(\lambda)=\begin{Bmatrix}\sqrt{a^2+\lambda}\\ \sqrt{b^2+\lambda}\\ \sqrt{c^2+\lambda}\end{Bmatrix} \times(\lambda-\theta_1)(\lambda-\theta_2)\dots(\lambda-\theta_{(n-1)/2}), \tag{7}$$

and the other p functions are of the type

$$f_n^m(\lambda)=\sqrt{(a^2+\lambda)(b^2+\lambda)(c^2+\lambda)} \times(\lambda-\theta_1)(\lambda-\theta_2)\dots(\lambda-\theta_{(n-3)/2}). \tag{8}$$

The functions (7) and (8) are called **Lamé functions of the second species** and **of the fourth species**, respectively. Hence in either case we have $2n+1$ linearly independent Lamé functions.

When n is even, we obtain an ellipsoidal harmonic

$$\psi_n^m=\prod_{p=1}^{n/2}(\lambda-\theta_p)(\mu-\theta_p)(\nu-\theta_p)$$

by multiplying $f_n^m(\lambda)$, $f_n^m(\mu)$, and $f_n^m(\nu)$ belonging to the first family. Also, in this case, by setting

$$\Theta_p=\frac{x^2}{a^2+\theta_p}+\frac{y^2}{b^2+\theta_p}+\frac{z^2}{c^2+\theta_p}-1 =\frac{(\lambda-\theta_p)(\mu-\theta_p)(\nu-\theta_p)}{(a^2+\theta_p)(b^2+\theta_p)(c^2+\theta_p)}$$

we have

$$\psi_n^m=\Theta_1\Theta_2\dots\Theta_{n/2} \tag{9}$$

up to constant coefficients. Utilizing Lamé functions of the third species (instead of functions of the first species) and formula (1), we find that

$$\psi_n^m=(yz \text{ or } zx \text{ or } xy)\times\Theta_1\Theta_2\dots\Theta_{n/2-1}. \tag{10}$$

For even n, every ellipsoidal harmonic expressible in terms of polynomials in x, y, z of degree n can be written as a linear combination of the functions (9) and (10), which are called the **ellipsoidal harmonics of the first** and **of the third species**, respectively. Similarly for odd n, the **ellipsoidal harmonics of the second** and **of the fourth species**

$$\psi_n^m=(x \text{ or } y \text{ or } z)\Theta_1\Theta_2\dots\Theta_{(n-1)/2}, \tag{11}$$

$$\psi_n^m=xyz\Theta_1\Theta_2\dots\Theta_{(n-3)/2} \tag{12}$$

are composed of the Lamé functions of the second and fourth species, respectively.

For odd n, these forms are a complete system of ellipsoidal harmonics that are linearly independent and expressible in terms of polynomials in x, y, z of degree n.

The zeros $\xi_1, \xi_2, \dots, \xi_p$ of the Lamé functions are real, and $\xi_i\neq\xi_j$ $(i\neq j)$. They never coincide with any one of e_1, e_2, and e_3. If $e_1>e_2>e_3$, then $\xi_1,\dots,\xi_p$ all lie between e_1 and e_3. If m is an integer such that $0\leqslant m\leqslant p$, we have one and only one Lamé function (with the species given) with m of its zeros lying between e_1 and e_2 and the remaining $p-m$ zeros between e_2 and e_3 (**Stieltjes's theorem**). In this way, a complete system of linearly independent Lamé functions of the specified type may be obtained, since m assumes $p+1$ different values. When the constant B appearing in the differential equation (3) takes specific values so that the equation has Lamé functions of the first kind as its solution, (3) also yields a solution Λ such that $\Lambda\to\xi^{-(n+1)/2}$ as $\xi\to\infty$. This function Λ is called the **Lamé function of the second kind**.

D. Ellipsoids of Revolution (Spheroids)

When the fundamental ellipsoid is a spheroid

$$\frac{x^2+y^2}{a^2}+\frac{z^2}{c^2}=1,$$

it is convenient to use the **spheroidal coordinates** (ξ,η,φ) given by

$$\begin{aligned} x&=l\sqrt{(\xi^2-1)(1-\eta^2)}\cos\varphi,\\ y&=l\sqrt{(\xi^2-1)(1-\eta^2)}\sin\varphi,\\ z&=l\xi\eta,\qquad l=\sqrt{c^2-a^2}\end{aligned} \tag{13}$$

for $a^2<c^2$ (prolate) and

$$\begin{aligned} x&=l\sqrt{(\xi^2+1)(1-\eta^2)}\cos\varphi,\\ y&=l\sqrt{(\xi^2+1)(1-\eta^2)}\sin\varphi,\\ z&=l\xi\eta,\qquad l=\sqrt{a^2-c^2}\end{aligned} \tag{14}$$

for $a^2>c^2$ (oblate). The solutions of Laplace's equation, which are regular at all finite points, are given by

$$\psi=P_n^m(\xi)P_n^m(\eta)\begin{smallmatrix}\cos\\ \sin\end{smallmatrix}m\varphi$$

in the prolate case, and

$$\psi=P_n^m(i\xi)P_n^m(\eta)\begin{smallmatrix}\cos\\ \sin\end{smallmatrix}m\varphi$$

in the oblate case. Here P_n^m is the †associated Legendre function of the first kind. Solutions which are regular outside a finite ellipsoid can be composed of the †associated Legendre functions of the second kind, $Q_n^m(\xi)$ or $Q_n^m(i\xi)$ instead of $P_n^m(\xi)$ or $P_n^m(i\xi)$, respectively.

E. Spheroidal Wave Functions

Transforming the †Helmholtz equation in prolate spheroidal coordinates (13), we have

$$\frac{1}{(\xi^2-\eta^2)}\left\{\frac{\partial}{\partial\xi}\left((\xi^2-1)\frac{\partial\Psi}{\partial\xi}\right)+\frac{\partial}{\partial\eta}\left((1-\eta^2)\frac{\partial\Psi}{\partial\eta}\right)+\left(\frac{1}{\xi^2-1}+\frac{1}{1-\eta^2}\right)\frac{\partial^2\Psi}{\partial\varphi^2}\right\}+\kappa^2\Psi=0,\qquad \kappa=kl. \tag{15}$$

By separating variables in the form $\Psi = X(\xi)Y(\eta)\frac{\cos}{\sin}m\varphi$, we have the equations

$$\frac{d}{d\xi}\left((1-\xi^2)\frac{dX}{d\xi}\right)+\left(\lambda-\kappa^2\xi^2-\frac{m^2}{1-\xi^2}\right)X=0, \tag{16a}$$

$$\frac{d}{d\eta}\left((1-\eta^2)\frac{dY}{d\eta}\right)+\left(\lambda-\kappa^2\eta^2-\frac{m^2}{1-\eta^2}\right)Y=0, \tag{16b}$$

which X and Y, respectively, must satisfy. The only difference between equations (16a) and (16b) arises from the fact that the domain of (16a) is given by $1<\xi$ whereas the domain of (16b) is given by $-1<\eta<1$. For the oblate spheroid, utilizing formula (14) we have

$$\frac{1}{(\xi^2+\eta^2)}\left\{\frac{\partial}{\partial\xi}\left((\xi^2+1)\frac{\partial\Psi}{\partial\xi}\right)+\frac{\partial}{\partial\eta}\left((1-\eta^2)\frac{\partial\Psi}{\partial\eta}\right)+\left(\frac{1}{1-\eta^2}-\frac{1}{\xi^2+1}\right)\frac{\partial^2\Psi}{\partial\varphi^2}\right\}+\kappa^2\Psi=0. \tag{17}$$

By separating variables as before, $Y(\eta)$ satisfies the same equation as (16b), while $X(\xi)$ satisfies equation (16a) with ξ replaced by $i\xi$. All these equations are of the type

$$\frac{d}{dz}\left((1-z^2)\frac{du}{dz}\right)+\left(\lambda-\kappa^2z^2-\frac{m^2}{1-z^2}\right)u=0. \tag{18}$$

A solution of (18) is known as a **spheroidal wave function**. The equation (18) has ±1 as regular singular points and ∞ as an irregular singular point of class 1. Hence spheroidal wave functions behave like †Legendre functions in the interval $[-1,1]$ and like †Bessel functions in the neighborhood of ∞.

F. The Functions $pe_n^m(x)$ and $qe_n^m(x)$

When we write a solution of (18), it is customary to write x instead of z when z is contained in the interval $[-1,1]$. We denote solutions of (18) which are regular on the whole domain $-1\leqslant x\leqslant 1$ by $pe_n^m(x)$, and the corresponding eigenvalues by $\lambda_{n,m}$ (assuming the boundary condition stated in this paragraph concerning singularities). In particular, when $\kappa\to0$, equation (18) reduces to †Legendre's associated differential equation, the eigenvalues of λ become $n(n+1)$ (n is a positive integer), and the corresponding eigenfunctions become the associated Legendre functions of the first kind:

$$P_n^m(x)=(1-x^2)^{m/2}\frac{d^mP_n}{dx^m}. \tag{19}$$

Hence $pe_n^m(x)$ is a solution which tends to a constant multiple of $P_n^m(x)$ as $\kappa\to0$. Using a system of orthogonal functions $P_n^m(x)$, we can expand $pe_n^m(x)$ as

$$pe_n^m(x)=\sum_{l\geqslant m}A_{n,l}^mP_l^m(x),\qquad |l-n|=\text{even number}. \tag{20}$$

The coefficients A satisfy a recurrence formula

$$\left(\lambda_{n,m}-l(l+1)+\kappa^2\frac{2l^2+2l-1-2m^2}{(2l-1)(2l+3)}\right)A_{n,l}^m-\kappa^2\frac{(l-m-1)(l-m)}{(2l-3)(2l-1)}A_{n,l-2}^m-\kappa^2\frac{(l+m+1)(l+m+2)}{(2l+3)(2l+5)}A_{n,l+2}^m=0. \tag{21}$$

The functions $pe_n^m(x)$ and $pe_l^m(x)$ are orthogonal in the domain $-1<x<1$.

Another solution of (18) exists which corresponds to the same eigenvalue $\lambda_{n,m}$, is independent of $pe_n^m(x)$, and has the opposite parity:

$$qe_n^m(x)=\sum_{l\geqslant -m,\,l-n=\text{even}}A_{n,l}^mQ_l^m(x)+\sum_{j\geqslant m,\,j-n=\text{odd}}B_{n,j}^mP_j^m(x), \tag{22}$$

where the $A_{n,l}^m$ are the same as in (20) and are determined by the recurrence formula (21), while for $j\geqslant m+2$, the $B_{n,j}^m$ satisfy the recurrence formula

$$\left(\lambda_{n,m}-j(j+1)+\kappa^2\frac{2j^2+2j-1-2m^2}{(2j-1)(2j+3)}\right)B_{n,j}^m+\kappa^2\frac{(j-m-1)(j-m)}{(2j-3)(2j-1)}B_{n,j-2}^m+\kappa^2\frac{(j+m+1)(j+m+2)}{(2j+3)(2j+5)}B_{n,j+2}^m=0. \tag{23}$$

Since the associated Legendre function of the second kind

$$Q_l^m(x)=(1-x^2)^{m/2}d^mQ_l/dx^m$$

is of the form

$$Q_l^m(x)=P_l^m(x)\log\sqrt{(1+x)/(1-x)}+(1-x^2)^{-m/2}\times(\text{a polynomial in } x)$$

for $l \geqslant m$, the $qe_n^m(x)$ have $x = \pm 1$ as singular points.

By expressing the solution of equation (18) in integral form we find that $pe_n^m(x)$ satisfies an integral equation

$$i^{n-m} v_{n,m} pe_n^m(x)$$
$$= \int_{-1}^{1} (1-x^2)^{m/2} (1-\xi^2)^{m/2} e^{i\kappa x \xi} pe_n^m(\xi) d\xi, \quad (24)$$

where the coefficient $v_{n,m}$ is related to $P_n^m(0)$ or $P_n^{m\prime}(0)$.

In order to extend the domain of definition of $pe_n^m(x)$ and $qe_n^m(x)$ beyond the interval $[-1,1]$, we adopt, in the domain G obtained by deleting the interval $[-1,1]$ from the complex plane, the Heine-Hobson definition of the associated Legendre function,

$$P_n^m(z) = (z^2-1)^{m/2} d^m P_n / dz^m, \quad (25)$$

instead of N. M. Ferrers's definition (19), and construct a solution of (18) in G:

$$pe_n^m(z) = \sum_{l \geqslant m} A_{n,l}^m P_l^m(z),$$
$$|l-n| = \text{even number}, \quad (26)$$

which is like (20) and again satisfies the integral equation (24). From this we can obtain the expansion formula

$$pe_n^m(z) = \frac{\sqrt{2\pi}\,(z^2-1)^{m/2}}{v_{n,m}(\kappa z)^m}$$
$$\times \sum_{l \geqslant m} (-1)^{(l-n)/2} \frac{(l+m)!}{(l-m)!} A_{n,l}^m \frac{J_{l+1/2}(\kappa z)}{\kappa z},$$
$$|l-n| = \text{even number}. \quad (27)$$

Multiplying this by a constant, we define

$$je_n^m(z) = \sqrt{\frac{\pi}{2}} \frac{(z^2-1)^{m/2}}{z^m}$$
$$\times \sum_{l \geqslant m} (-1)^{(l-n)/2} F_{n,l}^m \frac{J_{l+1/2}(\kappa z)}{\sqrt{\kappa z}} \bigg/ \sum_{l \geqslant m} F_{n,l}^m,$$
$$F_{n,l}^m = \frac{(l+m)!}{(l-m)!} A_{n,l}^m.$$

This expression asymptotically assumes the form

$$je_n^m(z) \sim \sin(\kappa z - n\pi/2)/\kappa z$$

for $|z| \gg 1$. In a similar manner we find a solution

$$ne_n^m(z) = -\sqrt{\frac{\pi}{2}} \frac{(z^2-1)^{m/2}}{z^m}$$
$$\times \sum_{l \geqslant m} (-1)^{(l-n)/2} F_{n,l}^m \frac{N_{l+1/2}(\kappa z)}{\sqrt{\kappa z}} \bigg/ \sum_{l \geqslant m} F_{n,l}^m$$

having the asymptotic form

$$ne_n^m(z) \sim \cos(\kappa z - n\pi/2)/\kappa z.$$

Disregarding a constant factor, this coincides with the function defined by (22) with $Q_n^m(z) = (z^2-1)^{m/2} d^m Q_n / dz^m$ in place of the associated Legendre function of the second kind.

References

[1] M. J. O. Strutt, Lamésche- Mathieusche- und verwandte Funktionen in Physik and Technik, Erg. Math., Springer, 1932 (Chelsea, 1967).
[2] E. W. Hobson, The theory of spherical and ellipsoidal harmonics, Cambridge Univ. Press, 1931 (Chelsea, 1955).
[3] Y. Hagihara, Theories of equilibrium figures of a rotating homogeneous fluid mass, revised translation, NASA, Washington D.C., 1970. (Original in Japanese, 1933.)
[4] Y. Hagihara, Celestial mechanics, vol. 4, pts. 1, 2, Japan Society for the Promotion of Science, 1975.
[5] C. Flammer, Spheroidal wave functions, Stanford Univ. Press, 1957.
[6] J. A. Stratton, P. M. Morse, L. J. Chu, J. D. C. Little, and F. J. Corbató, Spheroidal wave functions, including tables of separation constants and coefficients, Wiley, 1956.

134 (XIV.3) Elliptic Functions

A. Elliptic Integrals

Let $\varphi(z)$ be a polynomial in z of degree 3 or 4 with complex coefficients and $R(z,w)$ a rational function in z and w. Then $R(z, \sqrt{\varphi(z)})$ is called an **elliptic irrational function**. An integral of the type $\int R\,dz$ is called an **elliptic integral**. The origin of the name comes from the integral that appears in calculating the arc length of an ellipse. Any elliptic integral can be expressed by a suitable change of variables as a sum of elementary functions and elliptic integrals of the following three kinds:

$$\int \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}},$$

$$\int \sqrt{\frac{1-k^2z^2}{1-z^2}}\,dz,$$

and

$$\int \frac{dz}{(1-a^2z^2)\sqrt{(1-z^2)(1-k^2z^2)}}$$

(→ Appendix A, Table 16.I). These three kinds of integrals are called **elliptic integrals of the first**, **second**, and **third kind**, respectively, in **Legendre-Jacobi standard form**. This classifica-

tion corresponds to that of †Abelian integrals. The constant k is called the **modulus** of these elliptic integrals, and a is called the **parameter**. Let the four zeros of $\varphi(z)$ be $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ (we take one of them as ∞ when $\varphi(z)$ is of degree 3). The †Riemann surface $\mathfrak{R}$ corresponding to the elliptic irrational function has the zeros $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ as †branch points with degree of ramification 1, and is of two sheets and of †genus 1. If the integrand does not have a pole with a †residue, then the integral is multivalued only because the value (called the **periodicity modulus**) of the integral taken along the normal section (basis of the homology group) is not equal to zero.

B. Elliptic Integrals of the First Kind

When R is a function without singularities other than branch points, only the topological structure of R gives rise to the multivaluedness of the integral of R. The standard form is

$$\int_0^z \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}} = \int_0^\varphi \frac{d\varphi}{\sqrt{1-k^2\sin^2\varphi}} = w = F(k,\varphi), \qquad (1)$$

where $z=\sin\varphi$. This integral is the inverse function of $\operatorname{sn} w$ (→ Section J). The periodicity moduli are $2iK'$ and $4K$, where

$$K=K(k)=\int_0^1 \frac{dz}{\sqrt{(1-2^2)(1-k^2z^2)}} = \int_0^{\pi/2} \frac{d\varphi}{\sqrt{1-k^2\sin^2\varphi}} = F\left(k,\frac{\pi}{2}\right),$$

$$K'=K(k'), \qquad k'^2=1-k^2.$$

We call $K(k)$ a **complete elliptic integral of the first kind** and $F(k,\varphi)$ an **incomplete elliptic integral of the first kind** (→ Appendix A, Table 16). Setting

$$\sin\varphi_1=\frac{(1+k')\sin\varphi\cos\varphi}{\sqrt{1-k^2\sin^2\varphi}}, \qquad k_1=\frac{1-k'}{1+k'}, \qquad (2)$$

we have the relation

$$F(k,\varphi)=(1+k_1)F(k_1,\varphi_1)/2,$$

which is called **Landen's transformation**. Since $k_1<k$ when $0<k<1$, this transformation reduces the calculation of elliptic integrals to those with smaller values of k. For two given positive numbers a and b, put $a_0=a$, $b_0=b$, and $a_{n+1}=(a_n+b_n)/2$, $b_{n+1}=\sqrt{a_nb_n}$. Then the sequences $\{a_n\}$ and $\{b_n\}$ converge rapidly to the common limit, which is called the **arithmetico-geometric mean** of a and b, and is denoted by $\operatorname{ag}(a,b)$. The complete elliptic integral satisfies the relation $K(k)=\pi/[2\operatorname{ag}(1,\sqrt{1-k^2})]$.

C. Elliptic Integrals of the Second Kind

When R has poles with residue zero, its integral has no singularities other than poles. The standard form is

$$F(z)=\int_0^z \sqrt{\frac{1-k^2z^2}{1-z^2}}\,dz = \int_0^\varphi \sqrt{i-k^2\sin^2\varphi}\,d\varphi = E(k,\varphi), \qquad (3)$$

where $z=\sin\varphi$. We have

$$F(z)=\int_0^u \operatorname{dn}^2 u\,du=\frac{\Theta'(u)}{\Theta(u)}+\frac{E}{K}u$$

if we set $z=\operatorname{sn} u$. Here, Θ is Jacobi's theta function, and

$$\Theta(u)=\vartheta_4\left(\frac{u}{2K},\frac{iK'}{K}\right),$$

where ϑ_4 is a theta function to be described in Section I and K, K' are the same as in the case of elliptic integrals of the first kind. The quantity

$$E=\int_0^1 \sqrt{\frac{1-k^2z^2}{1-z^2}}\,dz = \int_0^{\pi/2} \sqrt{1-k^2\sin^2\varphi}\,d\varphi = E\left(k,\frac{\pi}{2}\right)$$

is called a **complete elliptic integral of the second kind**.

D. Elliptic Integrals of the Third Kind

When R has poles with nonzero residue its integral has logarithmic singularities. In this case, residues also contribute to multivaluedness of the integral. The standard form is

$$F(z)=\int_0^z \frac{dz}{(1-a^2z^2)\sqrt{(1-z^2)(1-k^2z^2)}} = \int_0^\varphi \frac{d\varphi}{(1-a^2\sin^2\varphi)\sqrt{1-k^2\sin^2\varphi}},$$

and it is expressed as

$$F(z)=\frac{\operatorname{sn}\alpha}{\operatorname{cn}\alpha\operatorname{dn}\alpha}\left(\frac{1}{2}\log\frac{\Theta(u-\alpha)}{\Theta(u+\alpha)}+u\frac{\Theta'(\alpha)}{\Theta(\alpha)}\right)+u$$

if we set $z=\operatorname{sn} u$ and $a^2=k^2\operatorname{sn}^2\alpha$ (→ Appendix A, Table 16).

E. Elliptic Functions and Periodic Functions

Historically the elliptic function was first introduced as the inverse function of the elliptic

integral. However, since it has been realized that elliptic functions are characterized as functions with double periodicity, it is now customary to define them as doubly periodic functions.

If $f(x)$, defined on a linear space X, satisfies the relation $f(x+\omega)=f(x)$ for some $\omega\in X$ and all $x\in X$, the number ω is called a **period** of $f(x)$, and $f(x)$ with a period other than zero is called a **periodic function**. The set P of all periods of $f(x)$ forms an additive group contained in X. If a basis $\omega_1,\dots,\omega_n$ of the additive group P exists, its members are called **fundamental periods** of $f(x)$.

Any continuous, nonconstant, periodic function of a real variable has only one positive fundamental period and is called a **simply periodic function**. The †trigonometric functions are typical examples: $\sin x$ and $\cos x$ have the fundamental period 2π; $\tan x$ and $\cot x$ have the fundamental period π (→ 159 Fourier Series).

A single-valued nonconstant †meromorphic function of n complex variables cannot have more than $2n$ fundamental periods that are linearly independent on the real number field. A function of one complex variable with two fundamental periods is called a **doubly periodic function**.

Let ω, ω' be the fundamental periods of a doubly periodic function. For a given number a, the parallelogram with vertices a, $a+\omega$, $a+\omega'$, $a+\omega+\omega'$ is called the **fundamental period parallelogram**. The complex plane is covered with a network of congruent parallelograms, called **period parallelograms**, obtained by translating the fundamental period parallelogram through $m\omega+n\omega'$ $(m,n=0,\pm1,\pm2,\dots)$.

A doubly periodic function $f(u)$ meromorphic on the complex plane is called an **elliptic function**. For simplicity, we usually denote the fundamental periods of an elliptic function by $2\omega_1$ and $2\omega_3$, and introduce ω_2 defined by the relation $\omega_1+\omega_2+\omega_3=0$. The first, and therefore also higher, derivatives of any elliptic function are elliptic functions with the same periods. The set of all elliptic functions with the same periods forms a †field. The number of poles in a period parallelogram is finite. The sum of the orders of the poles is called the **order** of the elliptic function. An elliptic function with no poles in a period parallelogram is merely a constant (**Liouville's first theorem**). The sum of the residues of an elliptic function at its poles in any period parallelogram is zero (**Liouville's second theorem**). Hence there can be no elliptic function of order 1. An elliptic function of order n assumes any value n times in a period parallelogram (**Liouville's third theorem**). The sum of the zeros minus the sum of the poles is a period (**Liouville's fourth theorem**).

F. Weierstrass's Elliptic Functions

Weierstrass defined

$$\wp(u)=\frac{1}{u^2}+\sum{}'\left(\frac{1}{(u-\Omega)^2}-\frac{1}{\Omega^2}\right)$$

as the simplest kind of elliptic function. Here $\Omega=2m\omega_1+2n\omega_3$, with m, n integers. The summation $\sum'$ extends over all integral values (positive, negative, and zero) of m and n, except for $m=n=0$. $\wp(u)$ is an elliptic function of order 2 with periods $2\omega_1$ and $2\omega_3$, called a **Weierstrass $\wp$-function**. The following functions $\zeta(u)$ and $\sigma(u)$ are called the **Weierstrass zeta** and **sigma functions**, respectively:

$$\zeta(u)=\frac{1}{u}+\sum{}'\left(\frac{1}{u-\Omega}+\frac{u}{\Omega^2}+\frac{1}{\Omega}\right)$$

and

$$\sigma(u)=u\prod{}'\left(\left(1-\frac{u}{\Omega}\right)\exp\left(\frac{u}{\Omega}+\frac{u^2}{\Omega^2}\right)\right).$$

These have quasiperiodicity, expressed by

$$\zeta(u+2\omega_i)=\zeta(u)+2\eta_i, \tag{4}$$

$$\sigma(u+2\omega_i)=-e^{2\eta_i(u+\omega_i)}\sigma(u), \tag{5}$$

$$\eta_1+\eta_2+\eta_3=0,\qquad \eta_i=\zeta(\omega_i),\qquad i=1,2,3;$$

and they satisfy the relations

$$\wp(u)=-\zeta'(u) \tag{6}$$

and

$$\zeta(u)=\frac{d\log\sigma(u)}{du}=\frac{\sigma'(u)}{\sigma(u)}. \tag{7}$$

The function $\wp(u)$ is an even function of u, and $\zeta(u)$ and $\sigma(u)$ are odd functions of u. By considering the integral $\int\zeta(u)du$ once around the boundary of a fundamental period parallelogram, we have

$$\left.\begin{matrix}\eta_1\omega_3-\eta_3\omega_1\\ \eta_2\omega_1-\eta_1\omega_2\\ \eta_3\omega_2-\eta_2\omega_3\end{matrix}\right\}=\pm\frac{\pi}{2}i,\qquad \operatorname{Im}\left(\frac{\omega_3}{\omega_1}\right)\gtrless 0, \tag{8}$$

which is called the **Legendre relation**.

The derivative

$$\wp'(u)=-2\sum 1/(u-\Omega)^3$$

of a $\wp$-function is an elliptic function of order 3 and bears the following relation to $\wp(u)$:

$$\begin{aligned}(\wp'(u))^2&=4(\wp(u))^3-g_2\wp(u)-g_3\\&=4(\wp(u)-e_1)(\wp(u)-e_2)(\wp(u)-e_3),\end{aligned}$$

$$g_2=60\sum{}'1/\Omega^4,\qquad g_3=140\sum{}'1/\Omega^6,$$

$$e_i=\wp(\omega_i),\qquad i=1,2,3. \tag{9}$$

By differentiating this relation successively, we see that $\wp^{(n)}(u)$ is expressed as a polynomial in $\wp(u)$ if n is an even number and as a product of polynomials in $\wp(u)$ and $\wp'(u)$ if n is an odd number.

In particular, writing $\wp(u)=z$ in (9) we find that the $\wp$-function is the inverse function of the elliptic integral

$$u=\int_\infty^z \frac{dz}{\sqrt{4z^3-g_2z-g_3}}$$

(→ Appendix A, Table 16.IV).

Any elliptic function can be expressed in terms of Weierstrass's functions. Specifically, let the poles of $f(u)$ and their orders be $a_1, a_2, \ldots, a_m$ and $h_1, h_2, \ldots, h_m$, respectively, and let the principal part in the expansion of $f(u)$ near the pole a_k be

$$\sum_{j=1}^{h_k}\frac{A_{kj}}{(u-a_k)^j}, \qquad k=1,2,\ldots,m. \tag{10}$$

Then we obtain

$$f(u)=C+\sum_{k=1}^{m}\left(A_{k1}\zeta(u-a_k)+\sum_{j=1}^{h_k}\frac{(-1)^jA_{kj}}{(j-1)!}\wp^{(j-2)}(u-a_k)\right), \tag{11}$$

where C is a constant depending on $f(u)$. This can be reduced to

$$f(u)=A+B\wp'(u)$$

by using the addition theorems (→ Appendix A, Table 16) for $\wp$ and zeta functions, where A and B are rational functions of $D(u)$. Therefore, given two elliptic functions with the same periods, after expressing them as rational functions of $\wp$ and $\wp'$ in the above form and eliminating $\wp$ and $\wp'$, we obtain an algebraic equation with constant coefficients. In particular, for any elliptic function $f(u)$ we obtain an †algebraic differential equation of the first order by using this method, with $f'(u)$ an elliptic function with the same periods. Furthermore, the functions $f(u+v)$, $f(u)$, and $f(v)$ satisfy an algebraic equation. Thus for any elliptic function an algebraic addition theorem holds.

G. Elliptic Functions of the Second Kind

As an extension of the definition of elliptic functions, if a †meromorphic function f satisfies the relations

$$f(u+2\omega_1)=\mu_1 f(u), \qquad f(u+2\omega_3)=\mu_3 f(u) \tag{12}$$

(μ_1 and μ_3 are constants) with the fundamental periods $2\omega_1, 2\omega_3$, we call $f(u)$ an **elliptic function of the second kind**. What we have called simply an elliptic function can now be called an **elliptic function of the first kind**. For constants ρ and v, the function

$$f(u)=e^{\rho u}\sigma(u-v)/\sigma(u) \tag{13}$$

is an example of an elliptic function of the second kind. In this case

$$\mu_i=e^{2\rho\omega_i-2v\eta_i}, \qquad i=1,2. \tag{14}$$

Furthermore, for given constants μ_1 and μ_3, any elliptic function of the second kind is expressed as the product of an elliptic function of the first kind and the function (13) with ρ and v determined by (14).

H. Elliptic Functions of the Third Kind

If a meromorphic function f satisfies

$$f(u+2\omega_i)=e^{a_iu+b_i}f(u), \qquad i=1,3 \tag{15}$$

(a_i and b_i are constants) with periods $2\omega_1, 2\omega_3$, we call it an **elliptic function of the third kind** (→ 3 Abelian Varieties I).

The Weierstrass sigma function $\sigma(u)$ is an example of an elliptic function of the third kind. The functions σ_1, σ_2, and σ_3, defined by the equations

$$\sigma_i(u)=-\frac{e^{\eta_iu}\sigma(u-\omega_i)}{\sigma(\omega_i)}, \qquad i=1,2,3, \tag{16}$$

are also elliptic functions of the third kind. In the case of elliptic functions of the second and third kinds, $2\omega_1$ and $2\omega_3$ are not periods in the strict sense defined earlier, but are conveniently referred to as the periods. The functions σ_i $(i=1,2,3)$ are called **cosigma functions**.

I. Theta Functions

The **theta functions**, or more strictly, **elliptic theta functions**, are defined by

$$\vartheta_1(v,\tau)=2\sum_{n=0}^{\infty}(-1)^nq^{(n+1/2)^2}\sin(2n+1)\pi v,$$

$$\vartheta_2(v,\tau)=2\sum_{n=0}^{\infty}q^{(n+1/2)^2}\cos(2n+1)\pi v,$$

$$\vartheta_3(v,\tau)=1+2\sum_{n=1}^{\infty}q^{n^2}\cos 2n\pi v,$$

$$\vartheta_4(v,\tau)=1+2\sum_{n=1}^{\infty}(-1)^nq^{n^2}\cos 2n\pi v, \tag{17}$$

where $q=e^{i\pi\tau}$, $\operatorname{Im}\tau>0$. We call (17) the **q-expansion formulas** of the theta functions, and we sometimes write ϑ_0 in place of ϑ_4. A theta function is an elliptic function of the third kind with periods 1 and τ. Any elliptic function can be expressed as a quotient of theta functions (→ Appendix A, Table 16 for specific exam-

ples). The q-expansion formula is quite suitable for numerical computation because of its rapid convergence; its terms decrease as the n^2 powers of q as $n \to \infty$.

An elliptic function with the fundamental periods $2\omega_1$ and $2\omega_3$ can also be viewed as having the periods $2\omega_1' = 2\omega_3$, $2\omega_3' = -2\omega_1$. Consequently, theta functions formed with the parameter $\tau = \omega_3/\omega_1$ can be expressed in terms of the parameter $\tau' = \omega_3'/\omega_1' = -\omega_1/\omega_3 = -1/\tau$, and we have

$$\vartheta_1(v,\tau) = iA\vartheta_1(v/\tau, -1/\tau),$$

$$\vartheta_2(v,\tau) = A\vartheta_4(v/\tau, -1/\tau),$$

$$\vartheta_3(v,\tau) = A\vartheta_3(v/\tau, -1/\tau),$$

$$\vartheta_4(v,\tau) = A\vartheta_2(v/\tau, -1/\tau),$$

$$A = \sqrt{i/\tau}\exp(-\pi i v^2/\tau). \tag{18}$$

These are called **Jacobi's imaginary transformations**. If $\mathrm{Im}(-1/\tau) \gg 1$, then $|q| \fallingdotseq 1$, and therefore the series in the q-expansion formula converges slowly. However, by an imaginary transformation we get $\mathrm{Im}(\tau) \gg 1$ and $|q| \fallingdotseq 0$, so that computations become much easier.

Each of the theta functions satisfies, as a function of two variables u and τ, the following partial differential equation of the heat-conduction type:

$$\partial^2 \vartheta(u,\tau)/\partial u^2 = 4\pi i \partial \vartheta(u,\tau)/\partial \tau \tag{19}$$

(→ Appendix A, Table 16.II).

J. Jacobi's Elliptic Functions

C. G. J. Jacobi defined elliptic integrals as inverse functions of elliptic integrals of the first kind in the Legendre-Jacobi standard form (1). They are, in the above notation,

$$\mathrm{sn}\, w = \sqrt{e_1 - e_3}\,\frac{\sigma(u)}{\sigma_3(u)} = \frac{\vartheta_3(0)\vartheta_1(v)}{\vartheta_2(0)\vartheta_4(v)}, \tag{20}$$

$$\mathrm{cn}\, w = \frac{\sigma_1(u)}{\sigma_3(u)} = \frac{\vartheta_4(0)\vartheta_2(v)}{\vartheta_2(0)\vartheta_4(v)}, \tag{21}$$

$$\mathrm{dn}\, w = \frac{\sigma_2(u)}{\sigma_3(u)} = \frac{\vartheta_4(0)\vartheta_3(v)}{\vartheta_3(0)\vartheta_4(v)}, \tag{22}$$

where $w = \sqrt{e_1 - e_3}\, u$ and $v = u/2\omega_1$.

These functions satisfy the relations

$$\mathrm{sn}^2 w + \mathrm{cn}^2 w = 1, \qquad k^2 \mathrm{sn}^2 w + \mathrm{dn}^2 w = 1, \tag{23}$$

where

$$k^2 = \frac{e_2 - e_3}{e_1 - e_3} = \frac{(\vartheta_2(0))^4}{(\vartheta_4(0))^4}. \tag{24}$$

The constants k and $k' = \sqrt{1-k^2}$ are called the **modulus** and the **complementary modulus**, respectively. Furthermore, the relation $d\,\mathrm{sn}\, w/dw = \mathrm{cn}\, w\, \mathrm{dn}\, w$ holds. The function $z = \mathrm{sn}\, w$ is the inverse function of the elliptic integral (1) (→ Appendix A, Table 16.III).

References

[1] S. Lang, Elliptic functions, Addison-Wesley, 1973.
[2] D. W. Masser, Elliptic functions and transcendence, Springer, 1975.
[3] C. G. J. Jacobi, Fundamenta nova theoriae functionum ellipticarum (1829) (C. G. J. Jacobis Gesammelte Werke I, G. Reimer, 1881, 49–239).
[4] G. H. Halphen, Traité des fonctions elliptiques et de leurs applications I, II, III, Gauthier-Villars, 1886–1891.
[5] A. Hurwitz and R. Courant, Vorlesungen über allgemeine Funktionentheorie und elliptische Funktionen, Springer, third edition, 1929.
[6] H. E. Rauch and A. Lebowitz, Elliptic functions, theta functions and Riemann surfaces, Williams & Wilkins, 1973.
[7] H. Hancock, Elliptic integral, Dover, 1958.

135 (II.3) Equivalence Relations

A. General Remarks

Suppose that we are given a relation R between elements of a set X such that for any elements x and y of X, either xRy or its negation holds. The relation R is called an **equivalence relation** (on X) if it satisfies the following three conditions: (1) xRx, (2) xRy implies yRx, and (3) xRy and yRz imply xRz. Conditions (1), (2), and (3) are called the **reflexive**, **symmetric**, and **transitive laws**, respectively. Together, they are called the **equivalence properties**. Condition (1) can be replaced by the following: (1') For each x there exists an x' such that xRx'. The relation "x is equal to y" is an equivalence relation. If xRy means that x and y are in X, then R is also an equivalence relation. An equivalence relation is often denoted by the symbol $\sim$. The relations of congruence and similarity between figures are equivalence relations. If X is the set of integers and $x \equiv y$ means that $x - y$ is even, then the relation $\equiv$ is an equivalence relation.

B. Equivalence Classes and Quotient Sets

Let R be an equivalence relation. "xRy" is read: "x and y are **equivalent**" (or "x is **equivalent to** y"). The subset of X consisting of all elements equivalent to an element a is called the **equivalence class** of a. By (1), (2), and (3), each equivalence class is nonempty, the equivalence class of a contains a, and different equivalence classes do not overlap. Namely, X is decomposed into a †disjoint union of equivalence classes. This †partition is called the **classification** of X with respect to R. For example, the set of integers is classified into the equivalence class of even numbers and that of odd numbers by the relation $\equiv$. Conversely, since the relation "x and y belong to the same member of a partition" is an equivalence relation, we can regard any partition as a classification. An element chosen from an equivalence class is called a **representative** of the equivalence class. In the example we can take 0 and 1 as the representatives of equivalence classes of even and odd numbers, respectively.

X/R denotes the set of equivalence classes of X with respect to R, and is called the **quotient set** of X with respect to R. The mapping $p: X \to X/R$ that carries x in X into the equivalence class of x is called **canonical surjection** (or **projection**). The idea of equivalence relations can be generalized to deal with the case when X is a †class.

C. Stronger and Weaker Equivalence Relations

Let R and S be two equivalence relations on X. If xRy always implies xSy, then we say that R is **stronger** than S, S is **weaker** than R, the classification with respect to R is **finer** than the one with respect to S, or the classification with respect to S is **coarser** than the one with respect to R. The relations "x is equal to y" and "x and y are in X" are the strongest and the weakest equivalence relations, respectively. Any two equivalence relations on X are ordered by their strength, and the set of equivalence relations on X forms a †complete lattice with respect to this ordering.

References

[1] N. Bourbaki, Eléments de mathématique I, Théorie des ensembles, ch. 2, Actualités Sci. Ind., 1212c, Hermann, second edition, 1960; English translation, Theory of sets, Addison-Wesley, 1968.
Also → references to 381 Sets.

136 (XVII.15) Ergodic Theory

A. General Remarks

The origin of ergodic theory was the so-called **ergodic hypothesis**, which provided the foundation for classical statistical mechanics as created by L. Boltzmann and J. Gibbs toward the end of the 19th century (→ 402 Statistical Mechanics). Attempts by various mathematicians to give a rigorous proof of the hypothesis resulted in the **recurrence theorem** of H. Poincaré and C. Carathéodory and the **ergodic theorems** of G. D. Birkhoff and J. von Neumann, which marked the beginnings of ergodic theory as we know it today. As the theory developed it acquired close relationships with other branches of mathematics, for example, the theory of dynamical systems, probability theory, functional analysis, number theory, differential topology, and differential geometry.

The principal object of modern ergodic theory is to study properties of †measurable transformations, particularly transformations with an invariant measure. In most cases, the transformations studied are defined on a **Lebesgue measure space with a finite** (or **σ-finite**) **measure**. A Lebesgue measure space with a finite measure (σ-finite measure) is a †measure space that is measure-theoretically isomorphic to a bounded interval (to the real line) with the usual †Lebesgue measure, possibly together with an at most countable number of atoms. It is known that any separable complete †metric space with a complete regular Borel †probability measure is a Lebesgue measure space with a finite measure. We assume, unless stated otherwise, that the measure space $(X, \mathscr{B}, m)$ is a Lebesgue measure space. All the subsets of X mentioned are assumed measurable, and a pair of sets or functions that coincide almost everywhere are identified. We use the abbreviation "a.e." to denote "†almost everywhere."

B. Ergodic Theorems

Let $(X, \mathscr{B}, m)$ be a σ-finite measure space. A transformation φ defined on X is called **measurable** if for every $B \in \mathscr{B}$, $\varphi^{-1}(B) \in \mathscr{B}$. A †bijective transformation φ on X is called **bimeasurable** if both φ and φ^{-1} are measurable. A measurable transformation φ is called **measure-preserving** (or equivalently, the measure m is **invariant** under φ) if $m(\varphi^{-1}(B)) = m(B)$ holds for every B. It is called **nonsingular** if $m(\varphi^{-1}(B)) = 0$ whenever $m(B) = 0$, and **ergodic**

if $m((\varphi^{-1}(B)\cup B)-(\varphi^{-1}(B)\cap B))=0$ implies either $m(B)=0$ or $m(X-B)=0$.

The **mean ergodic theorem** of von Neumann (*Proc. Nat. Acad. Sci. US*, 18 (1932)) states that if φ is a measure-preserving transformation on $(X,\mathscr{B},m)$, then for every function f belonging to the †Hilbert space $L_2(X)=L_2(X,\mathscr{B},m)$ (→ 168 Function Spaces), the sequence

$$A_nf(x)=\frac{1}{n}\left(\sum_{k=0}^{n-1}f(\varphi^k x)\right)$$

converges in the †norm of $L_2(X)$ as $n\to\infty$ to a function f^* that satisfies $f^*(\varphi x)=f^*(x)$ a.e. The **individual** (or **pointwise**) **ergodic theorem** of Birkhoff (*Proc. Nat. Acad. Sci. US*, 17 (1931)) states that for every f belonging to the †function space $L_1(X)$, the sequence $A_nf(x)$ converges a.e. to f^*. From either of these theorems it follows that for any set E satisfying $\varphi^{-1}(E)=E$ and $m(E)<\infty$, the limit function f^* satisfies $\int_E f^*dm=\int_E f dm$. In particular, if $m(X)=1$ and φ is ergodic, then the limit f^* equals the constant $\int f dm$ a.e. This fact therefore gives a mathematical justification to the ergodic hypothesis, which states that the "time mean" $(\sum_{k=0}^{n-1}f(\varphi^k x))/n$ of what is observable over a sufficiently long time can be replaced by the "phase mean" $\int f dm$. Both von Neumann's theorem and Birkhoff's theorem were subsequently generalized in various directions by many authors.

(1) Mean ergodic theorems are concerned with the †strong convergence of the sequence of averages $A_n=(\sum_{k=0}^{n-1}T^k)/n$ of the iterates of a †bounded linear operator T on some †Banach space. A generalization of von Neumann's theorem due to F. Riesz, K. Yosida, and S. Kakutani dispenses with the assumptions that the linear operator T is induced by a measure-preserving transformation φ and that T acts on the Hilbert space $L_2(X)$. A version of this generalization states that if a linear operator T defined on a Banach space $\mathscr{X}$ satisfies the conditions

(i) $\sup_{n\geqslant 1}\left\|\frac{1}{n}\sum_{k=0}^{n-1}T^k\right\|<\infty$,

(ii) $\lim_{n\to\infty}\frac{1}{n}\|T^n\|=0$,

then for an element $f\in\mathscr{X}$ the sequence of averages A_nf converges strongly to an element $f^*\in\mathscr{X}$ if and only if there exists a subsequence converging †weakly to f^*. From this theorem of Riesz, Yosida, and Kakutani follows the L_p-mean ergodic theorem $(1<p<\infty)$ for so-called **Markov operators**: If T is a linear operator defined on each of the Banach spaces $L_p(X)$ $(1\leqslant p\leqslant\infty)$ by means of the formula $Tf(x)=\int f(y)P(x,dy)$, where $P(x,B)$ is the †transition probability of a †Markov process on $(X,\mathscr{B})$ leaving the measure m invariant (i.e., $\int P(x,B)dm=m(B)$ for every $B\in\mathscr{B}$) (→ 261 Markov Processes), then for every f belonging to $L_p(X)$ $(1<p<\infty)$ the sequence A_nf converges in the norm of $L_p(X)$ to a limit function f^*.

(2) Birkhoff's ergodic theorem has been extended to the following individual ergodic theorem by E. Hopf (1954): If T is a †positive linear operator mapping $L_1(X)$ into $L_1(X)$ and $L_\infty(X)$ into $L_\infty(X)$ with $\|T\|_1\leqslant 1$ and $\|T\|_\infty\leqslant 1$, then for every f in $L_1(X)$ the sequence A_nf converges a.e. to a limit f^*. If T is a Markov operator, then T maps each $L_p(X)$ into itself and satisfies $\|T\|_p\leqslant 1$ for each p $(1\leqslant p\leqslant\infty)$, and therefore Hopf's ergodic theorem applies to such T. Special cases of this theorem were proved earlier by J. Doob and by Kakutani. Later, N. Dunford and J. Schwartz showed that the assumption of the positivity of T can be dispensed with in Hopf's theorem. For a positive linear operator T on $L_1(X)$ satisfying $\|T\|_1\leqslant 1$, R. Chacon and D. Ornstein (1960) proved that the **ratio ergodic theorem** holds: For every pair of functions f and g in $L_1(X)$ with $g\geqslant 0$ a.e., $\lim_{n\to\infty}\sum_{k=0}^{n-1}T^kf(x)/\sum_{k=0}^{n-1}T^kg(x)$ exists and is finite a.e. on the set $\{x\,|\,\sum_{k=0}^{n-1}T^kg(x)>0\}$. This theorem extends earlier results of Hopf and W. Hurewicz dealing with special classes of operators arising from measurable transformations. Hopf's ergodic theorem can be deduced from the Chacon-Ornstein theorem, while it is known that there are positive operators T on $L_1(X)$ satisfying $\|T\|_1\leqslant 1$ for which $\lim_{n\to\infty}A_nf$ fails to exist on a set of positive measure for some $f\in L_1(X)$. This shows that the assumption $\|T\|_\infty\leqslant 1$ is crucial in Hopf's theorem.

(3) As was the case in the original proof by Birkhoff of his ergodic theorem, every known proof of an individual ergodic theorem depends crucially on the so-called **maximal ergodic lemma** (or **maximal inequality**). For the case of a positive linear operator T on $L_1(X)$ with $\|T\|_1\leqslant 1$, Hopf proved the relevant maximal ergodic lemma: If $E(f)$ is the set $\{x\,|\,\sup_{n\geqslant 1}A_nf(x)>0\}$ for each f in $L^1(X)$, then $\int_{E(f)}f dm\geqslant 0$. Hopf's original proof of this lemma was quite intricate, but A. Garcia (1965) succeeded in giving an extremely simple and elegant proof. From the maximal ergodic lemma the following so-called **dominated ergodic theorem** can also be obtained: Let T be a linear operator mapping each $L_p(X)$ into $L_p(X)$ satisfying $\|T\|_p\leqslant 1$ for each p $(1\leqslant p\leqslant\infty)$. If for f in $L_p(X)$ we let $\tilde{f}(x)=\sup_{n\geqslant 1}|A_nf(x)|$, then if $1<p<\infty$ we have

$$\int\tilde{f}(x)^p dm\leqslant\frac{2^p p}{p-1}\int|f(x)|^p dm,$$

while if $p=1$ and $m(X)<\infty$ we have

$$\int \tilde{f}(x)\,dm \leqslant 2\left[m(X)+\int |f(x)|\log^{+}|f(x)|\,dm\right].$$

This theorem was obtained first by N. Wiener (1939) for the special case of T induced by a measure-preserving transformation. M. Akcoglu (1975) proved an isometric "dilation" theorem for positive contractions (i.e., positive linear operators T for which $\|T\|\leqslant 1$) in the space $L_p(X)$ for $1<p<\infty$, by means of which he was able to deduce a dominated ergodic theorem for an arbitrary positive contraction in $L_p(X)$ from the corresponding theorem for a positive linear isometry in $L_p(X)$ proved earlier by A. Ionescu-Tulcea. From this theorem of Akcoglu one can obtain the following individual ergodic theorem: If T is a positive contraction in $L_p(X)$ for some p with $1<p<\infty$, then for every f in $L_p(X)$, the sequence $A_n f$ converges a.e.

(4) Both mean and individual ergodic theorems can be extended without difficulty to a continuous time parameter semigroup $\{T_t \mid t\geqslant 0\}$ of bounded linear operators such that $T_t T_s = T_{t+s}$ ($T_0=I$), under a suitable continuity assumption on T_t with respect to t, by replacing the discrete time average $(\sum_{k=0}^{n-1} T^k)/n$ with $(\int_0^t T_s\,ds)/t$. Further extensions to n-parameter semigroups were obtained by N. Wiener and by N. Dunford and A. Zygmund. For mean ergodic theorems, even further extensions were possible to †amenable semigroups of bounded linear operators. For 1-parameter semigroups, the behavior of the mean at zero, $(\int_0^t T_s\,ds)/t$ as $t\downarrow 0$ (**local ergodic theorem**) or $\lambda\int_0^\infty e^{-\lambda s}T_s f\,ds$ as $\lambda\downarrow 0$ or $\lambda\uparrow\infty$ (**Abelian ergodic theorem**), has also been investigated by Wiener, U. Krengel, E. Hille, Yosida, and others. Abelian ergodic theorems are related to properties of the †resolvent of the semigroup $\{T_t\}$ (→ 378 Semigroups of Operators and Evolution Equations). Further extensions of all these theorems in various directions have been given by many authors; for these extensions and related topics → [5–7].

(5) J. F. C. Kingman (1968) proved an interesting and useful extension of (both mean and individual) ergodic theorems, called the **subadditive ergodic theorem**. A real-valued †stochastic process $\{X_{i,k} \mid 0\leqslant i<k,\ k=1,2,3,\dots\}$ is called a **subadditive process** if it satisfies the following conditions: (i) Whenever $i<j<k$, $X_{i,k}\leqslant X_{i,j}+X_{j,k}$. (ii) The †joint distribution of $\{X_{i+1,j+1}\}$ is the same as that of $\{X_{i,j}\}$. (iii) The †expectation $g_k=E(X_{0,k})$ exists, and satisfies $g_k\geqslant -Ak$ for some constant A and for all $k>1$. Kingman proved that if $\{X_{i,k}\}$ is a subadditive stochastic process, then the limit $\xi=\lim_{n\to\infty}(1/n)X_{0,n}$ exists a.e. and in the mean, and $E(\xi)=\lim_{n\to\infty}(1/n)g_n$. Note that if φ is a measure-preserving transformation on a finite measure space $(X,\mathscr{B},m)$ and if f is an element of $L_1(X)$, then $X_{i,k}=\sum_{j=i}^{k-1} f\circ\varphi^j$ for $0\leqslant i<k$, where $k=1,2,3\dots$ defines a subadditive process (in fact, $X_{i,k}=X_{i,j}+X_{j,k}$ for $i<j<k$). The **multiplicative ergodic theorem**, proved by V. Oseledec, which plays a significant role in ergodic theory of †nonhyperbolic smooth dynamical systems and which has found applications also in †algebraic groups, can be derived from the subadditive ergodic theorem [8, 9].

C. Recurrence and Invariant Measures

In this section we assume that the measure space $(X,\mathscr{B},m)$ is nonatomic. A nonsingular measurable transformation φ defined on $(X,\mathscr{B},m)$ is called **recurrent (infinitely recurrent)** if for every set B and for almost all $x\in B$, there exists an $n\in\mathbf{Z}^+$ (infinitely many $n\in\mathbf{Z}^+$) such that $\varphi^n(x)\in B$. A set W is called **wandering under** φ if $\varphi^{-n}(W)\cap\varphi^{-k}(W)=\varnothing$ for $n\neq k$. The transformation φ is called **conservative** if no sets of positive measure are wandering under φ, and **incompressible** if $B\supset\varphi^{-1}B$ implies $m(B-\varphi^{-1}B)=0$. The following statements about a nonsingular measurable transformation φ are equivalent: (i) φ is recurrent; (ii) φ is infinitely recurrent; (iii) φ is incompressible; (iv) φ is conservative. An immediate consequence of this is the following **recurrence theorem** of Poincaré (in the form formulated by Carathéodory): A measure-preserving transformation on a finite measure space is infinitely recurrent. In fact, in order for a nonsingular measurable transformation φ to be recurrent (and hence infinitely recurrent) it is sufficient that there exist a finite measure μ invariant under φ and equivalent to (i.e., mutually †absolutely continuous with) the given measure m.

The **invariant measure problem** is one of the basic problems in ergodic theory and is formulated abstractly in the following way: Given a nonsingular measurable transformation φ on a σ-finite measure space $(X,\mathscr{B},m)$, find necessary and sufficient conditions for the existence of a finite (or σ-finite) measure invariant under φ and equivalent to m. The given measure m specifies only the class of equivalent measures among which an invariant measure is to be found. Therefore we can assume without loss of generality that m is a finite measure. For the remainder of this section, unless we explicitly state otherwise, we always mean by an invariant measure the one that is equivalent to m.

The Poincaré recurrence theorem states that φ being recurrent is necessary for the existence

of a finite invariant measure. The recurrence of φ is, however, not sufficient, since an ergodic transformation with an infinite but σ-finite invariant measure is recurrent and has no finite invariant measure. A necessary and sufficient condition for the existence of a finite invariant measure was given by a theorem of A. Hajian and Kakutani (1964) which states: φ has a finite invariant measure if and only if φ has no **weakly wandering sets** (a set W is called **weakly wandering under** φ if there exists an infinite subset $\{n_k\}$ of $\mathbf{Z}^+$ such that $\varphi^{-n_k}W \cap \varphi^{-n_j}W = \varnothing, k \neq j$). Hajian also proved that a bimeasurable transformation φ has a finite invariant measure if and only if φ is **strongly recurrent** in the following sense: For every set E with $m(E)>0$, there exists a positive integer $k=k(E)$ such that

$$\max_{0 \leqslant j \leqslant k} m(\varphi^{n-j}E \cap E)>0$$

for every $n \in \mathbf{Z}$.

H. Fürstenberg [11] obtained the following striking extension of the Poincaré recurrence theorem: If a bimeasurable transformation φ possesses a finite invariant measure, then for every set E with $m(E)>0$ and for every integer $k \geqslant 2$, there exists $n \geqslant 1$ such that $m(E \cap \varphi^n E \cap \varphi^{2n} E \cap \ldots \cap \varphi^{(k-1)n} E)>0$. From this theorem one can deduce a difficult theorem of E. Szemerédi on †arithmetic progressions which states: Any subset of the integers having positive †upper density contains arithmetic progressions of arbitrary length. In fact, it is not difficult to show that the theorems of Fürstenberg and of Szemerédi are mutually equivalent, and for this reason the theorem of Fürstenberg is sometimes referred to as the **ergodic Szemerédi theorem**.

For a nonsingular bimeasurable transformation φ, a pair of sets A and B are said to be **countably equivalent under** φ if there exist countable decompositions $\{A_k \mid k \in \mathbf{Z}^+\}$ and $\{B_k \mid k \in \mathbf{Z}^+\}$ for A and B, respectively, and an infinite subset $\{n_k\}$ of $\mathbf{Z}$ such that $\varphi^{n_k} A_k = B_k$ for each k. A and B are said to be **finitely equivalent under** φ if finite decompositions $\{A_k\}$ and $\{B_k\}$ can be chosen. It was proved by Hopf that (i) φ is recurrent if and only if no set of positive measure is finitely equivalent under φ to one of its proper subsets, and (ii) φ has a finite invariant measure if and only if no set of positive measure is countably equivalent under φ to one of its proper subsets. It can be shown that if φ is ergodic, then φ has no σ-finite invariant measure if and only if every pair of sets of positive measure are countably equivalent under φ.

The first example of a transformation admitting no σ-finite invariant measure was constructed by Ornstein in 1960. Since then, a number of simpler examples have been obtained by L. Arnold, Brunel, and others. It is now known that there are many different types of transformations having no σ-finite invariant measure (→ Section F). Furthermore, it was shown by Ionescu-Tulcea that in the group of all nonsingular bimeasurable transformations with a suitable metric, those having a σ-finite invariant measure form a subset of the †first category.

Various extensions of results on the invariant measure problem for nonsingular transformations to the case of †Markov processes having nonsingular transition probabilities were obtained by Y. Ito, J. Neveu, S. Foguel, and others [12].

For investigation of detailed properties of particular transformations arising from classical dynamical systems, problems in number theory, and so on, rather than the existence of invariant measures it is more important to determine a specific form of an invariant measure with desirable properties and to develop methods to decide when such a measure is unique. Various people have considered special classes of transformations; these workers have been able to obtain explicit descriptions of invariant measures with nice properties for the transformations in question and have derived interesting consequences. Most important among these is the so-called **Gibbs measure**, introduced and investigated by Ya. Sinaĭ. He obtained this notion by generalizing the concept of the **equilibrium Gibbs distribution**, which plays a prominent role in †statistical mechanics. It is defined in the following way: Let X be a compact metric space, φ a homeomorphism on X, and μ_0 a probability measure on X invariant under φ. For a function g belonging to $L_\infty(X)$ and for $m, n>0$ let

$$\Xi_{m,n}(g \mid \mu_0)=\int_X \exp\left(\sum_{k=-n}^{m} g(\varphi^k x)\right) d\mu_0(x),$$

and form a sequence of probability measures $\mu_{m,n}(g)$ absolutely continuous with respect to μ_0, for which the †Radon-Nikodym derivative

$$\frac{d\mu_{m,n}(x \mid g)}{d\mu_0(x)}=\frac{\exp \sum_{k=-n}^{m} g(\varphi^k x)}{\Xi_{m,n}(g \mid \mu_0)}$$

holds. A measure that is a limit point, in the sense of †weak convergence, of the sequence of measures $\mu_{m,n}(g)$ is called a Gibbs measure constructed from μ_0 and g. It is clear that a Gibbs measure is invariant under φ. For mixing topological Markov shifts (→ Section D), Sinaĭ showed that if one starts with the invariant measure μ_0 of **maximal entropy** (→ Section H), the existence of which was earlier shown by W. Parry, one can determine a class of functions g for which the Gibbs measure

is unique, i.e., $\mu(g)=\lim_{m,n\to\infty}\mu_{m,n}(g)$ exists. Sinaĭ further investigated the properties of the unique measure $\mu(g)$ in detail. R. Bowen and D. Ruelle, defining the Gibbs measure somewhat differently, investigated the existence and uniqueness of such measures, thereby recapturing the results of Sinaĭ in the case of mixing topological Markov shifts. With the aid of **Markov partitions** (→ Section G), these results on Gibbs measures were carried over to the case of †Anosov and "†axiom A" diffeomorphisms, and they provided essential tools in the investigation of the ergodic behavior of these transformations. For details → [13, 14].

The explicit form of the density of the invariant measure with respect to the Lebesgue measure for the transformation associated with †continued fraction expansion was already known to Gauss, and one can draw from it numerous conclusions about metric properties of continued fractions. The way in which Gauss was able to determine this invariant measure, however, was never explained. Recently, Sh. Ito, H. Nakada, and S. Tanaka (*Keio Eng. Rep.*, 30 (1977)) developed an interesting method to describe the mechanism used to arrive at this density function for the invariant measure for the continued fraction transformation. They employed similar methods in subsequent work to determine explicit density functions for invariant measures for other related number-theoretic transformations and for certain classes of continuous mappings over an interval; by means of the explicit forms of invariant measures they were able to describe the metric properties of these transformations in detail.

D. Examples and Construction of Measure-Preserving Transformations

Examples of measure-preserving transformations appear in many different contexts. We describe some of the important ones.

(1) Let G be a †locally compact Abelian group satisfying the second axiom of †countability (→ 423 Topological Groups), $\mathscr{B}$ a σ-algebra of Borel subsets of G, and m its †Haar measure (normalized if G is compact). Then $(G,\mathscr{B},m)$ is a Lebesgue measure space. For a fixed element $g_0\in G$, define the transformation $\varphi_{g_0}:G\to G$ by $\varphi_{g_0}(g)=g+g_0$. Then φ_{g_0} is a bijective measure-preserving transformation on $(G,\mathscr{B},m)$ and is called the **rotation** on G by the element g_0. If G is compact, then the rotation φ_{g_0} is ergodic if and only if the cyclic subgroup generated by the element g_0 is dense in G. If this happens, the element g_0 is called the **topological generator** of G. A group is called **monothetic** if it has a topological generator.

(2) If φ is a group †endomorphism of a compact Abelian group G, then φ preserves the Haar measure m. If φ is a group †automorphism, then it is a bijective measure-preserving transformation on $(G,\mathscr{B},m)$. A continuous group automorphism φ induces a group automorphism φ^* of the character group G^*. The measure-preserving transformation φ is ergodic if and only if every character except the identity has an infinite orbit under the induced automorphism φ^*. When the group is the n-dimensional torus $\mathbf{T}^n$, a continuous group automorphism φ is uniquely represented by an $n\times n$ matrix with integer entries and with determinant ±1. In this case, φ is ergodic if and only if no roots of unity appear among the †eigenvalues of the representing matrix.

(3) Let $(Y,\mathscr{A})$ be a measurable space, let $(Y_n,\mathscr{A}_n)=(Y,\mathscr{A})$ for each $n\in\mathbf{Z}$, and define $(Y^*,\mathscr{A}^*)$ to be the †product measurable space $(\prod_{n\in\mathbf{Z}}Y_n,\prod_{n\in\mathbf{Z}}\mathscr{A}_n)$. The transformation φ defined on $(Y^*,\mathscr{A}^*)$ by

$$y^*=(\dots,y_{-1},y_0,y_1,\dots)\to\varphi(y^*)=(\dots,y'_{-1},y'_0,y'_1,\dots)$$

with $y'_n=y_{n+1}$ for each n is called the **shift transformation**. Let μ be a probability measure on $(Y,\mathscr{A})$ such that $(Y,\mathscr{A},\mu)$ is a Lebesgue measure space, let $\mu_n=\mu$ for each n, and define μ^* to be the †product measure $\prod_{n\in\mathbf{Z}}\mu_n$ on $(Y^*,\mathscr{A}^*)$. Then $(Y^*,\mathscr{A}^*,\mu^*)$ is a Lebesgue measure space, and the shift transformation φ is a bijective measure-preserving transformation. Considered with the product measure, φ is called a **generalized Bernoulli shift**. When the set Y is at most countable and the measure μ on Y is given by a sequence $\{p_j\}$ of positive numbers with $\sum p_j=1$, φ is called a **Bernoulli shift**. Suppose that $P(y,A)$ is a Markov transition function on $(Y,\mathscr{A})$ and π is a probability measure invariant under $P(y,A)$. Then we can define the **Markov measure** π^* on the product space $(Y^*,\mathscr{A}^*)$ by setting

$$\pi^*(E^*)=\int_{A_{i+s}}\dots\int_{A_{i+1}}\int_{A_i}\pi(dy_0)P(y_0,dy_1)\times P(y_1,dy_2)\dots P(y_{s-1},dy_s),$$

for a cylinder set

$$E^*=\Bigl(\prod_{j<i}Y_j\Bigr)\times\Bigl(\prod_{i\leqslant k\leqslant i+s}A_k\Bigr)\times\Bigl(\prod_{n>i+s}Y_n\Bigr),$$

and extending it to all of $\mathscr{A}^*$. The shift transformation φ preserves the Markov measure π^*. Considered as a measure-preserving transformation on $(Y^*,\mathscr{A}^*,\pi^*)$, φ is called a **Markov shift**.

A generalized Bernoulli shift is always ergodic. A Markov shift is ergodic if and only if the corresponding Markov process is †irreducible, which is the case if and only if the following property is satisfied: For every pair of sets A and B with $\pi(A)\pi(B)>0$, there exists an $n\in\mathbf{Z}^+$ such that $\int_A P^n(y,B)\,d\pi>0$.

There are other measures besides the product measure and the Markov measure that can be defined on the product space $(Y^*,\mathscr{A}^*)$ and are invariant under the shift transformation φ. For example, if Y is a Borel subset of $\mathbf{R}$ and $\mathscr{A}$ is the σ-algebra of Borel subsets of Y, then any †stationary stochastic process taking values in Y induces such a measure on $(Y^*,\mathscr{A}^*)$. When considered with a measure of this type, the shift transformation φ is called the **shift associated with the stationary process.** Properties of the shift associated with a †stationary Gaussian process have been investigated by G. Maruyama, I. Girsanov, H. Totoki, and others. In particular, it is known that the shift is ergodic if and only if the †spectral measure for the †covariance function of the associated Gaussian process is continuous [15].

(4) Other important examples of measure-preserving transformations arise from classical dynamical systems, which will be described in Section G.

(5) There are several ways of constructing new measure-preserving transformations from given ones. We describe important cases.

(i) Let φ be a nonsingular, measurable, recurrent transformation (not necessarily measure-preserving) on a σ-finite measure space $(X,\mathscr{B},m)$, and let A be a set of positive measure. For $x\in A$, let $n(x)=\min\{n\in\mathbf{Z}^+\mid \varphi^n(x)\in A\}$. The transformation $\varphi_A:A\to A$ defined by $\varphi_A(x)=\varphi^{n(x)}(x)$ is a nonsingular measurable transformation on the measure space $(A,\mathscr{B}\cap A,m_A)$, where $m_A(B)=m(A\cap B)/m(A)$, and it is measure-preserving if φ is. We call φ_A the **transformation induced by φ on A.** It is ergodic if φ is ergodic.

(ii) Let φ be a nonsingular measurable transformation on a σ-finite measure space $(X,\mathscr{B},m)$, and suppose that $\{A_n\}$ is a countable (possibly finite) partition of X. Define a function $f:X\to\mathbf{Z}^+$ by setting $f(x)=n$ for $x\in A_n$, and let $\tilde{X}=\{(x,j)\mid x\in X,\ 1\leqslant j\leqslant f(x)\}$ be a subspace of the product measure space $(X\times\mathbf{Z}^+,\mathscr{B}\times\mathscr{C},m\times\mu)$, where $\mathscr{C}$ is the σ-algebra of all subsets of $\mathbf{Z}^+$, and μ is the measure on $(\mathbf{Z}^+,\mathscr{C})$ defined by $\mu(\{n\})=1$ for each $n\in\mathbf{Z}^+$. The transformation $\tilde{\varphi}$ defined on $\tilde{X}$ by $\tilde{\varphi}(x,j)=(x,j+1)$ if $1\leqslant j<f(x)$ and $=(\varphi(x),1)$ if $j=f(x)$ is a nonsingular measurable transformation on the measure space $(\tilde{X},(\mathscr{B}\times\mathscr{C})\cap\tilde{X},(m\times\mu)_{\tilde{X}})$, and it is measure-preserving if φ is. We call $\tilde{\varphi}$ the **transformation built from φ with the ceiling function f.** If φ is ergodic and $m(A_n)\to0$ as $n\to\infty$, then $\tilde{\varphi}$ is also ergodic.

(iii) Suppose that ψ is a measure-preserving transformation on $(X,\mathscr{B},m)$ and φ_x is a measure-preserving transformation on $(Y,\mathscr{A},\mu)$ for each $x\in X$. Assume that the mapping $(x,y)\to\varphi_x(y)$ is measurable with respect to the σ-algebras $\mathscr{B}\times\mathscr{A}$ and $\mathscr{A}$. The transformation θ defined on the product space $(X\times Y,\mathscr{B}\times\mathscr{A},m\times\mu)$ by $\theta(x,y)=(\psi(x),\varphi_x(y))$ is measure-preserving and is called the **skew product** of ψ and $\{\varphi_x\}$. If $\varphi_x=\varphi$ for all $x\in X$, then we get a **direct product** transformation $\theta(x,y)=(\psi(x),\varphi(y))$.

(iv) A measure-preserving transformation φ on $(Y,\mathscr{A},\mu)$ is said to be a **factor transformation** (or a **homomorphic image**) of a measure-preserving transformation ψ on $(X,\mathscr{B},m)$ if there exists a measurable transformation η from X onto Y such that $m\circ\eta^{-1}=\mu$ and $\varphi\eta=\eta\psi$. If $\mathscr{B}'$ is a σ-subalgebra of $\mathscr{B}$ and ψ leaves $\mathscr{B}'$ invariant (i.e., $\psi^{-1}\mathscr{B}'\subset\mathscr{B}'$), then ψ induces a factor transformation φ on the measure space $(X,\mathscr{B}',m)$. Conversely, if a measure-preserving transformation φ on $(Y,\mathscr{A},\mu)$ is a factor transformation of ψ on $(X,\mathscr{B},m)$ via a mapping η, then $\mathscr{B}'=\eta^{-1}\mathscr{A}$ is a σ-subalgebra of $\mathscr{B}$ invariant under ψ.

(6) A one-parameter family $\{\varphi_t\mid t\in\mathbf{R}\}$ of bijective measure-preserving transformations on a measure space $(X,\mathscr{B},m)$ is called a **flow**. A flow is called **continuous** if the mapping $t\to T_t$ is †weakly continuous where $\{T_t\}$ is the one-parameter family of †unitary operators on $L^2(X)$ induced by the flow $\{\varphi_t\}$. A flow is called **measurable** if the mapping $(t,x)\to\varphi_t(x)$ is a measurable transformation of $\mathbf{R}\times X$ into X. A measurable flow is continuous. A. Vershik and Maruyama proved that for any continuous flow there exists a measurable flow, unique in a specified sense, which is spatially isomorphic (in the sense specified in Section E) to the given flow.

Important examples of flows are given by classical dynamical systems (→ Section G), and by continuous-time stationary stochastic processes.

An important tool in the study of flows is provided by the theorem of W. Ambrose and Kakutani: Every measurable ergodic flow without a fixed point is spatially isomorphic to an S-**flow**. A measurable flow $\{\varphi_t\}$ is called an **S-flow (special flow** or **flow built under a function)** if there exist a measure-preserving transformation φ of a measure space $(X,\mathscr{B},m)$ and an $\mathbf{R}^+$-valued function f on $(X,\mathscr{B},m)$ such that each φ_t is a measure-preserving transformation on the subspace $\tilde{X}=\{(x,u)\mid x\in X,\ 0\leqslant u\leqslant f(x)\}$ of the product measure space

$(X \times \mathbf{R}^+, \mathscr{B} \times \mathscr{M}, m \times \lambda)$ given by

$$\varphi_t(x,u)$$
$$= (x, u+t) \quad \text{if} \quad -u \leqslant t < -u + f(x),$$
$$= \left(\varphi^n(x), u+t-\sum_{k=0}^{n-1} f(\varphi^k(x)) \right) \quad \text{if}$$
$$-u + \sum_{k=0}^{n-1} f(\varphi^k(x)) \leqslant t < -u + \sum_{k=0}^{n} f(\varphi^k(x)),$$
$$= \left(\varphi^{-n}(x), u+t+\sum_{k=-n}^{-1} f(\varphi^k(x)) \right) \quad \text{if}$$
$$-u - \sum_{k=-n}^{-1} f(\varphi^k(x)) \leqslant t < -u - \sum_{k=-n+1}^{-1} f(\varphi^k(x)),$$

$n \geqslant 1$. Here $\mathscr{M}$ is the σ-algebra of Borel subsets of $\mathbf{R}^+$, and λ is the usual Lebesgue measure.

E. Isomorphism Problems

In this section we assume that the Lebesgue measure spaces considered are probability spaces. For simplicity, following the common usage among Russian mathematicians, we call a measure-preserving transformation on $(X, \mathscr{B}, m)$ an **endomorphism** and a bijective measure-preserving transformation an **automorphism**.

An automorphism φ_1 (a flow $\{\varphi_t^{(1)}\}$) on $(X_1, \mathscr{B}_1, m_1)$ is said to be **spatially isomorphic** (or **metrically isomorphic**) to an automorphism φ_2 (a flow $\{\varphi_t^{(2)}\}$) on $(X_2, \mathscr{B}_2, m_2)$ if there exist sets N_1 and N_2 with $m_1(N_1) = m_2(N_2) = 0$ and a bijective measurable transformation θ from $X_1 - N_1$ to $X_2 - N_2$ such that $m_2 \circ \theta = m_1$ and $\theta\varphi_1 = \varphi_2\theta$ ($\theta\varphi_t^{(1)} = \varphi_t^{(2)}\theta$ for each t).

Classification of automorphisms and flows into isomorphism classes constitutes the central problem of modern ergodic theory. Properties of automorphisms and flows that are preserved under spatial isomorphisms are called **isomorphism invariants** (or **metric invariants**). There are several isomorphism invariants that are essential to the study of the isomorphism problem. We describe these below for the case of automorphisms. There are corresponding invariants for flows as well.

(1) Spectral Invariants. Two automorphisms φ_1 and φ_2 are said to be **spectrally isomorphic** if the †unitary operators T_1 and T_2 induced by φ_1 and φ_2 on the Hilbert spaces $L_2(X_1)$ and $L_2(X_2)$, respectively, are unitarily equivalent (i.e., there exists an isometric isomorphism V of $L_2(X_1)$ onto $L_2(X_2)$ such that $VT_1 = T_2V$). Properties preserved under spectral isomorphisms are called **spectral invariants** (or **spectral properties**). If φ_1 and φ_2 are spatially isomorphic, it is clear that they are spectrally isomorphic; but the converse is not true in general.

(i) The property of φ being ergodic is a spectral property since φ is ergodic if and only if the number 1 is a simple eigenvalue of the induced unitary operator T. If φ is ergodic, then the set of all eigenvalues of the induced operator T forms a subgroup of the circle group, each eigenvalue is simple, and each †eigenfunction has constant absolute value. If the †spectrum of the induced operator T consists entirely of eigenvalues, φ is said to have **discrete spectrum** (or **pure point spectrum**). A theorem due to von Neumann and P. Halmos —the first theorem on the question of isomorphism—states that two ergodic automorphisms φ_1 and φ_2 with discrete spectra are spatially isomorphic if and only if they are spectrally isomorphic, which is the case if and only if the induced operators T_1 and T_2 have the same set of eigenvalues. Furthermore, every ergodic automorphism with discrete spectrum is spatially isomorphic to an ergodic rotation on a compact Abelian group [2].

Analogous results were obtained by L. Abramov for a bigger class of automorphisms, namely, for ergodic automorphisms having so-called **quasidiscrete spectra**.

(ii) An automorphism φ is ergodic if and only if for every pair of sets A, B,

$$\lim_{n\to\infty} \frac{1}{n}\left(\sum_{k=0}^{n-1} m(\varphi^k(A) \cap B) \right) = m(A)m(B).$$

Strengthening this condition, we can define φ to be **weakly mixing** if for every pair of sets A, B,

$$\lim_{n\to\infty} \frac{1}{n}\left(\sum_{k=0}^{n-1} |m(\varphi^k(A) \cap B) - m(A)m(B)| \right) = 0;$$

strongly mixing if

$$\lim_{k\to\infty} m(\varphi^k(A) \cap B) = m(A)m(B);$$

and **k-fold mixing** if for arbitrary choice of sets $A_j, j = 0, 1, \ldots, k$,

$$\lim m(A_0 \cap \varphi^{n_1}A_1 \cap \varphi^{n_2}A_2 \ldots \cap \varphi^{n_k}A_k)$$
$$= m(A_0)m(A_1)\ldots m(A_k),$$

where the limit is taken as $n_1, n_2, \ldots, n_k \to \infty$ in such a way that $n_1 < n_2 < \ldots < n_k$ and $\min_{1 \leqslant j \leqslant k}(n_j - n_{j-1}) \to \infty$. The property of an automorphism φ being weakly mixing or strongly mixing is a spectral property. For instance, φ is weakly mixing if and only if the number 1 is a simple eigenvalue and is the only eigenvalue of the induced operator T. It is also known that φ is weakly mixing if and only if the direct product automorphism $\varphi \times \varphi$ is ergodic. The set of all weakly mixing automorphisms forms a dense †G_δ-set in the group

of all automorphisms on $(X, \mathscr{B}, m)$ considered with the so-called weak topology (Halmos's theorem). On the other hand, it was shown by V. Rokhlin that the set of all strongly mixing automorphisms is a set of first category with respect to the weak topology. However, there are only a few known examples of automorphisms that are weakly mixing but not strongly mixing.

(iii) An automorphism φ is said to have **countable Lebesgue spectrum** if the maximal spectral type of the induced unitary operator T restricted to the †orthocomplement of the subspace of constant functions in $L_2(X)$ is equivalent to the Lebesgue measure and its †multiplicity is countably infinite.

(iv) An automorphism φ is called a **K-automorphism** (or **Kolmogorov automorphism**) if there exists a σ-subalgebra $\mathscr{B}_0$ of $\mathscr{B}$ such that (a) $\varphi\mathscr{B}_0 \supset \mathscr{B}_0$ and $\varphi\mathscr{B}_0 \neq \mathscr{B}_0$, (b) $\bigvee_{n\in\mathbf{Z}}\varphi^n\mathscr{B}_0 = \mathscr{B}$, and (c) $\bigwedge_{n\in\mathbf{Z}}\varphi^n\mathscr{B}_0 = \mathscr{N}$, where $\mathscr{N}$ is the σ-subalgebra of $\mathscr{B}$ consisting of null sets and their complements. The notion of a **K-flow** (or **Kolmogorov flow**) is defined similarly. K-automorphisms are k-fold mixing for all orders k and have countable Lebesgue spectra.

Generalized Bernoulli shifts are all K-automorphisms. An ergodic Markov shift is a K-automorphism if and only if it is strongly mixing, which is the case if and only if the corresponding Markov process is †irreducible and †aperiodic (→ 260 Markov Chains B). A continuous group automorphism of a compact Abelian group is a K-automorphism if and only if it is ergodic. In particular, a continuous group automorphism of the n-dimensional torus $\mathbf{T}^n$ is a K-automorphism if and only if no roots of unity appear among the eigenvalues of the representing matrix. Automorphisms and flows arising from classical dynamical systems also provide examples of K-automorphisms and K-flows. In particular, a †geodesic flow on a surface of negative curvature is a K-flow, and each automorphism (except the identity) of this flow is a K-automorphism. For the shift transformation φ associated with a stationary Gaussian process, it was shown by Maruyama [15] that φ is (a) weakly mixing if and only if it is ergodic, (b) strongly mixing if and only if the covariance function of the associated Gaussian process tends to 0 as $n\to\infty$, and (c) a K-automorphism if and only if the †spectral measure of the covariance function is absolutely continuous with respect to the Lebesgue measure.

(v) Examples of automorphisms having various types of spectra have been constructed by a number of authors by using stationary Gaussian processes and the theory of approximation developed by A. Katok and A. Stepin [16].

(vi) An ergodic automorphism with quasi-discrete spectrum has a mixed spectrum, that is, the spectrum of the induced unitary operator T has a continuous component and eigenvalues in addition to 1. Anzai (1951) constructed a special class of skew product automorphisms having mixed spectra and showed that in this class there are automorphisms that are spectrally isomorphic but not spatially isomorphic. However, the question of whether two spatially nonisomorphic automorphisms exist among automorphisms having the same purely continuous spectrum remained unanswered for a long time, until in 1958 Kolmogorov (*Dokl. Akad. Nauk SSSR*, (5) 119 (1958)) settled it affirmatively by using a new isomorphism invariant called **entropy**.

2. Generators and Entropy. (i) By a **partition** $\xi = \{A_\lambda\}$ of the space X we mean a collection of sets A_λ such that $A_\lambda \cap A_{\lambda'} = \varnothing$ whenever $\lambda \neq \lambda'$ and $\bigcup A_\lambda = X$. We denote by ε the partition of X into individual points, and by ν the trivial partition $\{X\}$. A partition into a finite (countable) number of sets is called a finite (countable) partition. A partition ξ is said to be finer than another partition ζ (or ζ is coarser than ξ) if for every $A\in\xi$ there is a set $B\in\zeta$ such that $A\subset B$. For a collection $\{\xi_\alpha\}$ of partitions of X, we denote by $\bigvee_\alpha\xi_\alpha$ the coarsest partition that is finer than each ξ_α, and by $\bigwedge_\alpha\xi_\alpha$ the finest partition that is coarser than each ξ_α. If $\xi_k = \{A_{k,n}\}$ is a sequence of countable (or finite) partitions, then $\bigvee_k\xi_k$ is precisely the partition of X into nonempty intersections of the form $\bigcap_k A_{k,n_k}$ with $A_{k,n_k}\in\xi_k$ for each k. With a partition ξ of X we associate a σ-subalgebra $\mathscr{B}(\xi)$ of $\mathscr{B}$ which is the σ-algebra of all $\mathscr{B}$-measurable sets that are a union of elements in ξ. Two partitions ξ and ζ are said to coincide a.e. if $\mathscr{B}(\xi) = \mathscr{B}(\zeta)$ a.e. (i.e., for every $A\in\mathscr{B}(\xi)$ there exists a set B in $\mathscr{B}(\zeta)$ such that $m(A \cup B - A\cap B) = 0$ and conversely).

(ii) Suppose that φ is an endomorphism of $(X, \mathscr{B}, m)$ and ξ a partition of X. By $\varphi^{-1}\xi$ we mean the partition $\{\varphi^{-1}(A) \mid A\in\xi\}$. If φ is an automorphism, we also define $\varphi\xi = \{\varphi(A) \mid A\in\xi\}$. A partition ξ is called a **generator** for an endomorphism φ if $\bigvee_{n=0}^{\infty}\varphi^{-n}\xi = \varepsilon$ a.s. If $\bigvee_{n=-\infty}^{\infty}\varphi^n\xi = \varepsilon$ a.s., ξ is called a **two-sided generator** for an automorphism φ.

An endomorphism φ is said to be **periodic at a point** $x\in X$ if there exists a positive integer n such that $\varphi^n(x) = x$ and **aperiodic** if the set of points of periodicity has measure zero. If the measure space $(X, \mathscr{B}, m)$ is nonatomic, then every ergodic endomorphism on it is aperiodic. A theorem of Rokhlin states that every aperiodic automorphism φ has a countable two-sided generator. This implies that every such φ is spatially isomorphic to the shift

transformation on the infinite product space $(Y^*, \mathscr{A}^*)$ considered with some invariant measure μ^*, where each coordinate space $Y_j = Y$ has at most a countable number of points. Krieger improved this result by showing that if an ergodic automorphism φ has finite entropy, then φ has a finite two-sided generator [17].

(iii) For a finite or countable partition $\xi = \{A_n\}$, define the **entropy** $H(\xi)$ **of the partition** to be $-\sum_n m(A_n)\log(m(A_n))$ (the logarithms here and below are natural logarithms). We denote by $\mathscr{Z}$ the set of all partitions ξ with $H(\xi) < \infty$. If $\xi \in \mathscr{Z}$, then for any endomorphism φ the limit

$$h(\varphi, \xi) = \lim_{n\to\infty} \frac{1}{n}\left(H\left(\bigvee_{k=0}^{n} \varphi^{-k}\xi\right)\right)$$

exists and is finite. The **entropy** $h(\varphi)$ **of the endomorphism** φ is defined to be $\sup\{h(\varphi, \xi) \mid \xi \in \mathscr{Z}\}$ and is an isomorphism invariant. Properties of entropy have been investigated extensively since the notion was introduced by Kolmogorov. We cite a few results.

(a) If a partition $\xi \in \mathscr{Z}$ is a generator for an endomorphism φ or a two-sided generator for an automorphism φ, then $h(\varphi) = h(\varphi, \xi)$ (Sinaĭ's lemma). (b) For every integer n, $h(\varphi^n) = |n| h(\varphi)$, and for a measurable flow $\{\varphi_t\}$, $h(\varphi_t) = |t| h(\varphi_1)$ for every real number t. (c) If an automorphism φ is periodic, then $h(\varphi) = 0$. (d) If φ_1 is a factor transformation of φ_2, then $h(\varphi_1) \leqslant h(\varphi_2)$. (e) $h(\varphi_1 \times \varphi_2) = h(\varphi_1) + h(\varphi_2)$. (There is a more complicated formula (due to Rokhlin) for the entropy of a skew product automorphism.) (f) If φ is a recurrent automorphism and φ_A is the automorphism induced by φ on a subset A with $m(A) > 0$, then $h(\varphi_A) = h(\varphi)/m(A)$ (**Abramov's formula**). (g) If φ is a Bernoulli shift with probability distribution $\{p_n\}$, then $h(\varphi) = -\sum_n p_n \log p_n$. (h) If φ is a Markov shift based on the Markov transition probability P_{ij} (defined on a countable or finite state space) and an invariant measure π_i, then

$$h(\varphi) = -\sum_i \sum_j \pi_i P_{ij} \log P_{ij}.$$

(i) If φ is ergodic and has a pure point spectrum, or more generally, has a quasidiscrete spectrum, then $h(\varphi) = 0$. (j) For an ergodic group automorphism φ on an n-dimensional torus, $h(\varphi) = \sum \log|\lambda|$, where the sum is taken over all eigenvalues λ of modulus > 1 of the representing matrix. (k) If an automorphism φ has positive entropy, then in $L_2(X)$ there exists a subspace invariant under the induced unitary operator T such that the spectrum of T restricted to this subspace is countable Lebesgue (Rokhlin's theorem). It follows from (k) that automorphisms with †singular spectra or spectra of finite multiplicity must have zero entropy. In proving assertion (g), Kolmogorov established for the first time the fact that there are uncountably many spatially nonisomorphic Bernoulli shifts.

(iv) An automorphism φ is said to have **completely positive entropy** if $h(\varphi, \xi) > 0$ for every partition $\xi \neq \nu$. It was shown by Rokhlin and Sinaĭ that an automorphism φ has completely positive entropy if and only if φ is a K-automorphism. M. Pinsker proved that for every automorphism φ there exists a partition, called the **Pinsker partition**, that is invariant under φ and such that the factor transformation of φ with respect to this partition has zero entropy and is the largest among the factor transformations of φ with zero entropy.

(v) Rokhlin showed that $h(\varphi) = 0$ for an endomorphism φ if and only if all of its factor transformations are automorphisms. An endomorphism φ is called **exact** if $\bigwedge_{n=0}^{\infty} \varphi^{-n}\varepsilon = \nu$ a.e. Rokhlin introduced a way to associate with each endomorphism a certain automorphism, called the **natural extension**, which reflects the properties of the endomorphism. For example, an endomorphism and its natural extension are simultaneously ergodic or nonergodic, are mixing of the same order, and have equal entropy. The natural extension of an exact endomorphism is a K-automorphism.

(vi) Automorphisms φ_1 and φ_2 are said to be **weakly isomorphic** if each of them is a factor transformation of the other. Sinaĭ proved (1964) that for each ergodic automorphism with positive entropy, there exists a factor automorphism having the same entropy and isomorphic to a Bernoulli shift, and hence it follows in particular that Bernoulli shifts with the same entropy are weakly isomorphic. Ornstein (*Adv. in Math.*, 4 (1970)) went further and succeeded in proving the following remarkable result: Two Bernoulli shifts with equal entropy are spatially isomorphic. Partial results in this direction were obtained earlier by L. Meshalkin, J. Blum, and D. Hanson. In the proof of Ornstein's theorem, essential use was made of the following theorem of C. Shannon and B. McMillan, which plays a fundamental role in information theory (→ 213 Information Theory): Suppose that φ is an ergodic endomorphism on $(X, \mathscr{B}, m)$ and ξ is a partition of X. For a point $x \in X$, let $A_n(x)$ denote the element in the partition $\bigvee_{k=0}^{n-1} \varphi^{-k}\xi$ that contains x. Then for almost all x,

$$\lim_{n\to\infty}\left(-\frac{1}{n}\log m(A_n(x))\right)$$

exists and equals $h(\varphi, \xi)$.

(vii) Techniques and ideas developed by Ornstein in his proof of the isomorphism theorem have been refined and extended further by himself, B. Weiss, N. Friedman, M. Smorodinsky, and others, and numerous re-

sults were subsequently obtained on the isomorphism problem. We describe below some of the main results and basic concepts; for more detailed accounts → [21–24].

A sequence $\{\xi_n\}$ of partitions is said to be **independent** if for every choice of sets $A_1, \ldots, A_s$ with $A_k \in \xi_{n_k}$, $m(\bigcap_{k=1}^s A_k) = \prod_{k=1}^s m(A_k)$ whenever $n_1, n_2, \ldots, n_s$ are all distinct. For a fixed $\varepsilon > 0$, two partitions ξ and ζ are said to be ε-**independent** if

$$\sum_{A \in \xi, B \in \zeta} |m(A \cap B) - m(A)m(B)| < \varepsilon.$$

A pair (φ, ξ), where φ is an automorphism of $(X, \mathscr{B}, m)$ and ξ is a finite (or countable) partition of X, is called a **process** (on X). $\mathscr{B}(\bigvee_{n=-\infty}^{\infty} \varphi^n \xi)$ is a σ-subalgebra of $\mathscr{B}$ invariant under φ, and φ restricted to this σ-subalgebra is a factor automorphism of φ and is isomorphic to a shift transformation on an infinite product space, where each coordinate space has the same number of points as the number of atoms in ξ. A process (φ, ξ) is called a **Bernoulli process** (or an **independent process**) if the sequence of partitions $\{\varphi^n \xi \mid n \in \mathbf{Z}\}$ is independent, and a **weak Bernoulli (W.B.) process** if for every $\varepsilon > 0$ there exists a $k > 0$ such that the partitions $\bigvee_{i=k}^{k+n} \varphi^i \xi$ and $\bigvee_{i=-n}^{0} \varphi^i \xi$ are ε-independent for all $n \geqslant 0$. Let J be a finite set and α, β be two functions from the set $\{1, 2, \ldots, N\}$ to J. By $d_N(\alpha, \beta)$ we denote $(1/N) \# \{n \mid \alpha(n) \neq \beta(n)\}$, where $\#$ denotes the cardinality. d_N defines a metric on the space $J^{\{1,2,\ldots,N\}}$ called the **Hamming distance**. Now, for a process (φ, ξ) on X with $\xi = \{A_j \mid j \in J\}$, we define for $x, y \in X$, $d_N^{\xi}(x, y)$ to be equal to $d_N(\xi_\varphi^N(x), \xi_\varphi^N(y))$, where for $x \in X$, $\xi_\varphi^N(x)$ is the point $(j(x), j(\varphi(x)), \ldots, j(\varphi^{N-1}(x))) \in J^{\{1,2,\ldots,N\}}$ and $j(x)$ is the index $j \in J$ for which $x \in A_j$. A process (φ, ξ) is called **very weak Bernoulli (V.W.B.)** if for any $\varepsilon > 0$ there exists an N_0 such that whenever $N > N_0$ there is a set G with $m(G) > 1 - \varepsilon$ belonging to $\mathscr{B}(\bigvee_{n=-\infty}^{0} \varphi^n \xi)$ satisfying the following condition: for any $A \in \mathscr{B}(\bigvee_{n=-\infty}^{0} \varphi^n \xi)$ with $A \subset G$ and $m(A) > 0$, there exists a probability measure ν on $X \times X$ satisfying (i) $\nu(E \times X) = m(E)$ and $\nu(X \times F) = m_A(F)$ for all $E, F \in \mathscr{B}$ and (ii) $\int_{X \times X} d_N^{\xi}(x, y) d\nu < \varepsilon$. It is not difficult to show that W.B. processes are V.W.B. For processes (φ, ξ) on $(X, \mathscr{B}, m)$ and $(\bar{\varphi}, \bar{\xi})$ on $(\bar{X}, \bar{\mathscr{B}}, \bar{m})$, where both ξ and $\bar{\xi}$ are indexed by the same finite set J, we define $\bar{d}_N((\varphi, \xi), (\bar{\varphi}, \bar{\xi}))$ to be $\inf\{\int_{X \times \bar{X}} d_N(\xi_\varphi^N(x), \bar{\xi}_{\bar{\varphi}}^N(\bar{x})) d\rho(x, \bar{x})\}$, where inf is taken over all probability measures ρ on $X \times \bar{X}$ satisfying $\rho(E \times \bar{X}) = m(E)$ for all $E \in \mathscr{B}$ and $\rho(X \times \bar{E}) = \bar{m}(\bar{E})$ for all $\bar{E} \in \bar{\mathscr{B}}$. $\bar{d}_N$ decreases with N, and we define $\bar{d}((\varphi, \xi), (\bar{\varphi}, \bar{\xi}))$ to be $\lim_{N \to \infty} \bar{d}_N((\varphi, \xi), (\bar{\varphi}, \bar{\xi}))$. Finally, a process (φ, ξ) is said to be **finitely determined (F.D.)** if for any $\varepsilon > 0$ there exists a $\delta > 0$ and N such that if $(\bar{\varphi}, \bar{\xi})$ is any process satisfying (i) ξ and $\bar{\xi}$ are indexed by the same set J, (ii) $h(\varphi, \xi) - h(\bar{\varphi}, \bar{\xi}) < \delta$, and (iii) $\bar{d}_N((\varphi, \xi), (\varphi, \bar{\xi})) < \delta$, then $\bar{d}((\varphi, \xi), (\bar{\varphi}, \bar{\xi})) < \varepsilon$. Let us call a partition ξ an **F.D. generator** for an automorphism φ if ξ is a two-sided generator for φ and the process (φ, ξ) is finitely determined. A general isomorphism theorem proved by Ornstein states: If automorphisms φ and $\bar{\varphi}$ have the same entropy and both have F.D. generators, then φ and $\bar{\varphi}$ are isomorphic. Ornstein and Weiss proved further that the process (φ, ξ) is F.D. if and only if it is V.W.B. From these basic results a number of results can be deduced. (a) Nontrivial factors of Bernoulli shifts are Bernoulli. (b) Strongly mixing Markov shifts have two-sided weak Bernoulli generators and hence are isomorphic to Bernoulli shifts. (c) Every Bernoulli shift can be embedded in a flow, which implies among other things that Bernoulli shifts have roots of all orders. (d) It was shown by Y. Katznelson that every ergodic group automorphism of an n-dimensional torus has a two-sided generator which can be shown to be V.W.B., and hence is also isomorphic to a Bernoulli shift. R. Adler and B. Weiss earlier proved by entirely different methods that on a 2-dimensional torus, two ergodic group automorphisms having the same entropy are spatially isomorphic. They have done this by constructing a two-sided generator ξ for such automorphisms which is also a Markov partition (→ Section G).

(viii) Among further positive results there are the following: (a) Except for a trivial change in the time scale, any two Bernoulli flows are isomorphic (Ornstein). (b) Every ergodic group automorphism of a compact group is isomorphic to a Bernoulli shift (Thomas and Miles; Lind). (c) A number of automorphisms and flows arising from classical dynamical systems are shown to be Bernoulli (→ Section G). (d) Examples of exact endomorphisms arise in connection with problems in number theory; for example, †continued fraction expansion and β-expansion. The natural extensions of the exact endomorphisms associated with continued fraction expansion and β-expansion are now known to be spatially isomorphic to Bernoulli shifts.

(ix) Since many of the automorphisms that have been shown to be K-automorphisms are now known to be isomorphic to Bernoulli shifts, it is natural to expect that every K-automorphism is in fact Bernoulli. However, Ornstein (1973) constructed an example of a K-automorphism that is not a Bernoulli shift. A lot of work has since been done to investigate how bad K-automorphisms can be. It turns out that K-automorphisms share almost none of the finer properties of Bernoulli shifts. For example: (a) There are uncountably many

nonisomorphic K-automorphisms of the same entropy (Ornstein and P. Shields). (b) There is a K-flow that is not Bernoulli, and there are uncountably many nonisomorphic K-flows having the same entropy at time one (Smorodinsky). (c) There exists a K-automorphism not isomorphic to its inverse (Ornstein and Shields). (d) There exists an automorphism that cannot be written as a direct product of a K-automorphism and an automorphism with zero entropy. This example shows that a conjecture made earlier by Pinsker was false (Ornstein). (e) There are weakly isomorphic K-automorphisms that are not isomorphic (Polit and Rudolph).

F. Weak Equivalence and Monotone Equivalence

(1) Weak Equivalence. In order to construct examples of †factors in the theory of von Neumann algebras (→ 308 Operator Algebras), F. Murray and von Neumann considered various ergodic groups of bimeasurable nonsingular transformations on a finite measure space. In this context a group $\mathscr{G}=\{g\}$ of transformations is called ergodic if $m((g^{-1}(B)\cup B)-(g^{-1}(B)\cap B))=0$ for every $g\in\mathscr{G}$ implies either $m(B)=0$ or $m(X-B)=0$. A measure μ is said to be invariant under the group $\mathscr{G}=\{g\}$ if μ is invariant under every transformation g in $\mathscr{G}$. Murray and von Neumann's construction (the so-called **group measure space construction**) gives a type II_1 †factor if the group admits a finite equivalent invariant measure, a type II_∞ factor if it admits an infinite (but σ-finite) equivalent invariant measure, and a type III factor if it has no σ-finite equivalent invariant measure. In this connection we note that Hajian and Y. Itô extended the theorem of Hajian and Kakutani and proved that an arbitrary group $\mathscr{G}=\{g\}$ of nonsingular bimeasurable transformations admits a finite invariant measure if and only if no set of positive measure is weakly wandering under the group $\mathscr{G}$ (a set W is said to be **weakly wandering under a group** $\mathscr{G}$ if there exists an infinite subset $\{g_n \mid n\in\mathbf{Z}^+\}$ of $\mathscr{G}$ such that $g_n(W)\cap g_k(W)=\varnothing$ for $n\neq k$). Analogously to the terminologies used in the theory of von Neumann algebras, we define ergodic countable group $\mathscr{G}$ (or an ergodic bimeasurable nonsingular transformation φ) to be of **type** II_1, II_∞, or III if $\mathscr{G}$ (or φ) has a finite equivalent invariant measure, a σ-finite infinite equivalent invariant measure, or no σ-finite equivalent invariant measure, respectively.

For a bimeasurable nonsingular transformation φ, define the **full group** $[\varphi]$ to be the group of all bimeasurable nonsingular transformations ψ such that, for some $n=n(x)$, $\psi(x)=\varphi^n(x)$ for almost all x. Two transformations φ_1 and φ_2 are said to be **weakly equivalent** if there exists a bimeasurable nonsingular transformation θ such that $\theta[\varphi_1]\theta^{-1}=[\varphi_2]$. H. Dye proved that any pair of type II_1 transformations are weakly equivalent. It easily follows from this that the same is true for type II_∞ transformations. Krieger showed, on the other hand, that among type III transformations there are uncountably many weakly nonequivalent ones. In fact, Krieger [28] introduced an invariant for weak equivalence, called the **ratio set**, by means of which he classified type III transformations into mutually weakly nonequivalent subclasses III_λ for $0\leqslant\lambda\leqslant1$. Krieger showed further that (a) for $0<\lambda\leqslant1$, every pair of transformations in the class III_λ is weakly equivalent; (b) in the class III_0 there are uncountably many mutually weakly nonequivalent transformations; and (c) two ergodic transformations are weakly equivalent if and only if the corresponding factors constructed via the group measure space construction are †*-isomorphic. An ergodic flow of nonsingular transformations (not necessarily measure-preserving) called the **associated flow** was constructed independently by Krieger [29] and by T. Hamachi, Y. Oka, and M. Osikawa [30], and was shown to give another effective invariant for weak equivalence. Krieger showed that the mapping which assigns to each ergodic transformation its associated flow gives a †bijective mapping between the set of all weak equivalence classes of ergodic type III_0 transformations and the set of all isomorphism classes of aperiodic conservative ergodic flows. In this connection the theorem of U. Krengel and I. Kubo on the representation of ergodic flows, extending the theorem of Ambrose and Kakutani mentioned in Section D, plays a significant role.

A bimeasurable nonsingular transformation θ is called a **normalizer** of another transformation φ if it satisfies $\theta[\varphi]=[\varphi]\theta$. The set of all normalizers $\mathscr{N}(\varphi)$ of a transformation φ forms a group called the normalizer group which contains the full group $[\varphi]$ as a subgroup. One can introduce a suitable topology to $\mathscr{N}(\varphi)$ to make it a complete separable metrizable group. Hamachi [31] has shown that, for type III transformations φ, the quotient group $\mathscr{N}(\varphi)/[\varphi]^-$, where $[\varphi]^-$ denotes the closure of $[\varphi]$, is algebraically and topologically isomorphic to the †commutant of the associated flow.

Results obtained by Krieger and others mentioned above were motivated in part by corresponding developments in von Neumann algebra theory due mainly to A. Connes (→ 308 Operator Algebras). From the recent deep result of Connes on the uniqueness of †approximately finite-dimensional factors of type II_∞

it follows that every approximately finite-dimensional factor with the exception of type III_1 is $*$-isomorphic to a factor constructed from an ergodic transformation via the group measure space construction.

(2) Monotone Equivalence. The theory of weak equivalence discussed above deals with the structure of orbits of a transformation or of groups of transformations. In fact, transformations φ and ψ are weakly equivalent if and only if there exists a bimeasurable nonsingular transformation θ mapping the φ-orbit of almost all x onto the ψ-orbit of $\theta(x)$. A somewhat more stringent notion of equivalence dealing with orbit structure is called **monotone equivalence** or **Kakutani equivalence**. We say that measurable flows $\{\varphi_t\}$ and $\{\bar{\varphi}_t\}$ of measure-preserving transformations on finite measure spaces $(X, \mathscr{B}, m)$ and $(\bar{X}, \bar{\mathscr{B}}, \bar{m})$, respectively, are monotonely equivalent if there exists a bimeasurable measure-preserving transformation θ on X to $\bar{X}$ such that for almost all $x \in X$ and all $t \in R$, $\theta\varphi_t(x) = \bar{\varphi}_{\tau(t,x)}\theta(x)$, where $\tau(t, x)$ is a monotone increasing function of t. Two S-flows (→ Section D) built over the same base transformation with different ceiling functions are monotonely equivalent, and it can be shown that monotonely equivalent ergodic flows are isomorphic to S-flows built over the same base transformation. This induces an equivalence relation on transformations, also called monotone equivalence or Kakutani equivalence: Two transformations φ and $\bar{\varphi}$ are monotonely equivalent if they can serve as base transformations of the same (or equivalent) flow. This notion of equivalence was introduced by Kakutani in [32], where he showed that φ and $\bar{\varphi}$ are monotonely equivalent if and only if there are sets $E \subset X$ and $\bar{E} \subset \bar{X}$ such that the induced transformations (→ Section D) φ_E and $\bar{\varphi}_{\bar{E}}$ are isomorphic. Abramov's formula mentioned in Section E implies that there are at least three classes of transformations (and flows) that are mutually nonequivalent: transformations (flows) of zero, finite, and infinite entropy. Nothing much was done in this equivalence theory for a number of years until in 1975 J. Feldman and E. Satayev independently showed that there are monotonely nonequivalent transformations within each class of zero, positive, and infinite entropy. Since then extensive work has been done by Feldman, Satayev, A. Katok, Ornstein, Weiss, D. Rudolph, M. Ratner, and others, and numerous results have been obtained. For detailed accounts of this development → [22, 24, 33, 34]. The main idea employed by Feldman and Satayev was to introduce a new metric called an **$\bar{f}$-metric** and then to define a notion corresponding to V.W.B. of the isomorphism theory described in Section E by using an $\bar{f}$- instead of a $\bar{d}$-metric. They then showed that this property, called **loosely Bernoulli (L.B.)** by Feldman and **monotonely very weak Bernoulli (M.V.W.B.)** by Satayev, stays invariant under monotone equivalence, and that there exist transformations with and without this property. An $\bar{f}$-metric is defined by starting off with an f_N-**metric** on the space $J^{\{1,2,\ldots,N\}}$ instead of the Hamming distance d_N and proceeding in exactly same way as for the definition of $\bar{d}$, namely, by extending f_N to f_N^ξ and $\bar{f}_N$ and finally to $\bar{f}$. For α and β in $J^{\{1,2,\ldots,N\}}$, $f_N(\alpha, \beta)$ is defined by setting $1 - f_N(\alpha, \beta)$ equal to $1/N$ times the maximal integer n for which there are positive integers $j_1 < j_2 < \ldots < j_n$, $k_1 < k_2 < \ldots < k_n$ with $\alpha(j_i) = \beta(k_i)$, $i = 1, \ldots, n$. Ornstein and Weiss discovered that by substituting $\bar{f}$ for $\bar{d}$ one can develop a theory of monotone equivalence that parallels the isomorphism theory described in Section E. One can define a process (φ, ξ) to be **finitely fixed (F.F.)** by substituting $\bar{f}$ for $\bar{d}$ in the definition of F.D. process, and, as was mentioned earlier, to be L.B. by doing the same in the definition of the V.W.B. process. Ornstein and Weiss proved: (a) If φ and $\bar{\varphi}$ have zero entropy, finite positive, or infinite entropy and if both have F.F. generators, then they are monotonely equivalent. (b) Let φ have an F.F. generator. If $h(\varphi) = 0$, then φ is monotonely equivalent to an irrational rotation of the circle. If $0 < h(\varphi) < \infty$, then φ is monotonely equivalent to a Bernoulli shift of finite entropy. If $h(\varphi) = \infty$, then φ is monotonely equivalent to a Bernoulli shift of infinite entropy. (c) If φ has an F.F. generator, then (φ, ξ) is F.F. for all nontrivial partitions ξ. (d) (φ, ξ) is F.F. if and only if it is L.B. It is known further that within each entropy class there exist uncountably many monotonely nonequivalent transformations. A measurable flow $\{\varphi_t\}$ is called L.B. if it can be represented as an S-flow built over an L.B. transformation. The L.B. flows of zero entropy are those monotonely equivalent to a †Kronecker flow on a 2-dimensional torus, and L.B. flows of finite positive (infinite) entropy are those monotonely equivalent to the Bernoulli flow of finite (infinite) entropy. The direct product of an L.B. flow and a Bernoulli flow is L.B., while it was shown by M. Ratner that the horocycle flow (→ Section G) is L.B. but its direct product with itself is not.

G. Classical Dynamical Systems

By a **classical dynamical system** we mean a †diffeomorphism or a flow generated by a smooth †vector field on some †differentiable

manifold M^n. Such a system is nonsingular with respect to a measure defined by any †Riemannian metric on M^n. For a fixed Riemannian metric, we call measures **smooth** if they have a smooth density with respect to the measure given by the metric.

(1) Among classical dynamical systems, **geodesic flows** have been investigated most extensively. Let $\mathcal{T}_1(M)$ be the unitary †tangent bundle over the manifold M^n. A point $(x, e) \in \mathcal{T}_1(M)$ defines a unique †geodesic through x in the direction of e. The geodesic flow on $\mathcal{T}_1(M)$ is the flow defined by $\varphi_t(x, e) = (x_t, e_t)$, where x_t is the point in M^n reached from x after time t under a motion with unit speed along the geodesic determined by (x, e), and e_t is the unit vector at x_t tangent to the geodesic. The classical †Liouville theorem in this context implies that the measure on $\mathcal{T}_1(M)$ that is the product of the measure on M^n induced by the metric and the Lebesgue measure on the $(n-1)$-dimensional sphere gives a smooth invariant measure for the geodesic flow. A wide class of systems arising from mechanics can be described as geodesic flows.

Hopf and G. Hedlund proved that if the manifold M^n is compact and has constant negative curvature, then the geodesic flow is strongly mixing. Later, by using the theory of group †representations, I. Gel'fand and S. Fomin proved that the spectrum of a geodesic flow on a compact manifold of constant negative curvature is Lebesgue, and is even countable Lebesgue in the case where the manifold is of dimension 2. F. Mautner and later L. Auslander, L. Green, and F. Hahn extended this algebraic method to flows obtained under the action of some one-parameter subgroup of a †Lie group acting on its †homogeneous space and obtained extensive results for the case of †nilpotent and some †solvable Lie groups [35].

(2) The flow on an n-dimensional torus defined by

$$\varphi_t(x_1, x_2, \ldots, x_n)$$
$$= (x_1 + \omega_1 t, x_2 + \omega_2 t, \ldots, x_n + \omega_n t)$$

is called a **translational flow** or a **Kronecker flow**. The numbers $\omega_1, \omega_2, \ldots, \omega_n$ are called **frequencies**. Every orbit of $\{\varphi_t\}$ is dense in the torus if and only if the frequencies are linearly independent over **Z**. The motion under a translational flow with independent frequencies is called a **quasiperiodic** motion. A translation flow for a quasiperiodic motion has discrete spectrum.

(3) Sinaĭ obtained a useful criterion for a classical dynamical system to be a K-system. Let M^n be compact, and suppose that $\{\varphi_t\}$ is a flow on M^n defined by a smooth vector field and preserving some smooth measure μ. A one-parameter group $\{\psi_t\}$ of transformations of the space M^n, given by a vector field, is called the flow **transversal** to the flow $\{\varphi_t\}$ if (i) the decomposition of the space M^n into the trajectories of the flow $\{\psi_t\}$ is invariant under $\{\varphi_t\}$; (ii) the limit $\lim_{s \to 0} \lim_{t \to 0} (W_s(t, x) - t)/ts = \alpha(x)$ exists for the function $W_s(t, x)$, which is defined to be the time length of the segment $\{\varphi_s \psi_u(x) \mid 0 \leqslant u \leqslant t\}$ of the trajectory of the flow $\{\psi_t\}$. Sinaĭ's fundamental theorem states that if a flow $\{\varphi_t\}$ is ergodic and has a transversal ergodic flow $\{\psi_t\}$ for which $\int \alpha(x)\, d\mu < 0$, then $\{\varphi_t\}$ is a K-flow. If $\alpha(x) < 0$, then we can even drop the assumption that $\{\varphi_t\}$ is ergodic. If $\int \alpha(x)\, d\mu > 0$, the theorem holds for the flow $\{\varphi_{-t}\}$.

A geodesic flow on a 2-dimensional manifold of constant negative curvature always has a transversal flow, called a **horocycle flow**. The ergodicity of a horocycle flow was proved by Hedlund. It follows from Sinaĭ's fundamental theorem, therefore, that a geodesic flow on a surface of constant negative curvature is a K-flow. Sinaĭ proved even more: A geodesic flow on any surface of negative curvature is a K-flow. There is an extension of the notion of transversal flow to higher dimensions, called **transversal field**. Using this notion Sinaĭ proved that a geodesic flow on a manifold (of any dimension) of constant negative curvature is a K-flow. Finally, Ornstein and Weiss established that geodesic flows on compact manifolds of negative curvature are Bernoulli.

(4) D. Anosov considered a class of flows and diffeomorphisms satisfying a condition that characterizes unstable motions such as geodesic flows on a manifold of negative curvature. They are now called **Anosov flows** (or **Y-flows**) and **Anosov diffeomorphisms** (or **Y-diffeomorphisms**) (→ 126 Dynamical Systems). Anosov proved that if an Anosov flow has a smooth invariant measure, then it is ergodic and either it has a continuous nonconstant eigenfunction or it is a K-flow. Anosov diffeomorphisms with smooth invariant measures are K-automorphisms. Sinaĭ constructed for †transitive Anosov diffeomorphisms (and for Anosov flows) a special partition of the underlying manifold M called a **Markov partition** having desirable properties. The importance of such partitions lies in the fact that they enable one to represent such diffeomorphisms as Markov shifts. Starting with a measure invariant for such a Markov shift having the maximal entropy (→ Section H), Sinaĭ constructed a Gibbs measure (→ Section C), which turned out to be unique in this case and gave rise to a natural invariant measure μ for the corresponding diffeomorphism. He proved further that the diffeomorphism φ considered as an

automorphism of the measure space $(M, \mathscr{B}, \mu)$ is a K-automorphism. It was shown subsequently by R. Azencott that φ is in fact Bernoulli with respect to μ. Sinaĭ investigated further the uniqueness of the invariant measure attaining the maximal entropy for transitive Anosov diffeomorphisms, and he was able to show among other things that the set of transitive C^∞-Anosov diffeomorphisms that do not have an invariant measure absolutely continuous with respect to the measure induced by the Riemannian metric on M contains an open dense subset. The methods employed by Sinaĭ in these investigations were extended further by D. Ruelle, Bowen, and others. Bowen, in particular, was able to construct Markov partitions for a wider class of diffeomorphisms, namely, those satisfying the so-called **axiom A** introduced earlier by S. Smale, and characterized the Gibbs measures for diffeomorphisms in this class by means of the **variational principle** (→ Section H). For a more detailed account of the results of Sinai and Bowen → [13, 14].

(5) An important example of a system that is neither an Anosov nor an "axiom A" system because of nonsmoothness has been studied by Sinaĭ: the simplest mechanical model due to Boltzmann and Gibbs of an ideal gas, which is described as a system generated by tiny rigid spherical pellets moving inside a rectangular box and colliding elastically. Sinaĭ succeeded in proving that this is a K-system, thereby giving an affirmative answer to the classical question of the ergodicity of the basic model of statistical mechanics [38]. It was shown subsequently by M. Aizenman, S. Goldstein, and J. Lebowitz that this system is Bernoulli. Sinaĭ also investigated another nonsmooth system of classical importance: the system describing the motion of a billiard ball on a square table with a finite number of convex obstacles. He showed that such a system, is a K-system. Ornstein and G. Gallavotti then showed that Sinaĭ's methods in fact show that the system is Bernoulli. Sinaĭ's methods were extended further by L. Bunimovich and I. Kubo to study properties of billiard systems in more complicated domains. For more detailed account of these matters → [24]. The results obtained by Sinaĭ and Bowen for Anosov and "axiom A" systems were extended to even wider class of systems by A. Stepin, R. Sacksteder, M. Brin, and Ya. Pesin (→ [24]).

H. Miscellany

(1) An arbitrary aperiodic automorphism can be approximated by periodic automorphisms. More precisely, if φ is an aperiodic automorphism, then for any positive integer n and $\delta > 0$, there exists a periodic automorphism ψ of period n such that $m(\{x \mid \varphi(x) \neq \psi(x)\}) < (1/n) + \delta$ (theorem of Halmos and Rokhlin). The question as to how quickly this approximation can be carried out has been investigated in detail by Katok, Stepin, and others. The rate of this approximation was shown to have a close relationship with the entropy and spectral properties of the automorphism φ. By utilizing this relationship, various examples of automorphisms with specified spectral properties have been constructed [16].

(2) When an arbitrary automorphism φ is given on a Lebesgue measure space $(X, \mathscr{B}, m)$, there exists a unique decomposition $\xi = \{A_\lambda\}$ of X such that (i) each A_λ is invariant under φ and (ii) except for a negligible set (in a specified sense) of A_λ, each A_λ is turned (in a natural manner) into a Lebesgue measure space, and the restriction of φ to A_λ is an ergodic automorphism. This decomposition is called the **ergodic decomposition** of X with respect to φ. There is a corresponding decomposition with respect to a flow. A formula also exists that enables us to compute the entropy $h(\varphi)$ in terms of the entropies of ergodic components of φ.

(3) Let X be a compact metric space and $\varphi: X \to X$ a †homeomorphism. N. Krylov and N. Bogolyubov showed that there always exists on X a Borel probability measure μ that is invariant under φ. Let $\mathscr{P}$ be the collection of all Borel probability measures on X, and let $\mathscr{P}_\varphi$ be the subset of $\mathscr{P}$ consisting of those invariant under φ. Then $\mathscr{P}$ and $\mathscr{P}_\varphi$ are both convex sets compact with respect to the †weak * topology. If $\mathscr{E}_\varphi$ is the set of all extreme points in $\mathscr{P}_\varphi$, then by the †Kreĭn-Milman theorem $\mathscr{E}_\varphi$ is not empty. A measure μ in $\mathscr{P}_\varphi$ belongs to $\mathscr{E}_\varphi$ if and only if φ is ergodic with respect to μ. When the set $\mathscr{E}_\varphi$ consists of a single element, φ is called **uniquely ergodic**; φ is called **minimal** if for every point $x \in X$, the orbit of x under $\varphi = \mathrm{orb}_\varphi(x) = \{\varphi^n(x) \mid n \in \mathbf{Z}\}$ is dense in X. φ is called **strictly ergodic** if it is both minimal and uniquely ergodic. A theorem of J. Oxtoby states that φ is strictly ergodic if and only if for every continuous real-valued function f on X, the sequence of averages $(\sum_{k=0}^{n-1} f(\varphi^k(x)))/n$ converges uniformly to a constant $M(f)$. There are homeomorphisms that are minimal or uniquely ergodic but not strictly ergodic.

(4) R. Jewett proved that any weakly mixing measure-preserving transformation on a Lebesgue space is spatially isomorphic to a strictly ergodic transformation. Krieger extended the result by showing that if the entropy of an ergodic transformation φ is finite, then φ is spatially isomorphic to a strictly ergodic transformation. Similar results were

obtained for flows by K. Jacobs, M. Denker, and E. Eberlein (→ [39]).

(5) A topological analog of the notion of entropy, called **topological entropy**, was introduced by Adler, A. Konheim, and J. McAndrew. This is defined as follows: For every open covering $\mathscr{A}$ of a compact topological space X, let $N(\mathscr{A})$ be the number of sets in the minimal subcovering of $\mathscr{A}$. For open coverings $\mathscr{A}$ and $\mathscr{B}$, let $\mathscr{A} \vee \mathscr{B}$ be the open covering $\{A \cap B \mid A \in \mathscr{A}, B \in \mathscr{B}\}$. For any open covering $\mathscr{A}$ and a continuous mapping φ on X, the limit $\lim_{n\to\infty}(\log N(\mathscr{A} \vee \varphi^{-1}\mathscr{A} \vee \ldots \vee \varphi^{-(n-1)}\mathscr{A}))/n = h_{\mathrm{top}}(\varphi, \mathscr{A})$ exists. Topological entropy $h_{\mathrm{top}}(\varphi)$ of the continuous transformation φ is now defined by $h_{\mathrm{top}}(\varphi) = \sup\{h_{\mathrm{top}}(\varphi, \mathscr{A}) \mid \mathscr{A}$ an open covering of $X\}$.

L. Goodwyn showed that $h_{\mathrm{top}}(\varphi) \geqslant h_\mu(\varphi)$ for any φ-invariant probability measure μ, where $h_\mu(\varphi)$ is the measure-theoretic entropy of φ regarded as a μ-preserving transformation. T. Goodman went further and succeeded in proving that $h_{\mathrm{top}}(\varphi) = \sup\{h_\mu(\varphi) \mid \mu$ a φ-invariant probability measure$\}$. In connection with these results there is interest in the question of the existence and uniqueness of an invariant measure for φ with **maximal entropy**, i.e., a φ-invariant measure μ for which $h_\mu(\varphi) = h_{\mathrm{top}}(\varphi)$. Such a measure does not always exist, and even if it does it may not be unique. The notions of topological entropy and of measures with maximal entropy are generalized by Ruelle in the following way. For an open covering $\mathscr{A}$, a continuous mapping φ, and a real-valued continuous function g on X, let $Z_n(\mathscr{A}, \varphi, g)$ be equal to $\inf\{\sum_{O\in\Gamma} \sup_{x\in O} \exp \sum_{k=0}^{n-1} g(\varphi^k(x))\}$, where the inf is taken over all subcoverings Γ of the covering $\mathscr{A} \vee \varphi^{-1}\mathscr{A} \vee \ldots \vee \varphi^{-(n-1)}\mathscr{A}$. Then a finite limit $P(\mathscr{A}, \varphi, g) = \lim_{n\to\infty} 1/n \log Z_n(\mathscr{A}, \varphi, g)$ exists, and the quantity $P(\varphi, g) = \sup\{P(\mathscr{A}, \varphi, g) \mid \mathscr{A}$ an open covering of $X\}$ is called **topological pressure**. When $g = 0$, $P(\varphi, g)$ reduces to $h_{\mathrm{top}}(\varphi)$. Ruelle proved for †expansive mappings φ that $P(\varphi, g) = \sup\{h_\mu(\varphi) + \int g\, d\mu \mid \mu$ is a φ-invariant probability measure$\}$. This assertion is called the **variational principle for the topological pressure**. The variation principle was proved for general continuous mappings φ by P. Walters. If a φ-invariant measure μ satisfies $P(\varphi, g) = h_\mu(\varphi) + \int g\, d\mu$, then μ is called an **equilibrium state** for g with respect to φ. It is known that for expansive mappings φ every continuous function g on X has an equilibrium state.

(6) Application of ergodic theory to problems in analytic number theory has been made by several authors. Ergodic or mixing properties of particular measure-preserving transformations that arise in connection with various problems in number theory have been exploited to give answers to these problems. New and more striking applications of ideas of ergodic theory to different types of questions in number theory have been started by Yu. Linnik, Fürstenberg, W. Veech, T. Kamae, and others (→ [40, 41]).

7. Most of the results discussed in this article dealt with the action of a cyclic group of transformations or of one-parameter flow. There are significant extensions of many of these results to different types of group actions. For recent developments → [22].

References

[1] E. Hopf, Ergodentheorie, Springer, 1937.
[2] P. R. Halmos, Lectures on ergodic theory, Publ. Math. Soc. Japan, 1956.
[3] P. Billingsley, Ergodic theory and information, Wiley, 1965.
[4] K. Jacobs, Neuere Methoden und Ergebnisse der Ergodentheorie, Springer, 1960.
[5] N. Dunford and J. T. Schwartz, Linear operators I, Interscience, 1958.
[6] K. Yosida, Functional analysis, Springer and Kinokuniya, second edition, 1968.
[7] U. Krengel, Recent progress on ergodic theorems, Soc. Math. France, Astérisque, 50 (1977), 151–192.
[8] J. F. C. Kingman, Subadditive ergodic theory, Ann. Probability, 1 (1973), 883–909.
[9] D. Ruelle, Ergodic theory of differentiable dynamical systems, Publ. Math. Inst. HES, 50 (1979), 27–58.
[10] N. A. Friedman, Introduction to ergodic theory, Van Nostrand, 1969.
[11] H. Fürstenberg, Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions, J. Analyse Math., 31 (1977), 204–256.
[12] S. R. Foguel, Ergodic theory of Markov processes, Van Nostrand, 1969.
[13] Ya. G. Sinaĭ, Gibbs measures in ergodic theory, Russian Math. Surveys, 27 (1972), 21–69. (Original in Russian, 1972.)
[14] R. Bowen, Equilibrium states and the ergodic theory of Anosov diffeomorphisms, Lecture notes in math. 470, Springer, 1975.
[15] G. Maruyama, The harmonic analysis of stationary stochastic processes, Mem. Fac. Sci. Kyushu Univ., (A) 4 (1949), 45–106.
[16] A. B. Katok and A. M. Stepin, Approximations in ergodic theory, Russian Math. Surveys, 22 (5) (1967), 77–102. (Original in Russian, 1967.)
[17] W. Krieger, On entropy and generators of measure-preserving transformations, Trans. Amer. Math. Soc., 149 (1970), 453–464.
[18] V. A. Rokhlin (Rohlin), New progress in

the theory of transformations with invariant measure, Russian Math. Surveys, 15 (4) (1960), 1–22. (Original in Russian, 1960.)
[19] V. A. Rokhlin, Lectures on the entropy theory of measure-preserving transformations, Russian Math. Surveys, 22 (5) (1967), 1–52. (Original in Russian, 1967.)
[20] W. Parry, Entropy and generators in ergodic theory, Benjamin, 1969.
[21] D. S. Ornstein, Ergodic theory, randomness, and dynamical systems, Yale Univ. Press, 1974.
[22] D. S. Ornstein, A survey of some recent results on ergodic theory, Math. Assoc. Amer. Studies in Math., 18 (1978), 229–262.
[23] P. Shields, The theory of Bernoulli shifts, Chicago Lectures in Math., Univ. of Chicago Press, 1973.
[24] A. B. Katok, Ya. G. Sinaĭ, and A. M. Stepin, Theory of dynamical systems and general transformation groups with invariant measure, J. Soviet Math., 7 (1977), 974–1065. (Original in Russian, 1975.)
[25] J. Moser, E. Phillips, and S. Varadhan, Ergodic theory, a seminar, NYU Lecture Notes, 1975.
[26] A. M. Vershik and S. A. Yuzvinskiĭ, Dynamical systems with invariant measure, Progress in Math. VIII, Plenum, 1970, 151–215. (Original in Russian, 1967.)
[27] H. A. Dye, On groups of measure-preserving transformations I, II, Amer. J. Math., 81 (1959), 119–159; 85 (1963), 551–576.
[28] L. Sucheston (ed.), Contributions to ergodic theory and probability, Lecture notes in math. 160, Springer, 1970.
[29] W. Krieger, On ergodic flows and the isomorphism of factors, Math. Ann., 223 (1976), 19–70.
[30] T. Hamachi, Y. Oka, and M. Osikawa, Flows associated with ergodic non-singular transformation groups, Publ. Res. Inst. Math. Sci., 11 (1975), 31–50.
[31] T. Hamachi, The normalizer group of an ergodic automorphism of type III and the commutant of an ergodic flow, J. Functional Anal., 40 (1981), 387–403.
[32] S. Kakutani, Induced measure preserving transformations, Proc. Imp. Acad. Tokyo, 19 (1943), 635–641.
[33] A. B. Katok, Monotone equivalence in ergodic theory, Math. USSR-Izv., 11 (1977), 99–146. (Original in Russian, 1977.)
[34] E. A. Satayev, An invariant of monotone equivalence determining the quotients of automorphisms monotonely equivalent to a Bernoulli shift, Math. USSR-Izv., 11 (1977), 147–169. (Original in Russian, 1977.)
[35] L. Auslander, L. Green, F. Hahn, et al., Flows on homogeneous spaces, Ann. Math. Studies, Princeton Univ. Press, 1963.
[36] V. I. Arnold and A. Avez, Ergodic problems of classical mechanics, Benjamin, 1968.
[37] Ya. G. Sinaĭ, Introduction to ergodic theory, Math. Notes, Princeton Univ. Press, 1976.
[38] Ya. G. Sinaĭ, On the foundations of the ergodic hypothesis for a dynamical system of statistical mechanics, Sov. Math. Dokl., 4 (1963), 1818–1822. (Original in Russian, 1963.)
[39] M. Denker, C. Grillenberger, K. Sigmund, Ergodic theory on compact spaces, Lecture notes in math. 527, Springer, 1976.
[40] Yu. V. Linnik, Ergodic properties of algebraic fields, Springer, 1968.
[41] W. A. Veech, Topological dynamics, Bull. Amer. Math. Soc., 83 (1977), 775–830.

137 (VI.18)
Erlangen Program

When F. Klein succeeded K. G. C. von Staudt as professor at the Philosophical Faculty of Erlangen University in 1872, he gave an inauguration lecture entitled "Comparative Consideration of Recent Geometric Researches," which later appeared as an article [1]. In it he developed a penetrating idea, now called the Erlangen program, in which he utilized group-theoretic concepts to unify various kinds of geometries that until that time had been considered separately.

The concept of transformation is not new; it was, however, not until the 18th century that the concept of transformation groups was recognized as useful. The theory of †invariants of linear groups and the †Galois theory of algebraic equations attracted attention in the 19th century. In the same century, †projective geometry made remarkable progress, for example, when A. Cayley and E. Laguerre discovered that metrical properties of Euclidean and †non-Euclidean geometries can be interpreted in the language of projective geometry. Cayley proclaimed, "All geometry is projective geometry." After learning geometry under J. Plücker, Klein made the acquaintance of S. Lie. Both men understood the importance of the group concept in mathematics. Lie studied the theory of †continuous transformation groups, and Klein studied discontinuous transformation groups from a geometric standpoint. Klein was thus led to the idea of the Erlangen program, which provided a bird's-eye view of geometry.

Klein's idea can be summarized as follows: A space S is a given set with some geometric structure. Let a transformation group G of S be given. A subset of S, called a **figure**, may

have various kinds of properties. The study of the properties that are left invariant under all transformations belonging to G is called the **geometry of the space S subordinate to the group G**. Let this geometry be denoted by (S, G). Two figures of S are said to be **congruent** in (S, G) if one of them is mapped to the other by a transformation of G. The geometry (S, G) is actually the theory of invariants of S under G, with the term **invariants** to be understood in a wider sense; it means both invariant quantities and invariant properties or relations.

Replacing G in (S, G) by a subgroup G' of G, we obtain another geometry (S, G'). A series of subgroups of G gives rise to a series of geometries. For instance, let A be a figure of S. The elements of G leaving A invariant form a subgroup $G(A)$ of G that operates on $S' = S - A$. We thus obtain a geometry $(S', G(A))$ in which A is called an **absolute figure**. In this way, many geometries are obtained from projective geometry. Klein gave numerous examples. It is noteworthy that he mentioned even the groups of †rational and †homeomorphic transformations.

Klein's idea not only synthesized the geometries known at that time, but also became a guiding principle for the development of new geometries.

In 1854 G. F. B. Riemann published his epochmaking idea of **Riemannian geometry**. This geometry has a metric, but in general lacks congruence transformations (isometries). Thus Riemannian geometry is a geometry that is not included in the framework of the Erlangen program. The importance of Riemannian geometry was acknowledged when it was used by A. Einstein in 1916 as a foundation of his general theory of relativity. H. Weyl, O. Veblen, and J. A. Schouten discovered geometries that are generalizations of affine, projective, and †conformal geometries in the same way as Riemannian geometry is a generalization of Euclidean geometry. It became necessary to establish a theory that reconciled the ideas of Klein and Riemann; E. Cartan succeeded in this by introducing the notion of †connection (→ 80 Connections). However, the Erlangen program, which gave an insight into the essential character of classical geometries, still maintains its role as one of the guiding principles of geometry.

References

[1] F. Klein, Vergleichende Betrachtungen über neuere geometrische Forschungen, Math. Ann., 43 (1893), 63–100 (Gesammelte mathematische Abhandlungen, Springer, 1921, vol. 1, 460–497).

[2] G. Fano, Kontinuierliche geometrische Gruppen, Enzykl. Math., Leipzig, 1907–1910, Geometrie III, pt. 1, AB4b, 289–388.

[3] F. Klein, Vorlesungen über nicht-Euklidische Geometrie, Springer, 1928.

[4] F. Klein, Vorlesungen über höhere Geometrie, Springer, 1926.

[5] E. Cartan, Les récentes généralisations de la notion d'espace, Bull. Sci. Math., 48 (1924), 294–320.

[6] E. Cartan, La théorie des groupes et les recherches récentes de géometrie différentielle, Enseignement Math., 24 (1925), 1–18; Proc. Internat. Congr. Math., Toronto, 1 (1928), 85–94.

138 (XV.3) Error Analysis

A. General Remarks

The data obtained by observations or measurements in astronomy, geodesy, and other sciences do not usually give exact values of the quantities in question. The **error** is the difference between the approximation and the exact value. The **theory of errors** originated from systematic work with data accompanied by errors, and the statistical treatment of experimental data was the main concern in the beginning stages (→ 397 Statistical Data Analysis). However, due to the recent development of high-speed computers it has become possible to carry out computations on a tremendously large scale, and the detailed analysis of errors has become an absolute necessity in modern numerical computation. Hence the analysis of errors in relation to numerical computation has become the center of research in error theory.

B. Errors

One rarely makes a mistake in counting a small number of things; therefore the exact value of the count can be determined. On the other hand, the exact value in decimals is never obtainable for a continuous quantity, say length, no matter how fine measurements are made, and a large or small stochastic error is thus inevitable in measuring a continuous quantity. A discrete finite quantity is a **digital quantity**, and a continuous quantity is an **analog** quantity. The natures of these two quantities are quite different. The values of a digital quantity are distributed on some discrete set, while the values of an analog quantity are distributed with a continuous

probability. Thus there is the possibility of error even for treatment of digital quantities, although checking the results for these quantities is easy. It is preferable to regard digital quantities as being analog quantities if the possible values are densely distributed.

On the other hand, when an appropriate †analog computer is not available, analog quantities receive treatment similar to digital quantities. They are expressed as x times some unit, and x is expanded in the decimal or binary systems. An approximation to such an expansion is obtained by rounding off a numeral at some place, the position depending on the capacity for computation by available methods. There are two ways of rounding off numbers, the **fixed point method** and the **floating point method**. The former specifies the place of digits where the rounding off is made, and the latter essentially specifies the number of significant digits.

Classification of errors. (1) **Errors of input data** are the errors included in input data themselves. Such input-data errors include the errors that occur when we represent constants such as $1/3$, $\sqrt{2}$, π by finite decimals. (2) **Truncation errors** occur in approximate expressions for the computation formulas under consideration. (3) **Roundoff errors** occur in taking some finite number of digits from the earlier digits in the numerical value at each step. If the computation of an infinite number of digits were actually possible, no errors of this type would appear. Recently, it has been considered more preferable to call this "computational error."

The difference between fixed point and floating point rounding off is that the former is better suited for operations of addition and subtraction and the latter is better for multiplication and division. In fixed point rounding off, if a number is multiplied many times by numbers less than 1, a so-called **underflow** may occur, and many digits may disappear; a great deal of information can thus be lost. In computation for scientific research that involves frequent multiplication and division, floating point rounding off is preferable. It should be noted that rounding off for addition and subtraction may also cause a critical loss of information. This phenomenon is called **canceling** digits. For instance, in the subtraction $7.6325071 - 7.6318425 = 0.0006646$, where the subtrahend and minuend share several early significant digits, the difference loses those digits. Thus, relative errors may be magnified tremendously. By taking a large number of significant digits, such a situation may be avoided to some extent. So-called **high-precision computation** shows its effectiveness in such cases. Similarly, when a small number b is added to a large number a, the result may be just a and the information of b could be lost completely. This kind of **loss of information** can often cause serious trouble.

C. Methods of Error Analysis

In order to analyze the **propagation of errors**, let us assume that all numbers are carried to infinitely many digits so that no roundoff error occurs. Suppose that we are to evaluate the function $y=f(x_1,\ldots,x_n)$ when $x_1,x_2,\ldots,x_n$ are assigned. Let η be the truncation error of an approximate expression. If an input error δ_i for x_i exists, then the corresponding error for y is

$$\eta+\sum_{i=1}^{n}\frac{\partial f}{\partial x_i}\delta_i.$$

Moreover, suppose that at the final step we round off to get a result with a finite number of digits, by which an error ε is introduced. Then the final error δ for y is

$$\delta=\varepsilon+\eta+\sum_{i=1}^{n}\frac{\partial f}{\partial x_i}\delta_i.$$

This procedure is performed for each step needed in the computation. If $y=f(x_1,\ldots,x_n)$ is a specified step, then the input error δ_i for that step involves all the errors arising before that step, i.e., δ_i is an **accumulated error. Forward analysis** is a method to estimate the total accumulated error from the initial input data. It is usually quite difficult to obtain precise estimates by means of this method. In contrast, J. H. Wilkinson proposed the following **backward analysis**. Here, the computational value y is considered as the exact result for the modified initial data $\tilde{x}_1,\ldots,\tilde{x}_n$, say $y=f(\tilde{x}_1,\ldots,\tilde{x}_n)$, and the estimates for $|x_i-\tilde{x}_i|$ are given. For example, in the binary floating point arithmetic of u bits, we always have the relation:

computational value of $a\pm b$

$=a(1+\delta)\pm b(1+\varepsilon)$,

with $|\delta|, |\varepsilon|\leqslant 2^{-u}$, even when cancellation or loss of information occurs. Wilkinson has made a deep investigation of error analysis for linear computation, algebraic equations, and eigenvalue problems by means of backward analysis [4, 5].

D. Proliferation of Errors

The phenomena usually called "accumulation of errors" should more appropriately be called the **proliferation** of errors, where the algorithm itself includes a particular mechanism to increase a small error indefinitely. An example is the recursion formula for †Bessel functions

stated as

$$J_{n+1}(x)=(2n/x)J_n(x)-J_{n-1}(x).$$

It is customary to compute $J_n(x)$ by this formula, starting with the values of $J_0(x)$ and $J_1(x)$, with x given. By putting $J_{n-1}(x)=y_n$, $J_n(x)=z_n$, the recursion formula can be regarded as a linear transformation of the point $P_n(y_n, z_n)$ in a plane into another point $P_{n+1}(y_{n+1}, z_{n+1})$, where

$$y_{n+1}=z_n, \qquad z_{n+1}=-y_n+(2n/x)z_n.$$

The †eigenvalues λ_1, λ_2 of this †difference equation satisfy the following: As long as $n<|x|$, we have $|\lambda_1|=|\lambda_2|=1$, while if $n>|x|$, λ_1 is greater than 1 and increases rapidly as n tends to infinity. Consequently, even the slightest discrepancy in the position of P_n gives rise to a greatly magnified error in the result [3].

Many studies have been made of the propagation of errors and of instability phenomena in the numerical solution of ordinary differential equations (→ 303 Numerical Solution of Ordinary Differential Equations; [2]).

References

[1] J. von Neumann and H. H. Goldstine, Numerical inverting of matrices of higher order, Bull. Amer. Math. Soc., 53 (1947), 1021–1099.
[2] P. Henrici, Discrete variable methods in ordinary differential equations, Wiley, 1962.
[3] T. Uno, The problem of error propagation (in Japanese), Sûgaku, 15 (1963), 30–40.
[4] J. H. Wilkinson, Rounding errors in algebraic processes, Prentice-Hall, 1963.
[5] J. H. Wilkinson, Algebraic eigenvalue problem, Clarendon Press, 1965.
[6] L. B. Rall (ed.), Error in digital computation I, II, Wiley, 1965, 1966.

139 (VI.3) Euclidean Geometry

A. History

Attempts to construct axiomatically the geometry of ordinary 3-dimensional space were undertaken by the ancient Greeks, culminating in Euclid's *Elements* (→ 187 Greek Mathematics). The **fifth postulate** of Euclid's *Elements* requires that two straight lines in a plane that meet a third line, as shown in Fig. 1, in angles α, β whose sum is less than 180°, have a common point. In the *Elements*, two straight lines in a plane without a common point are said to be **parallel**. It can be proved from other axioms in the *Elements* that if $\alpha+\beta=180°$, the two lines l and l' in Fig. 2 are parallel. Hence given a line l and a point P not lying on l,

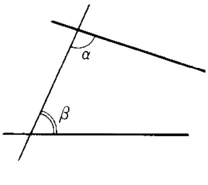


Fig. 1

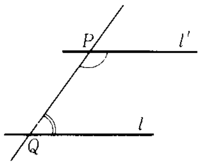


Fig. 2

there exists a line l' passing through P that is parallel to l. The fifth postulate ensures the uniqueness of the parallel l' passing through the given point P. For this reason, the fifth postulate is also called the **axiom of parallels**. Utilizing this axiom, we can prove the well-known theorems on parallel lines, the sum of interior angles of triangles, etc. The axiom plays an important role in the proof of the Pythagorean theorem in the *Elements*. The axiom is also called **Euclid's axiom**.

However, Euclid states this axiom in a quite complicated form, and unlike his other axioms, it cannot be verified within a bounded region of the space.

Many mathematicians tried in vain to deduce it from other axioms. Finally the axiom was shown to be independent of other axioms in the *Elements* by the invention of non-Euclidean geometry in the 19th century (→ 285 Non-Euclidean Geometry).

The term *Euclidean geometry* is used in contrast to *non-Euclidean geometry* to refer to the geometry based on Euclid's axiom of parallels as well as on other axioms explicit or implicit in Euclid's *Elements*. It was in the 19th century that a complete system of Euclidean geometry was explicitly formulated (→ 155 Foundations of Geometry). From the standpoint of present-day mathematics, it would be natural to define first the group of motions by the axiom of free mobility due to H. Helmholtz (→ Section B) and then, following F. Klein, to define Euclidean geometry as the study of properties of spaces that are invariant under the groups of these motions (→ 137 Erlangen Program).

B. Group of Motions

Let P be an †ordered field and A^n the n-dimensional †affine space over P. Let B^r be an r-dimensional affine subspace of A^n, B^{r-1} an $(r-1)$-dimensional subspace of B^r, B^{r-2} an $(r-2)$-dimensional subspace of B^{r-1}, etc. In the sequence of subspaces $B^r, B^{r-1}, \ldots, B^0$, each B^k-B^{k-1} consists of two †half-spaces ($k=r, r-1, \ldots, 1$). Let H^k be one of these half-

spaces. Then the sequence of half-spaces H^r, $H^{r-1}, \ldots, H^1$ is called an r-dimensional **flag**, denoted by $\mathfrak{H}^r$ $(n \geqslant r \geqslant 1)$, and B^r and H^r are called the **principal space** and the **principal half-space** of $\mathfrak{H}^r$, respectively. If f is a †proper affine transformation of A^n, $f(H^r)$, $f(H^{r-1})$, $\ldots, f(H^1)$ form an r-dimensional flag $\mathfrak{K}^r$. We write $f(\mathfrak{H}^r)=\mathfrak{K}^r$.

Let $\mathfrak{A}^n$ be the group of all proper affine transformations of A^n. The subgroup $\mathfrak{B}^n$ of $\mathfrak{A}^n$ with the following two properties is called the **group of motions**, and any element of $\mathfrak{B}^n$ is called a **motion** (or **congruent transformation**). (1) Let r be an integer between 1 and n, and let $\mathfrak{H}^r$, $\mathfrak{K}^r$ be any two r-dimensional flags. Then there exists an element f of $\mathfrak{B}^n$ that carries $\mathfrak{H}^r$ to $\mathfrak{K}^r$: $f(\mathfrak{H}^r)=\mathfrak{K}^r$. (2) Let f, g be two elements of $\mathfrak{B}^n$ with $f(\mathfrak{H}^r)=\mathfrak{K}^r$, $g(\mathfrak{H}^r)=\mathfrak{K}^r$, and let p be any point on the principal space of $\mathfrak{H}^r$. Then $f(p)=g(p)$, that is, f, g have the same "effect" on the principal space. In particular, when $r=n$, then $f=g$. That $\mathfrak{A}^n$ possesses a subgroup $\mathfrak{B}^n$ with properties (1) and (2) is called the **axiom of free mobility**.

When $n=1$, it is easy to see that the elements of $\mathfrak{B}^1$ are only those elements f of $\mathfrak{A}^1$ that can be expressed in the form $f(x)=\pm x+a$ $(a \in P)$. When $n \geqslant 2$, P must satisfy the following condition in order that a subgroup $\mathfrak{B}^n$ with properties (1) and (2) exists in $\mathfrak{A}^n$: If a, $b \in P$, then P contains an element x such that $x^2=a^2+b^2$. When this condition is satisfied, the ordered field P is called a **Pythagorean field**. Every †real closed field (e.g., the field **R** of real numbers) is Pythagorean. If $\mathfrak{B}^n$ exists, its uniqueness is assured by (1) and (2). Furthermore, if P contains a square root of every positive element (this condition is satisfied, for example, by **R**), then conditions (1) and (2) are reducible to the case $r=n$ only, i.e., conditions (1) and (2) for other values of r follow from (1) and (2) with $r=n$. Hereafter, we assume the existence of $\mathfrak{B}^n$.

Suppose that we have $A^n \supset B^r \supset B^k$ $(n \geqslant r \geqslant k \geqslant 0)$, and let $\mathfrak{H}^r$ be a flag with the principal space B^r: $\mathfrak{H}^r=(H^r, \ldots, H^k, \ldots, H^1)$. Let $\mathfrak{K}^r$ be another flag with the same principal space B^r: $\mathfrak{K}^r=(K^r, \ldots, K^k, \ldots, K^1)$, where we suppose that $H^j=K^j$ for $k \geqslant j \geqslant 1$, whereas for $r \geqslant i \geqslant k+1$, we suppose that H^i and K^i are different half-spaces on B^i divided by B^{i-1}. The flag $\mathfrak{K}^r$ is denoted by $\mathfrak{H}^r_k$. An element f of $\mathfrak{B}^n$ with $f(\mathfrak{H}^r)=\mathfrak{H}^r_k$ is called a **symmetry** (or **reflection**) of B^r with respect to B^k. It leaves every point on B^k invariant, and its effect on B^r is determined only by B^k independently of the choice of half-spaces in $\mathfrak{H}^r$ and $\mathfrak{K}^r$ (subject to the conditions mentioned above). In particular, the symmetry of A^n with respect to a point $A^0=p$ is called a **central symmetry** with respect to the **center** p; and the symmetry of A^n with respect to a hyperplane $A^{n-1}=h$ is called a **hyperplanar symmetry**. They are uniquely determined by p and h, respectively, and are denoted by S_p and $S(h)$, respectively. If $H(A^n)$ is the set of all hyperplanes of A^n, then $\mathfrak{B}^n$ is generated by $\{S(h) \mid h \in H(A^n)\}$. Furthermore, if p, q are two points of A^n, the composite S_pS_q is a parallel translation by $2 \cdot \overrightarrow{pq}$ (Fig. 3). The parallel translations generate a normal subgroup $\mathfrak{T}^n$ of $\mathfrak{B}^n$. For $p, q \in A^n$, the element of $\mathfrak{T}^n$ that carries p to q is denoted by τ_{pq}. The

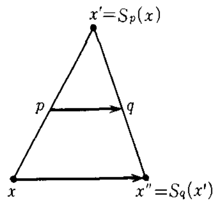


Fig. 3

subgroup of $\mathfrak{B}^n$ that leaves a point p of A^n invariant is denoted by O^n_p. Obviously, we have $O^n_q=\tau_{pq}\, O^n_p \tau^{-1}_{pq}$. Thus all the O^n_p (for $p \in A^n$) are isomorphic. We call O^n_p the **orthogonal group** around p and any element of O^n_p an **orthogonal transformation** around p. More generally, any element of $\mathfrak{B}^n$ that leaves a subspace A^k of A^n invariant is called an **orthogonal transformation around the subspace** A^k.

An element of $\mathfrak{B}^n$ that preserves the **orientation** of A^n, i.e., is represented by a †proper affinity with a positive determinant, is called a **proper motion**. Proper motions form a subgroup $\mathfrak{B}^n_0$ of $\mathfrak{B}^n$. Rotations are, by definition, orthogonal transformations belonging to $\mathfrak{B}^n_0$. Sometimes $\mathfrak{B}^n_0$ is called the group of motions; then $\mathfrak{B}^n$ is called the group of motions in the wider sense. In this article, however, we shall continue to use the terminology introduced above.

The study of the properties of A^n invariant under $\mathfrak{B}^n$ is n**-dimensional Euclidean geometry**. Since $\mathfrak{A}^n \supset \mathfrak{B}^n$, every proposition in affine geometry (→ 7 Affine Geometry) can be considered a proposition in Euclidean geometry, but there are many propositions that are proper to Euclidean geometry. Sometimes the subgroup of $\mathfrak{A}^n$ generated by $\mathfrak{B}^n$ and the **homotheties** of A^n, i.e., elements of $\mathfrak{A}^n$ represented by †scalar matrices, is called the **group of motions in the wider sense**, and the study of properties of A^n invariant under this group is called n-dimensional **Euclidean geometry in the wider sense**.

C. Length of Segments

Two figures F, F' in A^n are said to be **congruent** if there exists an $f \in \mathfrak{B}^n$ such that $f(F)=F'$.

Then we write $F \equiv F'$. The congruence relation is an †equivalence relation. Let $s = \overline{pq}$, $s' = \overline{p'q'}$ be two †segments in A^n. We say that s, s' have equal **length** when $s \equiv s'$. Length is an attribute of the equivalence class of segments. The length of s is denoted by $|s|$. All segments of the form $\overline{pp}$ are congruent, and we define $|\overline{pp}| = 0$. If we are given a length and a †half-line starting from a point p, we can find a unique point q on it such that $|\overline{pq}|$ = the given length (Fig. 4). Let r be a point on the extension of $\overline{pq}$. The length $|\overline{pr}|$ is then uniquely determined by $|pq|$ and $|qr|$. It is defined as the

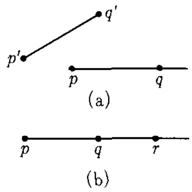


Fig. 4

sum of the lengths: $|\overline{pr}| = |\overline{pq}| + |\overline{qr}|$. With respect to the addition thus defined, lengths of segments in A^n form a †commutative semigroup with the cancellation law, which can be extended to an †Abelian group M with 0 as the identity element (→ 190 Groups P).

Let $|s| \neq 0$ and $|s'|$ be any length. On a half-line starting from p, we can find points q, r with $|\overline{pq}| = |s|$, $|\overline{pr}| = |s'|$. Then the element $pr/pq = \lambda \in P$ (→ 7 Affine Geometry) is a positive element of P uniquely determined by $|s|$ and $|s'|$. We call λ the **measure** of $|s'|$ with the **unit** $|s|$ and denote it by $|s'|:|s|$. If P is †Archimedean, λ can be represented by a real number (→ 149 Fields N). We have $(|s'| + |s''|):|s| = (|s'|:|s|) + (|s''|:|s|)$, $(|s''|:|s'|)(|s'|:|s''|) = |s''|:|s|$ (if $|s'| \neq 0$). Thus the mapping $|s'| \rightarrow |s'|:|s|$ sends the additive semigroup of lengths to that of the positive elements of P. This is actually an isomorphism, which can be extended to an isomorphism of M onto the additive group of the field P.

Let $\varphi(X, X', \ldots, X^\alpha)$ be a †homogeneous rational function of a finite number of variables $X, X', \ldots, X^\alpha$. If a relation $\varphi(\lambda, \lambda', \ldots, \lambda^\alpha) = 0$ holds for $\lambda = |s|:|s_0|$, $\lambda' = |s'|:|s_0|, \ldots, \lambda^\alpha = |s^\alpha|:|s_0|$, where $|s_0|$ is a length $\neq 0$, then $\varphi(\lambda_1, \lambda'_1, \ldots, \lambda^\alpha_1) = 0$ holds also for $\lambda_1 = |s|:|s_1|$, $\lambda'_1 = |s'|:|s_1|, \ldots, \lambda^\alpha_1 = |s^\alpha|:|s_1|$, where $|s_1|$ is any other length $\neq 0$. Hence, in this case, the expression $\varphi(|s|, |s'|, \ldots, |s^\alpha|) = 0$ is meaningful.

D. Angles and Their Measure

An **angle** $\angle AOB$ is a figure constituted by two half-lines OA, OB starting from the same point O but belonging to different straight lines. The point O is called the **vertex**, and the two half-lines OA, OB are called the **sides** of $\angle AOB$ (Fig. 5). Two congruent angles are said to have the same **measure**, denoted by $|\angle AOB|$ or sometimes simply by α. Let $\mathfrak{H}^2 = (H^2, H^1)$ (H^1 = half-line QR) be a given 2-dimensional flag and the given measure of an angle. Then we can find a unique half-line QP in the given half-plane H^2 such that $|\angle PQR| = \alpha$ (Fig. 6). The angle $\angle PQR$ is said to "belong" to $\mathfrak{H}^2$. Let K^2 be the half-plane separated by the line $P \cup Q$ containing the half-line QR. Then $H^2 \cap K^2$ is called the **interior** of the angle

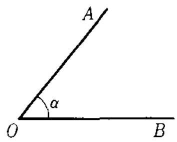


Fig. 5

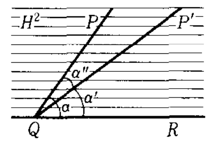


Fig. 6
$\alpha' + \alpha'' = \alpha$.

$\angle PQR$. Let $\angle P'QR$ be another angle belonging to $\mathfrak{H}^2$. If the interior of the latter angle is a subset of the interior of $\angle PQR$ and $QR \neq QP'$, then $|\angle PQR|$ is said to be greater than $|\angle P'QR|$, and we write $|\angle PQR| > |\angle P'QR|$. Actually, it can be shown that $>$ is a relation between the measures of $\angle PQR$ and $\angle P'QR$, and that the set of measures of angles forms a †linearly ordered set with respect to the relation $\geqslant$ defined in the obvious way. When $|\angle PQR| > |\angle P'QR|$, we write $|\angle PQR| = |\angle PQP'| + |\angle P'QR|$. Actually, these are relations between measures of angles. Furthermore, if the measure α of an angle is given, the set of measures of angles $\leqslant \alpha$ forms a linearly ordered set order-isomorphic to a segment and satisfying: (i) If $\beta < \gamma$ then there exists a δ such that $\beta + \delta = \gamma$; (ii) $\beta + \delta = \delta + \beta$; $(\beta_1 + \beta_2) + \beta_3 = \beta_1 + (\beta_2 + \beta_3)$ if all these sums exist; and (iii) $\beta_1 + \delta = \beta_2 + \delta$ implies $\beta_1 = \beta_2$. When P is Archimedean, these properties imply that the measure of angles $\leqslant |\angle PQR|$ can be represented by positive real numbers $\leqslant k$ (k is any given positive number) such that the relations of ordering an addition are preserved.

This one-to-one correspondence between the measures of angles and a subset S of the interval $(0, k]$ of real numbers can be extended to a correspondence between the measures of **general angles** and the subset of $\mathbf{R}$ obtained from S. When $P = \mathbf{R}$, then we have $S = (0, k]$, and any real number appears as a measure of a general angle. We can choose $\angle PQR$ and the positive number k arbitrarily, but it is customary to choose them as follows. Suppose we are given an angle $\angle AOB$. Let the extensions of the half-lines OA and OB in the opposite directions be OA' and OB', respectively. The angles $\angle AOB$ and $\angle A'OB$ are called **supplementary angles** of each other, and so are

$\angle AOB$ and $\angle AOB'$. The angles $\angle AOB$ and $\angle A'OB'$ are called **vertical angles** to each other (Fig. 7). Any angle is congruent to its vertical angle, and an angle that is congruent to its supplementary angle has a fixed measure.

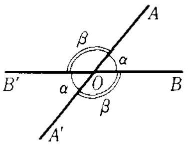


Fig. 7

Such an angle (or its measure) is called a **right angle**. In the description of the measurement of angles, we usually consider the case where the special angle $\angle PQR$ is a right angle, and we set $k=\pi/2$. (The existence and uniqueness of the right angle can be proved.) An angle that is greater (smaller) than a right angle is called an **obtuse (acute) angle**. A general angle whose measure is twice (four times) a right angle is called a **straight angle (perigon)**. Sometimes we choose as the "unit angle" 1/90 of a right angle, which is called a **degree** (hence a right angle = 90 degrees, denoted by 90°); 1/60 of a degree is called a **minute** ($1°=60$ minutes, denoted by $60'$), and 1/60 of a minute is called a **second** ($1'=60$ seconds, denoted by $60''$). If, as usual, we put the right angle equal to $\pi/2$, then the unit angle is $(2/\pi)$(right angle). This is called a **radian**, and 1 radian $=180°/\pi= 57°17'44.806\ldots'' \fallingdotseq 57.3°$.

If a straight line m intersects two straight lines l, l', eight angles $\alpha, \beta, \gamma, \delta, \alpha', \beta', \gamma', \delta'$ appear, as in (Fig. 8). In this figure, α and α', β and β', γ and γ', and δ and δ' are called **corresponding angles**, while α and γ', β and δ', γ and α', and δ and β' are called **alternate angles** to each other. When l and l' are parallel, each of these angles is congruent to its corresponding or alternate angle.

The **Pythagorean theorem** asserts that if a triangle $\triangle ABC$ is given for which $\angle ABC$ is a right angle (Fig. 9), then $|AB|^2+|BC|^2=|CA|^2$ (which makes sense since $X^2+Y^2-Z^2$ is a homogeneous polynomial).

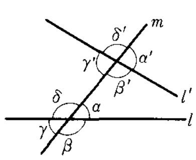


Fig. 8

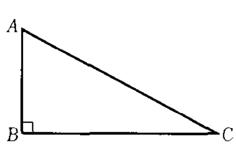


Fig. 9

E. Rectangular Coordinates

When two straight lines l, m intersect, two pairs of vertical angles appear. If one of these angles is a right angle, then all are. Then we say that l and m are **orthogonal** (or **perpendicular**) to each other, and write $l \perp m$. Let l be a line and A^r an r-dimensional subspace of A^n $(1 \leqslant r \leqslant n-1)$ intersecting l at a point $O=A^r \cap l$. If l is orthogonal to all lines on A^r passing through O, then l is said to be orthogonal to A^r, and we write $l \perp A^r$ (Fig. 10). If A^{n-1} is any hyperplane in A^n, then there exists a unique line l through a given point P of A^n that is orthogonal to A^{n-1}; this l is called the **perpendicular** to A^{n-1} through P, and the

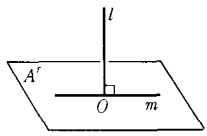


Fig. 10

intersection $l \cap A^{n-1}$ is called the **foot of the perpendicular** through P. When A^{n-1} is given, the mapping from A^n to A^{n-1} assigning to every point P of A^n the foot of the perpendicular through P is called the **orthogonal projection** from A^n to A^{n-1}.

Let $\mathfrak{H}^n=(H^n, H^{n-1}, \ldots, H^1)$ be an n-dimensional flag of A^n and O the initial point of the half-line H^1. Then we can find a point E_i in H^i $(i=1,2,\ldots,n)$ such that $O \cup E_i \perp O \cup E_j$ $(i \neq j, i,j=1,2,\ldots,n)$. Moreover, if $|e|$ is any unit of length, then E_i can be chosen uniquely so that $|OE_i|=|e|$ $(i=1,2,\ldots,n)$. Then O, $E_1, \ldots, E_n$ are †independent points in A^n, and we have $A^n=O \cup E_1 \cup \ldots \cup E_n$. Thus we have a †frame $\Sigma=(O; E_1, \ldots, E_n)$ of A^n with O as origin and the E_i as unit points. Such a frame is called an **orthogonal frame**. A coordinate system with this frame, called an **orthogonal coordinate system adapted to** $\mathfrak{H}^n$, is uniquely determined by $\mathfrak{H}^n$. A motion is characterized as an †affinity sending one orthogonal frame onto another or onto itself.

Utilizing an orthogonal coordinate system, the lengths of segments and the measures of angles can be expressed simply. Let $(x_1, \ldots, x_n)$ be the coordinates of X with respect to such a coordinate system. Then the length of the segment $|OX|$ (with $|e|$ as unit) is equal to $(\sum_{i=1}^n x_i^2)^{1/2}$, and when Y is another point, with coordinates $(y_1, \ldots, y_n)$, $O \neq X$, $O \neq Y$, then we have

$$\cos|\angle XOY|=\frac{\sum_{i=1}^n x_i y_i}{(\sum_{i=1}^n x_i^2)^{1/2}(\sum_{i=1}^n y_i)^{1/2}}.$$

In particular, we have $O \cup X \perp O \cup Y$ if and only if $\sum_{i=1}^n x_i y_i=0$.

We may write $\mathbf{x}=\overrightarrow{OX}$ for the †location vector of X. Then the †affinity $A\mathbf{x}+\mathbf{b}$ is a motion if and only if A is an †orthogonal matrix. Thus the †inner product $(\mathbf{x}, \mathbf{y})$ is invariant

under motions; it therefore has meaning in Euclidean geometry. If we put $|\mathbf{x}|=(\mathbf{x},\mathbf{x})^{1/2}$, then the right-hand sides of the formulas for $|OX|$ and $\cos|\angle XOY|$ can be written as $|\mathbf{x}|$ and $(\mathbf{x},\mathbf{y})/(|\mathbf{x}|\cdot|\mathbf{y}|)$. More generally, we have $|XY|=|\mathbf{y}-\mathbf{x}|$. This is the **Euclidean distance** (or simply **distance**) between X and Y. Then A^n becomes a †metric space with this distance, i.e., a Euclidean space. Historically, the notion of metric spaces was introduced in generalizing Euclidean spaces (→ 273 Metric Spaces).

F. Area and Volume

The subset I^n of A^n consisting of points $(x_1,\dots,x_n)$ with respect to an orthogonal coordinate system, with $0\leqslant x_i\leqslant 1$, $i=1,\dots,n$, is called an n-dimensional **unit cube**. A function m that assigns to †polyhedra in the wider sense $P, Q, \dots$ in A^n nonnegative real numbers $m(P)$, $m(Q), \dots$ is called an n-dimensional **volume** if it satisfies the following four conditions: (1) $m(\varnothing)=0$. (2) $m(P\cup Q)+m(P\cap Q)=m(P)+m(Q)$. (3) If P is sent to Q by a translation, then $m(P)=m(Q)$. (4) $m(I^n)=1$. It has been proved that such a function is unique and has the property that $P\equiv Q$ implies $m(P)=m(Q)$. Thus the concept of volume can be defined in the framework of Euclidean geometry. More generally, if the affinity $f(\mathbf{x})=A\mathbf{x}+\mathbf{b}$ sends P onto Q, then $m(Q)=cm(P)$, where c is the absolute value of the determinant $|A|$. If P is covered by a finite number of hyperplanes, then $m(P)=0$, and if P is a †parallelotope with n independent edges $\mathbf{a}_1,\dots,\mathbf{a}_n$, then $m(P)=\operatorname{abs}|\mathbf{a}_1,\dots,\mathbf{a}_n|$, where $|\mathbf{a}_1,\dots,\mathbf{a}_n|$ is the determinant of the $n\times n$ matrix with $\mathbf{a}_i$ as the ith column vector, and $\operatorname{abs}x$ is the absolute value of the real number x. If P is an †n-simplex whose vertices have location vectors $\mathbf{x}_0,\mathbf{x}_1,\dots,\mathbf{x}_n$, then we have

$$m(P)=\frac{1}{n!}\operatorname{abs}\begin{vmatrix}1 & 1 & \dots & 1\\ \mathbf{x}_0 & \mathbf{x}_1 & \dots & \mathbf{x}_n\end{vmatrix}.$$

The volume of any polyhedron can be obtained by dividing it into n-simplexes and summing their volumes. If P is an r-dimensional polyhedron in A^n, then the r-dimensional volume of P is obtained by dividing P into r-simplexes and summing their r-dimensional volumes (in the respective r-dimensional Euclidean spaces containing them). In particular, when $r=1$, we speak of **length** (e.g., the length of a broken line), and when $r=2$, of **area**. If V is the r-dimensional volume of an r-dimensional parallelotope with r edges $\mathbf{a}_1,\dots,\mathbf{a}_r$, we have the formula

$$V^2=\begin{vmatrix}(\mathbf{a}_1,\mathbf{a}_1) & \dots & (\mathbf{a}_1,\mathbf{a}_r)\\ & \dots & \\ (\mathbf{a}_r,\mathbf{a}_1) & \dots & (\mathbf{a}_r,\mathbf{a}_r)\end{vmatrix}.$$

The notion of measure of point sets other than polyhedra is a generalization of the notion of volume of polyhedra (→ 270 Measure Theory).

G. Orthonormalization

Let $O, A_1,\dots,A_n$ be $n+1$ †independent points in A^n. Then the n vectors $\overrightarrow{OA_i}=\mathbf{a}_i$, $i=1,\dots,n$, are independent. The points $O, A_1,\dots,A_n$ determine an n-dimensional flag $\mathfrak{H}^n$ of A^n as follows. Let H_1 be the half-line OA_1, H_2 be the half-plane on the plane $O\cup A_1\cup A_2$ separated by the line $O\cup A_1$ in which A_2 lies, ..., H_n be the half-space on A^n separated by the hyperplane $O\cup A_1\cup\dots\cup A_{n-1}$ in which A_n lies. Let $\mathbf{b}_1,\dots,\mathbf{b}_n$ be the unit vectors of the rectangular coordinate system adapted to $\mathfrak{H}^n$. Suppose further that we are given a rectangular coordinate system. Then $\mathbf{b}_1,\dots,\mathbf{b}_n$ can be obtained from $\mathbf{a}_1,\dots,\mathbf{a}_n$ by the following procedure, called **orthonormalization** (E. Schmidt): First put $\mathbf{b}_1=\mathbf{a}_1/|\mathbf{a}_1|$, so that $|\mathbf{b}_1|=1$. Then $\mathbf{c}_2=\mathbf{a}_2-(\mathbf{a}_2,\mathbf{b}_1)\mathbf{b}_1$ satisfies $(\mathbf{b}_1,\mathbf{c}_2)=0$, $\mathbf{c}_2\neq 0$. Put $\mathbf{b}_2=\mathbf{c}_2/|\mathbf{c}_2|$. Then we have $(\mathbf{b}_1,\mathbf{b}_2)=0$, $|\mathbf{b}_2|=1$. When $\mathbf{b}_1,\dots,\mathbf{b}_{i-1}$ are obtained in this way, so that $(\mathbf{b}_j,\mathbf{b}_k)=\delta_{jk}$ for $1\leqslant j,k\leqslant i-1$, then $\mathbf{c}_i=\mathbf{a}_i-(\mathbf{a}_i,\mathbf{b}_1)\mathbf{b}_1-\dots-(\mathbf{a}_i,\mathbf{b}_{i-1})\mathbf{b}_{i-1}$ satisfies $(\mathbf{b}_j,\mathbf{c}_i)=0$, $\mathbf{c}_i\neq 0$. Hence $\mathbf{b}_i=\mathbf{c}_i/|\mathbf{c}_i|$ added to $\mathbf{b}_1,\dots,\mathbf{b}_{i-1}$ retains the property $(\mathbf{b}_j,\mathbf{b}_k)=\delta_{jk}$ for $1\leqslant j,k\leqslant i$, and this procedure can be continued to $i=n$.

Two vectors $\mathbf{u}$, $\mathbf{v}$ are called **orthogonal** (denoted $\mathbf{u}\perp\mathbf{v}$) if $(\mathbf{u},\mathbf{v})=0$, and $\mathbf{u}$ is called **normalized** when $|\mathbf{u}|=1$. Thus any two of the vectors $\mathbf{b}_1,\dots,\mathbf{b}_n$ are orthogonal, and each of them is normalized. Between given vectors $\mathbf{a}_1,\dots,\mathbf{a}_n$ and $\mathbf{b}_1,\dots,\mathbf{b}_n$ we have the relation $\{\mathbf{a}_1,\dots,\mathbf{a}_i\}$ (= the linear space generated by $\mathbf{a}_1,\dots,\mathbf{a}_i$) $=\{\mathbf{b}_1,\dots,\mathbf{b}_i\}$, $i=1,\dots,n$.

Let $\mathfrak{M}_1$, $\mathfrak{M}_2$ be two subspaces of the linear space $\mathfrak{M}$ of the vectors of the Euclidean space A^n. If any element of $\mathfrak{M}_1$ is orthogonal to any element of $\mathfrak{M}_2$, then $\mathfrak{M}_1$ and $\mathfrak{M}_2$ are called **orthogonal** and written $\mathfrak{M}_1\perp\mathfrak{M}_2$. For any proper subspace $\mathfrak{M}_1$ of $\mathfrak{M}$, it can be shown by the method of orthonormalization that there exists a unique proper subspace $\mathfrak{M}_2$ of $\mathfrak{M}$ such that $\mathfrak{M}=\mathfrak{M}_1\cup\mathfrak{M}_2$, $\mathfrak{M}_1\perp\mathfrak{M}_2$. Such a subspace $\mathfrak{M}_2$ is called the **orthocomplement** of $\mathfrak{M}_1$ (with respect to $\mathfrak{M}$). Then $\mathfrak{M}_1\cap\mathfrak{M}_2=\{0\}$ follows, and hence $\mathfrak{M}=\mathfrak{M}_1+\mathfrak{M}_2$. Every element $\mathbf{a}$ of $\mathfrak{M}$ is therefore written uniquely in the form $\mathbf{a}_1+\mathbf{a}_2$, $\mathbf{a}_1\in\mathfrak{M}_1$, $\mathbf{a}_2\in\mathfrak{M}_2$; we call $\mathbf{a}_1$ the $\mathfrak{M}_1$-component of $\mathbf{a}$ and $\mathbf{a}_2$ the **orthogonal component** of $\mathbf{a}$ with respect to $\mathfrak{M}_1$. The mapping from $\mathfrak{M}$ to $\mathfrak{M}_1$ assigning $\mathbf{a}_1$ to $\mathbf{a}$ is called the **orthogonal projection** from $\mathfrak{M}$ to $\mathfrak{M}_1$; it is a linear and †idempotent mapping.

H. Distance between Subspaces

Since the Euclidean space A^n is a metric space, the distance is defined between any two nonempty subsets of A^n (→ 273 Metric Spaces). Let A^r, B^s be two subspaces of dimensions r, s of A^n, and let d be the distance between them. Then it can be shown that there exist points $p \in A^r$, $q \in B^s$ such that $d = \overline{pq}$, and if $p' \in A^r$, $q' \in B^s$ are any other points with $d = \overline{p'q'}$, then $\overrightarrow{pq} = \overrightarrow{p'q'}$. In particular, when $r=0$ and $s=n-1$ (i.e., when $A^r = p$ is a point and $B^s = B^{n-1}$ is a hyperplane), the distance d can be obtained as follows: If $(\mathbf{a}, \mathbf{x}) = b$ is an equation of B^{n-1} and $\mathbf{p}$ is the location vector of p, then $d = |(\mathbf{a}, \mathbf{p}) - b|/|\mathbf{a}|$. If $|\mathbf{a}| = 1$ in this equation of B^{n-1}, then d is given simply by $|(\mathbf{a}, \mathbf{p}) - b|$. An equation $(\mathbf{a}, \mathbf{x}) = b$ of a hyperplane is said to be in **Hesse's normal form** if $|\mathbf{a}| = 1$.

I. Spheres and Subspaces

The set of points in a Euclidean space lying at a fixed distance r from a given point is called the **sphere** of **radius** r with **center** at the given point. If $\mathbf{p}$ is the location vector of the center of this sphere with respect to a given rectangular coordinate system, then the equation of the sphere is $|\mathbf{x} - \mathbf{p}| = r$ or $(\mathbf{x}, \mathbf{x}) - 2(\mathbf{p}, \mathbf{x}) + (\mathbf{p}, \mathbf{p}) - r^2 = 0$. The set of points lying at equal distances from $k+1$ points with location vectors $\mathbf{p}_0, \mathbf{p}_1, \ldots, \mathbf{p}_k$ $(k \geqslant 1)$ is a linear subspace of the space (which may be $\varnothing$ or the entire space). If these points are independent, then the subspace has dimension $n-k$, where n is the dimension of the entire space. In particular, if these points are vertices of an n-dimensional †simplex, then there is a unique sphere passing through them, called the **circumscribing sphere** of the simplex. In this case, the simplex is said to be **inscribed** in the sphere. If $\mathbf{p}_0, \mathbf{p}_1, \ldots, \mathbf{p}_n$ are location vectors of the vertices of the simplex, then the equation of the circumscribing sphere of the simplex is given by

$$\begin{vmatrix} 1 & 1 & \ldots & 1 & 1 \\ \mathbf{p}_0 & \mathbf{p}_1 & \cdots & \mathbf{p}_n & \mathbf{x} \\ \mathbf{p}_0^2 & \mathbf{p}_1^2 & \cdots & \mathbf{p}_n^2 & \mathbf{x}^2 \end{vmatrix} = 0.$$

When $n=2$ or 3, there are many classical results concerning the circumscribing circle of a triangle, the circumscribing sphere of a simplex, and other figures related to a triangle or a simplex.

References

[1] G. D. Birkhoff and R. Beatley, Basic geometry, Scott, Foresman, 1941 (Chelsea, third edition, 1959).
[2] E. E. Moise, Elementary geometry from an advanced standpoint, Addison-Wesley, 1963.
[3] H. Weyl, Mathematische Analyse des Raumproblems, Springer, 1923 (Chelsea, 1960).
[4] F. Klein, Vorlesungen über höhere Geometrie, Springer, third edition, 1926 (Chelsea, 1957).
[5] J. Dieudonné, Algèbre linéaire et géométrie élémentaire, Hermann, second corrected edition, 1964; English translation, Linear algebra and geometry, Hermann, 1969.
[6] G. Choquet, L'enseignement de la géométrie, Hermann, 1964.

140 (VI.4) Euclidean Spaces

A space satisfying the axioms of Euclidean geometry is called a **Euclidean space**. An †affine space having as †standard vector space an n-dimensional Euclidean †inner product space over a real number field $\mathbf{R}$ is an n-dimensional Euclidean space E^n. In an n-dimensional Euclidean space E^n, we fix an †orthogonal frame $\Sigma = (O, E_1, \ldots, E_n)$, $e_i = \overrightarrow{OE_i}$, $(e_i, e_j) = \delta_{ij}$. The frame Σ determines †rectangular coordinates $(x_1, x_2, \ldots, x_n)$ of each point in E^n. We can thus establish a one-to-one correspondence between E^n and $\mathbf{R}^n = \{(x_1, \ldots, x_n) | x_i \in \mathbf{R}\}$. In this sense we identify E^n and $\mathbf{R}^n$ and usually call $\mathbf{R}^n$ itself a Euclidean space. The 1-dimensional space $\mathbf{R}^1$ is a straight line, and the †Cartesian product of n copies of $\mathbf{R}^1$ is an n-dimensional Euclidean space (or **Cartesian space**). Given points $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ in the Euclidean space $\mathbf{R}^n$, the †distance $d(x, y)$ between them is given by

$$\sqrt{(y_1 - x_1)^2 + \ldots + (y_n - x_n)^2}.$$

Thus the distance $d(x, y)$ supplies $\mathbf{R}^n$ with the structure of a †metric space. We call x_i the ith coordinate of the point x, the point $(0, \ldots, 0)$ the **origin** of $\mathbf{R}^n$, and the set of points $\{x | -\infty < x_i < \infty;\ x_j = 0, j \neq i\}$ the x_i-**axis** (or i**th coordinate axis**). For an integer m such that $-1 \leqslant m \leqslant n$, we define m-dimensional †subspaces in $\mathbf{R}^n$; a -1-dimensional subspace is the empty set, a 0-dimensional subspace is a point, and a 1-dimensional subspace is a straight line. If we take an orthogonal frame, an m-dimensional subspace is represented as an $\mathbf{R}^m$ (→ 139 Euclidean Geometry; 7 Affine Geometry).

As a †topological space, $\mathbf{R}^n$ is †locally compact and †connected. A bounded closed set in $\mathbf{R}^n$ is †compact (**Bolzano-Weierstrass theorem**).

Given a point $a = (a_1, \ldots, a_n)$ in $\mathbf{R}^n$ and a real positive number r, the subset $\{x | d(x, a) \leqslant r\}$ of

$\mathbf{R}^n$ is called an n-dimensional **solid sphere, solid n-sphere, ball, n-ball, disk,** or **n-disk** with **center** a and **radius** r, its †interior $\{x \mid d(x,a)<r\}$ an n-dimensional **open sphere, open n-sphere, open ball, open n-ball, open disk,** or **open n-disk**, and its †boundary $\{x \mid d(x,a)=r\}$ an $(n-1)$-dimensional **sphere** or **$(n-1)$-sphere**. In particular, a 2-dimensional solid sphere is called a **circular disk**, its interior an **open circle**, and its boundary a **circumference**. A disk or a circumference is sometimes called simply a **circle**. The family of n-dimensional open spheres with center a gives a base for a neighborhood system of the point a. Suppose that we are given a sphere S and two points x, y on S. The points x, y are called **antipodal points** on the sphere S if there exists a straight line L passing through the center of S such that $S \cap L = \{x, y\}$. The segment (or the length of the segment) whose endpoints are antipodal points is called the **diameter** of the solid sphere (or of the sphere). The notion of †diameter (→ 273 Metric Spaces) of a solid sphere or of a sphere considered as a subset of the metric space $\mathbf{R}^n$ coincides with the notion of diameter of the corresponding set defined above. When $n \geqslant 3$, the intersection of a sphere and a 2-dimensional plane passing through the center of the sphere is called a **great circle** of the sphere. For m such that $1 \leqslant m \leqslant n$, we consider an m-dimensional solid sphere or an $(m-1)$-dimensional sphere in an m-dimensional plane $\mathbf{R}^m$. These spheres are also called m-dimensional solid spheres or $(m-1)$-dimensional spheres in $\mathbf{R}^n$.

In particular, the solid sphere of radius 1 having the origin as its center is called the **unit disk, unit ball,** or **unit cell**, and its boundary is called the **unit sphere**. (In particular, when we deal with the 2-dimensional space $\mathbf{R}^2$, we use the term *circle* instead of *sphere*, as in **unit circle**.) The points $(0, \ldots, 0, 1)$ and $(0, \ldots, 0, -1)$ are called the **north pole** and **south pole** of the unit sphere, respectively. The $(n-2)$-dimensional sphere, which is the intersection of the unit sphere and the hyperplane $x_n = 0$, is called the **equator**; the part of the unit sphere that is "above" this hyperplane (i.e., in the half-space $x_n \geqslant 0$) is called the **northern hemisphere**, and the part that is "below" the hyperplane (i.e., in the half-space $x_n \leqslant 0$) the **southern hemisphere**.

Let a_i, b_i be real numbers satisfying $a_i < b_i$ $(i = 1, 2, \ldots, n)$. The subset $\{x \mid a_i < x_i < b_i, i = 1, 2, \ldots, n\}$ of $\mathbf{R}^n$ is called an **open interval** of $\mathbf{R}^n$, and the subset $\{x \mid a_i \leqslant x_i \leqslant b_i\}$ a **closed interval**. They are sometimes called **rectangles** (when $n=2$), **rectangular parallelepipeds**, or **boxes**. An open interval is actually an open set of $\mathbf{R}^n$, and a closed interval is a closed set. In particular, the closed interval $\{x \mid 0 \leqslant x_i \leqslant 1, i = 1, 2, \ldots, n\}$ is called the **unit cube** (or **unit n-cube**) of $\mathbf{R}^n$. We can take the set of open intervals as base for a neighborhood system of $\mathbf{R}^n$.

All the †convex closed sets (for example, closed intervals) having interior points in $\mathbf{R}^n$ are homeomorphic to an n-dimensional solid sphere. A topological space I^n that is homeomorphic to an n-dimensional solid sphere is called an n-dimensional **(topological) solid sphere, (topological) n-cell,** or **n-element**. A topological space S^{n-1} homeomorphic to an $(n-1)$-dimensional sphere is called an $(n-1)$-dimensional **topological sphere** (or simply $(n-1)$-dimensional sphere. The spaces I^n and S^{n-1} are †orientable †topological manifolds whose orientations are determined by assigning the generators of the (relative) †homology groups $H_n(I^n, \dot{I}^n)$ and $H_{n-1}(S^{n-1})$, respectively (both are infinite cyclic groups). By means of the †boundary operator $\partial: H_n(I^n, \dot{I}^n) \rightarrow H_{n-1}(S^{n-1})$, the orientation of I^n or S^{n-1} determines that of the other.

References

See references to 139 Euclidean Geometry.

141 (XXI.22) Euler, Leonhard

Leonhard Euler (April 15, 1707–September 18, 1783) was born in Basel, Switzerland. In his mathematical development he was greatly influenced by the Bernoullis (→ 38 Bernoulli Family). He was invited to the St. Petersburg Academy in 1726 and remained there until 1741, when he was invited to Berlin by Frederick the Great (1712–1786). Euler was active at the Berlin Academy until 1766, when he returned to St. Petersburg. Already having lost the sight of his right eye in 1735, he now became blind in his left eye also. This, however, did not impede his research in any way, and he continued to work actively until his death in St. Petersburg.

Euler was the central figure in the mathematical activities of the 18th century. He was interested in all fields of mathematics, but especially in analysis in the style of †Leibniz, which had been passed down through the Bernoullis and was developed by him into a form that led to the mathematics of the 19th century. Through his work analysis became more easily applicable to the fields of physics and dynamics. He developed calculus further and dealt formally with complex numbers. He also contributed to such fields as †partial differential equations, the theory of †elliptic func-

tions, and the †calculus of variations. He contributed much to the progress of algebra and theory of numbers in this period, and also did pioneering work in topology. He had, however, little of the concern for rigorous foundations that characterized the 19th century. He was the most prolific mathematician of all time, and his collected works are still incomplete, though some seventy volumes have already been published.

References

[1] L. Euler, Opera omnia, ser. 1, vol. 1–29, ser. 2, vol. 1–30, ser. 3, vol. 1–13, Teubner and O. Füssli, 1911–1967.

142 (XV.11) Evaluation of Functions

A. History

By the evaluation of a function $f(x)$ we mean the application of algorithms for obtaining the approximate value of the function. The evaluation methods are classified roughly into two groups: (1) evaluation of functions using approximation formulas, and (2) evaluation using microprogramming techniques. The advent of high-speed computers has brought about drastic changes in the evaluation of functions. Before the introduction of high-speed computers in the 1940s, mathematical tables had played a prominent role. The first issue (published in 1943) of the journal *Mathematical tables and other aids to computation* (MTAC), the predecessor of the journal *Mathematics of computation*, was primarily concerned with tables of mathematical functions. One of the aims of this journal was to facilitate the exchange of information on errors in the tables. The tables, obtained in the past by tedious hand calculation, can now be easily prepared by high-speed computers, and there was a period when more accurate and extensive tables were published one after another. It is ironic, however, that high-speed computers revolutionized numerical analysis and prompted a shift of emphasis in the field, beginning in the 1950s, away from the use of numerical tables and toward exploration of the most efficient methods of approximation of the functions, thus causing a rapid decrease in the need for tables.

Recently, a significant trend in computer design has replaced the conventional logic control section with "stored logic," or microprogrammed control, stored in high-speed, nondestructive Read-Only Storage (ROS). For example, the microprogrammed control used for the unified COordinate Rotation DIgital Computer (**CORDIC**) algorithm is effective in calculating elementary functions because of its simplicity, accuracy, and capability of high-speed execution via parallel processing. It is not clear whether the applications of approximation formulas heretofore in use will be superseded by microprogramming techniques in all kinds of computers. However, the value of the approximation formulas recognized in the 1950s has been declining insofar as elementary functions are concerned. Recently, as a method for the evaluation of functions based on a new viewpoint, some "unrestricted" algorithms have been proposed by Brent [1], which are useful for the computation of elementary and special functions when the required precision is not known in advance or when high accuracy is necessary. It is expected that methods for the evaluation of functions using approximation formulas will be further developed as microprogramming techniques and unrestricted algorithms see wider use.

B. Evaluation of Functions Using Approximation Formulas

Suppose we approximate a function $f(x)$ by a function $g(x)$ using the following class of functions $p_i(x)$ and $q_j(x)$. Let continuous functions $p_1(x), \ldots, p_n(x)$ and $q_1(x), \ldots, q_m(x)$ defined on a closed interval $[a, b]$ satisfy the following conditions: (i) $p_1, \ldots, p_n$ and $q_1, \ldots, q_m$ are both linearly independent; (ii) there exist at most a finite number of zeros for $\sum_{j=1}^{m} b_j q_j(x)$ in $[a, b]$ for any choice of $b_1, b_2, \ldots, b_m$, except for the case $b_1 = b_2 = \ldots = b_m = 0$; (iii) there is a continuous function $g(x)$ with a nonzero denominator in $[a, b]$ representable as

$$g(x) = \sum_{i=1}^{n} a_i p_i(x) \Big/ \sum_{j=1}^{m} b_j q_j(x), \tag{1}$$

where a_i, $i = 1, \ldots, n$, and b_j, $j = 0, 1, \ldots, m$, are constants. Then $g(x)$ is called a **generalized rational function** based on a class of functions $\{p_i(x)\}$, $i = 1, \ldots, n$, and $\{q_j(x)\}$, $j = 1, \ldots, m$. If $p_i(x) = x^{i-1}$ and $q_j(x) = x^{j-1}$, (1) is reduced to a rational function. If $m = 1$ and $q_1(x) = 1$, (1) is a linear combination of $p_i(x)$ and is called an approximation of linear type to $f(x)$; and further, if $p_i(x) = x^{i-1}$, (1) is reduced to a polynomial. The crux of the approximation problem lies in the criterion to be used in choosing the approximate constants in (1). There are three methods for choosing them, which lead to three types of approximation of major importance: (i) interpolatory approximations, (ii) †least-squares approximations, and (iii)

min-max error approximations (sometimes called **best approximations**). As a major aim of computer approximation of a function is to make the maximum error as small as possible, the third type of approximation has been used for digital computers.

For every continuous function $f(x)$, there always exist min-max error approximations of the form (1). In generating a min-max error approximation in practice, we essentially depend upon the following conditions and theorems. Let a function space **F** be a d-dimensional linear space. If for $f \in \mathbf{F}$ which is not identically zero, there exist at most $(d-1)$ zeros of $f(x)$ in $[a, b]$, then **F** is called the **unisolvent** space or the **Haar space**.

Let one of the best approximations to a continuous function $f(x)$ be $g(x)$. If the linear space **D** of all the functions of the form $\{\sum a_i p_i(x) + \sum b_k g(x) q_k(x)\}$ is unisolvent, then the best approximation is unique. For the case where $g(x)$ is an approximation of linear type to $f(x)$, Haar [2] proved the following theorem.

Let $g(x)$ be the best approximation to $f(x)$ and $g(x) = \sum_{i=1}^{m} a_i p_i(x)$. A necessary and sufficient condition for the best approximation to be always unique is that the m-dimensional linear space generated by the functions $p_1, p_2, \ldots, p_m$ be unisolvent. In particular, we have the best unique polynomial approximation (one variable) of $f(x)$ defined on $[a, b]$.

Let the necessary and sufficient condition of the previous theorem be satisfied in a d-dimensional linear space **D**. A necessary and sufficient condition for $g(x)$ given by (1) to be the best approximation of $f(x)$ is that there exist points $a \leqslant x_0 < x_1 < \ldots < x_{d-1} < x_d \leqslant b$, called deviation points, such that $(-1)^i (f(x_i) - g(x_i)) = \mu$ $(i = 0, 1, \ldots, d)$.

A great number of algorithms are known by means of which one can calculate the best approximation $g(x)$ of a function $f(x)$ for the given values of n and m and for $p_i(x) = x^{i-1}$ and $g_j(x) = x^{j-1}$ in (1). However, the following three kinds of algorithms are used most frequently for generating min-max approximations on computers: Remes's second algorithm [3], the differential correction algorithm, and Yamauti's folding-up method.

C. Evaluation of Elementary Functions Based on Microprogramming

Chen [6] has given a general algorithm for calculating an elementary function $z = f(x)$ as follows. Let $F(x, y) = y g(x) + h(x)$, where y is some parameter for evaluating $z_0 = f(x_0) = F(x_0, y_0)$. We assume that (x_0, y_0) has been given and that z_0 is unknown. Suppose that a new point (x_{k+1}, y_{k+1}) is obtained from (x_k, y_k) according to a pair of transformations, $x_{k+1} = \varphi(x_k, y_k)$ and $y_{k+1} = \psi(x_k, y_k)$, keeping the value of $F(x_0, y_0)$ invariant. If x_k is forced to converge to x_ω and if $g(x_\omega) = 1$ and $h(x_\omega) = 0$, then $z_0 = f(x_0) = F(x_0, y_0) = F(x_1, y_1) = \ldots = F(x_\omega, y_\omega) = y_\omega$. In this procedure it is necessary to determine $g(x)$, $h(x)$, $\varphi(x)$, and $\psi(x)$ for the given $f(x)$. Most of the elementary functions can be evaluated by Chen's algorithm, which is essentially identical with Specker's Sequential Table Look-up (**STL**) method based on addition formulas [7], e.g., $\log x = \log(x\alpha) - \log\alpha$ or $e^x = e^{x-\beta} e^\beta$. The iteration equations (2) of CORDIC given later can also be derived by the STL method with complex numbers.

The use of coordinate rotation to evaluate elementary functions is not new. In 1956 and 1959 Volder [8] described the CORDIC for the calculation of trigonometric functions, multiplication, division, and conversion between the binary and the decimal and r-adic number systems. Daggett, in 1959, discussed the use of CORDIC for decimal-binary conversion. It was recognized by Walther [9] in 1971 that these algorithms could be merged into one unified algorithm. Consider coordinate systems parameterized by m in which the radius R and angle A of the vector $P = (x, y)$ are defined as $R = (x^2 + my^2)^{1/2}$ and $A = m^{-1/2} \arctan(m^{1/2} y/x)$. The basis of Walther's algorithm is the coordinate rotation in a linear $(m = 0)$, circular $(m = 1)$, or hyperbolic $(m = -1)$ coordinate system, depending on which function is calculated. Iteration equations of CORDIC are as follows. The point $(x_{i+1}, y_{i+1}, z_{i+1})$ is obtained from the point (x_i, y_i, z_i) by means of the transformation

$$x_{i+1} = x_i + m y_i \delta_i,$$

$$y_{i+1} = y_i - x_i \delta_i, \qquad (2)$$

$$z_{i+1} = z_i + \alpha_i,$$

where m is a parameter for the coordinate system, $\alpha_i = m^{-1/2} \arctan(m^{1/2} \delta_i)$, and δ_i is a suitable value, e.g., $\pm 2^{-i}$. The angle A_{i+1} and radius R_{i+1} are $A_{i+1} = A_i - \alpha_i$ and $R_{i+1} = R_i \times K_i \equiv R_i \times (1 + m\delta_i^2)^{1/2}$. After n iterations we find $A_n = A_0 - \alpha$ and $R_n = R_0 \times K$, and then

$$x_n = K\{x_0 \cos(\alpha m^{1/2}) + y_0 m^{1/2} \sin(\alpha m^{1/2})\},$$

$$y_n = K\{y_0 \cos(\alpha m^{1/2}) - x_0 m^{-1/2} \sin(\alpha m^{1/2})\},$$

$$z_n = z_0 + \alpha,$$

where $\alpha = \sum_{i=0}^{n-1} \alpha_i$ and $K = \prod_{i=0}^{n-1} K_i$. These relations are summarized in Table 1 for $m = 1$, $m = 0$, and $m = -1$ in the following special cases. (i) The value of A is forced to converge to zero; $y_n \to 0$. (ii) The value of z is forced to converge to zero; $z_n \to 0$.

Table 1. Input and Output Functions for CORDIC

Function	m	Input x_0	y_0	z_0	Quantity to be 0	Output
$\sin t$	1	$1/K$	0	t	$z_n \to 0$	$y_n \to \sin t$
$\cos t$	1	$1/K$	0	t	$z_n \to 0$	$x_n \to \cos t$
$\tan^{-1} t$	1	1	t	0	$y_n \to 0$	$z_n \to \tan^{-1} t$
xz	0	x	0	z	$z_n \to 0$	$y_n \to xz$
y/x	0	x	y	0	$y_n \to 0$	$z_n \to y/x$
$\sinh t$	-1	$1/K_{-1}$	0	t	$z_n \to 0$	$y_n \to \sinh t$
$\cosh t$	-1	$1/K_{-1}$	0	t	$z_n \to 0$	$x_n \to \cosh t$
$\tanh^{-1} t$	-1	1	t	0	$y_n \to 0$	$z_n \to \tanh^{-1} t$

$$K_1 = \prod_{j=0}^{n-1} (1+\delta_j^2)^{1/2}, \quad K_{-1} = \prod_{j=0}^{n-1} (1-\delta_j^2)^{1/2}$$

D. Fast Fourier Transform (FFT)

When a function $f(x)$ can be taken to be periodic, it is advantageous to use trigonometric polynomials as least-squares approximating functions. The summations arising in least-squares approximations based on the trigonometric polynomials play an important role in various applications, and a quite efficient algorithm for the evaluation of such sums was developed by Cooley and Tukey [10] in 1965, known as the **fast Fourier transform (FFT)**. Let x_m, $(0 \leqslant m \leqslant N-1)$ be a set of complex numbers, and consider

$$X_n = (1/N) \sum_{m=0}^{N-1} x_m \exp(-2\pi imn/N) \qquad (0 \leqslant n \leqslant N-1). \quad (3)$$

Equation (3) is often called the **discrete Fourier transform (DFT)** of the sequence x_m, it being analogous to the continuous †Fourier transform,

$$X_n = (1/T) \int_0^T x(t) \exp(-2\pi int/T)\,dt \qquad (0 \leqslant n \leqslant N), \quad (4)$$

where $x(t)$ is a periodic function of t with period T. X_n is called the nth Fourier coefficient of a set of N equally spaced samples of size N for the function $x(t)$ $(t = jT/N, 0 \leqslant j \leqslant N)$. In the same way as for the continuous Fourier transform, the discrete transform can be inverted to yield

$$x_n = \sum_{m=0}^{N-1} X_m \exp(2\pi inm/N) \qquad (0 \leqslant n \leqslant N-1).$$

Here x_n is called the coefficient of the inverse Fourier transform, and X_m and x_n thus form a transform pair.

The FFT algorithm for a finite sequence of length N in (3) is based on the fact that the calculation of (3) can be performed in stages by using the direct product decomposition if $N = N_1 N_2$, with N_1 and N_2 relatively prime. For example, the 2-dimensional Fourier transform coefficient is given by

$$X_{n_1,n_2} = [1/(N_1 N_2)] \times \sum_{m_1=0}^{N_1-1} \sum_{m_2=0}^{N_2-1} \exp(-2\pi i(n_1 m_1/N_1 + n_2 m_2/N_2)) x_{m_1,m_2}$$
$$(0 \leqslant n_1 \leqslant N_1 - 1,\ 0 \leqslant n_2 \leqslant N_2 - 1).$$

If by an elementary operation we mean one complex multiplication and one complex addition, we can evaluate X_{n_1,n_2} through $(N_1 N_2)^2$ such operations using Horner's scheme. By the direct product decomposition method, however, the $(N_1 N_2)^2$ operations can be reduced to only $N_1 N_2 (N_1 + N_2)$ operations. Because the matrix corresponding to the transformation mentioned above is a direct product of $N_1 \times N_1$ and $N_2 \times N_2$ matrices, we can perform the calculations in two stages: first to obtain $\xi_{m_1 n_2}$ for $0 \leqslant m_1 \leqslant N_1 - 1$ and $0 \leqslant n_2 \leqslant N_2 - 1$ and then to obtain $X_{n_1 n_2}$ for $0 \leqslant n_2 \leqslant N_2 - 1$ and $0 \leqslant n_1 \leqslant N_1 - 1$. We have

$$\xi_{m_1,n_2} = (1/N_2) \sum_{m_2=0}^{N_2-1} \exp(-2\pi i n_2 m_2/N_2) x_{m_1,m_2},$$
$$X_{n_1,n_2} = (1/N_1) \sum_{m_1=0}^{N_1-1} \exp(-2\pi i n_1 m_1/N_1) \xi_{m_1,n_2}.$$

This direct product decomposition method is well known for 2- or 3-dimensional Fourier transforms. Even for a 1-dimensional Fourier transform of length $N = N_1 N_2$, if N_1 and N_2 are relatively prime, one can use the method of direct product decomposition. Even when N_1 and N_2 are not prime, we can use a method of "pseudo"-direct product decomposition to reduce the number of operations. Namely, we can put $m = m_1 + N_1 m_2$ $(0 \leqslant m_1 < N_1, 0 \leqslant m_2 < N_2)$; $n = N_2 n_1 + n_2$ $(0 \leqslant n_1 < N_1, 0 \leqslant n_2 < N_2)$. Then, similarly as before, X_n in (3) can be rewritten as

$$X_n = (1/N_1) \sum_{m_1=0}^{N_1-1} \exp(-2\pi i n_1 m_1/N_2) \times \exp(-2\pi i m_1 n_2/(N_1 N_2)) \xi_{m_1,n_2},$$

where

$$\xi_{m_1,n_2}=(1/N_2)\sum_{m_2=0}^{N_2-1}\exp(-2\pi i n_2 m_2/N_2)x_{m_1+N_1 m_2}, \quad (5)$$

which is a DFT of length N_2. If we put

$$\tilde{\xi}_{m_1,n_2}=\exp(-2\pi i m_1 n_2/(N_1 N_2))\xi_{m_1,n_2}, \quad (6)$$

then

$$X_n=(1/N_1)\sum_{m_1=0}^{N_1-1}\exp(-2\pi i n_1 m_1/N_1)\tilde{\xi}_{m_1,n_2}. \quad (7)$$

Formula (7) is nothing but a DFT of length N_1. Thus an FFT of length $N=N_1 N_2$ can be calculated by decomposing it into three stages as follows: (i) obtain N_1 transforms of length N_2 in (5); (ii) multiply ξ_{m_1,n_2} by $\exp(-2\pi i m_1 n_2/N)$ in (6) (phase rotation); (iii) calculate N_2 transforms of length N_1 in (7). If either or both of N_1 and N_2 can be factored further so that, e.g., $N=N_1 N_2=N_1 N_{21} N_{22}=\ldots$, an FFT of length N_2 can be calculated similarly by decomposing it, and so on, and in this way one can reduce the total number of operations. This is the principle of FFT pointed out by H. Takahasi. When N is a power of two, it can be shown that the FFT algorithm requires approximately $N\log_2 N/2$ operations.

E. Padé Approximation

Let $f(x)\sim c_0+c_1x+c_2x^2+\ldots$ be a formal power series. For any pair of nonnegative integers (p,q), we define the (p,q)th **Padé approximation** of $f(x)$ as follows: The Padé approximation is a rational function

$$(a_0+a_1x+a_2x^2+\ldots+a_px^p)/(b_0+b_1x+b_2x^2+\ldots+b_qx^q)$$

satisfying the condition that all terms in the formal power series

$$(b_0+b_1x+\ldots+b_qx^q)(c_0+c_1x+\ldots)-(a_0+a_1x+\ldots+a_px^p)$$

should vanish up to the term x^{p+q}. An infinite matrix whose (p,q)th entry is the (p,q)th Padé approximation is called the **Padé table** for $f(x)$. The Padé approximation is uniquely determined, provided that every **Hankel determinant**

$$\begin{vmatrix} c_p & c_{p+1} & \cdots & c_{p+q} \\ c_{p+1} & c_{p+2} & \cdots & c_{p+q+1} \\ \cdots & \cdots & \cdots & \cdots \\ c_{p+q} & c_{p+q+1} & \cdots & c_{p+2q} \end{vmatrix}$$

never vanishes.

When we expand $f(x)$ into a continued fraction of Stieltjes type, say,

$$f(x)\sim\frac{c_0|}{|1}+\frac{\alpha_1x|}{|1}+\frac{\alpha_2x|}{|1}+\ldots,$$

then its $2p$th and $(2p+1)$th approximate fractions are the (p,p)th and $(p+1,p)$th Padé approximation of $f(x)$, respectively.

References

[1] R. P. Brent, Unrestricted algorithms for elementary and special functions, IFIP, 1980, 613–619.
For the theory of approximation,
[2] A. Haar, Die Minkowskische Geometrie und die Annäherung an stetige Funktionen, Math. Ann., (1918), 294–311.
[3] A. Ralston, Rational Chebyshev approximation by Remes' algorithms, Numer. Math., 7 (1965), 322–330.
[4] E. W. Chenney, Introduction to approximation theory, McGraw-Hill, 1966.
[5] M. J. D. Powell, Approximation theory and methods, Cambridge Univ. Press, 1981.
For microprogramming techniques,
[6] T. C. Chen, The automatic computation of exponential, logarithms, ratios and square roots, IBM Res. Rep. RJ 970 (1972), 32.
[7] W. H. Specker, A class of algorithms for $\ln x$, $\exp x$, $\sin x$, $\cos x$, $\tan^{-1}x$ and $\cot^{-1}x$, IEEE Trans. Elec. Comp., EC-14 (1965), 85–86.
[8] J. E. Volder, Binary computation algorithms for coordinate rotation and function generation, Convair Rep. IAR-I 148 Aeroelectronics group, 1956.
[9] J. S. Walther, A unified algorithm for elementary functions, Proc. AFIPS, 1971, Spring Joint Comp. Conf., 379–385.
For fast Fourier transforms,
[10] J. W. Cooley and J. W. Tukey, An algorithm for the machine calculation of complex Fourier series, Math. Comp., 19 (1965), 297–301.
[11] E. C. Brigham, The fast Fourier transform, Prentice-Hall, 1974.
For the Padé approximation,
[12] H. Padé, Sur la représentation approchée d'une fonction par des fractions rationnelles, Ann. Sci. Ecole Norm. Sup., (3) 9 (1892), 1–92.
[13] G. A. Baker, Jr., The Padé approximation 1, 2, Encyclopedia of Mathematics and Its Applications, vols. 13, 14, Addison-Wesley, 1981.
For tables of approximation formulas,
[14] C. Hastings, Jr., Approximations for digital computers, Princeton Univ. Press, 1955.
[15] J. F. Hart (ed.), Computer approximation, Wiley, 1968.

143 (XI.18) Extremal Length

A. General Remarks

The notable relation between the lengths of certain families of curves in a plane domain and the area of the domain has long been recognized and utilized in function theory. L. V. Ahlfors and A. Beurling formulated this relation by introducing the notion of extremal length for families of curves [1]. Although there are various definitions of extremal length, they are essentially the same except for one due to J. Hersch [3] and A. Pfluger [2].

The image of an interval under a continuous mapping is called a curve. We say that it is **locally rectifiable** if every †continuous arc of the curve is †rectifiable. Let C be a finite or countable collection of locally rectifiable curves in a plane and ρ, $0 \leqslant \rho \leqslant \infty$, be a †Baire function defined in the plane. Represent C in terms of arc length s (→ 246 Length and Area), and set $\langle C, \rho \rangle = \int_C \rho\, ds$. For a family Γ of finite or countable collections C, ρ is called **admissible** if $\langle C, \rho \rangle \geqslant 1$ for every $C \in \Gamma$. If no $C \in \Gamma$ consists of a finite or countable number of points, then $\rho \equiv \infty$ is always admissible for Γ. Call $\inf \{\iint \rho^2\, dx\, dy\}$, where ρ runs over admissible Baire functions, the **module** of Γ, and denote it by $M(\Gamma)$. The reciprocal $\lambda(\Gamma) = 1/M(\Gamma)$ is called the **extremal length** of Γ. If no Baire functions are admissible for Γ, we set $\lambda(\Gamma) = 0$. The extremal length is defined equivalently in two other ways as follows: Let ρ be a nonnegative Baire function, and put $L(\Gamma, \rho) = \inf \{\langle C, \rho \rangle \mid C \in \Gamma\}$, then $\lambda(\Gamma) = \sup L(\Gamma, \rho)^2 / \iint \rho^2\, dx\, dy$, where ρ runs over nonnegative Baire functions. Next, let Φ be the collection of nonnegative Baire functions ρ such that $\iint \rho^2\, dx\, dy \leqslant 1$; then $\lambda(\Gamma) = \sup \{L(\Gamma, \rho)^2 \mid \rho \in \Phi\}$. We obtain the same value for $\lambda(\Gamma)$ if we require an admissible ρ to be †lower semicontinuous. If ρ is required to be continuous, then the **extremal length defined by Hersch and Pfluger** is obtained. As is shown in example (1) of Section B, there is a case where the two definitions actually differ.

If an admissible ρ yields $M(\Gamma) = \iint \rho^2\, dx\, dy$, then $\rho |ds|$ is called an extremal metric. Beurling gave a necessary and sufficient condition for a metric to be extremal [4].

We list four properties of extremal length: (1) $\lambda(\Gamma_1) \geqslant \lambda(\Gamma_2)$ if $\Gamma_1 \subset \Gamma_2$. (2) $M(\bigcup_n \Gamma_n) \leqslant \sum_n M(\Gamma_n)$. (3) Let $\{\Gamma_n\}$ and Γ be given. Suppose that there are mutually disjoint measurable sets $\{E_n\}$ such that each $C_n \in \Gamma_n$ is contained in E_n. If each element of $\bigcup_n \Gamma_n$ contains at least one $C \in \Gamma$, then $M(\Gamma) \geqslant \sum_n M(\Gamma_n)$, and hence

$$M\left(\bigcup_n \Gamma_n\right) = \sum_n M(\Gamma_n).$$

If each $C \in \Gamma$ contains at least one $C_n \in \Gamma_n$ for every n, then $\lambda(\Gamma) \geqslant \sum_n \lambda(\Gamma_n)$. (4) Let f be an analytic function in a domain Ω and $\{C\}$ be given in Ω. Denote by $f(C)$ the image of C by f. Then $\lambda(\{C\}) \leqslant \lambda(\{f(C)\})$. The equality holds if f is one-to-one. This shows that $\lambda(\{C\})$ is conformally invariant.

B. Extremal Distance

Let Ω be a domain in a plane, $\partial\Omega$ its boundary, and X_1, X_2 sets on $\Omega \cup \partial\Omega$. The extremal length of the family of curves in Ω connecting points of X_1 and points of X_2 is called the **extremal distance** between X_1 and X_2 (relative to Ω) and is denoted by $\lambda_\Omega(X_1, X_2)$.

Example (1). Let $\Omega = \{z \mid |z| < 2\}$, $X_1 = \partial\Omega$, and X_2 be a countable set in $|z| < 1$ such that the set of accumulation points of X_2 coincides with $|z| = 1$. Then $\lambda_\Omega(X_1, X_2) = \infty$, but the extremal distance in the sense of Hersch and Pfluger is equal to $(2\pi)^{-1} \log 2$.

Example (2). In a rectangle with sides a and b, the extremal distance between the sides of length a is b/a.

Example (3). Let Ω be an annulus $r_1 < |z| < r_2$. The extremal distance λ between the two boundary circles of Ω is equal to $(2\pi)^{-1} \log(r_2/r_1)$. The extremal length of the family of curves in Ω homotopic to the boundary circles is equal to $1/\lambda$.

Example (4). Let Ω be a domain in the extended z-plane such that $\infty \in \Omega$. Let $z_0 \in \Omega$ and $\{|z - z_0| = r\} \subset \Omega$, and denote by λ_r the extremal distance between $\{|z - z_0| = r\}$ and a set $X \subset \partial\Omega$ relative to Ω. Then $\lambda_r - (2\pi)^{-1} \log r$ increases with r. We call the limit the **reduced extremal distance** and denote it by $\tilde{\lambda}_\Omega(X, \infty)$. †Robin's constant for †Green's function in Ω with pole at $z = \infty$ is equal to $2\pi\tilde{\lambda}_\Omega(\partial\Omega, \infty)$.

Extremal length is also defined on Riemann surfaces. Some classical conformal invariants can be given in a generalized form in terms of extremal length. The notion of extremal length has applications in various branches of function theory, such as †conformal and †quasiconformal mappings, the †Phragmén-Lindelöf theorem, the †coefficient problem, the †type problem of Riemann surfaces, and studies of the boundary properties of functions of finite Dirichlet integrals. It is also applied to problems in differential geometry. Extending the notion of extremal length, M. Ohtsuka considered **extremal length with weight**, and B. Fuglede introduced the notion of **generalized**

module in higher-dimensional spaces [8]. These notions have useful properties and applications.

References

[1] L. V. Ahlfors and A. Beurling, Conformal invariants and function-theoretic null-sets, Acta Math., 83 (1950), 101–129.
[2] A. Pfluger, Extremallängen und Kapazität, Comment. Math. Helv., 29 (1955), 120–131.
[3] J. Hersch, Longuers extrémales et théorie des fonctions, Comment. Math. Helv., 29 (1955), 301–337.
[4] M. Ohtsuka, Dirichlet problem, extremal length and prime ends, Van Nostrand, 1970.
[5] Y. Kusunoki, Some classes of Riemann surfaces characterized by the extremal length, Proc. Japan Acad., 32 (1956).
[6] M. Ohtsuka, On limit of BLD functions along curves, J. Sci. Hiroshima Univ., 28 (1964).
[7] T. Fuji'i'e, Boundary behaviors of Dirichlet functions, J. Math. Kyoto Univ., 10 (1970).
[8] B. Fuglede, Extremal length and functional completion, Acta Math., 98 (1957), 171–219.
[9] J. A. Jenkins, Univalent functions and conformal mapping, Erg. Math., Springer, 1958.
[10] J. Väisälä, Lectures on n-dimensional quasiconformal mappings, Lecture notes in math. 229, Springer, 1971.

F

144 (XXI.23) Fermat, Pierre de

Pierre de Fermat (August 20, 1601–January 12, 1665) was born into a family of leather merchants near Toulouse, France. He became an attorney and in 1631 a member of the Toulouse district assembly. When not engaged in such work, he did research in mathematics, so that he consigned his results only to his correspondence or to unpublished manuscripts. The manuscripts were published posthumously by his son in 1679 and are known as *Varia opera mathematica.* His research into number theory, stimulated by Bachet's (1581–1638) translation of the *Arithmetika* of Diophantus (published in 1621), made Fermat's name immortal and initiated modern number theory. He posed the famous Fermat's Problem, which has yet to be solved (→ 145 Fermat's Problem). He began analytic geometry by studying the theory of †conic sections of Apollonius, and utilizing this theory he dealt with the notions of tangent lines, maximal (minimal) values of functions, and quadrature, which made him a pioneer in calculus. He also wrote a precursory work in the theory of probability in the course of his correspondence with †Pascal. †Fermat's principle is important in the field of optics, where it is known as the law of least action. Unlike †Descartes, he emphasized the revival rather than the criticism of Greek mathematics.

References

[1] P. Fermat, Oeuvres I–V and supplement, P. Tannery and C. Henry (eds.), Gauthier-Villars, 1891–1922.

145 (V.16) Fermat's Problem

The **last theorem of Fermat** (c. 1637) asserts that if n is a natural number greater than 2, then

$$x^n+y^n=z^n \tag{1}$$

has no rational integral solution x, y, z with $xyz\neq 0$. In the case $n=2$, equation (1) has integral solutions called **Pythagorean numbers** (→ 118 Diophantine Equations). Fermat read a Latin translation of Diophantus' *Arithmetika,* in which the problem of finding all Pythagorean numbers is treated. In his personal copy of that book, Fermat wrote his assertion as a marginal note at the point at which $n=2$ in equation (1) is treated and added the famous words, "I have discovered a truly remarkable proof of this theorem which this margin is too small to contain." It is not known whether Fermat actually had a proof. Fermat's problem asks for a proof or disproof of this conjecture, which itself has not been solved despite centuries of efforts by many mathematicians; but its study has promoted remarkable advances in number theory. In particular, E. E. Kummer's theory of ideal numbers and the development of the theory of †cyclotomic fields were originally conceived in treating Fermat's problem.

In this article, we consider only those integral solutions x, y, z of equation (1) with $xyz\neq 0$ that are relatively prime. We also restrict ourselves to the cases $n=l$ (odd prime) and $n=4$, without loss of generality.

For smaller values of n, the nonsolvability of equation (1) was proved long ago, for $n=3$ by L. Euler (1770), and later again by A. M. Legendre; for $n=4$ by Fermat and Euler; for $n=5$ by P. G. L. Dirichlet and Legendre (1825); and for $n=7$ by G. Lamé (1839). S. Germain and Legendre found some results on more general cases, but the most remarkable result was obtained by Kummer (*J. Reine Angew. Math.,* 40 (1850), *Abh. Akad. Wiss. Berlin* (1857)).

Let l be an odd prime, ζ a primitive lth root of unity, and h the †class number of the cyclotomic field $\mathbf{Q}(\zeta)$. Then the class number h_2 of the real subfield $\mathbf{Q}(\zeta+\zeta^{-1})$ of $\mathbf{Q}(\zeta)$ divides h. We call $h_1=h/h_2$ and h_2 the †first and †second factors of h, respectively.

(1) If l is †regular, that is, if $(h,l)=1$, then $x^l+y^l=z^l$ has no solution (Kummer, 1850).

There are infinitely many irregular primes [3]; those under 100 are 37, 59, and 67. There are 7,128 regular primes and 4,605 irregular primes between 3 and 125,000. It is not yet known whether there are infinitely many regular prime numbers, although the beginning part of the sequence of natural numbers contains a larger number of these than the number of irregular prime numbers. The condition $(l,h)=1$ is equivalent to saying that the numerators of †Bernoulli numbers B_{2m} $(m=1,2,\ldots,(l-3)/2)$ are not divisible by l (Kummer, 1850).

Kummer obtained a result on irregular primes (1857) which was improved later as follows. Note that if l is not regular then h_1 is divisible by l (Kummer, 1850) (→ 14 Algebraic Number Fields).

(2) If $(h_2,l)=1$ and the numerators of Bernoulli numbers B_{2ml} $(m=1,2,\ldots,(l-3)/2)$ are not divisible by l^3, then $x^l+y^l=z^l$ has no solution (H. S. Vandiver, *Trans. Amer. Math. Soc.,* 31 (1929)). By computation Vandiver confirmed that $x^l+y^l=z^l$ has no solution for

$l<619$. At present, this procedure has been continued for $l \leqslant 125{,}080$ using computers by the method of D. H. Lehmer, E. Lehmer, and H. S. Vandiver, *Proc. Nat. Acad. Sci US*, 40 (1954) (S. S. Wagstaff, *Math. Comp.*, 32 (1978)).

When the condition $(xyz, l)=1$ or $(xyz, l)=l$ is added, we speak of Case I or Case II, respectively. The following theorems hold for Case I.

(3) If $x^l+y^l=z^l$ has a solution in Case I, then

$$B_{2m}f_{l-2m}(t) \equiv 0 \pmod{l}, \quad m=1,2,\ldots,(l-3)/2, \tag{2}$$

holds for $-t=x/y$, y/x, y/z, z/y, x/z, and z/x, where $f_m(t)=\sum_{r=0}^{l-1} r^{m-1} t^r$, and B_m is the mth Bernoulli number. This is called **Kummer's criterion** (D. Mirimanov, 1905).

A simplification of the above result is

(4a) If $x^l+y^l=z^l$ has a solution in Case I, then

$$(2^{l-1}-1)/l \equiv 0 \pmod{l}$$

(A. Wieferich, *J. Reine Angew. Math.*, 136 (1909)). This result created a sensation at the time of its publication. It was first shown that 1093 and 3511 are the only primes with $l<3700$ for which the above congruence holds; it is presently known that no other l with $l \leqslant 6 \times 10^9$ satisfies this congruence. The criterion (4a) was gradually improved by Mirimanov (1910, 1911), P. Furtwängler (1912), Vandiver (1914), G. Frobenius (1914), F. Pollaczeck (1917), T. Morishima (1931), and J. B. Rosser (1940, 1941). For example:

(4b) If $x^l+y^l=z^l$ has a solution in Case I, then

$$(m^{l-1}-1)/l \equiv 0 \pmod{l} \tag{3}$$

holds for all m with $2 \leqslant m \leqslant 43$. By means of this result, Rosser (1941) showed for $l<41{,}000{,}000$, and D. H. Lehmer and E. Lehmer (*Bull. Amer. Math. Soc.*, 47 (1941)) showed for $l<253{,}747{,}889$ that $x^l+y^l=z^l$ has no solution in Case I.

We have hitherto been concerned with rational integral solutions of $x^l+y^l=z^l$. We may also consider the problem of proving or disproving that $\alpha^l+\beta^l=\gamma^l$ has no solution α, β, γ with $\alpha\beta\gamma \neq 0$ in the ring of †algebraic integers of $\mathbf{Q}(\zeta)$. Case I means the impossibility of

$$\alpha^l+\beta^l+\gamma^l=0, \quad (\alpha\beta\gamma, l)=1, \tag{4}$$

and Case II means the impossibility of

$$\alpha^l+\beta^l=\varepsilon\lambda^{nl}\gamma^l, \quad (\alpha\beta\gamma, l)=1, \tag{5}$$

where n is a natural number, ε is a †unit in $\mathbf{Q}(\zeta)$, and $\lambda=1-\zeta$. We have the following results:

(1*) If $(h,l)=1$, then neither equation (4) nor equation (5) has a solution (Kummer, 1850).

(2*) Under the same conditions as in statement (2), equation (4) has no solution. If we additionally restrict α, β, γ to relatively prime integers of $\mathbf{Q}(\zeta+\zeta^{-1})$ and replace λ by $(1-\zeta)(1-\zeta^{-1})$, then equation (5) also has no solution (Vandiver, 1929).

(4b*) If equation (4) has solutions α, β, γ in $\mathbf{Q}(\zeta)$, then congruence relation (3) holds for all m with $2 \leqslant m \leqslant 43$ (Morishima, 1934).

When l is sufficiently large, we have the results of M. Krasner (*C. R. Acad. Sci. Paris* (1934)) and Morishima (*Proc. Japan Acad.*, 11 (1935)).

Bibliographies are given in Vandiver and Wahlin [1] and Vandiver [2].

References

[1] H. S. Vandiver and G. E. Wahlin, Algebraic numbers II, Bull. Nat. Res. Council, no. 62, 1928.
[2] H. S. Vandiver, Fermat's last theorem, Amer. Math. Monthly, 53 (1946), 555–578.
[3] P. Ribenboim, 13 lectures on Fermat's last theorem, Springer, 1979.

146 (XX.30) Feynman Integrals

A. Introduction

As the S-matrix or †Green's function in quantum field theory is usually prohibitively difficult to calculate, perturbative expansions in terms of coupling constants have been employed since the beginning of the theory (→ 386 S-Matrices). R. P. Feynman (*Phys. Rev.*, 76 (1949)) invented a way of calculating the series in terms of **Feynman integrals**. His method drastically simplified the preceding method due to S. Tomonaga and J. S. Schwinger, even though, as was later shown, the two methods are theoretically equivalent (F. J. Dyson, *Phys. Rev.*, 75 (1949)). A Feynman integral is an integral associated with a **Feynman graph** according to the **Feynman rule** explained in Section B. Feynman integrals inherit the troublesome problem of divergence, and some recipe which systematically provides them with a definite meaning is needed. Such a recipe is given by the renormalization theory of Tomonaga, Schwinger, Feynman, and Dyson. A mathematically rigorous renormalization theory was given by N. N. Bogolyubov and O. S. Parasiuk (*Acta Math.*, 97 (1957)), later supplemented by K. Hepp (*Comm. Math.

Phys., 2 (1966)). See also W. Zimmermann (*Comm. Math. Phys.*, 15 (1969)) and S. A. Anikin et al. (*Theor. Math. Phys.*, 17 (1973)). Furthermore, E. R. Speer [1] gave a mathematically convenient recipe of renormalization (under the condition that massless particles are irrelevant). Although the series expansion in coupling constants is a divergent series even after renormalization (→ 386 *S*-Matrices), the study of Feynman integrals has given much insight into the qualitative aspects of the *S*-matrix, and in particular, into its analytic structure (e.g., R. J. Eden et al. [2]). In this respect the discovery of the Landau-Nakanishi equations, which describe the location of singularities of Feynman integrals, was crucially important (L. D. Landau, N. Nakanishi, and J. Bjorken, 1959; → Section C). Later, R. E. Cutkosky found a formula which gives important information concerning the ramification of Feynman integrals near their singularities (→ Section C). It gave impetus to J. Leray's mathematical study of Feynman integrals from the viewpoint of integration of multivalued analytic functions (Leray, *Bull. Soc. Math. France*, 87 (1959)). Such studies were subsequently carried out by D. Fotiadi, M. Froissart, J. Lascoux, F. Pham, etc.; → [3–5] and references cited there for this topic. An extensive study by G. Ponzano, T. Regge, Speer, and J. M. Westwater on the monodromy structure of Feynman integrals is closely related to the studies by Pham and others (→ Regge in [6] and references cited there). On the other hand, the progress of microlocal analysis has thrown new light on the Feynman integrals and has given a unified foundation to these various other studies (→ Section C; also Pham, M. Kashiwara, and T. Kawai in [6], M. Sato et al. in [6] and references cited there).

B. Definitions

First, the notion of **Feynman graphs** is introduced. A Feynman graph is sometimes called a **Feynman diagram**. A Feynman graph G consists of finitely many points (called vertices) $\{V_j\}_{j=1,\dots,n'}$, finitely many 1-dimensional segments (called internal lines) $\{L_l\}_{l=1,\dots,N}$ and finitely many half-lines (called external lines) $\{L_r^e\}_{r=1,\dots,n}$, all of which are located in a 4-dimensional affine space. Each of the endpoints W_l^+ and W_l^- of L_l and the endpoint of L_r^e coincide with some vertex V_j. A four-vector $p_r=(p_{r,0},p_{r,1},p_{r,2},p_{r,3})$ is associated with each external line L_r^e and a constant $m_l\geqslant 0$ is associated with each internal line L_l. For simplicity, we usually suppose, in addition, that each internal line and each external line are oriented, that $W_l^+\neq W_l^-$ and that G is connected. The orientation of a line is indicated by the symbol →. Given an orientation, we define the incidence number $[j:l]$ to be $+1$ or -1 according to whether L_l ends or starts from V_j. In other cases, $[j:l]$ is defined to be zero. The incidence number $[j:r]$ is defined in the same way but with L_r^e replacing L_l (Fig. 1).

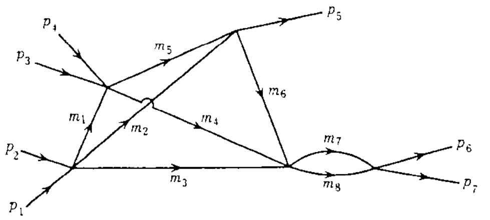


Fig. 1
In this example of a Feynman graph, the internal lines L_2 and L_4 do not intersect and this diagram should be drawn in $\mathbf{R}^4$, not in $\mathbf{R}^2$; this is indicated by ⤫. For convenience, multiple lines such as L_7 and L_8 are usually drawn in a curvilinear manner, as shown.

The **Feynman rule** associates the following integral $F_G(p)$ with each Feynman graph G:

$$F_G(p)=\int\frac{\prod_{j=1}^{n'}\delta^4(\sum_{r=1}^{n}[j:r]p_r+\sum_{l=1}^{N}[j:l]k_l)}{\prod_{l=1}^{N}(k_l^2-m_l^2+\sqrt{-1}\,0)}\times\sum_{l=1}^{N}d^4k_l, \quad (1)$$

where $k_l^2=k_{l,0}^2-\sum_{v=1}^{3}k_{l,v}^2$. Here $1/(k_l^2-m_l^2+\sqrt{-1}\,0)$ means $\lim_{\varepsilon\downarrow 0}(1/(k_l^2-m_l^2+\sqrt{-1}\,\varepsilon)$ (→ 125 Distributions and Hyperfunctions).

Here we consider the case where the interaction Lagrangian density does not contain differential operators (i.e., direct coupling) and all relevant particles are spinless. In general, we should multiply the integrand of $F_G(p)$ by a matrix of polynomials of the p_r and k_l.

$F_G(p)$ has the form $\delta(\sum_{j,r}[j:r]p_r)f_G(p)$; we often investigate $f_G(p)$ instead of $F_G(p)$. The function $f_G(p)$ is studied on $M=_{\text{def}}\{p\in\mathbf{R}^{4n}\mid\sum_{j,r}[j:r]p_r=0\}$, and is called a **Feynman amplitude**. The integral (1) is not well defined as it stands because of the following problems: (a) Is the product appearing in the integrand well defined? (b) Is the integral convergent? The first problem is not serious if $m_l\neq 0$ (→ 274 Microlocal Analysis E) However, the second problem, called the **ultraviolet divergence**, is serious. The renormalization procedure is intended to overcome this difficulty. When some m_l is equal to zero, even the first problem, called the **infrared divergence**, is serious. See D. R. Yennie et al. (*Ann. Phys.*, 13 (1961)) and T. Kinoshita (*J. Math. Phys.*, 3 (1962)) for analyses of the infrared divergenee. In this article we always assume for simplicity that every m_l is strictly positive, even though

such an assumption is too restrictive from the physical viewpoint.

Calculations of Feynman integrals are often done by means of the **parametric representation** of the integral (1), of the form

$$\delta\left(\sum_{j,r}[j:r]p_r\right)\int_0^1 \ldots$$
$$\times\int_0^1 \frac{\delta(1-\sum_{l=1}^N\alpha_l)\prod_{l=1}^N d\alpha_l}{U(\alpha)^2(V(p,\alpha)+\sqrt{-1}\,0)^{-N+2n'-2}},$$

where $U(\alpha)$ and $V(p,\alpha)$ are determined by the topological structure of the graph G. See [7] for the derivation of this formula and the treatment of the integral written in this form. It is useful not only for the study of the singularity structure of $F_G(p)$ but also for the study of spectral representations, etc. (→, e.g., [7, 8]). Note that several different notations are used in the literature for the parametric representation of the integral. Hence one should be careful in referring to papers which use parametric representations.

C. Analytic Properties of Feynman Integrals

The celebrated result of Landau (*Nuclear Phys.*, 13 (1959)), Nakanishi (*Prog. Theor. Phys.*, 22 (1959)) and Bjorken (thesis, Stanford Univ., 1959) asserts that the singularities of a Feynman amplitude $f_G(p)$ are confined to the subset $L^+(G)$ (called a positive-α **Landau-Nakanishi variety**) of $M=\{p\in\mathbf{R}^{4n}\mid\sum_{j,r}[j:r]p_r=0\}$, defined by the following set of equations (called **Landau-Nakanishi equations**), where u_r, w_j, and k_l are real four-vectors and α_l is a real number, all of which are to be eliminated to define relations among the p_r (note, however, that a positive-α Landau-Nakanishi variety is not strictly a subvariety of M, because of the constraint (2e):

$$u_r=\sum_{j=l}^{n'}[j:r]w_j \qquad (r=1,\ldots,n), \tag{2a}$$

$$\sum_{r=1}^{n}[j:r]p_r+\sum_{l=1}^{N}[j:l]k_l=0 \qquad (j=1,\ldots,n'), \tag{2b}$$

$$\sum_{j=1}^{n'}[j:l]w_j=\alpha_l k_l \qquad (l=1,\ldots,N), \tag{2c}$$

$$\alpha_l(k_l^2-m_l^2)=0 \qquad (l=1,\ldots,N), \tag{2d}$$

$$\alpha_l\geqslant 0, \tag{2e}$$

with some

$$\alpha_l\neq 0. \tag{2f}$$

Usually a Landau-Nakanishi variety (resp., equation) is called a **Landau variety** (resp., **equation**) for short. The equation (2a) is conventional; (2b) represents the energy-momentum conservation law at the vertex V_j; (2d) corresponds to the mass-shell constraint (if $\alpha_l\gneqq 0$) on the internal line L_l. Since (2c) entails $\sum_l\varepsilon_l(C)\alpha_l k_l=0$ for a closed circuit (=loop) C of G for some set of values $\varepsilon_l(C)=\pm 1$ or 0 with $\varepsilon_l(C)$ being 0 if L_l does not belong to C, the equation (2c) is usually called the closed-circuit condition. By attaching w_j to the vertex V_j and associating $\alpha_l k_l$ to the internal line L_l, we get a diagram representing a multiple collision of classical point particles, since the relations (2a)–(2f) are just the classical conditions for such a collision (S. Coleman and R. E. Norton, *Nuovo Cimento*, 38 (1965)). The fact that the Landau-Nakanishi equations admit such an interpretation is neither accidental nor superficial in view of the †macroscopic causality of the S-matrix. Note also that there is another interpretation of the Landau-Nakanishi equations, which emphasizes their resemblance to †Kirchhoff's law (Nakanishi, *Prog. Theor. Phys.*, 23 (1960)). Such a resemblance can be used to study the structure of Feynman integrals from the viewpoint of †graph theory. See [7] and the references cited there for this topic.

An important observation by Pham and Sato (1973), which opened a way to the microlocal analysis of Feynman integrals and the S-matrix, is the following: If we consider the Landau-Nakanishi equations to define a subvariety of S^*M, the †spherical cotangent bundle of M, by eliminating only w, k, and α, then the resulting variety describes the †singularity spectrum of $f_G(p)$. More precisely, $\mathrm{S.S.}f_G(p)$ is confined to the set $(p;\sqrt{-1}u)$, where $(p;u)$ satisfies the Landau-Nakanishi equations. The rigorous proof of this statement was given by Sato et al. in [6]. The subset of S^*M or $\sqrt{-1}S^*M$ thus obtained is denoted by $\mathscr{L}^+(G)$ and is also called a positive-α Landau-Nakanishi variety. It is noteworthy that the microlocalization of the classical result of Landau, Nakanishi, and Bjorken had essentially been achieved in a less sophisticated manner by D. Iagolnitzer and H. P. Stapp (*Comm. Math. Phys.*, 14 (1969)) in the framework of S-matrix theory. The variety defined by (2a)–(2d) and (2f) is denoted by $L(G)$ or $\mathscr{L}(G)$ and is a Landau-Nakanishi variety. In a neighborhood of p^0 in $L^+(G)$, $f_G(p)$ is the boundary value of a holomorphic function $\tilde{f}_G(p)$ whose domain of definition is determined by u-vectors (→ 274 Microlocal Analysis E). Furthermore, in simple cases one can verify that $\tilde{f}_G(p)$ can be analytically continued to define a holomorphic function on the universal covering space $\widetilde{U-L(G)^{\mathbf{C}}}$ of $U-L(G)^{\mathbf{C}}$ for a complex neighborhood U of p^0, where $L(G)^{\mathbf{C}}$ denotes a complexification of $L(G)$. Hence we can discuss the difference of

$\tilde{f}_G(\gamma p)$ and $\tilde{f}_G(p)$, where γ denotes a loop encircling $L(G)^C$. Cutkosky (*J. Math. Phys.*, 1 (1960)) observed that it can be expressed (in simple cases) by the integral obtained by replacing $1/(k_l^2-m_l^2+\sqrt{-1}0)$ in the right-hand side of (1) by $-2\pi\sqrt{-1}\delta^+(k_l^2-m_l^2)=_{\text{def}} -2\pi\sqrt{-1}\,Y(k_{l,0})\delta(k_l^2-m_l^2)$, where $Y(k_{l,0})$ denotes the †Heaviside function. A result of this type is called a **discontinuity formula** and is now obtained for the S-matrix itself in a suitably modified manner, i.e., the discontinuity formula holds beyond the framework of perturbation theory (→ 386 S-Matrices). As is mentioned in 386 S-Matrices, the discontinuity formula is closely related to Sato's conjecture on the †holonomic character of the S-matrix (T. Kawai and Stapp in [6]). Actually, M. Kashiwara and Kawai (in [6]) proved that $f_G(p)$ satisfies a †holonomic system of linear differential equations whose characteristic variety is confined to the extended Landau-Nakanishi variety $L(G)$. They further proved that the system has regular singularities (→ 274 Microlocal Analysis G). Their result gives, on one hand, a precise version of Regge's statement to the effect that $f_G(p)$ is a generalization of a †hypergeometric function (in *Battelle Rencontres*, C. DeWitt and J. A. Wheeler (eds.), Benjamin, 1968), and, on the other hand, a rigorous proof of the fact that $\tilde{f}_G(p)$ is a Nilsson class function. This fact is closely related to the works of D. Fotiadi, J. Lascoux, Pham, and others. Kashiwara and Kawai (*Comm. Math. Phys.*, 54 (1977)) also showed that the holonomic character of the Feynman amplitude is an important clue for understanding the so-called hierarchical principle, which had been proposed and studied in connection with the †Mandelstam representation by the Cambridge group (Eden et al. [2]). Furthermore, Kashiwara et al. (*Comm. Math. Phys.*, 60 (1978)) gave a useful expression of $f_G(p)$ at several physically important points by analyzing the microlocal structure of the holonomic system that $f_G(p)$ satisfies. Thus the use of †(micro-) differential equations in analyzing Feynman amplitudes has turned out to be effective in understanding their singularity structures in a unified manner.

References

[1] E. R. Speer, Generalized Feynman amplitudes, Princeton Univ. Press, 1969.
[2] R. J. Eden, P. V. Landshoff, D. I. Olive, and J. C. Polkinghorne, The analytic S-matrix, Cambridge Univ. Press, 1966.
[3] R. C. Hwa and V. L. Teplitz, Homology and Feynman integrals, Benjamin, 1966.
[4] F. Pham, Introduction a l'étude topologique de singularités de Landau, Gauthier-Villars, 1967.
[5] S. Lefschetz, Applications of algebraic topology, Springer, 1975.
[6] Publ. Res. Inst. Math. Sci., 12, suppl. (1977) (Proc. Oji Seminar on Algebraic Analysis, 1976).
[7] N. Nakanishi, Graph theory and Feynman integrals, Gordon & Breach, 1971. (Original in Japanese.)
[8] I. T. Todorov, Analytic properties of Feynman diagrams in quantum field theory, Pergamon, 1971. (Original in Russian, 1966.)

147 (IX.13)
Fiber Bundles

A. General Remarks

E. Stiefel [2] introduced certain †diffeomorphism invariants of †differentiable manifolds by considering a field of a finite number of linearly independent vectors attached to each point of a manifold; and H. Whitney [3] obtained the notion of fiber bundles as a compound idea of a manifold and such a field of tangent vectors. S. S. Chern [4] emphasized the global point of view in differential geometry by recognizing the relation between the notion of †connections (due to E. Cartan) and the theory of fiber bundles. The theory of fiber bundles is also applied to various fields of mathematics, for example, the theory of †Lie groups, †homogeneous spaces, †covering spaces, and general vector bundles, vector bundles of class C^r, or analytic vector bundles.

Homological properties of fiber bundles are studied by means of †spectral sequences, and cohomology structures of several homogeneous spaces and several characteristic classes are determined explicitly by means of †cohomology operations. Also, the group $K(X)$, formed by equivalence classes of vector bundles over a finite †CW complex X, is a †generalized cohomology group, treated in †K-theory, in which further development is expected (→ 237 K-Theory).

B. Definitions

Let E, B, F be topological spaces, $p: E\to B$ a continuous mapping, and G an †effective left topological †transformation group of F. If there exist an †open covering $\{U_\alpha\}$ $(\alpha\in\Lambda)$ of B and a homeomorphism $\varphi_\alpha: U_\alpha\times F\approx p^{-1}(U_\alpha)$ for each $\alpha\in\Lambda$ having the following three properties, then the system $(E, p, B, F, G, U_\alpha, \varphi_\alpha)$ is

called a **coordinate bundle**: (1) $p\varphi_\alpha(b,y)=b$ $(b\in U_\alpha, y\in F)$. (2) Define $\varphi_{\alpha,b}:F\approx p^{-1}(b)$ $(b\in U_\alpha)$ by $\varphi_{\alpha,b}(y)=\varphi_\alpha(b,y)$; then $g_{\beta\alpha}(b)=\varphi_{\beta,b}^{-1}\circ\varphi_{\alpha,b}\in G$ for $b\in U_\alpha\cap U_\beta$. (3) $g_{\beta\alpha}:U_\alpha\cap U_\beta\rightarrow G$ is continuous. We say that this bundle is **equivalent** to a coordinate bundle $(E,p,B,F,G,U'_\mu,\varphi'_\mu)$ if $\bar{g}_{\mu\alpha}(b)=\varphi'^{-1}_{\mu,b}\circ\varphi_{\alpha,b}\in G$ $(b\in U_\alpha\cap U'_\mu)$ and $\bar{g}_{\mu\alpha}:U_\alpha\cap U'_\mu\rightarrow G$ is continuous. An equivalence class $\xi=(E,p,B,F,G)$ of coordinate bundles is called a **fiber bundle** (or G-**bundle**), and E is called the **total space** (or **bundle space**), p the **projection**, B the **base space**, F the **fiber**, and G the **bundle group** (or **structure group**). Also, U_α of a coordinate bundle $(E,p,B,F,G,U_\alpha,\varphi_\alpha)$ belonging to the class ξ is called the **coordinate neighborhood**, φ_α the **coordinate function**, and $g_{\beta\alpha}$ the **coordinate transformation** (or **transition function**).

Let $\xi=(E,p,B,F,G)$ and $\xi'=(E',p',B',F,G)$ be two fiber bundles with the same fiber and group. A continuous mapping $\Psi:E\rightarrow E'$ is called a **bundle mapping (bundle map)** from ξ to ξ' if the following two conditions are satisfied: (1) There is a continuous mapping $\psi:B\rightarrow B'$ with $p'\circ\Psi=\Psi\circ p$. (2) $\psi_{\mu\alpha}(b)=\varphi'^{-1}_{\mu,b'}\circ\Psi\circ\varphi_{\alpha,b}\in G$ $(b\in U_\alpha\cap\psi^{-1}(V'_\mu), b'=\psi(b))$, and $\psi_{\mu\alpha}:U_\alpha\cap\psi^{-1}(V'_\mu)\rightarrow G$ is continuous, where $\{U_\alpha,\varphi_\alpha\}$ and $\{V'_\mu,\varphi'_\mu\}$ are pairs of coordinate neighborhoods and functions of ξ and ξ', respectively. Moreover, if ψ is a homeomorphism, then Ψ is also a homeomorphism and Ψ^{-1} is a bundle mapping.

Let $\xi=(E,p,B,F,G)$ and $\xi'=(E',p',B,F,G)$ be two fiber bundles with the same base space, fiber, and group. If there is a bundle mapping $\Psi:E\rightarrow E'$ such that $\psi:B\rightarrow B$ as described before is the identity mapping, then we say that ξ is **equivalent** (or **isomorphic**) to ξ' and write $\xi\equiv\xi'$. Take the same coordinate neighborhoods $\{U_\alpha\}$, and let $g_{\beta\alpha}$ and $g'_{\beta\alpha}$ be the coordinate transformations of ξ and ξ', respectively. Then $\xi\equiv\xi'$ if and only if there are continuous mappings $\lambda_\alpha:U_\alpha\rightarrow G$ with $g'_{\beta\alpha}(b)=\lambda_\beta(b)g_{\beta\alpha}(b)\lambda_\alpha(b)^{-1}$ $(b\in U_\alpha\cap U_\beta)$.

For a system $\{g_{\beta\alpha}\}$ of coordinate transformations of a fiber bundle, we have $g_{\gamma\beta}(b)g_{\beta\alpha}(b)=g_{\gamma\alpha}(b)$ $(b\in U_\alpha\cap U_\beta\cap U_\gamma)$. Conversely, given an open covering $\{U_\alpha\}$ and a system of $g_{\beta\alpha}:U_\alpha\cap U_\beta\rightarrow G$ satisfying, this condition, there is a unique G-bundle (E,p,B,F,G) with $\{g_{\beta\alpha}\}$ as a system of coordinate transformations. Actually, E is the †identification space of $\tilde{E}=\{(b,y,\alpha)\mid b\in U_\alpha\}\subset B\times F\times\Lambda$ obtained by identifying two points (b,y,α), (b',y',β) with $b=b'$, $y'=g_{\beta\alpha}(b)\cdot y$, and p is defined by $p\{(b,y,\alpha)\}=b$, where the index set $\Lambda=\{\alpha\}$ is considered a discrete space.

C. Principal Fiber Bundles

A fiber bundle $\eta=(P,q,B,G,G)$ is called a **principal fiber bundle** (or simply **principal bundle**) if G operates on G by left translations. This is also defined by the following conditions: G is a right †topological transformation group of P, and there exist an open covering $\{U\}$ of B and homeomorphisms φ: $U\times G\approx q^{-1}(U)$ with $q\varphi(b,g)=b$, $\varphi(b,g)\cdot g'=\varphi(b,gg')$ $(b\in U;g,g'\in G)$. A bundle mapping $\Psi:P\rightarrow P'$ between two principal bundles $\eta=(P,q,B,G)$ and $\eta'=(P',q',B',G)$ is also defined as a continuous mapping Ψ with $\Psi(x\cdot g)=\Psi(x)\cdot g$.

D. Associated Fiber Bundles

Let $\eta=(P,q,B,G)$ be a principal bundle, and let F be a topological space having G as an effective left topological transformation group. Then G is a right topological transformation group of the product space $P\times F$ by $(x,y)\cdot g=(x\cdot g,g^{-1}\cdot y)$ $(x\in P,y\in F,g\in G)$. Consider the †orbit space $P\times_G F=(P\times F)/G$, and define the continuous mapping $p:P\times_G F\rightarrow B$ by $p\{(x,y)\}=q(x)$. Then $\eta\times_G F=(P\times_G F,p,B,F,G)$ is a fiber bundle, called the **associated fiber bundle** of η with fiber F. On the other hand, η is called an **associated principal bundle** of $\xi=(E,p,B,F,G)$ if $\xi\equiv\eta\times_G F$. A principal bundle η having the same coordinate transformations as ξ is an associated principal bundle of ξ, and two fiber bundles are equivalent if and only if their associated principal bundles are equivalent. Therefore, given a fiber bundle ξ, there exists a principal bundle η such that $\xi=\eta\times_G F$.

E. Examples of Fiber Bundles

(1) Product bundle. $(B\times F,p_1,B,F,G)$, where p_1, the projection of the product space, is called a **product bundle** if there is just one coordinate neighborhood B and the coordinate function is the identity mapping of $B\times F$. A bundle that is equivalent to a product bundle is called a **trivial bundle**.

(2) A †covering $(\tilde{Y},p,Y)$ is a fiber bundle whose fiber is the discrete space $p^{-1}(y_0)$ $(y_0\in Y)$, and the structure group is a factor group of the †fundamental group $\pi_1(Y,y_0)$. In particular, a †regular covering is a principal bundle.

(3) Hopf bundle. Let Λ be the real number field $\mathbf{R}$, the complex number field $\mathbf{C}$, or the quaternion field $\mathbf{H}$, $\lambda=\dim_{\mathbf{R}}\Lambda$, and Λ^{n+1} the $(n+1)$-dimensional linear space over Λ. Identify two points $(z_0,\ldots,z_n)$, $(z'_0,\ldots,z'_n)$ of the subspace $\Lambda^{n+1}-\{0\}$ (0 is the origin) if there is a $z\in\Lambda$ such that $z_i=z'_iz$ $(i=0,\ldots,n)$. Then we obtain the identification space $P^n(\Lambda)$, called the n-dimensional **projective space over** Λ. Let S^n_Λ $(=S^{\lambda(n+1)-1}$, the $(\lambda(n+1)-1)$-sphere) be the unit sphere in Λ^{n+1}. Then S^0_Λ is the topological

transformation group of S_Λ^n by the product of Λ, and the orbit space S_Λ^n/S_Λ^0 is $P^n(\Lambda)$. Furthermore, $(S_\Lambda^n, q, P^n(\Lambda), S_\Lambda^0)$ (q is the projection) is a principal bundle called the **Hopf bundle** (or **Hopf fibering**). These comments are valid also for $n=\infty$. When $n=1$, $P^1(\Lambda)$ is homeomorphic to S^λ, and the Hopf bundle is $(S^{2\lambda-1}, q, S^\lambda, S^{\lambda-1})$ ($\lambda=1,2,4$). A Hopf bundle is defined similarly for $\lambda=8$ using the †Cayley algebra, and the projection $q: S^{2\lambda-1}\rightarrow S^\lambda$ ($\lambda=2,4,8$) is the **Hopf mapping (Hopf map)**.

(4) Let G be a topological group, H its closed subgroup, and $r: G\rightarrow G/H$, $r(g)=gH$ the natural projection. If there exist a neighborhood U of $r(H)\in G/H$ and a continuous mapping $f: U\rightarrow G$ such that $r\circ f$ is the identity mapping, then we say that H has a **local cross section** f in G, and $(G/K, p, G/H, H/K, H/K_0)$ is a fiber bundle for any closed subgroup K of H (where p is the natural projection $gK\rightarrow gH$ and K_0 is the largest normal subgroup of H contained in K). The associated principal bundle of the latter fiber bundle is $(G/K_0, p, G/H, H/K_0)$. Any closed subgroup H of a †Lie group G has a local cross section in G; hence the above bundles can be obtained.

F. Vector Bundles

A system $\xi=(E,p,B)$ of topological spaces E, B and a continuous mapping $p: E\rightarrow B$ is called an n-dimensional real **vector bundle** if the following two conditions are satisfied: (1) $p^{-1}(b)$ is a real vector space for each $b\in B$. (2) There exist an open covering $\{U_\alpha\}$ ($\alpha\in\Lambda$) of B and a coordinate function $\varphi_\alpha: U_\alpha\times F\approx P^{-1}(U_\alpha)$ for each $\alpha\in\Lambda$, where $F=\mathbf{R}^n$; furthermore, the $\varphi_{\alpha,b}$: $\mathbf{R}^n\approx p^{-1}(b)$ are isomorphisms of vector spaces. In this case, $g_{\beta\alpha}(b)=\varphi_{\beta,b}^{-1}\circ\varphi_{\alpha,b}: \mathbf{R}^n\approx\mathbf{R}^n$ ($b\in U_\alpha\cap U_\beta$) is an element of the †general linear group $GL(n,\mathbf{R})$. Hence a vector bundle is a fiber bundle with fiber $\mathbf{R}^n$ and group $GL(n,\mathbf{R})$, and the converse is also true. A 1-dimensional vector bundle is called a **line bundle**. A vector bundle $\xi'=(E',p',B)$ is called a **subbundle** of a vector bundle $\xi=(E,p,B)$ if $E'\subset E$, $p|E'=p'$, and $p'^{-1}(b)$ is a vector subspace of $p^{-1}(b)$ for each $b\in B$.

Let ξ_1 and ξ_2 be two vector bundles of dimension n_1 and n_2 with the same base space B. Let E be the union of the direct sum $p_1^{-1}(b)+p_2^{-1}(b)$ for $b\in B$, and define $p: E\rightarrow B$ by $p(p_1^{-1}(b)+p_2^{-1}(b))=b$. Take the same coordinate neighborhoods U_α for ξ_1 and ξ_2, and define $\varphi_\alpha: U_\alpha\times\mathbf{R}^{n_1+n_2}\rightarrow p^{-1}(U_\alpha)$ by $\varphi_\alpha(b,y)=(\varphi_{\alpha,b}^1+\varphi_{\alpha,b}^2)(y)$ ($y\in\mathbf{R}^{n_1+n_2}=\mathbf{R}^{n_1}+\mathbf{R}^{n_2}$), where the $\varphi_\alpha^i: U_\alpha\times\mathbf{R}^{n_i}\approx p_i^{-1}(U_\alpha)$ are the coordinate functions of ξ_i. Then E is topologized by taking the family $\{\varphi_\alpha(O)\}$ (O is open in $U_\alpha\times\mathbf{R}^{n_1+n_2}$) as the †open base, and we obtain an (n_1+n_2)-dimensional real vector bundle (E,p,B), denoted by $\xi_1\oplus\xi_2$ and called the **Whitney sum** of ξ_1 and ξ_2. Similarly, we can define the **tensor product** $\xi_1\otimes\xi_2$, the **p-fold exterior power** $\bigwedge^p\xi$ (or **bundle $\xi^{(p)}$ of p-vectors**), and the **bundle of homomorphisms** $\mathrm{Hom}(\xi_1,\xi_2)$, of dimension n_1n_2, $\binom{n}{p}$, and n_1n_2, respectively, using the tensor product $\mathbf{R}^{n_1}\otimes\mathbf{R}^{n_2}=\mathbf{R}^{n_1n_2}$, the p-fold †exterior power $\bigwedge^p\mathbf{R}^n=\mathbf{R}^{\binom{n}{p}}$ (the space $(\mathbf{R}^n)^{(p)}$ of p-vectors in $\mathbf{R}^n$), and the space of homomorphisms $\mathrm{Hom}(\mathbf{R}^{n_1},\mathbf{R}^{n_2})=\mathbf{R}^{n_1n_2}$. (For the last one, we use $\mathrm{Hom}((\varphi_{\alpha,b}^1)^{-1},\varphi_{\alpha,b}^2)$ instead of $\mathrm{Hom}(\varphi_{\alpha,b}^1,\varphi_{\alpha,b}^2)$.) $\mathrm{Hom}(\xi,\varepsilon^1)=\xi^*$ is called the **dual (vector) bundle** of ξ, where ε^1 is the trivial line bundle. If we use coordinate transformations, $\oplus$, $\otimes$, $\bigwedge^p$, and ξ^* are obtained by the direct sum, †Kronecker product, matrix of †p-minors, and †transpose of matrices, respectively. If ξ_2 is a subbundle of ξ_1, the **quotient bundle** ξ_1/ξ_2 of dimension n_1-n_2 is defined by using the quotient vector space $\mathbf{R}^{n_1}/\mathbf{R}^{n_2}=\mathbf{R}^{n_1-n_2}$, and $\xi_2\oplus(\xi_1/\xi_2)$ is equivalent to ξ_1. These operations preserve the equivalence relation of bundles. Also, $\oplus$ and $\otimes$ are commutative up to equivalence and satisfy the associative and distributive laws. For each ξ having a finite-dimensional †CW complex as base space, there is a ξ' such that $\xi\oplus\xi'$ is trivial.

Using the complex number field $\mathbf{C}$ or the quaternion field $\mathbf{H}$ instead of the real number field $\mathbf{R}$, we can define similarly the **complex vector bundle** or the **quaternion vector bundle** and the operations $\oplus$, $\otimes$, etc. For a complex vector bundle ξ, the **complex conjugate bundle** $\bar{\xi}$ is defined by the complex conjugate of matrices.

(5) Tangent bundles, tensor bundles. Let M be an n-dimensional †differentiable manifold of class C^r. Consider the †tangent vector space $T_p(M)$ at $p\in M$, set $T(M)=\bigcup_{p\in M}T_p(M)$, and define $\pi: T(M)\rightarrow M$ by $\pi(T_p(M))=p$. For a †coordinate neighborhood U_p of p with local coordinate system $(x_1,\ldots,x_n)$, each point of $\pi^{-1}(U_p)$ is represented by $\sum_{i=1}^n f_i\partial/\partial x_i$, and $\pi^{-1}(U_p)$ has a coordinate system $(x_1,\ldots,x_n, f_1,\ldots,f_n)$. Hence $T(M)$ is a C^{r-1}-manifold, and $\mathfrak{T}(M)=(T(M),\pi,M,\mathbf{R}^n,GL(n,\mathbf{R}))$ is an n-dimensional real vector bundle. $\mathfrak{T}(M)$ is called the **tangent (vector) bundle**, its dual bundle $\mathfrak{T}^*(M)$ the **cotangent (vector) bundle**, and the tensor product $\mathfrak{T}(M)\otimes\ldots\otimes\mathfrak{T}^*(M)\otimes\ldots$ a **tensor bundle** of M. The line bundle $\bigwedge^n\mathfrak{T}^*(M)$ is called the **canonical bundle** of M.

For a †complex manifold M, $T(M)$ is a complex manifold and $\mathfrak{T}(M)$ is a complex vector bundle. Therefore these bundles are defined as complex bundles.

(6) Tangent r-frame bundle. In the preceding example, the space of all †tangent r-frames of M is a bundle space with base space M and

group $GL(n, \mathbf{R})$. It is called the **tangent r-frame bundle** (or the bundle of tangent r-frames) of M.

G. The Classification Problem

For a fiber bundle $\xi=(E,p,B,F,G)$ and a continuous mapping $\psi: B' \to B$, consider the subspace $E'=\{(x,b')\in E\times B' \mid p(x)=\psi(b')\}$ of $E\times B'$ and the projections $p': E'\to B'$ and $\Psi: E'\to E$. Then $\psi^{\#}\xi=(E',p',B',F,G)$ is a fiber bundle, and Ψ is a bundle mapping from $\psi^{\#}\xi$ to ξ; $\psi^{\#}\xi$ is called the **induced bundle** of ξ by ψ. Let $\{U_\alpha\}$ and $\{g_{\beta\alpha}\}$ be the systems of coordinate neighborhoods and transformations of ξ. Then $\{\psi^{-1}(U_\alpha)\}$ and $\{g_{\beta\alpha}\circ\psi\}$ are corresponding systems of $\psi^{\#}\xi$. If $\Psi: E'\to E$ is a bundle mapping from ξ' to ξ having $\psi: B'\to B$ as the mapping of base spaces, then $\xi'\equiv\psi^{\#}\xi$. Also, we have $\psi^{\#}\xi_1\equiv\psi^{\#}\xi_2$ if $\xi_1\equiv\xi_2$; $(\psi\circ\psi')^{\#}\xi\equiv\psi'^{\#}(\psi^{\#}\xi)$. If ξ is a principal bundle, then $\psi^{\#}\xi$ is also principal, and $\psi^{\#}(\eta\times_G F)\equiv(\psi^{\#}\eta)\times_G F$. For a †paracompact space B', $\psi_1^{\#}\xi\equiv\psi_2^{\#}\xi$ if $\psi_1, \psi_2: B'\to B$ are †homotopic.

For a topological group G, a principal bundle $\xi(n,G)=(E(n,G),p,B(n,G),G)$ is called an **n-universal bundle** if $E(n,G)$ is †n-connected ($n\leqslant\infty$); its base space $B(n,G)$ is called an **n-classifying space** of G. In particular, $\xi(\infty,G)=\xi_G=(E_G,p,B_G,G)$ is called simply a **universal bundle** and B_G a **classifying space** of G. Then we have the **classification theorem**: Let B be a CW complex with $\dim B\leqslant n$; then the set of equivalence classes of principal G-bundles with base space B is in one-to-one correspondence with the †homotopy set $\pi(B;B(n,G))$ of continuous mappings of B into $B(n,G)$. Such a correspondence is given by associating with the induced bundle $\psi^{\#}\xi(n,G)$ a continuous mapping $\psi: B\to B(n,G)$, called the **characteristic mapping or classifying mapping (characteristic map** or **classifying map**) of $\psi^{\#}\xi(n,G)$. Furthermore, if G is an effective left topological transformation group of F, the set of equivalence classes of G-bundles with base space B and fiber F is in one-to-one correspondence with $\pi(B;B(n,G))$. The correspondence is given by associating $\psi^{\#}(\xi(n,G)\times_G F)$ to ψ.

H. Construction of Universal Bundles

For an arbitrary topological group G, J. W. Milnor [6] constructed a universal bundle (E_G, p, B_G, G) in the following manner. The **join** $E_G=G\circ\ldots\circ G\circ\ldots$ of countably infinite copies of G is defined as follows: A point e of E_G is the symbol $t_{i_1}g_{i_1}\oplus\ldots\oplus t_{i_m}g_{i_m}$ $(1\leqslant i_1< i_2<\ldots<i_m, m=1,2,3,\ldots)$, where $t_{i_1},\ldots,t_{i_m}$ are real numbers satisfying $t_{i_k}\geqslant 0$, $t_{i_1}+\ldots+t_{i_m}=1$, and g_{i_k} is an element of the i_kth copy of G, $k=1,\ldots,m$. Here we can omit $t_{i_k}g_{i_k}$ if $t_{i_k}=0$. Regard E_G as a topological space with a weak topology such that the coordinate functions $e\mapsto t_i$, $e\mapsto g_i$ are continuous. Define the right action of G on E_G by $(t_{i_1}g_{i_1}\oplus\ldots\oplus t_{i_m}g_{i_m})\cdot g=t_{i_1}(g_{i_1}\cdot g)\oplus\ldots\oplus t_{i_m}(g_{i_m}\cdot g)$. Let B_G and $p: E_G\to B_G$ be the identification space of E_G by this action of G and the identification mapping. Then (E_G,p,B_G,G) is a **universal bundle** for G, and B_G is called the **classifying space** for G. E_G and B_G are sometimes written as EG and BG. In particular, a classifying space B_G is a countable CW complex for any countable CW group G (i.e., a topological group that is a countable CW complex such that the mapping $g\to g^{-1}$ of G into G and the product mapping $G\times G\to G$ are both cellular). The following examples of classifying spaces for Lie groups are also useful. Note that every CW complex B_G of a given G has the same †homotopy type.

I. Examples of Universal Bundles

(1) G is either $O(n)$, $U(n)$, or $Sp(n)$: Let Λ and λ be as in (3) of Section E. According as Λ is $\mathbf{R}$, $\mathbf{C}$, or $\mathbf{H}$, we let $U(n,\Lambda)$ be the †orthogonal group $O(n)$, the †unitary group $U(n)$, or the †symplectic group $Sp(n)$. Then the †Stiefel manifold $V_{m+n,m}(\Lambda)=U(m+n,\Lambda)/I_m\times U(n,\Lambda)$ (I_m is the unit element of $U(m,\Lambda)$) is $(\lambda(n+1)-2)$-connected. Hence the principal bundle $\xi(\lambda(n+1)-2, U(m+n,\Lambda))=(V_{m+n,m}(\Lambda), M_{m+n,m}(\Lambda), U(m,\Lambda))$ from (4) of Section E is a $(\lambda(n+1)-2)$-universal bundle of $U(m,\Lambda)$, where the base space $M_{m+n,m}(\Lambda)=U(m+n,\Lambda)/U(m,\Lambda)\times U(n,\Lambda)$ is the †Grassmann manifold.

(2) G is either $O(\infty)$, $U(\infty)$, or $Sp(\infty)$. The examples in (1) are valid for $m, n=\infty$. Consider the †inductive limit group $U(\infty,\Lambda)=\bigcup_n U(n,\Lambda)$ under the natural inclusion $U(n,\Lambda)\subset U(n+1,\Lambda)$, and supply the **infinite classical group** $U(\infty,\Lambda)$ with the weak topology (this means that a set O of $U(\infty,\Lambda)$ is open if and only if each $O\cap U(n,\Lambda)$ is open in $U(n,\Lambda)$). Then the **infinite Stiefel manifold** $V_{m+n,m}(\Lambda)$ and the **infinite Grassmann manifold** $M_{m+n,m}(\Lambda)$ ($m=\infty$ or $n=\infty$) are defined as before, and we have

$$M_{\infty,m}(\Lambda)=\bigcup_n M_{m+n,m}(\Lambda),$$

and so on. Furthermore, these manifolds are CW complexes, and $V_{\infty,m}(\Lambda)$ ($m\leqslant\infty$) is ∞-connected. Although $U(\infty,\Lambda)$ is not a Lie group, $U(m,\Lambda)\times U(n,\Lambda)$ has a local cross section in $U(m+n,\Lambda)$ for $m, n\leqslant\infty$ [7]. Therefore, setting $n=\infty$ in (1), $\xi(\infty, U(m,\Lambda))$ is a universal bundle of $U(m,\Lambda)$, and the infinite Grassmann manifold $M_{\infty,m}(\Lambda)$ is a classifying

space $B_{U(m,\Lambda)}$. Also, $\xi(\lambda(n+1)-2, U(\infty,\Lambda))$ in (1) is a $(\lambda(n+1)-2)$-universal bundle of $U(\infty,\Lambda)$ $(n\leqslant\infty)$.

(3) G is either $SO(m)$ or a general Lie group. For the †rotation group $SO(m)$, we have $\xi(n-1, SO(n))=(V_{m+n,m}(\mathbf{R}), p, \tilde{M}_{m+n,m}, SO(m))$ and $B_{SO(m)}=\tilde{M}_{\infty,m}$, where $\tilde{M}_{m+n,m}=SO(m+n)/SO(m)\times SO(n)$ is the oriented Grassmann manifold. For any compact Lie group G, we have $\xi(n-1,G)=(V_{m+n,m}(\mathbf{R}), p, O(m+n)/G\times O(n), G)$, where $G\subset O(m)$. For any connected Lie group G, we have $\xi(n,G)=\xi(n,G_1)\times_{G_1}G$, where G_1 is the maximum compact subgroup of G (since G/G_1 is homeomorphic to a Euclidean space, $\xi(n,G)$ reduces to $\xi(n,G_1)$).

J. Reduction of Fiber Bundles

Let G be a topological group and H its closed subgroup. We say that the structure group of a G-bundle ξ is **reduced** to H if ξ is equivalent to a G-bundle whose coordinate transformations take values in H. For a principal H-bundle $\eta_0=(P,q,B,H)$, the associated H-bundle $\eta_0\times_H G=(P\times_H G, p, B, G)$ with fiber G is defined, where H operates on G by the product of G; it is also a principal G-bundle if we define an operation of G on $P\times_H G$ by $\{(x,g)\}\cdot g'=\{(x,gg')\}$. For a principal G-bundle η, we say that η is **reducible** to an H-bundle if there is a principal H-bundle η_0 with $\eta=\eta_0\times_H G$, and we call η_0 a **reduced bundle** of η. It is easy to see that the group of a G-bundle ξ is reducible to H if and only if the associated principal G-bundle of ξ is reducible to H. Also, if η_0 is a reduced bundle of η, then $\psi^{\#}\eta_0$ is a reduced bundle of $\psi^{\#}\eta$.

Now, assume that H has a local cross section in G and G/H is ∞-connected. Then for an n-universal bundle $\xi(n,H)$ of H, $\xi(n,H)\times_H G$ is an n-universal bundle of G (n-connectedness of $E(n,H)\times_H G$ is shown by the †homotopy exact sequence of †fiber spaces). Therefore, by the classification theorem, the group of any G-bundle is reducible to H, and the equivalence classes of G-bundles are in one-to-one correspondence with those of H-bundles.

(1) A G-bundle is trivial if and only if its group is reducible to e (identity element). A $2n$-dimensional differentiable manifold M of class C^∞ has an †almost complex structure if and only if the group of the tangent bundle $\mathfrak{T}(M)$ is reducible to $GL(n,\mathbf{C})$, i.e., $\mathfrak{T}(M)$ is considered as an n-dimensional complex vector bundle.

(2) Since $GL(n,\mathbf{R})\approx O(n)\times\mathbf{R}^{n(n+1)/2}$ and $GL(n,\mathbf{C})\approx U(n)\times\mathbf{R}^{n^2}$, n-dimensional real (complex) vector bundles can be considered as $O(n)$ $(U(n))$-bundles with fiber $\mathbf{R}^n(\mathbf{C}^n)$.

K. Homotopy and Homology Theory of Bundles

Since a fiber bundle is a †locally trivial fiber space, the exact sequence and the spectral sequence of fiber spaces (→ 148 Fiber Spaces) are applicable to fiber bundles. For example, the cohomology structures of homogeneous spaces of classical groups have been determined by A. Borel, J.-P. Serre, and others.

(1) Characteristic class. For a classifying space B_G of a topological group G, we have an isomorphism $\pi_n(B_G)\cong\pi_{n-1}(G)$ of homotopy groups and the following classification theorem of fiber bundles over the n-sphere S^n. The set of the equivalence classes of principal G-bundles or G-bundles with fiber F over the base space S^n is in one-to-one correspondence with the set $\pi_{n-1}(G)/\pi_0(G)$ of equivalence classes under the operation of G on $\pi_{n-1}(G)$ given by the inner automorphisms of G; such a correspondence is given by associating with each principal G-bundle $\eta=(P,q,S^n,G)$ the class (called the **characteristic class** of η) containing the image $\Delta(\iota_n)$ of a generator $\iota_n\in\pi_n(S^n)$ by the homomorphism $\Delta:\pi_n(S^n)\cong\pi_n(P,G)\to\pi_{n-1}(G)$. Take U_1 and U_2 (the open sets of S^n such that the last coordinates t_{n+1} are $>-1/2$ and $<1/2$, respectively) as coordinate neighborhoods of η. Then the restriction $T=g_{12}|S^{n-1}$ represents the characteristic class of η, where $g_{12}:U_1\cap U_2\to G$ is the coordinate transformation and S^{n-1} is the equator of S^n.

(2) For the principal bundle $\eta=(SO(n+1), q, S^n, SO(n))$, the mapping $T:S^{n-1}\to SO(n)$ is given by

$$T(t_1,\ldots,t_n)=(I_n-2(t_it_j))\begin{pmatrix} I_{n-1} & 0 \\ 0 & -1\end{pmatrix}$$

(I_n is the unit matrix of degree n). Hence the †mapping degree of the composite $q'\circ T:S^{n-1}\to S^{n-1}$ (of T and the natural projection q': $SO(n)\to S^{n-1}$) is equal to 0 if n is odd and 2 if n is even. From this fact and the homotopy exact sequence, we have

$$\pi_n(V_{m+n,m}(\mathbf{R}))=\begin{cases}\mathbf{Z} & \text{if } m=1 \text{ or } n \text{ is even}\\ \mathbf{Z}_2=\mathbf{Z}/2\mathbf{Z} & \text{if } m>1 \text{ and } n \text{ is odd}\end{cases}$$

for the real Stiefel manifold $V_{m+n,m}(\mathbf{R})$, which is $(n-1)$-connected.

(3) Sphere bundles. An $O(n+1)$-bundle with fiber S^n is called an **n-sphere bundle**. The set of equivalence classes of n-sphere bundles with base space S^m is in one-to-one correspondence with $\pi_{m-1}(O(n+1))/\pi_0(O(n+1))$. For example, any 1-sphere bundle over S^m $(m\geqslant 3)$ and any n-sphere bundle over S^3 is trivial. Every 3-sphere bundle over S^4 is equivalent to one of $\{\xi_{m,n}\mid m$ an integer, n a positive integer$\}$, where

$\xi_{m,n}$ is defined as follows: Let $\rho, \sigma: S^3 \to O(4)$ be defined by $\rho(q)q' = qq'q^{-1}$, $\sigma(q)q' = qq'$ (q, q' are †quaternions of norm 1). Then these mappings represent generators of $\pi_3(O(4)) \cong \pi_3(S^3 \times S^3) \cong \mathbf{Z} + \mathbf{Z}$, and $\xi_{m,n}$ is the 3-sphere bundle over S^4 corresponding to the element $m\{\rho\} + n\{\sigma\} \in \pi_3(O(4))$ (i.e., to the mapping $f_{m,n}: S^3 \to O(4)$ defined by $f_{m,n}(q)(q') = q^{m+n}q'q^{-m}$). (Here we use the fact that the operation of the element $r \in O(4)$ $(r(q) = q^{-1})$ is given by $r\rho r^{-1} = \rho$, $r\sigma r^{-1} = \rho\sigma^{-1}$.)

L. Cross Sections

For a fiber bundle $\xi = (E, p, B, F, G)$, a cross section $f: B_0 \to E$ over a subspace B_0 $(\subset B)$ is a continuous mapping such that $p \circ f$ is the identity mapping of B_0; a cross section over B is called a **cross section** of ξ. A bundle ξ is trivial if and only if the associated principal bundle of ξ has a cross section. More generally, given a principal bundle $\eta = (P, q, B, G)$ and a closed subgroup H having a local cross section in G, η is reducible to H if and only if the associated bundle $\eta(\times_G(G/H) = (P/H, q', B)$, $G/H)$ with fiber G/H has a cross section.

Suppose that the base space B of the fiber bundle $\xi = (E, p, B, F, G)$ is a †polyhedron. We denote the †r-skeleton of B by B^r and consider the problem of extending cross sections $f_r: B^r \to E$ successively for $r = 0, 1, \ldots$. Clearly, there is a cross section f_0. For each r-simplex σ of B, we have $(p^{-1}(\sigma), p, \sigma, F) \equiv (\sigma \times F, p_1, \sigma, F)$ since σ is †contractible. Hence there is a bundle mapping $\varphi_\sigma: \sigma \times F \approx p^{-1}(\sigma)$ with $p \circ \varphi_\sigma = p_1$. Assume the existence of a cross section f_{r-1}: $B^{r-1} \to E$, and consider the mapping $h_{\dot\sigma} = p_2 \circ \varphi_\sigma^{-1} \circ (f_{r-1} | \dot\sigma): \dot\sigma \to F$ ($p_2: \sigma \times F \to F$ is the projection and $\dot\sigma$ is the boundary of σ). Then if $h_{\dot\sigma}$ is extensible to $h_\sigma: \sigma \to F$, an extended cross section $f_\sigma: \sigma \to E$ of $f_{r-1} | \dot\sigma$ is defined by $f_\sigma(b) = \varphi_\sigma(b, h_\sigma(b))$ $(b \in \sigma)$, and the extension $f_r: B^r \to E$ of f_{r-1} is defined by $f_r | \sigma = f_\sigma$, $f_r | B^{r-1} = f_{r-1}$. If $\pi_{r-1}(F) = 0$, for example, there is an extension h_σ of $h_{\dot\sigma}$ since $(\sigma, \dot\sigma) \approx (V^r, S^{r-1})$, and f_{r-1} is extensible to a cross section f_r.

Now assume that the base space B of a G-bundle $\xi = (E, p, B, F, G)$ is an †arcwise connected polyhedron and F is †$(n-1)$-connected. Then there is a cross section $f: B^n \to E$ constructed by the stepwise method of the previous paragraph. But if $\pi_n(F) \neq 0$, we have an obstruction to extending f over B^{n+1}. Now we explain how to measure this obstruction. Suppose that F is †n-simple. Then for each $(n+1)$-simplex σ of B, the mapping $h_{\dot\sigma}: \dot\sigma \to F$, defined by f as in the previous paragraph, determines a unique element $c(f)(\sigma)$ of the homotopy group $\pi_n(F)$. Hence we have a †cochain $c(f) \in C^{n+1}(B; \pi_n(F))$, and f is extensible to a cross section over B^{n+1} if and only if $c(f) = 0$. Thus there is a cross section over B^{n+1} if and only if the set of $c(f)$ for every cross section f over B^n contains the cochain 0; $\{c(f)\}$ is considered as a measure of the obstruction.

Let $w: I \to B$ be a †path. We consider the space $I \times F$ and the canonical projection p_1: $I \times F \to I$. Then there is a bundle mapping Ω: $I \times F \to E$ with $p \circ \Omega = w \circ p_1$, since $w^{\#}\xi \equiv (I \times F, p_1, I, F)$; and a homeomorphism $w_{\#}: F \approx F$ is defined by $w_{\#} = \varphi_{0,b_0}^{-1} \circ \Omega_0 \circ \Omega_1^{-1} \circ \varphi_{1,b_1}$, where $b_\varepsilon = w(\varepsilon)$ $(\varepsilon = 0, 1)$, $\varphi_\varepsilon: U_\varepsilon \times F \approx p^{-1}(U_\varepsilon)$ is a coordinate function of a coordinate neighborhood $U_\varepsilon \ni b_\varepsilon$, and $\Omega_\varepsilon(= \Omega | \varepsilon \times F): F \approx p^{-1}(b_\varepsilon)$. The homeomorphism $w_{\#}$ induces an isomorphism $w_{\#}: \pi_n(F) \cong \pi_n(F)$, and $\pi_n(F)$ forms a †local coefficient on B. Then the cochain $c(f)$ is a †cocycle with the local coefficient $\pi_n(F)$, called the **obstruction cocycle** of f. Furthermore, the set $\{c(f)\}$ for every cross section $f: B^n \to E$ is a cohomology class $c^{n+1}(\xi) \in H^{n+1}(B; \pi_n(F))$ (local coefficient), and $c^{n+1}(\xi)$ is called the **primary obstruction** to the construction of a cross section. There is a cross section over the $(n+1)$-skeleton B^{n+1} if and only if $c^{n+1}(\xi) = 0$. The local coefficient $\pi_n(F)$ is trivial if B is †simply connected or, for example, if the structure group G of ξ is connected (in which case ξ is called an **orientable fiber bundle**); when this is true, $c^{n+1}(\xi)$ is an element of $H^{n+1}(B; \pi_n(F))$, where $\pi_n(F)$ is not a local coefficient. Furthermore, if $c^{n+1}(\xi) = 0$ and $\pi_i(F) = 0$ $(n < i < m)$, then the secondary obstruction $\mathbf{O}^{m+1}(\xi) \subset H^{m+1}(B; \pi_m(F))$ is defined similarly (→ 305 Obstructions).

M. Stiefel-Whitney Classes

Let $\xi = (E, p, B, F, O(n))$ be an $O(n)$-bundle over an arcwise connected polyhedron B and let ξ^0 be its associated principal bundle. Consider the Stiefel manifold $V_{n,n-k} = V_{n,n-k}(\mathbf{R}) = O(n)/I_{n-k} \times O(k)$, which is $(k-1)$-connected, and the associated bundle $\xi^k = \xi^0 \times_{O(n)} V_{n,n-k}$ with fiber $V_{n,n-k}$. The primary obstruction $W_{k+1}(\xi) = c^{k+1}(\xi^k) \in H^{k+1}(B; \pi_k(V_{n,n-k}))$ $(k = 0, 1, \ldots, n-1)$ is called the **Stiefel-Whitney class** of ξ. We have $2W_{k+1}(\xi) = 0$ unless $k = n-1$ and k is odd. Hence we usually consider $W_{k+1}(\xi) \in H^{k+1}(B; \mathbf{Z}_2)$. ξ is orientable, i.e., the group of ξ is reducible to $SO(n)$, if and only if $W_1(\xi) = 0$. The **Stiefel-Whitney classes of an n-dimensional †differentiable manifold** M are defined to be those of the tangent bundle $\mathfrak{T}(M)$. Since the orientability of M coincides with that of $\mathfrak{T}(M)$, M is orientable if and only if $W_1(M) = 0$. The condition $W_{k+1}(M) = 0$ is necessary for the existence of a continuous field of orthonormal tangent $(n-k)$-frames over M (if $k = n-1$ this condition is also suffi-

cient). Also, when M is closed, $W_n(M)$ is equal to $\chi(M)\mu$, where μ is the fundamental cohomology class of M and $\chi(M)$ is the †Euler characteristic of M (→ 56 Characteristic Classes B).

N. Chern Classes

For a $U(n)$-bundle $\xi=(E,p,B,F,U(n))$, the primary obstruction $C_{k+1}(\xi)=c^{2k+2}(\xi^k)\in H^{2k+2}(B;\mathbf{Z})$ $(k=0,1,\ldots,n-1)$ of the associated bundle $\xi^k=\xi^0\times_{U(n)}V_{n,n-k}(\mathbf{C})$ is called the **Chern class** of ξ. If we consider ξ as an $O(2n)$-bundle by $U(n)\subset O(2n)$, then $W_{2k+1}(\xi)=0$ and $W_{2k}(\xi)=C_k(\xi)\ (\mathrm{mod}\ 2)$. The **Chern classes** of a real $2n$-dimensional almost complex manifold are defined to be those of the tangent bundle $\mathfrak{T}(M)$ (→ 56 Characteristic Classes C).

O. Bundles of Class C^r, Analytic Bundles

A fiber bundle $\xi=(E,p,B,F,G)$ is called a **fiber bundle of class C^r** $(r=0,1,\ldots\infty,\omega)$ if E, B, F are differentiable manifolds of class C^r, G is a Lie group and a transformation group of F of class C^r, and p and the coordinate functions are differentiable mappings of class C^r. Bundles of class C^0 are usual G-bundles, and those of class C^ω are **real analytic fiber bundles**. Similarly, **complex analytic fiber bundles** are defined by the notions of complex manifolds, †complex Lie groups, and †holomorphic mappings. For example, the universal bundles $\xi(n-1,O(m))$ and $\eta(2n,U(m))$ are real and complex analytic principal bundles, respectively, and the tangent bundle $\mathfrak{T}(M)$ of a C^{r+1} (or complex) manifold is a C^r (or complex analytic) vector bundle. The operations of the Whitney sum, etc., are defined analogously for these vector bundles.

The equivalence of C^r (complex analytic) bundles is defined by means of bundle mappings that are C^r-differentiable (holomorphic). Bundles of class C^r $(r\leqslant\infty)$ are classified by C^r mappings into a classifying space in the same manner as for bundles of class C^0. Also, the connection of class C^r (→ 80 Connections) in C^r bundles is an important notion.

For complex analytic bundles, a similar classification has been obtained for restricted spaces by K. Kodaira, Serre, S. Nakano [8], and others. The classification of complex analytic bundles over a †Stein manifold is reduced to that of bundles of class C^0 (**Oka's principle** [9]), and similar results are valid for C^ω-manifolds. The complex analytic (or holomorphic) connection does not necessarily exist, and M. F. Atiyah [10] found the condition for its existence and its relation with Chern classes.

P. Microbundles

A system $x:B\xrightarrow{i}E\xrightarrow{j}B$ of topological spaces E, B and continuous mappings i,j is called an n-dimensional **microbundle** over B if for each $b\in B$, there exist a neighborhood U of b, a neighborhood V of $i(U)$, and a homeomorphism $h:V\approx U\times\mathbf{R}^n$ with $h\circ i|U=i_1$, $j|V=p_1\circ h$ $(i_1:U\approx U\times 0\subset U\times\mathbf{R}^n$, and $p_1:U\times\mathbf{R}^n\to U$ is the projection). Let $H_0(n)$ be the topological group of all homeomorphisms of $\mathbf{R}^n$ onto itself fixing the origin with compact-open topology. Then the equivalence classes of n-dimensional microbundles over B are naturally in one-to-one correspondence with the equivalence classes of $H_0(n)$-bundes with base space B and fiber $\mathbf{R}^n$ [13].

In the category of polyhedra and PL (†piecewise linear) mappings, the notion of a **PL microbundle** can be defined in the same manner. The structural group of n-dimensional PL microbundles is defined, in a generalized sense, as a complete †semi-simplicial complex [11]. The **tangent PL microbundle** is defined for any †topological (PL) manifold. J. Milnor classified †smoothings of PL manifolds by means of PL microbundles [11] and then showed that the †tangent vector bundles and its †Pontryagin classes of smooth manifolds are not topologically invariant [12].

For a †PL embedding $f:M\to N$ between PL manifolds, if there is a neighborhood E of $f(M)$ in N and a PL mapping $p:E\to M$ so that a diagram $v:M\xrightarrow{f}E\xrightarrow{p}M$ is a PL microbundle, then v is called a **normal PL microbundle** of f. In this case, f is †locally flat.

There is a locally flat PL embedding between PL manifolds which admits no normal PL microbundle [14].

Q. Block Bundles

As the normal bundle theory for locally flat PL embeddings of PL manifolds, the concept of block bundle was introduced independently by C. P. Rourke and B. J. Sanderson [15], M. Kato [16], and C. Morlet [17]. Let E be a †polyhedron, and let K be a cell complex. A set $\{E_\sigma\mid\sigma\in K\}$ of PL balls in E indexed on K is called a **q-block structure** of E if the following three conditions are satisfied: (1) $\bigcup_{\sigma\in K}E_\sigma=E$; (2) for each $\sigma\in K$, there is a †PL homeomorphism $h_\sigma:\sigma\times I^q\to E_\sigma$ such that $h_\sigma(\tau\times I^q)=E_\tau$ for each face τ of σ, where $I=[-1,1]$; and (3) if $E_\sigma\cap E_\rho\neq\varnothing$, then $E_\sigma\cap E_\rho=E_\tau$, where $\tau=\sigma\cap\rho$.

For $\sigma\in K$, E_σ is called the **block** over σ, and h_σ in (2) is called a **trivialization** of E_σ. Then a triple $(E,K,\{E_\sigma\mid\sigma\in K\})$ is referred to as a q-**block bundle** over K and is denoted by ξ/K. Another block bundle $\xi'/K=(E',K,\{E'_\sigma\mid\sigma\in$

$K\})$ over the same complex K is said to be **isomorphic** with ξ/K if there is a PL homeomorphism $g: E \to E'$, called an **isomorphism**, such that $g(E_\sigma) = E'_\sigma$ $(\sigma \in K)$. A PL embedding $i: |K| \to E$ is a **zero-section** of ξ/K if for each $\sigma \in K$, there is a trivialization $h_\sigma: \sigma \times I^q \to E_\sigma$ such that $h_\sigma(x, 0) = i(x)$ $(x \in \sigma)$. In this case we say that ξ/K is a block bundle with a zero-section $i: |K| \to E$. There is a unique zero-section of ξ/K up to isomorphism of ξ/K onto itself. For every †locally flat PL embedding between PL manifolds M and W of codimension q and for any cell division K of M, a †derived neighborhood N of $f(M)$ in W admits a unique q-block bundle $\nu/K = (N, K, \{N_\sigma \mid \sigma \in K\})$ with $f: M \to N$ as a zero-section up to isomorphism respecting the zero-section [15]. The block bundle ν/K is called a **normal block bundle** of $f: M \to W$.

References

[1] N. E. Steenrod, The topology of fiber bundles, Princeton Univ. Press, 1951.
[2] E. Stiefel, Richtungsfelder und Fernparallelismus in n-dimensionalen Mannigfaltigkeiten, Comment. Math. Helv., 8 (1936), 3–51.
[3] H. Whitney, Topological properties of differentiable manifolds, Bull. Amer. Math. Soc., 43 (1937), 785–805.
[4] S. S. Chern, Some new view-points in differential geometry in the large, Bull. Amer. Math. Soc., 52 (1946), 1–30.
[5] F. Hirzebruch, Topological methods in algebraic geometry, Springer, third edition, 1966.
[6] J. W. Milnor, Construction of universal bundles II, Ann. Math., (2) 63 (1956), 430–436.
[7] J. C. Moore, Espaces classfiants, Sém. H. Cartan, 12, no. 5 (1959–1960), Ecole Norm. Sup., Secrétariat Mathématique.
[8] S. Nakano, On complex analytic vector bundles, J. Math. Soc. Japan, 7 (1955), 1–12.
[9] H. Grauert, Analytische Faserungen über holomorph-vollständigen Räumen, Math. Ann., 135 (1958), 263–273.
[10] M. F. Atiyah, Complex analytic connections in fiber bundles, Trans. Amer. Math. Soc., 85 (1957), 181–207.
[11] J. W. Milnor, Microbundles and differentiable structures, Lecture notes, Princeton Univ., 1961.
[12] J. W. Milnor, Microbundles I, Topology, 3 (1964), Suppl. 1, 53–80.
[13] J. Kister, Microbundles are fiber bundles, Bull. Amer. Math. Soc., 69 (1963), 854–857.
[14] C. P. Rourke and B. J. Sanderson, An imbedding without a normal bundle, Inventiones Math., 3 (1967), 293–299.
[15] C. P. Rourke and B. J. Sanderson, Block-bundles I–III, Ann. Math., (2) 87 (1968).
[16] M. Kato, Combinatorial prebundles I, II, Osaka J. Math., 4 (1967).
[17] C. Morlet, Les méthodes de la topologie différentiélle dans l'étude des variétés semi-lineairs, Ann. Sci. École Norm. Sup., 1 (1968), 313–394.
[18] D. Husemoller, Fiber bundles, McGraw-Hill, 1966.

148 (IX.10) Fiber Spaces

A. General Remarks

J.-P. Serre [1] generalized the concept of fiber bundles to that of fiber spaces by utilizing the covering homotopy property (→ Section B). He applied the theory of †spectral sequences, due to J. Leray, to the (cubic) †singular (co)-homology groups of fiber spaces. These are quite useful for determining (co)homology structures and homotopy groups of topological spaces, and are now of fundamental importance in algebraic topology.

B. Definitions

Let $p: E \to B$ be a continuous mapping between topological spaces, and let X be a topological space. Then we say that p has the **covering homotopy property** with respect to X if for any mapping $f: X \to E$ and †homotopy $g_t: X \to B$ $(0 \leqslant t \leqslant 1)$ with $p \circ f = g_0$, there is a homotopy $f_t: X \to E$ $(0 \leqslant t \leqslant 1)$ with $f_0 = f$ and $p \circ f_t = g_t$. We call (E, p, B) a **fiber space** (or **fibration**) if p has the covering homotopy property with respect to every cube $I^n = \{(x_1, \ldots, x_n) \mid 0 \leqslant x_i \leqslant 1\}$, $n = 0, 1, \ldots$ (then p has the covering homotopy property with respect to every †CW complex). Then E is called the **total space**, p the **projection**, B the **base space**, and $F_b = p^{-1}(b)$ the **fiber** over $b \in B$.

Let E, B, F be topological spaces and $p: E \to B$ a continuous mapping. We call (E, p, B, F) a **locally trivial fiber space** if for each $b \in B$, there exist an open neighborhood U of b and a homeomorphism $\varphi: U \times F \to p^{-1}(U)$ with $p \circ \varphi(b', y) = b'$ $(b' \in U, y \in F)$. In this case, p has the covering homotopy property with respect to each †paracompact space; hence (E, p, B) is a fiber space. A †fiber bundle is clearly a locally trivial fiber space.

C. Path Spaces

Another important example of a fiber space is a path space. A **path** in a topological space X is a continuous mapping $w: I \rightarrow X$ ($I = [0,1]$). Given subsets A_0 and A_1 of X, the **path space** $\Omega(X; A_0, A_1)$ is the space of all paths $w: I \rightarrow X$ satisfying $w(\varepsilon) \in A_\varepsilon$ ($\varepsilon = 0, 1$) topologized by †compact-open topology. Define $p_\varepsilon: \Omega(X; A_0, A_1) \rightarrow A_\varepsilon$ by $p_\varepsilon(w) = w(\varepsilon)$ ($\varepsilon = 0, 1$). Then $(\Omega(X; A_0, A_1), p_\varepsilon, A_\varepsilon)$ is a fiber space; in fact, p_ε has the covering homotopy property with respect to every topological space. In particular, the total space of the fiber space $(\Omega(X; X, *), p_0, X)$ ($* \in X$) is †contractible, and the fiber $p_0^{-1}(*) = \Omega(X; *, *) = \Omega X$ over $*$ is the †loop space of X with base point $*$. For a continuous mapping $f: Y \rightarrow X$, consider the space $E_f = \{(y, w) \in Y \times \Omega(X; X, X) \mid f(y) = w(0)\}$ and the continuous mapping $p: E_f \rightarrow X$ defined by $p(y, w) = w(1)$. Then $Y \subset E_f$, and Y is a †deformation retract of E_f; furthermore, (E_f, p, X) is a fiber space with $f = p \mid Y$.

D. Homotopy Groups of Fiber Spaces

For †homotopy groups of fiber spaces, the **Hurewicz-Steenrod isomorphism theorem** holds: Let (E, p, B) be a fiber space and $F = p^{-1}(*)$ the fiber over the base point $* \in B$. Then $p_*: \pi_n(E, F) \rightarrow \pi_n(B)$ is an isomorphism for $n \geqslant 2$ and a bijection for $n = 1$. By this theorem, we have the **homotopy exact sequence** of a fiber space:

$$\ldots \rightarrow \pi_{n+1}(B) \xrightarrow{\Delta} \pi_n(F) \xrightarrow{i_*} \pi_n(E) \xrightarrow{p_*} \pi_n(B) \rightarrow \ldots.$$

Furthermore, the more general exact sequence

$$\ldots \rightarrow \pi(Z; \Omega B)_0 \rightarrow \pi(Z; F)_0 \xrightarrow{i_*} \pi(Z; E)_0 \xrightarrow{p_*} \pi(Z; B)_0$$

is valid for each CW complex Z, where $\pi(Z; X)_0$ denotes the †homotopy set of mappings from Z to X relative to the base point.

Example (1). A **cross section** of a fiber space (E, p, B) is a continuous mapping $f: B \rightarrow E$ with $p \circ f = 1$. If (E, p, B) has a †cross section or the fiber F is a †retract of E, then $\pi_n(E) \cong \pi_n(B) + \pi_n(F)$. If F is contractible in E, then $\pi_n(B) \cong \pi_n(E) + \pi_{n-1}(F)$ ($n \geqslant 2$).

Example (2). (E, p, B) is called an **n-connective fiber space** if B is †arcwise connected, E is †n-connected, and $p_*: \pi_i(E) \rightarrow \pi_i(B)$ is an isomorphism, for $i > n$. For each arcwise connected space B and integer n, there is such a fiber space.

Example (3). For a CW complex X, there are topological spaces X_n and continuous mappings $f_n: X \rightarrow X_n$, $q_{n+1}: X_{n+1} \rightarrow X_n$ ($n = 0, 1, \ldots$) with the following four properties: (i) X_n ($0 \leqslant n \leqslant m$) is a point if X is m-connected; (ii) $f_{n*}: \pi_i(X) \cong \pi_i(X_n)$ ($i \leqslant n$); (iii) (X_n, q_n, X_{n-1}) is a fiber space, and its fiber is an †Eilenberg-MacLane space $K(\pi_n(X), n)$; (iv) $q_n \circ f_n$ is †homotopic to f_{n-1}. Such a system $\{X_n, f_n, q_n\}$ is called the **Postnikov system** of X and is in a sense considered a decomposition of X into Eilenberg-MacLane spaces.

E. Spectral Sequences of Fiber Spaces

The cohomological properties of fiber spaces are obtained mainly from the following results (which are valid similarly for homology except for properties of products). Assume that the base space B of a given fiber space (E, p, B) is †simply connected and the fiber $F = p^{-1}(*)$ is arcwise connected, and let R be a commutative ring with unit. Then the **spectral sequence** (of singular cohomology) **of the fiber space** (E, p, B) (with coefficients in R) is defined to be a sequence $\{E_r, d_r\}$ satisfying the following properties:

(i) $E_r = \sum_{p,q} E_r^{p,q}$ are bigraded R-modules, and $d_r = \sum_{p,q} d_r^{p,q}$ are R-linear differentials such that $d_r(E_r^{p,q}) \subset E_r^{p+r, q-r+1}$.

(ii) $E_{r+1}^{p,q} = \operatorname{Ker} d_r^{p,q} / \operatorname{Im} d_r^{p-r, q+r-1}$, which means that $E_{r+1} = H(E_r)$.

(iii) $E_r^{p,q} = 0$ for $p < 0$ or $q < 0$, $E_r^{p,q} = E_{r+1}^{p,q} = \ldots = E_\infty^{p,q}$ for $r > \max(p, q+1)$.

(iv) E_r has a product for which $E_r^{p,q} \cdot E_r^{p',q'} \subset E_r^{p+p', q+q'}$ and $d_r(u \cdot v) = (d_r u) \cdot v + (-1)^{p+q} u \cdot d_r v$ ($u \in E_r^{p,q}$). Furthermore, the induced product in $H(E_r)$ coincides with the product in E_{r+1}.

(v) E_∞ is the †bigraded module associated with some filtration of the cohomology module $H^*(E; R)$; that is, $H^n(E; R) = D^{0,n} \supset D^{1,n-1} \supset \ldots \supset D^{n,0} \supset D^{n+1,-1} = 0$ and $E_\infty^{p,q} = D^{p,q} / D^{p+1,q-1}$. Furthermore, the †cup product $\smile$ in $H^*(E; R)$ satisfies $D^{p,q} \smile D^{p',q'} \subset D^{p+p', q+q'}$ and coincides with the given product in E_∞.

(vi) $E_2^{p,q} = H^p(B; H^q(F; R))$, and the product in E_2 coincides with the cup product in $H^*(B; H^*(F; R))$.

(vii) The composition of $H^n(B; R) = E_2^{n,0} \rightarrow E_3^{n,0} \rightarrow \ldots \rightarrow E_{n+1}^{n,0} = E_\infty^{n,0} = D^{n,0} \subset H^n(E; R)$ is equal to p^*, and the composition of $H^n(E; R) = D^{0,n} \rightarrow E_\infty^{0,n} = E_{n+2}^{0,n} \subset \ldots \subset E_3^{0,n} \subset E_2^{0,n} = H^n(F; R)$ is equal to i^* ($i: F \subset E$), where each $\rightarrow$ is the projection onto the quotient module. In the sequence

$$H^{n-1}(F; R) \xrightarrow{\partial^*} H^n(E, F; R) \xleftarrow{p^*} H^n(B; R),$$

we have $\partial^{*-1}(\operatorname{Im} p^*) = E_n^{0,n-1}$, $\operatorname{Coim} p^* = E_n^{n,0}$, and $d_n: E_n^{0,n-1} \rightarrow E_n^{n,0}$ is equal to the **transgression** $\tau^* = p^{*-1} \circ \partial^*: \partial^{*-1}(\operatorname{Im} p^*) \rightarrow \operatorname{Coim} p^*$. Each element of $\partial^{*-1}(\operatorname{Im} p^*)$ is called **transgressive**.

In the following examples, we assume that R is a principal ideal ring.

Example (4). Let k be a commutative

field, and assume that $\dim_k H_*(B;k)$ and $\dim_k H_*(F;k)$ are finite. Then for the †Poincaré polynomial $P_X(t)=\sum_n b_n t^n$, $b_n=\dim_k H_n(X;k)$, we have $P_E(t)=P_B(t)P_F(t)-(1+t)\varphi(t)$, where $\varphi(t)$ is a polynomial with nonnegative coefficients (Leray). In particular, for the †Euler characteristic $\chi(X)=P_X(-1)$, we have $\chi(E)=\chi(B)\chi(F)$. Also, if $i_*:H_n(F;k)\rightarrow H_n(E;k)$ is monomorphic for each $n\geqslant 0$, then $P_E(t)=P_B(t)P_F(t)$.

Example (5). Isomorphism theorem: If $H_n(B;R)=0$ $(0<n<r)$ and $H_n(F;R)=0$ $(0<n<s)$, then $p_*:H_n(E,F;R)\rightarrow H_n(B;R)$ is isomorphic for $0<n<r+s$ and epimorphic for $n=r+s$, and we have the following **homology exact sequence**:

$$\ldots\rightarrow H_n(F;R)\xrightarrow{i_*}H_n(E;R)\xrightarrow{p_*}H_n(B;R)\xrightarrow{\tau_*}H_{n-1}(F;R)\rightarrow\ldots,\quad n<r+s$$

(similarly for the cohomology).

Example (6). Assume that $H^n(F;R)\cong H^n(S^r;R)$ (S^r is the r-sphere, $r\geqslant 1$). Let $\tilde{E}=M_p$ be the †mapping cylinder of $p:E\rightarrow B$ and $\tilde{p}:\tilde{E}\rightarrow B$ be the continuous mapping defined by p. Then the **Thom-Gysin isomorphism** $\tilde{g}$: $H^{n-r-1}(B;R)\cong H^n(\tilde{E},E;R)$ $(n\geqslant 0)$ with $\tilde{g}(\alpha)=\tilde{p}^*(\alpha)\smile\tilde{g}(1)$ holds (→ 114 Differential Topology G). Also, we have the **Gysin exact sequence**:

$$\ldots\rightarrow H^n(B;R)\xrightarrow{p^*}H^n(E;R)\rightarrow H^{n-r}(B;R)\xrightarrow{g}H^{n+1}(B;R)\rightarrow\ldots,$$

where g satisfies $g(\alpha)=\alpha\smile\Omega=\Omega\smile\alpha$ $(\Omega=g(1)\in H^{r+1}(B;R))$. Here Ω is equal to the image of a generator of $H^r(F;R)=R$ by the transgression τ^*, and $2\Omega=0$ if r is even. (These results hold also for $r=0$ and $R=\mathbf{Z}_2$.)

Example (7). Assume that $H^n(B;R)\cong H^n(S^r;R)$ $(r\geqslant 2)$. Then we have $H^{n-r}(F;R)\cong H^n(E,F;R)$ $(n\geqslant 0)$ and the **Wang exact sequence**:

$$\ldots\rightarrow H^n(E;R)\xrightarrow{i^*}H^n(F;R)\xrightarrow{\theta}H^{n-r+1}(F;R)\rightarrow H^{n+1}(E;R)\rightarrow\ldots,$$

where θ satisfies $\theta(\alpha\smile\beta)=\theta(\alpha)\smile\beta+(-1)^{n(r-1)}\alpha\smile\theta(\beta)$ $(\alpha,\beta\in H^n(F;R))$.

Example (8). For a field k of odd characteristic, if $H^n(E;k)=0$ $(n>0)$ and the algebra $H^*(F;k)$ is generated by a finite number of elements of odd degree, then $H^*(F;k)\cong\bigwedge_k(x_1,\ldots,x_l)$ (the †exterior algebra) and $H^*(B;k)\cong k[y_1,\ldots,y_l]$, where $y_i=\tau^*(x_i)$ (A. Borel).

References

[1] J.-P. Serre, Homologie singulière des espaces fibrés, Ann. Math., (2) 54 (1951), 425–505.

[2] S. T. Hu, Homotopy theory, Academic Press, 1959.

[3] A. Borel, Sur la cohomologie des espaces fibrés principaux et des espaces homogènes de groupes de Lie compacts, Ann. Math., (2) 57 (1953), 115–207.

[4] E. H. Spanier, Algebraic topology, McGraw-Hill, 1966.

[5] G. W. Whitehead, Elements of homotopy theory, Springer, 1978.

149 (III.6) Fields

A. Definition

A set K having at least two elements is called a **field** if two operations, called †addition $(+)$ and †multiplication $(\cdot)$, are defined in K and satisfy the following three axioms.

(1) For any two elements a, b of K, the sum $a+b$ is defined; the associative law $(a+b)+c=a+(b+c)$ and the commutative law $a+b=b+a$ hold; and there exists for arbitrary a, b a unique element x such that $a+x=b$; that is, K is an †Abelian group with respect to the addition (the †identity element of this group is denoted by 0 and is called the **zero element** of K).

(2) For any two elements a, b of K, the product ab $(=a\cdot b)$ is defined; the associative law $(ab)c=a(bc)$ and the commutative law $ab=ba$ hold; and there exists for arbitrary a, b with $a\neq 0$ a unique element x such that $ax=b$, that is, the set K^* of all nonzero elements of K is an Abelian group with respect to multiplication. K^* is called the **multiplicative group** of K, while the identity element of K^* is denoted by 1 and is called the **unity element** (**unit element** or **identity element**) of K.

(3) The distributive law $a(b+c)=ab+ac$ holds. In other words, a field is a †cummutative ring whose nonzero elements form a group with respect to multiplication.

A noncommutative ring whose nonzero elements form a group is called a **noncommutative field** (**skew field** or **s-field**). It should be noted that sometimes a field is defined as a ring whose nonzero elements form a group without assuming the commutativity of that group, and in this case our "field" defined before is called a commutative field. (The term "skew field" is sometimes used to mean either a commutative or a noncommutative field.) In this article we limit ourselves to commutative fields (for noncommutative fields → 29 Associative Algebras).

B. General Properties

Since a field K is a commutative ring, we have properties such as $a0 = 0a = 0$, $(-a)b = a(-b) = -ab$ for elements a, b in K. If a †subring k of K is a field, we say that k is a **subfield** of K or K is an **overfield** (**extension field** or simply **extension**) of k. If a field K has no subfield other than K, K is called a **prime field**.

A map f of a field K into another field K' is called a (field) **homomorphism** if it is a ring homomorphism, i.e., if it satisfies $f(a+b) = f(a)+f(b)$, $f(ab) = f(a)f(b)$. Since a field is †simple as a ring, every (field) homomorphism is an injection unless it maps everything to zero. A homomorphism of K into K' is called an **isomorphism** if it is a bijection, and K and K' are called isomorphic if there exists an isomorphism of K onto K'. An isomorphism of K onto itself is called an **automorphism** of K.

If there is a natural number n such that the sum $n1 = \overbrace{1+\ldots+1}^{n}$ of the unity element 1 is 0, then the minimum of such n is a prime number p, called the **characteristic** of K. On the other hand, if there is no natural number n such that $n1 = 0$, we say that the characteristic of K is 0.

C. Examples of Fields

The rational number field $\mathbf{Q}$ consisting of all rational numbers, the real number field $\mathbf{R}$ consisting of all real numbers, and the complex number field $\mathbf{C}$ consisting of all complex numbers, are all fields of characteristic 0. A subfield of the complex number field $\mathbf{C}$ is called a **number field**. The rational number field is a prime field, and every prime field of characteristic 0 is isomorphic to the rational number field. For the ring $\mathbf{Z}$ of all rational integers, the residue class ring modulo a prime number p is a field $\mathbf{Z}/p\mathbf{Z} = \{0, 1, 2, \ldots, p-1 \pmod{p}\}$ of characteristic p, called the **residue class field** for p. Thus $\mathbf{Z}/p\mathbf{Z}$ is a prime field, and every prime field of characteristic p is isomorphic to $\mathbf{Z}/p\mathbf{Z}$. If the number of the elements of a field K is finite, K is called a **finite field**. $\mathbf{Z}/p\mathbf{Z}$ is an example of a finite field.

D. Extensions of a Field

In order to express that K is an extension field of k, we often use the notation K/k. Subfields of K containing k are called **intermediate fields** of K/k. Consider two extensions K_1/k_1 and K_2/k_2, and let $\varphi: K_1 \rightarrow K_2$ be an isomorphism which induces an isomorphism $\psi: k_1 \rightarrow k_2$. Then we call φ an **extension** of ψ. Suppose that k_1, K_2 are given fields and K_2 contains a subfield k_2 isomorphic to k_1. Then there exist an extension field K_1 of k_1 and an isomorphism $\varphi: K_1 \rightarrow K_2$ which is an extension of the given isomorphism $\psi: k_1 \rightarrow k_2$; construction of the field K_1 is often called the embedding of k_1 into K_2. When K_1 and K_2 are extensions of k, an isomorphism: $K_1 \rightarrow K_2$ is called a k-**isomorphism** if it leaves every element of k invariant.

In an extension K/k, let S be a subset of K. The smallest intermediate field of K/k containing S is called the field obtained by **adjoining** S to k or the field **generated** by S over k, denoted by $k(S)$. The field $k(S)$ consists of those elements in K each of which is a rational expression in a finite number of elements of S with coefficients in k. An extension field $k(t)$ obtained by adjoining a single element t to k is called a **simple extension** of k, and in this case t is called a **primitive element** of the extension. The †rational function field $k(X)$ with coefficient field k is a simple extension of k with a primitive element X.

When subfields k_λ $(\lambda \in \Lambda)$ of a field K are given, the smallest subfield of K containing all these subfields exists and is called the **composite field** of the k_λ.

E. Algebraic and Transcendental Extensions

An element α of an extension field K of a field k is called an **algebraic element** over k if α is a †zero point of a nonzero polynomial, say, $f(X) = a_0 + a_1 X + \ldots + a_n X^n$ with coefficients in k. If α is not algebraic over k, then α is called a **transcendental element** over k. An algebraic element α is always a root of an irreducible polynomial over k which is uniquely determined up to a constant factor $(\in k^*)$ and is called the **minimal polynomial** of α over k. K is called an **algebraic extension** of k if all elements of K are algebraic over k; otherwise we call K a **transcendental extension** of k. If K_1 is an algebraic extension of K and K is an algebraic extension of k, then K_1 is also an algebraic extension of k. In an arbitrary extension field K of k, the set of all algebraic elements over k forms an algebraic extension field of k. A simple extension $k(t)$ with a transcendental element t is isomorphic to the rational function field of one variable with coefficient field k. If t is an algebraic element over k, $k(t)$ is isomorphic to the †residue class field of the polynomial ring $k[X]$ modulo the minimal polynomial $f(X)$ of t over k.

F. Finite Extensions

An extension field K of a field k is called a **finite extension** if K has no infinite set of elements that are †linearly independent over k,

i.e., if K is a finite-dimensional linear space over k. The dimension of the linear space over k is called the **degree** of K over k and is denoted by $(K:k)$ (or $[K:k]$). If K is a finite extension of k and L is a finite extension of K, then L is also a finite extension of k and $(L:K)(K:k)=(L:k)$. Every finite extension field of k is an algebraic extension of k and is obtained by adjoining a finite number of algebraic elements to k. Conversely, every field obtained by adjoining a finite number of algebraic elements to k is a finite extension of k. If $K=k(\alpha)$ with an algebraic element α, then $(K:k)$ is equal to the degree of the minimal polynomial of α over k, also called the **degree** of α over k. Every element of $k(\alpha)$ is expressed as a polynomial in α with coefficients in k. On the other hand, for any nonconstant polynomial $f(X)$ of $k[X]$ there exists a simple extension $k(\alpha)$ such that α is a root of $f(X)$.

G. Normal Extensions

An algebraic extension field K of a field k is called a **normal extension** of k if every irreducible polynomial of $k[X]$ which has a root in K can always be decomposed into a product of linear factors in $K[X]$. An extension field K of k is called a **splitting field** of a (nonconstant) polynomial $f(X)\in k[X]$ if $f(X)$ can be decomposed as a product of linear polynomials, i.e., $f(X)=c(X-\alpha_1)(X-\alpha_2)\ldots(X-\alpha_n)$, $c\in k$, $\alpha_i\in K$. A splitting field K of $f(X)$ ($f(X)\in k[X]$) is called a **minimal splitting field** of $f(X)$ if any proper subfield L of K ($K\supset L\supset k$) is not a splitting field of $f(X)$. A minimal splitting field of $f(X)$ is obtained by adjoining all the zero points of $f(X)$. A finite extension field of k is a normal extension if and only if it is a minimal splitting field of a polynomial of $k[X]$. For any given (nonconstant) polynomial $f(X)\in k[X]$, there exists a minimal splitting field of $f(X)$, and all minimal splitting fields of $f(X)$ are k-isomorphic.

H. Separable and Inseparable Extensions

An algebraic element α over k is called a **separable element** or an **inseparable element** over k according as the minimal polynomial of α over k is †separable or †inseparable. An algebraic extension K of k is called a **separable extension** of k if all the elements of K are separable over k; otherwise, K is called an **inseparable extension**. An element α is separable with respect to k if and only if the minimal polynomial of α over k has no double root in its splitting field. If α is inseparable, then k has nonzero characteristic p, and the minimal polynomial $f(X)$ of α can be decomposed as $f(X)=(X-\alpha_1)^{p^r}(X-\alpha_2)^{p^r}\ldots(X-\alpha_m)^{p^r}$, $r\geqslant 1$, where $\alpha_1,\alpha_2,\ldots,\alpha_m$ are distinct roots of $f(X)$ in its splitting field; between the degree n of $f(X)$ and the number m of distinct roots of $f(X)$, the relation $n=mp^r$ holds. In particular, if $\alpha^{p^r}\in k$ for some r, we call α a **purely inseparable element** over k. An algebraic extension of k is called **purely inseparable** if all elements of the field are purely inseparable over k. In an algebraic extension K of k the set of all separable elements forms an intermediate field K_0 of K/k. The field K_0 is called the **maximal separable extension** of k in K. If K is inseparable over k. i.e., if $K\neq K_0$, then the characteristic of k is $p\neq 0$, and K is purely inseparable over K_0. The degrees $d=[K_0:k]$ and $p^r=[K:K_0]$ are denoted by $[K:k]_s$ and $[K:k]_i$, respectively. A separable extension of a separable extension of k is also separable over k, and every finite separable extension of k is a simple extension.

If no inseparable irreducible polynomial in $k[X]$ exists, we call k a **perfect field**; otherwise, an **imperfect field**. Every field of characteristic 0 is a perfect field. A field of characteristic p ($\neq 0$) is perfect if and only if for each $a\in k$ the polynomial X^p-a has a root in k. Every algebraic extension of a perfect field is a separable extension and a perfect field. Any imperfect field has an inseparable, in fact purely inseparable, proper extension.

I. Algebraically Closed Fields

If every nonconstant polynomial of $k[X]$ can be decomposed into a product of linear polynomials of $k[X]$, or equivalently, if every irreducible polynomial of $k[X]$ is linear, k is called an **algebraically closed field**; k is algebraically closed if and only if k has no algebraic extension field other than k, and hence every algebraically closed field is perfect. For any given field k there exists an algebraically closed algebraic extension field of k unique up to k-isomorphisms (E. Steinitz); hence we call such a field the **algebraic closure** of k. To proceed further, suppose that we are given a field k and its extension K. If there is no algebraic element of K over k outside of k, i.e., if k is the intersection of K and the algebraic closure of k, then we say that k is **algebraically closed** in K. The complex number field is an algebraically closed field (C. F. Gauss's fundamental theorem of algebra; → 10 Algebraic Equations).

J. Conjugates

Let k be a field and K an algebraic extension of k. Two elements α, β of K are called **conju-**

gate over k if they are roots of the same irreducible polynomial of $k[X]$ (or equivalently, if the minimal polynomials of α and β with respect to k coincide); in this case we call the subfields $k(\alpha)$, $k(\beta)$ **conjugate fields** over k. The conjugate fields $k(\alpha)$ and $k(\beta)$ are k-isomorphic under an isomorphism σ such that $\sigma(\alpha)=\beta$. In particular, if K is a normal extension of k, the number of conjugate elements of an element α of K is the number of distinct roots of the minimal polynomial $f(X)$ of α, which is independent of the choice of a normal extension K containing k. The element α is separable if and only if the number of conjugate elements in K is the same as the degree of $f(X)$. On the other hand, $k(\alpha)$ is normal over k if and only if $k(\alpha)$ coincides with all its conjugate fields.

Let α be a separable algebraic element over k, and let $\alpha_1=\alpha, \alpha_2, \ldots, \alpha_n$ be conjugate elements of α over k. The product $A=\alpha_1\alpha_2\ldots\alpha_n$ and sum $B=\alpha_1+\alpha_2+\ldots+\alpha_n$ are elements of k. Indeed, if $f(X)=X^n+c_1X^{n-1}+\ldots+c_n$ is the minimal polynomial of α with respect to k, we have $A=(-1)^n c_n$, $B=-c_1$, and A and B are called the **norm** and the **trace** of α, respectively, denoted by $A=N(\alpha)$, $B=Tr(\alpha)$. Let K be a finite separable extension of degree n over k, and let α be an element of K. Then the degree m of the minimal polynomial of α is a divisor of n; that is, $n=mr$ with a positive integer r. We define the norm and the trace of α with respect to K/k by $N_{K/k}(\alpha)=N(\alpha)^r$, $Tr_{K/k}(\alpha)=rTr(\alpha)$, respectively. Then these quantities satisfy $N_{K/k}(\alpha\beta)=N_{K/k}(\alpha)N_{K/k}(\beta)$, $Tr_{K/k}(\alpha+\beta)=Tr_{K/k}(\alpha)+Tr_{K/k}(\beta)$ for $\alpha, \beta\in K$. (For the Galois theory of algebraic extensions → 172 Galois Theory.)

K. Transcendental Extensions

Let K be an extension of k and $u_1, \ldots, u_n$ be elements of K. An element v of K is said to be **algebraically dependent** on the elements $u_1, u_2, \ldots, u_n$ if v is algebraic over the field $k(u_1, u_2, \ldots, u_n)$. A subset S of K is called **algebraically independent** over k if no $u\in S$ is algebraically dependent on a finite number of elements of S different from u itself; S is called a **transcendence basis** of K over k if S is algebraically independent and K is algebraic over $k(S)$. Furthermore, if K is separable over $k(S)$, then S is called a **separating transcendence basis** of K over k. There always exists an algebraically independent basis of K over k, and the †cardinal number of S depends only on K and k; this cardinal number is called the **transcendence degree** (or **degree of transcendency**) of K over k. (When S is an infinite set, we sometimes say that the transcendence degree is infinite.) In particular, if $K=k(S)$ with an algebraically independent S, K is called a **purely transcendental extension** of k.

An extension K of k is called a **separably generated extension**, or simply a **separable extension**, if every finitely generated intermediate field of K/k has a separating transcendence basis over k. If K itself has a separating transcendence basis over k, then K is separably generated, but not conversely.

A purely transcendental extension field of k having a finite transcendence degree n is also called a **rational function field in n variables** over k, and a finite extension of such a rational function field is called an **algebraic function field in n variables** over k.

Let K and L be extension fields of k, both contained in a common extension field. We say that K and L are **linearly disjoint** over k if every subset of K linearly independent over k is also linearly independent over L, or equivalently, if every subset of L linearly independent over k is also linearly independent over K. An algebraic function field $K=k(x_1, x_2, \ldots, x_n)$ over k (whose transcendence degree is $\leqslant n$) is called a **regular extension** of k if K and the algebraic closure $\bar{k}$ of k are linearly disjoint. In order that K be regular over k it is necessary and sufficient that k be algebraically closed in K and that K be separably generated over k.

L. Derivations

A map D of a field K into itself is called a **derivation** of K if it satisfies $D(a+b)=D(a)+D(b)$ and $D(ab)=aD(b)+bD(a)$ for all $a, b\in K$. The set of elements c of K for which $D(c)=0$ is a subfield. If the characteristic of K is $p(\neq 0)$, then $D(x^p)=0$ for all $x\in K$. Let k be a subfield of K. A derivation D of K is called a **derivation over** k if $D(c)=0$ for all $c\in k$; the totality of derivations over k is a †k-module. If $K=k(x_1, x_2, \ldots, x_n)$ is an algebraic function field over k, then the k-module of derivations over k has finite dimension $s(\leqslant n)$ and we can choose s suitable elements $u_1, u_2, \ldots, u_s$ of K such that K is separably algebraic over $k(u_1, u_2, \ldots, u_s)$. Generally, the transcendence degree r of K over k does not exceed s, and K is separably generated over k if and only if $r=s$.

M. Finite Fields

Finite fields were first considered by E. Galois (1830), so they are also called **Galois fields**. There is no finite noncommutative field (**Wedderburn's theorem**, J. H. M. Wedderburn, *Trans. Amer. Math. Soc.*, 6 (1905)). A simple

proof for this was given by E. Witt (*Abh. Math. Sem. Univ. Hamburg*, 8 (1931)). The characteristic of a finite field is a prime p, and the number of elements of the field is a power of p. Conversely, for any given prime number p and natural number α, there exists a finite field with p^α elements. Such a field is unique up to isomorphism, which we denote by $GF(p^\alpha)$ or $\mathbf{F}_q(q=p^\alpha)$. For any positive integer m, $GF(p^{m\alpha})$ is an extension field of $GF(p^\alpha)$ of degree m and a †cyclic extension. Every element a of $GF(p^\alpha)$ satisfies $a^{p^\alpha}=a$; hence a has its pth root in $GF(p^\alpha)$. Therefore every finite field is perfect. The multiplicative group of $GF(p^\alpha)$ is a cyclic group of order $p^\alpha-1$.

N. Ordered Fields and Real Fields

A field K is called an **ordered field** if there is given a †total order in K such that $a>b$ implies $a+c>b+c$ for all c and $a>b$, $c>0$ implies $ac>bc$. The characteristic of an ordered field is always 0. An element a of K is called a **positive element** or a **negative element** according as $a>0$ or $a<0$. For an element a of K the **absolute value** of a, denoted by $|a|$, is a or $-a$ according as $a\geqslant 0$ or $a\leqslant 0$. If we define neighborhoods of a by the sets $\{x \mid a-\varepsilon<x<a+\varepsilon\}$ with positive elements ε, K becomes a †Hausdorff space. If, for any two positive elements a, b of K, there exists a natural number n such that $na>b$, then we call K an **Archimedean ordered field**. Two ordered fields are called **similarly isomorphic** if there exists an isomorphism between them under which positive elements are always mapped to positive elements. The rational number field and the real number field are examples of Archimedean ordered fields, while every Archimedean ordered field is similarly isomorphic to a subfield of the real number field. (For the structure of non-Archimedean ordered fields, see [8].) A field k is called a **formally real field** (or simply **real field**) if -1 (1 is the unity element of k) cannot be expressed as a finite sum of squares of elements of k. The real number field is a model of formally real fields. More generally, every ordered field is a formally real field. A formally real field is called a **real closed field** if no proper algebraic extension of it is a formally real field. The real number field is a real closed field. The algebraic closure of a real closed field is obtained by adjoining a root of the polynomial X^2+1. If a is a nonzero element of a real closed field, then either a or $-a$ can be a square of an element of the field. Every real closed field can be made an ordered field in a unique way, namely, by defining squares of nonzero elements to be positive elements. Since it is known that every formally real field is a subfield of a real closed field, it follows that every formally real field is an ordered field and therefore is of characteristic 0. The problem of constructing an ordered field out of a formally real field is closely related to the existence of valuations of a certain type [8]. The notion of formally real fields was introduced by E. Artin (*Abh. Math. Sem. Univ. Hamburg*, 5 (1927)). By making use of the theory of formally real fields, Artin succeeded in solving affirmatively †Hilbert's 17th problem, which asked whether every positive definite rational expression (i.e., a rational expression with real coefficients that takes positive values for all real variables) can be expressed as a sum of squares of rational expressions. More precisely, it was shown by A. Pfister that every positive definite function in $\mathbf{R}(X_1, \dots, X_n)$ is a sum of at most 2^n squares.

References

[1] E. Steinitz, Algebraische Theorie der Körper, J. Reine Angew. Math., 137 (1910), 167–309.
[2] E. Steinitz, Algebraische Theorie der Körper, Walter de Gruyter, 1930 (Chelsea, 1950).
[3] H. Hasse, Höhere Algebra I, II, Sammlung Göschen, Walter de Gruyter, 1926–1927.
[4] B. L. van der Waerden, Algebra I, Springer, seventh edition, 1966.
[5] A. A. Albert, Fundamental concepts of higher algebra, Univ. of Chicago Press, 1956.
[6] N. Bourbaki, Eléments de mathématique, Algèbre, ch. 5, Actualités Sci. Ind., 1102b, Hermann, second edition, 1959.
[7] N. Jacobson, Lectures in abstract algebra III, Van Nostrand, 1964.
[8] M. Nagata, Field theory, Dekker, 1977.

150 (XX.28) Field Theory

A. History

When a quantity $\psi(x)$, such as velocity, is defined at every point x in a certain region of a space, we say that a field of the quantity ψ is given. This general concept is used in many branches of science. Here we confine ourselves to some branches of physics, in particular to the quantum theory of fields, which describes †elementary particles.

The †theories of elasticity and †hydrodynamics (in particular, concerning †Euler's

equation of motion) deal with displacement and velocity fields, respectively. However, a field in a vacuum (ether), which is quite different from a field in a space filled with matter, first became a subject of physics in †electromagnetism. M. Faraday (1837) introduced the electromagnetic field and discovered its fundamental laws, and J. C. Maxwell (1837) completed the mathematical formulation. On the basis of this formalism, A. Einstein (1905) established the theory of †relativity and later developed the general theories of relativity and of gravity.

Although quantum theory originated from the problem of blackbody radiation, the quantum theory of the electromagnetic field was first developed by P. A. M. Dirac (1927) after the development of †quantum mechanics. Along similar lines P. Jordan and E. P. Wigner (1927) quantized the matter wave (electron field), and W. Heisenberg and W. Pauli (1929) developed the quantum theory of wave fields in general. Subsequently, Jordan and Pauli (1927), Pauli (1939), S. Tomonaga (1943), J. Schwinger (1948), and others reformulated the theory in a relativistically covariant manner. Quantum electrodynamics, dealing with an electromagnetic field interacting with electrons (and positrons), has given excellent agreement with experimental measurements when formulated in this way. Divergence difficulties inherent to quantized field theory were bypassed by the renormalization procedure of Tomonaga, Schwinger, R. P. Feynman, and F. J. Dyson (1947).

On the other hand, H. Yukawa (1934) applied the concept of the quantized field to the interpretation of nuclear force and predicted the existence of π-mesons. Many kinds of new particles have since been found, including π-mesons and muons. The field theory of π-mesons can explain the qualitative features of the meson-nucleon system. Similar theories can be formulated for other types of unstable particles that have been found in cosmic rays since 1949.

During the progress of meson theory, various types of fields were investigated, and a general theory of elementary particles was developed. Dirac (1936) proposed a general wave equation for elementary particles, and Pauli and M. Fierz (1939) proved the connection of spin and statistics (→ the end of Section D). Schwinger (1951) derived quantum-mechanical equations of motion and commutation relations from a unified variational principle. For the cases where no interaction is present, a general theory of elementary particles consistent with the requirements of relativity and of quantum theory was established as the theory of **free fields**.

B. Relativistically Covariant Classical Fields

Relativistically covariant fields $\varphi_r^\alpha(x)$ on the level of classical theory are functions of the space-time point x with either real or complex values depending on the index α (which distinguishes different fields) [1]. A finite-dimensional representation D^α of $SL(2, C)$ (→ 258 Lorentz Group) on either a real or complex vector space is assigned to each α, and the index r of $\varphi_r^\alpha(x)$ refers to components in this representation space. Each model is specified in terms of a **Lagrangian density** $\mathscr{L}(x)$ that is a function (typically a polynomial) of $\varphi_r^\alpha(x)$, $\partial_\mu\varphi_r^\alpha(x)$ (∂_μ denoting $\partial/\partial x^\mu$, $\mu=0, \ldots, 3$) and their complex conjugates. $\mathscr{L}$ is taken to be invariant under the replacement of $\varphi_r^\alpha(x)$ and $\partial_\mu\varphi_r^\alpha(x)$ by $\sum_s D^\alpha(A)_{rs}\varphi_s^\alpha(x)$ and $\sum_{s,\nu} D^\alpha(A)_{rs}\Lambda(A)_\mu^\nu\partial_\nu\varphi_s^\alpha(x)$ for all $A\in SL(2, C)$ (called the **Lorentz invariance** of $\mathscr{L}$) and possibly under the replacement of $\varphi_r^\alpha(x)$ and $\partial_\mu\varphi_r^\alpha(x)$ by $\sum_\beta U^{\alpha\beta}\varphi_r^\beta(x)$ and $\sum_\beta U^{\alpha\beta}\partial_\mu\varphi_r^\beta(x)$ (called an **internal symmetry**).

The fields are supposed to satisfy the following partial differential equation, called the **field equation**, and are obtained as the †Euler equation of the variational problem for the action integral $I=\int\mathscr{L}(x)dx$:

$$\partial\mathscr{L}(x)/\partial\varphi_r^\alpha(x)-\sum_{\mu=0}^{3}(\partial/\partial x^\mu)(\partial\mathscr{L}(x)/\partial(\partial^\mu\varphi_r^\alpha(x))=0,$$

$$\partial\mathscr{L}(x)/\partial\bar{\varphi}_r^\alpha(x)-\sum_{\mu=0}^{3}(\partial/\partial x^\mu)(\partial\mathscr{L}(x)/\partial(\partial^\mu\bar{\varphi}_r^\alpha(x)))=0$$

(the bar indicates the complex conjugate), where the second equation is identical to the first for a real field. For complex fields, $\mathscr{L}(x)$ is chosen such that $I=\bar{I}$ (possibly except for a surface term) in order to ensure that the above two equations are complex conjugates of each other.

The invariance of I under a Lie group of (local and/or volume-preserving point) transformations of fields implies a conservation law for a certain quantity (**Noether's theorem**; → e.g., E. L. Hill, *Rev. Mod. Phys.*, 23 (1951)). For translational invariance ($x^\mu\to x^\mu+a^\mu$), the **energy-momentum (stress) tensor**

$$T^\mu{}_\nu(x)\equiv-\delta_{\mu\nu}\mathscr{L}(x)+\sum_{\alpha,r}\{\partial\mathscr{L}(x)/\partial(\partial_\mu\varphi_r^\alpha(x))\partial_\nu\varphi_r^\alpha(x)+\partial\mathscr{L}(x)/\partial(\partial_\mu\bar{\varphi}_r^\alpha(x))\partial_\nu\bar{\varphi}_r^\alpha(x)\}$$

(the complex conjugate terms in all equations to be suppressed for real fields) satisfies the differential conservation law $\sum_{\mu=0}^{3}\partial_\mu T^\mu{}_\nu(x)=0$ as a consequence of the field equations, which in turn implies the time-independence of the

following quantity provided that $T^k{}_\mu(x)$ vanishes sufficiently fast at spatial infinity:

$$H=\int T^0{}_0(x)d^3\mathbf{x},\quad P_k=\int T^0{}_k(x)d^3\mathbf{x}\quad (x^0=t).$$

These are the total field Hamiltonian (energy) and momentum.

In a general situation, $T^{\mu\nu}(=\sum_\rho g^{\nu\rho}T^\mu{}_\rho$, where $g^{\mu\nu}$ is the Minkowski metric tensor with the signature $1,-1,-1,-1$) is not necessarily symmetric in μ and ν. F. J. Belinfante (*Physica*, 6 (1939); 7 (1940)) has given a formula (by a change of $\mathscr{L}(x)$ via so-called 4-divergence) for obtaining a symmetrized energy-momentum tensor $\theta^\mu{}_\nu$ for which H and P_k are equal to those defined from the above $T^\mu{}_\nu$ and for which the differential conservation law is satisfied.

Lorentz invariance implies the conservation law $\sum_\mu \partial_\mu M^{\mu\nu\lambda}(x)=0$ for the **angular momentum density**

$$\begin{aligned}M^{\mu\nu\lambda}(x)&=x^\nu T^{\mu\lambda}(x)-x^\lambda T^{\mu\nu}(x)\\&\quad+\sum_{\alpha,r,s}\{\partial\mathscr{L}/\partial(\partial_\mu\varphi_r^\alpha(x))S_{rs}^{\alpha\nu\lambda}\varphi_r^\alpha(x)\\&\qquad+\partial\mathscr{L}/\partial(\partial_\mu\bar{\varphi}_r^\alpha(x))\bar{S}_{rs}^{\alpha\nu\lambda}\bar{\varphi}_r^\alpha(x)\}\\&=x^\nu\Theta^{\mu\lambda}(x)-x^\lambda\Theta^{\mu\lambda}(x)\end{aligned}$$

and the time-independence of the total angular momentum

$$M^{\nu\lambda}=\int M^{0\nu\lambda}d^3\mathbf{x}\qquad(x^0=t),$$

where $S_{rs}^{\alpha\mu\nu}$ is antisymmetric in μ, ν and $D^\alpha(A)_{rs}=\delta_{rs}+\sum_{\mu,\nu}S_{rs}^{\alpha\mu\nu}\varepsilon_{\mu\nu}/2$ for $\Lambda(A)_{\mu\nu}=g_{\mu\nu}+\varepsilon_{\mu\nu}$ with an infinitesimal $\varepsilon_{\mu\nu}$. (We have $(\Lambda(A)x)^\mu=\sum_\nu\Lambda(A)^\mu_\nu x^\nu$ and $\Lambda(A)_{\mu\nu}=\sum_\rho g_{\mu\rho}\Lambda(A)^\rho{}_\nu$.)

A continuous one-parameter group $U(\rho)$ of internal symmetries implies the conservation law $\sum_\mu\partial_\mu J^\mu(x)=0$ for the **4-current density** (charge ($\mu=0$) and current ($\mu=1,2,3$) densities)

$$\begin{aligned}J^\mu(x)=\sum_{\alpha\beta r}&\{\partial\mathscr{L}/\partial(\partial_\mu\varphi_r^\alpha(x))\lambda^{\alpha\beta}\varphi_r^\beta(x)\\&+\partial\mathscr{L}/\partial(\partial_\mu\bar{\varphi}_r^\alpha(x))\bar{\lambda}^{\alpha\beta}\bar{\varphi}_r^\beta(x)\}\end{aligned}$$

and the time-independence of the **charge** $\int J^0(x)d^3\mathbf{x}$ $(x^0=t)$, where $U'(0)^{\alpha\beta}=\lambda^{\alpha\beta}$. An example of $U(\rho)$ is the multiplication by $\exp iQ^\alpha\rho$ (called a **gauge transformation of the first kind**), where Q^α is an integer with $Q^\alpha=0$ for real fields φ_r^α.

Examples of Lagrangian densities for noninteracting fields (called **free Lagrangian densities**) are:

real scalar field:
$(\sum_{\mu\nu}g^{\mu\nu}\partial_\mu\varphi(x)\partial_\nu\varphi(x)-m^2\varphi(x)^2)/2$,
complex scalar field:
$(\sum_{\mu\nu}g^{\mu\nu}\partial_\mu\bar{\varphi}(x)\partial_\nu\varphi(x)-m^2\bar{\varphi}(x)\varphi(x))$,

Dirac field:

$$\begin{aligned}(i/2)\textstyle\sum_{\mu rs}&(\psi_r(x)^\dagger\gamma^\mu_{rs}\partial_\mu\psi_s(x)-\partial_\mu\psi_r(x)^\dagger\gamma^\mu_{rs}\psi_s(x))\\&-m\textstyle\sum_r\psi_r(x)^\dagger\psi_s(x)\end{aligned}$$

$(\psi_r(x)^\dagger=\sum_s\bar{\psi}_s(x)\gamma^0_{sr}$, $\gamma^\mu\equiv(\gamma^\mu_{rs})$ are †Dirac's γ-matrices),

massive real vector field:

$$\begin{aligned}-(1/4)&\sum_{\kappa\lambda\mu\nu}g^{\kappa\mu}g^{\lambda\nu}(\partial_\kappa A_\lambda(x)-\partial_\lambda A_\kappa(x))\\&\times(\partial_\mu A_\nu(x)-\partial_\nu A_\mu(x))+\frac{m2}{2}\sum_{\mu\nu}g^{\mu\nu}A_\mu(x)A_\nu(x),\end{aligned}$$

where D^α is trivial for scalar fields, $[1,0]\oplus[0,1]$ for the Dirac field, and $[1,1]$ for vector fields.

For interacting fields, some interaction parts are added to the sum of such free Lagrangian densities of relevant fields. Some examples are:

$P(\varphi)$(or $g\varphi^4$) interaction:
$P(\varphi(x))$ (or $g\varphi(x)^4$),
Yukawa-type interaction:
$g\sum_{\mu rs}\psi_r(x)^\dagger\gamma^\mu_{rs}\psi_s(x)A_\mu(x)$,
Fermi interactions:

$$\begin{aligned}g\textstyle\sum_{kk'}G_{kk'}&(\textstyle\sum_{rs}\psi_r^\alpha(x)^\dagger\mathcal{O}^k_{rs}\psi_s^\beta(x))\\&\times(\textstyle\sum_{r's'}\psi_{r'}^{\alpha'}(x)^\dagger\mathcal{O}^{k'}_{r's'}\psi_{s'}^{\beta'}(x))\end{aligned}$$

($\mathcal{O}^k=1$ (scalar, no index k), γ^μ (vector, $G_{kk'}=g_{\mu\mu'}$), $\gamma^\mu\gamma^\nu$ (tensor, $G_{kk'}=g_{\mu\mu'}g_{\nu\nu'}$), $\gamma^5\gamma^\mu$ (pseudovector, $G_{kk'}=g_{\mu\mu'}$, $\gamma^5=i\gamma^0\gamma^1\gamma^2\gamma^3$), γ^5 (pseudoscalar, no index k).

C. Heuristic Theory of Quantized Fields

For free fields, a quantization procedure similar to the usual quantum mechanics (→ 351 Quantum Mechanics; 377 Second Quantization) leads to the following type of canonical commutation or anticommutation relations among fields (and their first time derivatives) at time 0; for example,

real scalar field:
$[\varphi(0,\mathbf{x}),\dot{\varphi}(0,\mathbf{y})]_-=i\delta^3(\mathbf{x}-\mathbf{y})$,
$[\varphi(0,\mathbf{x}),\varphi(0,\mathbf{y})]_-=[\dot{\varphi}(0,\mathbf{x}),\dot{\varphi}(0,\mathbf{y})]_-=0$,
Dirac field:
$[\bar{\psi}_r(0,\mathbf{x}),\psi_s(0,\mathbf{y})]_+=\delta_{rs}\delta^3(\mathbf{x}-\mathbf{y})$,
$[\psi_r(0,\mathbf{x}),\psi_s(0,\mathbf{y})]_+=[\bar{\psi}_r(0,\mathbf{x}),\bar{\psi}_s(0,\mathbf{y})]_+=0$.

($[A,B]_\pm=AB\pm BA$, $\delta_{rs}=0$ for $r\neq s$ and $\delta_{rr}=1$, $\delta^3(\mathbf{x}-\mathbf{y})=\prod_{k=1}^3\delta(x^k-y^k)$.) The free-field equations then lead to the following 4-dimensional commutation relations:

real scalar field:
$[\varphi(x),\varphi(y)]_-=-i\Delta_m(x-y)$,
Dirac field:
$[\psi_r(x),\psi_s(y)^\dagger]_+=(\sum_\mu\gamma^\mu_{rs}\partial_\mu-im\delta_{rs})\Delta_m(x-y)$,
massive real vector field:
$[A_\mu(x),A_\nu(y)]_-=i(g_{\mu\nu}+m^{-2}\partial_\mu\partial_\nu)\Delta_m(x-y)$.

Here Δ_m is the †invariant distribution. There is a unique representation of such relations for a

field (called the **Fock representation**) with a vector Ω (called the **free vacuum** vector) which is annihilated by $\int \varphi(x)f(x)dx$, by $\int \bar{\psi}_r(x)f(x)dx$ and $\int \bar{\psi}_r(x)f(x)dx$, or by $\int A_\mu(x)f(x)dx$ whenever $\int e^{ip\cdot x}f(x)dx=0$ for $p^0>0$ (here, $p\cdot x=\sum g_{\mu\nu}p^\mu x^\nu$), and which is cyclic (i.e., polynomials of (smeared-out) fields generate a dense subset of Ω). The free fields in the Fock representation satisfy the Wightman axioms (→ Section D).

In the Fock representation of canonical fields a translationally invariant vector must be a free vacuum vector, up to multiplication by a complex number. In order to construct a model of a translationally invariant interaction among canonical fields with a unique vector of minimal energy (a vector which is called the true or interacting vacuum and which must be translationally invariant if unique), one must look for some other suitable representation. Such a no-go theorem is called **Haag's theorem**.

The earliest formulation of interacting quantized fields was developed heuristically by formal manipulation in the Fock representation, described in textbooks of **quantum field theory** [2–6]. It can be mathematically justified if the so-called **cutoff** is introduced by limiting the space to a finite volume, possibly changing the space into a lattice and smoothing out fields (effectively cutting off the high-energy part of the interaction) (A. M. Jaffe, O. E. Lanford III, and A. S. Wightman, *Comm. Math. Phys.*, 15 (1969)). The full theory is then expected to be obtained by taking a limit of the true vacuum expectation values as various cutoff parameters are removed. This is the aim of constructive field theory (→ Section F), which has been achieved for some space-time models of dimension 2 and 3.

In quantum field theory, the S-matrix is given in terms of (the mass-shell restriction of the Fourier transform of) the vacuum expectation value of the time-ordered product of fields, called τ-functions (→ Section E). In the heuristic approach, it is given in terms of the following Gell-Mann–Low formula (its imaginary time version being mathematically used in constructive field theory): If the field $\varphi(x)=e^{iHx^0}\varphi_0((0,\mathbf{x}))e^{-iHx^0}$ with $H=H_0+H_I$ (H_0 is the free Hamiltonian as the generator of the time translation for the free field $\varphi_0(x)$ and H_I is the interaction Hamiltonian, such as $P(\varphi_0)$) and $x_1^0>\ldots>x_n^0$, then

$$(\Omega,\varphi(x_1)\ldots\varphi(x_n)\Omega)$$
$$=\lim_{T\to\infty}(e^{iHT}\Omega_0,\varphi(x_1)\ldots\varphi(x_n)e^{-iHT}\Omega_0)$$
$$/(\Omega_0,e^{-2iHT}\Omega_0),$$

where Ω_0 is the free vacuum. The numerator can be written as

$$(\Omega_0,U(T,x_1^0)\varphi_0(x_1)U(x_1^0,x_2^0)\ldots$$
$$\varphi_0(x_n)U(x_n^0,T)\Omega_0)$$

in terms of $U(t,s)=e^{itH_0}e^{-i(t-s)H}e^{-isH_0}$ and the canonical field φ_0 at time 0. The covariant perturbation series due to Tomonaga, Schwinger, Feynman, and Dyson is obtained by substituting the following expansion, which is the iteration of the Duhamel formula:

$$U(t,s)$$
$$=\sum_{n=0}^{\infty}(-i)^n\int_s^t dt_1\ldots\int_s^{t_{n-1}}dt_n H_I(t_1)\ldots H_I(t_n).$$

Here, $H_I(t)\equiv e^{itH_0}H_Ie^{-itH_0}$. Each term can be represented by connected †Feynman diagrams (the denominator canceling out all disconnected graphs) and computed according to the †Feynman rule, yielding †Feynman integrals. Each expression so obtained (formally) may be a divergent integral (in the absence of the cutoff), in which case one tries to cancel it out by modifying the original Hamiltonian with the alteration of parameters such as mass and coupling constant (by an amount called the **renormalization constant**, which is divergent in the absence of the cutoff) or possibly by the addition of terms again involving renormalization constants. If this can be achieved in terms of a finite number of renormalization constants, the model or the Hamiltonian is called **renormalizable** [7]. If the divergent integral appears only in a finite number of graphs (counting the same subgraph of an infinite number of different diagrams as one graph), the model is called **super-renormalizable**.

D. Axiomatic Quantum Field Theory

Relativistically covariant fields $\varphi_\alpha(x)$ on a Hilbert space $\mathscr{H}$ are called **Wightman fields** if the following four **Wightman axioms** are fulfilled [8–10]:

(1) The fields $\varphi_\alpha(x)$ are **operator-valued distributions**: For each C^∞-function f of rapid decrease on the Minkowski space ($f\in\mathscr{S}(\mathbf{R}^4)$), $\varphi_\alpha(f)$ is an operator defined on a common domain D dense in $\mathscr{H}$ and satisfying $\varphi_\alpha(f)D\subset D$ (the domain of $\varphi_\alpha(f)^*$ contains D), $\varphi_\alpha(f)^*|D=\varphi_{\bar\alpha}(\bar f)$ for another index $\bar\alpha$ and $(\Psi,\varphi_\alpha(f)\Phi)$ for any Ψ, and Φ in D is linear and continuous in f relative to the topology of the Schwartz space $\mathscr{S}(\mathbf{R}^4)$.

(2) **Relativistic covariance**: There exists a continuous unitary representation $U(a,A)$ ($a\in\mathbf{R}^4$ and $A\in SL(2,\mathbf{C})$) of the universal covering group $\tilde{\mathscr{P}}_+^\uparrow$ of the †restricted inhomogeneous Lorentz group $\mathscr{P}_+^\uparrow$ on $\mathscr{H}$ such that $U(a,A)D\subset$

D and

$$U(a, A)\varphi_\alpha(f)U(a, A)^* = \sum_\beta D(A)_{\alpha\beta}\varphi_\beta(f_{a,A}),$$

where $f_{a,A}(x) = f(\Lambda(A)^{-1}(x-a))$ and $(D(A)_{\alpha\beta})$ is a finite-dimensional representation of $SL(2, \mathbf{C})$. The index α usually consists of †undotted and dotted indices, interchanged in $\bar{\alpha}$, and of other indices, interchanged among them in $\bar{\alpha}$.)

(3) **Locality**: If the supports of f and g are mutually spacelike, then

$$\varphi_\alpha(f)\varphi_\beta(g) = \varepsilon(\alpha, \beta)\varphi_\beta(g)\varphi_\alpha(f)$$

on D, where $\varepsilon(\alpha, \beta) = \pm 1$. If $\varepsilon(\alpha, \beta) = -1$ exactly when $D(A)_{\alpha\alpha} = D(A)_{\beta\beta} = -1$ for $A = -1$ (i.e., when both φ_α and φ_β are Fermi fields), they are said to satisfy the **normal commutation relations**.

(4) **Spectrum conditions**: Let $U(a, 1) = e^{ia \cdot P}$. The joint spectrum of P^μ $(\mu = 0, 1, 2, 3)$ is in the forward cone $\bar{V}_+ = \{p \in \mathbf{R}^4 \mid p \cdot p \geqslant 0, p^0 \geqslant 0\}$, with a point spectrum of multiplicity 1 at $p = 0$. A vector Ω belonging to the point spectrum 0 of P^μ is called the true (or interacting) vacuum and is required to be in D.

Usually D is taken to be minimal, namely, D is the linear hull of Ω and $\varphi_{\alpha_1}(f_1)\dots\varphi_{\alpha_n}(f_n)\Omega$ with all possible $n, \alpha_1, \dots, \alpha_n, f_1, \dots, f_n$. By means of the nuclear theorem, it is possible to define a linear operator $\varphi_{\alpha_1 \dots \alpha_n}(f)$ for $f(x_1, \dots, x_n)$ in $\mathscr{S}(\mathbf{R}^{4n})$, linear and continuous in f on D such that it coincides with $\varphi_{\alpha_1}(f_1)\dots\varphi_{\alpha_n}(f_n)$ if $f(x) = \prod_{j=1}^n f_j(x_j)$. If such operators are introduced, then the linear hull of Ω and $\varphi_{\alpha_1 \dots \alpha_n}(f)\Omega$ is taken to be D.

Under the foregoing axiom, the vacuum expectation values of the products of field operators define tempered distributions $W_\alpha(x) = W_{\alpha_1 \dots \alpha_n}(x_1, \dots, x_n)$, called **Wightman functions**, such that

$$W_\alpha(\prod f_j) = (\Omega, \varphi_{\alpha_1}(f_1)\dots\varphi_{\alpha_n}(f_n)\Omega).$$

The notion of a connected graph in perturbation theory corresponds to the following notion of **truncated Wightman functions** W_α^T:

$$W_\alpha^T(I) = \sum_m (-1)^{m-1}(m-1)! \sum_{\{I_k\}} \prod_{k=1}^m W_\alpha(I_k),$$

$$W_\alpha(I) = \sum_m \sum_{\{I_k\}} \prod_{k=1}^m W_\alpha^T(I_k),$$

where I indicates a set of variables $x_j \in M, j \in I$, $\{I_k\}$ is a partition of I into m subsets, with the order of the x's in each I_k remaining the same as in I, and the sum over $\{I_k\}$ extending over all possible partitions. If the spectrum of $\{P^\mu\}$ in $\Omega^\perp$ is contained in $\bar{V}_+^m = \{p \in M \mid p \cdot p \geqslant m^2, p^0 \geqslant m\}$ for some $m > 0$ (the mass gap), then W_α^T has an exponential clustering property at spatial infinity. For example $W_\alpha^T(x_1 + a_1, \dots, x_n + a_n)e^{m'R} \to 0$ as a distribution in x if $m' < m$, $a_j^0 = 0$ for all j and $R = \max|\mathbf{a}_j - \mathbf{a}_k| \to \infty$.

If the mass operator $(P \cdot P)^{1/2}$ has an isolated point spectrum at m with the eigenspace $\mathscr{H}_1(m)$, then there are sufficiently many n, $\alpha = (\alpha_1, \dots, \alpha_n)$, and $f(x_1, \dots, x_n)$ such that $\varphi_\alpha(f)\Omega \in \mathscr{H}_1(m)$, and the linear hull of such vectors are dense in $\mathscr{H}_1(m)$. In terms of the f_j satisfying $\varphi_{\alpha_{(j)}}(f_j)\Omega \in \mathscr{H}_1(m_j)$ (some of the m's may coincide) and the solutions $g_j(x) = \int \tilde{g}_j(p)\exp i(\mathbf{px} - \omega_j(\mathbf{p})x^0)d^3\mathbf{p}$, $\omega_j(p) = (\mathbf{p}^2 + m_j^2)^{1/2}$, of the Klein-Gordon equation, then the following limits, called **out** and **in states**, exist:

$$\Psi^{\substack{\text{out}\\ \text{in}}}(h_1 \dots h_n) = \lim_{t \to \pm\infty} Q_1(t, g_1)\dots Q_n(t, g_n)\Omega,$$

$$Q_j(t, g_j) = \int \varphi_{\alpha_{(j)}}(x_1 + (t, \mathbf{x}), \dots, x_{n_j} + (t, \mathbf{x})) \times f_j(x_1 + \dots + x_{n_j})g_j(t, \mathbf{x})\,dx_1 \dots dx_{n_j}d\mathbf{x},$$

$$h_j = (2\pi)^3 \tilde{g}_j(P)\varphi_{\alpha_j}(f_j)\Omega \in \mathscr{H}_1(m_j).$$

It defines the S-matrix elements as follows:

$$S(h_1 \dots h_n; h_1' \dots h_{n'}') = (\Psi^{\text{out}}(h_1 \dots h_n), \Psi^{\text{in}}(h_1' \dots h_{n'}')).$$

This is called **Haag-Ruelle scattering theory**, and the existence proof is based on the clustering property of W_α^T and an asymptotic estimate for the behavior of the g's for large t and all $\mathbf{x}$ (R. Haag, *Phys. Rev.*, 112 (1958); D. Ruelle, *Helv. Phys. Acta*, 35 (1962)). The out and in states can be interpreted in the limit of infinite future and past as the state where n particles are moving at velocities $\mathbf{v}_j$ related to the spectrum p^μ of P^μ through the relation $p^0 = m_j/(1 - \mathbf{v}_j^2)^{1/2}$, $\mathbf{p} = m_j\mathbf{v}_j/(1 - \mathbf{v}_j^2)^{1/2}$ (with probability amplitude proportional to $h_j(p)$) (H. Araki and R. Haag, *Comm. Math. Phys.*, 4 (1967)).

Since

$$(\Psi^{\text{out}}(h_1 \dots h_n), \Psi^{\text{out}}(h_1' \dots h_{n'}')) = \delta_{nn'} \sum_P \prod_{j=1}^n (h_j, h_{P(j)}'),$$

where the sum is over all permutations P, Ψ^{out} can be viewed as a unitary mapping from the †Fock space $\mathscr{H}^0$ over $\sum_m^\oplus \mathscr{H}_1(m)$ to the closed subspace $\mathscr{H}^{\text{out}}$ spanned by $\Psi^{\text{out}}(h_1 \dots h_n)$, $n = 0, 1, 2, \dots$ (Ω for $n = 0$, h_1 itself for $n = 1$). The free fields on $\mathscr{H}^{\text{out}}$ and $\mathscr{H}^{\text{in}}$ are called (out and in) **asymptotic fields**. Likewise, Ψ^{in} is a unitary mapping from $\mathscr{H}^0$ to $\mathscr{H}^{\text{in}}$. If $\mathscr{H}^{\text{in}} = \mathscr{H}^{\text{out}}$, then the matrix element S can be viewed as a matrix element of a unitary operator on $\mathscr{H}^0$, called an S-**matrix**. Its unitary transform by Ψ^{out} and by Ψ^{in} coincide and define a unitary operator, sometimes called an S-**operator**, on $\mathscr{H}^{\text{in}} = \mathscr{H}^{\text{out}}$. The equality $\mathscr{H} = \mathscr{H}^{\text{in}} = \mathscr{H}^{\text{out}}$ is called the **completeness of the scattering states** or **asymptotic completeness**.

If the scattering states are complete, then the following **LSZ asymptotic condition**, due

to Lehmann, Symanzik, and Zimmermann (*Nuovo Cimento*, 1 (1955)), holds:

$$\operatorname*{weak-lim}_{t\to\pm\infty} Q_j(t,g_j)=a^*_{\substack{\text{out}\\ \text{in}}}(h_j),$$

where $a^*(h)$ is the †creation operator ($a^*(h)\cdot\Psi(h_1\ldots h_n)=\Psi(hh_1\ldots h_n)$ for either out or in states), $\varphi_{\alpha_{(j)}}(f_j)\Omega\in\mathcal{H}_1(m_j)$ is not required, and $(\Omega,\varphi_{\alpha_{(j)}}(f_j)\Omega)=0$ is assumed instead. This leads to an explicit expression for the S-matrix elements in terms of τ-functions, which takes the following simple form if all particles have spin 0, all fields φ_α are scalar ($(D(A)_{\alpha\beta}=\delta_{\alpha\beta})$, and all n_j can be taken to be 1. First define the **connected part** $S_c(h_1\ldots h_n;h'_1\ldots h'_{n'})$ from S by exactly the same equation as truncated Wightman functions with S to be set to 0 if $n=0$, 1 or $n'=0$, 1 except when $n=n'=1$ ($S(h,h')=(h,h')$). Then

$$S_c(h_1\ldots h_l;h_{l+1}\ldots h_n)$$
$$=\int\left[\left(\prod_{j=1}^n(p_j\cdot p_j-m_j^2)\right)\tilde\tau_\alpha^T(p_1,\ldots,p_l,\right.$$
$$\left.-p_{l+1},\ldots,-p_n)\right]\prod_{j=1}^n(-iZ_j^{-1/2}h_j(\mathbf{p}_j)d\Omega_j(\mathbf{p}_j)),$$

where $p_j^0=(\mathbf{p}_j^2+m_j^2)^{1/2}$ (the set of such p is called the positive mass shell; there the vanishing factors $(p_j\cdot p_j-m^2)$ cancel the poles of $\tilde\tau_\alpha^T$), $d\Omega_{m_j}(\mathbf{p})$ is the Lorentz invariant measure $(2|p^0|)^{-1}d^3\mathbf{p}$ $(=\delta(p\cdot p-m_j^2)d^4p)$ on the positive mass shell, $h\in\mathcal{H}_1(m)$ is represented by a measurable function $h(\mathbf{p})$ with the inner product $(h,h')=\int\overline{h(\mathbf{p})}h'(\mathbf{p})d\Omega_m(\mathbf{p})$, Z_j is defined by $(h,\varphi_{\alpha_{(j)}}(f_j)\Omega)=Z_j^{1/2}(h,\tilde f_j)$ for all $h\in\mathcal{H}_1(m_j)$ and $\tilde f(p)=(2\pi)^{-3/2}\int e^{ip\cdot x}f(x)dx$, and $\tilde\tau_\alpha^T$ is the Fourier transform of the truncated **time-ordered function** (**τ-function**) defined by

$$\tilde\tau_\alpha^T(p_1,\ldots,p_n)=\int\tau_\alpha^T(x_1,\ldots,x_n)\exp i\left(\sum_{j=1}^n p_j\cdot x_j\right)\times\prod_{j=1}^n(2\pi)^{-3/2}d^4x_j,$$

$$\tau_\alpha^T(x)=\sum_P\theta(x^0;C_P)W_{P,\alpha}^T(x)\qquad(x=(x_1,\ldots,x_n)),$$

$$W_{P,\alpha}^T(x)=(\Omega,\varphi_{\alpha_{P(1)}}(x_{P(1)})\ldots\varphi_{\alpha_{P(n)}}(x_{P(n)})\Omega)^T,$$

in which the sum is over all permutations P, T indicates the truncated Wightman functions, $\theta(x^0;C_P)$ is the characteristic function of the cone $x_{P(1)}^0\geqslant x_{P(2)}^0\geqslant\ldots\geqslant x_{P(n)}^0$ (the time ordering) possibly smeared out by convolution with a C^∞-function of compact support (so that it can multiply a distribution), where the formula does not actually depend on the smearing functions. (Usually fields φ_α are normalized so that $Z_j=1$.)

The **r-functions** (O. Steinmann, *Helv. Phys. Acta*, 33 (1960); D. Ruelle, *Nuovo Cimento*, 19 (1961); H. Araki, *J. Math. Phys.*, 2 (1961)) are defined by

$$r_i(x)=\sum_P\theta(x^0;C_P/C_i)W_{P,\alpha}^T(x)$$
$$\left(=\sum_P\theta(x^0;C_P/C_i)W_{P,\alpha}(x)\right)$$

(often with an additional factor $(-i)^{n-1}$), where the Fourier-Laplace transform $\tilde\theta(q^0;C_p)$ of $\theta(x^0;C_p)$ is a rational function and $\theta(x^0;C_p/C_i)$ is the inverse Fourier transform of its boundary value as $\operatorname{Im}q^0\in C_i$ tending to 0, and the cone C_i and hence r_i is specified by a consistent choice of signs of $\sum_{j\in I}\operatorname{Im}q_j^0$ for all nonempty proper subsets I of $(1,\ldots,n)$. The Fourier transform of $r_i(x)$ is a boundary value of an analytic function common to all r_i and coincides with $\tilde\tau_\alpha^T$ for $p^0\in C_i$. Making use of this relation, some analyticity properties of the S-matrix (including TCP symmetry) have been proved ([11, 12]; J. Bros, H. Epstein, and V. Glaser, *Comm. Math. Phys.*, 1 (1965); Epstein, Glaser, and A. Martin, *Comm. Math. Phys.*, 13 (1969); Epstein, *J. Math. Phys.*, 8 (1967)).

The two-point function for scalar fields has the following simple expression in terms of †invariant distributions, sometimes called the **Källen-Lehmann representation** with the **Källen-Lehmann weight** ρ:

$$(\Omega,\varphi(x)\varphi(y)\Omega)=\int_0^\infty\Delta_\kappa(x-y)d\rho(\kappa^2).$$

Under the Wightman axiom (with the normal commutation relations), there exists an antiunitary operator Θ, called the **TCP operator**, which satisfies the relations

$$\Theta\Omega=\Omega,\qquad\Theta U(a,A)\Theta^{-1}=U(-a,A),$$
$$\Theta^2=U(0,-1),$$
$$\Theta\varphi_\alpha(x)\Theta^{-1}=(-1)^\eta(-i)^N\varphi_{\bar\alpha}(-x),$$

where η is the number of undotted indices in α, $N=0$ if $D(A)_{\alpha\alpha}=1$ for $A=-1$ (Bose fields), and $N=1$ if $D(A)_{\alpha\alpha}=-1$ for $A=-1$ (Fermi fields).

The choice of ±1 for $\varepsilon(\alpha,\alpha)$ in the locality axiom cannot be opposite to the normal commutation relations for any α except for the trivial case $\varphi_\alpha(x)=0$ ([8–10]; G. Lüders and B. Zumino, *Phys. Rev.*, 110 (1958); N. Burgoyne, *Nuovo Cimento*, 8 (1958)). This is called the **connection of spin and statistics**. For a general choice of ±1 for $\varepsilon(\alpha,\beta)$, the Wightman axioms imply existence of a certain number of **even-oddness conservation laws** (i.e., the existence of a unitary representation u of a group $(Z_2)^l$ for some l such that $u(g)\Omega=\Omega$, $u(g)\varphi_\alpha(f)u(g)^*=\chi_\alpha(g)\varphi_\alpha(f)$ for some characters χ_α on the group) so that the **Klein transforms** $\hat\varphi_\alpha(x)=u(g_\alpha)\varphi_\alpha(x)$ (for some choice of g_α) of the original fields satisfy the Wightman axioms with the normal

commutation relations ([8–10]; H. Araki, *J. Math. Phys.*, 2 (1961)).

E. Theory of Local Observables

Except for the technical assumption of operator-valued distributions, physical contents of the Wightman axioms have been formulated in terms of the von Neumann algebra $\mathscr{A}(\mathcal{O})$ for bounded open sets $\mathcal{O}$, generated by those observables which can be measured in the space-time region $\mathcal{O}$, as follows:
(1) **Isotony**: If $\mathcal{O}_1 \supset \mathcal{O}_2$, then $\mathscr{A}(\mathcal{O}_1) \supset \mathscr{A}(\mathcal{O}_2)$.
(2) **Covariance**: $U(a, A)\mathscr{A}(\mathcal{O})U(a, A)^* = \mathscr{A}(\Lambda(A)\mathcal{O} + a)$.
(3) **Locality**: If $Q_j \in \mathscr{A}(\mathcal{O}_j)$ and $\mathcal{O}_1$ is spacelike to $\mathcal{O}_2$, then $[Q_1, Q_2] = 0$. (No signal can propagate faster than the speed of light.)

In addition, the spectrum condition is as before (the stability of the vacuum) and, since we restrict our attention to the closed span of $\mathscr{A}(\mathcal{O})\Omega$ for all $\mathcal{O}$, Ω is assumed to be cyclic for $\bigcup_{\mathcal{O}} \mathscr{A}(\mathcal{O})$. (Then the latter is irreducible.) By treating $Q(x) = U(x, 1)QU(x, 1)^*$ for $Q \in \mathscr{A}(\mathcal{O})$ as a (noncovariant but localized) field, Haag-Ruelle scattering theory and the analyticity properties of the S-matrix described above for Wightman fields hold in exactly the same way in the theory of local observables. The notion of an algebra of local observables has been a concern of R. Haag since the late 1950s, with the consequent analogy to Wightman fields being demonstrated by Araki in his Zürich lectures of 1961–1962. Hence the above axioms are sometimes called **Haag-Araki axioms**.

With the help of the additional axiom $(\mathscr{A}(\mathcal{O}_1) \cup \mathscr{A}(\mathcal{O}_2))'' = \mathscr{A}(\mathcal{O}_1 \cup \mathcal{O}_2)$, called the **additivity**, the vacuum vector Ω is cyclic and separating for $\mathscr{A}(\mathcal{O})$ for any bounded open $\mathcal{O}$ (**Reeh-Schlieder theorem**) and $\mathscr{A}(\mathcal{O}) = \mathscr{A}(\hat{\mathcal{O}})$, where $\hat{\mathcal{O}}$ is the double cone $\{x \mid |x^0| + |\mathbf{x}| \leqslant L\}$ if $\mathcal{O} = \{x \mid |x^0| + |\mathbf{x}| \leqslant L, |\mathbf{x}| < \varepsilon\}$ for any $\varepsilon > 0$, for example (**Borchers theorem**).

A merit of the Haag-Araki axioms is that these axioms have direct physical interpretation. In particular they always imply the commutativity at spacelike separation of supports, in contrast to the anticommutativity for Fermi fields. This then necessitates the consideration of the representations associated with some other states, such as states with an odd number of fermions and representations of the C^*-algebra $\mathscr{A}$ of quasilocal observables (generated by all $\mathscr{A}(\mathcal{O})$), which are nonequivalent to the vacuum representation and are called **superselection sectors**. This viewpoint was introduced by Haag and D. Kastler (*J. Math. Phys.*, 5 (1964)) and the C^*-algebra version of the Haag-Araki axioms is called the **Haag-Kastler axioms**.

If $\mathcal{O}'$ denotes the causal complement of $\mathcal{O}$ (i.e., the set of all points spacelike to $\mathcal{O}$), then the assumption $\mathscr{A}(\mathcal{O})' = \mathscr{A}(\mathcal{O}')$ is called **duality** and is proved for a certain type of region, which includes the double cone in the case of free fields. With the assumption of the duality for double cones in the vacuum sector, S. Doplicher, Haag, and J. Roberts (*Comm. Math. Phys.*, 13 and 15 (1969); 24 (1971); 35 (1974)) succeeded in the analysis of superselection sectors and clarified the connection of spin and statistics in a much more satisfactory fashion, as well as the anticommutativity of intertwining operators for superselection sectors for Fermi statistics.

F. Constructive Field Theory

An effort to make mathematical sense out of the heuristic theory of quantized fields and to produce examples of Wightman fields and the associated system of local observables has been pushed forward by J. Glimm and A. Jaffe since the mid 1960s and is known as **constructive field theory**. Since 1972, the **Euclidean methods**, already known in some sense, have become extremely powerful central tools, are collectively known as **Euclidean field theory** [13–16].

The Wightman function $W(z_1 \ldots z_n)$ is analytic at the **Schwinger points** $z_j = (ix_j^0, \mathbf{x}_j)$ $(x \in \mathbf{R}^4)$ if $x_j \neq x_k$ for $j \neq k$, and its value $S(x_1 \ldots x_n) = W(z_1 \ldots z_n)$ is called the **Schwinger function**. The axioms for Schwinger functions equivalent to Wightman axioms are known as **Osterwalder-Schrader axioms** (*Comm. Math. Phys.*, 31 (1973); 42 (1975)). The positivity axiom reflecting the positive definite metric of the Hilbert space for Wightman fields is known as **O-S positivity** (or **T-positivity** or **reflection positivity**).

Since the Schwinger function is symmetric in its variables, it can be viewed as the expectation value of the product of (commuting) random fields, called **Euclidean fields**, if an additional positivity holds. This idea was put forward by K. Symanzik in the 1960s. E. Nelson then realized that Euclidean fields for free fields have the Markov property, and he developed Euclidean **Markov field theory**. A work of Guerra in 1972 utilizing **Nelson symmetry** revealed the extreme power of this approach, and the whole of constructive field theory has been studied in Euclidean formulation with remarkable results for super-renormalizable models in space-time of dimension 2 and 3. The key point is the **Feynman-Kac-Nelson formula**, which expresses Schwinger functions

as functional integrals and reveals a mathematical connection between Euclidean field theory and classical statistical mechanics.

G. Gauge Theory

†Electrodynamics in terms of the 4-vector potential $A_\mu(x)$ is invariant under the local gauge transformations $A_\mu(x) \to A_\mu(x) + \partial_\mu \Lambda(x)$ (and the associated transformation of the charged fields) for Λ satisfying the wave equation $\square\Lambda = 0$. This leads on one hand to complication in the canonical quantization of the fields $A_\mu(x)$ and, on the other hand, to the necessity of an indefinite inner-product space as exemplified by the **Gupta-Bleuler formalism** of quantum electrodynamics. In such a formalism, the **physical Hilbert space** (with positive definite metric) is introduced by considering a specific subspace (physical subspace) of a semidefinite metric and taking the quotient by its null subspace.

The corresponding theory with a noncommutative gauge group is known as the theory of **Yang-Mills fields**. In order to restore the formal unitarity of the S-matrix in the perturbation series in terms of Feynman diagrams, L. D. Faddeev and V. N. Popov (*Phys. Lett.*, 25B (1967)) introduced fictitious particles, called **Faddeev-Popov ghosts**. T. Kugo and I. Ojima (*Prog. Theoret. Phys.*, Suppl., 66 (1979)) developed a canonical quantization scheme for the Yang-Mills field which naturally introduces additional fields corresponding to the Faddeev-Popov ghosts and specifies the physical subspace by means of the simple condition that it be the kernel of the generator of **BRS transformations**, earlier introduced by C. Becchi, A. Rouet, and R. Stora.

Gauge theory can be formulated on a lattice of finite volume as a kind of classical statistical mechanics. This is known as **lattice gauge theory**, and the important issue currently being investigated is whether or not its limit, as the lattice interval tends to 0 (the continuum limit) and the volume tends to infinity, produces a nontrivial quantum theory of gauge fields.

References

[1] L. D. Landau and E. M. Lifshits, The classical theory of fields, Pergamon, 1975. (Original in Russian, 1973.)
[2] J. D. Bjorken and S. D. Drell, Relativistic quantum fields, McGraw-Hill, 1965.
[3] S. S. Schweber, An introduction to relativistic quantum field theory, Harper & Row, 1961.
[4] H. Umezawa, Quantum field theory, North-Holland, 1956.
[5] K. Nishijima, Fields and particles, Benjamin, 1969.
[6] N. N. Bogolyubov and D. V. Shirkov, Introduction to the theory of quantized fields, Wiley, third edition, 1980.
[7] K. Hepp, Théorie de la rénormalization, Springer, 1969.
[8] R. F. Streater and A. S. Wightman, PCT, spin and statistics, and all that, Benjamin, 1978.
[9] R. Jost, The general theory of quantized fields, Amer. Math. Soc., 1965.
[10] N. N. Bogolyubov, A. A. Logunov, and I. T. Todorov, Introduction to axiomatic quantum field theory, Benjamin, 1975. (Original in Russian, 1969.)
[11] C. DeWitt and R. Omnes (eds.), Dispersion relations and elementary particles, Wiley, 1960.
[12] M. Chretien and S. Deser (eds.), Axiomatic field theory, Gordon & Breach, 1966.
[13] G. Velo and A. S. Wightman (eds.), Constructive quantum field theory, Springer, 1973.
[14] B. Simon, The $P(\varphi)_2$ Euclidean (quantum) field theory, Princeton Univ. Press, 1974.
[15] J. Glimm and A. Jaffe, Quantum physics, Springer, 1981.
[16] B. Simon, Functional integration and quantum physics, Academic Press, 1979.
[17] A. Jaffe and C. Taubes, Vortices and monopoles, Birkhäuser, 1980.

151 (IV.4) Finite Groups

A. The Number of Finite Groups of a Given Order

A group is called a **finite group** if its order is finite (→ 190 Groups). Since the early years of the theory of finite groups, a major problem has been to find the number of distinct isomorphism classes of groups having a given order. It is almost impossible, however, to find a general solution to the problem unless the values of n are restricted to a (small) subset of the natural numbers. Let $f(n)$ denote the number of isomorphism classes of finite groups of order n. If p is a prime number, then $f(p)=1$ and any group of prime order is a cyclic group. If p is prime, then any group of order p^2 is an †Abelian group and $f(p^2)=2$. If p and q are distinct primes and $p>q$, then $f(pq)=2$ or 1 according as p is congruent to 1 modulo q or not. If $p\equiv 1 \pmod q$, there is a non-Abelian group of order pq as well as a cyclic group of order pq. For small n, the value of $f(n)$ is as

follows:

n	8	12	16	18	20	24	27	28	30	32	60
$f(n)$	5	5	14	5	5	15	5	4	4	51	13

For any n, $f(n)\geqslant 1$. When p is prime, $f(p^m)$ is known for $m\leqslant 6$: $f(p^3)=5$, $f(p^4)=15$ if $p>2$. For $f(p^5)$ see O. Schreier, *Abh. Math. Sem. Univ. Hamburg*, 4 (1926). For $f(2^6)$ see [11]. Set $f(p^m)=p^l$ and $l=Am^3$. Then $A\rightarrow 2/27$ as $m\rightarrow\infty$ (G. Higman, *Proc. London Math. Soc.*, 10 (1960); C. C. Sims, Symposium on Group Theory, Harvard, 1963).

B. Fundamental Theorems on Finite Groups

The following are some fundamental theorems useful in studying finite groups.

(1) The order of any subgroup of a finite group G divides the order of G (J. L. Lagrange). The converse is not necessarily true. If a finite group G contains a subgroup of order n for any divisor n of the order of G, then G is a †solvable group. Furthermore, if G contains a unique subgroup of order n for each divisor n of the order of G, then G is a cyclic group.

Let p be a prime number. Let the order of a finite group G be $p^n m$, where m is not divisible by p. A subgroup of order p^n of G is called a **p-Sylow subgroup** (or simply a **Sylow subgroup**) of G. The importance of this concept may be seen from the next theorem.

(2) A finite group contains a p-Sylow subgroup for any prime divisor p of the order of the group. Furthermore, p-Sylow subgroups are conjugate to each other. The number of distinct p-Sylow subgroups is congruent to 1 modulo p. In general, the number of distinct p-Sylow subgroups of G that contain a given subgroup whose order is a power of p is congruent to 1 modulo p (**Sylow's theorems**).

(3) A p-group is a †nilpotent group (a finite group is called a **p-group** if its order is a power of p). Thus any finite p-group G of order >1 contains a nonidentity element in the †center of G. Furthermore, any proper subgroup of G is different from its †normalizer. A paper by P. Hall (*Proc. London Math. Soc.*, (2) 36 (1933)) is a classic and fundamental work on p-groups.

A group G of order 8 with 2 generators σ, τ and relations $\sigma^4=1$, $\tau\sigma\tau^{-1}=\sigma^{-1}$, $\sigma^2=\tau^2$ is called the **quaternion group**. This group is isomorphic to the multiplicative group consisting of $\{\pm 1, \pm i, \pm j, \pm k\}$ in the †quaternion field. A **generalized quaternion group** is a group of order 2^n with 2 generators σ, τ and relations $\sigma^{2^{n-1}}=1$, $\tau\sigma\tau^{-1}=\sigma^{-1}$, $\tau^2=\sigma^{2^{n-2}}$. A non-Abelian group all of whose subgroups are normal subgroups is called a **Hamilton group**. A Hamilton group is the direct product of a quaternion group, an Abelian group of odd order, and an Abelian group of exponent 2 (i.e., $\rho^2=1$ for each element ρ).

Let G be a finite group, and let $G_0=G\supset G_1\supset\ldots\supset G_r=\{1\}$ be a †composition series of G. The set of isomorphism classes of the simple groups G_{i-1}/G_i, $i=1,2,\ldots,r$, is uniquely determined (up to arrangement) by the †Jordan-Hölder theorem. Thus the two most fundamental problems of finite groups are (i) the study of the simple groups and (ii) the study of a group with a given set of composition factors. The first is one of the leading problems of the theory, although it has been in a state of stagnation until rather recently (→ Section J). As to the second problem, initial works by H. Wielandt and others are under way (particularly in the direction of various generalizations of Sylow's theorems). For the class of finite solvable groups, the first problem has a rather trivial solution; only the second problem is important, and even in this case the theory seems to leave something to be desired.

C. Finite Nilpotent Groups

A finite group is nilpotent if and only if it is the †direct product of its p-Sylow subgroups, where p ranges over all the prime divisors of the order. Any maximal subgroup of a nilpotent group is normal. The converse holds for finite groups; that is, a finite group is nilpotent if and only if all its maximal subgroups are normal.

D. Finite Solvable Groups

One of the most profound results on finite groups, an affirmative answer to the longstanding **Burnside conjecture**, is the **Feit-Thompson theorem** (*Pacific J. Math.*, 13 (1963)): A finite group of odd order is solvable. The index of a maximal subgroup of a finite solvable group is a power of a prime number (E. Galois). But the converse is not true. The unique simple group of order 168 has the property that all maximal subgroups are of prime power index. A finite solvable group contains a self-normalizing nilpotent subgroup (i.e., a nilpotent subgroup H such that $N_G(H)=H$), and any two such subgroups are conjugate (R. W. Carter, *Math. Z.*, 75 (1960); cf. W. Gaschütz, *Math. Z.*, 80 (1963), for a generalization). Such a subgroup is called a **Carter subgroup** and is an analog of a Cartan subalgebra of a Lie algebra. But unlike Cartan subalgebras, most simple groups do not contain any self-normalizing nilpotent subgroups. A finite solvable group of order mn (m, n are relatively prime) contains a subgroup of order m; two subgroups of order m are conjugate; if l

is a divisor of m, then any subgroup of order l is contained in a subgroup of order m (P. Hall). The converse of the first part of this theorem is also true: A finite group is solvable if it contains a subgroup of order m for any decomposition of the order in the form mn, $(m,n)=1$ (P. Hall's solvability criterion). This generalizes the famous **Burnside theorem** asserting the solvability of a group of order $p^a q^b$, where both p and q are prime numbers. If the sequence of the quotient groups of a †principal series of a finite group G consists of cyclic groups, then the group G is called **supersolvable**. A finite group is supersolvable if and only if the index of any maximal subgroup is a prime number (B. Huppert, *Math. Z*, 60 (1954)). If p is the largest prime divisor of the order of a finite supersolvable group G, then a p-Sylow subgroup of G is a normal subgroup.

E. Hall Subgroups

A subgroup is called a **Hall subgroup** if its order is relatively prime to its index (see the theorems of P. Hall on finite solvable groups). There is no general theorem known on the existence of a Hall subgroup. If a finite group G has a normal Hall subgroup N, then G contains a Hall subgroup H that is a complement of N in the sense that $G=NH$ and $N\cap H=1$; furthermore, any two complements are conjugate (**Schur-Zassenhaus theorem**). The analog of Hall's theorem on finite solvable groups fails for nonsolvable groups. But if a finite group, solvable or not, contains a nilpotent Hall subgroup H of order n, then any subgroup of an order dividing n is conjugate to a subgroup of H (Wielandt, *Math. Z.*, 60 (1954); cf. P. Hall, *Proc. London Math. Soc.*, 4 (1954), for a generalization). There are some generalizations of these results for maximal π-subgroups which may not be Hall subgroups.

F. π-Solvable Groups

Let π be a set of prime numbers. Denote the set of prime numbers not in π by π'. A finite group is called a **π-group** if all the prime divisors of the order belong to π. A finite group is called **π-solvable** if any composition factor is either a π'-group or a solvable π-group. If $\pi=\{p\}$ consists of a single prime number p, we use terms such as *p-group* or *p-solvable* (instead of *$\{p\}$-solvable*). Let G be a π-solvable group. A series of subgroups $P_0=1\subset N_0 \subsetneqq P_1 \subsetneqq N_1 \subsetneqq \ldots \subsetneqq P_l \subset N_l=G$ defined by the properties that P_i/N_{i-1} is the maximal normal π-subgroup of G/N_{i-1} and N_i/P_i is the maximal normal π'-subgroup of G/P_i is called the **π-series** of G, and the integer l is called the **π-length** of G. A solvable group is π-solvable for any set π of prime numbers. A π-solvable group contains a Hall subgroup which is a π-group and also a Hall subgroup which is a π'-group; an analog of Hall's theorem on finite solvable groups holds. Hall and Higman (*Proc. London Math. Soc.*, 7 (1956)) discovered deep relations between the p-length of a p-solvable group and invariants of its p-Sylow subgroup. For example, the p-length is 1 if a p-Sylow subgroup is Abelian.

G. Permutation Groups

The set of all permutations on a set Ω of n elements forms a group of order $n!$ whose structure depends only on n. This group is called the **symmetric group of degree** n, denoted by S_n. Any subgroup of S_n is a **permutation group of degree** n. When it is necessary to mention the set Ω on which permutations operate, S_n may be denoted as $S(\Omega)$, and a subgroup of $S(\Omega)$ is called a permutation group on Ω. The set of n elements on which S_n operates is usually assumed to be $\{1,2,\ldots,n\}$, and an element σ of S_n is written as

$$\sigma=\begin{pmatrix}1 & 2 & \ldots & n\\ 1' & 2' & \ldots & n'\end{pmatrix},$$

for example, $\begin{pmatrix}1 & 2 & 3 & 4 & 5 & 6\\ 2 & 3 & 1 & 5 & 4 & 6\end{pmatrix}$,

where i' is the image of i by $\sigma: i'=\sigma(i)$. The element σ may be written as $(\ldots)(abc\ldots z)(\ldots)$, which means that σ cyclically maps a into b, b into c, and so on, and finally z back into a. In the example, $\sigma=(1\,2\,3)(4\,5)(6)$. It is customary to omit the cycle with only one letter in it, such as (6) in the example. With this convention, $\sigma=(1\,2\,3)(4\,5)$ may be an element of S_n for any $n\geqslant 5$, leaving all the letters $i\geqslant 6$ invariant. A **cycle** of length l is an element σ of S_n which moves l letters $a_1,\ldots,a_l$ cyclically and fixes all the rest; i.e., $\sigma=(a_1,\ldots,a_l)$. Then an expression such as $\sigma=(1\,2\,3)(4\,5)$ is the same as the product of two cycles $(1\,2\,3)$ and $(4\,5)$. In general, any permutation can be expressed as the product of mutually disjoint cycles (two cycles $(a_1,\ldots,a_l)$ and $(b_1,\ldots,b_m)$ are said to be disjoint if $a_i\neq b_j$ for all i and j). Furthermore, the expression of the permutation as the product of mutually disjoint cycles is unique up to the order in which these cycles are written. A cycle of length 2 is called a **transposition**. Any permutation may be written as a product of transpositions. This expression is not unique, but the parity of the number of transpositions in the expression is determined by the permutation. A permutation is called **even** if it is the product of an even number of transpositions

and **odd** otherwise. The symmetric group S_n contains the same number of even and odd permutations.

The totality of even permutations forms a normal subgroup of order $(n!)/2(n \geqslant 2)$, called the **alternating group of degree** n and usually denoted by A_n. An even permutation is the product of cycles of length 3. The alternating group A_5 of degree 5 is the nonsolvable group of minimal order. This fact was known to Galois. If $n \neq 4$, then the alternating group A_n is a simple group and a unique proper normal subgroup of S_n. If $n=4$, A_4 contains a normal noncyclic subgroup V of order 4. In this case A_4 and V are the only proper normal subgroups of S_4. A noncyclic group of order 4 is called a **(Klein) four-group**. If $n \geqslant 5$, the symmetric group S_n is not a solvable group. This is the group-theoretic ground for the famous theorem, proved by Ruffini, Abel, and Galois, which asserts the impossibility of an algebraic solution of a general algebraic equation of degree more than four. If $n \leqslant 4$, S_n is solvable. The group S_4 has a composition series with composition factors of orders 2, 3, 2, 2, and S_4 is realized as the group of motions in 3-dimensional space which preserve an octahedron. Hence S_4 is called the **octahedral group**. Similarly, $A_4(A_5)$ is realized as the group of motions in space which preserve a tetrahedron (icosahedron); thus A_4 is called the **tetrahedral group** and A_5 the **icosahedral group**.

These groups have been extensively studied in view of their geometric aspect. The group of motions of a plane which preserve a regular polygon is called a **dihedral group**. If a regular polygon has n sides, then the group has order $2n$. Sometimes a Klein four-group is included in the class of dihedral groups (for $n=2$). The dihedral groups, octahedral group, etc., are called **regular polyhedral groups**. A finite subgroup of the group of motions in 3-dimensional space is either cyclic or one of the regular polyhedral groups. A dihedral group is generated by two elements of order 2. Conversely, a finite group generated by two elements of order 2 is a dihedral group. This simple fact has surprisingly many consequences in the theory of finite groups of even order [15, ch. 9]. A dihedral group of order $2n$ contains a cyclic normal subgroup of order n, and hence is solvable.

If $n \neq 6$, every automorphism of S_n is inner. The order of the group of automorphisms of S_6 is twice the order of S_6. The index $(S_n:H)$ of a subgroup H of S_n is at least equal to n unless $H=A_n$. If $(S_n:H)=n$, then H is isomorphic to S_{n-1}. If $n \neq 6$, S_n contains a unique conjugate class of subgroups of index n. But S_6 contains two such classes, which are exchanged by an automorphism of S_6 (→ Section I).

H. Transitive Permutation Groups

A permutation group G on a set Ω is called a **transitive permutation group** if for any pair (a, b) of elements of Ω, there exists a permutation of G which sends a into b. Otherwise G is said to be **intransitive**. Let G be a transitive permutation group on a set Ω, and let a be an element of Ω. The totality of elements of G which leave a invariant forms a subgroup of G called the **stabilizer** of a (in G). The index of the stabilizer is equal to the number of elements of Ω, the degree of G. Thus the degree of a transitive permutation group G divides the order of G (a fundamental theorem).

The concept of orbits is important. Let G be a permutation group on a set Ω. A subset Γ of Ω is called an **orbit** of G if it is G-invariant and G acts transitively on Γ. In other words, a subset Γ of Ω is an orbit of G if the following two conditions are satisfied: (i) If $a \in \Gamma$ and $g \in G$, the image $g(a)$ also lies in Γ; and (ii) if a and b are two elements of Γ, there exists an element x of G such that $b=x(a)$. Thus each element x of G induces a permutation $\varphi_\Gamma(x)$ on Γ. The set of all these permutations $\varphi_\Gamma(x) (x \in G)$ forms a permutation group on Γ, which may be denoted by $\varphi_\Gamma(G)$. Then $\varphi_\Gamma(G)$ is transitive on Γ, and φ_Γ is a homomorphism of G onto $\varphi_\Gamma(G)$. Thus the number of elements in an orbit Γ is a divisor of the order of G. It is clear that the set Ω on which G acts is the union of mutually disjoint orbits $\Gamma_1, \ldots, \Gamma_r$ of G. This implies that the degree of G is the sum of the numbers of elements in the orbits Γ_i. The resulting equation often contains nontrivial relations. If φ_i denotes the homomorphism φ_{Γ_i} defined before, then G is isomorphic to a subgroup of the direct product of the groups $\varphi_i(G)$, $i=1, 2, \ldots, r$.

A transitive permutation group is called **regular** if the stabilizer of any letter is the identity subgroup $\{1\}$. A transitive permutation group is regular if and only if its order equals its degree. Any group can be realized as a regular permutation group (**Cayley's theorem**). A transitive permutation group which is Abelian is always regular.

Let G be a transitive permutation group on a set Ω. If the stabilizer of an element a of Ω is a maximal subgroup, G is called **primitive**, and otherwise **imprimitive**. A normal subgroup, which is $\neq \{1\}$, of a primitive permutation group is transitive. An imprimitive permutation group induces a decomposition of the set Ω into the union of mutually disjoint subsets $\Delta_1, \ldots, \Delta_s$ $(s>1)$ such that each Δ_i contains at least two elements, and if $x \in G$ maps an element a of Δ_i onto an element b of Δ_j, then x maps every element of Δ_i into Δ_j: $x(\Delta_i)=\Delta_j$. The set $\{\Delta_1, \ldots, \Delta_s\}$ is called a system of im-

primitivity. A subset Δ of Ω is called a **block** if $x(\Delta)\cap\Delta$ equals Δ or the empty set for all x in G. A block is called nontrivial if $\Delta\neq\Omega$ and Δ contains at least two elements. Each member of a system of imprimitivity is a nontrivial block. A transitive permutation group is primitive if and only if there is no nontrivial block.

A permutation group G on a set Ω is called k-**transitive** (or k-**ply transitive**, where k is a natural number) if for two arbitrary k-tuples $(a_1, \ldots, a_k)$ and $(b_1, \ldots, b_k)$ of distinct elements of Ω, there is an element of G which maps a_i into b_i for all $i=1, 2, \ldots, k$. If $k\geqslant 2$, G is called **multiply transitive**. A doubly transitive permutation group is always primitive. The symmetric group S_n of degree n is n-transitive, while the alternating group A_n is $(n-2)$-transitive for $n\geqslant 3$. Conversely, an $(n-2)$-transitive permutation group on $\{1, 2, \ldots, n\}$ is either S_n or A_n. For multiply transitive permutation groups which are simple, see the list in Section I. If $k\geqslant 6$, no k-transitive groups are known at present except S_n and A_n. If the Schreier conjecture (→ Section I) is true, then there are no 6-transitive permutation groups except S_n and A_n (Wielandt, *Math. Z.*, 74 (1960); H. Nagao, *Nagoya J. Math.*, 27 (1964); O'Nan, *Amer. Math. Soc. Notices*, 20 (1973)).

Two 5-transitive permutation groups other than S_n and A_n are known: the groups M_{12} and M_{24} of degrees 12 and 24, respectively, discovered by E. L. Mathieu in 1864 and 1871. The stabilizer of a letter in $M_{12}(M_{24})$ is a 4-transitive permutation group of degree 11 (23), denoted by $M_{11}(M_{23})$. No 4-transitive permutation groups other than S_n, A_n, M_i ($i=11, 12$, 23, and 24) are known. The groups M_{12}, M_{11}, M_{24}, M_{23} and the stabilizer M_{22} of a letter in M_{23} are called **Mathieu groups**. They are simple groups which have quite exceptional properties. For Mathieu groups, see E. Witt, *Abh. Math. Sem. Univ. Hamburg*, 12 (1938). A k-transitive permutation group G on Ω of degree n and order $n(n-1)\ldots(n-k+1)$ has the property that no nonidentity element of G leaves k distinct letters of Ω invariant. If $k\geqslant 4$, such a group is one of the following: S_k, A_{k+2}, M_{12}, and M_{11} (C. Jordan). For $k=2$ and 3, see H. Zassenhaus, *Abh. Math. Sem. Univ. Hamburg*, 11 (1936).

A multiply transitive permutation group G contains a normal subgroup S such that S is a non-Abelian simple group and G is isomorphic to a subgroup of the group Aut S of the automorphisms of S, except when the degree n of G is a power of a prime number and G contains a regular normal subgroup of order n which is an †elementary Abelian group (W. S. Burnside). Furthermore, in these exceptional cases, G is at most 2-transitive if n is odd, while G is at most 3-transitive if n is even but more than 4. The symmetric group S_4 of degree 4 is the only 4-transitive group that contains a proper solvable normal subgroup.

A transitive extension of a permutation group H on Ω is defined as follows. Let ∞ be a new element not contained in Ω. A **transitive extension** G of H is a transitive permutation group on the set $\{\Omega, \infty\}$ in which the stabilizer of ∞ is the given permutation group H on Ω. Transitive extensions do not exist for some H. Suppose that a permutation group H admits a transitive extension G that is primitive. If H is simple, then G is also simple unless the degree of G is a power of a prime number. Constructing transitive extensions has been an effective method for constructing sporadic simple groups.

Permutation groups of prime degree have been studied extensively since the last century, partly because of their connection with algebraic equations of prime degree. Let p be a prime number. A transitive permutation group of degree p is either multiply transitive and nonsolvable or has a normal subgroup of order p with factor group isomorphic to a cyclic group of order dividing $p-1$ (Burnside). Choose two cycles x and y of length p in S_p. If y is not a power of x, then the subgroup $\langle x, y\rangle$ generated by x and y is a multiply transitive permutation group which is simple. The structure of $\langle x, y\rangle$ is not known despite its simple definition. More attention has been paid to groups of degree p, where p is a prime number such that $(p-1)/2=q$ is another prime number. The problem is to decide if such nonsolvable groups contain the alternating group A_p. The Mathieu groups M_{11} and M_{23} are the only known exceptions for $p>7$. The search for additional exceptions has been aided by the development of high-speed computers. It is known that there is no exceptional group of degree $p=2q+1$ for $23<p\leqslant 4079$ (P. J. Nikolai and E. T. Parker, *Math. Tables Aids Comput.*, 12 (1958); see N. Ito, *Bull. Amer. Math. Soc.*, 69 (1963), for further results in this direction).

A **Frobenius group** is a nonregular transitive permutation group in which the identity is the only element leaving more than one letter invariant. A Frobenius group of degree n contains exactly $n-1$ elements which displace all the letters. These $n-1$ elements together with the identity form a regular normal subgroup of order n. This is a theorem of Frobenius; all the existing proofs depend on the theory of characters. The regular normal subgroup of a Frobenius group is nilpotent (J. G. Thompson, *Proc. Nat. Acad. Sci. US*, 45

(1959); [15, ch. 10; 16, ch. 3]). A **Zassenhaus group** is a transitive extension of a Frobenius group.

I. Finite Simple Groups

All simple groups of finite order were completely classified in February 1982 (→ Section J; [23]). These are divided into the following four classes: (1) cyclic groups of prime order, (2) alternating groups of degree $\geqslant 5$, (3) simple groups of Lie type, and (4) other simple groups.

The subclass (1) consists of cyclic groups of prime order p for any prime number p. Abelian simple groups belong to this subclass. The subclass (4) consists of twenty-six sporadic simple groups including five Mathieu groups. All sporadic groups, other than the five Mathieu groups, are of recent discovery.

Simple groups of Lie type are analogs of simple Lie groups, and include the classical groups as well as the exceptional groups and the groups of twisted type.

Classical groups are divided into four types: †linear, †unitary, †symplectic, and †orthogonal (→ 60 Classical Groups). Let $q=p^r$ be a power of a prime number p. Consider a vector space V of dimension $n\geqslant 2$ over the field $\mathbf{F}_q$ of q elements, except in the unitary case where V is a vector space of dimension $n\geqslant 2$ over the field $\mathbf{F}_{q^2}$ of q^2 elements. Let f be a nondegenerate form on V which is †Hermitian in the unitary case (with respect to the automorphism of order 2 of $\mathbf{F}_{q^2}$ over $\mathbf{F}_q$), †skew symmetric bilinear in the symplectic case, and †quadratic in the orthogonal case. In the orthogonal case, the dimension of V is assumed $\geqslant 3$. Consider the group of all linear transformations of V (linear case) of determinant 1, or the group of all linear transformations of determinant 1 which leave the form f invariant (in other cases). In the orthogonal case, take the commutator subgroup. With each of these groups, the factor group of it by its center is a simple group with a few exceptions.

There are several notations to denote these groups. E. Artin's notation for simple groups, which is reasonably descriptive and simple, follows the name of the simple group: n and q are as described in the preceding paragraph, g is the order of the simple group, and (a,b) denotes the greatest common divisor of two natural numbers a and b.

Linear simple group, $L_n(q)$:

$$g=q^{n(n-1)/2}\prod_{i=2}^{n}(q^i-1)/d, \qquad d=(n,q-1).$$

Unitary group, $U_n(q)$:

$$g=q^{n(n-1)/2}\prod_{i=2}^{n}(q^i-(-1)^i)/d, \qquad d=(n,q+1).$$

The structure of unitary groups does not depend on the form.

Symplectic group, $S_n(q)$, $n=2m$:

$$g=q^{m^2}\prod_{i=1}^{m}(q^{2i}-1)/d, \qquad d=(2,q-1).$$

In the symplectic case, the dimension n of the space V must be even, so $n=2m$, and the structure does not depend on the form.

Orthogonal group in odd dimension $n=2m+1$, $O_{2m+1}(q)$:

$$g=q^{m^2}\prod_{i=1}^{m}(q^{2i}-1)/d, \qquad d=(2,q-1).$$

The structure does not depend on the form in odd dimension.

Orthogonal groups in even dimension $n=2m$: There are two inequivalent forms, one with †index m (which is maximal) and the other with index $m-1$. The two orthogonal groups are denoted by $O_{2m}(\varepsilon,q)$, $\varepsilon=\pm 1$, where $\varepsilon=1$ if the form is of maximal index and -1 otherwise. Then

$$g=q^{m(m-1)}(q^m-\varepsilon)\prod_{i=1}^{m-1}(q^{2i}-1)/d, \qquad d=(4,q^m-\varepsilon).$$

The value of ε is determined by the form f: $\varepsilon=1$ if f is equivalent to $\sum_{i=1}^{m}x_{2i-1}x_{2i}$, and $\varepsilon=-1$ if $f\sim x_1^2+\gamma x_1x_2+x_2^2+\sum_{i=2}^{m}x_{2i-1}x_{2i}$, where the polynomial $t^2+\gamma t+1$ is irreducible over $\mathbf{F}_q$.

There are other ways to denote these groups. Let $X=X(*,*)$ be a group of nonsingular linear transformations of a vector space V. Two asterisks indicate two invariants, such as the dimension of V and the number of elements in the ground field. The notation SX stands for the subgroup of X consisting of linear transformations with determinant 1, and the notation PX stands for the factor group of the linear group X by its center. Thus PX is a subgroup of the group of all projectivities of the projective space formed by the linear subspaces of V. The following list is self-explanatory, except the last term in each row, which is the notation of L. E. Dickson [1]:

$L_n(q)=PSL(n,q)=LF(n,q)$

$U_n(q)=PSU(n,q)=HO(n,q^2)$

$S_{2n}(q)=PSp(n,q)=A(2n,q)$.

(LF: linear fractional group; HO: hyperorthogonal group; A: Abelian linear group.) If f is a nondegenerate quadratic form, then the sub-

group of $GL(n,q)$ consisting of all the elements leaving the form f invariant may be denoted by $O(n,q,f)$. Let $\Omega(n,q,f)$ denote the commutator subgroup of $O(n,q,f)$. Set $\varepsilon=1$ if f is of maximal index, and $\varepsilon=-1$ otherwise. Then

$$O_n(\varepsilon,q)=P\Omega(n,q,f).$$

Dickson's notation for orthogonal groups is complicated and seldom used.

Finite simple groups corresponding to Lie groups of some exceptional type were studied by Dickson early in this century, but C. Chevalley (*Tôhoku Math. J.*, (2) 7 (1955)) proved the existence, simplicity, and other properties of groups of any (exceptional) type over any field by a unified method. Simple Lie algebras over the field **C** of complex numbers are completely classified, and according to the classification theory they are in one-to-one correspondence with the †Dynkin diagrams. Let L be a simple Lie algebra (over **C**) corresponding to the †Dynkin diagram of type X (→ 248 Lie Algebras). Let $L=L_0+\sum L_\alpha$ be a Cartan decomposition of L, where α ranges over the †root system Δ of L. It is possible to choose a basis B of L with the following properties (†Chevalley's canonical basis): B consists of $e_\alpha\in L_\alpha(\alpha\in\Delta)$ and a basis of L_0; the structure constants of L with respect to B are all rational integers; the automorphism $x_\alpha(\xi)$ in the †adjoint group defined by

$$x_\alpha(\xi)=\exp(\xi\,\mathrm{ad}\,e_\alpha)\qquad(\xi\in\mathbf{C})$$

maps each element of B into a linear combination of elements of B with coefficients which are polynomials in ξ with integer coefficients. Thus the matrix $A_\alpha(\xi)$ representing the transformation $x_\alpha(\xi)$ with respect to B has coefficients which are polynomials in ξ with integer coefficients.

The elements of B span a Lie algebra $L_{\mathbf{Z}}$ over the ring **Z** of integers. Let F be a field and form $L_F=F\otimes_{\mathbf{Z}}L_{\mathbf{Z}}$. Then L_F is a Lie algebra over F, and the set B may be identified with a basis of L_F over F. Let t be any element of F, $A_\alpha(t)$ be the matrix obtained from $A_\alpha(\xi)$ by replacing the complex variable ξ by the element t, and finally $x_\alpha(t)$ be the linear transformation of L_F represented by the matrix $A_\alpha(t)$ with respect to B. The group generated by the $x_\alpha(t)$ for each root α and each element t of F is called the **Chevalley group** of type X over F. The commutator subgroups of the Chevalley groups are simple, with a few exceptions which will be stated after the complete list of simple groups of Lie type. Suppose that $X=A_n$, D_n, or E_6 (→ 248 Lie Algebras S). Then the Dynkin diagram of type X has a nontrivial symmetry; let it be $\alpha\rightarrow\beta$. Suppose that the field F has an automorphism σ of the same order as the order of the symmetry of the Dynkin diagram. Let θ be the automorphism of the Chevalley group which sends $x_\alpha(t)$ to $x_\beta(t^\sigma)$. Let U (resp. V) be the subgroup of the Chevalley group generated by $x_\alpha(t)$ with $\alpha>0$, $t\in F$ ($x_\beta(t)$, $\beta<0$), and let $U^1(V^1)$ be the subgroup consisting of all the elements of $U(V)$ which are left invariant by θ. The group generated by U^1 and V^1 is called the **group of twisted type**. If the order of σ is i, this group is said to be of twisted type ${}^iX$. In all but one case, the group of twisted type is simple (see R. Steinberg, *Pacific J. Math.*, 9 (1959)). The value i is 2 except when $X=D_4$. Since D_4 admits symmetries of orders 2 and 3, there are two twisted types. If $X=B_2$, G_2, or F_4, then the diagram has a symmetry. If the characteristic p of the ground field F is 2, 3, or 2 according as $X=B_2$, G_2, or F_4 and the field F has an automorphism σ such that $(t^\sigma)^\sigma=t^p$ for any $t\in F$, then a procedure similar to the one described before is applicable, and the group of twisted type X' is obtained (R. H. Ree, *Amer. J. Math.*, 83 (1961)). The group of twisted type is simple if the field F has more than three elements.

The following list contains all the simple groups of Lie type. For each classical group, we list the type followed by identification:

$A_n=L_{n+1}(q)\qquad(n\geqslant1)$

${}^2A_n=U_{n+1}(q)\qquad(n\geqslant1)$

$B_n=O_{2n+1}(q)\qquad(n>1)$

$C_n=S_{2n}(q)\qquad(n>2)$

$D_n=O_{2n}(1,q)\qquad(n>3)$

${}^2D_n=O_{2n}(-1,q)\qquad(n>3)$

For other groups the type of the group is followed by the customary name or notation, if any, and the order g:

B_2' **Suzuki group**, $Sz(q)$, $q=2^{2n+1}$

$g=q^2(q-1)(q^2+1)$

${}^3D_4$ $g=q^{12}(q^8+q^4+1)(q^6-1)(q^2-1)$

G_2 $g=q^6(q^6-1)(q^2-1)$

G_2' **Ree group**, $\mathrm{Re}(q)$, $q=3^{2n+1}$

$g=q^3(q^3+1)(q-1)$

F_4 $g=q^{24}(q^{12}-1)(q^8-1)(q^6-1)\times(q^2-1)$

F_4' $q=2^{2n+1}$ $g=q^{12}(q^6+1)(q^4-1)\times(q^3+1)(q-1)$

E_6 $dg=q^{36}(q^{12}-1)(q^9-1)(q^8-1)\times(q^6-1)(q^5-1)(q^2-1)$

$d=(3,q-1)$

$^2E_6$ $dg=q^{36}(q^{12}-1)(q^9+1)(q^8-1)$
$\times(q^6-1)(q^5+1)(q^2-1)$
$d=(3,q+1)$

E_7 $dg=q^{63}(q^{18}-1)(q^{14}-1)(q^{12}-1)$
$\times(q^{10}-1)(q^8-1)(q^6-1)(q^2-1)$
$d=(2,q-1)$

E_8 $g=q^{120}(q^{30}-1)(q^{24}-1)(q^{20}-1)$
$\times(q^{18}-1)(q^{14}-1)(q^{12}-1)$
$\times(q^8-1)(q^2-1)$.

B'_2: M. Suzuki, *Proc. Nat. Acad. Sci. US*, 46 (1960); G'_2 and F'_4: Ree, *Amer. J. Math.*, 83 (1961): G_2: Dickson, *Trans. Amer. Math. Soc.*, 2 (1901), *Math. Ann.*, 60 (1905); other Chevalley groups: Chevalley, *Tôhoku Math. J.*, (2) 7 (1955); twisted types: Steinberg, *Pacific J. Math.*, 9 (1959), J. Tits, *Séminaire Bourbaki* (1958), *Publ. Math. Inst. HES* (1959), D. Hertzig, *Amer. J. Math.*, 83 (1961), *Proc. Amer. Math. Soc.*, 12 (1961).

Nonsimple cases: $L_2(2)$, $L_2(3)$, $U_3(2)$, and $Sz(2)$ are solvable groups of orders 6, 12, 72, and 20, respectively. The groups $O_5(2)$, $G_2(2)$, $G'_2(3)$, and $F'_4(2)$ contain normal subgroups of indices 2, 2, 3, and 2, respectively. These normal subgroups are simple and identified as $L_2(9)$, $U_3(3)$, $L_2(8)$ in the first three cases. The normal subgroup of $F'_4(2)$ is not in the list of simple groups of Lie type and is quite exceptional (**Tits's simple group**, *Ann. Math.*, (2) 80 (1964)).

Isomorphisms between various simple groups: $L_2(q)=U_2(q)=S_2(q)=O_3(q)$; $O_5(q)=S_4(q)$; $O_4(1,q)=L_2(q)\times L_2(q)$; $O_4(-1,q)=L_2(q^2)$; $O_6(1,q)=L_4(q)$; $O_6(-1,q)=U_4(q)$; $O_{2n+1}(q)=S_{2n}(q)$ if q is a power of 2; $L_2(2)=S_3$; $L_2(3)=A_4$; $L_2(4)=L_2(5)=A_5$; $L_2(7)=L_3(2)$; $L_2(9)=A_6$; $L_4(2)=A_8$; $U_4(2)=S_4(3)$. If q is odd and $2n\geqslant 6$, then $S_{2n}(q)$ and $O_{2n+1}(q)$ have the same order but are not isomorphic. $L_3(4)$ and $L_4(2)$ have the same order but are not isomorphic. There is no other isomorphism or coincidence of orders among the known simple groups (Artin, *Comm. Pure Appl. Math.*, 8 (1955)).

The groups of the automorphisms of simple groups belonging to subclasses (1), (2), and (3) are known. For the simple groups of Lie type, see Steinberg, *Canad. J. Math.*, 10 (1960).

The following list of twenty-six groups consists of all the simple groups that belong to class (4):

Five Mathieu groups whose orders are

$M_{11}: g=7{,}920=2^4\cdot 3^2\cdot 5\cdot 11$

$M_{12}: g=95{,}040=2^6\cdot 3^3\cdot 5\cdot 11$

$M_{22}: g=443{,}520=2^7\cdot 3^2\cdot 5\cdot 7\cdot 11$

$M_{23}: g=10{,}200{,}960=2^7\cdot 3^2\cdot 5\cdot 7\cdot 11\cdot 23$

$M_{24}: g=244{,}823{,}040=2^{10}\cdot 3^3\cdot 5\cdot 7\cdot 11\cdot 23$.

The other twenty-one groups have been discovered since 1964. Each group is identified by the symbol $(x)_i$, indicating that it is the ith group discovered in the year 19x. The list continues with the name or names of discoverers, the order of the group, and a brief description.

$(64)_1$: Z. Janko, $g=175{,}560=2^3\cdot 3\cdot 5\cdot 7\cdot 11\cdot 19$. A subgroup of the Chevalley group $G_2(11)$. See *J. Algebra*, 3 (1966).

$(67)_1$: M. Hall and Z. Janko, $g=604{,}800=2^7\cdot 3^3\cdot 5^2\cdot 7$, a transitive extension of $U_3(3)$ of degree 100.

$(67)_2$: D. G. Higman and Sims, $g=44{,}352{,}000=2^9\cdot 3^2\cdot 5^3\cdot 7\cdot 11$, a transitive extension of M_{22} of degree 100. It is a normal subgroup of index 2 in the group of automorphisms of a certain graph with 100 vertices.

$(67)_3$: Suzuki, $g=448{,}345{,}497{,}600=2^{13}\cdot 3^7\cdot 5^2\cdot 7\cdot 11\cdot 13$, a transitive extension of $G_2(4)$; defined from the automorphism group of a graph of 1782 vertices.

$(67)_4$: J. McLaughlin, $g=898{,}128{,}000=2^7\cdot 3^6\cdot 5^3\cdot 7\cdot 11$, a transitive extension of $U_4(3)$; defined from a graph of 275 vertices.

$(68)_1$: G. Higman, Z. Janko, and J. McKay, $g=50{,}232{,}960=2^7\cdot 3^5\cdot 5\cdot 17\cdot 19$, a transitive extension of the group which is obtained from $L_2(16)$ by adjoining the field automorphism of order 2. The existence was verified by using a computer.

$(68)_2$, $(68)_3$, $(68)_4$: J. H. Conway,

$g=2^{21}\cdot 3^9\cdot 5^4\cdot 7^2\cdot 11\cdot 13\cdot 23$

$=4{,}157{,}776{,}806{,}543{,}360{,}000$,

$g=2^{18}\cdot 3^6\cdot 5^3\cdot 7\cdot 11\cdot 23$,

$g=2^{10}\cdot 3^7\cdot 5^3\cdot 7\cdot 11\cdot 23$.

The big group is obtained from the automorphism group of a lattice in 24-dimensional space, and the two smaller ones are subgroups of it. The lattice was defined by J. Leech in connection with a problem of close packing of spheres in 24 dimensions (*Canad. J. Math.*, 19 (1967)).

$(68)_5$: B. Fischer, $g=2^{17}\cdot 3^9\cdot 5^2\cdot 7\cdot 11\cdot 13=70{,}321{,}751{,}654{,}400$, a transitive extension of $U_6(2)$ derived by means of a certain graph.

$(69)_1$: D. Held and others, $g=2^{10}\cdot 3^3\cdot 5^2\cdot 7^3\cdot 17=4{,}030{,}387{,}200$.

$(69)_2$: B. Fischer, $g=2^{18}\cdot 3^{13}\cdot 5^2\cdot 7\cdot 11\cdot 13\cdot 17\cdot 23=4{,}089{,}470{,}473{,}293{,}004{,}800$.

$(69)_3$: B. Fischer, $g=2^{21}\cdot 3^{16}\cdot 5^2\cdot 7^3\cdot 11\cdot 13\cdot 17\cdot 23\cdot 29=1{,}255{,}205{,}709{,}190{,}661{,}721{,}292{,}800$.

$(71)_1$: R. N. Lyons and C. C. Sims, $g=2^8\cdot 3^7\cdot 5^6\cdot 7\cdot 11\cdot 31\cdot 37\cdot 67$.

The existence of $(69)_1$ and $(71)_1$ was verified by

using computers; $(69)_2$ and $(69)_3$ were derived by means of certain graphs. For a more detailed account of these simple groups, see J. Tits, *Séminaire Bourbaki* (1970), No. 375, and the references [18, 19, 20.]

$(72)_1$: A. Rudvalis, $g=2^{14}\cdot3^3\cdot5^3\cdot7\cdot13\cdot19$, a transitive extention of Tits's simple group, i.e., a normal subgroup of $F_4'(2)$ of index 2. Concerning this group $(72)_1$, see the article of J. H. Conway and D. B. Wales, *J. Algebra*, 27 (1973).

$(73)_1$: M. O'Nan, $g=2^9\cdot3^4\cdot5\cdot7^3\cdot11\cdot19\cdot31$. This group $(73)_1$ was discovered by O'Nan and the existence was verified by C. Sims, using a computer.

$(73)_2$: B. Fischer, $g=2^{41}\cdot3^{13}\cdot5^6\cdot7^2\cdot11\cdot13\cdot19\cdot23\cdot31\cdot47$.

$(73)_3$: B. Fischer and R. Griess, $g=2^{46}\cdot3^{20}\cdot5^9\cdot7^6\cdot11^2\cdot13^3\cdot17\cdot19\cdot23\cdot29\cdot31\cdot41\cdot47\cdot59\cdot71$.

$(74)_1$: J. G. Thompson, $g=2^{15}\cdot3^{10}\cdot5^3\cdot7^2\cdot13\cdot19\cdot31$.

$(74)_2$: K. Harada, $g=2^{14}\cdot3^6\cdot5^6\cdot7\cdot11\cdot19$.

The existence of $(73)_2$ was suggested by B. Fischer, and then that of $(73)_3$ by B. Fischer and R. Griess. Shortly after this, the existence of $(74)_1$ and $(74)_2$ was suggested by J. G. Thompson, and then Thompson and Harada proved the existence of these groups with the aid of P. Smith and S. Norton, using computers. The existence of $(73)_2$ was established in 1976 by Leon and Sims, using a computer. Very recently (July 1980), R. Griess has announced that the group $(73)_3$ is realized as a group of automorphisms of a 196,883-dimensional commutative nonassociative algebra over the rational numbers.

$(75)_1$: Z. Janko, $g=2^{21}\cdot3^3\cdot5\cdot7\cdot11^3\cdot23\cdot29\cdot31\cdot37\cdot43$. This group $(75)_1$ was discovered by Z. Janko. The existence was verified by using a computer.

For a more detailed account of the twenty-six sporadic simple groups, see [21].

Simple groups of order <1000 are A_5 $(g=60)$, $L_2(7)$ (168), $A_6(360)$, $L_2(8)$ (504), and $L_2(11)$ (660). All simple groups of order $\leqslant 20{,}000$ are known.

Among the known simple groups, the following multiply transitive permutation representations are known: Alternating groups A_n (degree n), A_8 and A_7 (degree 15), A_6 (degree 10), A_5 (degree 6), Mathieu groups M_i (degree i), M_{11} (degree 12), $L_n(q)$ (degree $(q^n-1)/(q-1)$), $L_2(p)$ (degree p for $p=5,7,11$), $U_3(q)$ (degree $1+q^3$), $Sz(q)$ (degree $1+q^2$), Re (3^n) (degree $1+3^{3n}$), $S_{2n}(2)$ (degrees $2^{n-1}(2^n\pm1)$), and the Higman-Sims group $(67)_2$ (degree 176) and, the Conway group $(68)_4$ (degree 276). Among them, A_n (degree n, $n\geqslant5$), M_i (degree i), M_{11} (degree 12), $L_2(2^m)$ (degree $1+2^m$), and $L_2(5)$ (degree 5) are triply transitive.

There are several remarkable properties of known finite simple groups which have been conjectured to hold for arbitrary finite simple groups. One of the most famous is the **Schreier conjecture**, which asserts that the group of outer automorphisms of a simple group is solvable. This has been verified for all known cases. Another conjecture says that a finite simple group is generated by two elements. This has also been verified for almost all known groups. In many cases, there is a generating set of two elements, one of which has order 2. There is no counterexample known to disprove the universal validity of this property. Except for $Sz(q)$, the orders of known simple groups are divisible by 12.

J. Classification of Finite Simple Groups

The objective of classification theory is to find the complete list of finite simple groups; this was accomplished in February 1982, following the series of important works mentioned below.

The order of a finite non-Abelian simple group is divisible by at least three distinct prime numbers (W. S. Burnside; → Section D). The order of a finite non-Abelian simple group is even (W. Feit and J. G. Thompson; → Section D). These theorems are special cases of the following theorem: If G is a finite non-Abelian simple group in which the normalizer of any solvable subgroup $\neq\{1\}$ is solvable, then $G=L_2(q)(q>3)$, $Sz(2^{2n+1})(n\geqslant1)$, A_7, $L_3(3)$, $U_3(3)$, M_{11}, or Tits's simple group. In particular, a minimal simple group is isomorphic to $L_2(p)$ ($p\equiv2$ or 3 (mod 5), $p>3$), $L_2(2^p)$, $L_2(3^p)$, $Sz(2^p)$, or $L_3(3)$, where p is a prime number (a finite non-Abelian simple group is called a minimal simple group if all proper subgroups are solvable). This theorem is proved in a series of papers by J. G. Thompson (*Bull. Amer. Math. Soc.*, 74 (1968), *Pacific J. Math.*, 33 (1970), 39 (1971), 48 (1973), 50 (1974), and 51 (1974)). The method which Thompson used in these papers has since been generalized in various ways by many authors to establish a number of important theorems. There are also some interesting consequences of this theorem concerning solvable groups. For example, a finite group is solvable if and only if every pair of elements generates a solvable subgroup.

Let G be a non-Abelian simple group of even order and S one of its 2-Sylow subgroups. Then S is neither cyclic nor a generalized quaternion group (W. S. Burnside, R. Brauer, and M. Suzuki). If S is a dihedral group, then $G=A_7$ or $L_2(q)$ (q odd $\geqslant5$) (D. Gorenstein and J. H. Walter). These theorems deal with the

cases where S is "small." The study in this direction has culminated in the classification of finite simple groups all of whose 2-subgroups are generated by at most four elements (D. Gorenstein and K. Harada, *Mem. Amer. Math. Soc.*, 147 (1974)).

If a 2-Sylow subgroup of a finite non-Abelian simple group G is an Abelian group, then $G = L_2(r)$ ($r \equiv 0$, 3 or 5 (mod 8), $r > 3$), or else G possesses an element of order 2 whose centralizer is isomorphic to $\mathbf{Z}_2 \times L_2(q)$ ($q \equiv 3$ or 5 (mod 8), $q > 3$) (J. H. Walter). In the latter case, G is called **a group of Janko-Ree type** (J-R type for short), and if $q \neq 5$, it is called a **group of Ree type**. If $q = 5$, then $G = (64)_1$, the Janko's simple group of order 175,560. The Ree groups $\mathrm{Re}(q)$ are groups of Ree type. Since the discovery of $\mathrm{Re}(q)$, it has long been an open problem to show that there are no other groups of Ree type. Very recently, the combined work of E. Bombiere, J. G. Thompson and others settled the problem (*Inventiones Math.*, 58 (1980)).

A surprisingly short proof of the Walter's theorem above is given by H. Bender (*Math. Z.*, 117 (1970)). Bender's method applies to a much larger class of groups. A subgroup A of a finite group G is said to be **strongly closed** if $A^g \cap N_G(A) \subset A$ for each $g \in G$. D. M. Goldschmidt proved (*Ann. Math.*, 99 (1974)) that if A is a strongly closed Abelian 2-subgroup, then the subgroup G_0 generated by the conjugates of A possesses a normal series $G_0 \supset G_1 \supset G_2$ with the properties: G_2 is of odd order, G_1/G_2 is a 2-group and is contained in the center of G_0/G_2, and either $G_0 = G_1$ or G_0/G_1 is the direct product of simple groups on the following list: $L_2(q)$ ($q \equiv 0$, 3 or 5 (mod 8), $q > 3$), $Sz(2^{2n-1})$, $U_3(2^n)$ ($n > 1$), and the groups of J-R type. Furthermore, $AG_2/G_2 \supset G_1/G_2$ and AG_1/G_1 is the center of a 2-Sylow subgroup of G_0/G_1. This theorem generalizes an earlier result of G. Glauberman (the so-called Z^*-theorem) which states that if A is a strongly closed subgroup of order 2, then the image of A in the quotient group G/K by the maximal normal subgroup of odd order is a normal subgroup and hence is contained in the center of G/K. These theorems of Glauberman and Goldschmidt are of fundamental importance in the study of finite simple groups since they provide an effective tool for showing that a given group has a normal Abelian 2-subgroup.

Glauberman obtained a criterion for the existence of a strongly closed Abelian p-subgroup $\neq \{1\}$ for some prime p [15, 22]. For any prime p, the quadratic group $Qd(p)$ is defined to be the semidirect product of the 2-dimensional vector space $V(2,p)$ over the field of p elements by the special linear group $SL(2,p)$ where the action of $SL(2,p)$ on $V(2,p)$ is taken to be the natural one. A finite group G contains a strongly closed Abelian p-subgroup $A \neq \{1\}$ if no section of G is isomorphic to $Qd(p)$ (a section is a quotient group of a subgroup). Furthermore, if p is odd, then we can choose as A a characteristic subgroup of a p-Sylow subgroup S of G which is determined only by the structure of S. Therefore a finite non-Abelian simple group has a section isomorphic to $Qd(2) = S_4$ except when it is one of the simple groups mentioned in Goldschmidt's theorem. This theorem generalizes an unpublished result of J. G. Thompson to the effect that 3 divides the order of finite non-Abelian simple groups except $Sz(2^{2n+1})$.

Let G be a 2-transitive permutation group on $n+1$ letters, and assume that the stabilizer H of a letter contains a normal subgroup K which is regular on the remaining n letters. Then G contains a normal subgroup N such that G is isomorphic to a subgroup of the automorphism group of N and either $N = L_2(q)$, $U_3(q)$, $Sz(q)$ or a group of Ree type, or else N is 2-transitive on the $n+1$ letters and no nonidentity element of N leaves two distinct letters invariant (the structure of N in the latter case is also known (H. Zassenhaus; → Section H). This theorem is proved by E. Shult for n even (*Illinois J. Math.*, 16 (1972)) and by C. Hering, W. M. Kantor, and G. M. Seitz for n odd (*J. Algebra*, 20 (1972)). Its proof depends on the work of many authors who considered various special cases, especially the work of H. Zassenhaus, W. Feit, N. Ito, and M. Suzuki on the classification of Zassenhaus groups, and the work of M. Suzuki on the case where n is even and H/K is of odd order (*Ann. Math.*, 79 (1964)). In this special case which Suzuki handles, the stabilizer H is of even order and $H \cap H^g$ is of odd order for any $g \in G - H$. If a proper subgroup H of an arbitrary finite group G has this property, then H is called a **strongly embedded subgroup**. Extending the work of Suzuki, H. Bender proved the following theorem (*J. Algebra*, 17 (1971)): If a finite group G possesses a strongly embedded subgroup, then either (i) a 2-Sylow subgroup of G is cyclic or a generalized quaternion group, or (ii) G possesses a normal series $G = G_0 \supset G_1 \supset G_2$ such that G_0/G_1 and G_2 are of odd order and $G_1/G_2 = L_2(2^n)$, $U_3(2^n)$, or $Sz(2^{2n-1})$ ($n > 1$). This theorem generalizes another theorem of Suzuki who reached the same conclusion under the assumption that two distinct 2-Sylow subgroups have only the identity element in common. Bender's theorem is of fundamental importance in the classification theory of finite simple groups, since a strongly embedded subgroup often appears as an obstacle to the proofs of classification

theorems. For a generalization of Bender's theorem, see a paper by M. Aschbacher (*Proc. Amer. Math. Soc.*, 38 (1973)), which also contains an alternative proof of Shult's theorem.

The theorem of Shult, Hering, Kantor, and Seitz may be interpreted as a classification of finite groups having a split (B, N)-pair of rank 1. Let G be a finite group and let B and N be subgroups of G such that (i) B and N generate G, (ii) $T = B \cap N$ is a normal subgroup of N, and (iii) $W = N/T$ is generated by a set S of elements of order 2 such that $sBs \neq B$ and $sBw \subset BwB \cup BswB$ for each $s \in S$ and each $w \in W$. The subgroups B and N are called a (B, N)-**pair** of G (the quadruplet (G, B, N, S) is called a **Tits system**), and the cardinality of the set S is called the rank of the (B, N)-pair. The (B, N)-pair is said to be split if B has a normal subgroup U such that $B = TU$ and $T \cap U = \{1\}$, and is said to be saturated if $T = \bigcap_{n \in N} B^n$. If a finite group G has a split saturated (B, N)-pair of rank 1, and if $Z = \bigcap_{g \in G} B^g$, then G/Z is a 2-transitive permutation group satisfying the assumption of the theorem of Shult, Hering, Kantor, and Seitz, and information is obtained on the structure of G. In general, the simple groups of Lie type are characterized as simple groups with certain (B, N)-pairs. For (B, N)-pairs of rank 2, see papers by P. Fong and G. M. Seitz (*Inventiones Math.*, 21 (1973), 24 (1974)). J. Tits has developed a satisfactory theory on finite groups having a (B, N)-pair of rank at least 3 (Lecture notes in math. 386, Springer).

Let G be a finite group generated by a conjugate class D of elements of order 2, and let π be the set of positive integers consisting of the orders of the products of two distinct elements of D. Furthermore, assume that G has no nontrivial solvable normal subgroup. B. Fischer proved (*Inventiones Math.*, 13 (1971)) that if $\pi = \{2, 3\}$, then G contains a normal subgroup isomorphic to one of the following groups: A_n, $S_{2n}(2)$, $O_{2n}(\pm 1, 2)$, $O_n(\pm 1, 3)$, $U_n(2)$, and the three Fischer's simple groups $(68)_5$, $(69)_2$, $(69)_3$. For a generalization of this theorem, see papers by M. Aschbacher (*Math. Z.*, 127 (1972), *J. Algebra*, 26 (1973)). The most powerful result in this direction is given by F. Timmesfeld (*J. Algebra*, 33 (1975), 35 (1975)): Suppose π consists of 2, 4, and odd positive integers. Furthermore, assume that if d and e are in D and de is of order 4, then $(de)^2 \in D$. Then $G = A_6$, $U_3(3)$, the Hall-Janko group $(67)_1$, $L_n(q)(n \geqslant 3)$, $O_{2n+1}(q)(n \geqslant 3)$, $O_{2n}(\pm 1, q)(n \geqslant 4)$, $G_2(q)$, ${}^3D_4(q)$, $F_4(q)$, ${}^2E_6(q)$, $E_6(q)$, $E_7(q)$, or $E_8(q)$, where $q = 2^m$.

Let G be a simple group of even order and let H be the centralizer of an element of order 2. Then the order of G is bounded by a function f of the order h of H (for example, we can choose $f(h) = \{h(h+1)\}!$; R. Brauer and K. A. Fowler, *Ann. Math.*, 62 (1955)). In particular, there exist only finitely many isomorphism classes of finite simple groups which contain an element of order 2 with a given centralizer H. This fact is a ground for Brauer's program of studying simple groups of even order in terms of the structure of the centralizers of elements of order 2. There are a number of important results concerning Brauer's program [20]. For example, nine sporadic simple groups were discovered in related works. Since 1973, Brauer's program has been improved greatly by M. Aschbacher, D. Gorenstein, and others, and the classification of finite simple groups was finally completed in February 1982 [21, 23].

References

[1] L. E. Dickson, Linear groups with an exposition of the Galois field theory, Teubner, 1901 (Dover, 1958).
[2] W. Burnside, Theory of groups of finite order, Cambridge Univ. Press, second edition, 1911.
[3] H. Zassenhaus, Lehrbuch der Gruppentheorie, Teubner, 1937; English translation, The theory of groups, Chelsea, 1958.
[4] A. Speiser, Die Theorie der Gruppen von endlicher Ordnung, Springer, third edition, 1937.
[5] W. Specht, Gruppentheorie, Springer, 1956.
[6] M. Suzuki, Structure of a group and the structure of its lattice of subgroups, Erg. Math., Springer, 1956.
[7] H. S. M. Coxeter and W. O. J. Moser, Generators and relations for discrete groups, Erg. Math., Springer, 1957.
[8] M. Hall, The theory of groups, Macmillan, 1959.
[9] C. W. Curtis and I. Reiner, Representation theory of finite groups and associative algebras, Interscience, 1962.
[10] W. R. Scott, Group theory, Prentice-Hall, 1964.
[11] M. Hall and J. K. Senior, The groups of order 2^n $(n \leqslant 6)$, Macmillan, 1964.
[12] H. Wielandt, Finite permutation groups, Academic Press, 1964.
[13] J. Dieudonné, Sur les groupes classiques, Actualités Sci. Ind., Hermann, 1948.
[14] J. Dieudonné, La géométrie des groupes classiques, Erg. Math., Springer, 1955.
[15] D. Gorenstein, Finite groups, Harper & Row, 1968. Second edition, Chelsea, 1980.
[16] D. Passman, Permutation groups, Benjamin, 1968.

[17] B. Huppert, Endliche Gruppen, Springer, 1967.
[18] R. Brauer and C. H. Sah (eds.), Theory of finite groups, a symposium, Benjamin, 1969.
[19] M. B. Powell and G. Higman, (eds.), Finite simple groups, Academic Press, 1971.
[20] W. Feit, The current situation in the theory of finite simple groups, Actes Congr. Intern. Math., 1970, Nice, Gauthier-Villars, vol. 1., p. 55–93.
[21] D. Gorenstein, The classification of finite simple groups I, Bull. Amer. Math. Soc., 1 (1979), 43–199.
[22] G. Glaubermann, Factorization in local subgroups of finite groups, Regional conference series in mathematics, no. 33 (1977).
[23] D. Gorenstein, Finite simple groups, Plenum, 1982.

152 (VII.18) Finsler Spaces

A. Definitions

Let $T(M)$ be the †tangent vector bundle of an n-dimensional †differentiable manifold M. An element of $T(M)$ is denoted by (x, y), where x is a point of M and y is a †tangent vector of M at x. Given a †local coordinate system $(x^1, \dots, x^n)$ of M, we can obtain a local coordinate system of $T(M)$ by regarding $(x^1, \dots, x^n, y^1, \dots, y^n) = (x^i, y^j)$ as coordinates of the pair $(x, y) \in T(M)$, where $(x^1, \dots, x^n)$ are coordinates of a point x of M and $y = \sum y^j \partial/\partial x^j$. A continuous real-valued function $L(x, y)$ defined on $T(M)$ is called a **Finsler metric** if the following conditions are satisfied: (i) $L(x, y)$ is differentiable at $y \neq 0$; (ii) $L(x, \lambda y) = |\lambda| L(x, y)$ for any element (x, y) of $T(M)$ and any real number λ; and (iii) if we put $g_{ij}(x, y) = (1/2)\partial^2 L(x, y)^2/\partial y^i \partial y^j$, the symmetric matrix $(g_{ij}(x, y))$ is positive definite. A differentiable manifold with a Finsler metric is called a **Finsler space**. There exists a Finsler metric on a manifold M if and only if M is †paracompact. We call $F(x, y) = L(x, y)^2$ the **fundamental form** of the Finsler space. When $F(x, y)$ is a quadratic form of $(y^1, \dots, y^n)$, $L(x, y)$ is a †Riemannian metric, and $F(x, y) = \sum_{i,j} g_{ij}(x) y^i y^j$. Therefore a Finsler space is a Riemannian space if and only if g_{ij} does not depend on y. The matrix g_{ij} is also called the **fundamental tensor** of the Finsler space ($i, j = 1, \dots, n$).

Thus the notion of a Finsler metric is an extension of that of Riemannian metric. The study of differentiable manifolds utilizing such generalized metrics was considered by B. Riemann, but he stated that a Riemannian metric is more convenient for the purpose since "only nongeometrical results can be obtained" by using Finsler metrics [7]. P. Finsler initiated the systematic study of Finsler metrics and extended to a Finsler space many concepts and theorems valid in the classical theory of curves and surfaces [5].

B. The Finsler Metric

In a Finsler space, the arc length of a curve $x = x(t)$ $(a \leqslant t \leqslant b)$ is given by $\int_a^b L(x, dx/dt)\,dt$. Therefore a †geodesic in a Finsler space is defined as a †stationary curve for the problem of †variation $\delta \int_a^b L(x, dx/dt)\,dt = 0$, and the differential equation of the geodesic is given by

$$\frac{d^2x^i}{dt^2} + \sum_{j,k} \gamma^i_{jk}\left(x, \frac{dx}{dt}\right)\frac{dx^j}{dt}\frac{dx^k}{dt} = 0,$$

where $\gamma^i_{jk}(x, y)$ is the †Christoffel symbol of g_{ij}, i.e.,

$$\gamma^i_{jk}(x, y) = \frac{1}{2}\sum_a g^{ia}\left(\frac{\partial g_{ja}}{\partial x^k} + \frac{\partial g_{ak}}{\partial x^j} - \frac{\partial g_{jk}}{\partial x^a}\right),$$

where $(g^{ij}(x, y))$ is the inverse matrix of $(g_{ij}(x, y))$.

The distance between two points in a Finsler space is defined, as in a Riemannian space, as the infimum of the lengths of curves joining the two points. Many properties of Riemannian spaces as metric spaces can be extended to Finsler spaces. The topology defined by the Finsler metric coincides with the original topology of the manifold. A Finsler space M is said to be †complete if every Cauchy sequence of M as a metric space is convergent. The following three conditions are equivalent: (i) M is complete; (ii) each bounded closed subset of M is compact; (iii) each geodesic in M is infinitely extendable. In a complete Finsler space, any two points can be joined by the shortest geodesic.

A diffeomorphism φ of a Finsler space M preserves the distance between an arbitrary pair of points if and only if the transformation on $T(M)$ induced by φ preserves the Finsler metric $L(x, y)$. Such a transformation is called an †isometry of the Finsler space. In the †compact-open topology the set of all isometries of a Finsler space is a †Lie transformation group of dimension at most $n(n+1)/2$. If a Finsler space admits the isometry group of dimension greater than $(n(n-1)+2)/2$, it is a Riemannian space of constant curvature [8].

C. The Theory of Connections

An important difference between a Finsler space and a Riemannian space relates to their

properties with respect to the theory of †connections. In the case of a Riemannian space, the Christoffel symbols constructed from the fundamental tensor are exactly the coefficients of a connection, whereas in the case of a Finsler space, the Christoffel symbols $\gamma^i_{jk}(x, y)$ do not define a connection, for the fundamental tensor g_{ij} depends not only on the points of the space but also on the directions of tangent vectors at these points.

When we consider notions such as tensors, etc., in a Finsler space M, it is generally more convenient to take the whole tangent vector bundle $T(M)$ into consideration rather than restricting ourselves to the space M. For example, let P be the †tangent n-frame bundle over a Finsler space M and $Q=p^{-1}(P)$ be the †principal fiber bundle over $T(M)$ induced from P by the projection p of $T(M)$ onto M. We call the elements of fiber bundles associated with Q tensors. In this sense, the fundamental tensor g_{ij} in a Finsler space is the covariant tensor field of order 2. Therefore it is natural to consider a connection in a Finsler space as a connection in the principal fiber bundle Q. The connection in a Finsler space defined by E. Cartan is exactly of this type [3]. Namely, he showed that by assigning to a connection in Q certain conditions related to the Finsler metric, we can determine uniquely a connection from the fundamental tensor so that the covariant differential of the fundamental tensor vanishes.

Cartan's introduction of the notion of connection produced a development in the theory of Finsler spaces that parallels the development in the theory of Riemannian spaces, and many important results have been obtained. O. Varga (1941) succeeded in obtaining a Cartan connection in a simpler way by using the notion of osculating Riemannian space. S. S. Chern (1943) studied general Euclidean connections that contain Cartan connections as a special case. Noticing that the tangent space of a Finsler space is a †normed linear space, H. Rund (1950) obtained many notions different from those of Cartan. However, as far as the theory of connections is concerned, the two theories do not seem to be essentially different. The theory of curvature in a Finsler space is more complicated than that in a Riemannian space because we have three curvature tensors in the Cartan connection. Using the fact that, in a local cross section of the tangent vector bundle of a Finsler space, a Riemannian metric can be introduced by the Finsler metric, L. Auslander (1955) [2] extended to Finsler spaces the results of J. L. Synge and S. B. Myers on the curvature and topology of Riemannian spaces. A. Lichnérowicz extended the †Gauss-Bonnet formula to Finsler spaces by considering an integral on the subbundle of the tangent vector bundle satisfying $L(x, y)=1$ [6]. M. H. Akbar-Zadeh studied †holonomy groups and transformation groups of Finsler spaces by using the theory of fiber bundles.

Connections of Finsler spaces have been investigated by many geometers, but most of them used methods considerably different from those of the modern theory of connections in principal fiber bundles. J. H. Taylor and Synge (1925) defined the covariant differential of a vector field along a curve. L. Berwald (1926) defined a connection from the point of view of the **general geometry of paths**. A curve on a manifold satisfying the differential equation

$$\frac{d^2x^i}{dt^2}+2G^i\left(x,\frac{dx}{dt}\right)=0$$

is called a **path**. The theory was originated by O. Veblen and T. Y. Thomas and generalized as above by J. Douglas. Characteristically, with respect to a Berwald connection, the covariant differential of the fundamental tensor does not vanish.

A Finsler space is a space endowed with a metric for line elements. As a dual concept, we have a **Cartan space**, which is endowed with a metric for **areal elements** [4]. A. Kawaguchi (1937) extended these notions further and studied a **space of line elements of higher order** (or **Kawaguchi space**).

References

[1] M. H. Akbar-Zadeh, Les espaces de Finsler et certaines de leurs généralisations, Ann. Sci. Ecole Norm. Sup., (3) 80 (1963), 1–79.
[2] L. Auslander, On curvature in Finsler geometry, Trans. Amer. Math. Soc., 79 (1955), 378–388.
[3] E. Cartan, Les espaces de Finsler, Actualités Sci. Ind., Hermann, 1934.
[4] E. Cartan, Les espaces métriques fondés sur la notion d'aire, Actualités Sci. Ind., Hermann, 1933.
[5] P. Finsler, Über Kurven und Flächen in allgemeinen Räumen, dissertation, Göttingen, 1918.
[6] A. Lichnérowicz, Quelques théorèmes de géométrie différentielle globale, Comment. Math. Helv., 22 (1949), 271–301.
[7] B. Riemann, Über die Hypothesen, welche der Geometrie zu Grunde liegen, Habilitationsschrift, 1854. (Gesammelte mathematische Werke, Teubner, 1876, 254–270; Dover, 1953.)
[8] H. C. Wang, On Finsler spaces with completely integrable equations of Killing, J. London Math. Soc., 22 (1947), 5–9.

[9] H. Rund, The differential geometry of Finsler spaces, Springer, 1959.
[10] M. Matsumoto, The theory of Finsler connections, Publ. Study Group Geom. 5., Okayama Univ., 1970.

153 (IX.7) Fixed-Point Theorems

A. General Remarks

Given a mapping f of a space X into itself, a point x of X is called a **fixed point** of f if $f(x) = x$. When X is a topological space and f is a continuous mapping, we have various theorems concerning the fixed points of f.

B. Fixed-Point Theorems for Polyhedra

(1) Brouwer Fixed-Point Theorem. Let X be a †simplex and $f: X \to X$ a continuous mapping. Then f has a fixed point in X (*Math. Ann.*, 69 (1910), 71 (1912)).

(2) Lefschetz Fixed-Point Theorem. Let $H_p(X)$ be the p-dimensional †homology group of a †finite polyhedron X (with integral coefficients), $T_p(X)$ the †torsion subgroup of $H_p(X)$, and $B_p(X) = H_p(X)/T_p(X)$. The continuous mapping $f: X \to X$ naturally induces a homomorphism f_* of the free **Z**-module $B_p(X)$ into itself. Let α_p be the †trace of f_* and $\Lambda_f = \sum_{p=0}^{n}(-1)^p \alpha_p$ ($n = \dim X$). We call this integer Λ_f the **Lefschetz number** f.

We have the **Lefschetz fixed-point theorem**: (i) Let f, g be continuous mapping sending X into itself. If f, g are †homotopic ($f \simeq g$), then $\Lambda_f = \Lambda_g$. (ii) If $\Lambda_f \neq 0$, then f has at least one fixed point in X (*Trans. Amer. Math. Soc.*, 28 (1926)).

The condition $\Lambda_f \neq 0$ is, however, not necessary for the existence of a fixed point of f. The Brouwer fixed-point theorem is obtained immediately from (i) and (ii). In particular, if the mapping f is homotopic to the identity mapping 1_X, then α_p is the pth †Betti number of X, and Λ_f is equal to the †Euler characteristic $\chi(X)$ of X. Hence, in this case, if $\chi(X) \neq 0$, then f has a fixed point.

When X is a compact oriented manifold without boundary the Lefschetz number Λ_f of f can be interpreted as the †intersection number of the graph of f and the diagonal of X.

More generally, let X and Y be compact oriented n-dimensional manifolds without boundary. If f and $g: X \to Y$ are continuous mappings, a point x of X such that $f(x) = g(x)$ is called a **coincidence point** of f and g. The intersection number $\Lambda_{f,g}$ of the graph of f and that of g is called the **coincidence number** of f and g. If $\Lambda_{f,g} \neq 0$, then f and g have at least one coincidence point. The coincidence number $\Lambda_{f,g}$ is also expressed as $\sum_{p=0}^{n}(-1)^p \mathrm{tr}(f^* \circ g_! \mid H^p(X))$, where $g_!$: $H^p(X) \to H^p(Y)$ is the †Gysin homomorphism of g.

Suppose that a finite group G acts on the manifolds X and Y. If $f: X \to Y$ is a mapping, a point x of X such that $\tau f(x) = f\tau(x)$ for all $\tau \in G$ is called an **equivariant point** of f. When G is a group of order 2 and acts on X nontrivially, the **equivariant point index** $\hat{\Lambda}_f$ is employed. This index was introduced by Nakaoka (*Japan. J. Math.* 4 (1978)), using the †equivariant cohomology. It has the property that $\hat{\Lambda}_f \neq 0$, implying that f admits an equivariant point. The prototype of this theorem is the Borsuk-Ulam theorem (*Fund. Math.* 20 (1933)), which states that a continuous mapping f: $S^n \to R^n$ always admits a point $x \in S^n$ such that $f(x) = f(-x)$.

(3) Lefschetz Number and Fixed-Point Indices. Suppose that $|K|$ is an n-dimensional homogeneous polyhedron (i.e., any simplex of K that is not a face of another simplex of K is of dimension n), and $f: |K| \to |K|$ is a continuous mapping. Then there exists a continuous mapping $g: |K| \to |K|$ homotopic to f and admitting only isolated fixed points $\{q_1, \ldots, q_r\}$, each of which is an inner point of an n-dimensional simplex of K. The †local degree λ_i of a mapping g at q_i is called the **fixed-point index** of g at q_i. Then $J_f = \sum_{i=1}^{r} \lambda_i$ does not depend on the choice of g and is equal to $(-1)^n \Lambda_f$.

(4) Singularities of a Continuous Vector Field. Let X be an n-dimensional †differentiable manifold and F a †continuous vector field on X that assigns a tangent vector x_p to each point p of X. A point p is called a **singular point** of F if x_p is the zero vector. The vector field F induces in a natural manner a continuous mapping $f: X \to X$ that is homotopic to the identity mapping 1_X. Then a fixed point of f is a singular point of F, and vice versa. When such a singular point p is isolated, there exists a †coordinate neighborhood N of p that is homeomorphic to an n-dimensional open ball such that x_q is nonzero for every point q in N except for $q = p$. Let N' be the boundary of N. Then we may consider $N' \cong S^{n-1}$ and a mapping $F \mid N'$ from S^{n-1} to $\mathbf{R}^n \setminus \{0\} \simeq S^{n-1}$. The †degree of a mapping $S^{n-1} \xrightarrow{F \mid N'} \mathbf{R}^n \setminus \{0\} \underset{\simeq}{\xrightarrow{\pi}} S^{n-1}$ is called the **index of the singular point** p. This index is equal to the fixed-point index λ_p of f at p. Hence, when X is compact and has no

boundary, the sum of indices of (isolated) singular points of F is equal to $(-1)^n\chi(X)$. In particular, a compact manifold X without boundary admits a continuous vector field with no singular point if and only if $\chi(X)=0$ (**Hopf's theorem**, *Math. Ann.*, 96 (1927)).

(5) **Poincaré-Birkhoff Fixed-Point Theorem.** In certain cases, a continuous mapping $f: X\to X$ of a finite polyhedron X into itself has fixed points even if $\Lambda_f=0$. For example, let X be the annular space $\{(r,\theta)\mid\alpha\leqslant r\leqslant\beta\}$ ((r,θ) are the polar coordinates of points in a Euclidean plane) and let $f: X\to X$ be a homeomorphism satisfying the following conditions: (i) there exist continuous functions $g(\theta)$, $h(\theta)$ such that $g(\theta)<\theta$, $h(\theta)>\theta$, $f(\alpha,\theta)=(\alpha,g(\theta))$, $f(\beta,\theta)=(\beta,h(\theta))$; (ii) there exists a continuous positive function $\rho(r,\theta)$ defined for $\alpha<r<\beta$ such that

$$\iint_D \rho(r,\theta)\,dr\,d\theta=\iint_D \rho(f(r,\theta))\,dr\,d\theta$$

for all measurable sets D.

Then f has at least two fixed points. This theorem was conjectured in 1912 by Poincaré, who hoped to apply it to solve the †restricted three-body problem. The theorem was later proved by G. D. Birkhoff (*Trans. Amer. Math. Soc.*, 14 (1913); see also M. Brown and W. D. Neumann, *Michigan Math. J.* 24 (1977)) and is called the **Poincaré-Birkhoff fixed-point theorem** or the **last theorem of Poincaré**.

C. Atiyah-Bott and Atiyah-Singer Fixed-Point Theorems

There is a far-reaching generalization of the Lefschetz formula given by Atiyah and Bott (*Ann. Math.*, (2) 86 (1967), 88 (1968)). Let M be a compact differentiable manifold without boundary and $f: M\to M$ a differentiable mapping with only simple fixed points; that is, it is assumed that $\det(1-df_p)\neq 0$ for each fixed point $p\in M$ of f, where df_p is the differential of f at the point p. The fixed points of f are finite in number. Suppose that an †elliptic complex over M

$$\mathscr{E}: 0\to\Gamma(E_0)\xrightarrow{d_0}\Gamma(E_1)\xrightarrow{d_1}\dots\xrightarrow{d_{l-1}}\Gamma(E_l)\to 0$$

and a sequence of smooth vector bundle homomorphisms $\varphi_i: f^*E_i\to E_i$ $(i=0,\dots,l)$ are given such that $d_iT_i=T_{i+1}d_i$ for each i, where $T_i:\Gamma(E_i)\to\Gamma(E_i)$ is defined by $T_is(x)=\varphi_is(f(x))$ for $s\in\Gamma(E_i)$. The sequence $T=(T_i)$ induces endomorphisms H^iT of the homology groups $H^i(\mathscr{E})$ of the elliptic complex $\mathscr{E}$. We define the Lefschetz number $L(T)$ by $L(T)=\sum_{i=0}^{l}(-1)^i\operatorname{tr}H^iT$. On the other hand, for a fixed point p of f, let $\varphi_{i,p}: E_{i,p}\to E_{i,p}$ denote the restriction of φ_i on the fiber $E_{i,p}$ of E_i over p. Under the circumstances mentioned above the Lefschetz number $L(T)$ is given by the formula

$$L(T)=\sum_p\frac{\sum_{i=0}^{l}(-1)^i\operatorname{tr}\varphi_{i,p}}{|\det(1-df_p)|},$$

where the summation is over the fixed points of f.

Here are some examples of the above formula. First, take as $\mathscr{E}$ the †de Rham complex and as φ_i the obvious one, i.e., the ith exterior power of the transpose of df. In this case, the formula reduces to the classical Lefschetz formula. As a second example consider the †Dolbeault complex

$$0\to\Lambda^{0,0}(M)\xrightarrow{\bar\partial}\Lambda^{0,1}(M)\xrightarrow{\bar\partial}\dots\xrightarrow{\bar\partial}\Lambda^{0,n}(M)\to 0$$

of a compact complex manifold M and a holomorphic mapping $f: M\to M$ with only simple fixed points. In this case the formula above reduces to one giving the Lefschetz number of the induced endomorphisms $f^*: H^{0,*}(M)\to H^{0,*}(M)$ of the †Dolbeault cohomology:

$$L(f^*)=\sum_p\frac{1}{\det_{\mathbf{C}}(1-df_p)},$$

where df_p is regarded as a holomorphic differential.

If the assumption that the mapping f has only simple fixed points is replaced by the one that f is a diffeomorphism of M contained in a compact †transformation group G, then there is also a generalized Lefschetz formula, given by Atiyah and Singer (*Ann. Math.*, (2) 87 (1968)). The fixed-point set of such a diffeomorphism is a closed submanifold of M (consisting of several components). Suppose that we are given an elliptic complex $\mathscr{E}$ over M and a lift of the G-action on M to $\mathscr{E}$. The latter implies that, if we define $T_i:\Gamma(E_i)\to\Gamma(E_i)$ by $T_is(x)=f^{-1}s(f(x))$ for $s\in\Gamma(E_i)$, then $d_iT_i=T_{i+1}d_i$ holds for each i. Under these circumstances, the Lefschetz number $L(T)$ can be expressed in the form

$$L(T)=\sum_{F_i}\nu(F_i),$$

where the summation is over the components $\{F_i\}$ of the fixed-point set M^f of f and where the number $\nu(F_i)$ is written explicitly in terms of the †symbol of the elliptic complex $\mathscr{E}$ with G-action, the characteristic classes of the manifold F_i, the characteristic classes of the normal bundle of F_i in M, and the action of $g=f^{-1}$ on the normal vectors. The formula is essentially a reformulation of the †Atiyah-Singer index theorem. In fact, $L(T)$ is the †analytic index of $\mathscr{E}$ evaluated on g and the number $\nu(F_i)$ is deduced from the †topological index of $\mathscr{E}$ using

the localization theorem. The most useful elliptic complexes are de Rham complexes, Dolbeault complexes, signature operators and Dirac operators. In the case of Dolbeault complexes, f is assumed to be an analytic automorphism, and the number $v(F_i)$ takes the form

$$v(F_i)=\frac{\prod_\theta \mathscr{U}^\theta \mathscr{T}(F_i)}{\det(1-g\,|\,N^*)}[F_i].$$

Here, the normal bundle N of F_i has a decomposition $N=\sum_\theta N(\theta)$ into the sum of complex vector bundles such that g acts on $N(\theta)$ as $e^{i\theta}$, and $\mathscr{U}^\theta$ is the characteristic class defined by

$$\mathscr{U}^\theta=\prod_j \frac{1-e^{-i\theta}}{1-e^{-x_j-i\theta}},$$

where the †Chern class of $N(\theta)$ is written as $c(N(\theta))=\prod_j(1+x_j)$; moreover $\mathscr{T}(F_i)$ denotes the †Todd class of the complex manifold F_i, and N^* denotes the dual bundle of N (→ 237 K-Theory H).

D. Fixed-Point Theorems for Infinite-Dimensional Spaces

Birkhoff and O. D. Kellogg generalized Brouwer's fixed-point theorem to the case of function spaces (*Trans. Amer. Math. Soc.*, 23 (1922)). Their result was utilized to show the existence of solutions of certain differential equations, and has led to a new method in the theory of functional equations.

J. P. Schauder obtained the following theorem: Let A be a closed convex subset of a Banach space, and assume that there exists a continuous mapping T sending A to a †countably compact subset $T(A)$ of A. Then T has fixed points (*Studia Math.*, 2 (1930)). This theorem is called the **Schauder fixed-point theorem.**

A. Tikhonov generalized Brouwer's result and obtained the following **Tikhonov fixed-point theorem** (*Math. Ann.*, 111 (1935)): Let R be a locally convex †topological linear space, A a compact convex subset of R, and T a continuous mapping sending A into itself. Then T has fixed points.

This theorem may be applied to the case where R is the space of continuous mappings sending an m-dimensional Euclidean space E^m into a k-dimensional Euclidean space E^k to show the existence of solutions of certain differential equations. For example, when $m=k=1$, consider the differential equation

$$dy/dx=f(x,y), \qquad y(x_0)=y_0.$$

We set $T(y)=y_0+\int_{x_0}^{x}f(t,y(t))\,dt$ to determine a continuous mapping $T:R\rightarrow R$. Then the fixed points of T are the solutions of the differential equation. Now we can apply the theorem of Tikhonov to show the existence of solutions.

On the other hand, when we are given problems of functional analysis, Schauder's fixed-point theorem is usually more convenient to apply than Tikhonov's theorem.

The following theorem, written in terms of functional analysis, is useful for applications: Let D be a subset of an n-dimensional Euclidean space, F the family of continuous functions defined on D, and $T:F\rightarrow F$ a mapping. Suppose that the following three conditions are satisfied: (i) For $f_1,f_2\in F$, $0<\lambda<1$ implies $\lambda f_1+(1-\lambda)f_2\in F$. (ii) If a series $\{f_k\}$ of functions in F converges uniformly in the wider sense to a function f, then $f\in F$; and furthermore, the series $\{Tf_k\}$ converges uniformly in the wider sense to Tf. (iii) The family $T(F)$ is a †normal family of functions on D. Then there exists a function $f\in F$ such that $Tf=f$.

Let R be a topological linear space and T a mapping assigning a closed convex subset $T(x)$ of R to each point x of R. A point x of R is called a **fixed point** of T if $x\in T(x)$. The mapping T is called **semicontinuous** if the condition $x_n\rightarrow a$, $y_n\rightarrow b$ ($y_n\in T(x_n)$) implies that $b\in T(a)$. In particular, if K is a bounded closed convex subset of a finite-dimensional Euclidean space R and T a semicontinuous mapping sending points of K into convex subsets of K, then T admits fixed points (**Kakutani fixed-point theorem**, *Duke Math. J.*, 8 (1941)). This result was further generalized to the case of locally convex topological linear spaces by Ky Fan (*Proc. Nat. Acad. Sci. US*, 38 (1952)).

References

[1] B. F. Brown, The Lefschetz fixed point theorem, Scott, Foresman, 1971.
[2] J. Cronin, Fixed points and topological degree in nonlinear Analysis, Amer. Math. Soc. Math. Surveys, no. 11, 1964.
[3] S. Lefschetz, Topology, Amer. Math. Soc. Colloq. Publ., vol. XII, 1930.
[4] D. R. Smart, Fixed point theorems, Cambridge tracts in math. 66, Cambridge Univ. Press, 1974.

154 (IX.21) Foliations

A. Introduction

A foliation is a kind of geometric structure on manifolds, such as a differentiable or complex structure. The study of foliations evolved from

investigation of the behavior of †orbits of a vector field and also of the solutions of †total differential equations. Through the early works of C. Ehresmann, G. Reeb, and A. Haefliger, together with the development of manifold theory in 1960s, it became an established field of mathematics. Since then, great progress has been made in this field, especially in its topological aspects. At the same time it turns out that many problems in foliation theory are deeply related not only to the geometry of manifolds but also to various other branches of mathematics, such as the theory of differential equations, functional analysis, and group theory.

B. Definitions and General Remarks

A **foliation** on a manifold can be defined within various categories: topological, C^r-differentiable ($1 \leqslant r \leqslant \infty$), real analytic, and holomorphic. For definiteness, however, we restrict ourselves to the C^r-differentiable category in what follows. Furthermore, all manifolds are assumed to be paracompact.

Let M be an n-dimensional C^∞-manifold, possibly with boundary. A **codimension** q, C^r-**foliation** of M ($0 \leqslant q \leqslant n, 0 \leqslant r \leqslant \infty$) is a family $\mathscr{F} = \{L_\alpha \mid \alpha \in A\}$ of arcwise connected subsets, called **leaves**, of M with the following properties: (i) $L_\alpha \cap L_{\alpha'} = \varnothing$ if $\alpha \neq \alpha'$; (ii) $\bigcup_{\alpha \in A} L_\alpha = M$; (iii) Every point in M has a local coordinate system (U, ψ) of class C^r such that for each leaf L_α the arcwise connected components of $U \cap L_\alpha$ are described by $x^{n-q+1} = \text{constant}, \ldots, x^n = \text{constant}$, where $x^1, x^2, \ldots, x^n$ denote the local coordinates in the system (U, ψ). In particular, every leaf of $\mathscr{F}$ is an $(n-q)$-dimensional †submanifold of M. The totality of integral submanifolds of a †completely integrable nonsingular system of †Pfaffian equations on $\mathbf{R}^n$, $\omega_i = a_{i1}(x)dx_1 + a_{i2}(x)dx_2 + \ldots + a_{in}(x)dx_n = 0$ ($i = 1, 2, \ldots, q$) forms a codimension q foliation, and the totality of integral curves of a nonsingular vector field of class C^r on M ($r \geqslant 1$) constitutes a codimension $n-1$ C^r-foliation.

Let $\mathscr{F}$ be a codimension q, C^r-foliation of M ($r \geqslant 1$). Then M admits a C^{r-1} †p-plane field consisting of all vectors tangent to the leaves of $\mathscr{F}$, and, dually, a C^{r-1} q-plane field ($p+q=n$). Denote the former by $\tau(\mathscr{F})$ and the latter by $\nu(\mathscr{F})$, and call them the **tangent bundle** and the **normal bundle** of $\mathscr{F}$, respectively. $\nu(\mathscr{F})$ is isomorphic to the quotient bundle $T(M)/\tau(\mathscr{F})$. $\mathscr{F}$ is called **transversely orientable** if $\nu(\mathscr{F})$ is an orientable vector bundle. A C^r-mapping f: $N \to M$ is said to be **transverse to the foliation** $\mathscr{F}$ if the composite mapping $T(N) \xrightarrow{df} T(M) \to T(M)/\tau(\mathscr{F})$ is epimorphic at each point of N. In this case, f induces a codimension q, C^r-foliation $f^{-1}(\mathscr{F})$ of N whose leaves are the arcwise connected components of $f^{-1}(L)$ ($L \in \mathscr{F}$).

In particular, if Q is a q-dimensional C^r-manifold and $f: N \to Q$ is a C^r-submersion, f induces a codimension q, C^r-foliation of N whose leaves are the arcwise connected components of $f^{-1}(x)$ ($x \in Q$).

A C^r p-plane field $\mathscr{X}$ on M is called **involutive** if, for any C^r vector fields X, Y on M such that $X_x, Y_x \in \mathscr{X}_x$ ($x \in M$), the †Lie bracket $[X, Y]$ satisfies $[X, Y]_x \in \mathscr{X}_x$. This condition is known as the **Frobenius integrability condition** for $\mathscr{X}$. If $\mathscr{X}$ is defined locally by q Pfaffian equations $\omega_1 = \ldots = \omega_q = 0$, the above condition is equivalent to the condition that there are C^{r-1} 1-forms θ_{ij} ($i,j = 1, \ldots, q$) such that $d\omega_i = \sum_{j=1}^q \theta_{ij} \wedge \omega_j$. A C^r p-plane field $\mathscr{X}$ is said to be **completely integrable** if it is a tangent bundle of some foliation. When $r \geqslant 2$, $\mathscr{X}$ is completely integrable if and only if it is involutive (**Frobenius theorem**) (→ 428 Total Differential Equations). There is a topological obstruction to the complete integrability of $\mathscr{X}$ (→ Section F).

A closed C^∞-manifold M admits a codimension 1, C^r-plane field if and only if the Euler number of M vanishes. In 1944, Reeb constructed a codimension 1, C^∞-foliation of the 3-sphere S^3 as follows [1]. Let $f(x)$ be a C^∞-even function defined on the open interval $(-1, 1)$, such that

$$\lim_{|x| \to 1} \frac{d^{(k)}}{dx^{(k)}} \frac{1}{f'(x)} = 0 \qquad (k = 0, 1, 2, \ldots).$$

The graphs of the equations $y = f(x) + c$ ($-1 < x < 1, c \in \mathbf{R}$) together with the lines $x = \pm 1$ constitute a codimension 1, C^∞-foliation of $[-1, 1] \times \mathbf{R}$. Then by rotating it around the y-axis in $\mathbf{R}^3$, we obtain a codimension 1, C^∞-foliation of $D^2 \times \mathbf{R}$, where D^2 denotes the closed 2-disk. The foliation is invariant under vertical translations and therefore defines a codimension 1, C^∞-foliation of $D^2 \times S^1$. This foliation is called the **Reeb component** of $D^2 \times S^1$. Since S^3 is a union of two solid tori intersecting in the common toroidal surface, the Reeb components of each solid torus constructed above define the so-called **Reeb foliation** of S^3 (Fig. 1).

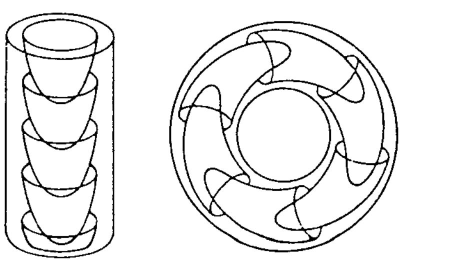


Fig. 1

Generalizing the above construction into a differential topological method, one obtains the following results: Every closed 3-dimensional manifold admits a codimension 1, C^∞-foliation (S. P. Novikov [3], W. Lickorish, J. Wood); every odd-dimensional sphere admits a codimension 1, C^∞-foliation (I. Tamura, A. Durfee, B. Lawson). On the other hand, every open manifold has a codimension 1, C^∞-foliation induced by a submersion over **R** (→ Section F).

Let M be a total space of a C^r-bundle over a C^∞-manifold B with fiber F. If F is a C^∞-manifold and the †structure group reduces to a totally disconnected subgroup of $\mathrm{Diff}^r F$, the group of all C^r-diffeomorphisms of F, then the local sections, which are defined in an obvious manner using the local triviality of this bundle, fit together to give leaves of a codimension q, C^r-foliation ($q=\dim F$). In this case, M is called a **foliated bundle** or a **flat F-bundle** over B. Each leaf of this foliation is diffeomorphic to a covering manifold of B and transverse to the fibers of the bundle $M\to B$. Foliated bundles exhibit a class of foliations; this is especially important in connection with the characteristic classes of foliations (→ Section G).

C. Holonomy

The notion of the holonomy of a leaf, given by Ehresmann, is a generalization of the †Poincaré mapping in †dynamical systems. Let $\mathscr{F}$ be a codimension q, C^r-foliation of M and L be a leaf of $\mathscr{F}$. Let $N(L)$ denote the total space of the normal disk bundle of L in M. Choose a C^r-immersion $i: N(L)\to M$ such that i restricted to the zero section of $N(L)$ is the natural inclusion and i maps the fibers of $N(L)$ transversely to the foliation $\mathscr{F}$. Then the induced foliation $i^{-1}(\mathscr{F})$ of $N(L)$ has the properties that the leaves are transverse to the fibers of $N(L)$ and the zero section of $N(L)$ is a leaf. If γ is an oriented loop in L based at a point $x_0\in L$, then there is a neighborhood U of 0 in the fiber over x_0 satisfying the following: for each point $x\in U$ there is a curve $\gamma_x:[0,1]\to N(L)$ having the properties: (i) $\gamma_x(0)=x$, (ii) $\mathrm{Im}(\gamma_x)$ lies on a leaf of $i^{-1}(\mathscr{F})$, and (iii) $\pi\circ\gamma_x(t)=\gamma(t)$ for any $t\in[0,1]$, where $\pi: N(L)\to L$ is the bundle projection. The family of curves $\{\gamma_x \mid x\in U\}$ gives a C^r-diffeomorphism H_γ from U to another open set of $\pi^{-1}(x_0)$, which assigns $\gamma_x(1)$ to x. Let G_q^r denote the group of †germs at 0 of all local C^r-diffeomorphisms of $\mathbf{R}^q$ which fix 0. The germ at 0 of the mapping H_γ depends only on the †homotopy class of γ, and, by identifying $\pi^{-1}(x_0)$ with $\mathbf{R}^q$, we obtain a homomorphism $h_L:\pi_1(L,x_0)\to G_q^r$. h_L is determined by L up to conjugacy and is called the **holonomy homomorphism**, or simply the **holonomy**, of the leaf L. The image of h_L is called the **holonomy group** of L. For $r\geqslant 1$, by differentiating each element of G_q^r, one has a homomorphism $dh_L:\pi_1(L,x_0)\to GL(q;\mathbf{R})$, called the **linear holonomy** of L. The holonomy of a proper leaf (→ Section D) completely characterizes the foliation of a neighborhood of it (Haefliger).

D. Topology of Leaves

Let $\mathscr{F}$ be a codimension q foliation of M. The **leaf topology** is a topology of M defined by requiring each connected component of the set of the form $U\cap L$ to be open, where U is an open set in M and $L\in\mathscr{F}$. Leaves are nothing but the connected components of M with respect to this topology. A leaf $L\in\mathscr{F}$ is called a **compact leaf** if L is compact in the leaf topology. In general, L is called **proper** if two topologies on L induced from the original and the leaf topologies on M coincide. Any compact leaf is proper. A leaf L is called **locally dense** if $\mathrm{Int}\,\bar{L}\neq\varnothing$. If a leaf is neither proper nor locally dense, it is called **exceptional**. Since we are assuming that M is paracompact, a leaf that is a closed subset of M is proper (Haefliger). There exists a codimension 1, C^1-foliation of the 2-torus T^2 that contains exceptional leaves. But in the C^r category ($r\geqslant 2$), such a foliation does not exist on T^2 (A. Denjoy, C. Siegel). In higher dimensions, there are examples of C^∞-foliations with exceptional leaves (R. Sacksteder). The following result is called the **Novikov closed leaf theorem** [3]: Any codimension 1, C^r-foliation ($r\geqslant 2$) of a closed 3-dimensional manifold M contains a Reeb component if either $\pi_1(M)$ is finite or $\pi_2(M)\neq 0$ ($M\neq S^1\times S^2, S^1\times RP^2$). In particular, every C^2-foliation of S^3 contains a compact leaf homeomorphic to T^2. The question of whether every codimension 2, C^r-foliation of S^3 has a compact leaf is known as the **Seifert conjecture**. There is a counterexample in the C^1 case (P. Schweitzer [7]), but it remains unsolved for $r\geqslant 2$ (→ 126 Dynamical Systems N).

A compact leaf L is said to be **stable** if it has an arbitrarily small open neighborhood that is a union of compact leaves. The following results are called the **Reeb stability theorems**: (1) Let L be a compact leaf of a C^r-foliation ($r\geqslant 0$). If the holonomy group of L is finite, then it is stable (Reeb [1]). (2) Let $\mathscr{F}$ be a transversely orientable codimension 1, C^r-foliation ($r\geqslant 1$) of a compact manifold M (tangent to the boundary). If there exists a compact leaf L with $H^1(L;\mathbf{R})=0$, then M is a fiber bundle over S^1 or $[0,1]$, and the leaves of $\mathscr{F}$ are the fibers of this bundle. In particular,

every leaf is compact and stable (Reeb [1], Thurston [8]).

The generalization of the stability theorem for proper leaves has been investigated by T. Inaba and P. Dippolito.

E. Haefliger Structures

Let $\mathscr{G}_q^r$ be the †pseudogroup consisting of all C^r-diffeomorphisms g of an open subset of $\mathbf{R}^q$ to another open subset of $\mathbf{R}^q$. We write Γ_q^r for the set of all germs $[g]_x$ of g at x, $x \in$ domain of g, $g \in \mathscr{G}_q^r$. The sheaf topology of Γ_q^r is the topology whose open base is the family of subsets of the form $\bigcup_{x \in \text{domain of } g}\{[g]_x\}$. With this topology and the multiplication induced from the composition of $\mathscr{G}_q^r$, Γ_q^r is a **topological groupoid**, i.e., a †groupoid whose multiplication and inverse mappings are continuous. If one identifies a point x in $\mathbf{R}^q$ with $[\mathrm{id}_{\mathbf{R}^q}]_x$, Γ_q^r contains $\mathbf{R}^q$ as a subspace.

A codimension q, C^r-**Haefliger structure** or a Γ_q^r**-structure** $\mathscr{H}$ on a topological space X is a maximal covering of X by open sets $\{U_i \mid i \in J\}$, such that for each pair $i, j \in J$, there is a continuous mapping $\gamma_{ij}: U_i \cap U_j \rightarrow \Gamma_q^r$ satisfying

$$\gamma_{ik}(x)=\gamma_{ij}(x)\circ\gamma_{jk}(x) \quad \text{for} \quad x \in U_i \cap U_j \cap U_k. \qquad (*)$$

Since $\gamma_{ii}(x)$ is the germ of the identity mapping for $x \in U_i$, the correspondence $x \rightarrow \gamma_{ii}(x)$ defines a continuous mapping $f_i: U_i \rightarrow \mathbf{R}^q \subset \Gamma_q^r$. A codimension q, C^r-foliation of a C^∞-manifold M is the same as a Γ_q^r-structure on M such that each f_i is a C^r-submersion (Haefliger [5]).

If $f: Y \rightarrow X$ is continuous and $\mathscr{H}$ is a Γ_q^r-structure on X, there is an induced Γ_q^r-structure $f^{-1}\mathscr{H}$ on Y which is defined by $\{f^{-1}(U_i), \gamma_{ij}\circ f \mid i, j \in J\}$. Two Γ_q^r-structures $\mathscr{H}_0$ and $\mathscr{H}_1$ on X are said to be **homotopic** if there exists a Γ_q^r-structure $\mathscr{H}$ on $X \times [0,1]$ such that $\mathscr{H}|_{X \times \{t\}} = \mathscr{H}_t$ $(t=0,1)$.

Let $\Gamma_q^r(X)$ be the set of homotopy classes of Γ_q^r-structures on X. There exists a space $B\Gamma_q^r$, called the **classifying space for** Γ_q^r**-structures**, such that there is a natural one-to-one correspondence between $\Gamma_q^r(X)$ and $[X, B\Gamma_q^r]$ for any paracompact space X, where $[A, B]$ denotes the set of homotopy classes of continuous mappings from A to B. By condition $(*)$ above, if $r>0$, the differentials $\{d\gamma_{ij}(x) \mid x \in U_i \cap U_j, i, j \in J\}$ define a q-dimensional vector bundle $\nu(\mathscr{H})$ over X, which is called the **normal bundle** of $\mathscr{H}$. The correspondence $\mathscr{H} \rightarrow \nu(\mathscr{H})$ gives a continuous mapping $\nu: B\Gamma_q^r \rightarrow BGL(q; \mathbf{R})$ among classifying spaces. If $r=0$, there is also a similar mapping $\nu: B\Gamma_q^0 \rightarrow B\mathrm{Top}_q$. Let $B\overline{\Gamma}_q^r$ be the homotopy fiber of the mapping ν. $B\overline{\Gamma}_q^r$ is a classifying space for the Γ_q^r-structures with trivialized normal bundles.

There is a tight connection between the topology of $B\overline{\Gamma}_q^r$ and the group structure of $\mathrm{Diff}^r(\mathbf{R}^q)$, which is stated below. Let $\mathrm{Diff}_K^r(\mathbf{R}^q)$ be the topological group of all C^r-diffeomorphisms of $\mathbf{R}^q$ that are identities outside some compact sets with the C^r topology. Let $\mathrm{Diff}_{K,\delta}^r(\mathbf{R}^q)$ denote $\mathrm{Diff}_K^r(\mathbf{R}^q)$ with the discrete topology and $B\overline{\mathrm{Diff}}_K^r(\mathbf{R}^q)$ denote the homotopy fiber of the natural mapping $B\,\mathrm{Diff}_{K,\delta}^r(\mathbf{R}^q) \rightarrow B\,\mathrm{Diff}_K^r(\mathbf{R}^q)$. Then there exists a continuous mapping $B\overline{\mathrm{Diff}}_K^r(\mathbf{R}^q) \rightarrow \Omega^q B\overline{\Gamma}_q^r$ that induces an isomorphism in the homology group with integer coefficient for $0 \leqslant r \leqslant \infty$ and $q \geqslant 1$, where Ω^q denotes the qth †loop space functor (J. N. Mather [12], Thurston [10]). Further, it has been proved that $\mathrm{Diff}_{K,\delta}^r(\mathbf{R}^q)$ is a †simple group (if $r \neq q+1$) and that $B\overline{\Gamma}_q^r$ is †$(q+1)$-connected for $r \neq q+1$ (Mather and Thurston, Haefliger [5]). The group $\mathrm{Homeo}_{K,\delta}\mathbf{R}^q$ is †acyclic (Mather), and hence $B\overline{\Gamma}_q^0$ is †contractible.

F. Existence and Classification of Foliations

Not every plane field on a manifold is isomorphic to a completely integrable field (R. Bott). Thus, in general, the existence of a plane field does not guarantee the existence of a foliation of a manifold. Haefliger and Thurston solved the existence and classification problems in foliation in terms of Haefliger structures as follows. Two codimension q foliations $\mathscr{F}_0$ and $\mathscr{F}_1$ of a C^∞-manifold M are said to be **concordant** if there is a codimension q foliation $\mathscr{F}$ on $M \times [0,1]$, that is transverse to $M \times \{t\}$ $(t=0,1)$ and induces there the given foliation $\mathscr{F}_t$ $(t=0,1)$. They are said to be **integrably homotopic** if one further requires that the foliation $\mathscr{F}$ be transverse to $M \times \{t\}$ for all $t \in [0,1]$ in the definition above. Similarly, two subbundles ξ_0, ξ_1 of $T(M)$ are said to be **concordant** if there is a subbundle ξ of $T(M \times [0,1])$ such that $\xi|_{M \times \{t\}} = \xi_t$ for $t=0,1$, and they are said to be **homotopic** if one further assumes that $\xi|_{M \times \{t\}}$ is a subbundle of $T(M \times \{t\})$ for all $t \in [0,1]$. The following theorem is of fundamental importance.

Theorem: Let M be an open (resp. closed) C^∞-manifold. Then for each $r=0,1,\ldots,\infty$, the integrable homotopy classes (resp. concordance classes) of codimension q, C^r-foliations of M are in a natural one-to-one correspondence with homotopy classes of Γ_q^r-structures $\mathscr{H}$ on M together with homotopy classes (resp. concordance classes) of subbundles of $T(M)$ isomorphic to $\nu(\mathscr{H})$. (M. L. Gromov, A. Phillips, Haefliger [5], Thurston [9]).

The following are consequences of the theorem: (i) A closed manifold M admits a codimension 1, C^∞-foliation if and only if the Euler number of M vanishes. (ii) If a manifold

admits a †q-frame field, then the associated q-plane field is homotopic to the normal bundle of a codimension q foliation of M. (iii) Every dimension q plane field on a C^∞-manifold is homotopic to the normal bundle of a C^0-foliation with C^∞ leaves.

G. Characteristic Classes of Foliations

Let $B\Gamma_q^r$ be the classifying space of Γ_q^r-structures. An element of the cohomology group $H^*(B\Gamma_q^r; \mathbf{R})$ is called a (real) **characteristic class of codimension** q, **C^r-foliations**. If $\mathscr{F}$ is a codimension q, C^r-foliation of M and $f: M \to B\Gamma_q^r$ is the classifying mapping for $\mathscr{F}$, then an element $\alpha(\mathscr{F}) = f^*\alpha \in H^*(M; \mathbf{R})$, $\alpha \in H^*(B\Gamma_q^r; \mathbf{R})$, is called the characteristic class of $\mathscr{F}$ corresponding to α. The first nontrivial characteristic classes of foliations are known as the **Godbillon-Vey classes** (C. Godbillon and J. Vey [13]) and can be defined as follows. Let $\mathscr{F}$ be a transversely oriented, codimension q, C^∞-foliation of M. Then there exists a †q-form Ω on M such that on a neighborhood U of each point of M, Ω is written as $\omega_1^U \wedge \ldots \wedge \omega_q^U$, where $\omega_1^U, \ldots, \omega_q^U$ are linearly independent 1-forms that vanish on leaves of $\mathscr{F}$. By the inte grability condition, there is a 1-form η such that $d\Omega = \eta \wedge \Omega$. Then the Godbillon-Vey class $\Gamma_{\mathscr{F}}$ of $\mathscr{F}$ is the de Rham cohomology class in $H^{2q+1}(M; \mathbf{R})$ represented by the closed $2q+1$ form $\eta \wedge (d\eta)^q$.

The following construction provides a wide class of characteristic classes of foliations (Bott and Haefliger [14], I. Bernstein and B. Rosenfeld [15]). Let J_k be the set of †k-jets at 0 of local C^∞-diffeomorphisms of $\mathbf{R}^q$ keeping 0 fixed. The set $\{J_k\}_{k=0}^\infty$ forms an †inverse system of †Lie groups with respect to the natural homomorphism $p_k: J_{k+1} \to J_k$, and each J_k ($k \geq 1$) contains $O(q)$ as a †maximal compact subgroup. Let P_k be the differentiable fiber bundle of k-jets at 0 of local diffeomorphisms of $\mathbf{R}^q$ whose domains contain 0. It is a †principal J_k-bundle over $\mathbf{R}^q$. Denote by $A(P_\infty)$ the †direct limit of the de Rham complexes of $\{P_k\}_{k=0}^\infty$, and let A be the subcomplex of $A(P_\infty)$ consisting of invariant forms with respect to the natural action of $\mathscr{G}_q^\infty$. A is canonically isomorphic to the †cochain complex $A(A_q)$ of continuous alternating forms on A_q, where A_q is the topological Lie algebra of †formal vector fields on $\mathbf{R}^q$ (→ 105 Differentiable Manifolds AA).

Now let $\mathscr{F}$ be a codimension q, C^∞-foliation of M. Let $P_k(\mathscr{F})$ denote the differentiable fiber bundle over M whose fiber over $x \in M$ is the space of k-jets at x of the C^∞-submersion $f: U \to \mathbf{R}^q$ from an open neighborhood U of x to $\mathbf{R}^q$, satisfying (i) $f(x) = 0$, (ii) $\operatorname{Ker} df = \tau(\mathscr{F})|_U$. This is a principal J_k-bundle over M, and its restriction to U is isomorphic to the pullback by f of the bundle P_k. Hence there are homomorphisms from the set of invariant forms on P_k to $A(P_k(\mathscr{F}))$ that induce a homomorphism $A \cong A(A_q) \to \varinjlim A(P_k(\mathscr{F}))$. This homomorphism is compatible with the action of $O(q)$ and hence induces a homomorphism of $O(q)$-basic subcomplexes. Thus one obtains a homomorphism $\varphi_{\mathscr{F}}: H^*(A_q; O(q)) \to \varinjlim H^*(A(P_k(\mathscr{F})); O(q)) \cong H^*(M; \mathbf{R})$. In fact, $\varphi_{\mathscr{F}}$ depends only on the 2-jet of the foliation $\mathscr{F}$, and one can think of it as a homomorphism $\varphi: H^*(A_q; O(q)) \to H^*(B\Gamma_q^r; \mathbf{R})$ ($r \geq 2$). The elements in $\operatorname{Im} \varphi$ are called the **smooth characteristic classes of foliations**.

Let WO_q be the differential graded algebra:

$$WO_q = E(u_1, u_3, \ldots, u_{2[(q+1)/2]-1}) \otimes \hat{\mathbf{R}}[c_1, \ldots, c_q],$$

where $du_i = c_i$, $dc_i = 0$, $\deg(u_i) = 2i-1$, $\deg(c_i) = 2i$, and E denotes the †exterior algebra over $\mathbf{R}$ generated by the u_i's, and $\hat{\mathbf{R}}$ denotes the †polynomial algebra of the c_i's truncated by the †ideal generated by elements of degree $> 2q$. There exists a homomorphism of differential graded algebras $WO_q \to A(A_q; O(q))$ which induces an isomorphism in cohomology (I. M. Gel'fand and D. B. Fuks). For a codimension q foliation $\mathscr{F}$ of M, the cohomology class determined by c_{2i} corresponds to the ith Pontryagin class of the normal bundle $\nu(\mathscr{F})$ of $\mathscr{F}$, and the cohomology class $u_1 c_1^q$ corresponds to the Godbillon-Vey class of $\mathscr{F}$. In particular, the subring of the cohomology ring $H^*(M; \mathbf{R})$ generated by the †Pontryagin classes of $\nu(\mathscr{F})$ is trivial for degree $> 2q$ (Bott's vanishing theorem).

Let N^{q+1} be a closed $(q+1)$-dimensional Riemannian manifold of †constant negative curvature. The total space $T_1 N$ of the unit tangent sphere bundle of N admits a codimension q, C^∞-foliation associated with the †geodesic flow of $T_1 N$ (†Anosov foliation). It has been shown that the Godbillon-Vey class of this foliation is nontrivial (R. Roussarie, F. Kamber and P. Tondeur, K. Yamato). It is known that many of the smooth characteristic classes are also nontrivial (Bott and Haefliger, Thurston, J. Heitsch).

A smooth characteristic class $\alpha \in H^*(B\Gamma_q^r; \mathbf{R})$ is called **rigid** if for any smooth one-parameter family $\{\mathscr{F}_t\}$ of codimension q foliations on a C^∞-manifold M, $d(\alpha(\mathscr{F}_t))/dt = 0$ holds. The elements in the image of the natural homomorphism

$$H^*(WO_{q+1}) \to H^*(WO_q) \to H^*(B\Gamma_q^r; \mathbf{R})$$

are rigid (Heitsch). On the other hand, the Godbillon-Vey class is not rigid. In fact, Thurston constructed a one-parameter family of codimension q foliations of a certain

$(2q+1)$-dimensional manifold for which the Godbillon-Vey class varies continuously. The characteristic classes of a simple foliation are often trivial. For example, the Godbillon-Vey class of a codimension 1 foliation of a closed manifold vanishes if it is almost without holonomy (i.e., no noncompact leaves have nontrivial holonomy) (M. Herman [16]; T. Mizutani, S. Morita, and T. Tsuboi [17]).

H. Further Topics

(1) **Transverse structures**. Let B be a q-dimensional manifold and $\mathscr{S}$ a geometric structure on it, and let $\mathscr{G}_{\mathscr{S}}$ denote the †pseudogroup generated by the local diffeomorphisms that preserve the structure $\mathscr{S}$. Replacing $\mathscr{G}_q^r$ by $\mathscr{G}_{\mathscr{S}}$ in the definition of a C^r-Haefliger structure, one obtains definitions of a **$\Gamma_{\mathscr{S}}$-structure** and a **$\Gamma_{\mathscr{S}}$-foliation**. A $\Gamma_{\mathscr{S}}$-foliation is called a **Riemannian foliation**, a transversely **real analytic foliation**, or a transversely **holomorphic foliation** if $\mathscr{S}$ is a Riemannian, real analytic, or complex structure on $\mathbf{R}^q$ ($\mathbf{C}^{q/2}$), respectively. The theories for many such foliations are analogous to those for C^r-foliations. For example, many results are known about the characteristic classes of Riemannian or holomorphic foliations. Haefliger showed that there is no codimension 1 real analytic foliation on a simply connected closed manifold and that the classifying space $B\bar{\Gamma}_1^{\omega}$ for codimension 1 transversely oriented transversely real analytic foliations has the homotopy type of a †$K(\pi,1)$-space for some uncountable †perfect group π [5].

(2) **Foliated cobordism**. Two closed oriented n-dimensional C^∞-manifolds M_0 and M_1 with codimension q, C^r-foliations are said to be **foliated cobordant** if there exist a compact oriented $(n+1)$-dimensional C^∞-manifold W with boundary $\partial W = M_1 \cup (-M_0)$ and a codimension q, C^r-foliation of W which is transverse to ∂W and induces the given foliations of M_0 and M_1. The resulting foliated cobordism classes $\{\mathscr{F}\}$ form a group $\mathscr{F}\Omega_{n,q}^r$ with respect to the disjoint union. It is known that $\mathscr{F}\Omega_{2,1}^\infty = \{0\}$ and that the Reeb foliation of S^3 is cobordant to zero. The characteristic classes of foliations provide invariants of foliated cobordisms. In particular the Godbillon-Vey number $\Gamma_{\mathscr{F}}[M]$ is an invariant of $\mathscr{F}\Omega_{2q+1,q}^r$ ($r \geqslant 2, q \geqslant 1$), and a result of Thurston mentioned in Section G implies that the homomorphism $\mathscr{F}\Omega_{2q+1,q}^r \to \mathbf{R}$ defined by $\{\mathscr{F}\} \to \Gamma_{\mathscr{F}}[M]$ is surjective.

(3) **Growth of leaves, transverse invariant measure**. Let $\mathscr{F}$ be a C^∞-foliation of a compact manifold. Fix a †Riemannian metric on M. Then each leaf L of $\mathscr{F}$ has the induced metric, and one has a function $f_L(r) = \mathrm{vol}(D(x,r))$, where $D(x,r)$ is the set of points $y \in L$ whose distance along L from a fixed point $x \in L$ is not greater than r. The growth type of the function f_L is determined only by L. Many papers have been published that deal with the relation between the behavior of leaves and their growth types in codimension 1 foliations (J. Cantwell and L. Conlon, G. Hector, T. Nishimori, N. Tsuchiya). On the other hand, as a generalization of the notion of asymptotic cycles in dynamical systems, the notion of **foliation cycles**, or equivalently a **transverse invariant measure**, has been defined (J. Plante [18], D. Ruelle, D. Sullivan [19]). The existence of a transverse invariant measure for a foliation is closely connected to the growth types of the leaves. The example of Denjoy in Section D leads to the study of †minimal sets of foliations and the structures of foliations (Hector, Cantwell and Conlon). The structure of a codimension 1 foliations which are almost without holonomy has been fairly well investigated (Sacksteder, Hector, H. Imanishi, R. Moussou, Roussaire).

(4) **Compact foliation**. A foliation whose leaves are all compact is called a **compact foliation**. D. Epstein proved that if $\mathscr{F}$ is a codimension 2, C^2-compact foliation of a closed 3-manifold, then the leaves of $\mathscr{F}$ are the fibers of a †Seifert fibration of M. In higher dimensions, the situation is more complicated (Sullivan, R. Edwards and K. Millett).

(5) **Foliated bundles**. There are many results on foliated bundles. In particular J. W. Milnor [20] and J. Wood obtained a condition for a circle bundle ξ over a closed surface Σ to have a foliation transverse to fibers. More precisely, if ξ and Σ are orientable, then ξ admits such a foliation if and only if $|X(\xi)| \leqslant -\min\{0, \chi(\Sigma)\}$, where X denotes the Euler number and χ the Euler-Poincaré characteristic. Kamber and Tondeur made an extensive study of characteristic classes of foliated bundles [21].

(6) **Transverse foliations**. Two foliations $\mathscr{F}$ and $\mathscr{G}$ of M are said to be **transverse** to each other if any two leaves K and L of $\mathscr{F}$ and $\mathscr{G}$ are transverse to each other. A foliated bundle has such foliations. D. Hardorp proved that on every orientable closed 3-manifold, there exists a triple of codimension 1 foliations that are pairwise transverse. Tamura and A. Sato classified the codimension 1 foliations that are transverse to the Reeb component of $D^2 \times S^1$.

References

[1] G. Reeb, Sur certains propriétés topologiques de variétés feuilletées, Actualité Sci. Ind., 1183, Hermann, 1952.
[2] C. Ehresmann, Sur la théorie des variétés

feuilletées, Univ. Roma Rend. Mat. Appl., 10 (1951), 64–82.
[3] S. P. Novikov, Topology of foliations, Amer. Math. Soc. Transl., 1967, 286–304. (Original in Russian, 1965.)
[4] I. Tamura, Foliations and spinnable structures on manifolds, Ann. Inst. Fourier, 23 (1973), 197–214.
[5] A. Haefliger, Homotopy and integrability, Lecture notes in math. 197, Springer, 1971, 133–163.
[6] A. Denjoy, Sur les courbes définies par les équations difféentielles à la surface du tore, J. Math. Pure Appl., 11 (1932), 333–375.
[7] P. Schweitzer, Counterexamples to Seifert conjecture and opening closed leaves of foliations, Ann. Math., (2) 100 (1974), 386–400.
[8] W. Thurston, A generalization of Reeb stability theorem, Topology, 13 (1974), 347–352.
[9] W. Thurston, Existence of codimension one foliations, Ann. Math., (2) 104 (1976), 249–268.
[10] W. Thurston, Foliations and groups of diffeomorphisms, Bull. Amer. Math. Soc., 80 (1974), 304–307.
[11] W. Thurston, Noncobordant foliations of S^3, Bull. Amer. Math. Soc., 78 (1972), 511–514.
[12] J. N. Mather, Integrability in codimension 1, Comment. Math. Helv., 48 (1973), 195–233.
[13] C. Godbillon and J. Vey, Un invariant des feuilletages de codimension un, C. R. Acad. Sci. Paris, 273 (1971), 92–95.
[14] R. Bott and A. Haefliger, On characteristic classes of Γ-foliations, Bull. Amer. Math. Soc., 78 (1972), 1039–1044.
[15] I. Bernstein and B. Rosenfeld, On characteristic classes of foliations, Functional Anal. Appl., 6 (1972), 60–62.
[16] M. Herman, The Godbillon-Vey invariant of foliations by planes of T^3, Lecture notes in math. 597, Springer, 1977, 294–307.
[17] T. Mizutani, S. Morita, and T. Tsuboi, The Godbillon-Vey classes of codimension one foliations which are almost without holonomy, Ann. Math., (2) 113 (1981).
[18] J. Plante, Foliations with measure-preserving holonomy, Ann. Math., (2) 102 (1975), 327–362.
[19] D. Sullivan, Cycles for the dynamical study of foliated manifolds and complex manifolds, Inventiones Math., 36 (1976), 225–255.
[20] J. W. Milnor, On the existence of a connection with curvature zero, Comment. Math. Helv., 32 (1958), 215–223.
[21] F. Kamber and P. Tondeur, Foliated bundles and characteristic classes, Lecture notes in math. 493, Springer, 1975.
[22] B. Lawson, Foliations, Bull. Amer. Math. Soc., 80 (1974), 369–418.
[23] B. Lawson, The quantitative theory of foliations, Regional conference series in math. 27, Conference Board of the Mathematical Sciences, 1977.

155 (VI.2) Foundations of Geometry

A. Introduction

Geometry deals with figures. It depends, therefore, on our spatial intuition, but our intuition lacks objectivity. The Greeks originated the idea of developing geometry logically, based on explicitly formulated axioms, without resorting to intuition. From this intention resulted Euclid's *Elements*, which was long considered the perfect model of a logical system. As time passed, however, mathematicians came to notice its imperfections. Since the 19th century especially, with the awakening of a more rigorous critical spirit in science and philosophy, more systematic criticism of the *Elements* began to appear. Non-Euclidean geometry was formulated after reexamination of Euclid's axiom of parallels; but it was also discovered that even as a foundation of Euclidean geometry, Euclid's system of axioms was far from perfect. Various systems of axioms for Euclidean geometry were proposed by mathematicians in the latter half of the 19th century, among them one by D. Hilbert [1], which became the basis of far-reaching studies.

B. Hilbert's System of Axioms

Hilbert took as undefined elements **points** (denoted by $A, B, C, \ldots$), **straight lines** (or simply **lines**, denoted by $a, b, c, \ldots$), and **planes** (denoted by $\alpha, \beta, \gamma, \ldots$). Between these objects there exist incidence relations (expressed in phrases such as "A lies on a," "a passes through A," etc.); order relations ("B is between A and C"); congruence relations; and parallel relations. The relations are subject to the following five groups of axioms:

(I) **Incidence axioms**: (1) For two points A, B, there exists a line a through A and B. (2) If $A \neq B$, the line a through A, B is uniquely determined. We write $a = A \cup B$ and call a the **join** of A, B. (3) Every line contains at least two different points. There exist at least three points that do not lie on a line. (4) If A, B, C are points not on a line, there exists a plane α through A, B, C. (We also say that A, B, C lie on α.) For every plane α, there exists at least one point A on α. (5) If A, B, C are points

not on a line, the plane α through A, B, C is uniquely determined. We write $\alpha = A \cup B \cup C$ and call α the **join** of A, B, C. (6) If A, B are two different points on a line a and if A, B lie on a plane α, then every point on a lies on α. (We say that a lies on α or α passes through a.) (7) If a point A lies on two planes α, β, there exists at least one other point B on α and β. (8) There exist at least four points not lying on a plane.

(II) **Ordering axioms**: (1) If B is between A and C, then A, B, C are three different points lying on a line; also, B is between C and A. (2) If A, C are two different points, then there exists a point B such that C is between A and B. (3) If B is between A and C, then A is not between B and C.

We define a **segment** as a set of two different points A, B, denoted as AB or BA, and we call A and B **ends** of this segment. The set of points between A, B is called the **interior** of AB, and the set of points of $A \cup B$ that are neither ends nor interior points of AB is called the **exterior** of AB.

(4) Let A, B, C be three points not lying on a line. If a line a on the plane $A \cup B \cup C$ does not pass through A, B, or C, but passes through a point of the interior of AB, then it also passes through a point of the interior of BC or CA (**Pasch's axiom**).

The following propositions are proved from the above axioms. Given n points $A_1, A_2, \ldots, A_n$ on a line ($n > 2$), we can rearrange them, if necessary, so that the point A_j is between A_i and A_k whenever we have $1 \leqslant i < j < k \leqslant n$. There are exactly two ways of arranging the points in this manner (**theorem of linear ordering**). Let O be a point on a line a, and let A, B be two points on the line different from O. Write $A \sim B$ when $A = B$ or O is not between A and B; write $A \nsim B$ otherwise. Then $\sim$ is an equivalence relation between points on the line different from O; from $A \nsim B$, $A \nsim C$ it follows that $B \sim C$. We say that A, B are on the same **side** or on different sides of O on a depending on whether $A \sim B$ or $A \nsim B$. Two subsets a' and a'' of a defined by $a' = \{A' \mid A \sim A'\}$, $a'' = \{A'' \mid A \nsim A''\}$ are called **half-lines** or **rays** on a with O as the extremity (or starting from O). Denoting by a, for simplicity, the set of points on a, we have $a = a' \cup \{O\} \cup a''$ (disjoint union).

Using axiom II.4, we can also prove the following: Let a be a line on α, and let A, B be two points on α not lying on a. If $A = B$ or if the interior of the segment AB has no point in common with a, we say that A, B are on the same **side** of a on α, and write $A \sim B$. Otherwise, we say that A, B are on different sides of a on α and write $A \nsim B$. Then $\sim$ is an equivalence relation between points on α not lying on a, and from $A \nsim B$, $A \nsim C$ follows $B \sim C$. The subsets $\alpha' = \{A' \mid A \sim A'\}$, $\alpha'' = \{A'' \mid A \nsim A''\}$ of α are called **half-planes** on α **bounded** by a. Again, denoting by α and a the set of points on α and the set on a, respectively, we obtain $\alpha = \alpha' \cup a \cup \alpha''$ (disjoint union).

(III) **Congruence axioms**: Two segments AB, $A'B'$ can be in a relation of **congruence**, expressed symbolically as $AB \equiv A'B'$. (Segments AB and $A'B'$ are then said to be **congruent**. Since the segment AB is defined as the set $\{A, B\}$, the four relations $AB \equiv A'B'$, $BA \equiv A'B'$, $AB \equiv B'A'$, $BA \equiv B'A'$ are equivalent.) This relation is subject to the following three axioms: (1) Let A, B be two different points on a line a, and A_1 a point on a line a_1 (a_1 may or may not be equal to a). Let a_1' be a ray on a_1 starting from A_1. Then there exists a unique point B_1 on a_1' such that $AB \equiv A_1B_1$. (2) From $A_1B_1 \equiv AB$ and $A_2B_2 \equiv AB$ follows $A_1B_1 \equiv A_2B_2$. (Hence it follows that $\equiv$ is an equivalence relation between segments.) (3) Let A, B, C be three points such that B is between A and C, and let A_1, B_1, C_1 be three points such that B_1 is between A_1 and C_1. Then from $AB \equiv A_1B_1$, $BC \equiv B_1C_1$ follows $AC \equiv A_1C_1$.

Now let h, k be two different lines in a plane α and through a point O, and let h', k' be the rays on h, k starting from O. The set of two such rays h', k' is called an **angle** in α, denoted by $\angle(h', k')$ or $\angle(k', h')$. This angle is also denoted by $\angle AOB$, where A, B are points of h', k', respectively. The rays h', k' are called the **sides** and the point O is called the **vertex** of this angle. Then h' is a subset of a half-plane on α bounded by k, and k' is a subset of a half-plane on α bounded by h. The intersection of these two half-planes is called the **interior** of this angle, and the subset of $\alpha - O$ consisting of points belonging to neither the inside nor the sides of the angle is called the **exterior** of the angle. Between two angles $\angle(h', k')$, $\angle(h_1', k_1')$ there may exist the relation of congruence, again expressed by the symbol $\equiv$, as in the case of segments, and subject to the following two axioms: (4) Let $\angle(h', k')$ be an angle on a plane α and h_1 be a line on α_1 (α_1 may or may not be equal to α). Let O_1 be a point on h_1, h_1' a ray on α_1 starting from O_1, and α_1' a half-plane on α_1 bounded by h_1. Then there exists a unique ray k_1' starting from O_1 and lying in α_1' such that $\angle(h', k') \equiv \angle(h_1', k_1')$. Moreover, $\angle(h', k') \equiv \angle(h', k')$ always holds. (Hence it follows that $\equiv$ is an equivalence relation between angles.) (5) Let both A, B, C and A_1, B_1, C_1 be triples of points not lying on a line. Then from $AB \equiv A_1B_1$, $AC \equiv A_1C_1$, and $\angle BAC \equiv \angle B_1A_1C_1$, it follows that $\angle ABC \equiv \angle A_1B_1C_1$.

(IV) **Axiom of parallels**: Suppose that a, b are two different lines. Then it follows from axiom I.2 that if a and b share a point P, such a point is the unique point lying on both a and

b. In this case we say that a, b **intersect** at P and write $a \cap b = P$. On the other hand, if a and b have no point in common and if a, b are on the same plane, we say that a, b are **parallel** and write $a /\!/ b$. If A, a are on a plane α and A is not on a, we can prove (utilizing axioms I, II, and III) that there exists a line b passing through A in α such that $a /\!/ b$. The axiom of parallels postulates the uniqueness of such a b.

(V) **Axioms of continuity**: (1) Let AB, CD be two segments. Then there exist a finite number of points $A_1, A_2, \ldots, A_n$ on $A \cup B$ such that $CD \equiv AA_1 \equiv A_1A_2 \equiv \ldots \equiv A_{n-1}A_n$ and B is between A and A_n (**Archimedes' axiom**). (2) The set of points on a line a (again denoted for simplicity by a) is "maximal" in the following sense: It should satisfy axioms II.1–II.3, III.1, V.1, and the theorem of linear ordering. If $\bar{a}$ is a set of points satisfying these axioms such that $\bar{a} \supset a$, then $\bar{a}$ should be $= a$ (**axiom of linear completeness**). Hence follows the **theorem of completeness**: the set of points, lines, and planes is maximal in the sense that it is not possible to add further points, lines, or planes to this set with the resulting set still satisfying axioms I–IV and V.1.

C. Consistency

In formulating the above axioms and proving their consistency, Hilbert assumed the consistency of the theory of real numbers (→ 156 Foundations of Mathematics). To prove consistency, Hilbert constructed a model for the above axioms using the method of analytic geometry. He defined points as triples of real numbers (x_1, x_2, x_3), lines and planes as sets of points satisfying suitable systems of linear equations, and relations of ordering, congruence, and parallelism in the usual way. It is easy to verify that such a system satisfies all the axioms I–V. Thus the consistency of these axioms is reduced to the consistency of the theory of real numbers (→ 35 Axiom Systems).

A model for I–IV and V.1 can be obtained in the countable field $\mathbf{R}_0$ of all real †algebraic numbers instead of $\mathbf{R}$. Then $\mathbf{R}_0$ can be further restricted to its subfield $\mathbf{P}_0$ defined as follows: Let F be an arbitrary field. An †extension of F of the form $F(\sqrt{1+\lambda^2})$ with $\lambda \in F$ is called a **Pythagorean extension** of F, and F is said to be a **Pythagorean field** if any Pythagorean extension of F coincides with F (e.g., $\mathbf{R}_0$ and $\mathbf{R}$ are Pythagorean). It is easily verified that I–IV are satisfied in the "analytic geometry over any Pythagorean field." On the other hand, we can construct a minimal Pythagorean field containing a given field (the **Pythagorean closure** of the field) in the same way as we construct the †algebraic closure of a field. The field $\mathbf{P}_0$ is defined as the Pythagorean closure of the field $\mathbf{Q}$ of rational numbers.

D. Independence of Axioms

In Hilbert's system, the axioms I and II are used to formulate further axioms. On the other hand, it can be shown that each of the groups III, IV, and V is independent from other axioms.

The independence of IV is shown by the consistency of non-Euclidean geometry (→ 285 Non-Euclidean Geometry). The following model shows the independence of III.5: In the analytic model for I–V, we replace the definition of distance between two points $(x_1, x_2, x_3), (y_1, y_2, y_3)$ by

$$((x_1 - y_1 + x_2 - y_2)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2)^{1/2}.$$

Then III.5 does not hold, while all other axioms remain satisfied. The independence of V.2 is shown by the geometry over $\mathbf{R}_0$ or $\mathbf{P}_0$. The independence of V.1 follows from the existence of the non-Archimedean Pythagorean field: the Pythagorean closure of any †non-Archimedean field (e.g., the field of rational functions of one variable over $\mathbf{Q}$ with a †non-Archimedean valuation) is such a field. A geometry in which V.1 does not hold is called a **non-Archimedean geometry**.

E. Completeness of the System of Axioms and Relations between Axioms

The †completeness of the system of axioms I–V can be shown by introducing coordinates in the geometry with these axioms and representing it as †Euclidean geometry of three dimensions. Axiom group V is essential for the introduction of coordinates over $\mathbf{R}$. Moreover, we have the following results:

(i) The geometry with the axioms I–IV can be represented as "Euclidean geometry" of three dimensions over a Pythagorean field, and the geometry with axioms I.1–I.3, II–IV can be represented as "Euclidean geometry" of two dimensions over a Pythagorean field.

(ii) The geometry with II, and a stronger axiom of parallels IV* (given a line a and a point A outside a, there exists one and only one line a' passing through A that is parallel to a) can be represented as an †affine geometry over a field K that is not necessarily commutative.

(iii) The field K is commutative if and only if the following holds: Suppose that in Fig. 1 $A' \cup B /\!/ A \cup B'$, $B' \cup C /\!/ B \cup C'$. Then it follows that $A' \cup C /\!/ A \cup C'$ (**Pascal's theorem**).

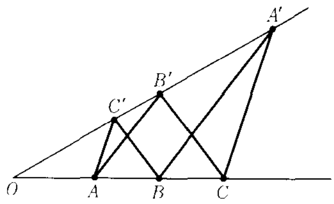


Fig. 1

(iv) The "two-dimensional geometry" with I.1–I.3, II, and IV* can be embedded in the "three-dimensional geometry" with axioms I, II, and IV* if and only if the following holds: Suppose that in Fig. 2 we have $A \cup B \,/\!/\, A' \cup B'$, $B \cup C \,/\!/\, B' \cup C'$. Then it follows that $C \cup A \,/\!/\, C' \cup A'$ (**Desargues's theorem**).

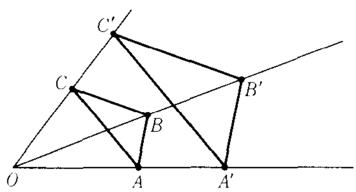


Fig. 2

(v) From I.1–I.3, II, IV*, and Pascal's theorem follows Desargues's theorem.

(vi) Desargues's theorem is independent of I.1–I.3, II, III.1–III.4, IV*, and V; that is, we can construct a **non-Desarguesian geometry** (a geometry in which Desargues's theorem does not hold) in which these axioms are satisfied.

Axioms I, II, and IV*, as well as the theorems of Pascal and Desargues, are propositions in affine geometry. Each has a corresponding proposition in †projective geometry, and the results concerning them can be transferred to the case of projective geometry (→ 343 Projective Geometry).

F. Polygons and Their Areas

Suppose that we are given a finite number of points A_i $(i=0,1,\ldots,r)$ in the geometry with axioms I and II. Then the set of segments (or, more precisely, the union of segments together with their interiors) A_iA_{i+1} $(i=0,1,\ldots,r-1)$ is called a **broken line** joining A_0 with A_r. In particular, if $A_0=A_r$, then this set is called a **polygon** with **vertices** A_i and **sides** A_iA_{i+1}. A polygon with r vertices is called an r-gon. (For $r=3,4,5,6$, r-gons are called **triangles, quadrangles, pentagons**, and **hexagons**, respectively.) A **plane polygon** is a polygon whose vertices all lie on a plane. A polygon is called **simple** if any three consecutive vertices do not lie on a line, and two sides A_iA_{i+1} and A_jA_{j+1} $(i \neq j)$ meet only when $j=i+1$ or $i=j+1$. In this article, we consider only simple plane polygons, and refer to them simply as polygons.

†Jordan's theorem implies that a polygon in the sense just defined divides the plane into two parts, its **interior** and its **exterior**. This special case of Jordan's theorem can be proved by I.1–I.3 and II only. A polygon P is divided into two polygons P_1, P_2 by a broken line joining two points on sides of P and lying in the interior of P (Fig. 3). In this case, we say that P is **decomposed** into P_1, P_2 and write $P=P_1+P_2$. We may again decompose P_1, P_2 and thereby arrive at a decomposition of the form $P=P_1+\ldots+P_k$. Axiom III is used to introduce the congruence relation $\equiv$ between polygons. Two polygons P, Q are called **decomposition-equal** if there exist decompositions $P=P_1+\ldots+P_k$, $Q=Q_1+\ldots+Q_k$ such that $P_1\equiv Q_1,\ldots,P_k\equiv Q_k$. This is expressed by PzQ. We call P, Q **supplementation-equal** if there exist two polygons P', Q' such that $(P+P')z(Q+Q')$, $P'zQ'$. This will be expressed by PeQ. If we assume IV, we can use result (i) of Section E. Let K be the ground field of the geometry (K is Pythagorean, hence †ordered). The **area** of polygon P is defined as the positive element $m(P)$ of K assigned to P such that $m(P+Q)=m(P)+m(Q)$, and $m(P)=m(P')$ if $P\equiv P'$. From PzQ or PeQ, it follows that $m(P)=m(Q)$. Under these axioms, it is proved that $m(P)=m(Q)$ implies PeQ. If we also assume V.1, then $m(P)=m(Q)$ implies PzQ. Thus the theory of area of polygons can be constructed without assuming axiom V.2, though this result cannot be generalized to higher-dimensional cases. For the case of three dimensions, we can construct two solids of the same volume that are not supplementation-equal [2, 7].

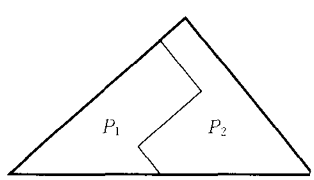


Fig. 3

G. Geometric Construction by Ruler and a Transferrer of Constant Lengths

The geometry with I–IV can be represented as 3-dimensional Euclidean geometry over a Pythagorean field. Conversely, all these axioms are valid in 3-dimensional Euclidean geometry over any Pythagorean field. Thus the minimal system of "quantities" whose existence is assured in geometry with these axioms is the field $\mathbf{P}_0$, the Pythagorean closure of $\mathbf{Q}$. Hilbert noticed that the existence of a

geometric object under axioms I–IV can be expressed as its constructibility by **ruler** (i.e., an instrument to draw a straight line joining two points) and a **transferrer of constant lengths**. The latter, for a constant length x, is an instrument that permits finding the point X on the given ray AB such that $AX=x$. It is not possible to construct by ruler and transferrer all the points that can be constructed by means of ruler and compass (→ 179 Geometric Construction). However, it is possible to construct all the lengths λx, where λ is any element of $\mathbf{P}_0$. Hilbert conjectured that an element of $\mathbf{P}_0$ can be characterized as a †totally positive algebraic number of degree 2^ν, $\nu \in \mathbf{N}$. This conjecture was proved by Artin [3].

H. Related Topics

While Hilbert's foundations are concerned with 3-dimensional Euclidean geometry, it is easy to generalize these results to the case of n-dimensional Euclidean geometry (→ 139 Euclidean Geometry). Also, for affine and projective geometries, there are well-organized systems of axioms (at least for the case of dimensions $\geqslant 3$). Hilbert [1, Appendix III] showed that plane †hyperbolic geometry can be constructed on a modified system of axioms, but for other non-Euclidean geometries (in particular, †elliptic geometries) there are no known systems of axioms as good as Hilbert's for the Euclidean case. On the other hand, Hilbert [1, Appendix IV] gives another method of constructing Euclidean geometry in characterizing the group of motions as the topological group with certain properties. G. Thomsen [4] rewrote Hilbert's system of axioms in group-theoretical language utilizing the fact that the group of motions is generated by symmetries with respect to points, lines, and planes. Finally, Hilbert's study of the foundations of geometry led him to research in the †foundations of mathematics.

References

[1] D. Hilbert, Grundlagen der Geometrie, Teubner, 1899, seventh edition, 1930.
[2] M. Dehn, Über den Rauminhalt, Math. Ann., 55 (1902), 465–478.
[3] E. Artin, Über die Zerlegung definiter Funktionen in Quadrate, Abh. Math. Sem. Univ. Hamburg, 5 (1926), 100–115 (Collected papers, Addison-Wesley, 1965, 273–288).
[4] G. Thomsen, Grundlagen der Elementargeometrie, Teubner, 1933.
[5] K. Reidemeister, Grundlagen der Geometrie, Springer, 1968.
[6] G. Hessenberg and J. Diller, Grundlagen der Geometrie, de Gruyter, 1967.
[7] V. G. Boltyanskiĭ, Hilbert's third problem, Winston, 1978.

156 (I.1) Foundations of Mathematics

A. General Remarks

The notion of †set, introduced toward the end of the 19th century, has proved to be one of the most fundamental and useful ideas in mathematics. Nonetheless, it has given rise to well-known †paradoxes. Based on this notion, R. Dedekind developed the theories of natural numbers [2] and real numbers [3], defining the latter as "cuts" of the set of rational numbers. Thus set theory served as a unifying principle of mathematics.

It has been noted, however, that some of the most commonly utilized arguments in set theory, which are at the same time the most useful in mathematics and belong almost to the basic framework of formal logic itself, resemble very much those which give rise to paradoxes. This fact has caused many critical mathematicians to question the very nature of mathematical reasoning. Thus a new field, **foundations of mathematics**, came into being at the beginning of this century. This field was divided at its inception into different doctrines according to the views of its initiators: **logicism** by B. Russell, **intuitionism** by L. E. J. Brouwer, and **formalism** by D. Hilbert. In set theory, which was the origin of this controversy, it was pointed out that the "definition" of set as given by G. Cantor was too naive, and axiomatic treatments of this theory were proposed (→ 33 Axiomatic Set Theory).

B. Logicism

Russell asserted that mathematics is a branch of logic and that paradoxes come from neglecting the "types" of concepts. According to his opinion, mathematics deals formally with structures independently of their concrete meanings. Science of this character has been called logic from antiquity. According to him, logic is the youth of mathematics, and mathematics is the manhood of logic. To construct mathematics from this standpoint, asserted Russell, ordinary language is lengthy and inaccurate, and some proper system of symbols should be used instead. Thus he tried to reconstruct mathematics using †symbolic logic.

Attempts to reorganize mathematics using logical symbols had formerly been made by G. Leibniz, who wrote *Dissertatio de arte combinatoria* in 1666, as well as by A. de Morgan, G. Boole, C. S. Peirce, E. Schröder, G. Frege, G. Peano, and others. Symbols used by the last two authors resemble those of today. Russell studied these works and published his own theory in a monumental joint work with A. N. Whitehead: *Principia mathematica* (3 vols., 1st ed. 1910–1913, 2nd ed. 1925–1927), in which the theories of natural numbers and real numbers as well as analytic geometry are developed from the fundamental laws of logic.

If this work had been completely successful, it could have eliminated any possibility of the intrusion of paradoxes into mathematics. However, the authors were forced to postulate an "unsatisfactory" axiom in order to construct mathematics. They introduced the notion of †type as follows: An object M defined as the set of all objects of a certain type belongs to a higher type than the types of the elements of M. This serves to eliminate certain paradoxes but brings about inconveniences such as the following. Suppose that we are trying to construct the theory of real numbers from that of rational numbers. Each real number can then be considered a †predicate about rational numbers. If this predicate contains only †quantifiers relating to variables running over all rational numbers, then the corresponding real number is said to be **predicative**, otherwise **impredicative**. According to Russell, the latter should have a higher type than the former, which makes the theory of real numbers exceedingly complicated. To avoid this difficulty, Russell proposed the **axiom of reducibility**, which says that every predicate can be replaced by a predicative one. With this rather artificial axiom Russell himself expressed dissatisfaction. Russell also postulated the †axiom of infinity and the †axiom of choice, which are also problematic. After examining the philosophical background of the book, H. Weyl wrote about *Principia mathematica*, "Mathematics is no more based on logic than the utopia built by the logician." Nevertheless, logic as formulated in this book, as well as the theory of types as developed by F. P. Ramsay in the school of Russell and Whitehead, is still an important subject of mathematical logic.

C. Intuitionism

The intuitionist claims that mathematical objects or truths do not exist independently from mathematically thinking spirit or intuition, and that these objects or truths should be directly seized by mental or intuitional activity. The philosophical standpoints of mathematicians such as L. Kronecker and H. Poincaré in the 19th century or E. Borel, H. Lebesgue, and N. N. Luzin at the turn of this century can be assimilated to intuitionism, but those of the latter three are often said to belong to **semi-intuitionism** or to **French empiricism**. Brouwer took a narrower standpoint, strongly antagonistic to Hilbert's formalism. Today the word "intuitionism" is generally interpreted in Brouwer's sense.

Brouwer sharply criticized the usual way of reasoning in mathematics and claimed that indiscriminate use of the **law of excluded middle** (or **tertium non datur**) $P \vee \neg P$ cannot be permitted. According to him, the proposition "Either there exists a natural number with a given property P, or else no such number exists" is to be regarded as proved only when an actual construction of a natural number with the property P is given or when the absurdity of the existence of such a natural number can be constructively proved. When neither of these two results can be shown, then one can say nothing about the truth of the above proposition. Thus the usual method of proof, known as the method of **reductio ad absurdum**, i.e., of proving a proposition P by proving its double negation $\neg\neg P$, is not generally considered valid. It is a difficult but important problem of mathematical logic to determine which parts of usual mathematics can be reconstructed intuitionistically, though it does not seem easy to reconstruct any part of mathematics elegantly from this standpoint.

D. Formalism

To eliminate paradoxes, Hilbert tried to apply his axiomatic method. From Hilbert's standpoint, any part of mathematics is a deductive system based on its axioms. In the deductive development, however, "logic," including set theory and elementary number theory, is used. Paradoxes appear already in such logic. Hilbert's idea was to axiomatize such logic and to prove its consistency. Thus one must first formalize the most elementary part of mathematics, including logic proper.

Hilbert proved the consistency of Euclidean geometry by assuming the consistency of the theory of real numbers. This is an example of a **relative consistency** proof, which reduces the consistency proof of one system to that of another. Such a proof can be meaningful only when the latter system can somehow be regarded as based on sounder ground than the former. To carry out the **consistency proof** of logic proper and set theory, one must reduce it to that of another system with sounder

ground. For this purpose, Hilbert initiated **metamathematics** and the **finitary standpoint**.

The finitary standpoint recognizes as its foundation only those facts that can be expressed in a finite number of symbols and only those operations that can be actually executed in a finite number of steps. Essentially, it does not differ from the standpoint of intuitionism. The methods based on this standpoint are also called **constructive methods**.

Metamathematics is also called **proof theory**. Its subject of research is mathematical proof itself. Hilbert was the first to insist on its importance. The theory is indispensable for consistency proofs of mathematical systems, but it may also be used for other purposes. In fact, the same idea can be seen in the †duality principle of projective geometry, which dates from long before Hilbert's proclamation of formalism. This is not a theorem of projective geometry deduced from its axioms; rather, it is a proposition about the theorems in projective geometry, based on the type of axioms and proofs in this subject.

According to Hilbert's method, one must develop proof theory from the finitary standpoint with the aim of proving the consistency of axiomatized mathematics. For this purpose, one must formalize the mathematical theory in question by means of symbolic logic. A theory thus formalized is called a **formal system**.

E. Some Results of Formalist Theory

One of the most remarkable results hitherto obtained with Hilbert's method is the consistency proof of pure number theory by G. Gentzen [7]. This consistency proof covers the largest domain for which an explicit consistency proof has so far been obtained. However, the methods of formalist proof theory have proved to be most effective in studying the logical structure of mathematical theories and have led to various results on the consistency of formalized mathematical systems, on symbolic logic, and on axiomatic set theory. We give some examples.

(1) Gödel's Incompleteness Theorem. K. Gödel [6] showed that if a system obtained by formalizing the theory of natural numbers is consistent, then this system contains a †closed formula A such that neither A nor its negation $\neg A$ can be proved within the system. He originally proved this under the assumption that the system is **ω-consistent**. This is a stronger condition for the system than simple consistency, but J. B. Rosser [13] succeeded in replacing this by the latter. This result shows the incompleteness not only of the usual theory of natural numbers but of any consistent theory (from the finitary standpoint) containing the theory of these numbers.

At the same time, Gödel also obtained the following important result: Let S be any consistent formal system containing the theory of natural numbers. Then it is impossible to prove the consistency of S by utilizing only arguments that can be formalized in S. This means that a consistency proof from the finitary standpoint of a formal system S inevitably necessitates some argument that cannot be formalized in S.

(2) Consistency Proofs for Pure Number Theory. Gentzen [7] called **pure number theory** the theory of natural numbers not depending on the free use of set theory (differing consequently from the usual theory of natural numbers based on †Peano axioms; → 294 Numbers) and proved its consistency. W. Ackermann [14] proved the consistency of a similar theory admitting the use of Hilbert's †ε-symbol. G. Takeuti [15] showed that Gentzen's result can be obtained as a corollary to his theorem extending †Gentzen's fundamental theorem on †predicate logic of the first order to a †theory of types of a certain kind.

According to the result of Gödel mentioned in (1) above, some reasoning outside pure number theory must be used to prove its consistency. In all consistency proofs of pure number theory mentioned above, †transfinite induction up to the first †ε-number ε_0 is used, but all the other reasoning used in these proofs can be presented in pure number theory. This shows that the legitimacy of transfinite induction up to ε_0 cannot be proved in this latter theory. A direct proof of this fact was given by Gentzen [13]. On the other hand, the legitimacy of transfinite induction up to an ordinal number $<\varepsilon_0$ can be proved within pure number theory.

Again, transfinite induction is not the only method by which to prove the consistency of pure number theory. Actually, Gödel [17] carried out the proof utilizing what he called computable functions of finite type on natural numbers and what we call primitive recursive functionals of finite type.

By restricting pure number theory further, one obtains weaker theories of natural numbers whose consistency can be proved with finitary methods without recourse to such methods as transfinite induction up to ε_0. M. Presburger [18] proved the consistency of a theory in which only the addition of numbers is considered an operation. Ackermann [19], J. von Neumann [20], J. Herbrand [21], and K. Ono [22] proved the consistency of theories in which some restrictions are placed

on the use of the axiom of †mathematical induction.

On the other hand, K. Schütte [23] gave a consistency proof for number theory including what he called "infinite induction" from a stronger standpoint than Hilbert's finitary one; he attempted to find a basis that makes such a proof possible.

(3) The Consistency of Analysis. No definitive result has yet been obtained from the standpoint of formalism, though many attempts are being made, among which a recent one by C. Spector [24] should be mentioned.

(4) Axiomatic Set Theory. There are different kinds of axiom systems (→ 33 Axiomatic Set Theory). To give a consistency proof for any of these systems is considered a very difficult problem today, but many interesting results are known concerning the relative consistency or independence of these axioms.

(5) The Skolem-Löwenheim Theorem. The metamathematical Skolem-Löwenheim theorem states: Given a consistent system of axioms stated in the first-order predicate logic whose cardinality is at most countable, there always exists an †object domain consisting of countable objects satisfying all these axioms.

For example, axiomatic set theory is stated in predicate logic of the first order, and the cardinality of its axioms is countable. Thus there exists an object domain consisting of countable objects satisfying all these axioms, provided that they are consistent. Such a domain is called a **countable model** of axiomatic set theory. On the other hand, from the axioms of this theory one can prove that there exists a family of sets that is more than countable. This should also hold in a model of the theory, in which each object represents a set. This situation is known as the **Skolem paradox**.

This does not imply, however, the inconsistency of axiomatic set theory. In fact, the term "countable" is to be interpreted in its mathematical sense when one says "there exists a family of sets that is more than countable," while it should be interpreted in its metamathematical sense when one speaks of a countable model of the theory. It is the confusion of these two different interpretations that leads to the "paradox."

(6) Skolem's Theorem on the Impossibility of Characterizing the System of Natural Numbers by Axioms. T. Skolem [27] proved that it is not possible to characterize the system of natural numbers by a countable system of axioms stated in the predicate logic of the first order. More precisely, given any consistent countable system of axioms satisfied by the system of natural numbers, there always exists another †linearly ordered system satisfying all these axioms and yet not isomorphic to the system of natural numbers as an ordered system.

Gödel's incompleteness theorem and the Skolem paradox, as well as this result, seem to indicate that there is a certain limit to the effectiveness of the formalist method. On the other hand, nonstandard analysis has originated in this result.

References

[1] S. C. Kleene, Introduction to metamathematics, Van Nostrand, 1952.
[2] R. Dedekind, Was sind und was sollen die Zahlen? Vieweg, 1888.
[3] R. Dedekind, Stetigkeit und irrationale Zahlen, Vieweg, 1872.
[4] B. Russell, Introduction to mathematical philosophy, Allen & Unwin and Macmillan, 1919.
[5] A. Heyting, Intuitionism, North-Holland, 1956.
[6] K. Gödel, Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I, Monatsh. Math. Phys., 38 (1931), 173–198.
[7] G. Gentzen, Die Widerspruchsfreiheit der reinen Zahlentheorie, Math. Ann., 112 (1936), 493–565.
[8] D. Hilbert and P. Bernays, Grundlagen der Mathematik, Springer, I, second edition, 1968; II, 1970.
[9] A. Tarski, Logic, semantics, metamathematics, Clarendon Press, 1956.
[10] S. C. Kleene, Mathematical logic, Wiley, 1967.
[11] J. R. Schoenfield, Mathematical logic, Addison-Wesley, 1967.
[12] O. Becker, Grundlagen der Mathematik in geschichtlicher Entwicklung, Alber, 1954.
[13] J. B. Rosser, Extensions of some theorems of Gödel and Church, J. Symbolic Logic, 1 (1936), 87–91.
[14] W. Ackermann, Zur Widerspruchsfreiheit der Zahlentheorie, Math. Ann., 117 (1940), 162–194.
[15] G. Takeuti, On the fundamental conjecture of GLC I, J. Math. Soc. Japan, 7 (1955), 249–275.
[16] G. Gentzen, Beweisbarkeit und Unbeweisbarkeit von Anfangsfällen der transfiniten Induktion in der reinen Zahlentheorie, Math. Ann., 119 (1943), 140–161.
[17] K. Gödel, Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes, Dialectica, 12 (1958), 280–287.

[18] M. Presburger, Über die Vollständigkeit eines gewissen Systems der Arithmetik ganzer Zahlen, in welchem die Addition als einzige Operation hervortritt, C. R. du I Congrès des Math. des Pays Slaves, Warsaw (1929), 92–101, 395.
[19] W. Ackermann, Begründung des "tertium non datur" mittels der Hilbertschen Theorie der Widerspruchsfreiheit, Math. Ann., 93 (1924), 1–36.
[20] J. von Neumann, Zur Hilbertschen Beweistheorie, Math. Z., 26 (1927), 1–46.
[21] J. Herbrand, Sur la non-contradiction de l'arithmétique, J. Reine Angew. Math., 166 (1931), 1–8.
[22] K. Ono, Logische Untersuchungen über die Grundlagen der Mathematik, J. Fac. Sci. Univ. Tokyo, 3 (1938), 329–389.
[23] K. Schütte, Beweistheoretische Erfassung der unendlichen Induktion in der Zahlentheorie, Math. Ann., 122 (1951), 369–389.
[24] C. Spector, Provably recursive functionals of analysis: a consistency proof of analysis by an extension of principles formulated in current intuitionistic mathematics, Recursive function theory, Amer. Math. Soc. Proc. Symp. Pure Math., 5 (1962). 1–27.
[25] L. Löwenheim, Über Möglichkeiten in Relativkalkül, Math. Ann., 76 (1915), 447–470.
[26] T. Skolem, Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre, 5 Congress der Skandinavischen Mathematiker, Helsingfors (1923), 217–232.
[27] T. Skolem, Über die Nichtcharakterisierbarkeit der Zahlenreihe mittels endlich oder abzählbar unendlich vieler Aussagen mit ausschliesslich Zahlenvariablen, Fund. Math., 23 (1934), 150–161.

157 (VI.22) Four-Color Problem

A. Brief History

It is obvious that four colors are necessary to color some geographical maps on a sphere (or plane), but are four colors sufficient to color every map? This is the so-called four-color problem. The precise formulation of this problem will be given in Section B. The conjecture was made by Francis Guthrie and communicated through his brother Frederick Guthrie to A. De Morgan in 1852 [4]. A. Cayley called attention to the problem in 1878. It became famous after J. P. Heawood pointed out a mistake in the proof by A. B. Kempe (1879). Heawood also studied the problem of coloring maps on arbitrary surfaces (→ Section E).

Though there have been various approaches to the four-color problem, the main stream of investigation has concentrated on obtaining an unavoidable set consisting of reducible arrangements (→ Section D) in order to correct the mistake made by Kempe. G. D. Birkhoff [7] first discovered the simplest nontrivial reducible arrangement, nowadays called Birkhoff's diamond (Fig. 1). Using all reducible arrangements known up to that time, P. Franklin showed that every map with up to 25 countries must contain a reducible arrangement, so that such a map is four-colorable [8]. This limit has been gradually increased; in 1975 a reducible arrangement for 25 countries was found.

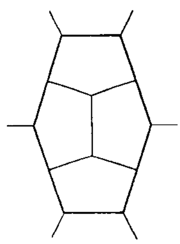


Fig. 1
Birkhoff's diamond.

H. Heesch invented the method of discharging (→ Section D), found criteria for reducibility, and finally conjectured the existence of an unavoidable set of reducible arrangements with several thousand elements, but this was too large to construct by hand. W. Haken and K. Appel with J. Koch worked with high-speed computers for a total of 1,200 CPU hours over a period of four and a half years and finally succeeded in constructing and checking an avoidable set of reducible arrangements with 1,834 elements, which proved the four-color problem affirmatively (1976; [9, 10]). For early investigations of the four-color problem → [5]; for results up to the 1930s → [3].

B. Precise Formulation of the Problem

To formulate the problem precisely, we must state the following two conditions: (i) Every country on a map is a †connected domain; a connected part of the sea is considered to be a country. (ii) Two countries sharing boundary lines must be colored differently. On the other hand, if two countries share only a finite number of points, then they may share the same color.

Actually we can modify the map so that no more than three countries meet at the same point. A map with this condition is called a **trivalent** or **cubic map**. In the study of the four-

color problem, we can restrict ourselves to trivalent maps without two-sided countries. In the remainder of this article, we shall assume that the maps considered have these properties.

In recent investigations, it has become customary to convert maps into their †dual graphs, where each country is replaced by its capital lying inside, and for each adjacent pair A, B (Fig. 2), the boundary is replaced by a line connecting the representative points of A and B in such a way that it meets the boundary only once.

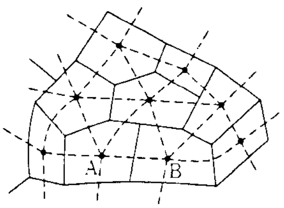


Fig. 2
Dual graph.

The formal extension of the four-color problem to higher dimensions is trivial, since we can easily construct arrangements needing arbitrary many colors for coloring. By converting to a dual graph, one can see that this is not surprising, since the planarity of a graph is a strong restriction, while the condition of representability of a linear graph in 3-dimensional space imposes no restriction.

C. Tait's Algorithm

Suppose that a planar trivalent map M is four-colored. Denote the four colors by 0, 1, 2, and 3. (In Fig. 3, we use A, B, C, and D instead of 0, 1, 2, and 3, respectively). Here we take the operation $a \oplus b \rightarrow c$. Represent the integers $a, b = 0, 1, 2,$ or 3 in the binary (2-adic) number system as 2-digit numbers $a_1 a_0$ and $b_1 b_0$, where a_i and b_i are 0 or 1. For each digit we take the operation $a_i \oplus b_i = c_i$ to be binary

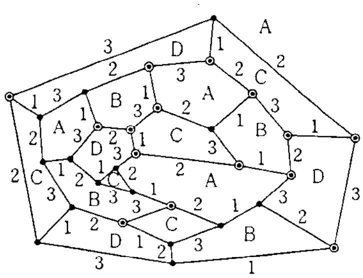


Fig. 3
An example of Tait's algorithm. At the vertices, • stands for + and ⊙ for −. Within the domains, A, B, C, and D stand for four colors corresponding to 0, 1, 2, and 3, respectively.

addition without carrying, i.e., the operation is that of the logical (exclusive) "or." The number $c = c_1 c_0$ in the binary system represents the result of the operation. To each boundary line, we give the number corresponding to $a \oplus b$, where a and b are the colors for the two countries meeting at the boundary. Then we have the numbering 1, 2, and 3 for each edge, where 1, 2, and 3 are labeled once and only once for each triple of edges emanating from each vertex. Such numbering for edges is called **Tait coloring** for edges. Then we give the signature + or − to each vertex, according as the order of the edge numbering is counterclockwise or clockwise. Then the algebraic sum of the signatures along the boundary of each country in the map is always a multiple of 3. Conversely, if we give the signature + or − to each vertex in such a way that the algebraic sum of the signatures is always a multiple of 3 along the boundary of each country, then we get a four-coloring of the map by reversing the above procedure. This is called **Tait's algorithm,** which shows that the four-coloring of a planar map is an †NP problem. As for five-coloring, it is known that an algorithm of polynomial complexity exists.

Tait's algorithm also shows that the four-color problem is equivalent to the following apparently elementary geometric proposition: "For any convex polyhedron, we can always cut near some of its vertices in such a way that the resulting polyhedron has only faces such that the number of sides is a multiple of 3." Many other equivalent statements of the four-color problem are known [1, 2].

D. Solution of the Four-Color Problem

For a given planar trivalent map, denote by V_n the number of the countries with n sides ($n \geqslant 3$). From Euler's theorem on polyhedra, we have immediately the relation

$$\sum_{n \geqslant 3} (6-n) V_n = 12. \qquad (1)$$

We easily see from this that every planar map must contain 3-, 4-, or 5-sided countries. A family **F** of the arrangements of countries with the property that every planar map must contain at least one arrangement belonging to **F** is called an **unavoidable set**. The family consisting of 3-, 4-, and 5-sided countries is the simplest example. In order to obtain an unavoidable set, Heesch invented the method of **discharging**. Let us assume the existence of a map M that contains no arrangement of a family **F**. We give a signed charge of $(6-n)$ to every n-gon in M. Next we divide and move the charges so that the pluses and the minuses cancel out. If all positive charges disappear,

according to the assumption of the existence of the map M, then we have a contradiction of (1), so we can conclude that no such map M exists.

A 3-sided country can be ignored in four-coloring. As for a 4-sided country A, Kempe proved (by means of a definite modification procedure) that after four-coloring outside A, we also have a total four-coloring including A. An arrangement of a country or countries A is called **reducible** if, as in the above case, after four-coloring outside A, we get a total four-coloring including A by suitable modifications. If we have an unavoidable set consisting only of reducible arrangements, then by definition, every planar map will be four-colorable.

Kempe believed that he had proved the reducibility of a pentagon (5-sided country), but unfortunately, he missed a particular case. Even though the pentagon is not reducible, much effort has been given to finding other reducible arrangements. Many useful criteria for reducibility have also been studied. If we have enough reducible arrangements, then we can either eventually obtain an unavoidable set, which proves the four-color problem affirmatively, or find a "minimal counter-example" without reducible arrangements, which disproves it.

Haken obtained an unavoidable set consisting of arrangements without certain kinds of reduction obstructions. (He called these "geographically good arrangements.") He firmly believed that, applying certain "probabilistic conjectures" to such arrangements, he would conquer the problem by considering arrangements of up to 14 countries. With repeated improvements on the discharging process and on the criteria for reducibility, he finally was able to conclude that his speculation was correct, as recounted in Section A. His investigation is not only an example of the use of high-speed computation in pure mathematics, but also one inviting reassessment of the meaning of mathematical proof.

E. Coloring Maps on Arbitrary Surfaces

On a †torus, seven colors are sufficient to color any map and that many colors are necessary to color some maps. Heawood (1890) investigated the problem of coloring maps on closed surfaces with †Euler characteristic $\chi<2$, orientable or not. The least number of colors sufficient to color any map on a surface is called the **chromatic number** of the surface. Heawood proved that the chromatic number is less than or equal to

$$p=[(7+\sqrt{49-24\chi})/2] \qquad (\chi<2), \qquad (2)$$

where $[\alpha]$ means the largest integer $a\leqslant\alpha$. After this was proved in various special cases, J. W. T. Youngs and G. Ringel finally proved in 1968 that the number p equals the chromatic number except for a Klein bottle (a non-orientable surface with $\chi=0$) [6]. Franklin proved in 1934 that the chromatic number for a Klein bottle is 6.

References

[1] O. Ore, The four-color problem, Academic Press, 1967.
[2] T. Saaty and P. Kainen, The four-color problem—assaults and conquest, McGraw-Hill, 1977.
[3] N. L. Biggs (ed.), Graph theory 1736–1936, Oxford Univ. Press, 1976.
[4] K. O. May, The origin of four-colour conjecture, Isis, 56 (1965), 346–348.
[5] W. W. R. Ball, Mathematical recreations and essays, Macmillan, 1892; revised twelfth edition, H. S. M. Coxeter (ed.), Toronto Univ. Press, 1958.
[6] G. Ringel, Map color theorem, Springer, 1974.
[7] G. D. Birkhoff, The reducibility of maps, Amer. J. Math., 35 (1913), 115–128.
[8] P. Franklin, The four-color problem, Amer. J. Math., 44 (1922), 225–236.
[9] K. Appel and W. Haken, Every planar map is four-colorable I. Discharging, Illinois J. Math., 21 (1977), 429–489.
[10] K. Appel, W. Haken, and J. Koch, Every planar map is four-colorable II. Reducibility, Illinois J. Math., 21 (1977), 490–567.

158 (XXI.24) Fourier, Jean Baptiste Joseph

Jean Baptiste Joseph Fourier (March 21, 1768–May 16, 1830) was born in Auxerre, France, the son of a tailor, and was orphaned at the age of eight. In 1790, he was appointed professor at the Ecole Polytechnique. In 1798, Napoleon took him on his Egyptian campaign together with G. Monge. On his return to France, he was made governor of the department of Isère. With the downfall of Napoleon, he lost his position; however, he was later appointed to the French Academy of Science as a result of his research on the transmission of heat. In 1826 he was elected a member of the Académie Française.

His research on heat transmission was begun in 1800. In 1811, he presented a prize-

winning solution to a problem put forth by the Academy of Science. He solved the equation for heat transmission under various †boundary conditions. Fourier also stated (without rigorous proof) that an arbitrary function could be represented by †trigonometric series (→ 159 Fourier Series), a statement that gave rise to subsequent developments in analysis.

References

[1] J. B. J. Fourier, Oeuvres I, II, G. Darboux (ed.), Gauthier-Villars, 1888–1890.

159 (X.22) Fourier Series

Throughout this article, we assume that $f(x)$, $g(x)$, ... are real-valued functions and that integrals are always Lebesgue integrals.

A. Introduction

The set of functions

$$1/\sqrt{2\pi}, \quad \cos x/\sqrt{\pi}, \quad \sin x/\sqrt{\pi}, \ldots,$$
$$\cos kx/\sqrt{\pi}, \quad \sin kx/\sqrt{\pi}, \ldots,$$

which is called the **trigonometric system**, is an †orthonormal system in $(-\pi, \pi)$ (→ 317 Orthogonal Functions). Let $f(x)$ be an element of $L_1(-\pi, \pi)$ (i.e., †Lebesgue integrable in $(-\pi, \pi)$). We put

$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos kt\,dt,$$
$$b_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin kt\,dt, \qquad k = 0, 1, \ldots, \tag{1}$$

and call a_k, b_k the **Fourier coefficients** of f. The formal series

$$\frac{1}{2}a_0 + \sum_{k=1}^{\infty}(a_k\cos kx + b_k\sin kx) \tag{2}$$

is called the **Fourier series** of f and is often denoted by $\mathfrak{S}(f)$. To indicate that a formal series $\mathfrak{S}(f)$, as above, is the Fourier series of a function f, we write

$$f(x) \sim \frac{1}{2}a_0 + \sum_{k=1}^{\infty}(a_k\cos kx + b_k\sin kx).$$

The sign $\sim$ means that the numbers a_k, b_k are connected with f by the formula (1); it does not imply that the series is convergent, still less that it converges to f. Generally, **trigonometric series** are those of form (2), where the a_k, b_k are arbitrary real numbers. Since the trigonometric series have period 2π, we assume that the functions considered are extended for all real x by the condition of periodicity $f(x+2\pi) = f(x)$. The study of the properties of the series $\mathfrak{S}(f)$ and the representation of f by $\mathfrak{S}(f)$ are major objects of the theory of Fourier series. Since $e^{ix} = \cos x + i\sin x$, if we set $2c_k = a_k - ib_k$, $c_{-k} = \bar{c}_k$ $(k = 0, 1, 2, \ldots)$, we have

$$c_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-ikt}\,dt, \qquad k = 0, \pm 1, \ldots.$$

Then $\mathfrak{S}(f)$ is represented by the **complex form** $\sum_{k=-\infty}^{\infty} c_k e^{ikx}$, and $\{e^{ikx}\}$ $(k = 0, \pm 1, \ldots)$ is an orthogonal system in $(-\pi, \pi)$. In this complex form, we take symmetric partial sums such as $\sum_{k=-n}^{n} c_k e^{ikx}$ $(n = 1, 2, \ldots)$.

Consider the power series $\frac{1}{2}a_0 + \sum_{k=1}^{\infty}(a_k - ib_k)z^k$ on the unit circle $z = e^{ix}$ in the complex plane. Its real part is the trigonometric series (2), and the imaginary part (with vanishing constant term) is

$$\sum_{k=1}^{\infty}(a_k\sin kx - b_k\cos kx), \tag{3}$$

which is called the **conjugate series** (or **allied series**) of f and is denoted by $\tilde{\mathfrak{S}}(f)$. In complex form, the conjugate series is $-i\sum_{k=-\infty}^{\infty}(\operatorname{sgn} k)c_k e^{ikx}$.

If f and g belong to $L_1(-\pi, \pi)$ and

$$f(x) \sim \sum_{k=-\infty}^{\infty} c_k e^{ikx}, \qquad g(x) \sim \sum_{k=-\infty}^{\infty} d_k e^{ikx},$$

then

$$\frac{1}{2\pi}\int_0^{2\pi} f(x-t)g(t)\,dt \sim \sum_{k=-\infty}^{\infty} c_k d_k e^{ikx}.$$

The function

$$f * g(x) = \frac{1}{2\pi}\int_0^{2\pi} f(x-t)g(t)\,dt$$

is called the **convolution** of f and g.

If f is †absolutely continuous, then the derivative $f'(x)$ satisfies

$$f'(x) \sim i\sum_{k=-\infty}^{\infty} kc_k e^{ikx}$$
$$= \sum_{k=1}^{\infty} k(b_k\cos kx - a_k\sin kx).$$

If $F(x)$ is an indefinite integral of f, then

$$F(x) - c_0 x \sim C + {\sum_{k=-\infty}^{\infty}}' \frac{c_k}{ik}e^{ikx}$$
$$= C + \sum_{k=1}^{\infty}\frac{a_k\sin kx - b_k\cos kx}{k},$$

where C is a constant of integration and the symbol $'$ indicates that the term $k = 0$ is omitted from the sum.

If $f \in L_1(-\pi, \pi)$, then the Fourier coefficients

c_n converge to 0 as $n \to \infty$ (**Riemann-Lebesgue theorem**). If f satisfies the †Lipschitz condition of order α $(0 < \alpha \leqslant 1)$, then $c_n = O(n^{-\alpha})$, and if f is of †bounded variation, then $c_n = O(n^{-1})$. When $f \in L_2(-\pi, \pi)$,

$$\frac{1}{\pi}\int_0^{2\pi} |f(x)|^2 dx = 2\sum_{k=-\infty}^{\infty} |c_k|^2 = \frac{1}{2}a_0^2 + \sum_{k=1}^{\infty}(a_k^2 + b_k^2),$$

which is called the **Parseval identity**. If $\sum |c_k|^2 < \infty$, then there exists a function $f \in L_2(-\pi, \pi)$ which has the c_k as its Fourier coefficients. This converse is implied by the †Riesz-Fischer theorem (→ Appendix A, Table 11.I).

B. Convergence Tests

The nth partial sums $s_n(x) = s_n(x; f)$ of the Fourier series $\mathfrak{S}(f)$ can be written in the form

$$s_n(x) = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x+t) D_n(t)\, dt,$$

where

$$D_n(t) = \{\sin(n+1/2)t\}/2\sin(t/2).$$

The function $D_n(t)$ is called the **Dirichlet kernel**. For a fixed point x we set $\varphi_x(t) = f(x+t) + f(x-t) - 2f(x)$; then

$$s_n(x) - f(x) = \frac{1}{\pi}\int_0^{\pi} \varphi_x(t) D_n(t)\, dt.$$

Hence if the integral on the right-hand side tends to zero as $n \to \infty$, $\lim_{n\to\infty} s_n(x) = f(x)$. If f vanishes in an interval $I = (a, b)$, then $\mathfrak{S}(f)$ converges uniformly in any interval $I' = (a + \varepsilon, b - \varepsilon)$ interior to I, and the sum of $\mathfrak{S}(f)$ is 0. This is called the **principle of localization**.

Here we give four convergence tests. (1) If f is of bounded variation, $\mathfrak{S}(f)$ converges at every point x to the value $\{f(x+0) + f(x-0)\}/2$. In addition, if f is continuous at every point of a closed interval I, $\mathfrak{S}(f)$ is uniformly convergent in I (**Jordan's test**). As a special case of this test, bounded functions having a finite number of maxima and minima and no more than a finite number of points of discontinuity have convergent Fourier series (**Dirichlet's test**). (2) If the integral $\int_0^{\pi} |\varphi_\chi(t)|/t\, dt$ is finite, then $\mathfrak{S}(f)$ converges at x to $f(x)$ (**Dini's test**). (3) If

$$\int_0^h |\varphi_x(t)|\, dt = o(h)$$

and

$$\lim_{\eta\to 0}\int_\eta^{\pi} \frac{|\varphi_x(t) - \varphi_x(t+\eta)|}{t}\, dt = 0,$$

then $\mathfrak{S}(f)$ converges at x to $f(x)$ (**Lebesgue's test**). Jordan's and Dini's tests are mutually independent, and both are included in Lebesgue's test, which, although not as convenient in certain cases, is quite powerful. (4) If $f(x)$ is continuous in (a, b) and its modulus of continuity satisfies the condition $\omega(\delta)\cdot\log(1/\delta) \to 0$ as $\delta \to 0$ in this interval, then $\mathfrak{S}(f)$ converges uniformly in $(a+\varepsilon, b-\varepsilon)$ (**Dini-Lipschitz test**).

C. Summability

Let $s_n(x)$ be the nth partial sum of the Fourier series $\mathfrak{S}(f)$, and $\sigma_n(x) = \sigma_n(x; f)$ be the first arithmetic mean ((C, 1)-mean) of $s_n(x)$ (i.e., $\sigma_n(x) = (s_0(x) + s_1(x) + \ldots + s_n(x))/(n+1)$). Then we have

$$\sigma_n(x) - f(x) = \frac{1}{\pi}\int_0^{\pi} \varphi_x(t) K_n(t)\, dt,$$

where

$$K_n(t) = \frac{1}{2(n+1)}\left(\frac{\sin((n+1)t/2)}{\sin(t/2)}\right)^2.$$

The expression $K_n(t)$ is called the **Fejér kernel**, and the $\sigma_n(x)$ are often called **Fejér means**. If the right and left limits $f(x \pm 0)$ exist, $\mathfrak{S}(f)$ is †(C, 1)-summable at the point x to the value $(f(x+0) + f(x-0))/2$. If f is continuous at every point of a closed interval I, $\mathfrak{S}(f)$ is uniformly (C, 1)-summable in I (**Fejér theorem**, 1904). As we explain in Section H, there exist continuous functions whose Fourier series are divergent at some points. Thus the summability of $\mathfrak{S}(f)$ is more important than its convergence. Fejér's theorem remains true if we replace (C, 1)-summability by (C, α)-summability $(\alpha > 0)$. More generally, if $f \in L_1(-\pi, \pi)$, then $\mathfrak{S}(f)$ is (C, α)-summable for $\alpha > 0$ to the value $f(x)$ almost everywhere (H. Lebesgue). Since (C, α)-summability $(\alpha > 0)$ implies †summability by Abel's method, the result of Fejér's theorem is valid for †A-summability. However, the direct study of A-summability is also important. Let $f(r, x)$ be Abel's mean of $\mathfrak{S}(f)$; that is,

$$f(r,x) = \frac{1}{2}a_0 + \sum_{k=1}^{\infty}(a_k \cos kx + b_k \sin kx) r^k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x+t) P(r,t)\, dt,$$

where $P(r, t) = (1 - r^2)/2(1 - 2r\cos t + r^2)$, $0 \leqslant r < 1$. We call $P(r, t)$ the **Poisson kernel**. The function $f(r, x)$ is †harmonic inside the unit circle and tends to $f(x)$ as $r \to 1$ almost everywhere. Hence $f(r, x)$ gives the solution of the †Dirichlet problem for the case of the unit circle.

D. The Gibbs Phenomenon

Let $f(x)$ be of bounded variation and not continuous at $0: f(0)=0, f(+0)=l>0, f(-0)=-l$. Then the partial sum $s_n(x)$ converges to $f(x)$ in the neighborhood of 0, but not uniformly. Moreover,

$$\lim_{n\to\infty} s_n\left(\frac{\pi}{n}\right)=lG,$$

$$G=\frac{2}{\pi}\int_0^{\pi}\frac{\sin t}{t}dt=1.1789\ldots.$$

Hence as x tends to 0 from above and n tends to ∞, the values $y=s_n(x)$ accumulate in the interval $[0, lG]$, while $s_n(-x)=-s_n(x)$ in the neighborhood of 0. This phenomenon is called the **Gibbs phenomenon**. If f is of bounded variation, then $\mathfrak{S}(f)$ exhibits the Gibbs phenomenon at every point of simple discontinuity of f. However, the (C, 1)-means of $\mathfrak{S}(f)$ do not exhibit this phenomenon.

E. Conjugate Functions

For any integrable $f\in L_1(-\pi,\pi)$, the integral $\tilde{f}(x)=$

$$\lim_{h\to 0}-\frac{1}{\pi}\int_h^{\pi}\left((f(x+t)-f(x-t))\Big/2\tan\frac{t}{2}\right)dt$$

exists almost everywhere. The function $\tilde{f}(x)$ is called the **conjugate function** of $f(x)$. The conjugate series $\tilde{\mathfrak{S}}(f)$ is (C,α)-summable ($\alpha>0$) to the value $\tilde{f}(x)$ at almost every point, and a fortiori summable by Abel's method. Even if $f\in L_1(-\pi,\pi)$, $\tilde{f}$ does not always belong to the class $L_1(-\pi,\pi)$. For example, $\sum_{n=2}^{\infty}\cos nx/\log n$ is the Fourier series of a function $f\in L_1(-\pi,\pi)$, but its conjugate series $\sum_{n=2}^{\infty}\sin nx/\log n$ is not the Fourier series of a function in $L_1(-\pi,\pi)$. However, if both f and $\tilde{f}$ are integrable, $\tilde{\mathfrak{S}}(f)=\mathfrak{S}(\tilde{f})$. If $f\in L_p$ $(p>1)$, then $\tilde{f}\in L_p$ and $\|\tilde{f}\|_p\leqslant A_p\|f\|_p$; also, $\tilde{\mathfrak{S}}(f)=\mathfrak{S}(\tilde{f})$. If $|f|\log^+|f|$ is integrable (such a function is said to belong to the **Zygmund class**), then $\tilde{f}$ is integrable and $\tilde{\mathfrak{S}}(f)=\mathfrak{S}(\tilde{f})$. Moreover, in this case there exist constants A and B such that

$$\int_0^{2\pi}|\tilde{f}|dx\leqslant A\int_0^{2\pi}|f|\log^+|f|dx+B.$$

If f is merely integrable, so is $|\tilde{f}|^p$ for any $0<p<1$, and $\|\tilde{f}\|_p\leqslant B_p\|f\|$ $(0<p<1)$. If $f\in\operatorname{Lip}\alpha$ $(0<\alpha<1)$, then $\tilde{f}\in\operatorname{Lip}\alpha$, but the theorem fails for $\alpha=0$ and $\alpha=1$. The conjugate function is important for convergence of partial sums of Fourier series.

F. Mean Convergence

The theorems on conjugate functions enable us to obtain some results for the †mean convergence of the partial sums s_n of $\mathfrak{S}(f)$. If $f\in L_p$ $(p>1)$, then $\|f-s_n\|_p\to 0$; if $f\in L_1$, then $\|f-s_n\|_p\to 0$, $\|\tilde{f}-\tilde{s}_n\|_p\to 0$ for every $0<p<1$. Also, if $|f|\log^+|f|\in L_1$, then $\|f-s_n\|\to 0$, $\|\tilde{f}-\tilde{s}_n\|\to 0$. As a corollary of this result, we obtain the following theorem, which is a generalization of the Parseval identity: If the Fourier coefficients of functions $f\in L_p$ and $g\in L_q$ $(1/p+1/q=1)$ are a_n, b_n and a_n', b_n', respectively, we have the Parseval formula

$$\frac{1}{\pi}\int_0^{2\pi}fg\,dx=\frac{1}{2}a_0a_0'+\sum_{n=1}^{\infty}(a_na_n'+b_nb_n'),$$

where this series is convergent.

G. Analytic Functions of the Class H_p

Let $p>0$. A complex function $\varphi(z)$ holomorphic for $|z|<1$ is said to belong to the class H_p (**Hardy class**) if there exists a constant M such that

$$\lim_{r\to 1}\frac{1}{2\pi}\int_0^{2\pi}|\varphi(re^{i\theta})|^p d\theta\leqslant M.$$

When $\varphi(z)\in H_p$, the nontangential limit $\varphi(e^{i\theta})=\lim_{z\to e^{i\theta}}\varphi(z)$ exists for almost all θ. We write this as $\varphi(e^{i\theta})=f(\theta)+i\tilde{f}(\theta)$, where $f(\theta)$, $\tilde{f}(\theta)$ belong to the class L_p. Also, $\tilde{f}(\theta)$ coincides with the conjugate function of $f(\theta)$ for $p\geqslant 1$. For $1<p<\infty$, H_p is isomorphic to L_p, but for $p=1$ and $p=\infty$, H_p and L_p are different classes. Using the theory of functions of H_p, we can discuss some properties of Fourier series. If $\varphi(e^{i\theta})=f(\theta)+i\tilde{f}(\theta)$ is of bounded variation, then $\varphi(e^{i\theta})$ is absolutely continuous and its Fourier series converges absolutely. We set

$$g(\theta)=\left(\int_0^1(1-r)|\varphi'(re^{i\theta})|^2dr\right)^{1/2},$$

$g^*(\theta)=$

$$\left(\int_0^1(1-r)dr\int_0^{2\pi}|\varphi'(re^{i(\theta-t)})|^2P(r,t)dt\right)^{1/2},$$

where $P(r,t)$ is the Poisson kernel. Then $g(\theta)\leqslant 2g^*(\theta)$, and there exist constants A_p, B, C, and A_μ such that

$$\int_0^{2\pi}|g(\theta)|^p d\theta\leqslant A_p\int_0^{2\pi}|\varphi(e^{i\theta})|^p d\theta,\qquad p>0,$$

$$\int_0^{2\pi}|g^*(\theta)|^p d\theta\leqslant A_p\int_0^{2\pi}|\varphi(e^{i\theta})|^p d\theta,\qquad p>1,$$

$$\int_0^{2\pi}|g^*(\theta)|d\theta\leqslant B\int_0^{2\pi}|\varphi|\log^+|\varphi|d\theta+C,$$

$$\int_0^{2\pi} |g^*(\theta)|^\mu d\theta \leqslant A_\mu \left(\int_0^{2\pi} |\varphi(e^{i\theta})| d\theta \right)^\mu, \quad 0<\mu<1.$$

Denote by $s_n(\theta)$ and $\sigma_n(\theta)$, respectively, the partial sums and arithmetic means of the Fourier series of $\varphi(e^{i\theta})$, and set

$$\gamma(\theta) = \left(\sum_{n=1}^{\infty} \frac{|s_n(\theta)-\sigma_n(\theta)|^2}{n} \right)^{1/2}$$

Then $0 \neq A_1 \leqslant g^*(\theta)/\gamma(\theta) \leqslant A_2 \neq \infty$. From these relations, we can prove that if the indices n_k satisfy the conditions $\beta > n_{k+1}/n_k > \alpha > 1$, $s_{n_k}(\theta)$ converges almost everywhere to $\varphi(e^{i\theta})$ for $\varphi(e^{i\theta}) \in H_p$ $(1 \leqslant p)$. If we set $\Delta_k(\theta) = \sum_{\nu=n_{k-1}+1}^{n_k} c_\nu e^{i\nu\theta}$, $\delta(\theta) = (\sum_{k=0}^{\infty} |\Delta_k(\theta)|^2)^{1/2}$, where $\varphi(e^{i\theta}) \sim \sum_{\nu=0}^{\infty} c_\nu e^{i\nu\theta}$, then $\|\delta(\theta)\|_p \leqslant A_p \|\varphi\|_p$ $(p>1)$. If $\varphi(z) \in H_p$ $(0<p<1)$, then $\sum c_n e^{in\theta}$ is $(C, p^{-1}-1)$-summable to $\varphi(e^{i\theta})$ almost everywhere. These functions and relations were introduced mainly by J. E. Littlewood and R. E. A. C. Paley and were later generalized by A. Zygmund. There are more precise results by E. Stein [7], G. Sunouchi [8], S. Yano [9], and others.

H. Almost Everywhere Convergence and Divergence

P. du Bois Reymond (1876) first showed that there exits a continuous function whose Fourier series diverges at a point, but the problem of whether Fourier series of continuous functions converge almost everywhere (the so-called **du Bois Reymond problem**) remained unsolved for many years. At last in 1966, L. Carleson [10] proved that the Fourier series of a function belonging to L_2 converges almost everywhere; hence the du Bois Reymond problem was solved affirmatively. Using Carleson's method, R. A. Hunt [12] proved that

$$\int_0^{2\pi} \left(\sup_n |s_n(x)| \right)^p dx \leqslant A_p \int_0^{2\pi} |f(x)|^p dx, \quad 1<p<\infty,$$

which implies that the Fourier series of $f \in L_p$ $(1<p<\infty)$ converges almost everywhere. Hunt also proved that

$$\int_0^{2\pi} \left(\sup_n |s(x)| \right) dx \leq A \int_0^{2\pi} |f(x)| (\log^+ |f(x)|)^2 dx + A.$$

Moreover, P. Sjölin proved that if

$$\int_0^{2\pi} |f| \cdot \log^+ |f| \cdot \log^+ \log^+ |f| dx < \infty,$$

then the Fourier series of $f(x)$ converges almost everywhere.

On the other hand, A. N. Kolmogorov gave an integrable function with Fourier series diverging everywhere (more precisely, $\limsup_{n\to\infty} |s_{2^n}(x)| = \infty$ almost everywhere). Using this example, J. Marcinkiewicz showed that there is an $f \in L_1$ such that $s_n(x; f)$ oscillates boundedly almost everywhere. Moreover, there exists an integrable function with integrable conjugate and almost everywhere diverging Fourier series [3]. For any given †null set E, we can construct a continuous function whose Fourier series diverges on every $x \in E$.

I. Absolute Convergence

The convergence of the series (1) $\sum(|a_n|+|b_n|)$ implies the absolute convergence of the trigonmometric series (2) $a_0/2 + \sum(a_n \cos nx + b_n \sin nx)$. Conversely, if the series (2) converges absolutely in a set of positive measure, the series (1) converges (**Denjoy-Luzin theorem**). For the absolute convergence of Fourier series, we have the following tests: If $f \in \mathrm{Lip}\,\alpha$ $(\alpha > 1/2)$, then $\mathfrak{S}(f)$ converges absolutely, but for $\alpha = 1/2$, this is no longer true. If $f(x)$ is of bounded variation and belongs to $\mathrm{Lip}\,\alpha$ $(\alpha>0)$, $\mathfrak{S}(f)$ converges absolutely.

Suppose that the Fourier series of a function $f(x)$ is absolutely convergent and the value of $f(x)$ belongs to an interval (a, b). If $\varphi(z)$ is a function of a complex variable holomorphic at every point of the interval (a, b), the Fourier series of $\varphi\{f(x)\}$ converges absolutely (**Wiener-Lévy theorem**). As a corollary we obtain that if $\mathfrak{S}(f)$ converges absolutely and $f(x) \neq 0$, then $\mathfrak{S}(1/f)$ converges absolutely. The converse of the Wiener-Lévy theorem was proved by Y. Katznelson [6]. For a given $\varphi(x)$ defined in $[-1, 1]$, if the Fourier series of $\varphi\{f(x)\}$ converges absolutely for every $f(x)$ with absolutely convergent Fourier series $(|f(x)| \leqslant 1)$, then $\varphi(z)$ is holomorphic at every point of the interval $[-1, 1]$.

Many problems concerning this topic still remain unsolved. In particular, the determination of the structure of the functions with absolutely convergent Fourier series has not been completed.

J. Sets of Uniqueness

If $a_n \cos nx + b_n \sin nx$ converges to 0 on a set of positive measure, then $a_n, b_n \to 0$ (**Cantor-Lebesgue theorem**). A point set $E \subset (0, 2\pi)$ is called a **set of uniqueness** (or **U-set**) if every trigonometric series converging to 0 outside E vanishes identically. A set that is not a U-set is

called a **set of multiplicity** (or **M-set**). G. Cantor showed that every finite set is a U-set, and W. H. Young showed that every denumerable set is a U-set. It is clear that any set E of positive measure is an M-set, but D. E. Men'shov showed that there are †perfect M-sets of measure 0. Moreover, N. K. Bari showed that there exist perfect sets of type U. However, the structure problem of sets of uniqueness has not yet been solved completely.

A set E is said to be of type H (or an H-set) if there exists a sequence of positive integers $n_1 < n_2 < \ldots$ and an interval I such that for each $x \in E$, no point of $\{n_k x\}_{k=1}^{\infty}$ is in I (mod 2π). H-sets are sets of uniqueness, a fact given by A. Rajchman. I. I. Pyatetskiĭ-Shapiro generalized H-sets to $H^{(m)}$-sets [3].

Lacunary trigonometric series are series in which very few terms differ from zero. Such series can be written in the form

$$\sum_{k=1}^{\infty} (a_k \cos n_k x + b_k \sin n_k x) = \sum_{k=1}^{\infty} A_{n_k}(x).$$

S. Sidon established some of the characteristic properties of such series; he generalized them further and obtained the notion of Sidon sets (→ 192 Harmonic Analysis). We often define a lacunary series more specifically as a series for which the n_k satisfy Hadamard's gaps; that is, $n_{k+1}/n_k > q > 1$. Then if $\sum_{k=1}^{\infty}(a_k^2 + b_k^2)$ is finite, the series $\sum_{k=1}^{\infty} A_{n_k}(x)$ converges almost everywhere. Conversely, if $\sum_{k=1}^{\infty} A_{n_k}(x)$ is convergent in a set of positive measure, then $\sum_{k=1}^{\infty}(a_k^2 + b_k^2)$ converges. This theorem is related to the Rademacher series and random Fourier series [4].

K. Multiple Fourier Series

Routine extensions to multiple Fourier series from the case of a single variable are easy, but significant results are difficult to obtain. Recently, however, there have been several important contributions in this field.

References

[1] A. Zygmund, Trigonometrical series, Warsaw, 1935.
[2] A. Zygmund, Trigonometric series I, II, Cambridge Univ. Press, second edition, 1959.
[3] G. H. Hardy and W. W. Rogosinski, Fourier series, Cambridge Univ. Press, 1950.
[4] N. K. Bari, A treatise on trigonometric series, Pergamon, 1964. (Original in Russian, 1958.)
[5] J.-P. Kahane and R. Salem, Ensembles parfaits et séries trigonométriques, Actualités Sci. Ind., Hermann, 1963.
[6] Y. Katznelson, Sur les fonctions opérant sur l'algèbre des séries de Fourier absolument convergentes, C. R. Acad. Sci. Paris, 247 (1958), 404–406.
[7] E. M. Stein, A maximal function with applications to Fourier series, Ann. Math., (2) 68 (1958), 584–603.
[8] G. Sunouchi, Theorems on power series of the class H^p, Tôhoku Math. J., (2) 8 (1956), 125–146.
[9] S. Yano, On a lemma of Marcinkiewicz and its applications to Fourier series, Tôhoku Math. J., (2) 11 (1959), 191–215.
[10] L. Carleson, On convergence and growth of partial sums of Fourier series, Acta Math., 116 (1966), 135–157.
[11] J.-P. Kahane and Y. Katznelson, Sur les ensembles de divergence des séries trigonométriques , Studia Math., 26 (1966), 305–306.
[12] R. A. Hunt, On the convergence of Fourier series, Orthogonal expansions and their continuous analogues, Proc. Conf. Southern Illinois Univ., 1967, Southern Illinois Univ. Press, 1968, 235–255.
[13] P. Sjölin, An inequality of Paley and convergence a.e. of Walsh-Fourier series, Ark. Mat., 7 (1968), 551–570.
[14] R. Salem, Algebraic numbers and Fourier analysis, Heath, 1963.
[15] J.-P. Kahane, Some random series of functions, Heath, 1968.

160 (X.23) Fourier Transform

A. Fourier Integrals

In this article we assume that $f(x)$ is a complex-valued function defined on $\mathbf{R} = (-\infty, \infty)$ and †(Lebesgue) integrable on any finite interval. If the integral

$$\int_{-\infty}^{\infty} f(x)e^{-ixt}\,dx = \lim_{\substack{A\to-\infty \\ B\to\infty}} \int_A^B f(x)e^{-ixt}\,dx$$

exists, it is called the **trigonometric integral** or **Fourier integral**. We have a general result: If $f(x) \in L_1(-\infty, \infty)$, $K(x)$ is bounded on $(-\infty, \infty)$, and $\int_0^T K(x)\,dx = o(T)$ $(T \to \pm\infty)$, then $\int_{-\infty}^{\infty} f(x)K(xt)\,dx$ exists and

$$\lim_{t\to\pm\infty} \int_{-\infty}^{\infty} f(x)K(xt)\,dx = 0.$$

In particular, it follows that if $f(x) \in L_1(-\infty, \infty)$, then $\int_{-\infty}^{\infty} f(x)e^{itx}\,dx$ exists and

$$\lim_{t\to\pm\infty} \int_{-\infty}^{\infty} f(x)e^{-itx}\,dx = 0$$

(Riemann-Lebesgue theorem).

B. Fourier's Integral Theorems

Suppose that $f(x)$ is of †bounded variation in an interval including x, or more generally satisfies the assumption for any one of the convergence tests for Fourier series (→ 159 Fourier Series B). Then **Fourier's single integral theorem**

$$\frac{1}{2}(f(x+0)+f(x-0))$$

$$=\lim_{A\to\infty}\frac{1}{\pi}\int_{-\infty}^{\infty}f(t)\frac{\sin A(t-x)}{t-x}dt \qquad (1)$$

holds if one of the following three conditions is satisfied: (1) $f(x)/(1+|x|)$ belongs to $L_1(-\infty,\infty)$; (2) $f(x)/x$ tends to zero monotonically as $x\to\pm\infty$; (3) $f(x)/x=g(x)\sin(px+q)$, where $g(x)$ tends to zero monotonically as $x\to\pm\infty$ (S. Izumi, 1934). The right-hand side of (1) is called **Dirichlet's integral**.

Let $f(x)$ be of bounded variation in an interval including x (or satisfy some other convergence test for Fourier series). Then **Fourier's double integral theorem**

$$\frac{1}{2}(f(x+0)+f(x-0))$$

$$=\frac{1}{\pi}\lim_{T\to\infty}\int_0^T dt\int_{-\infty}^{\infty}f(u)\cos t(u-x)du \qquad (2)$$

holds if one of the following three conditions is satisfied: (4) $f(x)\in L_1(-\infty,\infty)$; (5) $f(x)/(1+|x|)\in L_1(-\infty,\infty)$, and $f(x)$ tends to zero monotonically as $x\to\pm\infty$; (6) $f(x)/(1+|x|)\in L_1(-\infty,\infty)$ and $f(x)=g(x)\sin(px+q)$, where $g(x)$ tends to zero monotonically as $x\to\pm\infty$. If $f(x)\in L_1(-\infty,\infty)$, the formula

$$f(x)=\lim_{A\to\infty}\frac{1}{\pi A}\int_{-\infty}^{\infty}f(t)\frac{\sin^2 A(t-x)}{A(t-x)^2}dt$$

holds almost everywhere, and in particular at any x where $f(x)$ is continuous. More generally, the formula

$$f(x)=\lim_{A\to\infty}A\int_{-\infty}^{\infty}f(t)K(A(t-x))dt$$

$$=\lim_{A\to\infty}\int_{-\infty}^{\infty}f\left(x-\frac{t}{A}\right)K(t)dt \qquad (3)$$

holds at any point x where $f(x+0)$ and $f(x-0)$ exist, if $K(t)\in L_1(-\infty,\infty)$, $\int_{-\infty}^{\infty}K(t)dt=1$, $|K(t)|<M$, $K(t)=o(t^{-1})$ as $|t|\to\infty$, and $f(x)\in L_1(-\infty,\infty)$, or if $K(t)\in L_1(-\infty,\infty)$, $\int_{-\infty}^{\infty}K(t)dt=1$, and $f(x)$ is bounded. Similarly, the formula

$$\lim_{A\to 0}\int_0^{\infty}f\left(\frac{x}{A}\right)K(x)dx=\mathfrak{M}\{f\}\int_0^{\infty}K(x)dx$$

holds if $\mathfrak{M}\{f\}=\lim_{T\to\infty}T^{-1}\int_0^T f(t)dt$ exists, $K(x)$ is differentiable, $|x^2K(x)|\leqslant C$ $(1\leqslant x)$, and $x^{-1}\int_0^x|f(t)|dt\leqslant D$, where C, D are constants (**Wiener's formula**).

C. Fourier Transforms (→ Appendix A, Table 11.II)

Let $f(x)\in L_1(0,\infty)$. Then

$$F(t)=\sqrt{\frac{2}{\pi}}\int_0^{\infty}f(u)\cos ut\,du$$

is called the (**Fourier**) **cosine transform** of $f(x)$. Under the same condition as for the validity of (2), the **inversion formula** of the cosine transform $f(x)=\sqrt{2/\pi}\int_0^{\infty}F(t)\cos xt\,dt$ holds where we suppose that $f(x)=\frac{1}{2}(f(x+0)+f(x-0))$. If we define $f(-x)=f(x)$, then this is equivalent to the formula (2). Analogously,

$$G(t)=\sqrt{\frac{2}{\pi}}\int_0^{\infty}f(u)\sin ut\,du$$

is called the (**Fourier**) **sine transform** of $f(x)$. Under the same condition as for the validity of (2), we get the inversion formula $f(x)=\sqrt{2/\pi}\int_0^{\infty}G(t)\sin xt\,dt$. More generally, for any $f(x)\in L_1(-\infty,\infty)$,

$$F(t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}f(x)e^{-ixt}dx$$

is called the **Fourier transform** of $f(x)$. Under the same condition as for the validity of (2), the **Fourier inversion formula**

$$f(x)=\frac{1}{\sqrt{2\pi}}\lim_{T\to\infty}\int_{-T}^{T}F(t)e^{ixt}dt$$

holds. The cosine transform and the sine transform coincide with the Fourier transform when $f(x)=f(-x)$ and $-f(x)=f(-x)$, respectively.

If for any $f(x)$, $F(t)\in L_q(-\infty,\infty)$ $(1\leqslant q<\infty)$ exists, for which

$$\int_{-\infty}^{\infty}\left|\frac{1}{\sqrt{2\pi}}\int_{-T}^{T}f(x)e^{-ixt}dt-F(t)\right|^q dt\to 0$$

$$(T\to\infty)$$

is valid (i.e., $(1/\sqrt{2\pi})\int_{-T}^{T}f(t)e^{-ixt}dt$ converges to $F(x)$ as $T\to\infty$ in the mean of order q), then we say that $f(x)$ has the **Fourier transform** $F(t)$ **in** $L_q(-\infty,\infty)$. If $f(x)\in L_p(-\infty,\infty)$ $(1<p\leqslant 2)$, then $f(x)$ has the Fourier transform $F(t)$ in L_q $(1/p+1/q=1)$, and $F(t)$ has the Fourier transform $f(-x)$ in L_p (E. C. Titchmarsh). Moreover,

$$\int_{-\infty}^{\infty}|F(t)|^q dx$$

$$\leqslant\frac{1}{(2\pi)^{(q/2)-1}}\left(\int_{-\infty}^{\infty}|f(x)|^p dx\right)^{1/(p-1)}.$$

If $f(x)$, $G(x) \in L_p(-\infty, \infty)$ $(1 < p \leqslant 2)$ and their Fourier transforms in L_q are $F(t)$, $g(t)$, respectively, then the **Parseval identity**

$$\int_{-\infty}^{\infty} F(t)G(t)\,dt = \int_{-\infty}^{\infty} f(x)g(-x)\,dx$$

holds. The Fourier inversion formula holds, in the sense that

$$\int_0^x F(u)\,du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) \frac{e^{-ixt}-1}{-it}\,dt,$$

$$\int_0^x f(t)\,dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F(u) \frac{e^{ixu}-1}{iu}\,du.$$

The theory of Fourier transforms is valid for cosine and sine transforms. Specifically for $p = 2$, if $f \in L_2(0, \infty)$, then $\sqrt{2/\pi} \int_0^T f(x) \cdot \cos xt\,dt$ †converges in the mean in $L_2(0, \infty)$ to the cosine transform $F(t)$ as $T \to \infty$, and conversely, the cosine transform of $F(t)$ in L_2 is $f(x)$. The transforms $f(x)$, $F(x)$ are connected by the formulas

$$\int_0^x F(u)\,du = \sqrt{\frac{2}{\pi}} \int_0^{\infty} f(t) \frac{\sin xt}{t}\,dt,$$

$$\int_0^x f(t)\,dt = \sqrt{\frac{2}{\pi}} \int_0^{\infty} F(u) \frac{\sin xu}{u}\,du,$$

$$\int_0^{\infty} |f(x)|^2\,dx = \int_0^{\infty} |F(t)|^2\,dt.$$

The theory of Fourier transforms was generalized as follows by G. N. Watson (*Proc. London Math. Soc.*, 35 (1933)). We suppose that $\chi(x)/x \in L_2(0, \infty)$ and

$$\int_0^{\infty} \frac{\chi(xu)\chi(yu)}{u^2}\,du = \min(x, y). \tag{4}$$

If $f(x) \in L_2(0, \infty)$, then there exists an $F(t) \in L_2(0, \infty)$ such that

$$\int_0^x F(t)\,dt = \int_0^{\infty} \frac{\chi(xu)f(u)}{u}\,du,$$

and the inversion formula

$$\int_0^x f(u)\,du = \int_0^{\infty} \frac{\chi(xt)F(t)}{t}\,dt$$

and the Parseval identity

$$\int_0^{\infty} |f(x)|^2\,dx = \int_0^{\infty} |F(t)|^2\,dt$$

hold. $F(t)$ is called the **Watson transform** of $f(x)$. For any $f \in L_2$, equality (4) is necessary for the existence of the Watson transform $F(t)$ for which the inversion formula holds. S. Bochner (1934) generalized this theory further to †unitary transformations in L_2 (→ 192 Harmonic Analysis).

D. Conjugate Functions

Corresponding to Fourier's double integral theorem (2), the integral

$$\lim_{\lambda \to \infty} \frac{1}{\pi} \int_0^{\lambda} dt \int_{-\infty}^{\infty} \sin t(u-x) f(u)\,du$$

is called the **conjugate Fourier integral** of the integral in the right-hand side of (2). Formally, this is written $\lim_{\lambda \to \infty} t^{-1} \int_0^{\infty} (1 - \cos \lambda t) t^{-1} (f(x+t) - f(x-t))\,dt$. If $f(x)$ is a sufficiently regular function, the part involving $\cos \lambda t$ tends to 0 as $\lambda \to \infty$. Now let

$$g(x) = \frac{1}{\pi} \lim_{\substack{A \to \infty \\ \varepsilon \to 0}} \int_{\varepsilon}^{A} \frac{f(x+t) - f(x-t)}{t}\,dt.$$

For any $f \in L_1(0, \infty)$, the integral exists almost everywhere, and $g(x)$ is called the **conjugate function** or **Hilbert transform** of $f(x)$. If $f \in L_p$ $(p > 1)$, then $g(x) \in L_p$ also and we have

$$f(x) = -\frac{1}{\pi} \lim_{\substack{A \to \infty \\ \varepsilon \to 0}} \int_{\varepsilon}^{A} \frac{g(x+t) - g(x-t)}{t}\,dt$$

and $\int_{-\infty}^{\infty} |g(x)|^p\,dx \leqslant M_p \int_{-\infty}^{\infty} |f(x)|^p\,dx$, where M_p is a constant depending only on p. In particular,

$$\int_{-\infty}^{\infty} |f(x)|^2\,dx = \int_{-\infty}^{\infty} |g(x)|^2\,dx \quad \text{for } p = 2.$$

E. Boundary Functions of Analytic Functions

Suppose that a complex-valued function $f(z)$ $(z = x + iy)$ is †holomorphic for $y > 0$, $f(x + iy)$ converges as $y \to 0$ for almost all x to $f(x)$ (which is called the **boundary function**), and $f(x) \in L_p(-\infty, \infty)$ $(p \geqslant 1)$. Moreover, suppose that $f(z)$ is represented by †Cauchy's integral formula or †Poisson's integral of $f(x)$ on the real line. If either $f(x) \in L_p$ $(p \geqslant 1)$ and has $F(t)$ as its Fourier transform or $f(x)$ is an L_p-Fourier transform of $F(t) \in L_q$ $(q \geqslant 1)$, then a necessary and sufficient condition for the function $f(x)$ to be the boundary function of an analytic function is that $F(t)$ be 0 almost everywhere for $t > 0$ (N. Wiener, R. E. A. C. Paley, E. Hille, J. D. Tamarkin).

F. Generalized Fourier Integrals

Let $|f(x)|/(1 + |x|^k) \in L_1(-\infty, \infty)$ for a positive integer k and

$$L_k = L_k(t, x) = \begin{cases} \sum_{\nu=0}^{k-1} \frac{(-itx)^{\nu}}{\nu!}, & |x| \leqslant 1, \\ 0, & |x| > 1, \end{cases}$$

$$E_k(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x) \frac{e^{-itx} - L_k}{(-ix)^k}\,dx.$$

The function $E_k(t)$ or one that differs from $E_k(t)$ by a polynomial of degree at most k is called the k**th transform** of $f(x)$. We write formally $f(x)=\int_{-\infty}^{\infty}e^{ixt}d^kE_k(t)$. Actually, if we give an appropriate meaning to the integral, this formula itself is valid (H. Hahn, *Wiener Berichte*, 134 (1925); S. Izumi, *Tôhoku Science Rep.*, 23 (1935)). (For the theory and applications of kth transforms → [2, ch. 6].)

G. Applications of Fourier Transforms

Suppose that $f(x)\in L_1(-\infty,\infty)$, $F(t)$ is its Fourier transform, and $f(x)=o(e^{-\theta(x)})$, where $\theta(x)$ is positive and increasing. If $\int_1^{\infty}\theta(x)x^{-2}dx=\infty$ and $F(t)$ vanishes identically in an interval, then $F(t)\equiv 0$ over $(-\infty,\infty)$. If $\int_1^{\infty}\theta(x)x^{-2}dx<\infty$, then there exists a function $f(x)$ such that $F(t)$ vanishes identically in an interval, but $F(t)$ does not vanish identically in $(-\infty,\infty)$ (Wiener, Paley, N. Levinson). These results are applicable to the theory of †quasi-analytic functions.

Let $f(x)\in L_1(-\infty,\infty)$. Then a necessary and sufficient condition for any function in $L_1(-\infty,\infty)$ to be approximated as closely as we wish by linear combinations of the translations $\sum_{k=1}^{N}a_kf(x+h_k)$ of $f(x)$ with respect to the L_1-norm is that the Fourier transform of $f(x)$ does not vanish at any real number. When $f(x)\in L_2(-\infty,\infty)$, a necessary and sufficient condition that an arbitrary function in $L^2(-\infty,\infty)$ can be approximated as closely as we wish by $\sum_{k=1}^{N}a_kf(x+h_k)$ with respect to the L_2-norm is that the zeros of the Fourier transform of $f(x)$ have measure zero (Wiener). This result was used by Wiener to prove the **generalized Tauberian theorem**: Suppose that $g_1(x)\in L_1(-\infty,\infty)$ and its Fourier transform never vanishes. Moreover, let $g_2(x)\in L_1(-\infty,\infty)$ and $p(x)$ be bounded over $(-\infty,\infty)$. Then $\lim_{x\to\infty}\int_{-\infty}^{\infty}g_1(x-t)p(t)dt=A\int_{-\infty}^{\infty}g_1(t)dt$ implies that $\lim_{x\to\infty}\int_{-\infty}^{\infty}g_2(x-t)p(t)dt=A\int_{-\infty}^{\infty}g_2(t)dt$. Another type of Wiener theorem is concerned with †Stieltjes integrals. Suppose that

$$\sum_{n=-\infty}^{\infty}\sup_{n\leqslant x<n+1}|g_1(x)|<\infty$$

(hence $g_1(x)\in L_1$) and that the Fourier transform of $g_1(x)$ never vanishes. Moreover, let

$$\sum_{n=-\infty}^{\infty}\sup_{n\leqslant x<n+1}|g_2(x)|<\infty$$

and let $\int_x^{x+1}|d\alpha(t)|$ be bounded. Then

$$\lim_{x\to\infty}\int_{-\infty}^{\infty}g_1(x-t)d\alpha(t)=A\int_{-\infty}^{\infty}g_1(t)dt$$

implies

$$\lim_{x\to\infty}\int_{-\infty}^{\infty}g_2(x-t)d\alpha(t)=A\int_{-\infty}^{\infty}g_2(t)dt.$$

From these general theorems, we can prove various †Tauberian theorems about the summation of series. Also, these results were applied to the proof of the †prime number theorem by S. Ikehara and E. Landau (Wiener [1]). In the general Tauberian theorem, the boundedness of $p(t)$ can be replaced by one-sided boundedness (H. R. Pitt, 1938). In fact, the first form of the theorem still holds if we replace the condition by the following one: g_1 and g_2 are continuous, $g_1(x)\geqslant 0$, the Fourier transform of $g_1(x)$ does not vanish, $g_2(x)$ satisfies the conditions of the second theorem, and $p(x)\geqslant C$. (Concerning Fourier transforms on the topological groups → 192 Harmonic Analysis; and concerning Fourier transforms of the distributions → 125 Distributions and Hyperfunctions).

H. Fourier Transforms of Distributions

The Fourier transform is defined in higher-dimension $\mathbf{R}^n$ by

$$\mathscr{F}(f)=(\sqrt{2\pi})^{-n}\int e^{-ix\xi}f(x)dx,\qquad f\in L_1,$$

$$x\xi=x_1\xi_1+x_2\xi_2+\ldots+x_n\xi_n.$$

One denotes it by $\hat{f}(\xi)$. The inverse transform is defined by

$$\overline{\mathscr{F}}(g)=(\sqrt{2\pi})^{-n}\int e^{ix\xi}g(\xi)d\xi.$$

Differentiation under the integral sign gives

$$D^{\alpha}\hat{f}(\xi)=(\sqrt{2\pi})^{-n}\int e^{-ix\xi}(-ix)^{\alpha}f(x)dx,$$

$$D^{\alpha}=(\partial/\partial x_1)^{\alpha_1}\ldots(\partial/\partial x_n)^{\alpha_n},$$

under the assumption $(1+|x|)^{|\alpha|}f(x)\in L_1$ ($|\alpha|=\alpha_1+\ldots+\alpha_n$). Roughly speaking, the decreasing order of $f(x)$ when $|x|\to\infty$ is reflected in the differentiability of $\hat{f}(\xi)$. In the same way,

$$(i\xi)^{\alpha}\hat{f}(\xi)=(\sqrt{2\pi})^{-n}\int e^{-ix\xi}D^{\alpha}f(x)dx,$$

which shows that the differentiability of $f(x)$ is reflected in the decreasing order of $\hat{f}(\xi)$. The same statements can evidently be made for the relation between $g(\xi)$ and its inverse Fourier transform.

Let $\mathscr{S}$ be the space of rapidly decreasing functions $\varphi(x)$, i.e., such that for all positive integers k and $\alpha\geqslant 0$, $(1+|x|)^kD^{\alpha}\varphi(x)$ remains bounded (→ 125 Distributions and Hyperfunctions). From the facts above, it follows that the

linear mapping $f(x)\mapsto\mathcal{F}(f)$ is a topological isomorphism from $\mathcal{S}_x$ onto $\mathcal{S}_\xi$. Let $\mathcal{S}'$ be the dual space of $\mathcal{S}$. Usual function spaces in classical analysis are contained in $\mathcal{S}'$, for instance, the L_p-space and that of functions increasing at most of polynomial order at infinity.

Let $T\in\mathcal{S}_x'$. Then the mapping $\varphi\in\mathcal{S}_\xi\mapsto T(\mathcal{F}\varphi)$ is continuous, and $\mathcal{F}T\in\mathcal{S}_\xi'$ is defined by $\mathcal{F}T(\varphi)=T(\mathcal{F}\varphi)$. $\mathcal{F}$ is also a topological isomorphism from $\mathcal{S}_x'$ onto $\mathcal{S}_\xi'$. Take as an example the distribution p.v.$1/x$. This is no longer a function; however, its Fourier transform can be calculated as follows:

$$\lim_{\substack{\varepsilon\to 0\\ A\to\infty}}(\sqrt{2\pi})^{-1}\int_{\varepsilon\le|x|\le A}e^{-ix\xi}\frac{dx}{x}=\frac{2i}{\sqrt{2\pi}}$$

$$\times\lim_{A\to\infty}\int_0^A\frac{\sin(x\xi)}{x}dx=\begin{cases}\sqrt{\dfrac{\pi}{2}}i, & \xi>0,\\ -\sqrt{\dfrac{\pi}{2}}i, & \xi<0,\end{cases}$$

where the limit is taken with respect to the topology of $\mathcal{S}'$.

Using the definition above, classical results can be extended to $\mathcal{S}'$ almost automatically. For example, $\mathcal{F}(D^\alpha T)=(i\xi)^\alpha\mathcal{F}(T)$, $\mathcal{F}((-ix)^\alpha T)=D^\alpha\mathcal{F}(T)$. Moreover, if $T\in\mathcal{E}'$, then $\mathcal{F}T=\sqrt{2\pi}^{\,-n}T(e^{-ix\xi})$. In particular, $\mathcal{F}\delta=(2\pi)^{-n/2}$.

The **Plancherel theorem** says that if for $f(x)\in L_2$ one defines its Fourier transform by

$$\hat{f}(\xi)=\underset{A\to\infty}{\text{l.i.m.}}\sqrt{2\pi}^{\,-n}\int_{|x|\le A}e^{-ix\xi}f(x)\,dx,$$

then the correspondence $f(x)\mapsto\hat{f}(\xi)$ is a unitary mapping from L_2 onto L_2, i.e.,

$$\int|f(x)|^2\,dx=\int|\hat{f}(\xi)|^2\,d\xi.$$

This result can be extended. For any nonnegative integer m, the element of H^m (→ 168 Function Spaces) is characterized by its Fourier transform: $f(x)\in H^m$ if and only if $(1+|\xi|)^m\hat{f}(\xi)\in L_2$. Furthermore, for arbitrary real s, the space H^s can be defined as the set of all elements of $\mathcal{S}'$ whose Fourier transform $\hat{f}(\xi)$ satisfies $(1+|\xi|)^s\hat{f}(\xi)\in L_2$. H^s and H^{-s} are dual to each other.

For $f\in L_1$, $g\in L_2$, or $f, g\in L_2$, the convolution makes sense as a function, and it holds that $\mathcal{F}(f*g)=\mathcal{F}(f)\mathcal{F}(g)$. This relation can be extended to distributions, to state

$$\mathcal{F}(S*T)=(\mathcal{F}S)(\mathcal{F}T).$$

This holds for $(S,T)\in\mathcal{E}'\times\mathcal{S}'$, $\mathcal{D}_{L_2}'\times\mathcal{D}_{L_2}'$ ($\mathcal{D}_{L_2}'$ is the dual of $\mathcal{D}_{L_2}$), etc.

Fourier transforms are also often defined by

$$\hat{f}(\xi)=\int e^{-ix\xi}f(x)\,dx\quad\text{or}\quad=\int e^{-2\pi ix\xi}f(x)\,dx.$$

In the former case, the inversion formula takes the form

$$f(x)=(2\pi)^{-n}\int e^{ix\xi}\hat{f}(\xi)\,d\xi,$$

and the Parseval identity becomes

$$\int|f(x)|^2\,dx=(2\pi)^{-n}\int|\hat{f}(\xi)|^2\,d\xi.$$

References

[1] N. Wiener, The Fourier integral and certain of its applications, Cambridge Univ. Press, 1933.
[2] S. Bochner, Vorlesungen über Fouriersche Integrale, Akademische-Verlag, 1932; English translation, Lectures on Fourier integrals, Ann. Math. Studies, Princeton Univ. Press, 1959.
[3] E. C. Titchmarsh, Introduction to the theory of Fourier integrals, Clarendon Press, 1937.
[4] R. E. A. C. Paley and N. Wiener, Fourier transforms in the complex domain, Amer. Math. Soc. Colloq. Publ., 1934.
[5] S. Bocher and K. Chandrasekharan, Fourier transforms, Ann. Math. Studies, Princeton Univ. Press, 1949.
[6] M. J. Lighthill, Introduction to Fourier analysis and generalized functions, Cambridge Univ. Press, 1958.
[7] I. M. Gel'fand and G. E. Shilov (Šilov), Generalized functions I, Academic Press, 1964. (Original in Russian, 1958.)
[8] S. Bochner, Harmonic analysis and the theory of probability, Univ. of California Press, 1955.
[9] R. R. Goldberg, Fourier transforms, Cambridge Tracts, Cambridge Univ. Press, 1961.
[10] T. Carleman, L'intégrale de Fourier et questions qui s'y rattachent, Almqvist and Wiksell, 1944.
[11] L. Schwartz, Théorie des distributions, Hermann, 1966.
For formulas for Fourier transforms → references to 220 Integral Transforms.

161 (IV.3) Free Groups

A. General Remarks

A group F is called a **free group** if it is the †free product (→ 190 Groups M) of †infinite cyclic groups $G_1,\ldots,G_n$ generated by $a_1,\ldots,a_n$, respectively. Then n is called the **rank** of F. A

free product of †semigroups is defined similarly to that of groups, and the free product of infinite cyclic semigroups $G_i = \{1, a_i, a_i^2, \dots\}$ $(i = 1, \dots, n)$ is called a **free semigroup** generated by n elements a_i $(i = 1, \dots, n)$.

If a group G is generated by subgroups H_i $(i = 1, \dots, n)$ isomorphic to G_i, then G is a homomorphic image of the free product of the groups G_i. A subgroup $\neq \{e\}$ of the free product F of groups G_i is itself the free product of a free group and several subgroups, each of which is conjugate in F to a subgroup of some G_j (A. G. Kurosh, 1934). Notably, a subgroup $\neq \{e\}$ of a free group is itself a free group (O. Schreier, *Abh. Math. Sem. Univ. Hamburg*, 5 (1927)). A subgroup of index j of a free group of rank n is a free group of rank $1 + j(n-1)$ (Schreier).

Let F be the free group generated by n elements $a_1, \dots, a_n$, and let G be a group generated by n elements $b_1, \dots, b_n$. Then there is a homomorphism of F onto G. Let N be its kernel. If the class of a †word $w(a_1, \dots, a_n)$ belongs to N, then we have $w(b_1, \dots, b_n) = 1$. We call $w(b_1, \dots, b_n) = 1$ a **relation** among the generators $b_1, \dots, b_n$. If N is the minimal normal subgroup of F containing the classes of words $w_1(a_1, \dots, a_n), \dots, w_m(a_1, \dots, a_n)$, then the relations $w_1(b_1, \dots, b_n) = 1, \dots, w_m(b_1, \dots, b_n) = 1$ are called **defining relations** (or **fundamental relations**). If generators $a_1, \dots, a_n$ and words $w_1(a_1, \dots, a_n), \dots, w_m(a_1, \dots, a_n)$ are given, then there is a group generated by $a_1, \dots, a_n$ with defining relations $w_1(a_1, \dots, a_n) = 1, \dots, w_m(a_1, \dots, a_n) = 1$. In fact, let F be the free group generated by $a_1, \dots, a_n$ and N the minimal normal subgroup containing the classes of words $w_1(a_1, \dots, a_n), \dots, w_m(a_1, \dots, a_n)$. Then the factor group F/N is such a group. A free group is a group with an empty set of defining relations. In the preceding discussion, n and m are not necessarily finite. If both n and m are finite, then G is called **finitely presented**.

B. The Word Problem

If a finitely presented group G is given, then a general procedure has to be determined by which it can be decided, in a finite number of computational steps, whether a given word equals the identity element as an element of G. This is called the **word problem** (→ 190 Groups M). A solution to the word problem does not always exist (P. S. Novikov [8], 1955); in fact, there is a group with two generators and 32 defining relations for which the word problem cannot be solved (W. Boone [7]). However, it was shown by V. A. Tartakovskiĭ that the problem can be solved for a large class of groups. W. Magnus (1931) showed that it is solvable for any group with a single defining relation. The word problem is an example of decision problems (→ 97 Decision Problem). The word problem for groups is closely related to that for semigroups (A. M. Turing, 1937; E. L. Post, 1947; A. A. Markov, 1947). Similar problems for other algebraic systems can also be considered. The problem of determining a general procedure by which it can be decided, in a finite number of steps, whether two given words interpreted as elements of G can be transformed into each other by an (inner) automorphism of G is called the **transformation problem**.

Let F be a free group of rank n and $F = F_1 \supset \dots \supset F_r \supset F_{r+1} \supset \dots$ be the †lower central series of F. Then F_r/F_{r+1} is a †free Abelian group of rank $\mu_n(r) = (1/r)\sum_{d|r}\mu(r/d)n^d$, where μ is the †Möbius function (E. Witt). The intersection of all subgroups of F of finite index is the identity element.

C. The Burnside Problem

The original problem of Burnside is: If every element of a group G is of finite order (but not necessarily of bounded order) and G is finitely generated, is G a finite group? E. S. Golod [6] (1964) showed that this problem for p-groups has a negative solution. The following is the more usual form of the **Burnside problem**: If a group G is finitely generated and the orders of elements of G divide a given integer r, is G finite? Let F be a free group of rank n, N be the normal subgroup of F generated by all the rth powers x^r of elements of F, and $B(r, n) = F/N$. Then the problem is the same as the question of whether $B(r, n)$ is finite. For $r = 2, 3, 4, 6$ the group is certainly finite (I. N. Sanov, M. Hall). The **restricted Burnside problem** is the question whether the orders of finite factor groups of $B(r, n)$ are bounded. It was solved affirmatively for r a prime (A. I. Kostrikin [5], 1959).

A group generated by two generators x, y and satisfying the relations $x^u = y^v = (xy)^w = 1$ (where u, v, w are integers) is infinite if $1/u + 1/v + 1/w - 1 \leqslant 0$, and is of order g if $0 < 1/u + 1/v + 1/w - 1 = 2/g$.

There is also a finitely presented group which is isomorphic to its proper factor group (B. H. Neumann).

References

[1] A. G. Kurosh, The theory of groups I, II, Chelsea, 1960. (Original in Russian, 1953.)
[2] M. Hall, The theory of groups, Macmillan, 1959.

[3] H. S. M. Coxeter and W. O. J. Moser, Generators and relations for discrete groups, Erg. Math., Springer, 1957.
[4] W. Magnus, A. Karrass, and D. Solitar, Combinatorial group theory, Interscience, 1966.
[5] A. I. Kostrikin, The Burnside problem, Amer. Math. Soc. Transl., (2) 36 (1964), 63–100. (Original in Russian, 1959.)
[6] E. S. Golod, On nil-algebras and finitely approximable p-groups, Amer. Math. Soc. Transl., (2) 48 (1965), 103–106. (Original in Russian, 1964.)
[7] W. W. Boone, The word problem, Ann. of Math., (2) 70 (1959), 207–265.
[8] P. S. Novikov, On the algorithmic unsolvability of the word problem in group theory, Amer. Math. Soc. Transl., (2) 9 (1958), 1–122. (Original in Russian, 1955.)
Also → references to 190 Groups.

162 (XII.1) Functional Analysis

The origin of **functional analysis** can be traced to the 1887 work of V. Volterra. He stressed the important notion of operations or **operators**, that is, generalized functions of which the domains as well as the ranges are sets of functions. A typical example is the operator assigning to a function f its derivative f'. An operator is called a **functional** if its values are numbers, as in the case of the operator assigning to a function f the value $f'(a)$ or the value $\int_a^b f(t)\,dt$. In 1896, Volterra considered the operator mapping a continuous function f to a continuous solution φ of the integral equation $f(x)=\varphi(x)-\int_a^x K(x,y)\varphi(y)\,dy$, where $K(x,y)$ is a continuous function. Defining $I\varphi=\varphi$ and $(K\varphi)(x)=\int_a^x K(x,y)\,\varphi(y)\,dy$, he showed that φ is given by $\varphi=(I-K)^{-1}f=f+Kf+K^2f+\ldots$, where $K^nf=K(K^{n-1}f)$. Following this lead, I. Fredholm studied in 1900 the integral equation $f(x)=\varphi(x)-\lambda\int_a^b K(x,y)\varphi(y)\,dy$ containing a parameter λ. He proved the so-called †alternative theorem: For a given λ_0, the operator equation $(I-\lambda_0 K)\varphi=f$, $(K\varphi)(x)=\int_a^b K(x,y)\varphi(y)\,dy$, either admits a uniquely determined continuous solution φ for every continuous function f or else $(I-\lambda_0 K)\varphi_0=0$ admits a nontrivial continuous solution $\varphi_0\neq 0$. D. Hilbert discussed (1904–1910) a †continuous linear operator K defined on the †Hilbert space L_2 with values in L_2, and he called a complex number λ_0 a †spectrum of K if $(I-\lambda_0 K)$ does not have a continuous linear inverse. He proved that if K is †Hermitian, then K admits a †spectral resolution with real spectra only. One of his outstanding contributions was the discovery of the †continuous spectrum. In 1918, F. Riesz proved that Fredholm's alternative theorem holds for a †completely continuous linear operator in the space of continuous functions, and later the result was extended to Banach spaces.

In 1932, three important books by S. Banach [1], J. von Neumann [2], and M. H. Stone [3] were published. These books treated †closed linear operators that are not necessarily continuous. The notion of †Banach space was introduced: a †normed linear space complete with respect to the distance $\mathrm{dis}(x,y)=\|x-y\|$. By making use of the †Baire-Hausdorff theorem and the †Hahn-Banach theorem, Banach proved, for closed linear operators in Banach spaces, the fundamental principle consisting of the †open mapping theorem, the †closed graph theorem, the †uniform boundedness theorem, the †resonance theorem, and the †closed range theorem. These theorems were modified to be applicable in locally convex topological linear spaces by the Bourbaki group beginning in the late 1940s.

In 1929, von Neumann proved, as a mathematical foundation of quantum mechanics, that a closed linear operator T in a Hilbert space admits spectral resolution with real spectra if and only if T is a †self-adjoint operator (→ also Stone [3]). The condition that a closed linear operator is a †function of a self-adjoint operator was given by von Neumann, F. Riesz, and Y. Mimura (1934–1936). K. Friedrichs (1934) proved that a †semibounded linear operator admits a self-adjoint extension. T. Kato [4] (1950) proved that a Schrödinger-type Hermitian operator is †essentially self-adjoint.

Von Neumann's †mean ergodic theorem in the Hilbert space (1932) was extended to Banach spaces by K. Yosida, S. Kakutani, and Riesz in 1936. G. D. Birkhoff's †pointwise ergodic theorem (1931) was extended by N. Wiener (1939), Yosida (1940), E. Hopf (1954), N. Dunford (1955), R. V. Chacon and D. S. Ornstein (1960), and others in various ways. The †Abelian ergodic theorems were discussed by E. Hille and R. S. Phillips [5] and Yosida [6].

The notion of †Banach algebra was introduced by M. Nagumo in 1936. I. M. Gel'fand proved that a commutative Banach algebra with multiplicative unit e satisfying $\|e\|=1$ (= the normed ring) admits a representation by an algebra of complex-valued continuous functions (1941).

The †reflexivity as well as the †duality of Banach spaces were studied by S. Kakutani (1939), V. L. Shmul'yan (1940), and W. T. Eberlein (1941).

The notion of †vector lattice (= †Riesz space) was introduced in analysis by Riesz (1930). This was followed by the work of L. V. Kantrovitch (1935) and H. Freudenthal (1936). Kakutani gave two standard types of †Banach lattice: the †abstract (*M*) space (1940) and the †abstract (*L*) space (1941). M. Krein and S. Krein (1940) and Yosida and M. Fukamiya (1940) discussed the (*M*)-type vector lattices, and Yosida (1941), the (*L*)-type vector lattices. H. Nakano (1940–1941) and T. Ogasawara and F. Maeda (1942) studied the spectral resolution of the Banach lattice.

The connection between †Brownian motion and †potential theory was clarified by Kakutani (1942), and extended by J. L. Doob (1956) and G. A. Hunt (1957–1958).

The †one-parameter semigroup of continuous linear operators in Banach spaces was studied by Hille and Yosida, and they gave in 1948 a characterization of the †infinitesimal generator of such semigroups. Its dual was given by Phillips (1955). The one-parameter semigroup of nonlinear †contractive operators in Hilbert spaces was studied by Y. Kōmura (1967), who obtained a nonlinear version of the Hille-Yosida theorem. This has been extended considerably in Banach spaces by many scholars, e.g., Kato (1967), M. G. Crandall and A. Pazy (1969), Crandall and T. Liggett (1971), I. Miyadera and S. Oharu (1970), H. Brezis (1973), P. Bénilan (1973), and others.

In 1936, S. Sobolev gave a generalization of the notion of functions and their derivatives through integration by parts. This generalization has been extended by L. Schwartz (1945–) [9] to the notion of †distributions, which are continuous linear functionals defined on the function spaces $\mathscr{D}(\mathbf{R}^n)$ and $\mathscr{S}(\mathbf{R}^n)$, and this extension gives, e.g., a reasonable interpretation of Dirac's δ-function. Since 1959, Gel'fand [10] has been publishing, with his collaborators, books on the distribution theory pertaining to function spaces other than $\mathscr{D}(\mathbf{R}^n)$ and $\mathscr{S}(\mathbf{R}^n)$. M. Sato introduced (1959–1960) [11] the theory of †hyperfunctions, as a generalization of distribution theory. In the case of one independent variable, a hyperfunction f may be defined as a generalized boundary value on the real axis R of a holomorphic function F defined in $\mathbf{C}^1-\mathbf{R}^1$. Hyperfunction theory has been refined to †microlocal analysis and studied extensively by Sato, A. Martineau, H. Komatsu, P. Schapira, T. Kawai, M. Kashiwara, M. Morimoto, A. Kaneko, and others (→ [12]).

References

[1] S. Banach, Théorie des Opérations Linéaires, Warsaw, 1932.
[2] J. von Neumann, Mathematische Grundlagen der Quantenmechanik, Springer, 1932.
[3] M. H. Stone, Linear transformations in Hilbert spaces and their applications to analysis, Amer. Math. Soc., 1932.
[4] T. Kato, Perturbation theory for linear operators, Springer, second edition, 1976.
[5] E. Hille and R. S. Phillips, Functional analysis and semigroups, Amer. Math. Soc., 1957.
[6] K. Yosida, Functional analysis, Springer, sixth edition, 1980.
[7] N. Dunford and J. Schwartz, Linear operators, Interscience, I, 1958; II, 1963; III, 1971.
[8] H. Brezis, Opérateurs Maximaux Monotones et Semigroupes de Contractions dans les Espaces de Hilbert, American Elsevier, 1973.
[9] L. Schwartz, Théorie des Distributions, Hermann, 1966.
[10] I. M. Gel'fand, Generalized functions I–III (with G. E. Shilov), IV (with N. Ya. Vilenkin), V (with M. I. Graev and N. Ya. Vilenkin), Academic Press, 1964–1966.
[11] M. Sato, Theory of Hyperfunctions I, II, J. Fac. Sci. Univ. Tokyo, sec. I, 8 (1959), 139–193; 8 (1960), 387–437.
[12] M. Sato, T. Kawai, and M. Kashiwara, Microfunctions and pseudodifferential equations, Lecture notes in math. 287, Springer, 1973.

163 (XIII.16) Functional-Differential Equations

A. General Remarks

In many models, it is assumed that the future behavior of a system under consideration is governed only by its present state and not by past states. However, for various systems arising in practical problems we cannot ignore the effect of the past on the future. Such a phenomenon is often observed in population problems, epidemiology, chemical reactions, system engineering, and so on.

The description of such phenomena may involve **difference-differential equations**

$$\dot{x}(t)=f(t,x(t),x(t-h_1),\ldots,x(t-h_m)),\tag{1}$$

or **integrodifferential equations**

$$\dot{x}(t)=g(t,x(t))+\int_0^t f(t,s,x(t),x(s))\,ds.\tag{2}$$

An enormous variety of equations is discussed in the literature, but most of them can be

expressed in the form

$$\dot{x}(t)=f(t,x(s)), \tag{3}$$

which is usually called the **functional differential equation**; here s varies in an interval I_t and f depends on the function x of I_t. If t is the right or left endpoint of the interval I_t, then (3) is said to be of **retarded** or **advanced type**, respectively. Most of the results obtained have been from work on equations of retarded type. In this case, the maximal length of the interval I_t is called the **retardation (delay, lag, deviation,** etc.), and (3) is often called the **retarded differential equation, delay differential equation,** or **differential equation with lag (retardation, deviating argument**).

B. Historical Remarks

Functional differential equations of a sort were studied by Johann Bernoulli in the 18th century in connection with the string problem. Since then, a good deal of work has been done in this field by many people; some of this work was done before the beginning of this century. Among the investigators, Volterra [1, 2] is noteworthy for his systematic study on rather general equations related to problems of predator-prey populations and viscoelasticity, though his results were ignored by his contemporaries. In the early 1940s, Minorsky [3], in his famous study of ship stabilization, pointed out the importance of delay effects in control theory, with many of the modern issues first appearing in his work [4]. Mishkis [5] studied linear systems extensively, and Driver [6] gave a unified representation for functional differential equations. Important achievements were given by Krasovskiĭ [7] and Bellman and Cooke [8]. They laid the foundation for the qualitative theory of functional differential equations. Inspired by these works, many books (such as [9–14]) were published. Owing to these books together with many articles on this field, the theory of functional differential equations has become an important branch in the theory of differential equations.

Equations of advanced type are treated in, e.g., [5, 8], but qualitative theory for them has hardly been established. The equation

$$\dot{x}(t)=ax(t)+bx(\lambda t) \quad \text{on} \quad 0\leqslant t<\infty \tag{4}$$

is of advanced type if $\lambda>1$, and its analytic solution has been studied in detail. Equation (3), whose right-hand side involves differential operators, e.g., $\dot{x}(t)=f(t,x(t),x(t-h),\dot{x}(t-h))$, is said to be of **neutral type** [5, 8, 14], and is generally considered to be of neither retarded nor advanced type because of its distinctive features. However, Hale [14] originated the study of a class of equations of neutral type for which the general qualitative theory can be developed in the same manner as for equations of retarded type.

For the case of infinite retardation, proper choice of the space of functions x contained in the right-hand side of (3) is of great significance. A general treatment of the phase space for equations with infinite retardation in an axiomatic setting is given in [15].

C. Phase Space

For simplicity, let I_t always be the interval $[t-h,t]$ with a finite retardation $h>0$. A **solution** of (3) starting at $t=\tau$ is a function defined on $[\tau-h,a)$ for $a>\tau$ such that the solution coincides with a preassigned function (the **initial function**) on $[\tau-h,\tau]$ and is continuously differentiable and satisfies (3) on $[\tau,a)$. Since a solution starting at $t=\tau$ must be continuous for $t\geqslant\tau$, it is quite natural, as is done in much of the literature in order to develop a qualitative theory, to choose the space $\mathbf{C}([t-h,t],\mathbf{R}^n)$ of continuous $\mathbf{R}^n$-valued functions on $[t-h,t]$ as a space of functions x involved in the right-hand side of (3). Introducing a symbol x_t which represents an element of $C=C([-h,0],\mathbf{R}^n)$ defined by $x_t(s)=x(t+s)$ $(s\in[-h,0])$, we can rewrite (3) in a more convenient form:

$$\dot{x}(t)=f(t,x_t), \tag{E}$$

where the function $f(t,\varphi)$ is defined on $\mathbf{R}\times C$ (or on its subspace). Here C is considered to be a Banach space with the norm $\|\varphi\|=\max_{s\in[-h,0]}|\varphi(s)|$, $|\cdot|$ being a norm in $\mathbf{R}^n$. The space C is said to be the **phase space** of (E), and the initial condition can be written as

$$x_\tau=\xi, \qquad (\tau,\xi)\in\mathbf{R}\times C. \tag{5}$$

D. Initial Value Problem

Let $f(t,\varphi)$ be a continuous functions defined on an open domain $D\subset\mathbf{R}\times C$. The **initial value problem** is to find a solution of (E)–(5) for a given $(\tau,\xi)\in D$. Many fundamental results hold as for ordinary differential equations: (a) There always exists a solution of (E)–(5) under the continuity of f on D. Here, the solution means the one to the right. No general existence theorem is given for the solution to the left. (b) The solution of (E)–(5) is unique, if f satisfies the **Lipschitz condition**: $|f(t,\varphi)-f(t,\psi)|\leqslant L\|\varphi-\psi\|$ in a neighborhood of each point of D, where L is a constant depending on the neighborhood. (c) If f varies in the space $C(D,\mathbf{R}^n)$

equipped with the compact open topology and if the solution of (E)–(5) is unique for every $(\tau,\xi)\in D$, then the solution is continuous as a function of (τ,ξ,f). (d) If f is completely continuous on D, that is, f is a continuous mapping of bounded subsets of D into bounded sets in $\mathbf{R}^n$, then every solution can be continuable to the right as long as it remains bounded or stays away from the boundary of D. This assertion is no longer true in the absence of complete continuity.

To prove the existence theorem, Schauder's †fixed point theorem is utilized; the Picard successive method is also effective under the Lipschitz condition. Consider the case when $f(t,\varphi)$ can be written as $f(t,\varphi)=g(t,\varphi(0),\varphi)$, where $g(t,x,\varphi)$ is defined for $(t,x,\varphi)\in\mathbf{R}\times\mathbf{R}^n\times C$, and $g(t,x,\varphi)=g(t,x,\psi)$ if $\varphi(s)=\psi(s)$ on $[-h,-\delta]$, $\delta>0$. Equation (1) is one such case, where $h=\max h_k$, $\delta=\min h_x$. Then, $g(t,x,x_t)$ is a function of (t,x) alone on $[\tau,\tau+\delta]$ under the given condition (5), that is, (E) is reduced to an ordinary differential equation. This makes it possible to find a solution of (E)–(5) by matching successively the solutions of ordinary differential equations on $[\tau,\tau+\delta]$, $[\tau+\delta,\tau+2\delta]$, ... This is the **step-by-step method**, which is effective even for equations of neutral type with the same property.

Under uniqueness, the solution $x(t)$ of (E) induces a mapping $T(t,\tau):G(t,\tau)\to C$, $t\geqslant\tau$, which maps $x_\tau\in G(t,\tau)$ to $x_t\in C$, where $G(t,\tau)\subset C$ is the set of ξ for which the solution of (E)–(5) is continuable up to t. Then

$$T(t,s)T(s,\tau)=T(t,\tau)\quad\text{for}\quad t\geqslant s\geqslant\tau, \tag{6}$$

and $T(t,\tau)$ is strongly continuous in t. If (E) is **autonomous**, that is, $f(t,\varphi)$ is independent of t, then there exists a mapping $T(t)$, $t\geqslant 0$, satisfying $T(t-\tau)=T(t,\tau)$, and $\{T(t)\}_{t\geqslant 0}$ becomes a one-parameter semigroup.

E. Linear System

When $f(t,\varphi)$ is continuous on $I\times C$ for an interval I and is linear in φ, equation (E) is said to be a **linear system** (denoted by (L)). In this case, $f(t,\varphi)$ satisfies the Lipschitz condition with $L=L(t)$ on $I\times C$ for a continuous function $L(t)$ of I, which also implies that f is completely continuous. Thus the solution of (L)$-$(5) uniquely exists over $[\tau,\infty)\cap I$ for any $(\tau,\xi)\in I\times C$, and the mapping (the **fundamental** or **solution operator**) $T(t,\tau)$ is a †bounded linear operator on C for any $t,\tau\in I$, $t\geqslant\tau$, and †compact if $t\geqslant\tau+h$. $T(t,\tau)$ corresponds to the fundamental matrix for ordinary linear differential equations, but it is not invertible in general.

A continuous function $f(t,\varphi)$ linear in φ may be expressed in the form

$$f(t,\varphi)=\int_{-h}^{0}[d_s\eta(t,s)]\varphi(s), \tag{7}$$

where $\eta(t,s)$ is an $n\times n$ matrix defined on $I\times[-h,0]$, which is measurable in (t,s) and of †bounded variation in s with a †total variation less than $L(t)$, and the integration is that of Stieltjes with respect to s. From this fact it follows that the range of the initial functions can be extended to the space of †piecewise-continuous functions, and the solution $x(t)$ of the nonhomogeneous linear system $\dot{x}(t)=f(t,x_t)+p(t)$ through $(\tau,\xi)\in I\times C$ can be represented by the **constant variational formula**

$$x(t)=[T(t,\tau)\xi](0)+\int_\tau^t[T(t,s)\Gamma](0)p(s)\,ds,\qquad t\geqslant\tau,$$

where $\Gamma(s)=E$ (the unit matrix) for $s=0$, and $\Gamma(s)=O$ (the zero matrix) for $s<0$.

F. Autonomous Linear System

Let (L) be an autonomous linear system

$$\dot{x}(t)=f(x_t). \tag{AL}$$

In this case, (7) is not different from the Riesz representation theorem $f(\varphi)=\int_{-h}^{0}[d\eta(s)]\varphi(s)$. For (AL) the fundamental operator $T(t)$ plays an extremely important role. As was seen, $\{T(t)\}_{t\geqslant 0}$ is a one-parameter semigroup of bounded linear operators $T(t)$ which is strongly continuous in $t\geqslant 0$ and compact for $t\geqslant h$. Thus the asymptotic behavior of the solutions of (AL) are determined by the distribution of the spectra $\sigma(A)$ of the infinitesimal generator A of $T(t)$, which is given by $A\varphi=\dot{\varphi}$ with the domain $D(A)=\{\varphi\in C\mid\dot{\varphi}\in C$ and $\dot{\varphi}(0)=f(\varphi)\}$. The properties of $T(t)$ assert that: (a) $\sigma(A)$ consists of point spectra alone. (b) The number of $\Lambda_\alpha=\{\lambda\in\sigma(A)\mid\operatorname{Re}\lambda\geqslant\alpha\}$ is at most finite for any $\alpha\in R$. (c) For every $\lambda\in\sigma(A)$ the dimension of the generalized eigenspace of λ is finite. (d) The spectra of A coincide with the roots (**characteristic roots**) of the **characteristic equation** of (AL),

$$\det\left[\lambda E-\int_{-h}^{0}e^{\lambda s}d\eta(s)\right]=0,$$

together with their multiplicities. (e) $e^{\lambda t}\in\sigma(T(t))$, $t\geqslant 0$, if and only if $\lambda\in\sigma(A)$.

Let P_α, $\alpha\in R$, be the linear space spanned by the generalized eigenfunctions corresponding to a $\lambda\in\Lambda_\alpha$. Then, (a) P_α is invariant under $T(t)$, that is, $T(t)P_\alpha\subset P_\alpha$ for $t\geqslant 0$; (b) the restriction of $T(t)$ to P_α is invertible and hence extendable over $t\in R$; (c) if $\xi\in P_\alpha$, then $[T(t)\xi](0)=\sum_{\lambda\in\Lambda_\alpha}p_\lambda(t,\xi)e^{\lambda t}$, where $p_\lambda(t,\xi)$ are polynomials

in t and linear in ξ; (d) there is a direct-sum decomposition $C=P_\alpha+Q_\alpha$ such that Q_α is invariant under $T(t)$, the projection along $Q_\alpha \pi_\alpha$: $C\to P_\alpha$ is bounded, and there exist positive constants K, ε for which $\|T(t)\xi\|\leqslant Ke^{(\alpha-\varepsilon)t}\|\xi\|$, $t\geqslant 0$ if $\xi\in Q_\alpha$. Hence for any $\xi\in C$ the solution $x(t)$ of (AL) through ξ at $t=0$ satisfies

$$\left|x(t)-\sum_{\lambda\in\Lambda_\alpha} p_\lambda(t,\pi_\alpha\xi)e^{\lambda t}\right|\leqslant K(1+\|\pi_\alpha\|)e^{(\alpha-\varepsilon)t}\|\xi\|,\quad t\geqslant 0, \quad (8)$$

where $\|\pi_\alpha\|$ denotes the operator norm of π_α. The set $Q=\bigcap_{\alpha\in R}Q_\alpha$ may contain an element ξ other than the zero element, but $T(t)\xi$ must be identically zero for $t\geqslant hn$.

For a linear system (L) with $f(t,\varphi)$ ω-periodic in t, similar conclusions result, with $T(\omega,0)$ in place of A; and relation (8) holds, where Λ_α is defined similarly by replacing $\sigma(A)$ by $\{(1/\omega)\log\mu \mid \mu\in\sigma(T(\omega,0))\}$ and $p_\lambda(t,\xi)$ are polynomials with ω-periodic coefficients. This corresponds to †Floquet's theorem for ordinary periodic linear systems. $\mu\in\sigma(T(\omega,0))$ and $(1/\omega)\log\mu$ are said to be the **characteristic multiplier** and the **characteristic exponent**, respectively.

G. Stability Problem

The concept of stability can be defined and studied in the same spirit as for ordinary differential equations, and the Lyapunov second method also turns out to be effective. For instance, the zero solution of (E) with f defined on $D_H=[0,\infty)\times\{\varphi\in C \mid \|\varphi\|<H\}$ for $H>0$ is said to be **uniformly asymptotically stable** if there are a constant $\alpha>0$ and positive functions $\delta(\varepsilon)$ and $\sigma(\varepsilon)$ of $\varepsilon>0$ such that any solution $x(t)$ of (E) satisfies

$$|x(t)|<\varepsilon \text{ as long as } x(t) \text{ exists}, \quad (9)$$

whenever $\|x_\tau\|<\delta(\varepsilon)$ and $t\geqslant\tau$ or $\|x_\tau\|\leqslant\alpha$ and $t\geqslant\tau+\sigma(\varepsilon)$ for $\tau\geqslant 0$. In the above definition, if f is completely continuous on D_H and if $0<\varepsilon<H$, then the phrase "as long as $x(t)$ exists" is redundant, and (9) is equivalent to the claim that $x(t)$ exists for all $t\geqslant\tau$ and $|x(t)|<\varepsilon$. Under the Lipschitz condition and the complete continuity on f in (E), the zero solution of (E) is uniformly asymptotically stable if and only if there exists a continuous $\mathbf{R}$-valued function (**Lyapunov function**) $V(t,\varphi)$ defined on D_{H_1} for $0<H_1\leqslant H$ such that

(i) $\quad a(|\varphi(0)|)\leqslant V(t,\varphi)\leqslant b(\|\varphi\|)$,

(ii) $\quad D^+V(t,x_t)\leqslant -cV(t,x_t)$

as long as $(t,x_t)\in D_{H_1}$ for a solution $x(t)$ of (E), where $a(r)$, $b(r)$ are continuous functions with $a(r)>0$ for $r>0$, $b(0)=0$, $c>0$ is a constant, and D^+ denotes the upper-right Dini derivative. Furthermore, the Lyapunov function can be endowed with a Lipschitz condition $|V(t,\varphi)-V(t,\psi)|\leqslant L_1\|\varphi-\psi\|$ for some constant L_1 if f in (E) is uniformly Lipschitzian, that is, the coefficient L is constant over the domain. The Lipschitz condition together with (ii) makes it possible for a solution $y(t)$ of the perturbed equation

$$\dot y(t)=f(t,y_t)+P(t,y_t) \quad (10)$$

to satisfy

$$D^+V(t,y_t)\leqslant -cV(t,y_t)+L_1|P(t,y_t)|.$$

By using this fact, the stability of (10) can be discussed.

However, for functional differential equations it is quite a difficult problem to construct a suitable Lyapunov function for a given (E).

Many of the attempts to improve the sufficiency condition that have been made up to now are such that a stability property can be verified by means of a simple Lyapunov function. The main effort has been devoted to replacing condition (ii) by another type of condition. One of them is

(ii*) $\quad D^+V(t,x_t)\leqslant -c(|x(t)|)$

under the uniform boundedness of $f(t,\varphi)$, where $c(r)$ is a continous function with $c(r)>0$ for $r>0$. Another one is for the case when $V(t,\varphi)=W(t,\varphi(0))$, where $W(t,x)$ is defined for $(t,x)\in\mathbf{R}\times\mathbf{R}^n$. In this case, (ii) can be replaced by

(ii**) $\quad D^+V(t,x_t)\leqslant -cV(t,x_t)$

whenever $V(t+s,x_{t+s})\leqslant F(V(t,x_t))$ for $s\in[-h,0]$, where $c>0$ is a constant and $F(r)$ is a continuous function satisfying $F(r)>r$ for $r>0$. The condition (ii**) was given by B. S. Razumikhin and provides an easier way to construct a Lyapunov function.

For a linear system, uniform asymptotic stability implies $\|T(t,\tau)\|\leqslant 1/\delta(1)$ for $t\geqslant\tau$ and $\|T(t,\tau)\|\leqslant 1/2$ for $t\geqslant\tau+\sigma(\alpha/2)+h$. These facts together with (6) show that the zero solution is **exponentially stable**, that is, $|x(t)|\leqslant Ke^{-\nu(t-\tau)}\|x_\tau\|$ for $t\geqslant\tau$, where $\nu=(\log 2)/(\sigma(\alpha/2)+h)$, $K=2/\delta(1)$. From (8) it follows that the zero solution of (AL) is uniformly asymptotically stable if and only if

$$\{\lambda\in\sigma(A) \mid \operatorname{Re}\lambda\geqslant 0\}=\varnothing. \quad (11)$$

H. Equations of Neutral Type

To deal with an equation of neutral type, such as

$$\dot x(t)=f(t,x_t,\dot x_t), \quad (12)$$

a phase space such as $C^1=C^1([-h,0],\mathbf{R}^n)$ of continuously differentiable functions should be chosen. The solution of (E) will become smooth as time elapses if f in (E) is sufficiently smooth. However, this property cannot be expected for (12), and the solution of (12)–(5) need not belong to $\mathbf{C}^1$ even if $\xi\in C^1$ without a specific condition on ξ such as $\dot{\xi}(0)=f(\tau,\xi,\dot{\xi})$. Such a restriction on initial functions is an obstruction to the development of the general theory. Equations of the form

$$\frac{d}{dt}D(t,x_t)=f(t,x_t) \qquad \text{(N)}$$

cover a fairly general class of equations of neutral type, where $D(t,\varphi)$ and $f(t,\varphi)$ are defined on an open set $D\subset\mathbf{R}\times C$ and a **solution** of (N)–(5) means a continuous function $x(t)$ defined on $[\tau-h,a)$, $a>\tau$, which satisfies (5) and

$$D(t,x_t)=D(\tau,\xi)+\int_\tau^t f(s,x_s)\,ds \quad \text{on} \quad [\tau,a).$$

(E) corresponds to the case where $D(t,\varphi)=\varphi(0)$, while $\dot{x}(t)=G(t,\dot{x}_t)+g(t,x_t)$ is reduced to the form (N) if $G(t,\varphi)$ is linear in φ and continuously differentiable with respect to t by setting $D(t,\varphi)=\varphi(0)-G(t,\varphi)$ and $f(t,\varphi)=g(t,\varphi)-(\partial/\partial t)G(t,\varphi)$.

A continuous function $D(t,\varphi)$ linear in φ is said to be **atomic at** 0 if in the representation

$$D(t,\varphi)=\int_{-h}^0[d_s\mu(t,s)]\varphi(s)$$

(see equation (7)) $P(t)=\mu(t,0)-\mu(t,-0)$ exists and is nonsingular, which is equivalent to

$$D(t,\varphi)=P(t)\varphi(0)+\int_{-h}^0[d_s\mu_0(t,s)]\varphi(s) \qquad (13)$$

with a nonsingular matrix $P(t)$ and a matrix $\mu_0(t,s)$ whose total variation with respect to s on $[-\sigma,0]$ tends to 0 as $\sigma\to 0$ locally uniformly in t. A nonlinear function $D(t,\varphi)$ is said to be **atomic at** 0 if $D(t,\varphi)$ has a continuous †Fréchèt derivative $D_\varphi(t,\varphi)$ with respect to φ and $D_\varphi(t,\varphi)$ is atomic at 0. In equation (N), $D(t,\varphi)$ is always assumed to be atomic at 0, and many of the fundamental theorems for (E) are also valid for (N). Let $T(t)$ be the operator solution of the autonomous linear equation

$$\frac{d}{dt}D(x_t)=f(x_t). \qquad (14)$$

Then $\{T(t)\}_{t\geqslant 0}$ is a strongly continous semigroup of bounded linear operators, and the corresponding †infinitesimal generator A has the domain $D(A)=\{\varphi\in C\mid\dot{\varphi}\in C$ and $D(\dot{\varphi})=f(\varphi)\}$. The properties of the spectra $\sigma(A)$ are the same as for (AL), except that the characteristic equation is now given by

$$\det\left[\lambda(P+\int_{-h}^0 e^{\lambda s}\,d\mu_0(s))-\int_{-h}^0 e^{\lambda s}\,d\eta(s)\right]=0$$

and the property (b) is true if $\alpha>a_D$, where $D(\varphi)$, $f(\varphi)$ are assumed to be of the forms (13) and (7), respectively, without the argument t, and where a_D is given by

$$a_D=\max\left\{\operatorname{Re}\lambda\,\middle|\,\det\left[P+\int_{-h}^0 e^{\lambda s}\,d\mu_0(s)\right]=0\right\}.$$

Similarly, the decomposition $C=P_\alpha+Q_\alpha$ is possible if $\alpha>a_D$. Hence, to obtain the uniform asymptotic stability of the zero solution of (14), the condition $a_D<0$ should be assumed in addition to (11). The linear function $D(\varphi)$ is said to be **stable** if $a_D<0$. If $D(\varphi)$ is stable, then the Lyapunov method is also applicable, as a sufficient condition, to the stability problem of (N) with linear $D(\varphi)$ instead of general $D(t,\varphi)$, but in this case condition (i) for $V(t,\varphi)$ should be replaced by

$$a(|D(\varphi)|)\leqslant V(t,\varphi)\leqslant b(\|\varphi\|).$$

I. Infinite Retardation

There are some basic differences between cases of finite retardation and of infinite retardation. For example, in the definition of stability, the inequality $|x(t)|<\varepsilon$ in (9) can be replaced by $\|x_t\|<\varepsilon$ with no difference in the case of finite retardation, but this replacement yields a different concept deeply connected with the choice of the phase space in case of infinite retardation.

There are several ways to define a phase space for a functional differential equation with infinite retardation. One of those phase spaces generally used is a linear space X of functions: $(-\infty,0]\to\mathbf{R}^n$ with a seminorm $\|\cdot\|_X$ such that if a function $x:(-\infty,a)\to\mathbf{R}^n$ satisfies $x_\tau\in X$ and it is continuous on $[\tau,a]$, then

(i) $x_t\in X$ for all $t\in[\tau,a)$,

(ii) x_t is continuous as a function $[\tau,a)\to X$,

(iii) $m|x(t)|\leqslant\|x_t\|_X\leqslant K\max_{s\in[\tau,t]}|x(s)| + M(t-\tau)\|x_\tau\|_X$,

where m and K are positive constants and $M(t)$ is a continuous function.

If X satisfies the foregoing conditions and if $f(t,\varphi)$ is continuous on an open domain $D\subset\mathbf{R}\times X$, then the local properties (a)–(c) in Section D hold, and so does (d) when $D=\mathbf{R}\times X$. On the other hand, if X has a **fading memory**, namely,

(iv) $M(t)\leqslant Me^{-\mu t}$ in (iii) for positive constants M, μ,

then the procedure in Section F is applicable by restricting $\sigma(A)$ to $\sigma_\mu(A)=\{\lambda\in\sigma(A)\mid \mathrm{Re}\,\lambda> -\mu\}$ and, hence the zero solution of (AL) with infinite retardation is uniformly asymptotically stable under (11).

For $0\leqslant h\leqslant\infty$ and $0\leqslant\gamma<\infty$, both the space $C_{h,\gamma}$ of continuous functions $\varphi:(-h,0]\to\mathbf{R}^n$ with a finite limit $\lim_{s\to -h}e^{\gamma s}\varphi(s)$ and the space $M_{h,\gamma}$ of measurable functions $\varphi:(-h,0]\to\mathbf{R}^n$ with $\int_{-h}^{0}e^{\gamma s}|\varphi(s)|\,ds<\infty$ satisfy the foregoing conditions, where the norms are given by $\|\varphi\|_{C_{h,\gamma}}=\sup_{(-h,0]}e^{\gamma s}|\varphi(s)|$ and $\|\varphi\|_{M_{h,\gamma}}=|\varphi(0)| +\int_{-h}^{0}e^{\gamma s}|\varphi(s)|\,ds$. These spaces have a fading memory if $h<\infty$ or $\gamma>0$.

In much of the literature in which the equation of infinite retardation has a form like (2) or (4), the space CB of bounded continuous functions or the space C_0 of continuous functions with compact supports are sufficient as a phase space under the norm $\|\varphi\|= \sup_{s\leqslant 0}|\varphi(s)|$. However, it is to be noted that C_0 is not complete when CB does not satisfy condition (ii). When CB is chosen as the phase space, in order to show the existence theorem, $f(t,x_t)$ should be continuous for any bounded continuous function x in addition to the continuity in (t,φ). This condition is satisfied if $I_t =[g(t),t]$ in (3) for a continuous function $g(t)\leqslant t$. Equation (2) or (4) with $0\leqslant\lambda<1$ give rise to such a case, but $g(t)$ must be equal to 0 in (2) and (4) with $\lambda=0$. If $g(t)\to\infty$ as $t\to\infty$, the Razumikhin condition (ii**) $D^+V(t,x_t)\leqslant -cV(t,x_t)$ whenever $V(s,x_s)\leqslant F(V(t,x_t))$ $(s\in[g(t),t])$ for a Lyapunov function is effective, but one can conclude only that $x(t)\to 0$ as $t\to\infty$ without uniformity.

References

[1] V. Volterra, Sur la théorie mathématique des phénomènes héréditaires, J. Math. Pures Appl., 7 (1928), 249–298.
[2] V. Volterra, Théorie mathématique de la Lutte pour la Vie, Gauthier-Villars, 1931.
[3] N. Minorsky, Self-excited oscillations in dynamical systems possessing retarded actions, J. Appl. Mech., 9 (1942), 65–71.
[4] A. M. Zverkin, G. A. Kamenskiĭ, S. B. Norkin, and L. E. El'sgol'ts, Differential equations with retarded arguments (in Russian), Uspekhi Mat. Nauk., 17 (1962), 77–164.
[5] A. D. Mishkis, Lineare Differentialgleichungen mit nacheilendem Argument, Deutscher Verlag Wiss., 1955. (Original in Russian, 1957.)
[6] R. D. Driver, Existence and stability of solution of a delay-differential system, Arch. Rational Mech. Anal., 10 (1962), 401–426.
[7] N. N. Krasovskiĭ, Stability of motion, Stanford Univ. Press, 1963. (Original in Russian, 1959.)
[8] R. Bellman and K. L. Cooke, Differential-difference equations, Academic Press, 1963.
[9] L. E. El'sgol'ts and S. B. Norkin, Introduction to the theory and application of differential equations with deviating argument, Academic Press, 1971. (Original in Russian, 1963.)
[10] A. Halanay, Differential equations: Stability, oscillations, time lags, Academic Press, 1965.
[11] M. N. Oğuztöreli, Time-lag control systems, Academic Press, 1966.
[12] T. Yoshizawa, Stability theory by Liapunov's second method, Math. Soc. Japan, 1966.
[13] V. Lakshmikantham and S. Leela, Differential and integral inequalities II, Academic Press, 1969.
[14] J. K. Hale, Theory of functional differential equations, Springer-Verlag, 1977.
[15] J. K. Hale and J. Kato, Phase space for retarded equations with infinite delay, Funkcial. Ekvac., 21 (1978), 11–41.

164 (XII.18) Function Algebras

A. Definition [1–3]

Let $C(X)$ be the †Banach algebra of all continuous complex-valued functions on a compact Hausdorff space X with pointwise operations and †uniform norm. A **function algebra** (or **uniform algebra**) on X is a closed subalgebra A of $C(X)$ containing the constant functions and separating the points of X, i.e., for any $x,y\in X$, $x\neq y$, there exists $f\in A$ with $f(x)\neq f(y)$. If φ_x, $x\in X$, denotes the evaluation mapping $f\to f(x)$ of A, the correspondence $x\to\varphi_x$ is a homeomorphism of X into the †maximal ideal space $\mathfrak{M}(A)$ of A. Since $\hat{f}(\varphi_x)=f(x)$ for any $x\in X$ and $f\in A$, the †Gel'fand representation of A is an isometric isomorphism. By identifying x with φ_x, we regard X as a closed subset of $\mathfrak{M}(A)$ and the Gel'fand transform $\hat{f}$, $f\in A$, as a continuous extension of f to $\mathfrak{M}(A)$.

B. Examples [1, 3, 4]

For a compact plane set K let $P(K)$ $(R(K))$ be the subalgebra of all functions in $C(K)$ that can be approximated uniformly on K by polynomials (rational functions with poles off K). $A(K)$ denotes the subalgebra of all functions in $C(K)$ that are analytic in the interior of K. These are function algebras on K and $P(K)\subseteq R(K)\subseteq A(K)$. When K is the unit circle $\mathbf{T}=$

$\{|z|=1\}$, $P(\mathbf{T})$ is called the **disk algebra**, the most typical concrete example.

The theory of function algebras emerged from attempts to solve, by means of functional analysis, certain problems in complex analysis, especially problems of uniform approximation, e.g., when does $A(K)=P(K)$ or $A(K)=R(K)$ hold?

Another important example is given by the algebra $H_\infty(U)$ of all bounded analytic functions on a bounded plane domain U with the supremum norm $\|f\|_\infty=\sup\{|f(z)| \mid z\in U\}$ [5]. Since the Gel'fand representation is an isometric isomorphism of $H_\infty(U)$ into the algebra $C(\mathfrak{M}(H_\infty(U)))$, $H_\infty(U)$ is viewed as a function algebra on the space $\mathfrak{M}(H_\infty(U))$. Further examples result if we take K or U in a Riemann surface or in the complex n-space.

We list a few abstract function algebras that reflect certain relevant properties possessed by concrete examples. A is called a **Dirichlet** (resp. **logmodular**) **algebra** if the set $\mathrm{Re}\,A=\{\mathrm{Re}\,f \mid f\in A\}$ (resp. $\log|A^{-1}|=\{\log|f| \mid f, f^{-1}\in A\}$) is dense in $C_{\mathbf{R}}(X)$, the space of all continuous real-valued functions on X. It is called **hypo-Dirichlet** if the closure of $\mathrm{Re}\,A$ has finite codimension in $C_{\mathbf{R}}(X)$, and the linear span of $\log|A^{-1}|$ is dense in $C_{\mathbf{R}}(X)$.

In the following, A denotes a function algebra on X unless otherwise specified.

C. Boundary and Representing Measure [1–4]

A subset E of X is a **boundary** for A if for any $f\in A$ there exists $x\in E$ such that $|f(x)|=\|f\|$. A **closed boundary** is a boundary closed in X. G. E. Shilov proved that there is a smallest closed boundary, ∂A, which is called the **Shilov boundary** for A. A positive Borel measure μ on X is a **representing measure** for $\varphi\in\mathfrak{M}(A)$ if $f(\varphi)=\int f(x)\,d\mu(x)$ for all $f\in A$. Each $\varphi\in\mathfrak{M}(A)$ has a representing measure supported by ∂A. The **Choquet boundary**, $c(A)$, consists of all $x\in X$ such that the evaluation φ_x at x has a unique representing measure. Then $c(A)$ is a boundary, whose closure is ∂A. If X is metrizable, $c(A)$ is a G_δ set in X and supports a representing measure for every member of $\mathfrak{M}(A)$.

For $\varphi\in\mathfrak{M}(A)$, M_φ denotes the set of representing measures for φ. It is a †weak* compact convex subset of the space of measures on X. M_φ is a singleton if A is Dirichlet or logmodular. It is finite-dimensional if A is hypo-Dirichlet. The case $\dim M_\varphi<+\infty$ has been studied in detail. Extensive studies for the case $\dim M_\varphi=+\infty$ have been done only for concrete examples related to polydisks, infinitely connected domains, etc.

The notion of boundary reflects the †maximum modulus principle for analytic functions, and representing measures come from †Poisson's integral formula. The most relevant maximum principle is H. Rossi's **local maximum modulus principle**: For any closed set K in $\mathfrak{M}(A)$, we have $\|\hat{f}\|_K=\|\hat{f}\|_{\mathrm{bd}\,K\cup(K\cap\partial A)}$ for all $f\in A$, where $\mathrm{bd}\,K$ is the topological boundary of K and $\|\hat{f}\|_S=\sup\{|\hat{f}(s)| \mid s\in S\}$. A corresponding result for function spaces was considered by Y. Hirashita and J. Wada. Closely related to representing measures are the **orthogonal measures** for A, i.e., complex Borel measures μ on X such that $\int f\,d\mu=0$ for all $f\in A$; they are often useful in studying function algebras by means of the duality technique. The set of orthogonal measures is denoted by $A^\perp$.

D. Peak Sets [1–3]

A subset K of X is a **peak set** for A if there exists $f\in A$ such that $f(x)=1$ for $x\in K$ and $|f(y)|<1$ for $y\in X-K$. K is a **generalized peak set** if it is the intersection of peak sets. A point $x\in X$ is a **(generalized) peak point** if the set $\{x\}$ is a (generalized) peak set. The set of generalized peak points equals the Choquet boundary for A. If X is metrizable, the peak sets and generalized peak sets coincide. There exist X and A such that X is metrizable: $X=\mathfrak{M}(A)=c(A)$, but $A\neq C(X)$ (B. Cole). A subset E of X is **interpolating** for A if for any bounded continuous function u on E there exists $f\in A$ with $f|E=u$. Then a closed G_δ set K in X is an interpolating peak set if and only if $\mu\in A^\perp$ implies $|\mu|(K)=0$ (E. Bishop).

E. Antisymmetric Decomposition [1–3]

A subset F of X is a **set of antisymmetry** for A if every function in A which is real-valued on F is constant on F. Bishop's antisymmetric decomposition then appears as an extension of the †Weierstrass-Stone theorem on uniform approximation. It is a refinement of Shilov's decomposition and reads as follows: If $\{E_\alpha\}$ is the family of maximal sets of antisymmetry for A, it is a partition of X into generalized peak sets such that $f\in C(X)$ with $f|E_\alpha\in A|E_\alpha$ for all α belongs to A. An interesting connection was found by J. Tomiyama between the maximal antisymmetric decomposition of X relative to A and that of $\mathfrak{M}(A)$ relative to $\hat{A}$.

F. Parts and Analytic Structure [1–4]

A. M. Gleason defined an equivalence relation $\sim$ in $\mathfrak{M}(A)$ by setting $\varphi\sim\psi$ if $\sup\{|\hat{f}(\varphi)-\hat{f}(\psi)| \mid f\in A, \|f\|<1\}<2$. Each equivalence class for this relation is a **part** (or a **Gleason**

part) for A. Two points φ, ψ belong to the same part if and only if there exist mutually absolutely continuous representing measures μ for φ and ν for ψ such that $c^{-1} < d\mu/d\nu < c$ for some constant $c > 0$ (Bishop). An **analytic structure** in $\mathfrak{M}(A)$ is a pair (V, τ), with an †analytic set V in some open subset of $\mathbf{C}^n$ and a nonconstant continuous mapping $\tau: V \rightarrow \mathfrak{M}(A)$ such that $\hat{f} \circ \tau$ is analytic on V for all $f \in A$. For such a structure, $\tau(V)$ is always within a part. But there can be a nontrivial part with no analytic structure, as was shown by G. Stolzenberg. A topological characterization of parts was obtained by J. Garnett. Sample results in the positive direction are: (1) If the ideal $A_\varphi = \{f \in A \mid \hat{f}(\varphi) = 0\}$ is finitely generated, there is an analytic structure (V, τ) such that τ is a homeomorphism of V onto an open neighborhood of φ (Gleason); (2) if φ has a unique representing measure and if the part P of φ is not a singleton, there is a bijective continuous mapping τ of the open unit disk onto P such that $\hat{f} \circ \tau$ is analytic for $f \in A$ (J. Wermer, K. Hoffman and G. Lumer); (3) if A is hypo-Dirichlet and if a part P is not a singleton, P can be made into a 1-dimensional †analytic space so that each $f \in A$ is analytic on P (J. Wermer and B. V. O'Neill). Analytic structures in †polynomially convex hulls of curves in $\mathbf{C}^n$ have also been studied [3, 4].

G. Abstract Function Theory [1, 6–9]

Let $\varphi \in \mathfrak{M}(A)$, and choose $m \in M_\varphi$, which is fixed. The **generalized Hardy class** $H_p(m)$, $0 < p \leqslant \infty$, associated with A is the closure (weak* closure, if $p = \infty$) of A in the †L_p space $L_p(m)$ on the measure space (X, m). Under suitable restrictions on A, φ, or m, we can recapture some of the important classical facts, most of which have their origins in the works of A. Beurling, R. Nevanlinna, F. and M. Riesz, and G. Szegö. In this area, H. Helson and D. Lowdenslager came up with a powerful method using orthogonal projections in Hilbert space and gave together with S. Bochner's remark, a strong influence for subsequent development. The modification argument was then devised by Hoffman and Wermer, inspired by F. Forelli. After Hoffman's detailed study of logmodular algebras, Lumer observed that most results remain valid when φ has a unique representing measure. T. P. Srinivasan and J.-K. Wang (→ [8]) introduced the notion of weak* Dirichlet algebra and showed that some major theorems are mutually equivalent and are in fact measure-theoretic. With the Hoffman-Rossi complement, their result now states that for fixed $m \in M_\varphi$ the following are equivalent: (i) $A + \bar{A}$ is weak* dense in $L_\infty(m)$, i.e., A is a **weak* Dirichlet algebra** on (X, m); (ii) if $\mu \in M_\varphi$ is absolutely continuous with respect to m, then $\mu = m$; (iii) if a closed subspace M of $L_2(m)$ is simply invariant in the sense that $A_\varphi M \subseteq M$ and $A_\varphi M$ is not dense in M, then $M = qH_2(m)$ with $q \in L_\infty(m)$, $|q| = 1$ a.e.; (iv) the set $\log |H_\infty(m)^{-1}|$ coincides with $L_\infty(m; \mathbf{R})$, the set of real-valued elements in $L_\infty(m)$; (v) for every $w \in L_1(m)$, $w \geqslant 0$, we have $\inf\{\int |1 - f|^2 w\, dm \mid f \in A_\varphi\} = \exp(\int \log w\, dm)$; (vi) the linear functional $h \rightarrow \int h\, dm$ on $H_\infty(m)$ has a unique positive extension to $L_\infty(m)$. Further extensions were subsequently made by use of the conjugation operator (→ Section K) by T. Gamelin, H. König, Lumer, K. Yabuta, and others. Some more properties of weak* Dirichlet algebras have been obtained by T. Nakazi.

H. Generalized Analytic Functions [1, 10]

Let Γ be a dense subgroup of the additive group $\mathbf{R}$ of the reals with discrete topology, and let G be the †character group of Γ. Each $a \in \Gamma$, as a character of G, defines a continuous function χ_a on G. Let A be the closed subalgebra of $C(G)$ generated by $\{\chi_a \mid a \in \Gamma, a \geqslant 0\}$. A is a Dirichlet algebra but is far more difficult to describe than the disk algebra. The study of this algebra, especially that of invariant subspaces, has evolved from papers by Helson and D. Lowdenslager. Let σ be the †Haar measure of G and $H_2(\sigma)$ the closure of A in $L_2(\sigma)$. A closed subspace M of $L_2(\sigma)$ is called **invariant** if $\chi_a M \subseteq M$ for all $a \in \Gamma$, $a \geqslant 0$. M is called **doubly invariant** if $\chi_a M \subseteq M$ for all $a \in \Gamma$. Otherwise, it is called **simply invariant**. In fact, only the latter is interesting. Let $e_t \in G$ with $t \in \mathbf{R}$ be the character of Γ defined by $e_t(a) = e^{ita}$. Then the mapping $t \rightarrow e_t$ is a faithful representation of $\mathbf{R}$ into G. A **cocycle** on G is defined to be a Borel function B on $G \times R$ such that (i) $|B(x, t)| = 1$, (ii) $B(x + e_t, t) = \overline{B(x, s)} B(x, s + t)$ for $x \in G$ and $s, t \in \mathbf{R}$. Two cocycles are identified if they differ only on a null set in $G \times \mathbf{R}$. For a cocycle B, let M_B be the set of $f \in L_2(\sigma)$ such that $\overline{B(x, t)} f(x + e_t) \in H_2(dt/(1 + t^2))$ for almost all $x \in G$, where $H_2(dt/(1 + t^2))$ is the closure, in the space $L_2(dt/(1 + t^2))$ on $\mathbf{R}$, of the set of boundary value functions on $\mathbf{R}$ of bounded analytic functions on the upper half-plane. Then the mapping $B \rightarrow M_B$ is a bijection from the set of cocycles onto the set of simply invariant subspaces M of $L_2(\sigma)$ such that $M = \bigcap \{\chi_a M \mid a \in \Gamma, a < 0\}$. Moreover, $M_B = qH_2(\sigma)$ for some $q \in L_\infty(\sigma)$, $|q| = 1$ a.e. if and only if $B(x, t) = \overline{q(x)} q(x + e_t)$, i.e., B is a **coboundary**. When $\Gamma \neq \mathbf{R}$, there is a cocycle that is not a coboundary. Further studies have been done by Helson,

Gamelin, J. Tanaka, and others. On the other hand, F. Forelli observed that †flows in compact spaces give rise to a kind of analyticity. In the above case, the algebra A consists of all $f \in C(G)$ that are analytic with respect to the flow $T_t(x) = x + e_t$ for $x \in G$ and $t \in \mathbf{R}$. Generalized analytic functions induced by flows in general have been studied by Forelli and P. S. Muhly.

I. The Unit Disk [1, 6, 7, 11, 12]

Let A be the disk algebra $P(\mathbf{T})$, $\mathbf{D}$ the open unit disk $\{|z|<1\}$, and m the normalized Lebesgue measure on $\mathbf{T}$. The algebra A has been the most important model in the theory of function algebras, and we find here the origin of many abstract results. Some typical results: (i) every orthogonal measure for A is absolutely continuous with respect to m (F. and M. Riesz); (ii) a closed set E in $\mathbf{T}$ is an interpolating peak set for the †Gel'fand transform $\hat{A}$ of A if and only if $m(E)=0$ (W. Rudin and L. Carleson); (iii) A is maximal among the closed subalgebras of $C(\mathbf{T})$ (Wermer); (iv) a function $f \in C(\mathrm{Cl}\,\mathbf{D})$ belongs to $\hat{A}$ if f is analytic at every $z \in \mathbf{D}$ with $f(z) \neq 0$ (T. Radó).

The generalized Hardy class $H_p(m)$, $0 < p \leqslant \infty$, associated with A is viewed as the set of nontangential boundary value functions of elements in the classical †Hardy class $H_p(\mathbf{D})$. Here we find the origin of invariant subspace theorems: A closed subspace M $(\neq \{0\})$ of $H_2(m)$ is invariant, i.e., $AM \subseteq M$, if and only if $M = qH_2(m)$ with $q \in H_\infty(m)$, $|q|=1$ a.e. (Beurling).

The algebra $H_\infty = H_\infty(m)$ is a weak* Dirichlet algebra, whose Shilov boundary is identified with the maximal ideal space X of $L_\infty(m)$. We have $L_\infty(m; \mathbf{R}) = \log|(H^\infty)^{-1}|$, and *a fortiori* H_∞ is logmodular on X. The mapping $z \to \varphi_z$ embeds the disk $\mathbf{D}$ in $\mathfrak{M}(H_\infty)$ as an open set. The structure of $\mathfrak{M}(H_\infty)$ was studied in detail by I. J. Schark, Hoffman, and others. We finish with three remarkable results: (i) $\mathbf{D}$ is dense in $\mathfrak{M}(H_\infty)$ (Carleson). This is the **corona theorem** and was proved in the following equivalent form: For any $f_1, \ldots, f_n \in H_\infty(\mathbf{D})$ with $|f_1| + \ldots + |f_n| \geqslant \varepsilon > 0$ on $\mathbf{D}$, there exist $g_1, \ldots, g_n \in H_\infty(\mathbf{D})$ with $f_1 g_1 + \ldots + f_n g_n = 1$. A simple proof was discovered by T. Wolff [12]. (ii) The convex combinations of †Blaschke products are uniformly dense in the unit ball of $H_\infty(\mathbf{D})$ (D. Marshall) [12]. (iii) Every closed subalgebra B between H_∞ and $L_\infty(m)$ is a **Douglas algebra**, i.e., B is generated by H_∞ and the complex conjugates of a family of inner functions (S.-Y. Chang and Marshall) [11]. The proof of (iii) is an interesting application of the theory of †bounded mean oscillation.

J. Rational Approximation [1, 2]

The problem of rational (polynomial) approximation on a compact plane set K asks when $A(K) = R(K)$ $(A(K) = P(K))$ holds. An important tool for such problems is **Cauchy's transform** of measure $\mu: \hat{\mu}(\zeta) = \int (z-\zeta)^{-1} d\mu(z)$. Using this, one can show, for instance, the following: (1) $A(K) = P(K)$ if and only if K has a connected complement (S. N. Mergelyan); (2) $f \in C(K)$ belongs to $R(K)$ if each point $z \in K$ has a closed neighborhood V with $f|_{K \cap V} \in R(K \cap V)$ (Bishop); (3) $A(K) = R(K)$ if the diameters of the components of $\mathbf{C} - K$ are bounded away from zero (Mergelyan); (4) $C(K) = R(K)$ if and only if almost all points of K are peak points for $R(K)$ (Bishop). The last result cannot be extended much because there is a set K such that $A(K) \neq R(K)$, while $A(K)$ and $R(K)$ have the same peak points (A. M. Davie).

A complete characterization for $A(K) = R(K)$ was obtained by A. G. Vitushkin: (5) The following are equivalent: (i) $A(K) = R(K)$; (ii) for any bounded open set D in $\mathbf{C}$, $\alpha(D-K) = \alpha(D-K^0)$; (iii) for any $z \in \mathrm{bd}\, K$, there exists $r \geqslant 1$ such that $\limsup_{\delta \to 0} \alpha(\Delta(z;\delta) - K^0)/\alpha(\Delta(z; r\delta) - K) < +\infty$, where $\Delta(z;\delta)$ is the disk $\{w \in \mathbf{C} \mid |w-z| \leqslant \delta\}$ and $\alpha(E)$, for any bounded set E in $\mathbf{C}$, is the **continuous analytic capacity** of E, which is the supremum of $|f'(\infty)|$ for all continuous functions f on the Riemann sphere $\mathbf{C} \cup \{\infty\}$ such that $|f| \leqslant 1$, $f(\infty) = 0$, and f is analytic off a compact subset of E. As for uniform or asymptotic approximation on noncompact closed subsets in $\mathbf{C}$, Carleman's classical study has recently been extended by N. U. Arakelyan, A. A. Nersesyan, A. Stray, and others in an interesting way.

In connection with rational approximation, we should note detailed studies on pointwise bounded approximation in $H_\infty(U)$, U being a bounded open set in $\mathbf{C}$, by O. J. Farrell, L. Rubel and A. Shields, Gamelin and J. Garnett, and Davie.

K. Further Topics

(1) **Conjugation operator** [9, 13]. Take any $m \in M_\varphi$, $\varphi \in \mathfrak{M}(A)$. The **conjugation operator** then associates with each $u \in \mathrm{Re}\, A$ the unique element $^*u$ in $C_R(X)$ such that $u + i\,{}^*u \in A$ and $\int {}^*u\, dm = 0$. After the classical inequalities of Kolmogorov and M. Riesz, we consider the following conditions with constants c_p and d_p: (K) $(\int |{}^*u|^p dm)^{1/p} \leqslant c_p \int |u|\, dm$ for $0 < p < 1$; (M) $(\int |{}^*u|^p dm)^{1/p} \leqslant d_p (\int |u|^p dm)^{1/p}$ for $1 < p < \infty$. Let m be arbitrary. Then the inequality (K) is valid for $u \in \mathrm{Re}\, A$, $u > 0$. The inequality (M) is valid for all $u \in \mathrm{Re}\, A$ if p is an even integer $\geqslant 2$; it is

valid for all $u \in \operatorname{Re} A$, $u>0$, if p is not an odd integer. All remaining cases have counterexamples. On the other hand, M_φ always contains an m such that $\log|f(\varphi)| \leqslant \int \log|f|\, dm$ for $f \in A$ (Bishop). Such a representing measure is called a **Jensen measure** for φ. If m is Jensen, (K) is valid for all $u \in \operatorname{Re} A$ and all $0<p<1$, and (M) is valid for all $u \in \operatorname{Re} A$ and all $1<p<\infty$.

(2) **Riemann surfaces**. For a compact bordered Riemann surface R, let $A(R)$ be the algebra of functions, continuous on $\operatorname{Cl} R$ and analytic on R. Then $A(R)$ is hypo-Dirichlet on the border $\operatorname{bd} R$ of R, and many results for the disk algebra are extended to $A(R)$. Most of the basic results are described in [14]. The maximality of $A(R)$ in $C(\operatorname{bd} R)$ was obtained by H. Royden; extreme points of related Hardy classes were discussed by Gamelin and M. Voichick; and invariant subspaces were determined by Forelli, M. Hasumi, D. Sarason, and Voichick. A further extension to infinitely connected surfaces has been obtained by C. W. Neville, Hasumi, and M. Hayashi in the case of open Riemann surfaces R of Parreau-Widom type, which is defined as follows: Let $G(a,z)$ be †Green's function for R, and let $B(a,\alpha)$, $\alpha>0$, be the first †Betti number of the domain $\{z \in R \mid G(a,z)>\alpha\}$; then R is of **Parreau-Widom type** if $\int B(a,\alpha)\, d\alpha < +\infty$. For such surfaces the situation looks favorable: For instance, the Cauchy-Read theorem is valid, and the Brelot-Choquet problem concerning Green's lines is solved affirmatively.

As for the generalization of approximation theorems of Mergelyan and Arakelyan, we refer to the work of Bishop and L. K. Kodama for compact sets and to that of S. Scheinberg for noncompact closed sets.

(3) **Higher-dimensional sets**. Much attention has been paid to algebras of analytic functions on domains in $\mathbf{C}^n$, $n \geqslant 2$, e.g., polydisks, unit balls, and general pseudoconvex domains. Polydisk algebras and ball algebras have been studied extensively by W. Rudin, P. Ahern, Forelli, and many others [15, 16]. Approximation theorems of Mergelyan type were obtained by G. M. Henkin, N. Kerzman, and I. Lieb for strictly pseudoconvex domains with smooth boundary and by L. Hörmander and Wermer and L. Nirenberg and R. O. Wells, Jr., in the case of totally real manifolds. Further improvements have been obtained by R. M. Range, A. Sakai, and others.

References

[1] T. W. Gamelin, Uniform algebras, Prentice-Hall, 1969.

[2] A. Browder, Introduction to function algebras, Benjamin, 1969.

[3] E. L. Stout, The theory of uniform algebras, Bogden and Quigley, 1971.

[4] J. Wermer, Banach algebras and several complex variables, Markham, 1971.

[5] T. W. Gamelin, Lectures on $H^\infty(D)$, Universidad Nacional de La Plata, 1972.

[6] K. Hoffman, Banach spaces of analytic functions, Prentice-Hall, 1962.

[7] H. Helson, Lectures on invariant subspaces, Academic Press, 1964.

[8] F. Birtel (ed.), Function algebras (Proc. Int. Symposium on Function Algebras, Tulane Univ., 1965), Scott, Foresman, 1966.

[9] K. Barbey and H. König, Abstract analytic function theory and Hardy algebras, Lecture notes in math. 593, Springer, 1977.

[10] H. Helson, Analyticity on compact Abelian groups, Algebras in Analysis, J. H. Williamson (ed.), Academic Press, 1975, 1–62.

[11] D. Sarason, Function theory on the unit circle, Virginia Polytech. Inst. and State Univ., 1978.

[12] P. Koosis, Introduction to H_p spaces, Cambridge Univ. Press, 1980.

[13] T. W. Gamelin, Uniform algebras and Jensen measures, Cambridge Univ. Press, 1978.

[14] M. Heins, Hardy classes on Riemann surfaces, Lecture notes in math. 98, Springer, 1969.

[15] W. Rudin, Function theory in polydiscs, Benjamin, 1969.

[16] W. Rudin, Function theory in the unit ball of $\mathbf{C}^n$, Springer, 1980.

165 (II.4) Functions

A. History

Leibniz used the term *function* (Lat. *functio*) in the 1670s to refer to certain line segments whose lengths depend on lines related to curves. Soon the term was used to refer to dependent quantities or expressions. In 1718, Johann Bernoulli used the notation φx, and by 1734 the modern functional notation $f(x)$ had been used by Clairaut and by Euler, who defined functions as analytic formulas constructed from variables and constants (1728) [1]. †Cauchy stated (1821) [2]: "When there is a relation among many variables, which determines along with values of one of them the values of the others, we usually consider the others as expressed by the one. We then call

the one an 'independent variable,' and the others 'dependent variables.'" †Dirichlet considered a function of $x \in [a, b]$ in his paper (1837) [3] concerning representations of "completely arbitrary functions" and stated that there was no need for the relation between y and x to be given by the same law throughout an interval, nor was it necessary that the relation be given by mathematical formulas. A function was simply a correspondence in which values of one variable determined values of another.

B. Functions

Today, the word "function" is used generally in mathematics in the same sense as a †mapping (→ 381 Sets C) or, which is the same thing, a †univalent correspondence (→ 358 Relations B). But this word is sometimes used in a wider sense, to mean a general (not necessarily univalent) correspondence, called a **many-valued** (or **multivalued**) **function**; in that case a univalent correspondence is called a **single-valued function**.

Specialists in each branch of mathematics have their respective ways of using the word. In analysis, values of a function are often considered real or complex numbers; such functions are called **real-valued functions** or **complex-valued functions**, respectively. Furthermore, if the domain of the function is also a set of real or complex numbers, then it is called a **real function** or a **complex function**, respectively (→ 131 Elementary Functions; 84 Continuous Functions; 198 Holomorphic Functions). If the domain of a real- or complex-valued function is contained in a †function space, the function is often called a **functional**; the †distribution is an example. In algebra we often fix a †field, †ring, etc., and consider functions whose domains and ranges are in such algebraic systems. Special names are given to functions having special properties, which can be defined according to the structures of the domain and the range. For example, when both domain and range of a function f are sets of real numbers, f is called an **even function** if $f(t)=f(-t)$, and an **odd function** if $f(t)=-f(-t)$. A function f that preserves the order relation between real numbers, i.e., such that $t_1<t_2$ implies $f(t_1)\leqslant f(t_2)$, is called a †monotone increasing function.

A mapping from a set I to a set F of functions, $\varphi: I \rightarrow F$, is called a **family of functions indexed by** I (or simple **family of functions**), and is denoted, using the form f_λ instead of $\varphi(\lambda)$, by $\{f_\lambda\}_{\lambda \in I}$ or $\{f_\lambda\}(\lambda \in I)$. In particular, if I is the set of natural numbers, the family is called a **sequence of functions**.

C. Variables

A letter x, for which we can substitute a name of an element of a set X, is called a **variable**, and X is called the **domain** of the variable. An element of the domain of a variable x is called a **value** of x. In particular, if the domain is a set of real numbers or complex numbers, the variable is called a **real variable** or a **complex variable**, respectively. On the other hand, a letter that stands for a particular element is called a **constant**.

When the domain and range of a function f are X and Y, respectively, a variable x whose domain is X is called the **independent variable**, and a variable y whose domain is Y is called the **dependent variable**. Then we say y is a function of x, and write $y=f(x)$. When a concrete method is given by which we make a value of y correspond to each value of x, we say that y is an **explicit function** of x. When a function is determined only by a †binary relation such as $R(x, y)=0$, we say that y is an **implicit function** of x (→ 208 Implicit Functions).

Given functions f, g with an independent variable t, suppose that y is regarded as a function of x defined by relations $x=f(t)$, $y=g(t)$. Then we say that y is a function of x with the variable t as a **parameter**. A function whose range is a given set C with variable t as its independent variable is often called a **parametric representation** of C by t.

If the domain of a function f is contained in a Cartesian product set $X_1 \times X_2 \times \ldots \times X_n$, the independent variable is denoted by $(x_1, x_2, \ldots, x_n)$, and f is often called a **function of n variables** or a **function of many variables** (when $n \geqslant 2$).

D. Families and Sequences

A function whose domain is a set I, $\varphi: I \rightarrow X$, is called a **family indexed by** I (or simply **family**), and I is called the **index set**. In the case $\varphi(\lambda)=x_\lambda(\lambda \in I)$, the family is denoted by $\{x_\lambda\}_{\lambda \in I}$ or $\{x_\lambda\}(\lambda \in I)$. If the range X of a function φ is a set of points, a set of functions, a set of mappings, or a set of sets, then the family $\{x_\lambda\}_{\lambda \in I}$ is called a **family of points**, a **family of functions**, a **family of mappings**, or a **family of sets**, respectively. If the set I is a †directed set, the family is called a **directed family**. Generally, if J is a subset of I, the family $\{x_\lambda\}_{\lambda \in J}$ is called a **subfamily** of $\{x_\lambda\}_{\lambda \in I}$. In particular, if I is a finite or infinite set of natural numbers, the family indexed by I is called a **finite sequence** or **infinite sequence**, respectively. **Sequence** is a generic name for both, but in many cases it means an infinite sequence, and usually we

have $I=\mathbf{N}$. Then the value corresponding to $n \in \mathbf{N}$ is called the **nth term** or, generally, a **term**. For convenience, the 0th term is often used as well. If each term of a sequence is a number, a point, a function, or a set, the sequence is called a **sequence of numbers**, a **sequence of points**, a **sequence of functions**, or a **sequence of sets**, respectively. A sequence is usually denoted by $\{a_n\}$. If it is necessary to show the domain of n explicitly, the sequence is denoted by $\{a_n\}_{n\in I}$. If J is a subset of I, a sequence $\{a_n\}_{n\in J}$ is called a **subsequence** of the sequence $\{a_n\}_{n\in I}$. And if $I=\mathbf{N}$, the composite $\{a_{k_n}\}$ *of* $\{a_n\}$ and a sequence $\{k_n\}$ of natural numbers with $k_1<k_2<k_3\ldots$ is usually called a subsequence of $\{a_n\}$.

References

[1] L. Euler, Opera omnia, ser. I, Opera mathematica VIII: Introductio in analysin infinitorum I, Teubner, 1922.
[2] A. L. Cauchy, Cours d'analyse de l'Ecole Royale Polytechnique pt. 1, 1821 (Oeuvres ser. 2, III, Gauthier-Villars, 1897).
[3] G. P. L. Dirichlet, Über die Darstellung ganz willkürlicher Funktion durch Sinus- und Cosinusreihen, Repertorium der Physik, Bd. 1 (1837), 152–174 (Werke, Bd. 1, G. Reimer, 1889, I, p. 133–160).

166 (X.5) Functions of Bounded Variation

A. Monotone Functions

A function (or mapping) f from an †ordered set X to another ordered set Y is called a **monotone increasing (monotone decreasing) function** if

$$x_1<x_2 \quad \text{implies} \quad f(x_1)\leqslant f(x_2)$$

$$(x_1<x_2 \quad \text{implies} \quad f(x_1)\geqslant f(x_2)). \qquad (1)$$

A monotone increasing (decreasing) function is also called a **nondecreasing (nonincreasing)** function. In either case, the function f is called simply a **monotone function**. If X and Y are †totally ordered sets and the inequality $<$ ($>$) holds in (1) instead of $\leqslant$ ($\geqslant$), then f is called a **strictly (monotone) increasing (strictly (monotone) decreasing)** function. In either case, f is called simply a **strictly monotone function**.

In particular, when X and Y are subsets of the real line $\mathbf{R}$, a monotone function is continuous except for at most a countable number of points. Hence it is †Riemann integrable in a finite interval provided that it is bounded. A continuous real function $f(x)$ defined on an interval in $\mathbf{R}$ is †injective if and only if it is strictly monotone. In such a case, the range of the function $f(x)$ is also an interval, and the inverse function is also strictly monotone. Furthermore, a differentiable real function f defined on an interval is monotone if and only if its derivative f' is always $\geqslant 0$ (monotone increasing) or always $\leqslant 0$ (monotone decreasing). If $f'>0$ (<0), f is strictly monotone increasing (decreasing).

B. Functions of Bounded Variation

Let $f(x)$ be a real function defined on a closed interval $[a,b]$ in $\mathbf{R}$. Given a subdivision of the interval $a=x_0<x_1<x_2<\ldots<x_n=b$, we denote the sum of positive differences $f(x_i)-f(x_{i-1})$ by P and the sum of negative differences $f(x_i)-f(x_{i-1})$ by $-N$. Then we easily obtain

$$P-N=f(b)-f(a),$$

$$P+N=\sum_i |f(x_i)-f(x_{i-1})|.$$

The suprema of P, N, and $P+N$ for all possible subdivisions of $[a,b]$ are called the **positive variation**, the **negative variation**, and the **total variation** of the function $f(x)$ in the interval $[a,b]$, respectively. If any of these three values is finite, then all three values are finite. In such a case, the function $f(x)$ is called a **function of bounded variation**. Every function of bounded variation is bounded, but the converse is not true. The positive and negative variations $\pi(t)$, $\nu(t)$ of the function $f(x)$ in the interval $[a,t]$ are monotone increasing functions with respect to t, and we have

$$f(x)-f(a)=\pi(x)-\nu(x) \qquad (2)$$

if $f(x)$ is a function of bounded variation. Hence every function of bounded variation has both left and right limits at every point. A monotone function is a function of bounded variation, and the sum, the difference, or the product of two functions of bounded variation is also a function of bounded variation. Hence $f(x)$ is a function of bounded variation if and only if it is the difference of two monotone functions. The representation (2) (representing a function of bounded variation as the difference of two monotone increasing functions) is called the **Jordan decomposition** of the function $f(x)$. A function of bounded variation is Riemann integrable, continuous except for at

most a countable number of points, and differentiable †almost everywhere.

A continuous function defined on a closed interval is bounded but not necessarily of bounded variation (e.g., $f(x)=x\sin(1/x)$ ($x\in(0,1]$), 0 ($x=0$)). A discontinuous function may be a function of bounded variation on a closed interval (e.g., $\operatorname{sgn}(x)$). However, an †absolutely continuous function, a differentiable function with bounded derivative, or a function satisfying the †Lipschitz condition is a function of bounded variation on a closed interval.

The notion of functions of bounded variation was introduced by C. Jordan in connection with the notion of the length of curves (→ 246 Length and Area).

C. Lebesgue-Stieltjes Integral

Let $f(x)$ be a right continuous function of bounded variation on a closed interval $[a,b]$, and $f(x)=\pi(x)-\nu(x)$ the Jordan decomposition of $f(x)$. Then $\pi(x)$ and $\nu(x)$ are monotone increasing right continuous functions and hence define bounded measures $d\pi(x)$ and $d\nu(x)$ on $[a,b]$, respectively (→ 270 Measure Theory L (v)). The difference $d\pi-d\nu$ of these two measures is a †completely additive set function on $[a,b]$ which is often called the **(signed) Lebesgue-Stieltjes measure** induced by f, written df. For every function g integrable with respect to the measure $d\mu=d\pi+d\nu$, we define $\int_E g\,df$ to be equal to $\int_E g\,d\pi-\int_E g\,d\nu$ and call it the **Lebesgue-Stieltjes integral**. In this case, $h(x)=\int_{[a,b]}g\,df$ is of bounded variation on $[a,b]$ and the Lebesgue-Stieltjes measure $dh(x)$ is denoted by $g(x)df(x)$. If f_1 and f_2 are of bounded variation on $[a,b]$, then so is the product $f=f_1f_2$, and we have

$$df(x)=f_1(x\pm 0)df_2(x)+f_2(x\mp 0)df_1(x). \qquad (3)$$

The Lebesgue-Stieltjes integral for a continuous integrand is often called the **Riemann-Stieltjes integral**, because it can be defined in an elementary way similar to the definition of the Riemann integral.

The notion of bounded variation can also be defined for interval functions on $\mathbf{R}^n$ and set functions on an abstract space (→ 380 Set Functions).

References

[1] H. L. Royden, Real analysis, Macmillan, second edition, 1963.
[2] W. Rudin, Real and complex analysis, McGraw-Hill, second edition, 1974.

167 (XIV.7) Functions of Confluent Type

A. Confluent Hypergeometric Functions

If some singularities of an ordinary differential equation of †Fuchsian type are confluent to each other, we obtain a **confluent differential equation** whose solutions are called **functions of confluent type**. The equations that appear frequently in practical problems are the **confluent hypergeometric differential equations**

$$z\frac{d^2w}{dz^2}+(\gamma-z)\frac{dw}{dz}-\alpha w=0 \qquad (1)$$

and related equations. Equation (1) corresponds to the †hypergeometric differential equation for which a †regular singular point coincides with the point at infinity and is an †irregular singular point of class 1. For (1), $z=0$ is a regular singular point, and a series solution (radius of convergence ∞) is given by

$$F(\alpha,\gamma;z)={}_1F_1(\alpha,\gamma;z)$$
$$=\sum_{n=0}^{\infty}\frac{\alpha(\alpha+1)\dots(\alpha+n-1)}{n!\gamma(\gamma+1)\dots(\gamma+n-1)}z^n \qquad (|z|<\infty), \qquad (2)$$

where γ is not a nonpositive integer. The function ${}_1F_1$ in (2) is a †generalized hypergeometric function due to Barnes and is called a **hypergeometric function of confluent type** or **Kummer function**. If γ is not equal to a positive integer, the other solution of (1) independent of (2) is given by $z^{1-\gamma}F(1+\alpha-\gamma,2-\gamma;z)$ (→ Appendix A, Table 19.I).

B. Whittaker Functions

Equation (1) with $w=e^{z/2}z^{-\gamma/2}W$, $\gamma-2\alpha=2k$, $\gamma^2-2\gamma=4m^2-1$ reduces to **Whittaker's differential equation**

$$\frac{d^2W}{dz^2}+\left(-\frac{1}{4}+\frac{k}{z}+\frac{\frac{1}{4}-m^2}{z^2}\right)W=0. \qquad (3)$$

If $2m$ is not equal to an integer, (3) has two series solutions for any finite z:

$$M_{k,m}(z)=z^{(1/2)+m}e^{-z/2}F(\tfrac{1}{2}+m-k,1+2m;z),$$

$$M_{k,-m}(z)=z^{(1/2)-m}e^{-z/2}F(\tfrac{1}{2}-m-k,1-2m;z).$$

If $2m$ is an integer, since the functions $M_{k,m}$ and $M_{k,-m}$ are linearly dependent, E. T. Whittaker considered a solution of the form

$$W_{k,m}(z)=-\frac{1}{2\pi i}\Gamma(k+\tfrac{1}{2}-m)z^ke^{-z/2}$$
$$\times\int_{\infty}^{(0+)}(-t)^{-k-(1/2)+m}\left(1+\frac{t}{z}\right)^{k-(1/2)+m}e^{-t}dt.$$

If $k-\frac{1}{2}-m$ is equal to a negative integer, this integral does not exist. The function

$$W_{k,m}(z)=\frac{z^k e^{-z/2}}{\Gamma(\frac{1}{2}-k+m)}\int_0^\infty t^{-k-(1/2)+m}\times\left(1+\frac{t}{z}\right)^{k-(1/2)+m}e^{-t}dt$$

for $\mathrm{Re}(k-\frac{1}{2}-m)\leqslant 0$ is defined for any m, k, and for any z except when z is a negative real number. We call $M_{k,m}$ and $W_{k,m}$ the **Whittaker functions**. †Bessel functions are particular cases of these functions, and the relation

$$J_n(z)=\frac{z^{-1/2}}{2^{2n+(1/2)}i^{n+(1/2)}\Gamma(n+1)}M_{0,n}(2iz)$$

is satisfied. In Whittaker's differential equation, since $W_{-k,m}(-z)$ is also a solution and $W_{k,m}(z)/W_{-k,m}(-z)$ is not equal to a constant, $W_{k,m}(z)$ and $W_{-k,m}(-z)$ can be considered a pair of fundamental solutions (→ Appendix A, Table 19.II).

C. Parabolic Cylinder Functions

Putting $x=(\xi^2-\eta^2)/2$ and $y=\xi\eta$, the curves corresponding to $\xi=$constant and to $\eta=$ constant, respectively, constitute families of orthogonal parabolas. The curvilinear coordinates (ξ,η,z) in three dimensions are called **parabolic cylindrical coordinates**. By using parabolic coordinates, separating variables in Laplace's equation into the form $f(\xi)g(\eta)e^{i\alpha z}$, and making a simple transformation, we find that f and g satisfy a differential equation of the form

$$\frac{d^2F}{dz}+(n+\tfrac{1}{2}-\tfrac{1}{4}z^2)F=0. \tag{4}$$

By means of the Whittaker function $W_{k,m}(z)$, a solution $D_n(z)$ of (4) is represented by

$$D_n(z)=2^{n/2+(1/4)}z^{-1/2}W_{n/2+(1/4),-1/4}(\tfrac{1}{2}z^2).$$

Equation (4) is called **Weber's differential equation** or the **Weber-Hermite differential equation**, and $D_n(z)$ the **Weber function**. Another solution of (4) is $D_{-n-1}(iz)$ or $D_{-n-1}(-iz)$. The solutions of (4) are called **parabolic cylinder functions**. In particular, if n is equal to a nonnegative integer, then

$$H_n(z)=2^{-n/2}\exp(\tfrac{1}{2}z^2)D_n(\sqrt{2}\,z)$$

is the †Hermite polynomial of degree n. Solutions of differential equations for harmonic oscillators in quantum mechanics are of this form.

In general, suppose that three regular singular points are confluent to the point at infinity, and that they are reduced to an irregular singular point of class 2. Suppose further that there are no other singularities. Then differential equations of order 2 with these conditions are transformed into the form (4), whose solutions are represented by parabolic cylinder functions. Differential equations of the form (4) are reduced to confluent hypergeometric differential equations if z^2 is chosen as an independent variable (→ Appendix A, Table 20.III).

D. Indefinite Integrals of Elementary Functions

Since exponential and trigonometric functions can be represented by particular types of Kummer functions, their indefinite integrals that cannot be represented by elementary functions, e.g., †incomplete Γ-functions and the **error function** $\mathrm{Erf}\,z=\int_0^z\exp(-t^2)dt$, can be represented by Kummer or Whittaker functions. They are included in a family of †special functions of confluent type. The functions defined by

$$C(z)=\frac{1}{\sqrt{2\pi}}\int_0^z\frac{\cos t}{\sqrt{t}}dt,$$

$$S(z)=\frac{1}{\sqrt{2\pi}}\int_0^z\frac{\sin t}{\sqrt{t}}dt$$

are called **Fresnel integrals**, which are also represented in terms of the Whittaker function as

$$C(z)-iS(z)=\frac{1-i}{2}\left(1-\frac{e^{-\pi i/8}}{\sqrt{\pi}}z^{1/4}e^{-z^2/2}W_{-1/4,1/4}(iz)\right).$$

Fresnel integrals first appeared in the theory of the diffraction of waves. More recently they have been applied to designing highways for high-speed automobiles. Furthermore, the functions

$$C(u)=\int_0^u\cos\left(\frac{\pi}{2}s^2\right)ds,$$

$$S(u)=\int_0^u\sin\left(\frac{\pi}{2}s^2\right)ds$$

(obtained by a change of variables $z=\pi u^2/2$) are also called Fresnel integrals. Numerical tables are available for them. The curves $x=C$ and $y=S$ with a parameter z or u are called **Cornu's spiral** (Fig. 1). The functions

$$\mathrm{Li}\,x=\int_0^x\frac{dt}{\log t},\qquad \mathrm{Ei}\,x=\int_{-\infty}^x\frac{e^t}{t}dt,$$

where a †principal value must be taken at $t=0$ if $x>0$,

$$\mathrm{Si}\,x=\int_0^x\frac{\sin t}{t}dt,\qquad\text{and}\qquad \mathrm{Ci}\,x=-\int_x^\infty\frac{\cos t}{t}dt$$

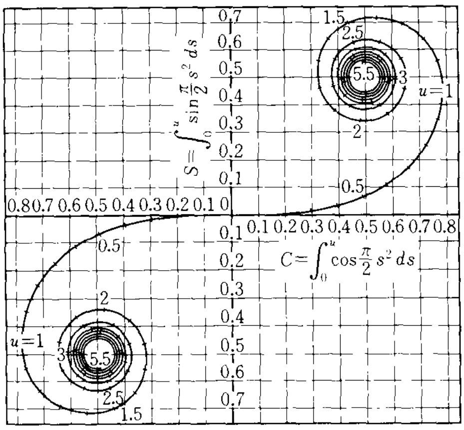


Fig. 1

are called the **logarithmic integral, exponential integral, sine integral,** and **cosine integral,** or **integral logarithm, integral exponent, integral sine,** and **integral cosine,** respectively. They satisfy the relations

$$\mathrm{Ei}\,x=\mathrm{Li}\,e^{x},\qquad \mathrm{Ei}\,ix=\mathrm{Ci}\,x+i\,\mathrm{Si}\,x+(\pi/2)i.$$

They have important applications: $\mathrm{Ei}\,x$ in quantum mechanics, $\mathrm{Si}\,x$ and $\mathrm{Ci}\,x$ in electrical engineering, and $\mathrm{Li}\,x$ in estimating the number of †primes less than x (→ 123 Distribution of Prime Numbers). $\mathrm{Li}\,x$ is also denoted by $\mathrm{li}\,x$ (→ Appendix A, Table 19.II).

E. Stokes's Equation

Consider a linear differential equation of the second order with five regular singular points including the point at infinity such that the difference of the characteristic indices at every singularity is equal to 1/2. Such equations are called **generalized Lamé's differential equations.** F. Klein and M. Bôcher have shown that every linear differential equation that is commonly treated in mathematical physics is represented by a confluent type of generalized Lamé's equation. Among these equations, if all five singularities are confluent to the point at infinity, the resulting equation is called **Stokes's differential equation,** which is applied to the investigation of diffraction. This is reduced to †Bessel's differential equation of order 1/3 by suitable transformations of the independent and dependent variables.

References

[1] H. Buchholz, The confluent hypergeometric function, with special emphasis on its applications, translated by H. Lichtblau and K. Wentzel, Springer, 1969. (Original in German, 1953.)

[2] L. J. Slater, Confluent hypergeometric functions, Cambridge Univ. Press, 1960.

For the logarithmic integral, etc.,

[3] N. Nielsen, Theorie der Integrallogarithmus und verwandter Transzendenten, Teubner, 1906.

[4] Mathematical tables I, British Assoc. Adv. Sci., London, 1931.

[5] Nat. Bur. Standards, Tables of sine, cosine and exponential integrals I, II, Washington, D. C., 1940.

[6] L. Lewin, Polylogarithms and associated functions, North-Holland, 1958, revised edition, 1981.

Also → references to 389 Special Functions.

168 (XII.6) Function Spaces

A. General Remarks

It is a general method in modern analysis to consider a set X of mappings of a space Ω into another space A as a space (→ 381 Sets) and its elements (namely, mappings of Ω into A) as points of the space X, and to investigate them as geometric objects. In particular, it is important to consider the case where Ω is a †topological space, a †measure space, or a †differentiable manifold and X is a set of real- or complex-valued functions defined on Ω and satisfying certain conditions, such as continuity, measurability, and differentiability. Such spaces are generally called **function spaces;** they usually form †topological linear spaces (→ 37 Banach Spaces; 197 Hilbert Spaces; 424 Topological Linear Spaces).

B. Examples of Function Spaces

The following are important examples of function spaces. Throughout this section, all functions are real- or complex-valued, and two functions on a measure space are identified whenever they are equal to each other †almost everywhere.

(1) The Function Spaces $C(\Omega)$, $C_\infty(\Omega)$, and $C_0(\Omega)$. The totality of continuous functions $f(x)$ defined on a compact †Hausdorff space Ω is denoted by $C(\Omega)$. Let $f+g$ and αf (α a real or complex number) be the functions $f(x)+g(x)$ and $\alpha f(x)$, respectively. Then $C(\Omega)$ forms a †linear space. Furthermore, define the norm of f by $\|f\|=\sup_{x\in\Omega}|f(x)|$. Then $C(\Omega)$ becomes a †Banach space since it is complete in (the

metric defined by) the norm. The norm is called the **supremum norm** or **uniform norm** because $\lim_{n\to\infty}\|f_n-f\|=0$ means that $f_n(x)$ converges to $f(x)$ uniformly on Ω as $n\to\infty$. Define $f\cdot g$ to be the function $f(x)g(x)$. Then clearly $\|f\cdot g\|\leqslant\|f\|\,\|g\|$. Hence $C(\Omega)$ is also a †Banach algebra.

Suppose that a subset R of $C(\Omega)$ satisfies the following three conditions: (i) R is an †algebra over the complex number field with respect to the addition and multiplication defined above and contains the function identically equal to one. (ii) For any two distinct points x and y of Ω, there exists a function $f\in R$ satisfying $f(x)\neq f(y)$. (iii) For any $f\in R$, there exists an $f^*\in R$ such that $f^*(x)=\overline{f(x)}$ on Ω. Then R is dense in $C(\Omega)$ with respect to the supremum norm (namely, in the sense of uniform convergence). This fact is known as the **Weierstrass-Stone theorem** (or **Stone-Gel'fand theorem**). A subset E of $C(\Omega)$ is †precompact (i.e., any sequence of functions in E contains a subsequence that converges uniformly on Ω) if and only if E is †uniformly bounded and †equicontinuous (**Ascoli-Arzelà theorem**). Since $C(\Omega)$ is not a †reflexive Banach space except for trivial cases (→ Section C), precompact sets and relatively compact sets are different in the †weak topology. For important characterizations of the latter sets → [2, 5].

When Ω is a topological space that is not necessarily compact, the totality of bounded continuous functions on Ω (denoted by $BC(\Omega)$) is also a Banach space with respect to the supremum norm $\|f\|=\sup_{x\in\Omega}|f(x)|$. Let Ω be a locally compact Hausdorff space. Then the space $C(\Omega)$ of all continuous functions on Ω is endowed with the topology of uniform convergence on the compact sets, i.e., the †locally convex topology defined by the †seminorms $\sup_{x\in K}|f(x)|$ as K ranges over the compact sets in Ω. $C(\Omega)$ is always a complete locally convex space. It is a †Fréchet space if Ω is σ-compact (i.e., Ω is the union of a countable family of compact sets). We denote by $C_\infty(\Omega)$ the subspace of all functions $f(x)\in C(\Omega)$ that converge to zero as x tends to infinity (i.e., given an $\varepsilon>0$, there is a compact set K such that $|f(x)|<\varepsilon$ for $x\notin K$). $C_\infty(\Omega)$ is a Banach space with the norm $\sup_{x\in\Omega}|f(x)|$. It can be regarded as a closed linear subspace of $BC(\Omega)$.

The totality of continuous functions with compact support is denoted by $C_0(\Omega)$ or $\mathscr{K}(\Omega)$, where the **support** (or **carrier**) of a function f is the †closure of the set $\{x\mid f(x)\neq 0\}$ in Ω and is usually denoted by supp$f$. If Ω is not compact, $C_0(\Omega)$ is not complete with respect to the supremum norm, but when Ω is σ-compact $C_0(\Omega)$ is complete with respect to the strongest locally convex topology with the property that for each compact set K in Ω the embedding in $C_0(\Omega)$ of the linear subspace $\mathscr{K}_K$ of all functions with support in K equipped with the supremum norm is continuous. Usually $\mathscr{K}(\Omega)$ denotes the space $C_0(\Omega)$ equipped with this topology.

(2) The Lebesgue Spaces $L_p(\Omega)$ ($0<p<\infty$). Let (Ω,μ) be a †measure space. We denote by $L_p(\Omega)$ the totality of †measurable functions $f(x)$ on Ω such that $|f(x)|^p$ is integrable. A function f defined on (Ω,μ) is called **square integrable** if $f\in L_2(\Omega,\mu)$. If Ω is the interval (a,b) equipped with the Lebesgue measure, it is sometimes denoted by $L_p(a,b)$. We define the norm $\|f\|_p$ by

$$\|f\|=\|f\|_p=\left(\int_\Omega |f(x)|^p\,d\mu(x)\right)^{1/p}.$$

If $1\leqslant p<\infty$, then $L_p(\Omega)$ is a Banach space. The triangle inequality for this norm is precisely the †Minkowski inequality. If $0<p<1$, the norm no longer satisfies the triangle inequality but does satisfy the quasinorm inequality $\|f+g\|_p\leqslant 2^{1/p-1}(\|f\|_p+\|g\|_p)$, and if $\sum\|f_n\|_p^p<\infty$, then $\sum f_n$ converges unconditionally in $L_p(\Omega)$. Hence $L_p(\Omega)$, $0<p<1$, is a †quasi-Banach space. If $\lim_{n\to\infty}\|f_n-f\|_p=0$, we say that the sequence $\{f_n\}$ **converges** to f **in the mean of order** p (or **in the mean of power** p), and write $\text{l.i.m.}_{n\to\infty}f_n=f$. If $\{f_n\}$ converges to f in the mean of order 2, we simply say that $\{f_n\}$ converges to f in the mean. (The notation l.i.m. means the **limit in the mean** and is used mostly when $p=2$.) For any $f,g\in L_2(\Omega)$, $(f,g)=\int_\Omega f(x)\overline{g(x)}\,d\mu(x)$ is well defined, by the †Schwarz inequality, and has the properties of the †inner product. Hence, $L_2(\Omega)$ is a †Hilbert space. If $1<p<\infty$, then $L_p(\Omega)$ is a †uniformly convex Banach space and is in particular †reflexive. Deepest results on $L_p(\mathbf{R}^n)$, $1<p<\infty$, are often derived from the **Littlewood-Palay theory** due to J. E. Littlewood and R. E. A. C. Paley, A. Zygmund [6], and E. M. Stein [7]. Its starting point is the inequality

$$A_p\|f\|_p\leqslant\|g(f)\|_p\leqslant A'_p\|f\|_p,$$

where $g(f)$ is the function

$$g(f)(x)=\left(\int_0^\infty |\text{grad}_x u(x,t)|^2 t\,dt\right)^{1/2}$$

obtained from the **Poisson integral**

$$u(x,t)$$
$$=\frac{\Gamma((n+1)/2)}{\pi^{(n+1)/2}}\int_{\mathbf{R}^n} t(t^2+|x-y|^2)^{-(n+1)/2}f(y)\,dy,$$

where $|x|=(x_1^2+\ldots+x_n^2)^{1/2}$.

The L_p spaces, $1<p<\infty$, are generalized in the following way. Let $\Phi(s)$ be a convex and nondecreasing function on $[0,\infty)$ satisfying $\Phi(0)=0$ and $\Phi(s)/s\to\infty$ as $s\to\infty$. Denote

by $L_\Phi(\Omega)$ ($L_\Phi^*(\Omega)$) the set of all functions $f(x)$ such that $\Phi(|f(x)|)$ is integrable ($\Phi(k|f(x)|)$ is integrable for some $k>0$). $L_\Phi(\Omega)=L_\Phi^*(\Omega)$ if $\Phi(2s)\leqslant C\Phi(s)$. $L_\Phi^*(\Omega)$ is a Banach space, called the **Orlicz space**, under the norm

$$\|f\|=\inf\left\{\lambda>0\,\middle|\,\int\Phi(\lambda^{-1}|f(x)|)\,d\mu(x)\leqslant 1\right\}.$$

$L_p(\Omega)$, $1<p<\infty$, is the Orlicz space for $\Phi(s)=s^p$.

(3) The Function Space $M(\Omega)=L_\infty(\Omega)$. Let Ω and μ be as in (2). A measurable function $f(x)$ on Ω is said to be **essentially bounded** if there exists a positive number α such that $|f(x)|\leqslant\alpha$ almost everywhere on Ω. The infimum of such α is called the **essential supremum** of f, denoted by $\operatorname{ess\,sup}_{x\in\Omega}|f(x)|$. The totality of essentially bounded measurable functions on Ω (denoted by $M(\Omega)$) is a Banach space with respect to the norm $\|f\|=\|f\|_\infty=\operatorname{ess\,sup}_{x\in\Omega}|f(x)|$. If $\mu(\Omega)<\infty$, then $M(\Omega)\subset L_p(\Omega)$ for any $p>0$, and $\|f\|_\infty=\lim_{p\to\infty}\|f\|_p$ for any $f\in M(\Omega)$. From this point of view, $M(\Omega)$ is also denoted by $L_\infty(\Omega)$ even when $\mu(\Omega)=\infty$. This is also the reason why the notation $\|\cdot\|_\infty$ is used for the norm in $M(\Omega)$.

(4) The Lorentz Spaces $L_{(p,q)}(\Omega)$ $(0<p,q\leqslant\infty)$. The Lebesgue spaces $L_p(\Omega)$, $0<p\leqslant\infty$, are **rearrangement invariant**. Namely, define for a measurable function f on a measure space (Ω,μ) the **distribution function** $\mu_f(s)=\mu\{x\in\Omega\,|\,|f(x)|>s\}$, $s>0$, and the **rearrangement** $f^*(t)=\inf\{s>0\,|\,\mu_f(s)\leqslant t\}$, $t>0$. Then $f\in L_p(\Omega)$ if and only if $f^*\in L_p(0,\infty)$ and $\|f\|_p=\|f^*\|_p$. Another important class of rearrangement invariant spaces are the **Lorentz spaces** $L_{(p,q)}(\Omega)$, $0<p$, $q\leqslant\infty$ (G. G. Lorentz, 1950; R. A. Hunt [8]), which is defined to be the quasi-Banach space of all measurable functions f on Ω such that

$$\|f\|_{(p,q)}=\|t^{1/p}f^*(t)\|_{L_q^*}<\infty,$$

where L_q^* is the L_q-space on $(0,\infty)$ relative to the measure dt/t. $L_{(p,p)}(\Omega)=L_p(\Omega)$ with equal norms. If $1<p<\infty$ and $1\leqslant q\leqslant\infty$, then $L_{(p,q)}(\Omega)$ is a Banach space under the equivalent norm $\|t^{1/p-1}\int_0^t f^*(s)\,ds\|_{L_q^*}$. Except for these cases, $L_{(p,q)}(\Omega)$ is not equivalent to a normed space in general [8]. If $q_0<q_1$, then $L_{(p,q_0)}(\Omega)\subset L_{(p,q_1)}(\Omega)$ with continuous embedding. In case $\mu(\Omega)<\infty$, $L_{(p_0,q_0)}(\Omega)\subset L_{(p_1,q_1)}(\Omega)$ for $p_0>p_1$ and any q_0, q_1. The Lorentz spaces play an important role in interpolation and approximation theory (→ 224 Interpolation of Operators).

(5) The Function Space $S(\Omega)$. Let (Ω,μ) be a measure space with $\mu(\Omega)<\infty$. Denote by $S(\Omega)$ the totality of measurable functions on Ω that take finite value almost everywhere. Then $\|f\|=\int_\Omega(|f(x)|/(1+|f(x)|))\,d\mu(x)$ for $f\in S(\Omega)$ has the properties of the †pseudonorm and $S(\Omega)$ is a †Fréchet space (in the sense of Banach) that is not locally convex in general. We have $\lim_{n\to\infty}\|f_n-f\|=0$ if and only if

$$\lim_{n\to\infty}\mu(\{x\,|\,|f_n(x)-f(x)|>\varepsilon\})=0$$

for any positive number ε. Convergence of this type is called **convergence in measure** (or **asymptotic convergence**) and is the same notion as †convergence in probability of a sequence of †random variables (→ 342 Probability Theory). If $\{f_n\}\in L_p(\Omega)$ converges to $f\in L_p(\Omega)$ in the mean of order p, then $\{f_n\}$ converges to f asymptotically, but the converse is not true in general. If a sequence $\{f_n\}\in S(\Omega)$ converges to $f\in S(\Omega)$ almost everywhere, then $\{f_n\}$ converges to f asymptotically. Any sequence $\{f_n\}$ that converges to f asymptotically contains a subsequence $\{f_{n_\nu}\}$ that converges to f almost everywhere.

(6) The Sequence Spaces c, c_0, l_p $(0<p<\infty)$, $m=l_\infty$, and s. The totality c (resp. c_0) of sequences $x=\{\xi_n\}$ that converges (resp. converges to zero) as $n\to\infty$ forms a Banach space with respect to the norm $\|x\|=\sup|\xi_n|$. c_0 (resp. c) is the space $C_\infty(\Omega)$ (resp. $C(\Omega)$) when Ω is (resp. the one-point compactification of) the discrete locally compact space $\{1,2,3,\dots\}$. The sequence space l_p, $0<p<\infty$ (resp. $m=l_\infty$), is defined to be the spaces $L_p(\Omega)$ (resp. $M(\Omega)$), where Ω is the space $\{1,2,3,\dots,n,\dots\}$, of which each point has unit mass, while s denotes the space $S(\Omega)$, where $\Omega=\{1,2,\dots,n,\dots\}$, provided with the measure assigning mass $1/2^n$ to the point n. s is the set of all sequences equipped with the topology of pointwise convergence (s is also used to denote the space of rapidly decreasing sequences; → Section (16)).

Assume that the space $L_2(\Omega)$ mentioned in (2) is †separable and that $\{\varphi_n\}$ is a †complete orthonormal set in $L_2(\Omega)$. Then putting

$$\xi_n=\int_\Omega f(x)\overline{\varphi_n(x)}\,d\mu(x)$$

(†Fourier coefficients) for any $f\in L_2(\Omega)$, we have $\{\xi_n\}\in l_2$ and $\sum_{n=1}^\infty|\xi_n|^2=\|f\|^2$. Conversely, for any $\{\xi_n\}\in l_2$ there exists an $f=\sum_{n=1}^\infty\xi_n\varphi_n\in L_2(\Omega)$ whose Fourier coefficients are the given ξ_n (**Riesz-Fischer theorem**). By means of this correspondence, separable spaces $L_2(\Omega)$ and l_2 are mutually isomorphic as Hilbert spaces. Sometimes we denote by $l_p(\Omega)$ the function space $L_p(\Omega)$, where Ω is an arbitrary set endowed with the measure assigning mass 1 to each point.

(7) The John-Nirenberg Space BMO. A locally integrable function $f(x)$ on $\mathbf{R}^n$ is said to be of **bounded mean oscillation** if

$$\|f\|_{BMO}=\sup |S|^{-1}\int_S |f(x)-f_S|\,dx<\infty, \qquad (1)$$

where the supremum is taken over all (solid) spheres S in $\mathbf{R}^n$, f_S is the mean $|S|^{-1}\int_S f(x)\,dx$, and $|S|$ denotes the measure of S (F. John and L. Nirenberg, 1961). The set $BMO(\mathbf{R}^n)$ of all functions on $\mathbf{R}^n$ of bounded mean oscillation forms a Banach space under the norm mentioned above if two functions f and g are identified whenever $f-g$ is equal to a constant almost everywhere. Condition (1) is equivalent to

$$\sup\left(|S|^{-1}\int_S |f(x)-f_S|^p\,dx\right)^{1/p}<\infty$$

for any $1\leqslant p<\infty$. There are constants B and $K>0$ such that

$$|\{x\in S \mid |f(x)-f_S|>\lambda\}| \leqslant B|S|\exp(-K\lambda/\|f\|_{BMO})$$

for any sphere S and $\lambda>0$. BMO is a slightly larger space than L_∞ (e.g. $\log|x|\in BMO$) and has better properties. For example, the †Calderón-Zygmund operators are bounded in $BMO(\mathbf{R}^n)$. We have

$$BMO(\mathbf{R}^n)=L_\infty(\mathbf{R}^n)+\sum_{j=1}^n R_jL_\infty(\mathbf{R}^n),$$

where R_j are the †Riesz transforms [9]. This is called thc **Fefferman-Stein decomposition**. The Riesz transforms can be replaced by more general families of singular integral operators (A. Uchiyama, *Acta Math.* (1982)).

(8) The Hardy Spaces H_p $(0<p<\infty)$. The classical theory of Hardy classes (→ 159 Fourier Series G) has been reconstructed by the real-analysis method and extended to higher-dimensional cases by E. M. Stein, G. Weiss, C. Fefferman, and others. According to their terminology the elements of the Hardy space H_p are (the complex linear combinations of) the real parts of the boundary values of holomorphic functions of the Hardy class.

Let $f\in\mathscr{S}'(\mathbf{R}^n)$ be a †tempered distribution. For a $\varphi\in\mathscr{S}(\mathbf{R}^n)$ define the radial maximal function M_φ^+f and the nontangential maximal function M_φ^*f relative to φ by

$$M_\varphi^+f(x)=\sup_{t>0}|\varphi_t*f(x)|,$$

$$M_\varphi^*f(x)=\sup_{|x-y|\leqslant t}|\varphi_t*f(y)|,$$

where $\varphi_t(x)=t^{-n}\varphi(x/t)$ and $*$ denotes convolution. Then the **Hardy space** $H_p(\mathbf{R}^n)$, $0<p<\infty$, is defined to be the space of all tempered distributions f which satisfy the following equivalent conditions: (i) $M_\varphi^+f\in L_p$ for a φ with $\int\varphi(x)\,dx\neq0$; (ii) $M_\varphi^*f\in L_p$ for a φ with $\int\varphi\,dx\neq0$; (iii) $M_\varphi^+f\in L_p$ for any φ; (iv) $M_\varphi^*f\in L_p$ for any φ. If a distribution f satisfies (one of) these conditions, then its Poisson integral $u(x,t)=\varphi_t*f(x)$ is a function, where $\varphi(x)=\mathrm{const}(1+|x|^2)^{-(n+1)/2}$, and its **radial maximal function** $u^+(x)=\sup_{t>0}|u(x,t)|$ and **nontangential maximal function** $u^*(x)=\sup_{|x-y|<t}|u(y,t)|$ both belong to L_p. Conversely if $u(x,t)$ is a harmonic function on the upper half-space $t>0$ and if either u^+ or u^* belongs to L_p, then its boundary value $f=u(\cdot,0)$ exists in the sense of tempered distribution and f satisfies the above conditions. Define the norm of an $f\in H_p(\mathbf{R}^n)$ by $\|u^*\|_p$. Then $H_p(\mathbf{R}^n)$ becomes a quasi-Banach space. If $1<p<\infty$, then $H_p(\mathbf{R}^n)=L_p(\mathbf{R}^n)$ with equivalent norms. $H_1(\mathbf{R}^n)$ is a Banach space strictly smaller than $L_1(\mathbf{R}^n)$. An $f\in L_1(\mathbf{R}^n)$ belongs to $H_1(\mathbf{R}^n)$ if and only if all the Riesz transforms R_jf are in $L_1(\mathbf{R}^n)$, and $\|f\|_{H_1}$ is equivalent to $\|f\|_1+\sum\|R_jf\|_1$. Similar characterizations of $H_p(\mathbf{R}^n)$ are also known for $p>0$ (Fefferman and Stein [9]).

Let Γ_j, $j=1,\dots,m$, be †proper convex open cones in $\mathbf{R}^n$ such that the †polars Γ_j° cover $\mathbf{R}^n$. Then an $f\in\mathscr{S}'(\mathbf{R}^n)$ belongs to $H_p(\mathbf{R}^n)$ if and only if there are holomorphic functions $F_j(x+iy)$ on $\mathbf{R}^n+i\Gamma_j$ such that $\sup\{\|F_j(\cdot+iy)\|_p \mid y\in\Gamma_j\}<\infty$ and $f=\sum F_j(\cdot+i\Gamma_j0)$ (D. L. Burkholder, R. F. Gundy, and M. L. Silverstein for $n=1$ and L. Carleson for $n>1$). Let $0<p\leqslant1$. A measurable function a on $\mathbf{R}^n$ is said to be a ***p*-atom** if there is a sphere S such that $\operatorname{supp}a\subset S$ and $\|a\|_\infty\leqslant|S|^{-1/p}$ and if $\int a(x)x^\alpha\,dx=0$ for all multi-indices α with $|\alpha|\leqslant n(p^{-1}-1)$. Here a **multi-index** α is an n-tuple $(\alpha_1,\dots,\alpha_n)$ of nonnegative integers, $|\alpha|=\alpha_1+\ldots+\alpha_n$ and $x^\alpha={x_1}^{\alpha_1}\ldots{x_n}^{\alpha_n}$. A distribution f belongs to $H_p(\mathbf{R}^n)$, $0<p\leqslant1$, if and only if there are a sequence of p-atoms a_j and a sequence of numbers $\lambda_j\geqslant0$ in l_p such that $f=\sum\lambda_ja_j$ in the sense of distributions, and the norm $\|f\|_{H_p}$ is equivalent to the infimum of $\|\lambda_j\|_{l_p}$ (R. R. Coifman for $n=1$ and R. H. Latter for $n>1$). The theory of H_p and BMO has been generalized to more general situations (→ Coifman and Weiss, Bull. Amer. Math. Soc. (1977)).

From now on we assume that Ω is a domain in the n-dimensional Euclidean space $\mathbf{R}^n$ (or more generally a differentiable manifold). D^α stands for ${D_1}^{\alpha_1}\ldots{D_n}^{\alpha_n}$, where $D_j=\partial/\partial x_j$.

(9) The Function Spaces $C^l(\Omega)$ and $C_0^l(\Omega)$ $(l=0,1,2,\dots,\infty)$. The totality of l-times continuously differentiable functions in Ω (namely, differentiable functions of †class C^l in Ω) is denoted by $C^l(\Omega)$. We say that a sequence $\{f_\nu\}$ of functions in $C^l(\Omega)$ converges to 0 in $C^l(\Omega)$ if $|D^\alpha f_\nu(x)|$ converges to 0 uniformly on

every compact subset of Ω for every α satisfying $0 \leqslant |\alpha| \leqslant l$ ($0 \leqslant |\alpha| < \infty$ if $l = \infty$). $C^l(\Omega)$ is a Fréchet space. The totality of functions in $C^l(\Omega)$ whose supports are compact subsets of Ω is denoted by $C_0^l(\Omega)$. We say that a sequence $\{f_\nu\}$ of functions in $C_0^l(\Omega)$ converges to 0 in $C_0^l(\Omega)$ if $\operatorname{supp} f_\nu$ ($\nu = 1, 2, \ldots$) is contained in a compact subset of Ω independent of ν and $\{f_\nu\}$ converges to 0 in $C^l(\Omega)$. $C_0^l(\Omega)$ is an †(LF)-space.

When Ω is a locally closed set in $\mathbf{R}^n$ (or a differentiable manifold with boundary), we denote by $C^l(\Omega)$ the totality of functions $f(x)$ on Ω together with their continuous formal derivatives $D^\alpha f(x)$, $|\alpha| \leqslant l$ ($|\alpha| < \infty$ if $l = \infty$) on Ω such that for every α and m with $|\alpha| \leqslant m \leqslant l$ ($m < \infty$ if $l = \infty$),

$$D^\alpha f(x) - \sum_{|\beta| \leqslant m - |\alpha|} D^{\alpha+\beta} f(y)(x-y)^\beta/\beta!$$
$$= o(|x-y|^{m-|\alpha|})$$

locally uniformly in Ω as $|x-y|$ tends to 0. Convergence in $C^l(\Omega)$ is defined in the same way as above.

If Ω is a closed set in $\mathbf{R}^n$ with a finite number of connected components in each of which two points x and y are connected by an arc of length $\leqslant C|x-y|$ with a constant C independent of x and y, then every f in $C^l(\Omega)$ can be extended to an $\tilde{f}$ in $C^l(\mathbf{R}^n)$ (**Whitney's extension theorem**).

(10) The Lipschitz Spaces Λ^s. Let $s > 0$, and let k be the least integer greater than s. The **Lipschitz** (or **Hölder**) **space** $\Lambda^s(\mathbf{R}^n)$ is the totality of functions $f(x)$ on $\mathbf{R}^n$ which satisfy

$$\|f\|_{\Lambda^s} = \sup_{x,y} \left| \sum_{j=0}^{k} \binom{k}{j} (-1)^j f(x+jy) \right| \Big/ |y|^s < \infty.$$

When $s < 1$, this is exactly the †Lipschitz (or Hölder) condition of order s. But when $s = 1$, it is strictly weaker than the †Lipschitz condition. A function $f \in \Lambda^1$ is said to be **smooth in the sense of A. Zygmund** [6]. Suppose $0 < h < s$ is an integer. Then a function f belongs to Λ^s if and only if it is h times continuously differentiable and all the derivatives $D^\alpha f$ of order h are in Λ^{s-h}. $\Lambda^s(\mathbf{R}^n)$ is a Banach space of functions modulo the polynomials of degree $\leqslant k-1$.

Suppose that $1 \leqslant q < \infty$ and f is a measurable function on $\mathbf{R}^n$ such that

$$\sup_S \inf_P \left(|S|^{-1-(sq/n)} \int_S |f(x) - P(x)|^q \, dx \right)^{1/q} < \infty,$$

where the supremum is taken over all spheres in $\mathbf{R}^n$ and the infimum over all polynomials of degree $\leqslant s$. Then $f(x)$ is equal to a function $\tilde{f}(x)$ in $\Lambda^s(\mathbf{R}^n)$ almost everywhere and the supremum is equivalent to the norm $\|\tilde{f}\|_{\Lambda^s}$. Conversely every $f \in \Lambda^s(\mathbf{R}^n)$ satisfies the above inequality (S. Campanato).

(11) The Sobolev Spaces $W_p^l(\Omega)$, $H^l(\Omega)$, and $H_0^l(\Omega)$ ($l = 0, 1, 2, \ldots$, $1 \leqslant p \leqslant \infty$ or $-\infty < l < \infty$, $1 < p < \infty$). Let $l \geqslant 0$ be an integer and $1 \leqslant p \leqslant \infty$. The Sobolev space $W_p^l(\Omega)$ is the totality of functions $f(x)$ such that for all α satisfying $|\alpha| \leqslant l$, the derivatives $D^\alpha f(x)$ in the sense of distribution (→ 125 Distributions and Hyperfunctions) belong to $L_p(\Omega)$ with respect to Lebesgue measure in Ω. $W_p^l(\Omega)$ is a Banach space with the norm

$$\|f\| = \|f\|_{W_p^l} = \left(\sum_{0 \leqslant |\alpha| \leqslant l} \int_\Omega |D^\alpha f(x)|^p \, dx \right)^{1/p}.$$

Clearly $W_p^0(\Omega) = L_p(\Omega)$. $W_2^l(\Omega)$ is a Hilbert space with respect to the inner product

$$(f, g) = \int_\Omega \sum_{0 \leqslant |\alpha| \leqslant l} D^\alpha f(x) \cdot \overline{D^\alpha g(x)} \, dx.$$

Sometimes $W_2^l(\Omega)$ is denoted by $H^l(\Omega)$. Its closed linear subspace obtained as the completion of $C_0^\infty(\Omega)$ is denoted by $H_0^l(\Omega)$. We have $H_0^0(\Omega) = W_2^0(\Omega) = L_2(\Omega)$. However, if $l \geqslant 1$, we have $H_0^l(\Omega) \subset W_2^l(\Omega)$, and identity does not hold unless $\Omega = \mathbf{R}^n$.

The definition of Sobolev spaces has been extended to those with fractional and also negative order $-\infty < s < \infty$ in many different ways. When $1 < p < \infty$, it is natural to define $W_p^s(\mathbf{R}^n)$ to be the space of all tempered distributions f on $\mathbf{R}^n$ such that $(1-\Delta)^{s/2} f = \mathscr{F}^{-1}((1 + |\xi|^2)^{s/2} \mathscr{F} f) \in L_p(\mathbf{R}^n)$. If $s > 0$, then $W_p^s(\mathbf{R}^n)$ thus defined coincides with the space of all $f \in L_p(\mathbf{R}^n)$ whose Poisson integral $u(x, t)$ satisfies

$$\left\| \left(\int_0^\infty t^{4k-2s-1} |\Delta^k u(\cdot, t)|^2 \, dt \right)^{1/2} \right\|_p < \infty$$

for some (and any) integer $k > s/2$.

The Poisson integral can be replaced by other regularizations, and thus the definition of Sobolev spaces of fractional orders is extended to arbitrary open set Ω with the cone condition (T. Muramatu [13]). Here Ω is said to satisfy the cone condition if there are a bounded and uniformly continuous mapping $\Psi: \mathbf{R}^n \to \mathbf{R}^n$ and an $\varepsilon > 0$ such that for any $x \in \Omega$ the convex hull of the ε-ball with center at $x + \Psi(x)$ and $\{x\}$ is included in Ω.

(12) The Besov Spaces $B_{p,q}^s$ ($-\infty < s < \infty$, $1 \leqslant p, q \leqslant \infty$). The effort to make the Sobolev embedding theorem [10] more precise led S. M. Nikol'skiĭ and O. V. Besov [11] to the other classes of "Sobolev spaces" of fractional order. Let $s > 0$ and $1 \leqslant p, q \leqslant \infty$. The **Besov space** $B_{p,q}^s(\mathbf{R}^n)$ is the totality of functions $f \in L_p(\mathbf{R}^n)$ such that

$$|f|_{B_{p,q}^s} = \left(\int \left\| \sum_{j=0}^{k} \binom{k}{j} (-1)^j f(\cdot + jy) \right\|_p^q \frac{dy}{|y|^{qs+n}} \right)^{1/q}$$
$$< \infty$$

for some (and any) integer $k>s$. In terms of the Poisson integral an $f \in L_p(\mathbf{R}^n)$ belongs to $B^s_{p,q}(\mathbf{R}^n)$ if and only if

$$\left(\int_0^\infty t^{(2k-s)q-1}\left\|\Delta^k u(\cdot,t)\right\|_p^q dt\right)^{1/p}<\infty$$

for some (and any) integer $k>s/2$. The first definition is obviously extended to an arbitrary domain Ω in $\mathbf{R}^n$. If Ω satisfies the cone condition, then the functions in $B^s_{p,q}(\Omega)$ are characterized similarly as above by using a suitable regularization $u(x,t)$ of $f(x)$ (Muramatu [13]). Let Ω be a domain with the cone condition. When $0<h<s$ is an integer, an f belongs to $B^s_{p,q}(\Omega)$ if and only if $f \in W^h_p(\Omega)$ and all the derivatives $D^\alpha f$ of order h are in $B^{s-h}_{p,q}(\Omega)$. $B^s_{p,q}(\Omega)$ is a Banach space with the norm $\|f\|_p + |f|_{B^s_{p,q}}$. If $p=2$, then $B^s_{2,2}(\Omega)$ coincides with the Sobolev space $W^s_2(\Omega)$. But in the other cases, $B^s_{p,q}(\Omega)$ is different from $W^s_p(\Omega)$ for any q. However, $B^s_{p,p}(\Omega)$ or $B^s_{p,\infty}(\Omega)$ is often called a Sobolev space of order s and denoted by $W^s_p(\Omega)$. Clearly we have $B^s_{\infty,\infty}=L_\infty \cap \Lambda^s$.

The **Besov space** $B^s_{p,q}(\Omega)$ of order $s \leqslant 0$ is defined to be the totality of distributions f which can be represented as $f=\sum_{|\alpha|\leqslant k} D^\alpha f_\alpha$ with $f_\alpha \in B^{s+k}_{p,q}(\Omega)$ for some (and any) integer k such that $s+k>0$. The norm is defined to be $\inf \sum_\alpha \|f_\alpha\|_{B^{s+k}_{p,q}(\Omega)}$. In terms of the Poisson integrals or other regularizations the same characterization holds as above.

If Ω is a domain in $\mathbf{R}^n$ with the cone condition, then the restriction $B^s_{p,q}(\mathbf{R}^n)\rightarrow B^s_{p,q}(\Omega)$ (resp. $W^s_p(\mathbf{R}^n)\rightarrow W^s_p(\Omega)$ for $1<p<\infty$) is a bounded linear surjection with a bounded right inverse [13].

From now on we denote by $B^s_{p,q}$, etc., $B^s_{p,q}(\Omega)$, etc. for a domain $\Omega \subset \mathbf{R}^n$ with the cone condition. If $q<r$, then $B^s_{p,q}\subset B^s_{p,r}$. If $1 \leqslant p \leqslant 2$, then $B^s_{p,p}\subset W^s_p \subset B^s_{p,2}$. If $2 \leqslant p \leqslant \infty$, then $B^s_{p,2}\subset W^s_p \subset B^s_{p,p}$. If $s>t$, then $B^s_{p,\infty}\subset B^t_{p,1}$.

Sobolev-Besov embedding theorems: (i) Let $p<p'$ and $s-n/p=s'-n/p'$. Then $B^s_{p,q}\subset B^{s'}_{p',q}$, $B^s_{p,p'}\subset W^{s'}_{p'}$, $W^s_p \subset B^{s'}_{p',p}$ and, if $p'<\infty$, then $W^s_p \subset W^{s'}_{p'}$. (ii) Let $0<s<n/p$ and $n/p'=n/p-s$. Then $B^s_{p,q}\subset L_{(p',q)}$ and $W^s_p \subset L_{(p',p)}(\subset L_{p'})$. (iii) Let $s=n/p$. Then $B^s_{p,1}\subset BC$ for any p and $B^s_{p,\infty}\subset BMO$ for $p<\infty$. (iv) Let $0 \leqslant n'<n$ and $0<s'=s-(n-n')/p$. Then there is a bounded **trace operator** $Tr: B^s_{p,q}(\mathbf{R}^n)\rightarrow B^{s'}_{p',q}(\mathbf{R}^{n'})$ that extends the restriction mapping $\mathscr{S}(\mathbf{R}^n)\rightarrow\mathscr{S}(\mathbf{R}^{n'})$. Tr is surjective and has a bounded linear right inverse [11–13].

(13) The Function Spaces $\mathscr{D}$, $\mathscr{E}$, $\mathscr{D}_{L_p}$, $\mathscr{B}$, and $\mathscr{S}$. The spaces of infinitely differentiable functions $C_0^\infty(\Omega)$ and $C^\infty(\Omega)$ are also denoted by $\mathscr{D}(\Omega)$ and $\mathscr{E}(\Omega)$, respectively. The totality of functions $f(x)$ in $C^\infty(\Omega)$ such that, for all α, $D^\alpha f(x)$ belongs to $L_p(\Omega)$ with respect to †Lebesgue measure is denoted by $\mathscr{D}_{L_p}(\Omega)$. The neighborhood $V_{l,\varepsilon}$ of 0 in $\mathscr{D}_{L_p}(\Omega)$ is defined to be the totality of functions $f(x)$ such that $\|D^\alpha f\|_p<\varepsilon$ for any α satisfying $|\alpha|\leqslant l$. In particular, $\mathscr{D}_{L_\infty}(\Omega)$ is also denoted by $\mathscr{B}(\Omega)$. ($\mathscr{B}(\Omega)$ is also used to denote the space of hyperfunctions.)

A function $f(x)$ is called a **rapidly decreasing C$^\infty$-function** if it belongs to $C^\infty(\mathbf{R}^n)$ and satisfies

$$\lim_{|x|\rightarrow\infty}|x|^k|D^\alpha f(x)|=0$$

for any α and any integer $k>0$. The totality of rapidly decreasing C$^\infty$-functions is denoted by $\mathscr{S}$. The neighborhood $V_{l,k,\varepsilon}$ of 0 in $\mathscr{S}$ is defined to be the totality of functions $f(x)$ such that $(1+|x|^2)^k|D^\alpha f(x)|<\varepsilon$ for any α satisfying $|\alpha|\leqslant l$.

The spaces in this section are †nuclear except for $\mathscr{D}_L$ and $\mathscr{B}$, and are employed in the theory of distributions [14] (→ 125 Distributions and Hyperfunctions). When $\Omega=\mathbf{R}^n$, we usually omit $(\mathbf{R}^n)$; for example, $\mathscr{D}(\mathbf{R}^n)$ and $\mathscr{D}_{L_p}(\mathbf{R}^n)$ are denoted by $\mathscr{D}$ and $\mathscr{D}_{L_p}$, respectively.

(14) The Function Spaces $\mathscr{D}_{\{M_p\}}$, $\mathscr{D}_{(M_p)}$, $\mathscr{E}_{\{M_p\}}$, $\mathscr{E}_{(M_p)}$, and $\mathscr{A}=C^\omega$. Let $\{M_p\}$ be a sequence of positive numbers satisfying the logarithmic convexity $M_p^2 \leqslant M_{p-1}M_{p+1}$. $\mathscr{E}_{\{M_p\}}(\Omega)$ (resp. $\mathscr{E}_{(M_p)}(\Omega)$) denotes the totality of C$^\infty$-functions f on Ω such that for any compact set K in Ω there are constants k and C (resp. for any $k>0$ there is a constant C) satisfying

$$|D^\alpha f(x)|\leqslant Ck^{|\alpha|}M_{|\alpha|}, \qquad x \in K, |\alpha|\geqslant 0.$$

$\mathscr{E}_{\{p!\}}(\Omega)$ is the totality of †real analytic functions on Ω and is denoted by $\mathscr{A}(\Omega)$ or $C^\omega(\Omega)$. If $\{M_p\}$ satisfies the **Denjoy-Carleman condition** $\sum M_p/M_{p+1}<\infty$, then $\mathscr{D}_{\{M_p\}}(\Omega)=\mathscr{D}(\Omega)\cap\mathscr{E}_{\{M_p\}}(\Omega)$ and $\mathscr{D}_{(M_p)}(\Omega)=\mathscr{D}(\Omega)\cap\mathscr{E}_{(M_p)}(\Omega)$ are dense in $\mathscr{D}(\Omega)$. Conversely, if $\mathscr{D}_{\{M_p\}}(\Omega)$ is different from $\{0\}$, then $\{M_p\}$ satisfies the Denjoy-Carleman condition. In this case an $f \in \mathscr{E}_{\{M_p\}}(\Omega)$ (resp. $\mathscr{E}_{(M_p)}(\Omega)$) is sometimes called an **ultradifferentiable function** of class $\{M_p\}$ (resp. (M_p)). The most important is the case where $M_p=p!^s$ for an $s>1$. Then an $f \in \mathscr{E}_{\{M_p\}}(\Omega)$ (resp. $\mathscr{E}_{(M_p)}(\Omega)$) is called a **function of Gevrey class** $\{s\}$ (resp. (s)). The topological properties of $\mathscr{A}(\Omega)$ (resp. $\mathscr{E}_{\{M_p\}}(\Omega)$, etc.) have been discussed by A. Martineau (resp. H. Komatsu).

For function spaces of S type → 125 Distributions and Hyperfunctions.

(15) The Function Spaces $\mathscr{O}(\Omega)$, $\mathscr{O}_p(\Omega)=A_p(\Omega)$, and $A(\Omega)$. Let Ω be an open set in $\mathbf{C}^n$. The totality of †holomorphic functions on Ω is denoted by $\mathscr{O}(\Omega)$. $\mathscr{O}(\Omega)$ is a †nuclear Fréchet space with the topology of uniform convergence on compact sets. It is a closed linear subspace of $C(\Omega)$ and also of $C^\infty(\Omega)$.

For any $p \geqslant 1$, the totality of functions f

holomorphic in Ω and satisfying $\int_\Omega |f(z)|^p dx\,dy < \infty$ ($z = x + iy$; $dx\,dy$ is Lebesgue measure), denoted by $\mathcal{O}_p(\Omega)$ or $A_p(\Omega)$, is a Banach space with respect to the norm $\|f\|_p = (\int_\Omega |f(z)|^p dx\,dy)^{1/p}$. In particular, it is a Hilbert space when $p = 2$.

The totality of functions bounded and continuous on the closure of Ω and holomorphic in Ω (denoted by $A(\Omega)$) is a Banach space with respect to the norm $\|f\| = \sup_{z \in \Omega} |f(z)|$.

(16) The Köthe Spaces $\bigcap \lambda(\alpha^{(k)})$ and $\sum \lambda^\times(\alpha^{(k)})$. Let $\{\alpha^{(k)}\}$ be an increasing sequence of sequences $\alpha^{(k)} = (\alpha_1^{(k)}, \alpha_2^{(k)}, \dots)$ of positive numbers. The **echelon space** $\bigcap \lambda(^{\propto(k)})$ of G. Köthe [15] is the totality of sequences $x = (\xi_1, \xi_2, \dots)$ such that $p^{(k)}(x) = \sum \alpha_i^{(k)} |\xi_i| < \infty$ for any k. It is a Fréchet space with the topology determined by the seminorms $p^{(k)}$. The **co-echelon space** $\sum \lambda^\times(\alpha^{(k)})$ is the totality of sequences $y = (\eta_1, \eta_2, \dots)$ such that $|\eta_i| \leqslant C\alpha_i^{(k)}$ for some C and k. The space s of **rapidly decreasing sequences** and the space s' of **slowly increasing sequences** are the echelon space and the co-echelon space for the sequence $\alpha_i^{(k)} = i^k$. If $\alpha_i^{(k)} = k^i$, then we obtain the space of power series with infinite radius of convergence and the space of convergent power series. More generally, let $\alpha = (\alpha_i)$ be a sequence of positive numbers. The echelon spaces for $\alpha_i^{(k)} = \exp(k\alpha_i)$ and $\exp(-k^{-1}\alpha_i)$ are called the **infinite type power series space** and the **finite type power series space** and are denoted by $\Lambda_\infty(\alpha)$ and $\Lambda_1(\alpha)$, respectively. Echelon spaces and co-echelon spaces have been employed to construct examples and counterexamples in the theory of †locally convex spaces by Köthe [15], Grothendieck, Y. and T. Kōmura, E. Dubinsky, D. Vogt, and others.

C. Dual Spaces

When Ω is a compact Hausdorff space, any bounded linear functional Φ on $C(\Omega)$ is expressed by the †Stieltjes integral

$$\Phi(f) = \int_\Omega f(x)\,d\varphi(x), \quad f \in C(\Omega), \tag{2}$$

with a †Radon measure φ, i.e., a (real- or complex-valued) †regular †countably additive set function defined on the Borel sets in Ω. Since φ is of bounded variation, the totality of such φ is denoted by $BV(\Omega)$. Conversely, any $\varphi \in BV(\Omega)$ gives a bounded linear functional on $C(\Omega)$ defined by (2), and $\|\Phi\|$ = the total variation of φ over Ω. Hence the dual space of $C(\Omega)$ is isomorphic to the Banach space $BV(\Omega)$ with the norm $\|\Phi\|$.

Let Ω be a locally compact Hausdorff space. Then the dual space of $C_\infty(\Omega)$ is again the Banach space $BV(\Omega)$ of Radon measures of bounded variation. On the other hand, the dual space of $C_0(\Omega)$ is the space of all Radon measures on Ω. N. Bourbaki takes this fact as the definition of measure.

Any bounded linear functional Φ on $L_1(\Omega)$ is expressed as

$$\Phi(f) = \int_\Omega f(x)\varphi(x)\,d\mu(x), \quad f \in L_1(\Omega), \tag{3}$$

with a suitable $\varphi \in M(\Omega)$; and $\|\Phi\| = \|\varphi\|_\infty$. Conversely, any $\varphi \in M(\Omega)$ defines a bounded linear functional on $L_1(\Omega)$ by means of (3). Accordingly, the dual space of $L_1(\Omega)$ is isomorphic to $M(\Omega)$.

The dual space of $L_p(\Omega)$ $(1 < p < \infty)$ is isomorphic to $L_q(\Omega)$, where q is the real number defined by $(1/p) + (1/q) = 1$ (accordingly, $1 < q < \infty$) and is called the **conjugate exponent** of p. Any bounded linear functional on $L_p(\Omega)$ is expressible by the formula in (3) (where $f \in L_p(\Omega)$) with $\varphi \in L_q(\Omega)$, and $\|\Phi\| = \|\varphi\|_q$.

The dual space of $M(\Omega)$ is isomorphic to the normed linear space of all (real- or complex-valued) finitely additive set functions φ defined on all measurable sets in Ω, of bounded variation over Ω, and absolutely continuous with respect to the measure μ given in Ω (i.e., $\mu(N) = 0$ implies $\varphi(N) = 0$). If $1 < p < \infty$ and $1 \leqslant q < \infty$, then the dual space of $L_{(p,q)}(\Omega)$ is isomorphic to $L_{(p',q')}(\Omega)$, where p' and q' are conjugate exponents of p and q, respectively.

If Ω is **nonatomic** (i.e., Ω has no set of positive measure that cannot be decomposed into two subsets of positive measure), then no continuous linear functionals exist other than zero on $S(\Omega)$ and on $L_p(\Omega)$ and $L_{(p,q)}(\Omega)$ for $0 < p < 1$.

The sequence spaces c_0, l_p $(1 \leqslant p < \infty)$, m, and s (the space defined in Section B (6)) are special cases of $C_\infty(\Omega)$, $L_p(\Omega)$, $M(\Omega)$, and $S(\Omega)$, respectively. Hence their dual spaces can be described explicitly. For example, the dual space of c_0 (resp. l_1) is l_1 (resp. m), and if $1 < p < \infty$, the dual space of l_p is l_q (where $(1/p) + (1/q) = 1$). The coupling of $x = (\xi_n)$ and $y = (\eta_n)$ is given by $\sum \xi_n \eta_n$. The dual space of s is the totality of sequences (η_n) such that $\eta_n = 0$ except for a finite number of n.

Let $VMO(\mathbf{R}^n)$ denote the closure of $C_0(\mathbf{R}^n)$ in $BMO(\mathbf{R}^n)$. Then the dual space of $VMO(\mathbf{R}^n)$ is identified with $H_1(\mathbf{R}^n)$. This can be considered to be a generalization of the †F. and M. Riesz theorem. On the other hand, the dual space of $H_1(\mathbf{R}^n)$ is $BMO(\mathbf{R}^n)$ [9]. Hardy spaces $H_p(\mathbf{R}^n)$, $0 < p < 1$, are not locally convex but have sufficiently many bounded linear functionals, and their dual spaces are identified with Lipschitz spaces $\Lambda^s(\mathbf{R}^n)$, where $s = n(p^{-1} - 1)$.

Let $1 < p, q < \infty$. Then the dual space of $W_p^s(\mathbf{R}^n)$ is isomorphic to $W_{p'}^{-s}(\mathbf{R}^n)$ and that of

$B^s_{p,q}(\mathbf{R}^n)$ to $B^{-s}_{p',q'}(\mathbf{R}^n)$, where p' and q' are conjugate exponents of p and q, respectively. The dual space $\mathscr{D}'$ of $\mathscr{D}$ is defined to be the space of †distributions; the dual spaces of $\mathscr{E}$, $\mathscr{D}_{L_p}$, and $\mathscr{S}$ are (algebraically) linear subspaces of $\mathscr{D}'$. Similarly, the dual spaces of $\mathscr{D}_{\{M_p\}}$ and $\mathscr{D}_{(M_p)}$ are called the spaces of †ultradistributions of classes $\{M_p\}$ and (M_p), respectively. The dual space of $\mathscr{A}$ is identified with the space of †hyperfunctions with compact support (→ 125 Distributions and Hyperfunctions). A continuous linear functional on $\mathscr{O}(\Omega)$ is called an **analytic functional**. For each analytic functional Φ there is a compact set $L\subset\Omega$ such that $|\Phi(f)|\leqslant C\sup_{z\in L}|f(z)|$. A compact set K is called a **porter** of Φ if every compact neighborhood L of K satisfies this condition. Porters are similar to supports of generalized functions, but an analytic functional does not necessarily have a smallest porter.

The dual space of an echelon space is the corresponding co-echelon space.

References

[1] S. Banach, Théorie des opérations linéaires, Warsaw, 1932 (Chelsea, 1963).
[2] N. Dunford and J. T. Schwartz, Linear operators, Interscience, I, 1958; II, 1963; III, 1971.
[3] K. Yosida, Functional analysis, Springer, 1965.
[4] J. Lindenstrauss and L. Tzafriri, Classical Banach spaces I, II, Erg. Math., 92 (1977), 97 (1979).
[5] A. Grothendieck, Critères de compacité dans les espaces fonctionneles généraux, Amer. J. Math., 74 (1952), 168–186.
[6] A. Zygmund, Trigonometric series, Cambridge Univ. Press, second edition, 1959.
[7] E. M. Stein, Singular integrals and differentiability properties of functions, Princeton Univ. Press, 1971.
[8] R. A. Hunt, On $L(p,q)$ spaces, Enseignement Math., (2) 12 (1966), 249–276.
[9] C. Fefferman and E. M. Stein, H^p spaces of several variables, Acta Math., 129 (1972), 137–193.
[10] S. L. Sobolev, Applications of functional analysis in mathematical physics, Amer. Math. Soc., 1963. (Original in Russian, 1950.)
[11] O. V. Besov, Investigations of a family of function spaces in connection with theorems of imbedding and extension, Amer. Math. Soc. Transl., Ser. 2, 40 (1964), 85–126. (Original in Russian, 1961.)
[12] H. Triebel, Interpolation theory, function spaces, differential operators, North-Holland, 1978.
[13] T. Muramatu, On Besov spaces and Sobolev spaces of generalized functions defined on a general region, Publ. Res. Inst. Math. Sci., 9 (1974), 325–396.
[14] L. Schwartz, Théorie des distributions, Hermann, revised edition, 1966.
[15] G. Köthe, Topological vector spaces I, Springer, 1969. (Original in German, 1966.)

169 (XI.19) Function-Theoretic Null Sets

A. General Remarks

By a **function-theoretic null set** we mean an exceptional set that appears in theorem as the one asserting that a certain property holds with a "small exception." We give below some of the more important examples of exceptional sets. For simplicity, we limit ourselves to the n-dimensional Euclidean space $\mathbf{R}^n$ $(n\geqslant 2)$.

B. Sets of Harmonic Measure Zero

Denote by χ_E the characteristic function of a set E on the boundary ∂D of a bounded domain D in $\mathbf{R}^n$. We call the †hypofunction $\underline{H}_{\chi_E}$ and †hyperfunction $\overline{H}_{\chi_E}$ (→ 120 Dirichlet Problem) the **inner** and **outer harmonic measures** of E (with respect to D), respectively. When they coincide, the function is called the **harmonic measure** of E. A necessary and sufficient condition for E to be of inner harmonic measure zero is that $u\leqslant 0$ hold in D whenever a †subharmonic function u bounded above in D satisfies $\limsup u(P)\leqslant 0$ as P tends to any point of $\partial D-E$. This theorem implies the following uniqueness theorem: If h is bounded and harmonic in D, if E is a set of inner harmonic measure zero on ∂D, and if $h(P)\to 0$ as P tends to any point of $\partial D-E$, then $h\equiv 0$. A necessary and sufficient condition for E to be of outer harmonic measure zero is that there exist a positive †superharmonic function v in D such that $v(P)\to\infty$ as P tends to any point of E. (Concerning the existence of a limit for a subharmonic or †harmonic function at every boundary point except those on a set of harmonic measure zero, → 193 Harmonic Functions and Subharmonic Functions.)

C. Sets of Capacity Zero

Although there are many kinds of capacity (→ 48 Capacity), here we consider only †logarithmic capacity and α-capacity $(\alpha>0)$. Let K be a nonempty compact set in $\mathbf{R}^n$. Set

$W(K)=\inf_\mu \iint \overline{PQ}^{-\alpha} d\mu(P) d\mu(Q)$, where μ runs through the class of nonnegative †Radon measures of total mass 1 supported by K, and write $C_\alpha(K)=(W(K))^{-1/\alpha}$. Define $C_\alpha(\varnothing)=0$ for an empty set $\varnothing$. For a general set $E\subset \mathbf{R}^n$, define the inner capacity by $\sup_{K\subset E} C_\alpha(K)$ and the outer capacity by the infimum of the inner capacity of an open set containing E. When the inner and outer capacities coincide, the common value is called the **α-capacity** (or **capacity of order** α) of E and is denoted by $C_\alpha(E)$. We denote the logarithmic capacity of E by $C_0(E)$. In order that $C_0(K)=0$ $(n=2)$ or the †Newtonian capacity $C_{n-2}(K)=0$ $(n\geqslant 3)$ for a compact set K, it is necessary and sufficient that the harmonic measure of K with respect to $G-K$ vanish for any bounded domain G containing K. Then K is removable for any harmonic function that is bounded or has a finite †Dirichlet integral in $G-K$. In general, K is said to be **removable** for a family F of functions if for any domain G containing K and $f\in F$ defined in $G-K$, there exists $g\in F$ defined in G such that $g=f$ in $G-K$. Let K be a compact set in $\mathbf{R}^2$ with $C_0(K)=0$ and G be a domain containing K. Let f be a holomorphic function defined in $G-K$ for which every point of K is an †essential singularity. Then the set of †exceptional values for f at every point of K is of logarithmic capacity zero. If f is †meromorphic in $|z|<1$ and

$$\iint \frac{|f'|^2}{(1+|f|^2)^2}(1-|z|)^\alpha dx\,dy<\infty \qquad (0\leqslant\alpha<1),$$

then f has a finite limit in any angular domain with a vertex at every point of $|z|=1$ except for those belonging to a set of α-capacity zero (logarithmic capacity zero if $\alpha=0$).

D. Hausdorff Measure

Let $\alpha>0$ and a set E be given in $\mathbf{R}^n$. Denote by σ a covering of E by a countable number of balls with radii $d_1, d_2, \ldots$, all of which are smaller than ε (>0). As $\varepsilon\to 0$, $\inf_\sigma \sum_k d_k^\alpha$ increases. The limit is called the **Hausdorff measure** of E of dimension α and is denoted by $\Lambda_\alpha(E)$. In order for a compact set K to be removable for the family of harmonic functions defined in a bounded domain and satisfying the †Hölder condition of order α, it is necessary and sufficient that $\Lambda_{n-2+\alpha}(K)=0$. Next, suppose that K is a compact set in a plane and the complement G of K with respect to the plane is connected. Set $\|f\|_q=(\iint_G |f|^q dx\,dy)^{1/q}$ for f holomorphic in G and q, $1\leqslant q<\infty$, and $\|f\|_\infty=\sup_G|f|$. Denote by H^q the family of $f\not\equiv 0$ with $\|f\|_q<\infty$. If p is defined by $1/p+1/q=1$, then $\Lambda_{2-p}(K)<\infty$ implies $H^q=\varnothing$ for q, $2<q<\infty$, and $\Lambda_1(K)=0$ implies $H^\infty=\varnothing$. Moreover, $H^2=\varnothing$ if and only if $C_0(K)=0$, and $H^q=\varnothing$ implies $C_{2-p}(K)=0$ for q, $2<q\leqslant\infty$ [1].

E. Null Sets Defined with Respect to Families of Functions

Conversely, L. V. Ahlfors and A. Beurling characterized the size of sets in a plane by means of families of functions [2]. Let D be a domain, and let f represent a holomorphic function in D. Fix a point z_0 in D. Set

$$\mathfrak{B}=\{f\,|\,|f|\leqslant 1\},$$

$$\mathfrak{D}=\left\{f\,\Big|\,\iint_D |f'|^2 dx\,dy\leqslant\pi\right\},$$

$$\mathfrak{E}=\{f\,|\,\text{the area of } R^c\geqslant\pi\},$$

where R^c is the complement of the †range R of $(f(z)-f(z_0))^{-1}$. Denote by $\mathfrak{S}\mathfrak{B}$, $\mathfrak{S}\mathfrak{D}$, $\mathfrak{S}\mathfrak{E}$ the families consisting of constants and †univalent functions in $\mathfrak{B}$, $\mathfrak{D}$, $\mathfrak{E}$, respectively. Use the notation $\mathfrak{F}$ to represent any one of these six families, and define $M_\mathfrak{F}=M_\mathfrak{F}(z_0;D)$ by $\sup\{|f'(z_0)|\,|\,f\in\mathfrak{F}\}$. Then $M_\mathfrak{B}=M_\mathfrak{E}\geqslant M_\mathfrak{D}=M_{\mathfrak{S}\mathfrak{E}}\geqslant M_{\mathfrak{S}\mathfrak{B}}=M_{\mathfrak{S}\mathfrak{D}}$, and $M_\mathfrak{F}(z_0;D)=0$ implies $M_\mathfrak{F}(z;D)=0$ for any $z\in D$.

Denote by $N_\mathfrak{F}$ the class of compact sets K such that the complement K^c of K is connected and $M_\mathfrak{F}(z;K^c)=0$. We call $K\in N_\mathfrak{F}$ a **null set of class** $N_\mathfrak{F}$. In order for K to be removable for $\mathfrak{B}$ or $\mathfrak{D}$, it is necessary and sufficient that $K\in N_\mathfrak{B}$ or $\in N_\mathfrak{D}$, respectively. We in general have

$$\{K\,|\,C_0(K)=0\}\subsetneqq N_\mathfrak{B}\subsetneqq N_\mathfrak{D}\subsetneqq N_{\mathfrak{S}\mathfrak{B}}.$$

If $\Lambda_1(K)=0$, then $K\in N_\mathfrak{B}$. There exists a set $K\in N_\mathfrak{B}$ with $\Lambda_1(K)>0$ (A. G. Vitushkin, *Dokl. Akad. Nauk SSSR*, 127 (1959); J. Garnett, *Proc. Amer. Math. Soc.*, 21 (1970)). When K is a subset of an analytic arc A, $K\in N_\mathfrak{B}$ implies $\Lambda_1(K)=0$, $N_\mathfrak{D}$ is equal to $N_{\mathfrak{S}\mathfrak{B}}$, and K belongs to N_D if and only if $C_0(A)=C_0(A-K)$. If an †analytic function has an essential singularity at every point of $K\in N_\mathfrak{B}$, then any compact subset of the set of exceptional values at every point of K belongs to $N_\mathfrak{B}$. A necessary and sufficient condition for $K\in N_\mathfrak{D}$ is either that the complement of any one-to-one †conformal image of K^c be of plane measure zero or that any †univalent analytic function in K^c be reduced to a †linear fractional function. If the union of at most a countable number of $N_\mathfrak{B}$ or $N_\mathfrak{D}$ sets is compact, it belongs to the same class. But it not true for $N_{\mathfrak{S}\mathfrak{B}}$ sets [3].

F. Analytic Capacity

For a compact set K, let D_K be the unbounded connected component of K^c. The quantity

$M_{\mathfrak{B}}(\infty, D_K)$ is called the **analytic capacity** of K and is denoted by $\alpha(K)$. If $\alpha(K)>0$, there exists an extremal function $f_0(z)=\alpha(K)z^{-1}+\ldots$, called the †Ahlfors function, that maps D_K onto a covering surface of the unit disk. In general, $\alpha(K)$ is not greater than the logarithmic capacity $C_0(K)$. If K is a continuum, then $\alpha(K)=C_0(K)$, and $\alpha(K)$ is attained by and only by $f_0(z)$, which maps D_K onto $|w|<1$ conformally and $z=\infty$ to $w=0$. For a linear set K, $\alpha(K)$ is equal to a quarter of its length (C. Pommerenke, *Arch. Math.*, 11 (1960)). The concept of analytic capacity is basic for the theory of rational approximation on compact sets.

References

[1] L. Carleson, Selected problems on exceptional sets, Van Nostrand, 1967.
[2] L. V. Ahlfors and A. Beurling, Conformal invariants and function-theoretic null-sets, Acta Math., 83 (1950), 101–129.
[3] L. Sario and K. Oikawa, Capacity functions, Springer, 1969.
[4] L. Sario and M. Nakai, Classification theory of Riemann surfaces, Springer, 1970.
[5] J. Garnette, Analytic capacity and measure, Lecture notes in math. 297, Springer, 1970.
[6] L. Zalcman, Analytic capacity and rational approximation, Lecture notes in math. 50, Springer, 1968.
[7] T. W. Gamelin, Uniform algebras, Prentice-Hall, 1969.

170 (IX.2) Fundamental Groups

A †continuous mapping f from the interval $I=\{t \mid 0 \leqslant t \leqslant 1\}$ into a topological space Y is called a **path** connecting the **initial point** $f(0)$ and the **terminal point** $f(1)$. In particular, a path satisfying $f(0)=f(1)=y_0$ is called a **loop** (or **closed path**) with y_0 as the **base point**. For a path f, the **inverse path** $\bar{f}$ of f is defined by $\bar{f}(t)=f(1-t)$. When the terminal point of f and the initial point of g coincide, the path F defined by $F(t)=f(2t)$ for $0 \leqslant t \leqslant 1/2$ and $F(t)=g(2t-1)$ for $1/2 \leqslant t \leqslant 1$ is called the **product** (or **concatenation**) of f and g, and is denoted by $f \cdot g$. With $[f]$ standing for the equivalence class of a path f under the relation of †homotopy relative to 0, 1 ($\in I$) (i.e., by homotopy with 0 and 1 being fixed), the inverse $[f]^{-1}=[\bar{f}]$ and the product $[f]\cdot[g]=[f\cdot g]$ are defined. In particular, in the set of homotopy classes of loops with one common base point y_0, the product is always defined, and the set forms a group $\pi_1(Y, y_0)$. This group is called the **fundamental group** (or **Poincaré group**) (H. Poincaré, 1985) of Y (with respect to y_0). If Y is †arcwise connected, then $\pi_1(Y, y_0) \cong \pi_1(Y, y_1)$ for an arbitrary pair of points y_0, y_1, and the structure of the group is independent of the choice of the base point. This group is denoted simply by $\pi_1(Y)$. A continuous mapping $\varphi:(Y, y_0)\rightarrow(Y', y_0')$ induces a homomorphism $\varphi_*:\pi_1(Y, y_0)\rightarrow\pi_1(Y', y_0')$ by sending $[f]$ to $\varphi_*[f]=[\varphi\circ f]$, and $(\varphi'\circ\varphi)_*=\varphi'_*\circ\varphi_*$ holds for the composite $\varphi'\circ\varphi$ of mappings. Thus $\pi_1(Y)$ is a †topological invariant of Y. If $\pi_1(Y)$ consists of only one class (the class of the constant path), we say that Y is **simply connected**. For example, cells and spheres S^n ($n \geqslant 2$) are simply connected. The famous †Poincaré conjecture states that a simply connected 3-dimensional compact †manifold is homeomorphic to the 3-dimensional sphere. We have **van Kampen's theorem**: Let P be a connected polyhedron, P_1 and P_2 be its connected subpolyhedra such that $P_1 \cap P_2$ is connected, and $P=P_1 \cup P_2$. Then $\pi_1(P)$ is isomorphic to the group (†amalgamated product) obtained from the †free product of $\pi_1(P_1)$ and $\pi_1(P_2)$ by giving the relations that the images of each element of $\pi_1(P_1\cap P_2)$ in $\pi_1(P_1)$ and in $\pi_1(P_2)$ are equivalent. Also, the fundamental group of the †product spaces is the direct product of the fundamental groups of the spaces involved. Any group is the fundamental group of some †CW complex. The Abelization $\pi_1/[\pi_1, \pi_1]$ of the fundamental group $\pi_1=\pi_1(Y)$ (Y arcwise connected) is isomorphic to the 1-dimensional integral †homology group $H_1(Y)$. For example, the fundamental groups of a circle S^1 and a †torus T^n are an infinite cyclic group and a free Abelian group of rank n, respectively; the fundamental group of a 1-dimensional CW complex is a free group; and the fundamental group of an orientable 2-dimensional closed surface of †genus p is a group having $2p$ generators $\{a_1, \ldots, a_p, b_1, \ldots, b_p\}$ and a relation $a_1 b_1 a_1^{-1} b_1^{-1} \ldots a_p b_p a_p^{-1} b_p^{-1}=1$. If x_0 is a fixed point of the circle S^1, then the fundamental group can be defined as the set of all homotopy classes of continuous mappings $f:(S^1, x_0)\rightarrow(Y, y_0)$.

Extending the definition of the fundamental group by replacing I, S^1 with I^n, S^n, we obtain the n-dimensional homotopy group (→ 202 Homotopy Theory; 91 Covering Spaces).

References

[1] H. Seifert and W. Threlfall, Lehrbuch der Topologie, Teubner, 1934 (Chelsea, 1965).

[2] S. Lefschetz, Introduction to topology, Princeton Univ. Press, 1949.
[3] W. S. Massey, Algebraic topology, an introduction, Springer, 1977.
[4] E. H. Spanier, Algebraic topology, McGraw-Hill, 1966.

G

171 (XXI.25) Galois, Evariste

Evariste Galois (October 25, 1811–May 31, 1832) was born in Bourg-la-Reine, a suburb of Paris. In 1828, while still in junior high school, he published a paper on periodic †continued fractions. Although he published four papers, his most important works were submitted to the French Academy of Science and either lost or rejected. He was unsuccessful in his attempt to enter the Ecole Polytechnique and instead entered the Ecole Normale Supérieure in 1829. Active in political affairs, he was expelled from school, imprisoned, and died in a duel soon after his release.

The night before the duel, he left his research outline and manuscripts to his friend, A. Chevalier. These were published by J. Liouville in *J. Math. Pures Appl.*, first series, 11 (1846). The contents include the concept of groups and what essentially became the †Galois theory of algebraic equations. The manuscript also contained expressions such as "theory of ambiguity," which seems to indicate that Galois intended to study the theory of algebraic functions along the same lines.

References

[1] E. Galois, Oeuvres mathématiques, E. Picard (ed.), Gauthier-Villars, 1897.
[2] F. Klein, Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert I, Springer, 1926 (Chelsea, 1956).
[3] R. Bourgne and J.-P. Azra, Ecrits et mémoires mathématiques d'Evariste Galois, Gauthier-Villars, 1962.

172 (III.7) Galois Theory

A. History

After the discovery of formulas giving the general solutions of algebraic equations of degrees 3 and 4 in the 16th century, efforts to solve equations of degree 5 remained unsuccessful. Early in the 19th century P. Ruffini and N. H. †Abel showed that a general algebraic solution is impossible. Shortly afterward, E. †Galois established a general principle concerning the construction of roots of algebraic equations by radicals. The principle was described in terms of the structure of a certain permutation group (the Galois group) of the roots of the equation. Even in this original form of the theory (**Galois theory**), Galois not only completed the research started by J. L. Lagrange, P. Ruffini, and Abel, but also made an epochal discovery that opened the way to modern algebra. J. W. R. Dedekind (*Werke* III, 1894) interpreted this result as a duality theorem concerning the automorphism groups of a field. It was shown later that Galois theory plays an important role in the general theory of commutative fields established by E. Steinitz. In the 1920s, W. Krull generalized the idea of Dedekind, using the concept of topological algebraic systems, and obtained the Galois theory of infinite algebraic extensions (*Math. Ann.*, 100 (1928)). The Galois theory gives a model for a successful theory summarizing the essentials of separable algebraic extensions and has led to analogous theories for other algebraic systems. For example, E. R. Kolchin constructed an analogous theory for differential fields where the Galois groups are algebraic groups (→ 113 Differential Rings, Galois theory of differential fields). Another line of development of this theory, also originated by Dedekind (*Werke* III, 1876/77), led to the Galois theory of rings, an object of active research by N. Jacobson (*Ann. Math.*, 41 (1940)), T. Nakayama, and others since the 1940s. Also, recently, the Galois theory for some general algebraic systems containing inseparable fields has been constructed by M. E. Sweedler, U. S. Chase, and others in which Galois groups are replaced by Hopf algebras or bialgebras (→ 203 Hopf algebras).

B. Definitions

Given a group G of †automorphisms of a given †field L, the subfield $F(G)=\{a\in L \mid a^\sigma=a, \sigma\in G\}$ is called the **invariant field** associated with G. For any extension fields L, L' of K, an isomorphism of L into L' whose restriction to K is the identity is called a **K-isomorphism**. If L is a †normal extension of K, any K-isomorphism of L into L' is a K-automorphism of L. If an algebraic extension L of K is normal, the group $G(L/K)$ of all K-automorphisms of L is called the **Galois group** of L/K. A †separable normal algebraic extension of K is called a **Galois extension** of K. Let L/K be a finite normal extension; then there exist intermediate fields M and N of L/K such that M/K is a Galois extension and N/K is a †purely inseparable extension and $L=M\otimes_K N$ (→ 277 Modules J). Further, $G(L/K)=G(L/N)\simeq G(M/K)$ and the order of $G(L/K)$ is equal to the †separable degree of L/K, i.e. $[M:K]$. A necessary and sufficient condition for L/K to

be a Galois extension is that the invariant field associated with $G(L/K)$ be K. A Galois extension L/K is called an **Abelian extension** or a **cyclic extension** when $G(L/K)$ is †Abelian or †cyclic, respectively (→ 149 Fields).

C. Fundamental Theorem of Galois Theory

Let L/K be a finite Galois extension and G its Galois group. Then there exists a †dual lattice isomorphism between the set of intermediate fields of L/K and the set of subgroups of G, under which an intermediate field M of L/K corresponds to the subgroup $H=G(L/M)$; conversely, a subgroup H of G corresponds to $M=F(H)$. The degree of extension $[L:M]$ is equal to the order of the corresponding subgroup H (in particular, $[L:K]$ is the order of G), and $[M:K]$ coincides with the index $(G:H)$. If subfields M and M' are †conjugate over K, then the corresponding subgroups $G(L/M)$ and $G(L/M')$ are conjugate to each other in G, and vice versa. In particular, M/K is a Galois extension if and only if the subgroup H corresponding to M is a †normal subgroup of G, and in this case, the Galois group $G(M/K)$ is isomorphic to the factor group G/H.

D. Extensions of a Ground Field

Let L/K be a finite Galois extension, K'/K any extension, and L' the †composite field of L and K'. Then L'/K' is also a Galois extension, and its Galois group is isomorphic to $G(L/L\cap K')$ by the restriction map.

E. Normal Basis Theorem

Let L/K be a finite Galois extension with Galois group G. Then there exists an element u of L such that $\{u^\sigma \mid \sigma\in G\}$ forms a basis for L over K called a **normal basis**. If we denote by $K[G]$ the †group ring of G over K, a †$K[G]$-module structure can be introduced in L by the operation $\sum a_\sigma\sigma(x)=\sum a_\sigma x^\sigma$; the existence of a normal basis implies that L is isomorphic to $K[G]$ itself as a $K[G]$-module, or in other words, that the K-linear representation of G by means of L is equivalent to the †regular representation of G.

F. Examples of Galois Extensions

(1) †Cyclotomic Fields. Let m be a positive integer not divisible by the †characteristic of K; ζ a †primitive mth root of unity, and $L=K(\zeta)$. Then L/K is an Abelian extension, and its Galois group is isomorphic to a subgroup of the †reduced residue class group $(\mathbf{Z}/m\mathbf{Z})^*$; in particular, if $K=\mathbf{Q}$, the subgroup coincides with $(\mathbf{Z}/m\mathbf{Z})^*$, by the irreducibility of †cyclotomic polynomials. Hence the degree $[\mathbf{Q}(\zeta):\mathbf{Q}]$ is equal to $\varphi(m)$, where φ is †Euler's function.

(2) Finite Fields. A †finite field K has nonzero characteristic p, and the number q of elements of K is a power of p. Also, K is uniquely determined by q (up to isomorphism), hence it is denoted by $GF(q)$ or $\mathbf{F}_q$. Thus $GF(q^n)$ is the only extension of $GF(q)$ of degree n; moreover, it is a cyclic extension.

(3) Kummer Extensions. Assume that K contains a primitive mth root ζ of unity and the characteristic of K is 0 or is not a divisor of m. Denote by K^* the multiplicative group of K. An extension L of K can be expressed in the form $L=K(\sqrt[m]{a_1},\dots,\sqrt[m]{a_r})$ $(a_i\in K)$ if and only if L/K is an Abelian extension and all $\sigma\in G(L/K)$ satisfy $\sigma^m=1$; in this case L/K is called a **Kummer extension** of **exponent** m. There exists a one-to-one correspondence between Kummer extension L of exponent m over K and finite subgroups $H/(K^*)^m$ of the factor group $K^*/(K^*)^m$, given by the relations $H=L^m\cap K^*$, $L=K(\sqrt[m]{H})$. Moreover, there exists a canonical isomorphism between $H/(K^*)^m$ and the †character group of $G(L/K)$, so that $H/(K^*)^m$ is isomorphic to $G(L/K)$. Let $L=K(\theta)$ be a cyclic Kummer extension of degree m of K, and let σ be a generator of the Galois group $G(L/K)$. Then the **Lagrange resolvent** $(\zeta,\theta)=\theta+\zeta\theta^\sigma+\dots+\zeta^{m-1}\theta^{\sigma^{m-1}}$ satisfies $(\zeta,\theta)^\sigma=\zeta^{-1}(\zeta,\theta)$, $(\zeta,\theta)^m\in K$, and θ and its conjugates can be expressed in terms of (ζ,θ). In particular, L is generated by (ζ,θ) over K.

(4) Artin-Schreier Extensions. Assume that K is of characteristic $p\neq 0$. For any element a of an extension of K, we denote by $\mathscr{P}a$ the element a^p-a and by $(1/\mathscr{P})a$ a root of $\mathscr{P}X-a=0$. A finite extension L of K is of the form $L=K((1/\mathscr{P})a_1,\dots,(1/\mathscr{P})a_r)$ $(a_i\in K)$ if and only if L/K is a Galois extension whose Galois group is an Abelian group of †exponent p; in this case, L/K is called an **Artin-Schreier extension**. There exists a one-to-one correspondence between Artin-Schreier extensions L over K and finite subgroups $H/\mathscr{P}K$ of the additive group $K/\mathscr{P}K$, given by the relations $H=\mathscr{P}L\cap K$, $L=K((1/\mathscr{P})H)$; moreover, $H/\mathscr{P}K$ is isomorphic to the character group of $G(L/K)$ (therefore also to $G(L/K)$ itself). More generally, for Abelian extensions L of **exponent** p^n (i.e., Galois extensions whose Galois groups are Abelian groups of exponent p^n), we obtain similar descriptions by using the additive

group of †Witt vectors of length n instead of K (→ 449 Witt Vectors).

G. Galois Group of an Equation

L/K is a finite Galois extension if and only if L is a †minimal splitting field of a †separable polynomial $f(X)$ in $K[X]$. In this case, we call $G(L/K)$ the **Galois group of the polynomial** $f(X)$ or **of the algebraic equation** $f(X)=0$. The equation $f(X)=0$ is called an **Abelian equation** or a **cyclic equation** if its Galois group is Abelian or cyclic, respectively, while $f(X)=0$ is called a **Galois equation** if L is generated by any root of $f(X)$ over K. Generally, $G(L/K)$ can be †faithfully represented as a permutation group of roots of $f(X)=0$. If this group is †primitive, then $f(X)=0$ is called a **primitive equation**. The index of the group in the group of all permutations of roots is called the **affect** of the equation $f(X)=0$; if the affect is 1, the equation $f(X)=0$ is called **affectless**. Let $u_1, \ldots, u_n$ be †algebraically independent elements over K. Then for the polynomial $F_n(X)=X^n-u_1X^{n-1}+\ldots+(-1)^n u_n$ in $K(u_1, \ldots, u_n)[X]$, the equation $F_n(X)=0$ is called a **general equation** of degree n. The Galois group of $F_n(X)=0$ is isomorphic to the †symmetric group $\mathfrak{S}_n$ of degree n, and if K is not of characteristic 2, then the quadratic subfield corresponding to the †alternating group $\mathfrak{A}_n$ is the field $K(\sqrt{D})$ obtained by adjoining the quadratic root of the †discriminant D of $F_n(X)$.

H. Solvability of an Algebraic Equation

Assume that K is of characteristic 0, $f(X)\in K[X]$, and L is the minimal splitting field of $f(X)$. We say that the equation $f(X)=0$ is **solvable by radicals** if there is a chain of subfields $K=L_0\subset L_1\subset\ldots\subset L_r=L$ such that $L_i=L_{i-1}(\sqrt[n_i]{a_i})$ with some $a_i\in L_{i-1}$, and this is the case if and only if the Galois group of $f(X)$ is †solvable (Galois). In particular, Abelian equations are solvable by radicals. Cyclic equations are solved by using the Lagrange resolvent, and theoretically the general solvable equation can be solved by repeating this procedure. Since $\mathfrak{S}_n$ is solvable if and only if $n\leqslant 4$, it follows that a general equation of degree n is solvable only if $n=1, 2, 3, 4$ (Abel). (For a method of solving these equations → 10 Algebraic Equations D.) Also, a polynomial is solvable by square roots if and only if the order of the Galois group is a power of 2. This fact enables us to answer some questions concerning geometric construction problems such as trisection of an angle or division of a circumference in equal parts (→ 179 Geometric Construction).

I. Infinite Galois Extensions

If a Galois extension L/K is infinite, then its Galois group is an infinite group. Let $\{M_\nu\}$ be the family of intermediate fields of L/K that are finite and normal over K, and put $H_\nu=G(L/M_\nu)$. Then by taking $\{H_\nu\}$ as a †base of a neighborhood system of the unity element, G becomes a †topological group. This topology is called the **Krull topology** (Krull, *Math. Ann.*, 100 (1928)). G is then isomorphic to the †projective limit of the family of finite groups $\{G/H\}$ and is †totally disconnected and †compact. There is a one-to-one correspondence between the set of intermediate fields of L/K and the set of closed subgroups of G given by the map (Galois group)$\leftrightarrow$(invariant field), and thus we have a generalization of Galois theory for finite extensions as described in Section C. Various theories, including the theory of Kummer extensions, can be generalized to the case of infinite extensions (→ 423 Topological Groups).

J. Galois Cohomology

Let L/K be a finite Galois extension and G its Galois group. Then both the additive group L and the multiplicative group L^* have G-module structures. The †cohomology groups of G with coefficient module L are 0 for all dimensions because of the existence of a normal basis (→ 200 Homological Algebra). As for the multiplicative group L^*, we have $\hat{H}^0(G, L^*)\cong K^*/\mathrm{N}(L^*)$ (N is the †norm $\mathrm{N}_{L/K}$), $H^1(G, L^*)=0$ (**Hilbert's theorem 90** or the **Hilbert-Speiser theorem**). In particular, if G is a cyclic group with generator σ, then every element a such that $\mathrm{N}(a)=1$ can be expressed in the form $a=b^{1-\sigma}$. $H^2(G, L^*)$ is isomorphic to the †Brauer group of †central simple algebras over K which have L as a †splitting field. In the case of number fields, a number of G-modules arise, such as †principal orders, †unit groups, †ideal groups, †idele groups, and so on, whose cohomological considerations are important (→ 6 Adeles and Ideles; 59 Class Field Theory). Further, let A be a group (not necessarily commutative), and suppose that G acts on A. We denote by ${}^\sigma a$ the action of $\sigma\in G$ on $a\in A$. A is called a **G-group** if ${}^\sigma(ab)={}^\sigma a{}^\sigma b$ for all $a, b\in A$ and $\sigma\in G$. Then, in the same way as for G-modules, we can define the **0th cohomology group** $H^0(G, A)$ to be the subgroup A^G of A consisting of all elements of A left fixed by G. The map $a:\sigma\mapsto a_\sigma$ from G into A such that $a_{\sigma\tau}=a_\sigma{}^\sigma a_\tau$ $(\sigma, \tau\in G)$ is called a **1-cocycle** with

values in A. Denote by $Z^1(G,A)$ the set of the 1-cocycles of G with values in A. Two 1-cocycles a and a' in $Z^1(G,A)$ are called **cohomologous** if there exists an element $b\in A$ such that $a'_\sigma=b^{-1}a_\sigma{}^\sigma b$ for all $\sigma\in G$. This is an equivalence relation in $Z^1(G,A)$ and the 1-**cohomology set** $H^1(G,A)$ of G with values in A is defined to be the set of cohomologous classes of $Z^1(G,A)$. $H^1(G,A)$ does not have the structure of a group in general, but there does exist an element called an identity, namely, the cohomologous class containing the trivial 1-cocycle. In general, a set X with an element p of X is called a **pointed set**; for any two pointed sets (X,p) and (Y,q), a map $f:X\rightarrow Y$ is called a **morphism of pointed sets** if $f(p)=q$. For pointed sets (X,p), (Y,q), and (Z,r), a sequence of morphisms $X\xrightarrow{f}Y\xrightarrow{g}Z$ of pointed sets is called **exact** if $\operatorname{Im}f=\operatorname{Ker}g$. The 1-cohomology set $H^1(G,A)$ with the identity can be regarded as a pointed set. Therefore, exact sequences of these sets make sense and possess some of the properties of cohomology groups in the commutative case. For example, let $1\rightarrow A\rightarrow B\rightarrow C\rightarrow 1$ be an exact sequence of G-groups (and A is central in B); then

$$H^0(G,A)\rightarrow H^0(G,B)\rightarrow H^0(G,C)$$
$$\rightarrow H^1(G,A)\rightarrow H^1(G,B)\rightarrow H^1(G,C)\ (\rightarrow H^2(G,A))$$

is an exact sequence of pointed sets. For a Galois extension L/K with Galois group G, a linear algebraic L-group defined over K (→ 13 Algebraic Groups) has naturally the structure of a G-group, and we have $H^1(G,GL_n(L))=0$. Applying the above exact sequence to $1\rightarrow SL_n(L)\rightarrow GL_n(L)\xrightarrow{\det}L^*\rightarrow 1$, we have $H^1(G,SL_n(L))=0$. Also, we have $H^1(G,Sp_{2n}(L))=0$ for any integer $n\geqslant 1$. It is difficult to define higher cohomology sets naturally, but various methods to define them have been obtained.

In many cases, we are more concerned with the †category of Galois extensions of K with K-isomorphisms between them than with a single extension L/K. In other words, we consider a †functor $L\rightarrow\mathfrak{F}(L)$ of the category of Galois extensions of K into the category of (Abelian) groups, and study the cohomology related to $G(L/K)$-module or $G(L/K)$-group structures derived from $\mathfrak{F}(L)$. In the case of infinite algebraic extensions, we consider the inductive limit of cohomology of subfields of finite degrees, making use of continuous cocycles of Galois groups relative to the Krull topology [8,9].

Let L/K be a Galois extension with Galois group G. Consider objects X, Y defined over K on which the extension of the ground field is defined (such as K-linear spaces with certain tensors on them, algebras over K (associative or not), K-varieties, or K-algebraic groups defined over K). When X and Y are isomorphic over L, Y is called a L/K-form of X. Let $E(L/K,X)$ be the set of all isomorphism classes of X over L. Let A be the group of all automorphisms over L of X. For a L/K-form Y of X and an isomorphism $f:X\rightarrow Y$ over L, since G acts on X and Y, one can define the isomorphism ${}^\sigma f:X\rightarrow Y$ over L that satisfies ${}^\sigma(f(x))={}^\sigma f({}^\sigma x)$. In particular, A is a G-group by this action. Further, for $\sigma\in G$, define $a_\sigma=f^{-1}\cdot{}^\sigma f\in A$, then $a:\sigma\mapsto a_\sigma$ is a 1-cocycle of G with values in A that corresponds to a L/K-form Y of X. This induces a bijection from $E(L/K,X)$ onto $H^1(G,A)$. Thus L/K-forms of X can be classified if one can determine $H^1(G,A)$. For example, $H^1(G,Sp_{2n}(L))=0$ is equivalent to saying that an L/K-form of a skew-symmetric bilinear form on a K-linear space of dimension $2n$ is unique up to K-isomorphisms. L/K-forms of a semisimple Lie algebra $\mathfrak{g}$ over K can be classified by the 1-cohomology set of the algebraic group $\operatorname{Aut}(\mathfrak{g}\otimes_K L)$. (For the L/K-forms of algebraic groups → 13 Algebraic Groups M.) This descent theory can also be discussed for more general categories.

K. Galois Theory of Rings

The theory of †centralizers in simple algebras can be interpreted as the theory of a certain Galois correspondence with respect to inner automorphism groups. Also, by using †crossed products, we can deduce from the theory of centralizers in simple algebras the Galois theory of commutative fields. On the other hand, Jacobson obtained a Galois theory of †division rings with respect to finite groups of outer automorphisms that is similar to the commutative case. Since then, many algebraists have proceeded with investigations that aim either at unifying these two theories by admitting inner automorphisms in the group of automorphisms, at extending the theory from division rings to general rings such as simple rings, †primitive rings, or †semiprimary rings, or at weakening the finiteness conditions. One principal method in these theories lies in considering first the endomorphism ring $\operatorname{Hom}_R(S,S)$ for an extension S/R and then the roles of endomorphisms (or derivations) in it [6] (→ 29 Associative Algebras).

Let K be a field of characteristic $p>0$ and L/K be a finite purely inseparable extension such that $L^p\subset K$. The set of all †derivations of L/K forms a restricted Lie algebra $D(L/K)$ over K. Then there exists a one-to-one and dual lattice isomorphic correspondence between the set of intermediate fields M of L/K

and the set of restricted Lie subalgebras H of $D(L/K)$, given by the relations $H = D(L/M)$ and $M = \{a \in L \mid d(a) = 0, d \in H\}$ (Jacobson [7]). Namely, in the case of purely inseparable extensions, derivations or higher derivations play the role of automorphisms, and the bialgebras defined by such derivations correspond to Galois groups of Galois extension. Now, the group algebra KG of a group G over K and the (restricted) universal enveloping algebra of a (restricted) Lie algebra over K both have the structure of a †Hopf algebra. From this fact, unifying the Galois theory of Galois extensions and Jacobson's theory for purely inseparable extensions and using bialgebras or Hopf algebras, one can construct Galois theories of more general objects containing certain nonseparable field extensions (see M. E. Sweedler, *Ann. Math.*, 87 and 88 (1968); S. U. Chase and Sweedler, *Lecture notes in math.* 97, Springer, 1969).

References

[1] J. W. R. Dedekind, Über die Permutationen des Körpers aller algebraischen Zahlen, Gesammelte mathematische Werke, Viewig, 1930–1932, vol. 2, 272–291.
[2] E. Artin, Galois theory, Univ. of Notre Dame Press, second edition, 1948.
[3] N. Bourbaki, Eléments de mathématique, Algèbre, ch. 5, Actualités Sci. Ind. 1102b, Hermann, second edition, 1959.
[4] N. G. Chebotarev, Grundzüge der Galois'schen Theorie, Noordhoff, 1950.
[5] B. L. van der Waerden, Algebra I, Springer, seventh edition, 1966.
[6] N. Jacobson, Structure of rings, Amer. Math. Soc. Colloq. Publ., 1968.
[7] N. Jacobson, Lectures in abstract algebra III, Van Nostrand, 1964; Springer, 1975.
[8] J.-P. Serre, Corps locaux, Actualités Sci. Ind., Hermann, 1962; Local fields, Springer, 1979.
[9] J.-P. Serre, Cohomologie galoisienne, Lectures notes in math. 5, Springer, 1974.

173 (XIX.9) Game Theory

A. Introduction and Historical Highlights

Game theory consists of mathematical models used in the study of decision making in situations involving conflict and cooperation. A conflict arises when each player in a game selects from a list of alternatives one which, possibly together with chance and random events, leads to various outcomes over which the players have different preferences; thus the behavior of one player aiming at his own favorable outcome might induce unfavorable outcomes for others. Although the potential outcomes usually bring about conflicts among the players, there may be room for cooperation among some of them. Game theory attempts to extract that which is common and essential to such situations, to handle them by means of mathematical methods, and to provide a normative guide to rational behavior for each of the players. Game theory thus goes beyond classical theories of probability and decision making that are sufficient to solve games involving just one player and chance.

Modern game theory started in 1944 with the publication of the monumental book by von Neumann and Morgenstern [1]. These authors presented many logical classifications of games, including a distinction between two-person and n-person games, between **constant-sum (zero-sum)** and **general-sum** games (depending upon whether the sum of payoffs to the players is constant (zero) or not), and between **noncooperative** and **cooperative** games (depending upon whether any collaboration among the players is prohibited or allowed). The second edition included a remarkable expected utility theory, which has become a mainstay of game theory. Von Neumann and Morgernstern also gave three representations of games. The first representation is the so-called extensive form; this representation has been slightly modified by Kuhn [2]. The second representation is the normal-form game; this is the form which the minimax theorem for two-person zero-sum games was established [3]. This theorem was generalized to n-person general-sum noncooperative games by Nash [4]. Finally, the characteristic-function form was the one in which the authors developed the theory of stable sets. Several other solution concepts, such as the core, the Shapley value, and the bargaining set, have since been defined for games in characteristic-function form. Historically, the development of game theory has been closely related to various areas of pure mathematics, such as analysis, topology, geometry, and the foundations of mathematics. A survey of game theory up to 1957 was presented by Luce and Raiffa [5]. In the 1950s and early 1960s, several major papers appeared in five issues of the *Annals of Mathematics Studies* [6]. Lucas [7] has presented a good survey of developments to 1972, and Shubik [8] has surveyed the development of the field through 1981. Current articles

appear mainly in the *International Journal of Game Theory*, and also in the journals in fields such as operations research, management science, economics, political science, and psychology.

B. The Extensive Form

An **n-person game in extensive form** is represented by a game †tree (i.e., a connected graph with no cycles (→ 186 Graph Theory)) having the following properties: There is one special vertex corresponding to the starting point of the game. Each nonterminal vertex corresponds either to a **move** of one of the n players or to a **chance move**. The edges ascending from a vertex denote the **alternatives** of the player at this vertex. For each terminal vertex there is an n-dimensional vector whose components represent the **payoffs** to each player. The state of a player's information at any stage can be described by certain subsets of the set of all his vertices, called **information sets**. At each of his moves, player i knows which information set he is in, but not which vertex he occupies within this set. A **local strategy** for player i is a †probability distribution over the set of all alternatives at each of his information sets. A **behavior strategy** for player i is a function which assigns a local strategy to all of his information sets. A **pure strategy** is a special behavior strategy that assigns a particular choice to each information set. Kuhn [2] showed the existence of pure optimal strategies for n-person general-sum noncooperative games with perfect information (i.e., games in which all information sets contain a single vertex) as a generalization of the result for two-person zero-sum games given in [1]). Kuhn also proved the existence of equilibrium behavior strategies for games with perfect recall (i.e., games in which player i at any of his information sets remembers all his prior moves but is not aware of the prior choices of the other players). The definition of "equilibrium strategy" is given in the next section. Because of the complexity of n-person general-sum games, research on them has been minimal since the work of Kuhn. Recently a new attack on them has begun; for example, the Nash equilibrium point in extensive forms has been reexamined by Selten [9].

C. The Normal Form (or Strategic Form)

An **n-person game in normal form** is specified by $\{N, \{X^i\}_{i\in N}, \{F^i\}_{i\in N}\}$, where $N=\{1,\ldots,n\}$ is a set of players, X^i is the set of player i's **strategies**, and F^i is a real-valued function on the Cartesian product $\prod_{i=1}^n X^i$, called player i's **payoff function**. Chance is incorporated by invoking an extra set X^0 of chance moves and a probability distribution on X^0. A two-person zero-sum game or **matrix game** is the simplest case in which the existence of an equilibrium point has been established. Equilibrium follows from von Neumann's **minimax theorem**: Let $M^1=\{1,\ldots,m_1\}$, $M^2=\{1,\ldots,m_2\}$ be the sets of pure strategies that may be chosen by two players. Let a_{ij} be player 1's payoff when strategies i and j are taken by players 1 and 2. A **mixed strategy** for player i is a probability distribution on M^i. Sets of mixed strategies for players 1 and 2 are thus given by $X^1=\{x\in R^{m_1} \mid \sum_{i=1}^{m_1} x_i=1,\ x_i\geqslant 0\ \forall i\in M^1\}$ and $X^2=\{x\in R^{m_2} \mid \sum_{i=1}^{m_2} x_i=1,\ x_i\geqslant 0\ \forall i\in M^2\}$. If players 1 and 2 use the mixed strategies x^1 and x^2, respectively, the expected payoff for player 1 is $F^1(x^1,x^2)=\sum_i\sum_j x_i^1 a_{ij} x_j^2$, which is a payoff function for player 1. Von Neumann [3] proved

$$\max_{x^1\in X^1}\min_{x^2\in X^2} F^1(x^1,x^2)=\min_{x^2\in X^2}\max_{x^1\in X^1} F^1(x^1,x^2).$$

We say that a pair $(\hat{x}^1,\hat{x}^2)$ satisfying the foregoing minimax theorem is an **equilibrium point**. This pair $(\hat{x}^1,\hat{x}^2)$ turns out to be a saddle point (→ 292 Nonlinear Programming A) of $F^1(x^1,x^2)$. The duality theorem (→ 255 Linear Programming B) is mathematically equivalent to this minimax theorem. A generalization of this equilibrium to n-person general-sum noncooperative games was presented by Nash [4]. An n-tuple $(\hat{x}^1,\ldots,\hat{x}^n)$ $(\hat{x}^i\in X^i)$ is a **Nash equilibrium** if, for each i,

$$F^i(\hat{x}^1,\ldots,\hat{x}^{i-1},\hat{x}^i,\hat{x}^{i+1},\ldots,\hat{x}^n)$$
$$\geqslant F^i(\hat{x}^1,\ldots,\hat{x}^{i-1},x^i,\hat{x}^{i+1},\ldots,\hat{x}^n) \quad \text{for all } x^i\in X^i.$$

That is, no player can improve his payoff by changing his strategy if all other players continue to use the same strategies. Nash demonstrated the existence of this equilibrium for n-person general-sum noncooperative games with finite pure strategies. The existence of Nash equilibria for wider classes of noncooperative games is proved by means of fixed-point theorems (→ 153 Fixed-Point Theorems). Recently, **multistage games**, such as supergames and stochastic games in which games are played repeatedly, have become a major research topic of noncooperative game theory (with work being done on both extensive and normal forms). In two-person general-sum games or **bimatrix games**, there arise possibilities of cooperation (or bargaining) between two players. A solution concept for such situations, the **Nash bargaining solution**, was proposed by Nash [10]; this solution was subsequently discussed by Luce and Raiffa [5].

D. The Characteristic-Function Form (or Coalitional Form)

An n-person **cooperative game in characteristic-function form** is given by a pair (N, v), where $N = \{1, \ldots, n\}$ is a set of players and v is a real-valued function on a set of all subsets of N with $v(\varnothing)=0$, called a **characteristic function**. Sometimes v is assumed to be **superadditive**, i.e., $v(S \cup T) \geqslant v(S)+v(T)$ $^{\forall}S, T \subseteq N$ with $S \cap T \neq \varnothing$. $v(S)$ represents the worth or the power achievable by the subset (coalition) S when its members cooperate regardless of the behavior(s) of the players in the complement of S. An n-dimensional vector $x=(x_1, \ldots, x_n)$ is said to be an **imputation** if it satisfies (i) $x_i \geqslant v(\{i\})$ $^{\forall}i \in N$ (individual rationality) and (ii) $\sum_{i \in N} x_i = v(N)$ (group rationality). The set of all imputations, denoted by A, represents all reasonable or realizable ways of distributing the available gains among the n players. An imputation x is said to **dominate** another imputation y if there is some nonempty coalition S such that (i) $x_i > y_i$ $^{\forall}i \in S$ and (ii) $\sum_{i \in S} x_i \leqslant v(S)$ (the effectiveness of S with respect to x). The original solution concept for games in characteristic-function form given by von Neumann and Morgenstern [1] is now called the **stable-set** or **von Neumann–Morgenstern solution**. A subset K of A is a stable set if (i) no dominance relation exists between any two elements of K (internal stability) and (ii) any imputation outside K is dominated by some imputation of K (external stability). The existence of stable sets was settled negatively by the ten-person example of Lucas [11]. This example, however, is rather specialized, and thus the existence of stable sets may yet be proved for a large class of games. Stable sets have been characterized for several classes of games, and these sets accurately reflect the coalition-forming processes among the players. Thus stable-set theory remains a major research topic in the theory of games in characteristic-function form. Many other solution concepts have been developed since that of the stable set. One of these is the **core**, defined by Gillies in [6] (its naive idea had already appeared in [1]). The core is the set $C=\{x \in A \mid \sum_{i \in S} x_i \geqslant v(S)$ $^{\forall}S \subseteq N\}$. This says that no coalition can protest against or block an imputation in the core on grounds that the coalition can expect more. For superadditive games the core coincides with the set of undominated imputations, and thus the core is a subset of any stable set if both exist. The condition for nonemptiness of the core was derived by Shapley [12] using the duality theorem. Another solution concept, defined by Shapley in [6], is known as the **Shapley value**. The Shapley value is a function $\varphi(v)=(\varphi_1(v), \ldots, \varphi_n(v))$ sending an arbitrary characteristic function v on N into an n-dimensional Euclidean space satisfying the conditions (i) $\varphi_{\pi i}(\pi v) = \varphi_i(v)$, where π is any permutation of N and $\pi v(\pi S)=v(S)$ $^{\forall}S \subseteq N$; (ii) $\sum_{i \in S} \varphi_i(v)=v(S)$ $^{\forall}S \subseteq N$ such that $v(T)=v(S \cap T)$ $^{\forall}T \subseteq N$; and (iii) $\varphi_i(v+w)=\varphi_i(v)+\varphi_i(w)$ $^{\forall}i \in N$ and for any two games v and w. These three axioms uniquely determine the value

$$\varphi_i(v)=\sum_{\substack{S \subseteq N \\ S \ni i}} \frac{(s-1)!(n-s)!}{n!}(v(S)-v(S-\{i\}))$$

for each i, where s is the number of players in S. The Shapley value is an imputation, and can be viewed as a fair-division solution since the three axioms are desirable properties for any equitable allocation scheme. From the above formula, the Shapley value can also be interpreted as the average of the marginal contributions of the players in a coalition.
A **bargaining-set** concept was proposed by Aumann and Maschler in [6]. This set describes what payoffs are stable once a particular **coalition structure** (a †partition of N) has formed. Briefly, a payoff associated with a coalition structure is stable or in a bargaining set if there is no objection to it from any player; or, even with an objection, if there exists a counterobjection to such an objection from other players. For details → [6]. Since there are many different ways to define objection and counterobjection, there are various types of bargaining sets. Some of them are known to be nonempty. Two additional solution concepts, the **kernel** and the **nucleolus**, derive from investigations into particular bargaining sets. The kernel, introduced by Davis and Maschler [13], is always a nonempty subset of its bargaining set. More important is the nucleolus defined by Schmeidler [14], which is a unique imputation in the kernel and thus in the bargaining set. It is also in the core if the latter is nonempty. The "excess" of any nonempty $S \subseteq N$ for an imputation x is defined by $e(x, S)=v(S)-\sum_{i \in S} x_i$. The excess represents the dissatisfaction or the complaint of the coalition S with respect to x. The nucleolus is the imputation which minimizes the largest excess. If we have a tie, i.e., if the maximum excess attains a minimum at several imputations, then the next largest excess is to be compared, and so on. That is, the nucleolus is the †lexicographical minimum in the ordering of these arrangements. It can be computed by solving a series of linear programs.

Several generalizations and variations of the classical von Neumann–Morgenstern formulation of games in characteristic-function form have also been investigated, such as the

games without side payments studied by Aumann and others (→ e.g., M. Shubik (ed.), *Essays in Mathematical Economics: In honor of Oskar Morgenstern*, Princeton Univ. Press, 1967); the **games in partition-function form** proposed by Thrall and Lucas (R. M. Thrall, and W. F. Lucas, *Naval Res. Logistics Quart.*, 10 (1963), 281–298); and the **games with infinitely many players**, in connection with which the Shapley value theory has become a major research topic [18].

E. Applications; Related Areas of Mathematics

Game theory has been applied to many fields, such as economics, political science, management science, operations research, information theory, and control theory, as well as to pure mathematics. Games in extensive form are now important tools for analyzing the effects of information, and thus for solving many decision problems with uncertainty. Noncooperative games in normal form and Nash equilibria have been used in the study of many phenomena, including oligopolistic markets (Friedman [15]), bidding processes, electoral competition, resource allocation, and arms control. Cooperative games have been successfully applied to economics, and the relation between the core and competitive equilibria sheds further light on the theory of competitive economy. It is generally true that competitive equilibria are contained in the core. Debreu and Scarf [16] demonstrated that if the number of players approach infinity in a certain manner, the core shrinks to the set of competitive equilibria. By working in measure-theoretic terms, Aumann [17] was able to identify the core with the set of competitive equilibria. It was also demonstrated by Aumann and Shapley [18] that the set of competitive equilibria (and hence also the core) converges to the Shapley value under such formulations, provided certain conditions are satisfied. Another major application of cooperative game theory has arisen in political science, wherein value-type solutions, such as the Shapley value, are widely used as indices of the power of each participant in various voting situations. Major applications are to problems of cost allocation for public goods such as water resources [19], public transportation systems, and telephone systems; in such applications the core, the Shapley value, and the nucleolus have all been employed. The books [8, 20–22] are good references to the most recent applications of game theory.

Game theory also has many close relations with various areas of pure mathematics. The following are typical examples. The study of certain types of games in extensive form has increased our understanding of the axiom of choice (→ 34 Axiom of Choice and Equivalents A) and other foundational questions. Nash equilibrium is closely related to fixed-point theorems (→ 153 Fixed-Point Theorems) and to separation theorems (→ 89 Convex Sets A). Cooperative game theory also has many connections to functional analysis and to convex analysis.

Finally, we mention that the study of differential games (→ 108 Differential Games) is also highly developed and widely used in areas such as economics and control theory.

References

[1] J. von Neumann and O. Morgenstern, Theory of games and economic behavior, Princeton University Press, 1944; second edition, 1947; third edition, 1953.
[2] H. W. Kuhn, Extensive games, Proc. Nat. Acad. Sci. US, 36 (1950), 570–576.
[3] J. von Neumann, Zur Theorie der Gessellschaftsspiele, Math. Ann. 100 (1928), 295–320.
[4] J. F. Nash, Noncooperative games, Ann. Math., 54 (1951), 286–295.
[5] R. Luce and H. Raiffa, Games and decisions: Introduction and critical survey, Wiley, 1957.
[6] A. W. Tucker et al. (eds.), Contributions to the theory of games, vols. I–IV, and Advanced studies in game theory: Ann. Math. Studies 24, 28, 39, 40, 52, Princeton University Press, 1950, 1953, 1957, 1959, 1964.
[7] W. F. Lucas, An overview of the mathematical theory of games, Management Sci., 18, no. 5, pt. 2 (1972), P-3–P-19.
[8] M. Shubik, Game theory in the social sciences: Concepts and solutions, MIT Press, 1982.
[9] R. Selten, Reexamination of the perfectness concept for equilibrium points in extensive games, Int. J. Game Theory, 4 (1975), 25–55.
[10] J. Nash, The bargaining problem, Econometrica, 18 (1950), 155–162.
[11] W. F. Lucas, A game with no solution, Bull. Amer. Math. Soc., 74 (1968), 237–239.
[12] L. S. Shapley, On balanced sets and cores, Naval Res. Logistics Quart., 14 (1967), 453–460.
[13] M. D. Davis and M. Maschler, The kernel of a cooperative game, Naval Res. Logistics Quart., 12 (1965), 223–259.
[14] D. Schmeidler, The nucleolus of a characteristic function game, SIAM J. Appl. Math., 17 (1969), 1163–1170.
[15] J. W. Friedman, Oligopoly and the theory of games, North-Holland, 1977.
[16] G. Debreu and H. Scarf, A limit theorem on the core of an economy, Int. Econ. Rev., 4 (1963), 235–246.

[17] R. J. Aumann, Markets with a continuum of traders, Econometrica, 32 (1964), 39–50.
[18] R. J. Aumann and L. S. Shapley, Values of non-atomic games, Princeton Univ. Press, 1974.
[19] M. Suzuki and M. Nakayama, The cost allocation of the cooperative resource development: A game-theoretical approach, Management Sci., 22 (1976), 1081–1086.
[20] P. C. Ordeshook (ed.), Game theory and political science, New York Univ. Press, 1978.
[21] S. J. Brams et al. (eds.), Applied game theory, Physica-Verlag, 1979.
[22] M. Shubik, A game-theoretic approach to political economy, MIT Press, 1984.

174 (XIV.4) Gamma Function

A. The Gamma Function

The function $\Gamma(x)$ was defined by L. Euler (1729) as the infinite product

$$\frac{1}{x}\prod_{n=1}^{\infty}\left(1+\frac{1}{n}\right)^{x}\left(1+\frac{x}{n}\right)^{-1}.$$

Legendre later called it the **gamma function** or **Euler's integral of the second kind**. The latter name is based on the fact that for positive real x, we have

$$\Gamma(x)=\int_0^{\infty} e^{-t}t^{x-1}\,dt.$$

This function satisfies the functional relation

$$\Gamma(x+1)=x\Gamma(x),$$

and hence for positive integral x, we have $\Gamma(x+1)=x!$. C. F. Gauss denoted the function $\Gamma(x+1)$ by $\Pi(x)$ or $x!$, even when x is not a positive integer. The function $x!$ is also called the **factorial function**. The gamma function can also be defined as the solution of the functional equation $\Gamma(x+1)=x\Gamma(x)$ satisfying the conditions

$$\Gamma(1)=1,\qquad \lim_{n\to\infty}\frac{\Gamma(x+n)}{\Gamma(n)n^{x}}=1.$$

Furthermore, we have

$$\frac{1}{\Gamma(x)}=xe^{Cx}\prod_{n=1}^{\infty}\left(1+\frac{x}{n}\right)e^{-x/n}.$$

This expression is known as the **Weierstrass canonical form**. Here C is **Euler's constant**

$$C=\lim_{r\to\infty}\left(1+\frac{1}{2}+\ldots+\frac{1}{n}-\log n\right)$$
$$=0.57721566490153286060651209\ldots,$$

which is conjectured to be †transcendental, but as yet even its irrationality remains unproved. However, it is known, that if C were rational, the numerator and the denominator would be integers of more than 30,000 digits (R. P. Brent, 1980).

The numerical value of C was calculated by Adams (1878) to 260 decimal places, and recently it has been calculated to more than 20,000 decimal places by means of an electronic computer. Seven thousand digits have been computed by W. A. Beyer and M. S. Waterman (*Math. Comp.*, 28 (1974)), and 20,000 digits have been computed by Brent et al. (1980).

$\Gamma(x)$ is †holomorphic on the complex x-plane except at the points $x=0,\ -1,\ -2,\ldots,$ where it has simple †poles. When $\mathrm{Re}\,x>0$, we have Hankel's integral representation

$$\Gamma(x)=-\frac{1}{2i\sin\pi x}\int_C(-t)^{x-1}e^{-t}\,dt,\qquad x\neq\text{integer},$$

where the contour C lies in the complex plane cut along the positive real axis, starting at ∞, going around the origin once counterclockwise, and ending at ∞ again.

Among various properties of this function (→ Appendix A, Table 17.I), the following two formulas are especially useful for numerical calculations: **Binet's formula**

$$\log\Gamma(x)=\left(x-\frac{1}{2}\right)\log x-x+\frac{\log 2\pi}{2}+2\int_0^{\infty}\frac{\arctan(t/x)}{e^{2\pi t}-1}\,dt,\qquad \mathrm{Re}\,x>0,$$

and the †asymptotic expansion formula that holds when $|\arg x|\leqslant(\pi/2)-\delta$ $(\delta>0)$,

$$\log\Gamma(x)\sim\left(x-\frac{1}{2}\right)\log x-x+\frac{\log 2\pi}{2}+\sum_{n=1}^{\infty}\frac{(-1)^{n-1}B_{2n}}{2n(2n-1)x^{2n-1}}$$

(**Stirling's formula**), where the B_n are †Bernoulli numbers. This last formula can be rewritten as

$$\Gamma(x+1)=x!\sim x^{x}e^{-x}\sqrt{2\pi x},$$

which is used for large positive integers x.

The integrals

$$\int_0^{\lambda}e^{-t}t^{x-1}\,dt,\qquad \int_{\lambda}^{\infty}e^{-t}t^{x-1}\,dt,\qquad \mathrm{Re}\,x>0,$$

are known as the **incomplete gamma functions** and are used in statistics, the theory of molecular structure, and other fields. The †exponential integral and the †error function (→ 167 Functions of Confluent Type D) are special cases of the incomplete gamma function.

B. Polygamma Functions

The derivatives of the logarithm of the gamma function are named the **digamma function** (or **psi function**) $\psi(x)=d\log\Gamma(x)/dx$; the **trigamma function** $\psi'(x)$; the **tetragamma function** $\psi''(x)$; the **pentagamma function** $\psi'''(x)$, etc. These functions are called **polygamma functions.** In particular, $\psi(x)$ is the solution of the functional equation

$$\psi(x+1)-\psi(x)=1/x, \qquad \psi(1)=-C,$$

$$\lim_{n\to\infty}(\psi(x+n)-\psi(1+n))=0.$$

C. The Beta Function

Euler's integral of the first kind

$$B(x,y)=\int_0^1 t^{x-1}(1-t)^{y-1}\,dt,$$

$$\operatorname{Re}x>0, \qquad \operatorname{Re}y>0,$$

is called the **beta function** and is an analytic function of two variables x, y. This function is related to the gamma function as follows:

$$B(x,y)=\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

If the upper limit 1 in the integral is replaced by α, the result is called the **incomplete beta function** $B_\alpha(x,y)$.

References

[1] N. Nielsen, Die Gammafunktion I, II, Chelsea, 1964 (two volumes in one).
[2] W. Shibagaki, Theory and applications of the gamma function, with tables of complex arguments (in Japanese), Iwanami, 1952.
[3] E. T. Whittaker and G. N. Watson, A course of modern analysis, Cambridge Univ. Press, fourth edition, 1958.
[4] K. Pearson, Tables of the incomplete Γ-function, Cambridge Univ. Press, second edition, 1968.
[5] E. Artin, The gamma function, Holt, Rinehart and Winston, 1964. (Original in German, 1931.)
Also → references to 389 Special Functions.

175 (XXI.26) Gauss, Carl Friedrich

Carl Friedrich Gauss (April 30, 1777–February 23, 1855) was born into a poor family in Braunschweig, Germany. From childhood, Gauss showed genius in mathematics. He gained the favor of Grand Duke Wilhelm Ferdinand and under his sponsorship attended the University of Göttingen. In 1797, on proving the †fundamental theorem of algebra, he received his doctorate from the University of Halle. From 1807 until his death, he was a professor and director of the Observatory at the University of Göttingen.

On March 30, 1796, he made the discovery that it is possible to draw a 17-sided †regular polygon with ruler and compass, which motivated his decision to devote himself to mathematics. The publication of his *Disquisitiones arithmeticae* in 1801 opened an entirely new era in number theory. In pure mathematics, he did excellent research on †non-Euclidean geometry, †hypergeometric series, the theory of functions of a complex variable, and the whole theory of †elliptic functions.

In the field of applied mathematics he made outstanding contributions to astronomy, geodesy, and electromagnetism; he also studied the †method of least squares, the theory of surfaces (→ 111 Differential Geometry of Curves and Surfaces), and the theory of †potential. He considered perfection in papers for publication to be of utmost importance; thus his published works are few relative to his amount of research. However, the scope of his work can be seen in his diary and letters, some of which are included in his complete works, comprising 12 volumes. He is generally considered the greatest mathematician of the first half of the 19th century.

References

[1] C. F. Gauss, Werke I–XII, Königliche Gesellschaft der Wissenschaften, Göttingen, 1863–1933.
[2] C. F. Gauss, Disquisitiones arithmeticae. German translation, Untersuchungen über höhere Arithmetik, Springer, 1889 (Chelsea, 1965); English translation, Yale Univ. Press, 1966.
[3] F. Klein, Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert I, II, Springer, 1926–1927 (Chelsea, 1956).

176 (XVII.13) Gaussian Processes

A. Gaussian Systems

A system $\mathbf{X}=\{X_\lambda(\omega)\mid\lambda\in\Lambda\}$ of real-valued random variables on a probability space

$(\Omega, \mathbf{B}, P)$ is said to be **Gaussian**, or $\mathbf{X}$ is a **Gaussian system**, if any finite linear combination of elements X_λ of $\mathbf{X}$ is a †Gaussian random variable.

Any subsystem of a Gaussian system is again a Gaussian system. In particular, the joint distribution of $(X_1, X_2, \ldots, X_n)$ for any finite subsystem $\{X_j \mid 1 \leqslant j \leqslant n\}$ of a Gaussian system $\mathbf{X}$ is a multidimensional Gaussian distribution. This distribution is supported on the whole of $\mathbf{R}^n$ or on a hyperplane (at most $(n-1)$-dimensional). Let m_j be the expectation of X_j, and let $V = (V_{j,k})$ be the (positive definite) †covariance matrix of $\{X_j\}$ given by

$$V_{j,k} = E\{(X_j - m_j)(X_k - m_k)\}, \qquad 1 \leqslant j, \quad k \leqslant n.$$

Suppose that the rank of the matrix V is r. Then the distribution of $(X_1, X_2, \ldots, X_n)$ is concentrated on an r-dimensional hyperplane of $\mathbf{R}^n$:

$$m + V\mathbf{R}^n, \qquad m = (m_1, m_2, \ldots, m_n).$$

When V is nondegenerate, that is, when $r = n$, the distribution is supported by the whole of $\mathbf{R}^n$ and has a density function of the form

$$(2\pi)^{-n/2} |V|^{-1/2} \exp\left[-\frac{1}{2}(x-m)V^{-1}(x-m)' \right],$$

where $x = (x_1, x_2, \ldots, x_n) \in \mathbf{R}^n$, $|V|$ and V^{-1} are the determinant and the inverse, respectively, of V, and where $(x-m)'$ denotes the (column) vector transpose to the (row) vector $(x-m)$. The above expression is the general form of the density of an n-dimensional Gaussian distribution. The characteristic function $\varphi(z)$ of this distribution is given by

$$\varphi(z) = \exp\left[i(m, z) - \frac{1}{2}(Vz, z) \right],$$

$$z = (z_1, z_2, \ldots, z_n) \in \mathbf{R}^n,$$

where (,) denotes the inner product on $\mathbf{R}^n$. This distribution is denoted by $N(m, V)$. If, in particular, $m = 0$ and if V is the identity matrix, then it is called the n-dimensional standard Gaussian distribution.

For a general Gaussian system $\mathbf{X} = \{X_\lambda \mid \lambda \in \Lambda\}$ we are given the mean vector $m_\lambda = E(X_\lambda)$, $\lambda \in \Lambda$, and the covariance matrix $V_{\lambda,\mu} = E\{(X_\lambda - m_\lambda)(X_\mu - m_\mu)\}$, $\lambda, \mu \in \Lambda$, which is positive definite; for any $n \geqslant 1$, complex numbers $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbf{C}$, and $\lambda_1, \lambda_2, \ldots, \lambda_n \in \Lambda$, we have

$$\sum_{j,k=1}^{n} \alpha_j \bar{\alpha}_k V_{\lambda_j, \lambda_k} \geqslant 0.$$

Conversely, given m_λ, $\lambda \in \Lambda$, and a positive definite $V = (V_{\lambda,\mu} \mid \lambda, \mu \in \Lambda)$, there exists a Gaussian system $\mathbf{X} = \{X_\lambda \mid \lambda \in \Lambda\}$, the mean vector and the covariance matrix of which coincide with (m_λ) and $V = (V_{\lambda,\mu})$, respectively. If there exists another system $\mathbf{X}'$ with the same property as $\mathbf{X}$, then $\mathbf{X}$ and $\mathbf{X}'$ have the same distribution.

A necessary and sufficient condition for the X_λ in a Gaussian system to be independent is that $V_{\lambda,\mu} = 0$ for every pair $\lambda \neq \mu$. For a Gaussian system $\mathbf{X} = \{X_n \mid n \geqslant 1\}$, the †convergence in probability of the sequence $\{X_n\}$ is equivalent to †convergence in the mean square. The limit of the sequence in this case is again Gaussian in distribution. Also for this system, the †almost sure convergence of the sequence implies the convergence in mean square.

There are many characterizations of Gaussian distributions and of Gaussian systems. (i) A necessary and sufficient condition for a distribution to be Gaussian is that cumulants (†semi-invariants) γ_k of all orders exist and satisfy $\gamma_k = 0$ for all $k \geqslant 3$. (ii) Gaussian distributions have a self-reproducing property: If X and Y are independent Gaussian random variables, then their sum $S = X + Y$ is also Gaussian. A converse to this property holds: If the sum is Gaussian, then, assuming that they are independent, X and Y are both Gaussian as well (P. Lévy, H. Cramér). (iii) If X_1 and X_2 are independent and if $Y_1 = aX_1 + bX_1$ and $Y_2 = cX_1 + dX_2$ are also independent, then both X_1 and X_2 are Gaussian except for the trivial case $b = c = 0$ or $a = d = 0$. (iv) If for X and Y there exist U independent of X and V independent of Y satisfying $Y = aX + U$, $X = bY + V$ for some constants a, b, then there are only three possibilities: (1) (X, Y) is Gaussian, (2) X and Y are independent, (3) there is an affine relation between X and Y. (v) Suppose that a distribution has a finite mean m and a density function of the form $f(x-m)$. If the †maximum likelihood estimate of the mean is always given by the arithmetic mean of the samples, then the distribution is Gaussian (C. F. Gauss).

For any element X_λ of a Gaussian system $\mathbf{X}$ and for a subsystem $\mathbf{X}'$ of $\mathbf{X}$, the conditional expectation $E(X_\lambda/\mathbf{B}')$ is the orthogonal projection of X_λ onto $\overline{\mathbf{X}'}$, where $\mathbf{B}'$ is the smallest σ-field with respect to which all the X_μ in $\mathbf{X}'$ are measurable and where $\overline{\mathbf{X}'}$ is the closed linear subspace of $L^2(\Omega, P)$ spanned by $\mathbf{X}'$.

A system $\mathbf{X} = \{X_\lambda \mid \lambda \in \Lambda\}$, $X_\lambda = (X_\lambda^1, \ldots, X_\lambda^d)$, of d-dimensional random variables is said to be Gaussian if the collection $\{X_\lambda^j \mid \lambda \in \Lambda, 1 \leqslant j \leqslant d\}$ is Gaussian.

B. Complex Gaussian Systems

Let Z be a complex-valued random variable with mean m, and denote it in the form $Z = X + iY + m$, $i = \sqrt{-1}$, X, Y real. If X and Y are independent and have the same Gaussian distribution with zero mean, then Z is called a

complex Gaussian random variable. A system $\mathbf{Z}=\{Z_\lambda \mid \lambda\in\Lambda\}$ of complex-valued random variables is said to be **complex Gaussian** if any finite linear combination $\sum c_j Z_{\lambda_j}$, $c_j\in\mathbf{C}$, is complex Gaussian. A **complex Gaussian system** has properties similar to those of a Gaussian system discussed above. For instance, two complex Gaussian random variables are independent if and only if they are uncorrelated. Convergence properties are also similar. Furthermore, one can define complex Gaussian systems consisting of higher-dimensional random variables.

C. Gaussian Processes

A real-valued stochastic process $\{X_t\}$ is called a **Gaussian process** or a **normal process** if it forms a Gaussian system. If $\{X_t\}$ is a complex Gaussian system, it is called a complex Gaussian process. The most important example of a Gaussian process is †Brownian motion $\{B_t \mid t\geqslant 0\}$ with the properties $E(B_t)=0$ and $E(B_t-B_s)^2=|t-s|$. There are several generalizations of this process, among them: (i) †Brownian motion with a d-dimensional time parameter, (ii) a †Wiener process with d-dimensional time parameter, which is a Gaussian system $\{X_a \mid a=(a_1,\dots,a_d)$, all $a_j\geqslant 0\}$ with $E(X_a)=0$ and $E(X_{a^1}X_{a^2})=\prod_j \min\{a_j^1,a_j^2\}$, $a^i=(a_1^i,\dots,a_d^i)$, $i=1,2$.

Since a multidimensional Gaussian distribution is completely determined by its †mean vector and the †covariance matrix, a Gaussian process is strongly stationary if it is weakly stationary (→ 395 Stationary Processes); and such a process is called a **stationary Gaussian process**. The mean value m and the spectral measure $F(d\lambda)$ are associated with a weakly stationary process. The measure $F(d\lambda)$ is symmetric with respect to the origin since each X_t is real-valued. Conversely, given such $F(d\lambda)$ and m, we can construct a real weakly stationary process $\{X_t\}$ with mean value m and spectral measure $F(d\lambda)$. Generally, such a process $\{X_t\}$ is not determined uniquely; however, if $\{X_t\}$ is Gaussian, then there exists only one stationary Gaussian process with given m and $F(d\lambda)$. In view of this fact, stationary Gaussian processes can be regarded as being typical among weakly stationary processes.

The discussions so far on stationary Gaussian processes are generalized to the case of stationary complex Gaussian processes, for which symmetry of $F(d\lambda)$ need not be assumed. The †random measure $\{M(\Lambda)\}$ in the †spectral decomposition of a complex Gaussian process $\{X_t\}$ again forms a complex Gaussian system. Weakly stationary Gaussian random distributions can also be introduced.

D. White Noise and Gaussian Random Distributions

A real-valued weakly stationary process with discrete parameter is called a **white noise** if the mean $m=0$ and the †covariance function $\rho(t)=1$ $(t=0)$; $=0$ $(t\neq 0)$. Obviously the †spectral measure $F(d\lambda)$ is the †Lebesgue measure $d\lambda$. In the continuous parameter case, a weakly stationary random distribution (→ 407 Stochastic Processes C) is called a **white noise** if $m=0$ and $\rho=\delta$ (δ is †Dirac's delta function: $\delta(\varphi)=\varphi(0)$). The spectral measure is, therefore, the Lebesgue measure. In both cases Gaussian white noise $\{X_t\}$ or $\{X_\varphi\}$ is determined uniquely. It has independent values at every point in the following sense: In the discrete parameter case, $X_{t_1},X_{t_2},\dots,X_{t_n}$ are mutually independent for any mutually distinct t_1, $t_2,\dots,t_n$. For a continuous parameter Gaussian white noise, often called simply white noise if no confusion occurs, $X_{\varphi_1},X_{\varphi_2},\dots,X_{\varphi_n}$ are mutually independent if the supports of $\varphi_1,\varphi_2,\dots,\varphi_n$ are disjoint. In the latter case Gaussian white noise is realized by taking the derivative of a Brownian motion. The †characteristic functional $c(\varphi)$ is given by

$$c(\varphi)=\exp\left[-\frac{1}{2}\|\varphi\|^2\right],$$

where $\|\ \|$ is the $L^2(\mathbf{R}^1)$-norm.

Characteristic functionals of general Gaussian random distributions are expressed in the form

$$c(\varphi)=\exp\left[im(\varphi)-\frac{1}{2}K(\varphi,\varphi)\right],$$

where m is the mean functional and K is the covariance functional.

E. Representations of Gaussian Processes

The family of †stochastic integrals $X_t=\int_a^t F(t,u)\,dB_u$, $t>a$ ($t>a$ if $a=-\infty$), based on Brownian motion, defines a Gaussian process $\{X_t \mid t\geqslant a\}$. The converse problem is, however, not obvious. Given a Gaussian process $\{X_t \mid t>a\}$ with $E(X_t)=0$, the problem is to find a Brownian motion $\{B_t\}$ and a kernel F that give a representation of the above form (P. Lévy [10]). A more specific problem discussed below is important. If, in addition, the representation is formed in such a way that $\mathbf{B}_t(X)=\mathbf{B}_t(B)$ holds for every t, then the representation is called **canonical**, where $\mathbf{B}_t(X)$ is the smallest σ-field with respect to which all the X_s, $s\leqslant t$, are measurable. The **canonical representation**, if it exists, is unique up to equivalence, i.e., $F(t,u)^2$ is unique. The existence of the canonical representation is, however, not

always guaranteed. Mean continuous, zero mean, purely nondeterministic stationary Gaussian processes have †backward moving average representations, which are nothing but the canonical representations. Once the canonical representation of $\{X_t\}$ is obtained, the kernel, together with known properties of $\{B_t\}$, tells us important properties of the given process, e.g., sample function properties, Markov properties, and so forth.

Let $\{X_t \mid t>a\}$ be a general Gaussian process with $E(X_t)=0$, and let $\mathbf{M}_t(X)$ be the subspace of $L^2(\Omega, P)$ spanned by the X_s, $s\leqslant t$. Assume that (i) $\{X_t\}$ is separable, i.e., $\mathbf{M}_\infty(X)$ is separable and (ii) $\{X_t\}$ is purely nondeterministic in the sense that $\bigcap_s \mathbf{M}_s(X)=\{0\}$. Then X_t is expressed as a sum of stochastic integrals:

$$X_t=\sum_j \int_a^t F_j(t,u)\,dB_u^j,$$

where the $\{B_u^j\}$, $j=1,2,\ldots,N$ (N can be ∞), are mutually independent †additive Gaussian processes. In addition, the B_u^j can be taken in such a way that the measures $dv^j(u)=E(dB_u^j)^2$, $j=1,2,\ldots,N$ is decreasing in j (i.e., $dv^1 \gg dv^2 \gg \ldots$) and that $\mathbf{B}_t(X)=\bigvee_j \mathbf{B}_t(B^j)$ holds for every t. Such a representation is called a **generalized canonical representation** of $\{X_t\}$ [3]. It always exists under the above assumptions (i) and (ii), although it is not unique when $N\geqslant 2$. The number N, called the **multiplicity** of $\{X_t\}$, is independent of the choice of a generalized canonical representation. It is a good question to ask when $\{X_t\}$ has a simple unit multiplicity or when $\{X_t\}$ has the canonical representation. No interesting answer to this question has been given so far except for stationary Gaussisn processes.

F. Gaussian Markov Properties

A †simple Markov Gaussian process $\{X_t \mid t\geqslant a\}$ has the canonical representation if it is separable and purely nondeterministic and if its covariance function never vanishes. It is of the form

$$X_t=f(t)\int_a^t g(u)\,dB(u),$$

$$f(t)\neq 0, \qquad \int_a^t g(u)^2\,du\neq 0 \quad \text{for} \quad t>a.$$

A generalization of the simple Markov property of a Gaussian process is given in the following manner (suggested by P. Lévy [10]). If the †conditional expectations $E(X_{t_k}/\mathbf{B}_t(X))$, for any choice of distinct $t_1,t_2,\ldots,t_{N+j}>t$ ($j\geqslant 0$) span an N-dimensional subspace of $L^2(\Omega,P)$, then the process $\{X_t\}$ is called an **N-ple Markov Gaussian process.** If the canonical representation of an N-ple Markov Gaussian process exists, then it is expressed in the form

$$X_t=\sum_{j=1}^N f_j(t)\int_a^t g_j(u)\,dB_u,$$

where $\det(f_j(t_i))$ never vanishes for any choice of distinct t_i ($>a$), $1\leqslant i\leqslant N$, and where the g_j are linearly independent vectors in $L^2([a,t])$ for any $t>a$. Conversely, a Gaussian process given by the above expression is N-ple Markov. A stationary N-ple Markov Gaussian process has a †spectral density function $f(\lambda)$ of a specific form, namely, it is expressed in the form [3]

$$f(\lambda)=|Q(i\lambda)/P(i\lambda)|^2;\ P,\ Q \text{ polynomials},$$
$$\deg P=N \ \text{ and } \ \deg Q<N$$

(**rational spectral density function**). Y. Okabe [14] proved that the roots of $P(x)=0$ are all real, and introduced a multiple Markov property of a stationary Gaussian process $\{X_t\}$ in a much wider sense to prove that $\{X_t\}$ enjoys this property if and only if it has a rational spectral density.

A somewhat restricted definition of multiple Markov properties for a Gaussian process uses differential operators. Assume that X_t is $(N-1)$-times differentiable with respect to the $L^2(\Omega,P)$-norm. If there is an Nth order differential operator $L=\sum_{k=0}^N a_k(t)D^{N-k}$ with $D=d/dt$ such that

$$LX_t=\dot{B}_t, \qquad \dot{B}_t=dB_t/dt \ \text{ white noise},$$

then $\{X_t\}$ is called an **N-ple Markov Gaussian process in the restricted sense**. Such a process is naturally N-ple Markov, and the canonical representation always exists. The kernel of the representation is the †Riemann function of the differential operator L that was used in the definition. The spectral density function corresponding to this process is of restricted form, namely, it is expressed in the form $1/|P(i\lambda)|^2$, where P is a polynomial of degree N, and $P(i\lambda)=0$ has no root in the lower half-plane.

Many attempts have been made to define a Markov property of a Gaussian random field, namely, a Gaussian system with a multidimensional time parameter; however, only two significant approaches are mentioned here. Let $\{X_a \mid a\in\mathbf{R}^d\}$ be a Gaussian random field with d-dimensional time parameter. H. P. McKean [12] gave a Markov property in the following manner. Let $\mathscr{F}_U$ be the σ-field generated by the X_x, $x\in U$, $U(\subset\mathbf{R}^d)$ open. For a closed set C, we set $\mathscr{F}_C=\bigcap_\varepsilon \mathscr{F}_{C_\varepsilon}$, where C_ε is the ε-neighbourhood of C. If, for any open set U, $\mathscr{F}_U$ and $\mathscr{F}_{U^c}$ become independent under the assumption that $\mathscr{F}_{\partial U}$ is known, then $\{X_a\}$ is called a **Markov Gaussian random field in the McKean sense.** It has been proved that a Brownian motion with d-dimensional time

parameter is Markov in this sense if d is odd, while it is not Markov if d is even. Further detailed investigations of Markov properties for Gaussian random fields have been given by L. Pitt, S. Kotani, Y. Okabe [17], and others.

Another definition of Markov properties of Gaussian †random fields, in fact those of Gaussian random distributions, has been given by E. Nelson [13] in connection with Euclidean fields (invariant under Euclidean motion) in quantum theory. Let X_φ be a Gaussian random field. For a smooth surface Γ in $\mathbf{R}^d$ let $\mathscr{F}_\Gamma$ be the σ-field generated by X_Φ, $\Phi = \psi \otimes \delta(n)$, ψ being a smooth function with compact support and n being normal to Γ. If $\mathscr{F}_D$ and $\mathscr{F}_{D^c}$ become independent under the assumption that $\mathscr{F}_\Gamma$, $\Gamma = \partial D$, is given, then $\{X_\varphi\}$ is called a **Markov Gaussian random field in the Nelson sense**. Given a Gaussian Euclidean field which is Markov in the Nelson sense, then under some reasonable assumptions, one can form a field (in the sense of †quantum field theory) satisfying the †Wightman axiom by changing imaginary time to real time.

G. Sample Function Properties

Some finer results about the smoothness of sample functions have been obtained for stationary Gaussian processes. G. A. Hunt proved that if the spectral measure $F(d\lambda)$ of a stationary Gaussian process $\{X_t\}$ has finite moment of order α, then almost all its sample functions satisfy the following †Hölder condition of order α [5]: For every finite interval I and every constant C, there exists a sufficiently small $h_0 = h_0(I, C)$ such that

$$|X_{t+h} - X_t| < C|h|^\alpha (\log 1/|h|)^{1/2}$$

for any $t \in I$ and $|h| < h_0$. The following result is due to Yu. K. Belyaev: Sample functions of a stationary Gaussian process are with probability 1 either continuous or unbounded on every finite interval.

Other conditions for continuity of sample functions are given in terms of the †covariance function. Let $v(s,t)$ be the covariance function of a Gaussian process $\{X_t \mid t \in [0,1]\}$. If v satisfies

$$|v(s_1, t) - v(s_2, t)| \leqslant c|s_1 - s_2|^\alpha$$

$$(0 < \alpha \leqslant 1, \quad v(0, s) = 0, \quad s_1, s_2 \in [0,1]),$$

then

$$P\left(\limsup_{\delta \downarrow 0, |t_1 - t_2| \leqslant \delta} \frac{|X_{t_1} - X_{t_2}|}{|t_1 - t_2|^{\alpha/2} |\log|t_1 - t_2||^{1/2}} < c_1 c^{1/2}/\alpha^2 \right) = 1,$$

where c_1 is a constant. This result is due to Z. Ciesielski.

Conditions for continuity of sample functions have also been given for Gaussian processes with multidimensional time parameters. For $\{X_t \mid t \in [0,1]^n\}$, set $d(s,t) = E[(X_s - X_t)^2]^{1/2}$. If there exists a function φ monotone increasing on some interval $0 < u < \alpha$ such that

$$d(s,t) \leqslant \varphi(\|s - t\|),$$

where $\| \ \|$ is the metric in $\mathbf{R}^n$, and such that

$$\int^{+\infty} \varphi(e^{-x^2})\, dx < \infty,$$

then almost all sample functions are continuous (X. Fernique [2]). Conversely, if there exists a monotone increasing function φ satisfying $d(s,t) = \varphi(\|s-t\|)$ and if sample functions are continuous with positive probability, then the above integral should be finite.

Suppose that $\{X_t\}$ is a stationary Gaussian field. R. M. Dudley [11] proved that if there exists a number $q > 1$ and a neighborhood $V = V(0)$ in $\mathbf{R}^n$ such that

$$\sum_k q^{-k} (\log N(V, q^{-k}))^{1/2} < \infty,$$

where $N(V, \varepsilon)$ is the minimum number of ε-balls (relative to the metric d) that cover V, then almost all sample fields of $\{X_t\}$ are continuous, and Fernique [18] proved that the converse statement is also true. Thus the problem of continuity of the sample fields of stationary Gaussian fields has been completely solved.

H. Gaussian Measures

Let $\{X_t, t \in T\}$, T an interval, be a Gaussian process. It gives a probability distribution on the measurable space $(\mathbf{R}^T, \mathscr{B})$, where $\mathscr{B}$ is the σ-field of Borel subsets of $\mathbf{R}^T$. Such a measure P is called a Gaussian measure. Let P^i, $i = 1, 2$, be Gaussian measures induced by Gaussian processes $\{X_t^i, t \in T\}$, $i = 1, 2$, respectively. Then for these measures concepts such as †absolute continuity, †equivalence, and †singularity are introduced in the usual manner. The most significant property for Gaussian measures is that two measures P^1 and P^2 are either equivalent or singular. A powerful criterion for testing this dichotomy is the †entropy of P^2 with respect to P^1, given by

$$H(P^2 \mid P^1) = \sup_\alpha \sum_i \left\{ P^2(A_i) \log \frac{P^2(A_i)}{P^1(A_i)} \right\},$$

where the sup is taken over all finite partitions $\alpha = \{A_1, \ldots, A_n\}$, $A_i \in \mathscr{B}$, $\bigcup A_i = \mathbf{R}^T$ (Yu. A. Rozanov [15]). The measures P^1 and P^2 are equivalent (resp. singular) if and only if

$H(P^2|P^1)<\infty$ (resp. $=\infty$). This statement can be paraphrased in terms of the Hilbert spaces $H(X^i)$, $i=1, 2$, spanned by $\{X_t^i\}$, $i=1, 2$, respectively. For simplicity, assume that $E(X_t^i)=0$. Then, P^1 and P^2 are equivalent if and only if the following two conditions hold: (i) A mapping A determined by $AX_t^i=X_t^2$ is extended to an invertible linear transformation from $H(X^1)$ to $H(X^2)$; (ii) There exists a symmetric operator B of †Hilbert-Schmidt type such that $(A\xi, A\eta)_2=((I-B)\xi,\eta)_1$, where $(\xi,\eta)_i=E(\xi\eta)$ is the inner product in $H(X^i)$. Next, suppose that $\{X_t^i\}$, $i=1, 2$, are not centered. Set $E(X_t^i)=m_t^i$, $i=1, 2$. Then P^1 and P^2 are equivalent if and only if $\{\bar{X}_t^i\}$ with $\bar{X}_t^i=X_t^i-m_t^i$ satisfy the following condition: (iii) in addition to conditions (i) and (ii) above, there exists an element ξ_0 in $H(X^1)$ such that $m_t^1-m_t^2=E[\xi_0X_t^1]$, $t\in T$.

Given equivalent Gaussian measures P^1 and P^2 on $(\mathbf{R}^T,\mathscr{B})$, $T=[0,t_0]$, it is in general not easy to obtain the †Radon-Nikodym derivative; however, if one of them, say P^1, is the †Wiener measure, then one can obtain $\varphi=dP^1/dP^2$ explicitly (M. Hitsuda [4]). Assume that P^2 is derived from $\{X_t\}$. By assumption, X_t has the †canonical representation

$$X_t=B_t-\int_0^t\left\{\int_0^s l(s,u)dB_u+a(s)\right\}ds, \qquad t\in T,$$

where $\{B_t\}$ is the †Brownian motion from which P^1 is assumed to be derived and where $l\in L^2(T^2)$, $a\in L^2(T)$. With this notation, we can write

$$\varphi=\frac{dP_1}{dP_2}=\exp\left[\int_0^{t_0}\left\{\int_0^s l(s,u)dB_u+a(s)\right\}dB_s -\frac{1}{2}\int_0^{t_0}\left(\int_0^s l(s,u)dB_u+a(s)\right)^2ds\right].$$

This result yields a remark about general Gaussian measures: If P^1 and P^2 are equivalent, $\{X_t^1, t\in T\}$ has a canonical representation if and only if $\{X_t^2, t\in T\}$ does.

I. Nonlinear Functionals

Start with nonlinear functionals of a Brownian motion $\{B_t, t\in\mathbf{R}^1\}$, and call them **Brownian functionals**. If $\{\varphi_n\}$ is a complete orthonormal system in $L^2(\mathbf{R}^1)$, then the collection $\{X_n=\int\varphi_n(u)dB_u\}$ of †stochastic integrals forms a Gaussian system of independent standard Gaussian random variables. A **Fourier-Hermite polynomial** based on $\{\varphi_n\}$ is a polynomial in X_n expressible as a finite product in the form

$$c\prod_j H_{n_j}(X_j/\sqrt{2}), \qquad c \text{ a constant.}$$

The sum $n=\sum n_j$ is the degree of this polynomial. If $n>0$, its mean is zero and the variance is $|c|^2\prod_j(n_j!\,2^{n_j})$. Taking c to be $\prod_j(n!\,2^{n_j})^{-1/2}$, the collection of all Fourier-Hermite polynomials based on $\{\varphi_n\}$ forms a complete orthonormal system in the Hilbert space (L^2) consisting of all Brownian functionals with finite variance. Denote by $\mathscr{H}_n$ the subspace of (L^2) spanned by all the Fourier-Hermite polynomials of degree n. Then the Hilbert space (L^2) admits a direct-sum decomposition:

$$(L^2)=\sum_{n=0}^{\infty}\oplus\mathscr{H}_n,$$

which is called the **Wiener-Itô decomposition** [7]. The subspace $\mathscr{H}_n$ is eventually independent of the choice of a complete orthonormal system $\{\varphi_n\}$. A member X of $\mathscr{H}_n$ is referred to as a **multiple Wiener integral** of degree n and is expressed in the form

$$X=\int\cdots\int_{\mathbf{R}^n} f(s_1,s_2,\ldots,s_n)dB_{s_1}dB_{s_2}\ldots dB_{s_n},$$

where f is a symmetric $L^2(\mathbf{R}^n)$-function. In addition, $E(X^2)=n!\|f\|^2$, $\|\ \|$ being the $L^2(\mathbf{R}^n)$-norm.

Brownian functionals can be expressed as functionals of white noise, so that the Hilbert space $L^2(\mu)$ is viewed as a realization of (L^2), where μ is the probability distribution of the Gaussian white noise introduced in Section D.

Nonlinear functionals of a general Gaussian process can be dealt with in a similar but somewhat more complicated manner. If the process has a canonical representation, then the functionals can easily be rewritten as Brownian functionals.

J. Applications to Prediction Theory

If a Gaussian process $\{X_t | t\geqslant a\}$ has the canonical representation

$$X_t=\int_a^t F(t,u)dB_u,$$

then the best †predictor $E(X_t/\mathbf{B}_s(X))$, $s<t$, is given by

$$\int_a^s F(t,u)dB_u.$$

This is in agreement with the best †linear predictor.

No systematic approach for nonlinear prediction theory has been established so far. We illustrate this theory only in a typical case that arises from a stationary Gaussian process. Let $\{X_t\}$ be a real-valued stationary Gaussian process. Without loss of generality, we can assume that $E(X_t)=0$. We consider the continuous parameter case, since the discrete parameter case is easier and can be inferred from it.

Let $\mathbf{M}_t(X)$ be the subspace of $L^2(\Omega, P)$ de-

fined in the same way as in Section E, and let $\mathbf{L}_t = L^2(\Omega, \mathbf{B}_t(X))$. If the process is purely non-deterministic in the sense that $\bigcap_t \mathbf{M}_t(X) = \{0\}$, then X_t has the backward moving average representation

$$X_t = \int_{-\infty}^{t} f(t-s)\, dB_s,$$

where $\{B_t\}$ is a Brownian motion. Furthermore, this representation is canonical, i.e., $\mathbf{B}_t(X) = \mathbf{B}_t(B)$. Since $\mathbf{B}$ is assumed to be $\bigvee_t \mathbf{B}_t$, every Y in (L^2) can be expressed in the form

$$Y = \sum_{p=0}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{s_3} \int_{-\infty}^{s_2} f_p(s_1, s_2, \ldots, s_p)\, dB_{s_1} dB_{s_2} \ldots dB_{s_p},$$

in terms of multiple Wiener integrals. The predictor of Y based on the known values $X_s, s \leqslant 0$, that is, the projection of Y on $\mathbf{L}_0$ is given by truncating the domain of the integral as

$$\sum_{p=0}^{\infty} \int_{-\infty}^{0} \cdots \int_{-\infty}^{s_3} \int_{-\infty}^{s_2} f_p(s_1, s_2, \ldots, s_p)\, dB_{s_1} dB_{s_2} \ldots dB_{s_p}.$$

K. Extrapolation and Interpolation

Extrapolation of a stationary Gaussian process is used to find the best estimate of the unknown values of a given process when we have been given the only some of the past values of the process. M. G. Kreĭn [8] obtained some results by using the theory of the inverse spectral problem discussed by I. M. Gel'fand and B. Levitan.

In contrast to extrapolation is the interpolation of a stationary Gaussian process. There is an important contribution to this problem by H. Dym and H. P. McKean [1] who developed the Kreĭn theory by applying the de Branges method to the theory of entire functions.

References

[1] H. Dym and H. P. McKean, Gaussian processes, function theory, and the inverse spectral problem, Academic Press, 1976.
[2] X. Fernique, Continuité des processus gaussiens, C. R. Acad. Sci. Paris, 258 (1964), 6058–6060.
[3] T. Hida, Canonical representations of Gaussian processes and their applications, Mem. Coll. Sci. Univ. of Kyoto, (A, Math.) 34 (1960), 109–155.
[4] M. Hitsuda, Representation of Gaussian processes equivalent to Wiener processes, Osaka J. Math., 5 (1968), 299–312.
[5] G. A. Hunt, Random Fourier transforms, Trans. Amer. Math. Soc., 71 (1951), 38–69.
[6] I. A. Ibragimov and Yu. A. Rozanov, Gaussian random processes, Springer, 1978. (Original in Russian, 1970.)
[7] K. Itô, Multiple Wiener integral, J. Math. Soc. Japan, 3 (1951), 157–169.
[8] M. G. Kreĭn, On the fundamental approximation problem in the theory of extrapolation and the filtration of stationary random processes, Selected Transl. in Math. Stat. Prob., 4 (1963), 127–131. (Original in Russian, 1954.)
[9] P. Lévy, Processus stochastiques et mouvement brownien, Gauthier-Villars, 1948; second edition, 1965.
[10] P. Lévy, A special problem of Brownian motion and a general theory of Gaussian random functions, Proc. 3rd Berkeley Symp. Math. Stat. Prob. II, Univ. of Calif. Press, 1956, 133–175.
[11] R. M. Dudley, The size of compact subset of Hilbert space and continuity of Gaussian processes, J. Functional Anal., 1 (1967), 290–330.
[12] H. P. McKean, Jr., Brownian motion with a several-dimensional time, Teor. Veroyatnost i Primenen., 8 (1963), 357–378.
[13] E. Nelson, Construction of quantum fields from Markov fields, J. Functional Anal., 12 (1973), 97–112.
[14] Y. Okabe, On the structure of splitting fields of stationary Gaussian processes with finite multiple Markovian property, Nagoya Math. J., 54 (1974), 191–213.
[15] Yu. A. Rozanov, Infinite-dimensional Gaussian distribution, Proc. Steklov Inst. Math., 108 (1971). (Original in Russian, 1968.)
[16] M. Yor, Existence et unicité de diffusion à valeurs dans un espace Hilbert, Ann. Inst. H. Poincaré, (B) 10 (1974), 55–88.
[17] Y. Okabe, Stationary Gaussian processes with Markovian property and M. Sato's hyperfunctions, Japan J. Math., 41 (1973), 69–122.
[18] X. Fernique, Regularité des trajectories des fonctions aléatoires gaussienes, Lecture notes in math. 480, Springer, 1974, 1–96.

177 (XIV.2) Generating Functions

A. General Remarks

A power series $g(t) = \sum_{n=0}^{\infty} a_n t^n$ in t that converges in a certain neighborhood of $t=0$ defines the sequence of numbers $\{a_n\}$. The function $g(t)$ is called the **generating function** of

the sequence. Similarly, the series $K(x,t)=\sum_{n=0}^{\infty}f_n(x)t^n$ that is convergent for x and t in a certain domain in (x,t)-space is called the **generating function** of the sequence of functions $\{f_n(x)\}$. Sometimes the function $g(t)=\sum_{n=0}^{\infty}(a_n/n!)t^n$ is called the **exponential generating function** of the sequence $\{a_n\}$. For example, the generating functions of the †binomial coefficients and †Legendre polynomials are $(1+t)^n$ and $(1-2tx+t^2)^{-1/2}$, respectively. When a generating function of $\{a_n\}$ or $\{f_n(x)\}$ is given, we can obtain a_n or $f_n(x)$ by integral expressions; for example, in the latter case we have

$$f_n(x)=\frac{1}{2\pi i}\int_C\frac{K(x,t)}{t^{n+1}}dt,$$

where the contour C is a sufficiently small circle going counterclockwise around the origin. A generating function can be continued beyond the domain of convergence of the power series. Simple generating functions are known for many important orthogonal systems of functions (→ 317 Orthogonal Functions). Generating functions are widely used because they enable us to derive analytically the properties of sequences of numbers or functions. For a system of numbers or functions depending on a continuous parameter instead of the integral parameter n, we define the generating function in the form of a †Laplace or †Fourier transform. In particular, for a probability †distribution function $F(x)$, the exponential generating function $f(t)=\int_{-\infty}^{\infty}e^{-tx}dF(x)$ for the moments $\{a_n\}$ of $F(x)$ is called the **moment-generating function** of $F(x)$. There is another kind of generating function for the sequence a_n: $\sum a_n n^{-s}$ in the form of †Dirichlet series; it is frequently used in number-theoretic problems.

B. Bernoulli Polynomials

A system of polynomials

$$B_n(x)=\sum_{k=0}^{n}\binom{n}{k}B_k(0)x^{n-k}$$

is defined by the generating function

$$\frac{te^{xt}}{e^t-1}=\sum_{n=0}^{\infty}B_n(x)\frac{t^n}{n!}.$$

$B_n(x)$ is called the **Bernoulli polynomial** of degree n. Since $B_k(0)$ is the coefficient of $t^k/k!$ in

$$\frac{t}{e^t-1}=\sum_{n=0}^{\infty}B_n(0)\frac{t^n}{n!},$$

we have $B_0(0)=1$, $B_1(0)=-1/2$, $B_2(0)=1/6,\ldots$; $B_{2n+1}(0)=0$ for $n\geqslant 1$; and $(-1)^{n-1}B_{2n}(0)>0$ for $n\geqslant 1$. The nth **Bernoulli number** B_n (→ Appendix B, Table 3) is usually defined by $|B_n(0)|$. Sometimes other definitions, such as $B_n=B_n(0)$ or $B_n=B_{2n}(0)$, are used. Bernoulli polynomials satisfy the relations

$$B_n(x+1)-B_n(x)=nx^{n-1},$$

$$dB_n(x)/dx=nB_{n-1}(x),$$

which are used in †interpolation. For example, a polynomial solution of the †difference equation $f(x+1)-f(x)=\sum_{n=0}^{m}a_nx^n$ is given by $f(x)=\sum_{n=0}^{m}a_nB_{n+1}(x)/(n+1)+$(arbitrary constant). In particular, we have $1^n+2^n+\ldots+p^n=(B_{n+1}(p+1)-B_{n+1}(1))/(n+1)$.

C. Euler Polynomials

A system of polynomials

$$E_n(x)=\sum_{k=0}^{\infty}\binom{n}{k}a_kx^{n-k}$$

is defined by the generating function

$$\frac{2e^{xt}}{e^t+1}=\sum_{n=0}^{\infty}E_n(x)\frac{t^n}{n!}.$$

We call $E_n(x)$ the **Euler polynomial** of degree n. Here a_k is defined by $a_k=E_k(0)$ and

$$\frac{2}{e^t+1}=\sum_{n=0}^{\infty}a_n\frac{t^n}{n!},$$

so that we have $a_0=1$, $a_1=-1/2$, $a_3=1/4,\ldots$; $a_{2n}=0$ for $n\geqslant 1$. The nth Euler number is sometimes taken as a_n, but more often it is defined by

$$E_n=(-1)^n\sum_{\mu=0}^{n}2^\mu\binom{n}{\mu}a_\mu,$$

i.e., by

$$\frac{2}{e^x+e^{-x}}=\operatorname{sech}x=\sum_{n=0}^{\infty}(-1)^n\frac{E_n}{n!}x^n$$

(→ Appendix B, Table 3). All the E_n are integers, $E_{2m+1}=0$ $(m=0,1,\ldots)$ and $E_{2m}>0$ $(m=0,1,\ldots)$; in the decimal expressions for E_n the last digit is 5 for E_{4m} $(m\geqslant 1)$ and 1 for E_{4m+2} $(m\geqslant 0)$. Sometimes E_{2n} is denoted by E_n. We have the relations

$$\sec x=\sum_{n=0}^{\infty}\frac{E_nx^n}{n!},$$

$$E_n(x)+E_n(x+1)=2x^n,$$

$$E_n(1-x)=(-1)^nE_n(x),$$

$$\frac{dE_n(x)}{dx}=nE_{n-1}(x),$$

and in particular,

$$-1^n+2^n-3^n+4^n-\ldots+(-1)^pp^n$$
$$=((-1)^pE_n(p+1)-E_n(1))/2.$$

D. Application to Combinatorics

Let $p(n)$ denote the number of ways of dividing n similar objects into nonempty class. This is called the **number of partitions** of n. Euler noticed that the following formula is valid for the generating function of $p(n)$:

$$1+\sum_{n=1}^{\infty} p(n)x^n=\left[\prod_{n=1}^{\infty}(1-x^n)\right]^{-1},$$

whence we obtain $p(n)=p(n-1)+p(n-2)-p(n-5)-p(n-7)+p(n-12)+p(n-15)-\ldots=\sum_{k=1}^{\infty}(-1)^{k-1}[p(n-(3k^2-k)/2)+p(n-(3k^2+k)/2)]$ (→ 328 Partitions of Numbers). Let $B(n)$ denote the number of ways of dividing n completely dissimilar objects into nonempty classes. These are called **Bell numbers**. The first few are 1, 1, 2, 5, 15, 52, 203, 877, 4140, 21147, ... [2]. For this sequence with $B(0)=1$, the formal generating function $\sum_{n=0}^{\infty} B(n)x^n$ does not converge except at $x=0$. However, the exponential generating function $g(x)=\sum_{n=0}^{\infty} B(n)(n!)^{-1}x^n$ is convergent for all complex numbers x and is equal to e^{e^x-1}. The differential equation $g'=e^x g$ gives rise to a recursive formula

$$B(n+1)=\sum_{k=0}^{n}\binom{n}{k}B(k).$$

References

[1] E. Netto, The theory of substitutions and its applications to algebra, revised and translated by F. N. Cole, second edition, Chelsea, 1964. (Original in German, 1882.)
[2] E. T. Bell, Algebraic arithmetic, Amer. Math. Soc., 1927.
[3] E. B. McBride, Obtaining generating functions, Springer, 1971.
Also → references to 389 Special Functions.

178 (VII.6) Geodesics

A. General Remarks

In the study of global differential geometry, geodesics play an essential role because the behavior of geodesics on a complete Riemannian m-manifold ($m\geqslant 2$) M without boundary heavily influences its topological structure. A significant merit of using geodesics is that one can make elementary and intuitive observations on M (as in Euclidean geometry), which often yields fruitful results.

Riemannian structure induces in a natural way a distance function d on M, and hence M is a metric space. A smooth curve γ of constant speed is by definition a **geodesic** if and only if every point on γ is contained in a nontrivial subarc whose length realizes the distance between its endpoints. M is equipped with the Levi-Civita connection ∇ (→ 80 Connections) by means of which a geodesic is characterized as an autoparallel curve. In local coordinates, γ is obtained as a solution of an ordinary differential equation of order two (→ 80 Connections L). It follows from the properties of solutions of the equation for geodesics that every point $p\in M$ has a neighborhood in which every two points can be joined by a unique shortest geodesic that depends smoothly on both of the endpoints. A classical result of J. H. C. Whitehead (*Quart. J. Math.*, 3 (1932)) states that every point $p\in M$ has a metric ball B centered at p such that every two points in B are joined by a unique shortest geodesic whose image is contained in B. Let M_p be the tangent space to M at p, and let $\tilde{M}_p\subset M_p$ be the set of all vectors $v\in M_p$ such that the solution of geodesic γ with initial condition $\gamma(0)=p$, $\dot{\gamma}(0)=v$ is well defined on $[0,1]$. $\tilde{M}_p$ contains an open neighborhood of the origin and is †star-shaped with respect to the origin. The exponential mapping $\exp_p:\tilde{M}_p\rightarrow M$ is defined to be $\exp_p v=\gamma(1)$. This mapping $\exp_p$ is smooth and has the maximal rank at the origin. Small balls around p are obtained as the image under the exponential mapping of the corresponding balls in $\tilde{M}_p$ centered at the origin, and the restriction of $\exp_p$ to these balls are diffeomorphisms. Thus the topology of M as a metric space is equivalent to the original one of M. The fundamental Hopf-Rinow theorem (*Comment. Math. Helv.*, 3 (1931)) states that (1) M is complete as a metric space if and only if $\tilde{M}_p=M_p$ for some $p\in M$; (2) M is complete if and only if every closed metric ball is compact; and (3) if M is complete, then every two points can be joined by a shortest geodesic, namely, a geodesic with length realizing the distance between the endpoints.

In the following discussion M is assumed to be connected, complete, and without boundary. Let $\gamma:[0,1]\rightarrow M$ be a geodesic. A piecewise smooth 1-parameter variation V along γ is a continuous mapping $V:[0,1]\times(-\varepsilon,\varepsilon)\rightarrow M$ with a finite partition $0=t_0<t_1<\ldots<t_k=1$ such that $V|[t_i,t_{i+1}]\times(-\varepsilon,\varepsilon)$ is smooth and $V(t,0)=\gamma(t)$ for all $t\in[0,1]$. Then the first and second variation formulas for V are

$$L'(0)=\langle Y,\dot{\gamma}\rangle|_0^1/L(\gamma)$$

and

$$L''(0)=\sum_{i=1}^{k}\left[\int_{t_{i-1}}^{t_i}(\langle Y_i',Y_i'\rangle-\langle R(Y_i,\dot{\gamma})\dot{\gamma},Y_i\rangle)dt\right.$$
$$\left.+\langle\nabla_{\partial/\partial s}Y_i,\dot{\gamma}\rangle|_{t_{i-1}}^{t_i}\right]\Big/L(\gamma),$$

respectively, where $Y_i(t)=\partial V(t,s)/\partial s|_{s=0}$ for $t\in[t_{i-1},t_i]$ is the variation vector field associated with $V|[t_{i-1},t_i]\times(-\varepsilon,\varepsilon)$, $L(s)$ is the length of the variation curve $t\to V(t,s)$, and $Y_i'=\nabla_{\dot\gamma}Y_i$ and R is the curvature tensor.

A smooth vector field J along a geodesic $\gamma:[0,1]\to M$ is called a **Jacobi field** if and only if it satisfies $J''+R(J,\dot\gamma)\dot\gamma=0$. The set of all Jacobi fields along γ forms a vector space isomorphic to $M_{\gamma(0)}\times M_{\gamma(0)}$ by the natural correspondence $J\to(J(0),J'(0))$. Every Jacobi field is associated with a 1-parameter geodesic variation V along γ. Namely, every variation curve of V is a geodesic, V is smooth, and $J(t)=\partial V(t,s)/\partial s|_{s=0}$ for all $t\in[0,1]$. Conversely, the variation vector field of any 1-parameter geodesic variation V along γ is a Jacobi field. Especially, if A is a tangent vector to M_p at $\dot\gamma(0)$, then $d(\exp_p)_{\dot\gamma(0)}A=J(1)$, where J is the Jacobi field associated with the 1-parameter geodesic variation $V(t,s)=\exp_p t(\dot\gamma(0)+sA)$, and A is identified under the canonical parallel translation in M_p. A point $\gamma(t_0)$ is said to be **conjugate** to $p=\gamma(0)$ along γ if and only if there exists a nontrivial Jacobi field along γ that vanishes at 0 and t_0. A theorem of Morse and Schoenberg states that if the sectional curvature K of M satisfies $0<A\leqslant K\leqslant B<\infty$, then every unit speed geodesic $\gamma:[0,l]\to M$ has a point conjugate to $\gamma(0)$ if $l\geqslant\pi/\sqrt{A}$ and has no point conjugate to $\gamma(0)$ if $l<\pi/\sqrt{B}$. Especially, every geodesic γ on M with $K\leqslant0$ contains no conjugate point on it, and $\exp_p$ is locally regular for any $p\in M$. Thus a complete and simply connected M with $K\leqslant0$ is diffeomorphic to $\mathbf{R}^m$ (Hadamard and Cartan).

Let γ be a unit speed geodesic with $\gamma(0)=p$. If $t_1>0$ satisfies $d(p,\gamma(t))=t$ for all $t\in[0,t_1)$ and $d(p,\gamma(t))<t$ for all $t>t_1$, then the point $\gamma(t_1)$ is called a **cut point** to p along γ. It appears no later than the first conjugate point to p along γ. The set $C(p)$ of all cut points to p is called the **cut locus** of p. Let U be the set of vectors $tv\in M_p$, $0\leqslant t<1$, where $\exp_p v$ is the cut point to p along $\gamma(t)=\exp_p tv$. U is a nonempty open set and $\exp_p|U:U\to M-C(p)$ is a diffeomorphism, where $M-C(p)$ is diffeomorphic to an m-disk. The cut locus possesses essential information on the topology of M.

The **metric comparison theorems** of H. Rauch (*Ann. Math.*, (2) 54 (1951)) and M. Berger (*Illinois J. Math.*, 6 (1962)) are stated as follows: Let $\inf_M K=\delta>-\infty$ and $\sup_M K=\Delta<\infty$, and let $\gamma:[0,1]\to M$ be a nontrivial geodesic. Let $M(c)$ be a complete and simply connected space form of constant curvature c. Fix geodesics $\gamma_\delta,\gamma_\Delta:[0,1]\to M(\delta)$, $M(\Delta)$, respectively, with the same speed as γ. If J, J_δ, and J_Δ are Jacobi fields along γ, γ_δ, and γ_Δ respectively such that $J(0)=J_\delta(0)=J_\Delta(0)=0$ and $\|J'(0)\|=\|J_\delta'(0)\|=\|J_\Delta'(0)\|$, then $\|J_\Delta(t)\|\leqslant\|J(t)\|\leqslant\|J_\delta(t)\|$ holds for all $t\in[0,t_\Delta]$, where $\gamma_\Delta(t_\Delta)$ is the first conjugate point to $\gamma_\Delta(0)$ along γ_Δ (Rauch). If the Jacobi fields satisfy $\|J(0)\|=\|J_\delta(0)\|=\|J_\Delta(0)\|$ and $J'(0)=J_\delta'(0)=J_\Delta'(0)=0$, then $\|J_\Delta(t)\|\leqslant\|J(t)\|\leqslant\|J_\delta(t)\|$ holds for all $t\in[0,t_\Delta']$, where t_Δ' is the first zero point of J_Δ (Berger). They are often applied to compare curve lengths as follows. Let $c:I\to\mathbf{R}^m$ be a piecewise smooth curve with $\|c(t)\|\leqslant t_\Delta$. Fix points $p\in M$, $p_\delta\in M(\delta)$, and $p_\Delta\in M(\Delta)$ with fixed isometric identifications of the tangent spaces with $\mathbf{R}^m$. Then $L(\exp_{p_\Delta}\circ c)\leqslant L(\exp_p\circ c)\leqslant L(\exp_{p_\delta}\circ c)$. If P, P_δ, and P_Δ are unit parallel fields along corresponding geodesics σ, σ_δ, $\sigma_\Delta:I\to M$, $M(\delta)$, $M(\Delta)$, respectively, which have the same speed, and if $\langle P,\dot\sigma\rangle=\langle P_\delta,\dot\sigma_\delta\rangle=\langle P_\Delta,\dot\sigma_\Delta\rangle$, then for every piecewise smooth function $f:I\to[0,t_\Delta']$ the curves $u(s)=\exp_{\sigma(s)}f(s)P(s)$ and $u_\delta(s)$, $u_\Delta(s)$ defined in the same way on $M(\delta)$, $M(\Delta)$, respectively, have lengths $L(u_\Delta)\leqslant L(u)\leqslant L(u_\delta)$. They play important roles in the theory of geodesics.

A geodesic triangle is a triple $(\gamma_0,\gamma_1,\gamma_2)$ of shortest geodesics (for convenience they are parametrized on $[0,1]$) such that $\gamma_i(1)=\gamma_{i+1}(0)$ for all $i=0,1,2$ with mod 3. The vertices p_0, p_1, p_2 and the angles θ_0, θ_1, θ_2 of a geodesic triangle $(\gamma_0,\gamma_1,\gamma_2)$ are defined by $p_{i+1}=\gamma_i(0)$ and $\theta_{i+1}=\cos^{-1}\{\langle\dot\gamma_i(0),-\dot\gamma_{i-1}(1)\rangle/\|\dot\gamma_i\|\,\|\dot\gamma_{i-1}\|\}$, where angles are always taken in $[0,\pi]$. The **triangle comparison theorem** of Toponogov (*Amer. Math. Soc. Transl.*, 37 (1964)) is stated as follows. Let $K\geqslant\delta>-\infty$ be satisfied on M. For any geodesic triangle $(\gamma_0,\gamma_1,\gamma_2)$ on M there is a geodesic triangle $(\tilde\gamma_0,\tilde\gamma_1,\tilde\gamma_2)$ on $M(\delta)$ such that $L(\tilde\gamma_i)=L(\gamma_i)$ and the angles $\tilde\theta_i$ of this triangle satisfy $\tilde\theta_i\leqslant\theta_i$ for $i=0,1,2$. It turns out that if $\delta>0$ then the circumference of any geodesic triangle on M does not exceed $2\pi/\sqrt{\delta}$.

For details of the basic facts stated above we refer the reader to [1–3].

B. Curvature and Fundamental Groups

The fundamental group $\pi_1(M)$ of M(→ 170 Fundamental Groups) is influenced by the curvature of M. A basic idea, going back to Hadamard (*J. Math. Pures Appl.*, 4 (1898)) and Cartan, states that every nontrivial †free homotopy class $\delta\in\pi_1(M)$ of loops on a compact M contains a closed geodesic of minimal length whose preimage under the covering projection $\pi:\tilde M\to M$ in the universal Riemannian covering $\tilde M$ is either closed (when δ is of finite order) or a straight line (when δ is of infinite order) that is translated along itself by the †deck transformation δ.

(1) By using the second variational formula, an even-dimensional compact M with $K>0$

either is simply connected or has $\pi_1(M)=\mathbf{Z}_2$ (Synge, *Quart. J. Math.*, 7 (1936)). A beautiful result of Myers (*Duke Math. J.*, 8 (1941)) states that if the Ricci curvature Ric of M is bounded below by $\delta>0$, then the diameter $d(M)$ is not greater than $\pi/\sqrt{\delta}$ and hence M is compact and $\pi_1(M)$ is finite. It is proved in [3] that if $K\geqslant 0$, then there exists a finite normal subgroup Φ of $\pi_1(M)$ such that $\pi_1(M)/\Phi$ is a Bieberbach group (→ 92 Crystallographic Groups). The splitting theorem (→ Section F(2)) is used in the proof.

(2) When $K\leqslant 0$ is satisfied on M, $\tilde{M}$ is diffeomorphic to $\mathbf{R}^m$, and hence $\pi_i(M)=0$ for all $i\geqslant 2$. A basic tool used in the study of $\pi_1(M)$ in this case is that the displacement function $\tilde{x}\to\tilde{d}(\tilde{x},\delta(\tilde{x}))$, $\tilde{x}\in\tilde{M}$, of every isometry δ on $\tilde{M}$ is convex [4], where a function on M is said to be **convex** if and only if its restriction to every geodesic is convex. A classical result of Preissmann (*Comment. Math. Helv.*, 15 (1943)) states that every nontrivial Abelian subgroup of $\pi_1(M)$ of a compact M with $K<0$ is an infinite cyclic subgroup. It is proved in [4] that if $K<0$ holds on M and if every deck transformation on $\tilde{M}$ translates some geodesic along itself, then $\pi_1(M)$ is a disjoint union of infinite cyclic subgroups, and any two commuting elements belong to the same cyclic subgroup. Moreover, if M is compact and $K\leqslant 0$ and if $\pi_1(M)$ is a direct product $\Gamma_1\times\Gamma_2$ such that $\pi_1(M)$ is †centerless, then M is isometric to the Riemannian product $M_1\times M_2$, and $\pi_1(M_i)=\Gamma_i$ holds for $i=1,2$ [5, 6]. It is shown in [7] that the fundamental group of M with $K\leqslant 0$ occurs as that of an $(m+1)$-dimensional M' with $K\leqslant c<0$, which is diffeomorphic to $M\times\mathbf{R}$. The warped product [4] is used to construct such a metric on $M\times\mathbf{R}$.

C. Cut Locus

The **injectivity radius** $i(M)$ of M is defined to be the infimum of the continuous function $x\to d(x,C(x))$, $x\in M$. As is seen in Section D, the estimate of injectivity radii provides many fruitful results on the topology of Riemannian manifolds. It follows from Synge's result that an even dimensional compact and orientable M with $0<K\leqslant 1$ has its injectivity radius $i(M)\geqslant\pi$ [3]. However, such an estimate cannot be obtained in odd dimensions. The examples discussed in [8] show that there are infinitely many homotopically distinct homogeneous Riemannian 7-manifolds of positive curvature. The injectivity radii of such examples are estimated precisely in [9], according to which there is no positive lower bound for them. The **sphere theorem** of Klingenberg (*Comment. Math. Helv.*, 35 (1961)) states that if M is compact and simply connected and if $1/4<K\leqslant 1$, then $i(M)\geqslant\pi$, and M is a topological sphere. This extends the pioneering work by H. Rauch (*Ann. Math.*, (2) 54 (1951)). A **rigidity theorem** of M. Berger (*Ann. Scuola Norm. Sup. Pisa*, 14 (1962)) states that if $\dim M$ is even, M is simply connected, and $1/4\leqslant K\leqslant 1$, then M is homeomorphic to S^m (if $d(M)>\pi$) or isometric to one of the symmetric spaces of compact type of rank 1 (if $d(M)=\pi$). A slight generalization of the triangle comparison theorem is used to prove that for given positive constants d, V, and S, there exists a constant $C_m(d,V,S)>0$ such that if K, $d(M)$ and the volume $v(M)$ satisfy $|K|<S$, $d(M)<d$, and $v(M)>V$, then $i(M)\geqslant C_m(d,V,S)$ [10].

D. Finiteness Theorem

Finiteness theorems are a natural extension of the sphere theorem. They provide *a priori* estimates for the number of various topological types of manifolds (for instance, homology, homotopy, homeomorphism, diffeomorphism types, etc.), which admit certain classes of Riemannian metrics characterized by geometric quantities. The basic idea of the estimates is to find a constant $c>0$ that depends only on geometric information which characterizes the class of manifolds so that if M belongs to the class, then $i(M)\geqslant c$. Then a number N is found from the information given *a priori* such that every element M of the class has an open cover of at most N balls whose radii are all less than c. It is proved in [11] that for given $\delta\in(0,1)$ and m, there exists a number $N(\delta,m)$ such that there are at most $N(\delta,m)$ homotopy types for the class of all simply connected $2m$-dimensional manifolds with $\delta\leqslant K\leqslant 1$. Furthermore, for given positive numbers d, V, and S, there are at most finitely many homeomorphism (or diffeomorphism) types for the class of all m-manifolds M, each of which has the property that $d(M)<d$, $v(M)>V$, and $|K|<S$ [10]. This result is applied to obtain a result in [12], which states that there exists for given m, $V>0$ and an $\varepsilon>0$ such that if M is compact, $-1-\varepsilon\leqslant K\leqslant -1$, and $v(M)>V$, then M admits a metric of constant negative curvature.

For an $\varepsilon\geqslant 0$, M is said to be **ε-flat** if and only if $\sup_M K\cdot d(M)^2\leqslant\varepsilon$ is satisfied for M. Every compact flat manifold is ε-flat for all $\varepsilon\geqslant 0$. M is called a **nilmanifold** if and only if it admits a transitive action of a †nilpotent Lie group. It is shown in [13] that for a compact M with $\dim M=m$, there exists a number $\varepsilon(m)>0$ such that if M is $\varepsilon(m)$-flat, then (1) there is a maximal nilpotent normal divisor $N\subset\pi_1(M)$, (2)

the order of $\pi_1(M)/N$ is bounded by a constant which depends only on m, and (3) the finite cover of M which corresponds to N is diffeomorphic to a nilmanifold. Moreover, if M is $\varepsilon(m)$-flat then M is diffeomorphic to $\mathbf{R}^m$. If M is $\varepsilon(m)$-flat and if $\pi_1(M)$ is commutative, then M is diffeomorphic to a †torus.

E. Uniqueness Theorem

Uniqueness of topological structures (as in the sphere theorem) of certain classes of compact Riemannian manifolds is discussed here.

The discovery of †exotic 7-spheres by J. Milnor (*Ann. Math.*, (2) 64 (1956)) gave rise to the question of whether in the sphere theorem the conclusion (homeomorphism to the sphere) could be replaced by diffeomorphism. Since the number of differentiable structures on a topological sphere depends on its dimension (→ 114 Differential Topology), it has been thought that in order to get the standard sphere in the sphere theorem the best possible restriction for the curvature might also depend upon dimension. The appropriate **differentiable pinching problem** is to find a sequence $\{\Delta_m\}$ so that if M is compact and simply connected and if $\delta \leqslant K \leqslant 1$ for some $\delta > \Delta_m$, then M is diffeomorphic to S^m, and Δ_m is the least possible with this property. D. Gromoll and Y. Shikata proved independently [14, 15] that $\Delta_m < 1$ holds for all $m \geqslant 2$. Later it was proved that there exists a $\delta_0 \in (1/4, 1)$ such that $\Delta_m \leqslant \delta_0$ holds for all $m \geqslant 2$ [16]. The **diffeotopy theorem**, which plays an essential role in the proof of finding such a δ_0, provides a sufficient condition for a diffeomorphism on S^{m-1} to be †isotopic to the identity mapping.

A different idea is put forth in [17], which imitates the Gauss normal mapping of a closed convex hypersurface in $\mathbf{R}^{m+1}$. It is proved that if the curvature operator is sufficiently closed to the identity, then M is diffeomorphic to S^m. This idea is used to obtain a diffeomorphism between M and a spherical space form. A generalization of the sphere theorem states that if M is compact and if $(d(M)/\pi)^2 \inf_M K > 1/4$, then M is a topological sphere [18].

F. Noncompact Manifolds

Let M be noncompact. It is due to the nature of noncompactness that through each point on M there passes a **ray** $\gamma: [0, \infty) \to M$, e.g., γ is a unit speed geodesic any of whose subarc is minimizing.

(1) Busemann Functions. A **Busemann function** $F_\gamma: M \to R$ with respect to a ray γ is defined by $F_\gamma(x) = \lim_{t \to \infty}[t - d(x, \gamma(t))]$. The original definition of it was used by H. Busemann to define a parallel axiom on straight G-spaces (→ Section H). F_γ has the property $F_\gamma^{-1}((-\infty, t_1]) = \{x \in F_\gamma^{-1}((-\infty, t_2]) \mid d(x, \partial F_\gamma^{-1}((-\infty, t_2])) \geqslant t_2 - t_1\}$ for any $t_2 \geqslant t_1$, and hence $F_\gamma(x) = t_2 - d(x, \partial F_\gamma^{-1}((-\infty, t_2]))$ for all t_2 and x with $F_\gamma(x) \leqslant t_2$. It has been proved, by using the second variation formula, that if $\mathrm{Ric} \geqslant 0$, $K \geqslant 0$ [19] or if, in the case where M is Kählerian, the holomorphic bisectional curvature [20] is nonnegative, then F_γ is subharmonic (sh), convex or †plurisubharmonic (psh), respectively, where the holomorphic bisectional curvature is defined as $R(X, JX, JY, Y)$ for the complex structure J and orthonormal vectors X and Y. If the holomorphic bisectional curvature [20] is positive, then F_γ is strictly psh. If $K \geqslant 0$, then the triangle comparison theorem implies that $F = \sup\{F_\gamma \mid \gamma(0) = p\}$ is convex and is an exhaustion [21], where the sup is taken over all rays emanating from a point $p \in M$, and a function f is said to be an **exhaustion** if and only if $f^{-1}((-\infty, a])$ is compact for all $a \in \mathbf{R}$. A well-known theorem of H. Grauert (*Math. Ann.*, 140 (1960)) states that if M admits a strictly psh exhaustion function, then M is a †Stein manifold (→ 21 Analytic Functions of Several Complex Variables). In this context, various conditions for curvatures on Kähler manifolds under which they become Stein manifolds have been found [20].

Let H be complete and simply connected, and $K \leqslant 0$. Busemann functions on H are differentiable of class C^2, and the †horosphere $F_\gamma^{-1}(\{t\})$ is a C^2-surface [22]. On a parabolic visibility manifold (→ Section F(3)) the negative of every Busemann function is C^1-convex without minimum [7].

(2) Ends and Splitting Theorems. Ends of a noncompact M are defined as follows: If A_1 and A_2 are compact subsets of M with $A_1 \subset A_2$, then any component of $M - A_2$ is contained in a unique component of $M - A_1$. An **end** is the limit of an inverse system $\{$components of $M - A; A\}$ directed by the inclusion relation as indicated above and indexed by $\{A \mid A$ compact$\}$.

It has been shown that if $\mathrm{Ric} > 0$ [23] (or $\geqslant 0$ [19]), then M has exactly one end (or at most two ends). A visibility manifold has at most two ends if it is not Fuchsian [7]. If M admits a locally nonconstant convex function, then M has at most two ends [24].

If M has more than one end, then there exists a straight line $\gamma: \mathbf{R} \to M$ which is by definition a nontrivial geodesic any of whose subarcs is minimizing. A classical result of Cohn-Vossen (*Mat. Sb.*, 1 (1936)) states that if $\dim M = 2$, $K \geqslant 0$ and if M contains a straight

line, then the total curvature is 0 and hence M is isometric to either a plane or a cylinder $S^1 \times \mathbf{R}$. Generalizations of this result state that if M admits k independent straight lines and if $K \geqslant 0$ (Toponogov, *Amer. Math. Soc. Transl.*, 37 (1964)) (or if $\mathrm{Ric} \geqslant 0$ [19]), then M is isometric to the Riemannian product $N \times \mathbf{R}^k$.

(3) Structure Theorems. The structure theorem of 2-dimensional M with $K>0$ was proved by Cohn-Vossen (*Compositio Math.*, 2 (1935)). It has been shown that M is diffeomorphic to $\mathbf{R}^m$ if $K>0$ [23] and that if $K \geqslant 0$, then there exists a compact totally geodesic submanifold S without boundary such that M is homeomorphic to the total space of the †normal bundle over S [21]. Here, homeomorphism can be replaced by diffeomorphism, and if $K>0$ outside a compact set of M, then M is diffeomorphic to $\mathbf{R}^m$ [26].

For a ray γ and a point x on H, if σ is a ray with $\sigma(0)=x$ and if $d(\gamma(t), \sigma(t))$, $t>0$, is bounded above, then σ is called an **asymptotic ray** (a ray asymptotic) to γ. The asymptotic relation is an equivalence relation on the set of all geodesics on H. A **point at infinity** of H is an equivalence class on geodesics on H, and the set of all points at infinity of H is denoted by $H(\infty)$. With a suitable topology, the set of all points at infinity of H constitutes a bounding sphere such that $\bar{H}=H \cup H(\infty)$ is a closed m-cell. For a properly discontinuous group D of isometries acting on H, a closed D-invariant limit set $L(D)$ is obtained in $H(\infty)$ as the set of all accumulation points of $\delta(p)$, $\delta \in D$. $M=H/D$ is a **visibility manifold** if and only if H satisfies the following axiom: any two distinct points on $H(\infty)$ can be joined by at least one geodesic. If $K \leqslant c<0$ is satisfied on M, then M is a visibility manifold [7]. By investigating limit sets, visibility manifolds are classified into three types [7]. (1) M is parabolic; M is diffeomorphic to $N \times \mathbf{R}$ and is characterized by the fact that it has a convex function without minimum. (2) M is axial; M is a vector bundle over S^1, and hence it is diffeomorphic to either $S^1 \times \mathbf{R}^{m-1}$ or the product of a Möbius strip with $\mathbf{R}^{m-2}$. (3) M is Fuchsian; M has more than two ends, and a strong algebraic restriction is imposed on the fundamental group of M.

(4) Convex Functions. Some elementary properties of C^∞ convex functions on M have been investigated and it was proved in [4] that if M admits a C^∞ convex function without minimum and if $K \leqslant 0$, then M is diffeomorphic to $N \times \mathbf{R}$, where N is a level hypersurface. When $K>0$, F can be replaced by a strongly convex exhaustion function. Namely, for every compact set $A \subset M$ there is a $\delta>0$ such that the second difference quotient along every geodesic at any point on A is bounded below by δ [20]. A standard †convolution smoothing procedure yields smooth convex approximations in neighborhoods of compact sets for every strongly convex function such that their second derivatives along every geodesic is positive. Global approximations can be constructed [27]. Let $\varphi: M \to R$ be a convex function which is not constant on any open set. It has recently been proved that (1) if $b>\inf_M \varphi \geqslant -\infty$, then there is a homeomorphism $H: \varphi^{-1}(\{b\}) \times (\inf_M \varphi, \infty) \to M - \{$minimum set of φ, if any$\}$ such that $\varphi \circ H(y, \alpha)=\alpha$ for all $y \in \varphi^{-1}(\{b\})$ and for all $\alpha \in (\inf_M \varphi, \infty)$; (2) if φ takes a minimum, then the continuous extension $\bar{H}: [\inf_M \varphi, \infty) \to M$ of H is proper and surjective [24].

G. Manifolds All of Whose Geodesics Are Closed

As is well known, every symmetric space of compact type of rank 1 (it is abbreviated CROSS as in [28]) has the property that all geodesics are simply closed and of the same length. Let M be a compact Riemannian manifold with the property that all geodesics on M are simply closed and that they have the same length (SC property). The problem discussed here is whether such an M is isometric (or at least the topology of such an M is equivalent) to some CROSS, or if it is possible to classify all such manifolds.

An example of a nontrivial SC-manifold was first constructed by Zoll (*Math. Ann.*, 57 (1903)) on S^2 as a surface of revolution in $\mathbf{R}^3$, which is not isometric to a standard sphere. Blaschke conjectured that every SC-structure on PR^2 is the standard real projective space. This has been solved affirmatively by L. W. Green (*Ann. Math.*, 78 (1963)). In higher dimensions it has been proved that every infinitesimal deformation of the standard SC-structure on PR^m is trivial [29]. A general result for a SC-manifold M states that the volume of M (with $\dim M=m$) with period 2π is the integral multiple of the volume of the standard unit m-sphere and this integer is a topological invariant [30]. For a point $p \in M$, if every geodesic segment with length l emanating from p is a simple geodesic loop at p, then M is called a **SC^p-manifold**. It has been proved that the †integral cohomology ring of every SC^p-manifold is isomorphic to one of the CROSS [31], which is a generalization of Bott's theorem (*Ann. Math.*, 60 (1954)).

An essential difference between the metrics of a Zoll surface and a standard sphere is seen

on the cut locus and the conjugate locus of a point. The (tangent) cut locus of any fixed point of a CROSS coincides with the (tangent) first conjugate locus; however, this does not hold on a Zoll surface. This observation gives rise to the definition of a Blaschke manifold: For distinct points $p, q \in M$, let $\Lambda(p,q) \subset M_q$ be the set of all unit vectors tangent to the minimizing geodesics from q to p. M is said to be a **Blaschke manifold at a point** p if for every $q \in C(p)$, $\Lambda(p,q)$ is a great sphere of the unit hypersphere in M_q centered at the origin. M is called a **Blaschke manifold** if it is so at every point of it. The following statements are equivalent [28]: (1) M is a Blaschke manifold at p, (2) the tangent cut locus C_p at p is spherical in M_p, (3) along every geodesic starting from p, the first conjugate point p appears at a constant length and the multiplicity of it is independent of the initial direction. Then such an M can be exhibited as $D \cup_a E$, where D is a closed ball with center p, E a C^∞-closed k-disk bundle over an $(m-k)$-dimensional C^∞-compact manifold with boundary ∂E diffeomorphic to S^{m-1}, and $a: \partial D \rightarrow \partial E$ an †attaching diffeomorphism. Conversely, if M is exhibited as $D \cup_a E$ then there exists a metric g on M such that (M,g) is a pointed Blaschke manifold at p, where p is the center of D. A generalized Blaschke's conjecture states that every Blaschke manifold is isometric to a CROSS. A partial solution to this conjecture has been obtained by M. Berger and states that if (S^m, g) is a Blaschke manifold, then it is isometric to a standard sphere [28].

H. *G*-Spaces

G-spaces were created by Busemann to show that many global theorems of differential geometry are independent of the Riemannian character of the metric and also of smoothness. This approach naturally led to novel questions and results. (Many facts hold even when the distance is not symmetric; → [34].) The symbol G suggests that the principal property of G-spaces is the existence of geodesics with the same properties as in complete Riemannian manifolds without boundary except for differentiability. A ***G*-space** X is defined as follows: (1) X is metric with (symmetric) distance d; (2) every bounded infinite set has at least one point of accumulation; (3) given two distinct points $p, r \in X$, there is a point $q \in X$ such that $p \neq q \neq r$ and $d(p,q) + d(q,r) = d(p,r)$; (4) to each point $x \in X$ a positive number ρ_x is assigned such that for any $p, q \in X$ with $d(p,x) < \rho_x$, $d(q,x) < \rho_x$ there is a point r with $d(p,q) + d(q,r) = d(p,r)$; (5) if $d(p,q) + d(q,r) = d(p,r)$ and if $d(p,q) + d(q,r') = d(p,r')$, then $d(q,r) = d(q,r')$ implies $r = r'$.

From (2) and (3) it follows that any two points on X can be joined by a curve called a **segment** whose length realizes the distance between them. A **geodesic arc** (or for simplicity **geodesic**) $\gamma: [a,b] \rightarrow X$ is a curve such that any subarc contained in a ρ_x-ball around some point $x \in X$ is a segment. From (4) and (5) it follows that every geodesic arc has a unique infinite extension to both sides. If a geodesic arc is extended infinitely to both sides, then it is called a **geodesic line**.

The absence of smoothness causes an essential difference in the fact that the distance function to a fixed point on X is not necessarily convex in a small ball around it, in contrast to the Whitehead theorem (→ Section A) for Riemannian manifolds (or †Finsler spaces). Here a function on X is called **convex** if and only if its restriction to every arc length–parametrized geodesic line is convex.

Every two points on X have neighborhoods which are homeomorphic to each other. Thus the topological dimension of X (→ 117 Dimension Theory) is well defined. A 1-dimensional X is either a circle S^1 or the whole real line. If $\dim X = 2$ [33] or if $\dim X = 3$ [36], then X is a topological manifold.

It should be noted that G-space theory not only proves known Riemannian theorems under weaker conditions, but also leads to many facts which were either not considered previously or well understood in (or thought to be very different from) the Riemannian case. Among many fruitful topics discussed by Busemann [33], only two are included below.

(1) *G*-Surfaces. For each point p on a 2-dimensional G-space $\mathscr{S}$, which is called a ***G*-surface**, let $S_p = \{x \in \mathscr{S} \mid d(x,p) = \rho\}$, where $0 < \rho < \rho_p/4$ is a fixed number. S_p is homeomorphic to a circle and bounds a 2-disk. An **angle** A at p is the set of all segments emanating from p and passing through the points on a connected subarc of S_p. An angular measure $|\cdot|$ for the angles at p is a nonnegative function which satisfies: (1) $|A| = \pi$ if and only if p is the midpoint of the segment joining two points on S_p which bounds A; (2) if the intersection of two angles A_1 and A_2 is a unique segment, then $|A_1 \cup A_2| = |A_1| + |A_2|$. If a triple of points (p_0, p_1, p_2) on $\mathscr{S}$ are contained in a sufficiently small ball, then the three segments joining them define a **triangle**, and each vertex p_i has the angle $|p_{i-1}p_ip_{i+1}|$ determined by the two segments. An **excess** $\varepsilon(p_0p_1p_2)$ of a triangle (p_0, p_1, p_2) is defined to be $\varepsilon(p_0p_1p_2) = |p_0p_1p_2| + |p_1p_2p_0| + |p_2p_0p_1| - \pi$. A degenerate triangle has excess 0. If a triangle (a,b,c), which consists of geodesic arcs, is simplicially decomposed

into a finite number of triangles (a_i, b_i, c_i), $i = 1, \ldots, k$, then $\varepsilon(abc) = \sum_{i=1}^{k} \varepsilon(a_i b_i c_i)$, and this is independent of the choice of finite decomposition. If $\mathscr{S}$ is compact and orientable, the **total excess** $\varepsilon(\mathscr{S})$ of $\mathscr{S}$ is defined by $\varepsilon(\mathscr{S}) = \sum_{i=1}^{k} \varepsilon(a_i b_i c_i)$, where $\mathscr{S}$ is simplicially decomposed into nondegenerate finite triangles (a_i, b_i, c_i), $i = 1, \ldots, k$. Then $\varepsilon(\mathscr{S}) = 2\pi\chi(\mathscr{S})$, where $\chi(\mathscr{S})$ is the †Euler characteristic of $\mathscr{S}$. The results obtained by Cohn-Vossen [35] on the †total curvature of complete open surfaces have also been generalized to G-surfaces with angular measure uniform at π. Although no angular measure leading to a true analog of the †Gauss-Bonnet theorem exists even in G-surfaces [34], many of its applications, in particular the results by Cohn-Vossen, can be proved by using angular measures which are uniform at π [33].

If every two points on X can be joined by a unique geodesic line, then all geodesic lines on X are simultaneously either straight lines or circles of the same length l. If in the latter case $\dim X > 1$, then X has a two-fold universal covering space $\tilde{X}$ whose geodesic lines are all circles of the same length $2l$, and they have the property that all geodesics through a point on $\tilde{X}$ meet at the same point at length l. For the Riemannian case it is a Blaschke manifold (→ Section G). If every two points on $\mathscr{S}$ can be joined by a unique geodesic line, then $\mathscr{S}$ is either a plane or a projective plane.

(2) *G*-Spaces with Nonpositive Curvature. Let (p_0, p_1, p_2) be a triple of points on X which are contained in a small ball. X, by definition, is of **nonpositive curvature** if and only if for any such triple of points

$$d(p_1, p_{i+1}) \geqslant 2d(p'_i, p'_{i+1}) \quad \text{for} \quad i = 0, 1, 2, \qquad (*)$$

where p'_i is the midpoint of the unique segment joining p_{i-1} to p_{i+1}. If the above inequality is strict for any nondegenerate small triangle, then X is said to be of **negative curvature**. X of curvature 0 is defined in the same way. It should be noted that a G-space of curvature 0 is a locally Minkowskian space [33]. The sectional curvature of a Riemannian manifold M is nonpositive if and only if (*) holds for all small triangles on it. Because of $d(p_0, p'_0) \leqslant d(p_0, p'_1) + d(p'_1, p'_0) \leqslant \{d(p_0, p_1) + d(p_0, p_2)\}/2$, the distance function to a point $x \in X$ of nonpositive curvature is convex on a small ball around x. X is called **straight** if and only if all nontrivial geodesic lines on X are straight lines. If X has nonpositive curvature, then the universal covering space $\tilde{X}$ is straight. Moreover, for any two geodesic lines $\gamma, \sigma: \mathbf{R} \to X$, the function $t \to d(\sigma(t), \gamma(t))$ is convex. In particular, the distance function to every fixed point on $\tilde{X}$ is convex, and (*) holds for any triangle on $\tilde{X}$. Thus $\tilde{X}$ is contractible. Every nontrivial element of the homotopy class of loops on X with base point $x \in X$ contains a geodesic loop at x which has the minimum length. This fact and the strict convexity of $t \to d(\sigma(t), \gamma(t))$ for any two arc length parametrized geodesic lines γ, σ on the universal covering $\tilde{X}$ of X with negative curvature imply that if X is compact, then every Abelian subgroup of the fundamental group $\pi_1(X)$ of X is infinite cyclic. This is a generalization of the Preissmann theorem (→ Section B).

References

[1] A. D. Alexandrov, Die innere Geometrie der konvexen Flächen, Akademie-Verlag, 1955.
[2] J. Cheeger and D. Ebin, Comparison theorems in Riemannian geometry, North-Holland, 1975.
[3] D. Gromoll, W. Klingenberg, and W. Meyer, Riemannsche Geometrie in Grossen, Lecture notes in math. 55, Springer, 1968.
[4] R. Bishop and B. O'Neill, Manifolds of negative curvature, Trans. Amer. Math. Soc., 145 (1969), 1–49.
[5] D. Gromoll and J. Wolf, Some relations between the metric structure and the algebraic structure of the fundamental group in manifolds of nonpositive curvature, Bull. Amer. Math. Soc., 77 (1971), 545–552.
[6] B. Lawson and S. T. Yau, On compact manifolds of nonnegative curvature, J. Differential Geometry, 7 (1972), 211–228.
[7] P. Eberlein and B. O'Neill, Visibility manifolds, Pacific J. Math., 46 (1973), 45–109.
[8] N. Wallach, Homogeneous Riemannian manifolds of strictly positive curvature, Ann. Math., (2) 75 (1972), 277–295.
[9] T. Sakai, Cut loci of Berger's sphere, Hokkaido Math. J., 10 (1981), 143–155.
[10] J. Cheeger, Finiteness theorems for Riemannian manifolds, Amer. J. Math., 92 (1970), 61–74.
[11] A. Weinstein, On the homotopy type of positively pinched manifolds, Arch. Math., 18 (1967), 523–524.
[12] M. Gromov, Manifolds of negative curvature, J. Differential Geometry, 13 (1978), 223–230.
[13] M. Gromov, Almost flat manifolds, J. Differential Geometry, 13 (1978), 231–241.
[14] D. Gromoll, Differenzierbare Struktur und Metriken positiver Krümmung auf Sphären, Math. Ann., 164 (1966), 353–371.
[15] Y. Shikata, On the differentiable pinching problem, Osaka J. Math., 4 (1967), 279–287.
[16] M. Sugimoto and K. Shiohama, On the

differentiable pinching problem, Math. Ann., 195 (1971), 1–16.
[17] E. Ruh, Curvature and differentiable structures on spheres, Comment Math. Helv., 46 (1971), 127–136.
[18] K. Grove and K. Shiohama, A generalized sphere theorem, Ann. Math., (2) 106 (1977), 201–211.
[19] J. Cheeger and D. Gromoll, The splitting theorem for manifolds of nonnegative Ricci curvature, J. Differential Geometry, 6 (1971), 119–129.
[20] R. Greene and H. Wu, Function theory on manifolds which possess a pole, Lecture notes in math. 699, Springer, 1979.
[21] J. Cheeger and D. Gromoll, On the structure of complete manifolds of nonnegative curvature, Ann. Math., (2) 96 (1972), 413–443.
[22] H. Im Hof and E. Heitze, Geometry of horospheres, J. Differential Geometry, 12 (1977), 481–492.
[23] D. Gromoll and W. Meyer, On complete open manifolds of positive curvature, Ann. Math., (2) 90 (1969), 75–90.
[24] R. Greene and K. Shiohama, Convex functions on complete noncompact manifolds; Topological structures, Inventiones Math., 63 (1981), 129–157.
[25] W. Poor, Some results on nonnegatively curved manifolds, J. Differential Geometry, 9 (1974), 583–600.
[26] H. Wu, An elementary proof in the study of nonnegative curvature, Acta Math., 142 (1979), 57–78.
[27] R. Greene and H. Wu, C^∞ convex functions and manifolds of positive curvature, Acta Math., 137 (1976), 209–245.
[28] A. Besse, Manifolds all of whose geodesics are closed, Springer, 1978.
[29] R. Michel, Problème d'analyse géométrique lies à la conjecture de Blaschke, Bull. Soc. Math. France, 101 (1973), 17–69.
[30] A. Weinstein, On the volume of manifolds whose geodesics are all closed, J. Differential Geometry, 9 (1974), 513–517.
[31] H. Nakagawa, A note on theorems of Bott and Samelson, J. Math. Kyoto Univ., 7 (1967), 205–220.
[32] C. T. Yang, Odd-dimensional Wiedersehen manifolds are spheres, J. Differential Geometry, 15 (1980), 91–96.
[33] H. Busemann, The geometry of geodesics, Academic Press, 1955.
[34] H. Busemann, Recent synthetic differential geometry, Springer, 1970.
[35] S. Cohn-Vossen, Kürzeste Wege und Totalkrümmung auf Flächen, Compositio Math., 2 (1935), 63–133.
[36] B. Krakus, Any 3-dimensional G-space is a manifold, Bull. Acad. Pol. Sci., 16 (1968), 737–740.

179 (VI.5) Geometric Construction

A. General Remarks

A **geometric construction problem** is a problem of drawing a figure satisfying given conditions by using certain prescribed tools only a finite number of times. If the problem is solvable, then it is called a **possible construction problem**; if it is unsolvable, even though there exist figures satisfying the given conditions, then it is an **impossible construction problem**. If there does not exist a figure satisfying the given conditions, then we say that the problem is **inconsistent**.

Among problems of geometric construction, the oldest and the best known are those of constructing plane figures by means of **ruler** and **compass**. In this article, we call these problems simply problems of elementary geometric construction. The following are some of the more famous problems of this kind; the first four are possible construction problems.

(1) Suppose that we are given three straight lines l, m, n and three points P, Q, R in a plane. Draw a triangle ABC in such a way that vertices A, B, C lie on l, m, n and sides BC, CA, AB pass through P, Q, R (**Steiner's problem**).

(2) Suppose that we are given a circle O and three points P, Q, R not lying on O. Draw a triangle ABC inscribed in O in such a way that the sides BC, CA, AB pass through P, Q, R (**Cramer-Castillon problem**).

(3) Draw a circle tangent to all of three given circles (**Apollonius' problem**).

(4) Suppose that we are given a triangle. Draw three circles inside this triangle in such a way that each is tangent to two sides of the triangle and any two of the circles are tangent to each other (**Malfatti's problem**).

(5) Let n be a natural number. For the division of the circumference of a circle into n equal parts (consequently, the construction of a †regular n-gon) to be a possible construction problem, it is necessary and sufficient that the representation of n as a product of prime numbers take the form $n=2^\lambda p_1 \dots p_k$, where $\lambda \geqslant 0$, $p_1, \dots, p_k$ are all different prime numbers of the form 2^h+1 (†Fermat number) (C. F. Gauss, 1801).

(6) The following are three famous impossible construction problems of Greek origin: (i) divide a given angle into three equal parts (**trisection of an angle**); (ii) construct a cube whose volume is double that of a given cube (**duplication of a cube** or the **Delos problem**); and (iii) construct a square whose area is that of a given circle (**quadrature of a circle**). P. L. Wantzel (1837) proved that problems (i) and

(ii) are impossible except for the special cases of $\pi/2$, $\pi/4$, etc. for (i); and C. L. F. Lindemann (1882) proved the impossibility of (iii) while proving that the number π is †transcendental.

B. Conditions for Constructibility

A problem of elementary geometric construction amounts to a problem of determining a certain number of points by drawing straight lines that pass through given pairs of points, and circles having given points as centers and passing through given points. Let (a_1, b_1), $(a_2, b_2), \ldots, (a_n, b_n)$ be rectangular coordinates of given points, and let K be the smallest †number field containing the numbers $a_1, \ldots, b_n$. Straight lines that join given pairs of points and circles that have given points as centers and that pass through given points are represented by equations of the first or second degree with coefficients belonging to K. Consequently, the coordinates of points of intersection of these straight lines and circles belong to a quadratic extension $K' = K(\sqrt{d})$ of K. Let A be the set of coordinates of the points that are to be determined. Then the problem is solvable if and only if any number α in A is contained in a field $L = K(\sqrt{d_1}, \sqrt{d_2}, \ldots, \sqrt{d_r})$, where $d_{i+1} \in K(\sqrt{d_1}, \ldots, \sqrt{d_i})$ $(i = 0, 1, \ldots, r-1)$. Thus L is a †normal extension field of K whose degree over K is a power of 2. Using this theorem we can prove the impossibility of trisection of an angle and duplication of a cube.

Since the 18th century, besides the problem of construction by ruler and compass, problems of construction by ruler alone or by compass alone have also been studied. We state here some of the more notable results: (1) If by drawing a straight line we mean the process of finding two different points on that line, then we can solve all the problems of elementary geometric construction by means of compass alone (G. Mohr, L. Mascheroni). (2) If by drawing a circle we mean the process of finding its center and a point on its circumference, and if a circle and its center are given, then we can solve any problem of elementary geometric construction by means of ruler alone. (3) It is not possible to find the center of a given circle by ruler alone (D. Hilbert). (4) It is impossible to bisect a given segment by ruler alone. (5) When two intersecting circles or concentric circles are given, we can find the centers of these circles by ruler alone. When nonintersecting and nonconcentric circles are given, it is not possible to find their centers by ruler alone (D. Cauer).

Cases have been considered in which the radius of a circle we can draw by compass or the length of a segment we can draw by ruler is required to satisfy certain conditions. Also, various considerations have been made concerning cases in which we can use tools other than ruler and compass. For example, it is known that although not all possible elementary geometric construction problems are solvable by ruler alone, all these problems are possible if we have either a pair of parallel rulers, a square, or a triangle having a fixed acute angle. If we use a square and a compass, then the trisection of an angle and the duplication of a cube are possible (L. Bieberbach). Also, when a conic section other than a circle is given, the trisection of an angle and the duplication of a cube become possible by ruler and compass (H. J. S. Smith and H. Kortum). By ruler and †transferrer of constant lengths, we can solve Malfatti's problem but not Apollonius' problem (Feldblum).

Even when a problem is possible, the method of construction may be rather complicated and impractical. In these cases, various methods of highly accurate approximate construction have been investigated.

References

[1] H. L. Lebesgue, Leçons sur les constructions géométriques, Gauthier-Villars, 1950.
[2] L. Bieberbach, Theorie der geometrischen Konstruktionen, Birkhäuser, 1952.

180 (XX.15) Geometric Optics

A. General Remarks

Geometric optics is a mathematical theory of light rays. It is not concerned with the properties of light rays as waves (e.g., their wavelength and frequency), but studies their properties as pencils of rays that follow three laws: the law of rectilinear propagation, the law of reflection (i.e., angles of incidence and reflection on a smooth plane are equal (Euclid)), and the law of refraction (i.e., if θ and θ' are angles of incidence and refraction of a light ray refracted from a uniform medium to a second uniform medium and if n, n' are the refractive indices of the first and the second medium, respectively, then $n \sin\theta = n' \sin\theta'$ (R. W. Snell, Descartes)). These three laws follow from **Fermat's principle**, which states that the path of a light ray traveling from a point A' to A in a medium with refractive index $n(P)$ at P is such that the integral $\int_{A'}^{A} n(P)\,ds$ attains its

extremal value, where ds is the line element along the path. This line integral is called the **optimal distance** from A' to A. Therefore Fermat's principle can be taken as a foundation of geometric optics and is, in a way, similar to the †variational principle in particle dynamics (**Maupertuis's principle**),

$$\delta\int\sqrt{2h-2U(P)}\,ds=0,$$

which is satisfied by the path of a particle of unit mass having constant energy h passing through a field of †potential $U(P)$. The quantity $\sqrt{2h-2U}$ corresponds to the refractive index n.

In an optical system, express the position of a point on the path of a light ray by orthogonal coordinates (x, y, z), and define the †Lagrangian $L=n\sqrt{1+\dot{x}^2+\dot{y}^2}$ ($\dot{x}=dx/dz$, $\dot{y}=dy/dz$), **optical direction cosines** $p=\partial L/\partial\dot{x}$, $q=\partial L/\partial\dot{y}$, and the †Hamiltonian $H=\dot{x}p+\dot{y}q-L=-\sqrt{n^2-p^2-q^2}$. Then the †canonical equations of the path are obtained as in particle dynamics; x, y and p, q are called †canonical variables. Due to the variational principle, the integral of the linear differential form $p\,dx+q\,dy-H\,dz=\omega_a$ along a light path is a function $S(A', A)$ of the endpoints A', A of the path, and the optical direction cosines and the Hamiltonians of the system at A and A' are given by

$$\frac{\partial S}{\partial x'}=-p',\quad \frac{\partial S}{\partial y'}=-q',\quad \frac{\partial S}{\partial z'}=H',$$

$$\frac{\partial S}{\partial x}=p,\quad \frac{\partial S}{\partial y}=q,\quad \frac{\partial S}{\partial z}=-H.$$

Hence we obtain †Hamilton-Jacobi differential equations

$$\left(\frac{\partial S}{\partial x'}\right)^2+\left(\frac{\partial S}{\partial y'}\right)^2+\left(\frac{\partial S}{\partial z'}\right)^2=n'^2,$$

$$\left(\frac{\partial S}{\partial x}\right)^2+\left(\frac{\partial S}{\partial y}\right)^2+\left(\frac{\partial S}{\partial z}\right)^2=n^2.$$

As a corollary to these relations we obtain **Malus's theorem**, which states that a pencil of light rays perpendicular to a common surface (locally) at a given moment is also perpendicular to a common surface (locally) after an arbitrary number of reflections and refractions.

Suppose that light rays travel from an object space into an image space through an optical apparatus. If all the rays starting from any one point of the object space converge to a point of the image space and if the mapping given by this correspondence is †bijective, then we say that this imaging is **perfect**. Examples of perfect imaging systems are realized by optical apparatus such as **Maxwell's fisheye** (having refractive index $n(r)=a/(b+r^2)$, where r denotes the distance from the center of the system) and **Luneburg's lens** ($n(r)=\sqrt{a-r^2}$).

Perfect imaging conserves optical distance, yields the relation $n(A')ds'=n(A)ds$, and gives a †conformal mapping, with the magnification inversely proportional to the refractive index.

B. Gauss Mappings

Consider an optimal system with a symmetrical axis of rotation, its **optical axis**. A ray of light that is near the optical axis and has a small inclination to the axis is called a **paraxial ray**. A mapping realizable by paraxial rays where the canonical variables x, y, p, q can be considered to be infinitesimal variables whose squares are negligible, is called a **Gauss mapping (Gauss map)**. When the positions of an object point and its image under a Gauss mapping are represented by homogeneous coordinates, the mapping is represented as a linear transformation, i.e., a †collineation, which maps a point to a point and a line to a line. A point in one space corresponding to the point at infinity in the other space is called a **focus**. If we take a focus as the origin of a coordinate system in each space and use the †homogeneous coordinates x_i such that $x=x_1/x_4$, $y=x_2/x_4$, $z=x_3/x_4$, then a Gauss mapping can be represented as $x_1=x'_1$, $x_2=x'_2$, $x_3=fx'_4$, $f'x_4=x'_3$. The ratio of x to x', i.e., the lateral magnification, is $x/x'=z/f=f'/z'$, where x' is the length of an object orthogonal to the axis and x is the length of its image. The distance f between a focus and a point where the lateral magnification is 1 (such a point is called a **principal point**) is called **focal length** in each space. The telescopic mapping, i.e., $x_1=x'_1$, $x_2=x'_2$, $x_3=ax'_3$, $x_4=bx'_4$, is also a Gauss mapping, in which the lateral magnification is constant.

C. Aberration

When a mapping is realized not only by paraxial rays but also by rays having larger inclinations, a departure from the Gauss mapping arises. This departure is generally called **aberration**. Suppose that a light ray that passes through the point (x', y', z') of a plane perpendicular to the optical axis at a fixed z' and has optical direction cosines p', q' is transformed by the optical apparatus into a light ray that passes through the point (x, y, z) of a plane perpendicular to the optical axis at a fixed z and has optical direction cosines p, q there. Then by the variational principle, $p\,dx+q\,dy-p'\,dx'-q'\,dy'=dW$ (dW is an †exact differential). Therefore the transformation $(x', y', p', q')\to(x, y, p, q)$ is a †canonical transformation. The

mapping can be described by

$$p=\frac{\partial W}{\partial x}, \quad q=\frac{\partial W}{\partial y},$$

$$p'=-\frac{\partial W}{\partial x'}, \quad q'=-\frac{\partial W}{\partial y'}$$

in terms of W and can also be represented in terms of $V=W+p'x'+q'y'$ or $U=W+p'x'+q'y'-px-qy$. For a given optical system, one of these functions W, U, V (called a **characteristic function** or **eikonal**) can be used to estimate the aberration. By developing such a characteristic function in power series of canonical variables and observing its terms of less than the fifth power, we can single out five kinds of aberration: spherical aberration, curvature of image field, distortion, coma, and astigmatism in a rotationally symmetric optical system. To eliminate these aberrations an optical system must satisfy Abbe's sine condition (the elimination of spherical aberration and coma), Petzval's condition (the elimination of astigmatism and curvature of image field), and the tangent condition (the elimination of distortion).

The path of a charged particle in an electromagnetic field can be treated in the same way as the path of a light ray. Let ε represent the specific charge of the particle, h the energy, A_0 the electrostatic potential, and A_x, A_y, A_z vector potentials. Then the index of refraction is $\sqrt{2(h-\varepsilon A_0)}+\varepsilon(A_x dx/ds+A_y dy/ds+A_z dz/ds)$. In this case, the index of refraction shows the anisotropy caused by the existence of the magnetic field. The paths of paraxial rays are determined by a set of linear differential equations of the second order, and the Gauss mapping is realized as in geometric optics.

References

[1] C. Carathéodory, Geometrische Optik, Springer, 1937.
[2] M. Herzberger, Modern geometrical optics, Interscience, 1958.
[3] R. K. Luneburg, Mathematical theory of optics, Univ. of California Press, 1964.
[4] O. N. Stavroudis, The optics of rays, wavefronts, and caustics, Academic Press, 1972.

181 (VI.1) Geometry

The Greek word for geometry, which means *measurement of the earth*, was used by the historian Herodotus, who wrote that in ancient Egypt people used geometry to restore their land after the inundation of the Nile. Thus the theoretical use of figures for practical purposes goes back to pre-Greek antiquity. Tradition holds that Thales of Miletus knew some properties of congruent triangles and used them for indirect measurement, and that the Pythagoreans had the idea of systematizing this knowledge by means of proofs (→ 24 Ancient Mathematics; 187 Greek Mathematics). †Euclid's *Elements* is an outgrowth of this idea [1]. In this work, we can see the entire mathematical knowledge of the time presented as a logical system. It includes a chapter (Book V) on the theory of quantity (i.e., the theory of positive real numbers in present-day terminology) and chapters on the theory of integers (Books VII–IX), but for the most part, it treats figures in a plane or in space and presents number-theoretic facts in geometric language.

Geometry in today's usage means the branch of mathematics dealing with spatial figures. In ancient Greece, however, all of mathematics was regarded as geometry. In later times, the French word *géomètre* or the German word *Geometer* was sometimes used as a synonym for *mathematician*. In a fragment of his *Pensées*, B. Pascal speaks of the *esprit de géométrie* as opposed to the *esprit de finesse*. The former means simply the mathematical way of thinking.

Algebra was introduced into Europe from the Middle East toward the end of the Middle Ages and was further developed during the Renaissance. In the 17th and the 18th centuries, with the development of analysis, geometry achieved parity with algebra and analysis.

As R. Descartes pointed out, however, figures and numbers are closely related [2]. Geometric figures can be treated algebraically or analytically by means of †coordinates (the method of **analytic geometry**, so named by S. F. Lacroix [3]); conversely, algebraic or analytic facts can be expressed geometrically. Analytic geometry was developed in the 18th century, especially by L. Euler [4], who for the first time established a complete algebraic theory of †curves of the second order. Previously, these curves had been studied by Apollonius (262–200? B.C.) as †conic sections. The idea of Descartes was fundamental to the development of analysis in the 18th century. Toward the end of that century, analysis was again applied to geometry. For example, G. Monge's contribution [5] can be regarded as a forerunner of †differential geometry.

However, we cannot say that the analytic method is always the best manner of dealing with geometric problems. The method of treating figures directly without using coordinates

is called **synthetic** (or **pure**) **geometry**. In this vein, a new field called †projective geometry was created by G. Desargues and B. Pascal in the 17th century. It was further developed in the 19th century by J.-V. Poncelet, L. N. Carnot, and others. In the same century, J. Steiner insisted on the importance of this field (→ 343 Projective Geometry).

On the other hand, the †axiom of parallels in Euclid's *Elements* has been an object of criticism since ancient times. In the 19th century, by denying the a priori validity of Euclidean geometry, J. Bolyai and N. I. Lobachevskiĭ formulated non-Euclidean geometry, whose logical consistency was shown by models constructed in both Euclidean and projective geometry (→ 285 Non-Euclidean Geometry).

In analytic geometry, physical spaces and planes, as we know them, are represented as 3-dimensional or 2-dimensional Euclidean spaces E^3, E^2. It is easy to generalize these spaces to n-dimensional Euclidean space E^n. A "point" of E^n is an n-tuple of real numbers $(x_1, \ldots, x_n)$, and the distance between two points $(x_1, \ldots, x_n)$, $(y_1, \ldots, y_n)$ is $((y_1-x_1)^2+\ldots+(y_n-x_n)^2)^{1/2}$. The geometries of E^2, E^3 are called **plane geometry** and **space** (or **solid**) **geometry**, respectively. The geometry of E^n is called **n-dimensional Euclidean geometry**. We obtain n-dimensional projective or non-Euclidean geometries similarly. F. Klein [7] proposed systematizing all these geometries in group-theoretic terms. He called a "space" a set S on which a group G operates and a "geometry" the study of properties of S invariant under the operations of G (→ 137 Erlangen Program).

B. Riemann [6] initiated another direction of geometric research when he investigated n-dimensional †manifolds and, in particular, †Riemannian manifolds and their geometries. Some aspects of Riemannian geometry fall outside of geometry in the sense of Klein. It was a starting point for the broad field of modern differential geometry, that is, the geometry of †differentiable manifolds of various types (→ 109 Differential Geometry).

The reexamination of the system of axioms of Euclid's *Elements* led to D. Hilbert's †foundations of geometry and to the axiomatic tendency of present-day mathematics. The study of algebraic curves, which started with the study of conic sections, developed into the theory of algebraic manifolds, the algebraic geometry that is now developing so rapidly (→ 12 Algebraic Geometry). Another branch of geometry is topology, which has developed since the end of the 19th century. Its influence on the whole of mathematics today is considerable (→ 114 Differential Topology; 426 Topology). Geometry has now permeated all branches of mathematics, and it is sometimes difficult to distinguish it from algebra or analysis. The importance of geometric intuition, however, has not diminished from antiquity until today.

References

[1] T. L. Heath, The thirteen books of Euclid's elements, Cambridge, 1908.
[2] R. Descartes, Géométrie, Paris, 1637 (Oeuvres, IV, 1901).
[3] S. F. Lacroix, Traité élémentaire de trigonométrie rectiligne et sphérique et d'application de l'algèbre à la géométrie, Bachelier, 1798–1799.
[4]L. Euler, Introductio in analysin infinitorum, Lausanne, 1748 (Opera omnia VIII, IX, 1922); French translation, Introduction à l'analyse infinitesimale I, II, Paris, 1835; German translation, Einleitung in die Analysis des Unendlichen, Springer, 1885.
[5] G. Monge, Application de l'analyse à la géométrie, fourth edition, Paris, 1809.
[6] B. Riemann, Über die Hypothese, welche der Geometrie zu Grunde liegen, Habilitationsschrift, 1854 (Gesammelte mathematische Werke, Teubner, 1876, 254–269; Dover, 1953).
[7] F. Klein, Vergleichende Betrachtungen über neuere geometrische Forschungen, Das Erlanger Programm, 1872 (Gesammelte mathematicshe Abhandlungen I, Springer, 1921, 460–497).

182 (V.10) Geometry of Numbers

A. History

H. Minkowski introduced the notions of lattice and convex set in the †algebraic theory of numbers. He developed a simple yet powerful method of arithmetic investigation using these geometric notions to simplify the analytic theory of †Diophantine approximation, which had been developed by P. G. L. Dirichlet and C. Hermite. His theory, the **geometry of numbers**, has continued its development and contributed to various fields of mathematics (→ 83 Continued Fractions).

B. Lattices

Let E^n be an n-dimensional Euclidean space identified with the linear space $\mathbf{R}^n$. For a point

P in E^n, we denote the corresponding vector in $\mathbf{R}^n$ by $v(P)={}^t(x_1, \ldots, x_n)$. A subset Λ of E^n is called an n-dimensional **(homogeneous) lattice** if there exists a basis $\{v_1, \ldots, v_n\}$ of $\mathbf{R}^n$ such that $\Lambda=\{P \in E \mid v(P)=\sum_{i=1}^n \lambda_i v_i, \lambda_i \in \mathbf{Z}\}$. The set of points $\{X_1, \ldots, X_n\}$ such that $v(X_i)=v_i$ $(i=1, \ldots, n)$ is called a **basis** of the lattice Λ. A typical example of a lattice is the point set corresponding to $\mathbf{Z}^n$ in $\mathbf{R}^n$. The †free module generated by v_i $(i=1, \ldots, n)$ is denoted by Λ^* and is called the **lattice group** of Λ. We have $\Lambda^* =\{v \in \mathbf{R}^n \mid v=v(P), P \in \Lambda\}$. If $\{u_1, \ldots, u_n\}$ is another basis of the free module Λ^*, then there exists an element (α_{ij}) of $GL(n, \mathbf{Z})$ (i.e., $\alpha_{ij} \in \mathbf{Z}$ and $|\det(\alpha_{ij})|=1$) such that $u_j=\sum_{i=1}^n \alpha_{ij} v_j$. Hence the quantity $|\det(v_1, \ldots, v_n)|$ is independent of the choice of the basis $\{v_1, \ldots, v_n\}$. We denote this quantity by $d(\Lambda)$ and call it the **determinant of the lattice**. We denote the minimum distance between the points belonging to Λ by $\delta(\Lambda)$.

A subset L of the space E^n is called an **inhomogeneous lattice** if there exists a homogeneous lattice Λ in E^n and a point P_0 in E^n such that $L=\{P \in E^n \mid v(P)-v(P_0) \in \Lambda^*\}$. Thus an inhomogeneous lattice is obtained from a homogeneous lattice by translation. In this article we restrict ourselves to the case of homogeneous lattices and henceforth omit the adjective "homogeneous."

Suppose we are given a sequence of lattices $\Lambda_1, \Lambda_2, \ldots$, in E^n with bases $\{X_i^{(1)}\}, \{X_i^{(2)}\}, \ldots$. If the sequence of points $X_i^{(\nu)}$ converges to X_i $(i=1, \ldots, n)$ and the set $\{X_1, \ldots, X_n\}$ forms a basis of a lattice Λ, we call Λ the **limit of the sequence** $\{\Lambda_\nu\}$; we also say that the sequence $\{\Lambda_\nu\}$ **converges** to the lattice Λ. In this case we have $d(\Lambda_\nu) \rightarrow d(\Lambda)$, $\delta(\Lambda_\nu) \rightarrow \delta(\Lambda)$. The notion of convergence of lattices gives rise to a topology of the space M_0 of all the lattices in E. A sequence $\{\Lambda_\nu\}$ of lattices is said to be *bounded* if there exist positive numbers c and c' such that $d(\Lambda_\nu) \leqslant c$, $\delta(\Lambda_\nu) \geqslant c'$ for all ν. A bounded sequence of lattices has a convergent subsequence.

Let S be a subset of the space E^n. A lattice Λ is called S-**admissible** if we have $\Lambda \cap S^i=\{O\}$, where S^i is the interior of S and O is the origin. We denote the set of S-admissible lattices by $A(S)$. Given a closed subset M of M_0, we put $\Delta(S \backslash M)=\inf_{\Lambda \in A(S) \cap M} d(\Lambda)$ if $A(S) \cap M$ is nonempty, while if $A(S) \cap M$ is empty, we put $\Delta(S \backslash M)=\infty$. When $M=M_0$, we write $\Delta(S \backslash M) =\Delta(S)$ and call it the **critical determinant** of S. Generally, a lattice Λ in M is said to be **critical in M with respect to** S if $\Lambda \in A(S)$ and $d(\Lambda)=\Delta(S \backslash M)$. Suppose that we have $0<\Delta(S \backslash M)<\infty$. Then for a lattice critical in M with respect to S to exist it is necessary and sufficient that there exist a bounded sequence $\{\Lambda_\nu\}$ such that $\Lambda_\nu \in M \cap A(S)$ and $d(\Lambda_\nu) \rightarrow \Delta(S \backslash M)$.

C. Successive Minima and Minkowski's Theorem

A subset S of the space E^n is called a **bounded star body** (symmetric with respect to the origin) if there exists a continuous function F defined on the space E satisfying the following four conditions: (i) $F(0)=0$; (ii) if $X \neq 0$, then $F(X)>0$; (iii) for an arbitrary real number t and a point X, we have $F(tX)=|t|F(X)$; (iv) $S=\{X \mid F(X) \leqslant 1\}$. A bounded closed †convex body that is symmetric with respect to the origin is a bounded star body. If we are given a star body S, the associated function F, and a lattice Λ, there exist a set of points $\{P_1, \ldots, P_n\}$ in Λ and a set of positive numbers $\{p_1, \ldots, p_n\}$ satisfying the following four conditons: (1) $v(P_1), \ldots, v(P_n)$ are linearly independent; (2) $F(P_i)=p_i$ $(i=1, \ldots, n)$; (3) $p_1 \leqslant \ldots \leqslant p_n$; (4) if P is a point in Λ and $v(P)$ is not contained in the subspace spanned by $\{v(P_1), \ldots, v(P_{m-1})\}$, then $F(P) \geqslant p_m$. The set $\{p_1, \ldots, p_n\}$ is uniquely determined by S and Λ. The numbers p_i are called the **successive minima** of S in Λ; the points P_i are the **successive minimum points** of S in Λ.

Minkowski's theorem: Let Λ be a lattice in a Euclidean space E^n and S a bounded subset of E^n. Then we have the following:

(I) If the volume $V(S)$ is larger than $d(\Lambda)$, then there exist points X_1 and X_2 in S such that $X_1 \neq X_2$ and $v(X_1)-v(X_2) \in \Lambda^*$. Suppose, moreover, that S is convex and symmetric with respect to the origin. Then, if $V(S)>2^n d(\Lambda)$, there exists a point X in $S \cap \Lambda$ different from the origin. Hence we have $2^n \Delta(S) \geqslant V(S)$ $(n=\dim E^n)$.

(II) Let S be a bounded closed convex body that is symmetric with respect to the origin, and let $p_1, \ldots, p_n$ be the successive minima of S in Λ. Then we have $p_1 \ldots p_n \cdot V(S) \leqslant 2^n d(\Lambda)$.

D. Minkowski-Hlawka Theorem

Suppose that we are given a subset S of the n-dimensional Euclidean space E^n such that the characteristic function $\chi(X)$ of S is †integrable in the sense of Riemann. Then we have the **Minkowski-Hlawka theorem**: (i) If $n \geqslant 2$ and S is open, then $\Delta(S) \leqslant V(S)$, and (ii) if, moreover, S is symmetric with respect to the origin, then $2\Delta(S) \leqslant V(S)$; (iii) if S is a symmetric star body with respect to the origin, then $2\zeta(n)\Delta(S) \leqslant V(S)$, where $\zeta(n)$ is the †Riemann zeta function.

A proof for the theorem was given by E. Hlawka (1944); (iii) and (iv) were conjectured by Minkowski. C. L. Siegel obtained another proof (1945), and C. A. Rogers simplified the original proof by Hlawka (1947). There are

results concerning the estimation of $\Delta(S)/V(S)$ for various subsets S.

E. Siegel's Mean Value Theorem

In an attempt to obtain a proof for the latter half of the Minkowski-Hlawka theorem, Minkowski observed the necessity of establishing the arithmetic theory of the linear transformation groups. Siegel was inspired by this observation and obtained the following theorem, **Siegel's mean value theorem**, which implies the Minkowski-Hlawka theorem: Let F be a †fundamental region of the group $SL(n, \mathbf{R})$ with respect to a discrete subgroup $SL(n, \mathbf{Z})$. Let ω be the †invariant measure on $SL(n, \mathbf{R})$ such that $\int_F d\omega = 1$ (→ 225 Invariant Measures). Let f be a bounded Riemann integrable function with compact †support defined on the space $\mathbf{R}^n$. Note that the lattice $\mathbf{Z}^n$ is stabilized by the subgroup $SL(n, \mathbf{Z})$. We have

$$\int_F \sum_{\substack{x \in \mathbf{Z}^n \\ x \neq 0}} f(g \cdot x)\, d\omega(g) = \int_{\mathbf{R}^n} f(x)\, dx,$$

where the right-hand side of the equation is the usual Riemann integral of the function f. A. Weil considered this theorem in a more general setting (*Summa Brasil. Math.*, 1 (1946)).

F. Diophantine Approximation

Minkowski initiated the notion of Diophantine approximation in reference to the problem of estimating the absolute value $|f(x)|$ of a given function f, where x varies in $\mathbf{Z}$ or in a given ring of †algebraic integers. (A †Diophantine equation is an equation $f(x) = 0$, where x varies in $\mathbf{Z}$.) Today Diophantine approximation (in the wide sense) refers to the investigation of the scheme of values $f(x)$, where x varies in a suitable ring of algebraic integers. The geometry of lattices is a powerful tool in this investigation. A typical problem in this field of study is that of approximating irrational numbers by rational numbers; here †continued fractions play an important role (→ Section G). For the problem of uniform distribution considered by H. Weyl, the analytic method, especially that of trigonometric series, is useful (→ Section H). **Dirichlet's drawer principle** (to put n objects in m drawers with $n > m$, it is necessary to put more than one object in at least one drawer) is one of the basic principles used in the theory of Diophantine approximation. Recently, the theory has been applied to the theory of †transcendental numbers and the theory of †Diophantine equations.

G. Approximation of Irrational Numbers by Rational Numbers

Given an irrational number θ, we have the problem of finding rational integers x (>0) and y such that $|\theta - y/x| < \varepsilon/x$, where ε is a given positive number. Suppose that we are given a positive integer N. Using Dirichlet's drawer principle we can show the existence of x $(\leqslant N)$ and y such that $|\theta - y/x| < 1/xN$. Let $M(\theta)$ be the supremum of positive numbers M such that the inequality $|\theta - y/x| < 1/Mx^2$ holds for infinitely many pairs of integers x, y. We have $1 \leqslant M(\theta)$ $(\leqslant \infty)$. Two irrational numbers θ and θ' are said to be **equivalent** if there exists an element $(a_{ij}) \in GL(2, \mathbf{Z})$ such that $\theta' = (a_{11}\theta + a_{12})/(a_{21}\theta + a_{22})$. In this case we have $M(\theta) = M'(\theta)$.

If the irrational number θ satisfies the quadratic equation $a\theta^2 + b\theta + c = 0$ (a, b, c are rational integers), then we have $M(\theta) = k^{-1}\sqrt{b^2 - 4ac}$, where $k = \min\{ax^2 + bxy + cy^2 \mid x, y \in \mathbf{Z}, x \neq 0, y \neq 0\}$. In general, for an irrational number θ of degree two, we have $M(\theta) \geqslant \sqrt{5}$. The equality $M(\theta) = \sqrt{5}$ holds if θ is equivalent to $\theta_1 = (1 + \sqrt{5})/2$. If θ is not equivalent to θ_1, then $M(\theta) \geqslant \sqrt{8}$; the equality holds if θ is equivalent to $\theta_2 = 1 + \sqrt{2}$. Similarly, we have $\theta_3, \theta_4, \ldots$; and $M(\theta_n) \to 3$ $(n \to \infty)$. If $M(\theta) < 3$, there exists a θ_n such that θ is equivalent to θ_n. The set of irrational numbers θ satisfying $M(\theta) = 3$ is uncountable (A. A. Markov [13]). We have no information about $M(\theta)$ for the general algebraic irrational number θ. Let $\mu(\theta)$ be the supremum of real numbers μ such that the inequality $|\theta - y/x| < 1/x^\mu$ holds for infinitely many pairs of integers x, y. Given a number $\kappa > 2$, it can be shown that the †Lebesgue measure of the set of real numbers θ such that $\mu(\theta) \geqslant \kappa$ is zero. If θ is a real algebraic number of degree n, then $\mu(\theta) \leqslant n$ (J. Liouville). Concerning $\mu(\theta)$, results have been obtained by A. Thue, Siegel, A. O. Gel'fond, and F. J. Dyson. K. F. Roth (1954) proved that $\mu(\theta) = 2$ (**Roth's theorem** [12]), which settled the problem of $\mu(\theta)$. Roth's theorem means that if κ is larger than 2, then there exist only a finite number of pairs x, y satisfying $|\theta - y/x| < 1/x^\kappa$. This can be generalized to the case of the approximation of an element θ that is algebraic over an A-field k by an element of the field k (S. Lang [9]). (An A-field is either an algebraic number field of finite degree or an algebraic function field in one variable over a finite constant field.)

In 1970 W. M. Schmidt [20] obtained theorems on simultaneous approximation which generalize Roth's theorem. Thus, if $\alpha_1, \ldots, \alpha_n$ are real algebraic numbers such that $1, \alpha_1, \ldots, \alpha_n$ are linearly independent over the

field of rational numbers, then for every $\varepsilon>0$ there are only finitely many positive integers q with

$$\|q\alpha_1\| \cdots \|q\alpha_n\| q^{1+\varepsilon}<1,$$

where $\|\xi\|$ denotes the distance from a real number ξ to the nearest integer; in particular, we have

$$|\alpha_i-p_i/q|<q^{-(n+1)/n-\varepsilon}, \qquad i=1,\dots,n,$$

for only finitely many n-tuples of rationals $p_1/q,\dots,p_n/q$. A dual to this result is as follows. Let $\alpha_1,\dots,\alpha_n$, ε be as before. Then there are only finitely many n-tuples of nonzero integers $q_1,\dots,q_n$ with

$$\|q_1\alpha_1+\dots+q_n\alpha_n\|\cdot|q_1\cdots q_n|^{1+\varepsilon}<1.$$

This last theorem can be used to prove that if α is an algebraic number, k a positive integer, and $\varepsilon>0$, then there are only finitely many algebraic numbers ω of degree at most k such that $|\alpha-\omega|<H(\omega)^{-k-1-\varepsilon}$, where $H(\omega)$ denotes the height of ω. See also [16, 22].

The work of Thue, Siegel, and Roth had the basic limitation of noneffectiveness. A. Baker (1968) succeeded in proving that for any algebraic number θ of degree $n\geqslant 3$ and any $\kappa>n$, there exists an effectively computable number $c=c(\theta,\kappa)>0$ such that $|\theta-y/x|>cx^{-n}\exp(\log x)^{1/\kappa}$ for all integers x, y $(x>0)$ [15]. This result is an immediate consequence of the following effective version of a classical theorem on binary Diophantine equations (Thue, 1909): Let $f=f(x,y)$ be an irreducible binary form of degree $n\geqslant 3$ with integer coefficients, and suppose that $\kappa>n$. Then for any positive integer m, all integer solutions x, y of the equation $f(x,y)=m$ satisfy $\max(|x|,|y|)<c\exp(\log m)^{\kappa}$, where $c>0$ is an effectively computable number depending on n, κ, and the coefficients of f. Baker obtained this result by making use of his theorems which give effective estimates of moduli of linear forms in the logarithms of algebraic numbers with algebraic coefficients. A typical theorem reads as follows: Let $\alpha_1,\dots,\alpha_n$ be nonzero algebraic numbers with $\log\alpha_1,\dots,\log\alpha_n$ linearly independent over the rationals, and let $\beta_0,\dots,\beta_n$ be algebraic numbers, not all 0, with degrees and heights at most d and H, respectively. Then for any $\kappa>n+1$, we have $|\beta_0+\beta_1\log\alpha_1+\dots+\beta_n\log\alpha_n|>c\exp(-(\log H)^{\kappa})$, where $c>0$ is an effectively computable number depending only on n, κ, $\log\alpha_1,\dots,\log\alpha_n$, and d [14, 19].

Results of this kind have many important applications in number theory. For instance, we obtain a generalization of the Gel'fond-Schneider theorem on transcendental numbers. Furthermore, the imaginary quadratic number fields of class number 1 can be completely determined on the basis of Baker's result. This was actually done by Baker (1966) and independently by H. M. Stark (1966) (→ 347 Quadratic Fields).

Refinements and generalizations of Thue's theorem on the finiteness of solutions of binary Diophantine equations have been obtained by Baker and his collaborators. (→ [17], and also [16]). Also, p- and $\mathfrak{p}$-adic analogs of Baker's results are known [18].

There are a number of unsettled problems on the irrationality of particular numbers, such as the †Euler constant C, π^e, or $\zeta(2k+1)$ with k a positive integer. R. Apéry (1978) proved that there exist a positive number ε and a sequence of positive integers $\{q_n\}$ such that $0<\|q_n\zeta(3)\|<e^{-\varepsilon n}$ for all $n\geqslant 1$, so that $\zeta(3)$ is irrational.

H. Uniform Distribution

Let θ be a real number, x a positive integer, and $[\theta x]$ the maximum integer not larger than θx. We write $(\theta x)=\theta x-[\theta x]\equiv\theta x \pmod 1$. Jacobi showed that if θ is irrational, then the set $\{(\theta x)\mid x\in\mathbf{N}\}$ is densely distributed in the interval $(0,1)$ ($\mathbf{N}$ is the set of positive integers). In general, let f be a real-valued function defined on $\mathbf{N}$. We say that $f(x) \pmod 1$ is **uniformly distributed** in the unit interval, or $f(x)$ is **uniformly distributed (mod 1)**, if the following condition is satisfied: Let α, β be an arbitrary pair of real numbers such that $0\leqslant\alpha<\beta\leqslant 1$, and let N be a given positive integer. Let $T(N)$ be the number of positive integers x such that $x\leqslant N$, $\alpha\leqslant(f(x))<\beta$, where $(f(x))=f(x)-[f(x)]$. Then $\lim_{N\to\infty}T(N)/N=\beta-\alpha$. In order for $f(x) \pmod 1$ to be uniformly distributed, it is necessary and sufficient that $\lim_{N\to\infty}N^{-1}\sum_{x=1}^{N}e^{2\pi i h f(x)}=0$ for any nonzero integer h (**Weyl's criterion**, 1914). Weyl proved that if θ is an irrational number, then $\theta x \pmod 1$ is uniformly distributed.

The following theorem, given by J. G. van der Corput, is often useful: Let f be a real-valued function defined on $\mathbf{N}$. Consider the function $f_h(x)=f(x+h)-f(x)$ for an arbitrary positive integer h. If $f_h(x) \pmod 1$ is uniformly distributed (mod 1) for all such h, then $f(x) \pmod 1$ is also uniformly distributed (mod 1).

Utilizing this theorem, it can be shown that if $f(x)=\theta_r x^r+\theta_{r-1}x^{r-1}+\dots+\theta_1 x$, where at least one of the coefficients θ_i is irrational, then $f(x) \pmod 1$ is uniformly distributed.

The notion of uniform distribution of sequences of real numbers has an analog in compact Hausdorff spaces and in various topological groups. A systemmatic treatment of such generalized notions of uniform distribution can be found in [21].

References

[1] H. Minkowski, Geometrie der Zahlen, Teubner, 1910 (Chelsea, 1953).
[2] H. Minkowski, Diophantische Approximationen, Teubner, second edition, 1927 (Chelsea, 1957).
[3] H. Minkowski, Gesammelte Abhandlungen I, II, Teubner, 1911 (Chelsea, 1967).
[4] J. F. Koksma, Diophantische Approximationen, Erg. Math., Springer, 1936 (Chelsea, 1950).
[5] H. Weyl, C. L. Siegel, and K. Mahler, Geometry of numbers, Mimeographed notes, Princeton, 1950.
[6] J. W. S. Cassels, An introduction to Diophantine approximation, Cambridge Univ. Press, 1957.
[7] J. W. S. Cassels, An introduction to the geometry of numbers, Springer, 1959.
[8] T. Schneider, Einführung in die transzendentalen Zahlen, Springer, 1957.
[9] S. Lang, Diophantine geometry, Interscience, 1962.
[10] C. L. Siegel, Über einige Anwendungen diophantischer Approximationen, Abh. Preuss. Akad. Wiss., no. 1 (1929) (Gesammelte Abhandlungen, Springer, 1966, vol. 1, 209–266).
[11] C. L. Siegel, A mean value theorem in geometry of numbers, Ann. Math., (2) 46 (1945), 340–347 (Gesammelte Abhandlungen, Springer, 1966, vol. 3, 39–46).
[12] K. F. Roth, Rational approximations to algebraic numbers, Mathematika, 2 (1955), 1–20.
[13] A. A. Markoff (A. A. Markov), Sur les formes quadratiques binaires indéfinies, Math. Ann., 15 (1879), 381–407.
[14] A. Baker, Linear forms in the logarithms of algebraic numbers I, II, III, IV, Mathematika, 13 (1966), 204–216; 14 (1967), 102–107; 14 (1967), 220–228; 15 (1968), 204–216.
[15] A. Baker, Contributions to the theory of Diophantine equations I, II, Philos. Trans. Roy. Soc. London, A 263 (1967–1968), 173–191, 193–208.
[16] A. Baker, Transcendental number theory, Cambridge Univ. Press, 1975.
[17] A. Baker and J. Coates, Integer points on curves of genus 1, Proc. Cambridge Philos. Soc., 67 (1970), 595–602.
[18] A. Brumer, On the units of algebraic number fields, Mathematika, 14 (1967), 121–124.
[19] N. I. Fel'dman, Linear form of logarithmic algebraic numbers (in Russian), Dokl. Akad. Nauk SSSR, 182 (1968), 1278–1279.
[20] W. M. Schmidt, Simultaneous approximation to algebraic numbers by rationals, Acta Math., 125 (1970), 189–201.
[21] L. Kuipers and H. Niederreiter, Uniform distribution of sequences, Interscience, 1974.
[22] W. M. Schmidt, Diophantine approximation, Lecture notes in math. 785, Springer, 1980.
[23] W. M. Schmidt, Approximation to algebraic numbers, Enseignement Math., 17 (1971), 188–253.
[24] C. F. Osgood (ed.), Diophantine approximation and its applications, Academic Press, 1973.

183 (VII.23) Global Analysis

Mathematics that treats, by using functional analytical techniques, various problems concerning the †calculus of variations, †singularities, infinite-dimensional Lie groups, or †nonlinear partial differential equations, such as equations of fluids or of gravitation in general relativity, may be called global analysis if it uses as a main tool an infinite-dimensional version of differential geometry and topology analogous to that for finite-dimensional manifolds. The term **global analysis** therefore has no precise definition. However, it can be said that it is analysis on manifolds, and the concept of †infinite-dimensional manifolds is the central abstract idea in it.

Suppose one considers a nonlinear differential operator on a finite-dimensional manifold. Then by using functional analytical techniques it often happens that its domain is neither a linear space nor its open subset but an infinite-dimensional manifold, and that such a nonlinear operator can be regarded as a †differentiable mapping between infinite-dimensional manifolds. The †differential at a point in that source manifold is called a **linearized operator**, to which one can apply various theories of linear functional analysis.

"Global analysis" seems to have first appeared in the literature in the late 1960s [1, 2]. However, the phrase "infinite-dimensional manifold" has been used widely since about 1960. By that early date, local theories of infinite-dimensional manifolds, sometimes called "general analysis," such as definitions of differentiability, the †implicit function theorem, †Taylor's theorem, the existence and uniqueness of †ordinary differential equations, and the †Frobenius theorem, had already been established in †Banach spaces [3]. Therefore, with these regarded as a local theory, the concept of infinite-dimensional manifolds could be defined naturally, and the concept of infinite-

dimensional Lie groups as well [4]. A few years later, R. Palais and S. Smale [5] and J. Eells and J. Sampson [6] showed that such concepts are useful in the calculus of variations, and R. Abraham and J. Robbin [7] remarked that †transversality theorems initiated by R. Thom can be easily proved by using an infinite-dimensional version of †Sard's theorem [8].

The so-called †Atiyah-Singer index theorem [9], announced in 1963, gave impetus to the field as people sought the theorem's most natural expression; the work finally resulted in the theorem classifying separable †Hilbert manifolds by homotopy type (→ 105 Differentiable Manifolds).

After the appearance of these theorems announcing the nonexistence of differential topology on separable infinite-dimensional Hilbert manifolds, global analysis moved toward more concrete problems and applications to various branches of mathematics. However, many of these applications are formulated not on †Banach or Hilbert manifolds, but on manifolds modeled on Fréchet spaces (†Fréchet manifolds) or on †nuclear spaces. For such situations we have in general no local theories. Neither the implicit function theorem nor the Frobenius theorem holds on such manifolds. However, since these theorems are crucial for nonlinear problems, various kinds of sufficient conditions for the validity of these theorems are being studied by many people (→ 286 Nonlinear Functional Analysis).

As for the calculus of variations, **Yamabe's problem** is being studied extensively, for this seems to be a typical problem not satisfying the so-called †Condition C. The original paper of H. Yamabe [13], insisting that every compact Riemannian manifold can be conformally deformed into a manifold of constant scalar curvature, contains a serious gap, and the problem is still open, though many cases are known where the statement holds (→ 364 Riemannian Manifolds H). The harmonic mappings defined in [6] are also being studied extensively (→ 195 Harmonic Mappings).

In differential geometry or the †general theory of relativity definitions of various curvatures, such as Riemannian, Ricci or scalar, or Gauss, can be sometimes regarded as nonlinear differential equations on manifolds. Proof of the global existence of solutions to these equations has long been sought, and several theorems have recently appeared [14–16].

Infinite-dimensional groups such as $GL(E)$, or $GL_c(E)$ with uniform topology, are called †Banach-Lie groups, and they have been studied in the operator calculus. On the other hand, infinite groups studied by S. Lie and E. Cartan were in fact infinite-dimensional germs of transformation groups defined on a neighborhood of a point in a manifold. These were not groups in the strict sense. Recently, H. Omori [17] has given a definition of abstract infinite-dimensional Lie groups that includes Banach-Lie groups and many infinite-dimensional transformation groups studied by Cartan. An application of these groups to †fluid dynamics can be found in [18]. Moreover, †unitary representation theories of these groups are now being constructed [19, 20].

Though global analysis consists at present of a rather disorganized combination of many nonlinear problems in analysis on manifolds and in mathematical physics, it is nevertheless one of the most active branches of mathematics.

References

[1] J. Eells, A setting for global analysis, Bull. Amer. Math. Soc., 72 (1966) 751–807.
[2] R. Palais, Foundations of global nonlinear analysis, Benjamin, 1968.
[3] J. Dieudonné, Foundations of modern analysis, Academic Press, 1960.
[4] G. Birkhoff, Analytic groups, Trans. Amer. Math. Soc., 43 (1938), 61–101.
[5] R. Palais and S. Smale, A generalized Morse theory, Bull. Amer. Math. Soc., 70 (1964), 165–172.
[6] J. Eells and J. Sampson, Harmonic mappings of Riemannian manifolds, Amer. J. Math., 86 (1964), 109–160.
[7] R. Abraham and J. Robbin, Transversal mappings and flows, Benjamin, 1967.
[8] S. Smale, An infinite-dimensional version of Sard's theorem, Amer. J. Math., 87 (1965), 861–866.
[9] M. Atiyah and I. Singer, The index of elliptic operators on compact manifolds, Bull. Amer. Math. Soc., 69 (1963), 422–433.
[10] N. H. Kuiper, The homotopy type of the unitary group of Hilbert space, Topology, 3 (1965), 19–30.
[11] D. Burghelea and N. H. Kuiper, Hilbert manifolds, Ann. Math., (2) 90 (1969), 335–352.
[12] J. Eells and D. Elworthy, Open embeddings of certain Banach manifolds, Ann. Math., (2) 91 (1970), 465–485.
[13] H. Yamabe, On deformation of Riemannian structures on compact manifolds, Osaka Math. J., 12 (1960), 21–37.
[14] T. Aubin, Equations differentielles non linéaires et problème de Yamabe concernant la courbure scalaire, J. Math. Pures Appl., 55 (1976), 269–296.

[15] H. Gluck, The generalized Minkowski problem in differential geometry in the large, Ann. Math., 96 (1972), 245–276.
[16] S. Yau, On the Ricci curvature of a compact Kähler manifold and the complex Monge-Ampère equation I, Comm. Pure Appl. Math., 31 (1978), 339–411.
[17] H. Omori, Infinite-dimensional Lie transformation groups, Lecture notes in math. 427, Springer, 1974.
[18] D. Ebin and J. Marsden, Groups of diffeomorphisms and the motion of an incompressible fluid, Ann. Math., (2) 94 (1970), 101–162.
[19] A. M. Vershik, I. M. Gel'fand, and M. I. Graev, Representations of the group of diffeomorphisms, Russian Math. Surveys, 30 (6) (1975), 1–50. (Original in Russian, 1975.)
[20] R. S. Ismagilov, Unitary representations of the group of diffeomorphisms of the space $\mathbf{R}^n$, $n \geqslant 2$, Functional. Anal. Appl., 9 (1975), 144–155. (Original in Russian, 1975.)

184 (XXI.27) Gödel, Kurt

Kurt Gödel (April 28, 1906–January 14, 1978) was born in Brno, Czechoslovakia (at that time Brünn, Austria-Hungary). He studied mathematics and physics at the University of Vienna, where he took the Ph.D. degree in 1930. After he had taught mathematics at the University of Vienna from 1933 to 1938, he was invited to the Institute for Advanced Study at Princeton, where he became professor in 1953. In 1976, he was named professor emeritus; he died in Princeton in 1978.

Gödel contributed important fundamental results covering all aspects of mathematical logic. Among his famous works are the proof of the †completeness of the first-order predicate calculus, the incompleteness of the consistent axiomatic system containing Peano's arithmetic (incompleteness theorem), and the †consistency of the axiom of choice and the generalized continuum hypothesis. He also introduced the notion of †recursive functions and found the Gödel solution of Einstein's equations of relativity. In addition to mathematical works, he left philosophical papers on set theory and the foundations of mathematics.

Reference

[1] K. Gödel, The consistency of the axiom of choice and of the generalized continuum hypothesis, Princeton Univ. Press, 1940.

185 (I.8) Gödel Numbers

A. General Remarks

K. Gödel [1] devised the following method to prove his incompleteness theorems (→ Section C).

Let $\mathfrak{S}$ be a †formal system. In this article, we call its basic symbols, †terms, †formulas, and formal proofs the "constituents" of $\mathfrak{S}$. Let g be an †injection from the constituents of $\mathfrak{S}$ into the natural numbers satisfying the following two conditions: (1) Given a constituent C, we can compute the value $g(C)$ in a finite number of steps. (2) Given a natural number n, there exists a finitary procedure to find out whether there exists a constituent C of $\mathfrak{S}$ such that $g(C)=n$; furthermore, when such a C exists, it can actually be specified in a finite number of steps.

If such a mapping g is given for the system $\mathfrak{S}$, then the mapping g is called a **Gödel numbering** and the number $g(C)$ is called the **Gödel number** of the constituent C (with respect to g).

B. An Example of Gödel Numbers

(1) Let $\alpha_0, \alpha_1, \ldots$ be the basic symbols of $\mathfrak{S}$. With each α_i we associate a distinct odd number q_i: $g(\alpha_i)=q_i$ $(i=0,1,\ldots)$. (2) Let F be a constituent of $\mathfrak{S}$. If F is constructed from any other constituents $F_0, F_1, \ldots, F_k$ of $\mathfrak{S}$ by a rule peculiar to $\mathfrak{S}$ (for convenience we write this $F=(F_0, F_1, \ldots, F_k)$), and if, for each F_i, $g(F_i)$ is already defined, then we put $g(F)=\langle g(F_0), g(F_1), \ldots, g(F_k)\rangle$, where $\langle a_0, a_1, \ldots, a_k\rangle$ denotes the number $p_0^{a_0} p_1^{a_1} \ldots p_k^{a_k}$ (p_i is the $(i+1)$st prime number). For example, suppose that $\mathfrak{S}$ contains 0, $=$, v_j (variables), and $\neg$ (negation) among the basic symbols, and let their Gödel numbers be 7, 9, 11^{j+1}, and 13, respectively. Since the formula $\neg\,(0=v_j)$ can be analyzed in the form $(\neg, (0, =, v_j))$, its Gödel number is $\langle 13, \langle 7, 9, 11^{j+1}\rangle\rangle$. For details → [1, 4].

C. Gödel's Incompleteness Theorems

By means of a Gödel numbering any metamathematical notion about the constituents of a formal system $\mathfrak{S}$ can be interpreted as a number-theoretic notion. For example, the notion "formula" is interpreted as the number-theoretic predicate Form(x) which means that x is the Gödel number of a formula. The provability of a formula is interpreted as the number-theoretic predicate Prov(x), which means that x is the Gödel number of a prov-

able formula; accordingly, for any formula A of $\mathfrak{S}$, the proposition $\text{Prov}(g(A))$ means that A is provable. This interpretation is called the **arithmetization** of metamathematics.

Let the formal system $\mathfrak{S}$ include formal elementary number theory. Then many metamathematically useful number-theoretic predicates can be expressed by the respective formulas of $\mathfrak{S}$; for example, there exist formulas $\mathbf{Form}(x)$ and $\mathbf{Prov}(x)$ expressing the predicates $\text{Form}(x)$ and $\text{Prov}(x)$, respectively. Furthermore, Gödel proved the existence of a closed formula U such that the formula $\neg\mathbf{Prov}(g(U)) \leftrightarrow U$ is provable. This closed formula U is one of the so-called **formally undecidable propositions**, and in fact it is shown that neither U nor $\neg U$ is provable in $\mathfrak{S}$ if $\mathfrak{S}$ is consistent in a strong sense. This result is called Gödel's **first incompleteness theorem**.

By use of the formulas $\mathbf{Form}(x)$ and $\mathbf{Prov}(x)$ the consistency of the formal system $\mathfrak{S}$ is expressed by the formula **Consis** which is an abbreviation of $\exists x(\mathbf{Form}(x) \wedge \neg\mathbf{Prov}(x))$. Gödel obtained the result that the formula **Consis** is not provable in $\mathfrak{S}$ if $\mathfrak{S}$ is consistent, on the basis of the following three facts: (1) U is not provable in $\mathfrak{S}$ if $\mathfrak{S}$ is consistent; (2) $\mathbf{Consis} \rightarrow \neg\mathbf{Prov}(g(U))$ is provable in $\mathfrak{S}$; (3) $\neg\mathbf{Prov}(g(U)) \rightarrow U$ is not provable in $\mathfrak{S}$. This result is called Gödel's **second incompleteness theorem**.

The method of arithmetization is important and useful in the study of mathematical logic. The notion of Gödel numbers of recursive functions is one of its applications (→ 356 Recursive Functions).

D. Tarski's Theorem Concerning Truth Definitions

Let a consistent formal system $\mathfrak{S}$ and a †model of $\mathfrak{S}$ be given. By means of a Gödel numbering, the truth notion of closed formula is interpreted as the number-theoretic predicate $\text{Tr}(x)$ which means that x is the Gödel number of a true closed formula; accordingly, for any closed formula A of $\mathfrak{S}$, the proposition $\text{Tr}(g(A))$ means that A is true.

In relation to the foregoing fact, if there exists a formula $\mathbf{Tr}(x)$ of a single variable and the formula $\mathbf{Tr}(g(A)) \leftrightarrow A$ is provable for every closed formula A, then that formula $\mathbf{Tr}(x)$ is called a **truth definition**. A. Tarski proved the fact that there is no truth definition in $\mathfrak{S}$ if the formal system $\mathfrak{S}$ is consistent.

References

[1] K. Gödel, Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I, Monatsh. Math. Phys., 38 (1931), 173–198; English translation, On formally undecidable propositions of Principia mathematica and related systems, Oliver & Boyd, 1962.

[2] A. Tarski, Der Wahrheitsbegriff in den formalisierten Sprachen, Studia Philosophica, 1 (1936), 261–405, translated from the Polish original 1933; English translation, The concept of truth in formalized languages, logic, semantics, metamathematics, Oxford Univ. Press, 1956.

[3] J. B. Rosser, Extentions of some theorems of Gödel and Church, J. Symbolic Logic, 1 (1936), 87–91.

[4] S. C. Kleene, Introduction to metamathematics, North-Holland and Noordhoff, 1952.

[5] S. C. Kleene, Mathematical logic, Wiley, 1967.

[6] J. R. Shoenfield, Mathematical logic, Addison-Wesley, 1967.

186 (XVI.12) Graph Theory

A. Overview of Graph Theory

Two aspects of graphs are the object of **graph theory**. One is that a graph expresses a binary relation over a set V, and the other is the fact that a graph is a CW-complex of 1 dimension, an object of study in algebraic topology. Because of its special natural as an object of 1 dimension, we can consider various concrete properties in detail. Hence graph theory has close connection with other areas, such as network theory, system theory, automata theory, and the theory of computational processes, and it has many useful applications.

The notion of a "graph" as currently used in graph theory was first discussed by L. Euler [1]. It is said that J. J. Sylvester coined the word "graph" as we presently understand it. However, up to now, there were few unifying principles, and graph theory seemed like a large collection of miscellaneous problems and ad hoc techniques. The terminology has not yet been standardized; different usages prevail in different schools. The resulting confusion can be seen in references books such as [2]–[5]. Recently, infinite graphs have also been studied. But here we restrict ourselves to finite graphs, since only they give a typical theory.

B. Definition of Graph

The notion of a **graph** $G=(V,E,\partial^+,\partial^-)$ is a composite notion of two finite sets V and E

and two maps $\partial^+: E \to V$ and $\partial^-: E \to V$. An element of V is called a **vertex**, and an element of E is called an **edge**. The maps $\partial^\pm$ are called **incidence relations**. The terms **point**, or **node** instead of vertex, and **arc**, **line**, or **branch** instead of edge have also been used (sometimes with slightly different meanings). For an edge $e \in E$, $\partial^+ e \in V$ is called the **initial vertex** and $\partial^- e \in V$ is called the **terminal vertex**. Both are called the **end vertices** of e. The inverse map $\delta^\pm$ of $\partial^\pm: V \to 2^E$ is defined by $\delta^\pm v = \{e \in E \mid \partial^\pm e = v\}$ and has the following properties: (i) If $v \neq v'$, then $\delta^+ v \cap \delta^+ v' = \delta^- v \cap \delta^- v' = \varnothing$. (ii) $\bigcup_{v \in V} \delta^+ v = \bigcup_{v \in V} \delta^- v = E$. Conversely, if the maps $\delta^\pm: V \to 2^E$ have the properties (i) and (ii), then there exist corresponding maps $\partial^\pm$. For a vertex $v \in V$, $|\delta^+ v|$ is called the **outdegree** or positive degree, $|\delta^- v|$ is called the **indegree** or negative degree, and the sum $|\delta^+ v| + |\delta^- v|$ is called the **degree**. We always have the relation $\sum_{v \in V} |\delta^+ v| = \sum_{v \in V} |\delta^- v| = |E|$, and $\sum_{v \in V}(|\delta^+ v| + |\delta^- v|) = 2|E|$. The number of vertices with odd degree is always even. Two end vertices of an edge are called mutually **adjacent**, and two edges with at least one common end vertex are also called mutually **adjacent**. An edge satisfying $\partial^+ e = \partial^- e$ is called a **self-loop**, and a vertex satisfying $\delta^+ v = \delta^- v = \varnothing$ is called an **isolated vertex**.

When permutation groups P_V operating over V and P_E operating over E are given, we can naturally define the permutation (π_V, π_E) over the graph $G(\pi_V \in P_V, \pi_E \in P_E)$. A graph is classified into equivalence classes by P_V, P_E. When P_V, P_E are all the permutations of V, E, respectively, each equivalence class is called an **unlabeled graph**, whereas when P_V, P_E consist only of an identity transformation, G is called a **labeled graph**.

For a given graph $G = (V, E, \delta^+, \delta^-)$ and a subset $E' \subset E$, we can define the **reoriented graph** of edges in E' by $G' = (V, E, \partial'^+, \partial'^-)$, where $\partial'^\pm e = \partial^\pm e$ if $e \notin E'$ and $\partial'^\pm e = \partial^\mp e$ if $e \in E'$. We have an equivalence relation by identifying the reoriented graphs. Each class by this equivalence relation is called an **undirected graph** or an **unoriented graph**. In contrast, the graph in the original sense is called a **directed graph** or an **oriented graph**.

C. Examples of Special Graphs

(1) A **complete graph** is a graph such that there exists one and only one edge with end vertices v and v' for every different pair of vertices v and v'. (2) A **bipartite graph** is a graph with a partition $V^+ \cup V^- = V$ such that $V^+ \cap V^- = \varnothing$ and for every $e \in E$, we have $\partial^+ e \in V^+$ and $\partial^- e \in V^-$. If there always exists an edge e with $\partial^+ e = v$, $\partial^- e = v'$ for every pair $v \in V^+$ and $v' \in V^-$, it is called a **complete bipartite graph**. (3) A **regular** graph is a graph whose degree at each vertex is the same. (4) A **partial graph** is defined as follows: Let E' be a subset of E in a graph $G = (V, E, \partial^+, \partial^-)$. The graph $G' = (V', E', \partial'^+, \partial'^-)$ is called a **partial graph** of G, where $V' = \partial^+ E' \cup \partial^- E'$, and $\partial'^\pm$ are the restrictions of $\partial^\pm$ on E', respectively. When $E' = E$, the partial graph is the graph obtained by deleting all isolated vertices from G. Similarly, for a subset V' of V, the graph $G'' = (V'', E'', \partial''^+, \partial''^-)$, defined by $E'' = \delta^+ V'' \cup \delta^- V''$ and with $\partial''^\pm$ being the restriction of $\partial^\pm$ on E'', is called a **section graph** of G.

D. Representation of a Graph

When we want to process by computer a problem concerning a graph, we must represent a graph in a suitable form (→ 96 Data Processing). A commonly used method is the following **list representation**: Let $V = \{v_1, v_2, \ldots, v_M\}$ and $E = \{e_1, e_2, \ldots, e_n\}$. For each $v_\alpha \in V$ $(\alpha = 1, \ldots, M)$, we arrange the edges in $\delta^+ v_\alpha$ and in $\delta^- v_\alpha$ in suitable orders. Then (i) for each $v_\alpha \in V$, record $|\delta^+ v_\alpha|$, $|\delta^- v_\alpha|$, B_α^+, $\bar{B}_\alpha^+$, B_α^-, and $\bar{B}_\alpha^-$, where $B_\alpha^\pm$ and $\bar{B}_\alpha^\pm$ are the first and the last index of the edges in $\delta^\pm v_\alpha$, respectively. We define the corresponding values to be 0 if the $\delta^\pm v_\alpha$ are empty. (ii) For each $e_\kappa \in E$, record ∂_κ^+, ∂_κ^-, E_κ^+, $\bar{E}_\kappa^+$, E_κ^-, and $\bar{E}_\kappa^-$, where $\partial_\kappa^\pm$ are the number of vertices $\partial^\pm e_\kappa$, and $E_\kappa^\pm$, $\bar{E}_\kappa^\pm$ are the number of edges immediately after or before e_κ among the edges with the same initial (for $^+$) or terminal (for $^-$) vertex as that of e_κ. If e_κ is the last or the first among such edges, we define the corresponding values to be 0.

E. Deformation of a Graph

Let $G = (V, E, \partial^+, \partial^-)$ be a graph. A graph obtained by **opening** an edge e is the graph $G' = (V, E - \{e\}, \partial'^+, \partial'^-)$, where $\partial'^\pm$ is the restriction of $\partial^\pm$ on $E - \{e\}$. The graph obtained by deleting all isolated vertices from G' is a partial graph of G over $E - \{e\}$. A graph obtained by **shortening** an edge e is the graph $G'' = (V'', E - \{e\}, \delta''^+, \partial''^-)$, where $V'' = (V - \{\partial^+ e, \partial^- e\}) \cup \{\hat{v}\}$, $\hat{v}$ being a new vertex not contained in V, and $\partial''^\pm = \varphi \cdot \partial^\pm$, $\varphi: V \to V''$ being defined by $\varphi(v) = v$ for $v \neq \partial^\pm e_\kappa$ and $= \hat{v}$ when $v = \partial^+ e_\kappa$ or $v = \partial^- e_\kappa$. A graph obtained by opening and shortening several different edges is similarly defined, and the result is independent of the order of opening or shortening process. A graph obtained by opening edge(s) or by shortening edge(s) or by both processes is called a **subgraph**, a **contraction**, or a **subcontraction**, respectively.

F. Connectedness

Let $G=(V,E,\partial^+,\partial^-)$ be a graph, and let v_α and v_β be two vertices. A **path** from v_α to v_β of **length** l is a sequence $P=(v_\alpha=v_{\alpha_0},\varepsilon_1 e_{\kappa_1},v_{\alpha_1},\varepsilon_2 e_{\kappa_2},\ldots,\varepsilon_l e_{\kappa_l},v_{\alpha_l}=v_\beta)$, where ε_i are $+1$ or -1, and for every $i=1,\ldots,l$, we have $\partial^+ e_{\kappa_i}=v_{\alpha_{i-1}}$, $\partial^- e_{\kappa_i}=v_{\alpha_i}$ if $\varepsilon_i=+1$, and $\partial^+ e_{\kappa_i}=v_{\alpha_i}$, $\partial^- e_{\kappa_i}=v_{\alpha_{i-1}}$ if $\varepsilon_i=-1$. When $v_\alpha=v_\beta$, it is called a **closed path** considering P as a cyclic sequence. If in the sequence P of a path no edge appears more than once, it is called a **simple path**. Similarly, if no vertex appears more than once, it is called **elementary**. If all $\varepsilon_i=+1$, it is called a **direct path. Direct closed paths**, etc., are similarly defined.

If we define $v\sim v'$ by the existence of a path from v to v', this $\sim$ is an equivalence relation. Let us denote the equivalence classes by $V_1,\ldots,V_s$. The section graph $G_i=(V_i,E_i,\partial_i^+,\partial_i^-)$ determined by V_i is called a **connected component**, or simply a **component**, of G. The sets $E_1,\ldots,E_s$ are mutually disjoint and the union is E. If we denote by $v\rightarrow v'$ the existence of a direct path from v to v', the relation $\rightarrow$ is a †pseudo-order. The relation $v\leftrightarrow v'$ defined by $v\rightarrow v'$ and $v'\rightarrow v$ is an equivalence relation in V, and the equivalence classes $\tilde{V}_1,\ldots,\tilde{V}_t$ or the section graphs $\tilde{G}_i=(\tilde{V}_i,\tilde{E}_i,\tilde{\partial}_i^+,\tilde{\partial}_i^-)$ determined by $\tilde{V}_i$ are called **strongly connected components** of G. $\tilde{E}_1,\ldots,\tilde{E}_t$ are mutually disjoint, but their union is not always E. Among the sets $\tilde{V}_1,\ldots,\tilde{V}_t$, we can define an order relation $\tilde{V}_i\rightarrow\tilde{V}_j$ by the existence of $v\in\tilde{V}_i$, $v'\in\tilde{V}_j$, $v\rightarrow v'$. Classification by $\leftrightarrow$ is a refinement of that by $\sim$. When $s=1$, the graph G is called **connected**, and when $t=1$, G is called **strongly connected**.

For a different pair $v,v'\in V$, a set $S\subset V-\{v,v'\}$ is called a **separator** of v and v' if every path from v to v' contains at least one vertex of S. When every separator S for any pair v, v' has at least k elements, the graph G is called k-**connected**. For $k\geqslant 3$, there are several variations of the definition of k-connectedness. 1-connectedness is equivalent to connectedness in the sense defined above.

A simple path containing all edges of a graph is called an **Euler path**. Euler direct paths, etc., are similarly defined. A graph with an Euler path is called an **Euler graph**. A graph is an Euler graph if and only if (i) it is connected, and (ii) the number of vertices with odd degrees is 0 or 2 (Euler's unicursal graph theorem [1]). Similar results are known for Euler closed paths or Euler direct paths. An elementary path containing all vertices of a graph is called a **Hamilton path**. Hamilton closed paths. etc., are similarly defined. The criteria for the existence of a Hamilton path for any given graph is unknown. This is known to be an †NP-complete problem (→ 71 Complexity of Computations). Various sufficient conditions or necessary and sufficient conditions for special graphs have been given (→ e.g., [6]).

G. Tieset, Cutset, Tree, and Cotree

For a graph $G=(V,E,\partial^+,\partial^-)$, we define its **incidence matrix** $[D_\kappa^\alpha$ $(\alpha=1,\ldots,M$ $(=|V|),\kappa=1,\ldots,n$ $(=|E|))]$ by defining D_κ^α to be 1 if $v_\alpha=\partial^+e_\kappa\neq\partial^-e_\kappa$, -1 if $v_\alpha=\partial^-e_\kappa\neq\partial^+e_\kappa$, and 0 if $\partial^+e_\kappa=\partial^-e_\kappa$ or $e_\kappa\notin\delta^+v_\alpha\cup\delta^-v_\alpha$. Similarly we define its **adjacement matrix** $[\Gamma_\beta^\alpha$ $(\alpha,\beta=1,\ldots,M)]$ by defining Γ_β^α to be 0 if v_α and v_β are not mutually adjacent and 1 if v_α and v_β are mutually adjacent.

A set of edges forming a closed path is called a **tieset**. A set of edges of the form $\{e\mid\partial^+e\in W,\partial^-e\in V-W\}\cup\{e\mid\partial^-e\in W,\partial^+e\in V-W\}$ for a partition $(W,V-W)$ of V is called a **cutset**. A maximal subset of edges containing no tiesets is called a **tree**, and a maximal subset of edges containing no cutsets is called a **cotree**. A tree is sometimes called a **spanning tree** of G. Every tree is the complement (with respect to E) of a cotree, and vice versa.

The number of elements of a tree is always the same and equal to the †rank m of the incidence matrix. This value m is called the **rank** of the graph G. Similarly, the number k of elements of a cotree is always the same, which is called the **nullity** or the **cyclomatic number** of the graph G. Always, $k=n-m$.

Let K be a field, K^n, K^M be vector spaces over K of dimensions n and M, respectively, and K^{n*}, K^{M*} be the †dual spaces of K^n, K^M. The incidence matrix defines two mutually contragradient linear mappings $\partial:K^n\rightarrow K^M$ and $\delta:K^{M*}\rightarrow K^{n*}$ with respect to their canonical bases. A minimal set among the family of supports (in E) of nonzero vectors in the kernel $\operatorname{Ker}\partial$ is a tieset corresponding to an elementary closed path. Similarly, a minimal set among the family of supports (in E) of nonzero vectors of the image $\operatorname{Im}\delta$ is a minimal cutset.

Let T $(\subset E)$ be a tree and $\bar{T}=E-T$ be a cotree. Let us renumber the edges so that $T=\{e_1,\ldots,e_m\}$, $\bar{T}=\{e_{m+1},\ldots,e_n\}$. For each $e_{m+p}\in\bar{T}$, there is a unique vector $R_p^\kappa(p=1,\ldots,k(=n-m))$ in $\operatorname{Ker}\partial$ whose support is in $\{e_{m+p}\}\cup T$ and whose e_{m+p}- equals $+1$. Similarly, for each $e_a\in T$ $(a=1,\ldots,m)$, there is a unique vector D_κ^a $(a=1,\ldots,m)$ in $\operatorname{Im}\partial$ whose support is in $\{e_a\}\cup\bar{T}$ and whose e_a- component equals $+1$. $\{R_1^\kappa,\ldots,R_k^\kappa\}$ is a basis of $\operatorname{Ker}\partial$, and $\{D_\kappa^1,\ldots,D_\kappa^m\}$ is a basis of $\operatorname{Im}\partial$. Both matrices $R_p^\kappa(k\times n)$ and $D_\kappa^a(m\times n)$ are **totally unimodular**, i.e., every minor determinant is 0, $+1$, or -1. Furthermore, the following relations hold: $R_p^{m+q}=\delta_p^q$ $(p,q=1,\ldots,k)$, $D_b^a=\delta_b^a$ $(a,b=1,\ldots,m)$, R_p^a+

$D^a_{m+p}=0$ $(a=1,\dots,m; p=1,\dots,k)$. The matrices R^κ_p and D^a_κ are called the **fundamental tieset matrix** and the **fundamental cutset matrix**, respectively, with respect to the tree-cotree pair $(T,\bar{T})$. The minor of the matrix R^κ_p consisting of all rows and k columns $\kappa_1,\dots,\kappa_k$ is $+1$ or -1 if and only if $\{e_{\kappa_1},\dots,e_{\kappa_k}\}$ is a cotree, and 0 otherwise. Similarly, the minor of the matrix D^a_κ consisting of all rows and m columns $\kappa_1,\dots,\kappa_m$ is $+1$ or -1 if and only if $\{e_{\kappa_1},\dots,e_{\kappa_m}\}$ is a tree, and 0 otherwise.

H. Planarity of a Graph

Let there be a natural one-to-one correspondence between the sets of edges of two graphs $G_i=(V_i,E_i,\partial_i^+,\partial_i^-)$ $(i=1,2)$, and suppose that under the correspondence a tree T_1 in G_1 corresponds to a tree T_2 in G_2 and that the fundamental cutset matrices D_1 and D_2 with respect to (T_i,E_i-T_i) are mutually equal. Then G_1 and G_2 are said to be **2-isomorphic**. The definition is equivalent to the one given by the equality of fundamental tieset matrices. 2-isomorphism is an equivalence relation. If G_1 and G_2 are 2-isomorphic, the families of trees, cotrees, tiesets and cutsets are mutually corresponding. The coincidence of one of the families is a sufficient condition for the 2-isomorphism of G_1 and G_2 as undirected graphs. A 3-connected graph has no 2-isomorphic graph other than itself.

Similarly, when a tree T_1 of G_1 corresponds to a cotree $\bar{T}_2$ of G_2 and if the fundamental cutset matrix of G_1 with respect to (T_1,E_1-T_1) is equal to the fundamental tieset matrix of G_2 with respect to (E_2-T_2,T_2) as matrices, then we say that G_2 is **dual** to G_1. In this case, G_1 is dual to G_2 also. If G_1 is dual to G_2 and G_2 is dual to G_3, then G_1 and G_3 are 2-isomorphic. The duality is a relation among the equivalence classes of graphs by 2-isomorphisms.

Any graph $G=(V,E,\partial^+,\partial^-)$ can be "drawn" in 3-dimensional Euclidean space in the following sense: Each vertex is a (distinct) point, and an edge e is an arc connecting two points ∂^+e and ∂^-e with a direction pointing from ∂^+e to ∂^-e in such a way that no two arcs intersect. A graph representable on a plane or on a 2-dimensional sphere in the above sense is called a **planar graph**. A graph G is planar if and only if it has a dual graph (H. Whitney [7]). Another necessary and sufficient condition is that as an undirected graph, neither the complete five-point graph K_5 nor the bipartite complete graph of three-three points $K_{3,3}$, appear in any subcontraction of the graph G. (This is a version of the criterion of C. Kuratowski [8].)

I. Coloring of Graphs

A **coloring** of the vertices of a graph $G=(V,E,\partial^+,\partial^-)$ is a mapping ψ from V to the set of integers N satisfying the condition $\psi(v)\neq\psi(v')$ for all adjacent vertices v and v'. $\gamma(G)=\min\{|\psi(V)|\mid\psi$ is a coloring of the vertices of $G\}$ is called the **chromatic number** of G. If the graph G is drawn over a 2-dimensional closed surface with a 1-dimensional †Betti number b, we have $\gamma(G)\leqslant\lfloor(7+\sqrt{1+24b})/2\rfloor$, where $\lfloor\ \rfloor$ denotes the integral part. Except for the †Klein bottle (with $b=2$), where $\gamma(G)\leqslant 6$, this is the best possible, i.e., there exists a graph whose $\gamma(G)$ equals the upper bound on the right-hand side. As for $b\geqslant1$, the inequality was first proved by P. J. Heawood (1890), and final results were established by J. W. T. Young and G. Ringel [10]. When $b=0$, $\gamma(G)\leqslant 5$ was shown by A. B. Kempe (1879), but the **four color conjecture**: "$\gamma(G)\leqslant4$?" has remained unsolved for more than a hundred years. The conjecture is believed to have been solved affirmatively recently through checking a huge number of cases on a large computer [11, 12].

A subset $W\subset V$ is called an **independent set** or an **internally stable set** if no two vertices in W are mutually adjacent. $\alpha(G)=\max\{|W|\mid W$ is an independent set of $G\}$ is called the **number of independence** of G. A subset W of V is called a **dominating set** or **externally stable set** of G if every vertex $v\in V$ is either $v\in W$ or adjacent to a vertex of W. $\beta(G)=\min\{|W|\mid W$ being a dominating set of $G\}$ is called the **number of domination** of G. For every graph G, we have $\alpha(G)\geqslant\beta(G)$ and $\gamma(G)\cdot\alpha(G)\geqslant|V|$.

J. Decision Problems and Graphs

There are many interesting topics in decision problems concerning graphs, especially from the standpoint of †complexity of computations (→ e.g., [13]). The following are some typical problems: (i) Problems for which algorithms of polynomial order are known: Is the given graph k-connected, strongly connected, a Euler graph, or a planar graph? (ii) †NP-complete problems: Is the given graph a Hamilton graph? Do we have $\alpha(G)=k$, $\beta(G)=k$, or $\gamma(G)=k$?

Let G_1 and G_2 be two graphs. The problem of whether they coincide as unlabeled graphs is called the **isomorphism problem**, and has been studied for many years in connection with problems concerning the structure of chemical compounds. Unfortunately, no algorithm of polynomial order is known; nor do we know whether this is an NP-complete problem. As for the isomorphism problem for

planar graphs, algorithms of polynomial order are known.

K. Perfectness Theorem

Let the number of independence and the chromatic number of a graph $G=(V,E,\partial^+,\partial^-)$ be $\alpha(G)$ and $\gamma(G)$, respectively. Let $V_1,\ldots,V_r$ be a disjoint decomposition of V. We denote by $\theta(G)$ the minimal number of r such that every section graph of G over V_i contains a complete graph. Furthermore, we denote by $\omega(G)$ the maximum value of $|W|$ for a subset $W\subset V$ such that the section graph of G over W contains a complete graph. We always have $\alpha(G)\leqslant\theta(G)$ and $\gamma(G)\leqslant\omega(G)$. G is called α-**perfect** if every section graph H of G satisfies $\alpha(H)=\theta(H)$. Similarly, G is called γ-**perfect** if $\gamma(H)=\omega(H)$ for every section graph H of G. The conjecture that γ-perfectness and α-perfectness are equivalent has recently been solved affirmatively [14]. This is called the **perfectness theorem**.

References

[1] L. Euler, Solutio problematis ad geometriam situs pertinentis, Commentarii Academiae Petropolitanae, 8 (1736), 128–140.
[2] F. Harary, Graph theory, Addison-Wesley, 1969.
[3] C. Berge, Graphes et Hypergraphes, Dunod, second edition, 1974.
[4] A. A. Zykov, Finite graph theory (in Russian), Nauka, 1969.
[5] M. Iri, Network flow, transportation and scheduling, Academic Press, 1969.
[6] H. Walther and H.-J. Voss, Über Kreise in Graphen, VEB Deutscher Verlag der Wissenschaften, 1974.
[7] H. Whitney, Non-separable and planar graphs, Trans. Amer. Math. Soc., 34 (1932), 339–362.
[8] C. Kuratowski, Sur le problème des courbes gauches en topologie, Fund. Math., 15 (1930), 271–283.
[9] F. Harary and W. Tutte, A dual form of Kuratowski's theorem, Canad. Math. Bull., 8 (1965), 17–20, 373.
[10] G. Ringel, Map color theorems, Springer, 1974.
[11] K. Appel and W. Haken, Every planar map is four colourable, I. Discharging, Illinois J. Math., 21 (1977), 429–489; K. Appel, W. Haken and J. Koch, II. Reducibility, Illinois J. Math., 21 (1977), 490–567.
[12] S. Hitotumatu, Four color problem (in Japanese), Kodansha, 1978.
[13] A. V. Aho, J. E. Hopcroft, and J. D. Ullman, The design and analysis of computer algorithms, Addison-Wesley, 1974.
[14] L. Lovász, Normal hypergraphs and perfect graph conjecture, Discrete Math., 2 (1972), 253–267.

187 (XXI.2) Greek Mathematics

It is generally believed that theoretical mathematics originated with the Greeks. The Greeks learned the arts of land surveying and commercial arithmetic from earlier civilizations; but they developed theoretical mathematics themselves, toward the middle of the 4th century B.C. The creation of a mathematics that transcends practical purposes was one of the most remarkable events in the history of human culture, and one that had an immense impact on the development of all branches of science. We owe the reestablishment of important Greek mathematical texts and the reconstruction of the development of Greek mathematics to the historians of the 19th century (the oldest extensive exposition of the history of mathematics before Euclid is due to Proklos (Proclus) (410–485)).

The earliest known Greek mathematicians are Thales of Miletus (c. 639–546 B.C.) and Pythagoras of Samos (fl. 510? B.C.). Both were Ionians, but the latter went to what is now southern Italy and founded a semireligious school whose members called themselves Pythagoreans. Their motto was "Everything is number"; their studies were called *mathema* ("what is learned") and consisted of music, astronomy, geometry, and arithmetic (the subject group called the **quadrivium**, which formed the core of medieval and later higher education) "for the purification of soul." Their research delved into the theories of proportion (in relation to music) and †polygonal numbers (triangular numbers, square numbers, etc.), and more generally into the theory of numbers and geometric algebra. It is said that they knew of the irrationality of $\sqrt{2}$, though no evidence of this has been found. Even after the demise of the Pythagorean school, its followers continued to promote mathematics in collaboration with the Academy of Plato.

Another significant school was the Eleatic. Among its members Zeno (c. 490–c. 430 B.C.) is especially important. †Zeno's paradoxes are arguments leading to absurdity. Some see within them the origin of logical reasoning and, consequently, of theoretical mathematics [3]. It is chronologically difficult to attribute

to Zeno the consideration of the continuum and irrational numbers, but we can find in him the impetus toward atomistic reasoning. The computation of the volume of pyramids (by dividing them into "atomistic" laminae) by Democritus (fl. 430? B.C.) and the atomistic calculation of the area of circles by Antiphon (fl. 430 B.C.) came shortly after the time of Zeno.

The middle decades of the 4th century B.C. are known as the Age of Pericles, the Golden Age of Athens. The †trisection of an angle, the †duplication of a cube, and the †quadrature of a circle, known at that time as the "**three big problems**" (→ 179 Geometric Construction), were studied by the Sophists. Hippias of Elis (fl. 420 B.C.), Hippocrates of Chios (fl. 430 B.C. in Athens), Archytas of Taras (c. 430–365 B.C.), Menaechmus (fl. 350 B.C.), and his brother Dinostratus (fl. 350 B.C.) solved these problems using conic sections and the **quadratrix** (a transcendental curve whose equation is $y = x\cot(\pi x/2)$).

By 400 B.C. Athens had lost its political influence, but it remained the center of Greek culture. It was during this time that Plato's Academy flourished, and Plato (427–347 B.C.) and his followers laid particular importance on mathematics. Archytas, Menaechmus, and Dinostratus belonged to or were closely associated with this school. During the first fifty years of the Academy, research in the following fields was pursued: methodology of mathematics or science in general (i.e., dialectics, analysis, synthesis); geometric reconstruction of Mesopotamian algebra; the theory of irrationals in relation to the geometrization of algebra (Theodorus of Cyrene (5th century B.C.), who was Plato's teacher, as well as Theaitetus of Athens (415?–369 B.C.) contributed to this study, and the general theory of proportion by Eudoxus of Cnidos (c. 408–c. 355 B.C.) also belongs to this field); the method of exhaustion (by Eudoxus); and studies of the "three big problems" and conic sections. It was this school in which the term *mathema* came to be used in its present sense of "mathematics" rather than in the sense of disciplines in general.

The conquests of Alexander the Great accelerated the already considerable cultural influence of Athens. Later, during the Ptolemaic period, the center of culture moved to Alexandria. The Mouseion at Alexandra, the combined library and university, is said to have possessed hundreds of thousands of volumes.

At Alexandria, Euclid (c. 300 B.C.) compiled his *Elements*, which became a model for scientific works for centuries to come—Newton's *Principia* as well as Spinoza's *Ethics* are modeled on it. Most historians say that Euclid's method derives from Aristotle (384–322 B.C.), who, after studying at Plato's Academy, founded a new school, the Peripatetics, whose doctrines are at many points opposed to those of the Academy. Others, however, see the origin of Euclid's axiomatic method in the Eleatics [5]; we can find prototypes of some parts of the *Elements* in both Oenopides, who lived during the time of Zeno, and in Hippocrates.

The third century B.C. was the Golden Age of Greek mathematics. Archimedes of Syracuse (c. 282–212 B.C.) was the greatest mathematician, mechanic, and technician of antiquity. He did important work in mathematics, studying the exact quadrature of the parabola. According to his *ephodos* (method), he would obtain a result by mechanical experiments and then prove it by the method of "exhaustion." He also computed the value of π; studied spirals and other curves, spheres, and circular cylinders; contributed to the development of statics and optics and their application; and had a profound influence on later mathematicians. During the same period, Apollonius of Perga (fl. 210 B.C.) wrote *Konikon biblia* (Books on Conics) in eight books, of which the last has been lost. The geometric theory of †conic sections contained in this work is not much different from the one we know today; it had a great influence on 17th-century scientists especially. Other mathematicians of this period worth noting are Eratosthenes (c. 275–195 B.C.), who conceived the †sieve method of finding prime numbers and who measured the earth, and Hipparchus (fl. 150 B.C.), called the father of astronomy, who made a table of sines.

Hellenistic influence began to decline in the first century B.C., and the influence of Alexandria decreased. The Mouseion burned in 48 B.C., but was rebuilt. Among the mathematicians of this time, we may count Heron (fl. 60? A.D.); Menelaus (fl. 100 A.D.), who wrote *Sphaerica*; Theon of Smyrna (fl. 125 A.D.); Ptolemy (fl. 150 A.D.), the author of *Almagest*; Nicomachus (50?–150? A.D.), the author of *Arithmetike eisagoge*; Diophantus (fl. 250? A.D.), whose career is not fully known but who wrote *Arithmetika*, of which six of the original thirteen books remained to influence †Fermat; and Pappus (fl. 300 A.D.), the last creative mathematician in Greece, who left eight books of the *Synagoge*, which influenced †Descartes and which still exist today.

The period following the fall of the Western Roman Empire was a difficult one for Greco-Egyptian science. The Mouseion was destroyed for the second time in 392 A.D. Theon of Alexandria (fl. 380) and his daughter, Hy-

patia (c. 370–415), were at that time working on commentaries on the classics. Among the few remaining works of the period is Proclus' (410–485) commentaries on the first book of Euclid's *Elements*. The Athenian Academy was closed in 529 by order of the Emperor Justinian; the last director was Simplicius, who commented on Aristotle. Soon afterward, Alexandria fell into the hands of the Moors, and many scholars fled as refugees to Constantinople, the capital of the Eastern Empire.

References

[1] T. L. Heath, A history of Greek mathematics I, II, Clarendon Press, 1921.
[2] T. L. Heath, A manual of Greek mathematics, Oxford Univ. Press, 1931 (Dover, 1963).
[3] M. Cantor, Vorlesungen über Geschichte der Mathematik I, Teubner, third edition, 1907.
[4] B. L. van der Waerden, Zenon und die Grundlagenkrise der griechischen Mathematik, Math. Ann., 117 (1940), 141–161.
[5] Á. Szabó, Anfänge des euklidischen Axiomensystems, Arch. History Exact Sci., 1 (1960), 37–106.
[6] J. L. Heiberg (ed.), Euclidis opera omnia I–XIII and suppl., Leipzig, 1883–1916.
[7] T. L. Heath, The thirteen books of Euclid's Elements I, II, III, Cambridge Univ. Press, 1908 (Dover, 1956).
[8] P. Ver Eecke (trans.), Les oeuvres complètes d'Archimède, Desclée de Brouwer & Cie, 1921 (Blanchard, 1961).
[9] P. Ver Eecke (trans.), Proclus Diadochus, les commentaires sur le premier livre des Eléments d'Euclide, Desclée de Brouwer & Cie, 1948 (Blanchard, 1959).
[10] P. Tannery, Le géométrie grecque, comment son histoire nous est parvenue et ce que nous en savons, Paris, 1887.
[11] B. L. van der Waerden, Science awakening, Noordhoff, 1954.
[12] M. Clagett, Greek science in antiquity, Abelard-Schuman, 1955 (Collier, 1963).
[13] O. Becker (ed.), Zur Geschichte der griechischen Mathematik, Darmstadt, 1965.
[14] Á. Szabó, Anfänge der griechischen Mathematik, Oldenbourg, 1969.
For the history of mathematics in general,
[15] R. C. Archibald, Outline of the history of mathematics, Mathematical Association of America, sixth edition, 1949.
[16] N. Bourbaki, Eléments d'histoire des mathématiques, Hermann, second edition, 1969.
[17] M. Cantor, Vorlesungen über Geschichte der Mathematik I–IV, Teubner, second edition, 1894–1908.
[18] E. Montucla, Histoire des mathématiques I–IV, Paris, new edition, 1799–1802.
[19] D. E. Smith, A source book in mathematics, McGraw-Hill, 1929 (Dover, 1959).
[20] D. J. Struik, A concise history of mathematics, Dover, third edition, 1967.

188 (XIII.28) Green's Functions

A. General Remarks

Green's functions are usually considered in connection with †boundary value problems for ordinary differential equations and also with †elliptic and †parabolic partial differential equations. For example, consider boundary value problems for the †Laplacian in 3-dimensional space: $L[u]=(\partial^2/\partial x_1^2+\partial^2/\partial x_2^2+\partial^2/\partial x_3^2)u$. Let D be a bounded domain with a smooth boundary S and the boundary condition B on S be either $u(x)=0$ $(x\in S)$ (the first kind) or $\partial u/\partial n+\beta u=0$ $(x\in S)$ (the third kind), where n is the outer normal of unit length, $\beta(x)\geqslant 0$, and $\beta(x)\not\equiv 0$. We say that the function $g(x_1,x_2,x_3;\xi_1,\xi_2,\xi_3)$ is the **Green's function** of L (or the partial differential equation $L[u]=0$) relative to the boundary condition B, when (i) $g(x,\xi)$ satisfies $L_x[g(x,\xi)]=0$ except for $x=\xi$; (ii) $g(x,\xi)=-1/4\pi r+\omega(x,\xi)$, where $r=(\sum_{i=1}^3(x_i-\xi_i)^2)^{1/2}$ and $\omega(x,\xi)$ is a regular function, i.e., of class C^ν for a suitable value ν; (iii) $g(x,\xi)$ satisfies the boundary condition B, i.e., $g(x,\xi)=0$, $x\in S$ (the first kind), or $(\partial/\partial n+\beta)\,g(x,\xi)=0$, $x\in S$ (the third kind). Conditions (i) and (ii) mean that $g(x,\xi)$ is a †fundamental solution of L, i.e., $L_x[g(x,\xi)]=\delta(x-\xi)$, where $\delta(x-\xi)$ is †Dirac's measure at the point $x=\xi$. Note that if $g(x,\xi)$ is a fundamental solution, then by adding any solution u of the equation $L[u]=0$ to g, we obtain another fundamental solution $g+u$. Thus Green's function is the fundamental solution that satisfies the given boundary condition. To be more precise, in the boundary value problem, if the boundary condition is of the first kind, $g(x,\xi)$ can be obtained by adding to a fundamental solution $-1/4\pi r$ the solution $\omega(x,\xi)$ of the following Dirichlet problem: $\Delta_x\omega(x,\xi)=0$, $\omega(x,\xi)=1/4\pi r$ $(x\in S)$. We remark that there are slightly different definitions for Green's function. For example, there are cases where $g(x,\xi)$ is defined by $g(x,\xi)=1/4\pi r+\omega(x,\xi)$ or by $g(x,\xi)=1/r+\omega(x,\xi)$.

In the case in the previous paragraph, Green's functions satisfying the boundary conditions are uniquely determined. In general, if we are given a Green's function, then

for any regular function $v(x)$ the function

$$u(x)=\int_D g(x,\xi)v(\xi)\,d\xi$$

represents the solution of $L[u]=v$ with the boundary condition B. More precisely, if $v(x)$ satisfies the †Hölder condition $|v(x)-v(x')|\leqslant L|x-x'|^\alpha$ $(0<\alpha\leqslant 1)$ (L, α positive constants), then $u(x)$ is of class C^2. Conversely, if $u(x)$ satisfies the equation $L[u]=v$ and the boundary condition B, it is represented by the formula for $u(x)$. This means that if we denote the operator that associates u to v by G, then G is the inverse operator of the Laplacian L with the boundary condition B, and Green's function is the †integral kernel of the operator G. Using this property, the boundary value problem relative to L can be reduced to a problem of †integral equations. For example, the differential equation with the boundary condition B containing the complex parameter λ, $L[u]+\lambda u=f$, is equivalent to the integral equation $u+\lambda G[u]=G[f]$, which is obtained by letting G act from the left on the above differential equation. In this way, the problem can be simplified.

In the case of general boundary value problems for higher-order elliptic operators, Green's functions are defined in the same way as before (→ 189 Green's Operator). The important case is when L and the boundary condition B define a †self-adjoint operator. In this case, Green's function is symmetric $(g(x,\xi)=g(\xi,x))$. To obtain Green's function is not easy in general. However, in some cases such functions can be obtained fairly easily (→ Appendix A, Table 15.VI).

B. Self-Adjoint Ordinary Differential Equations of the Second Order

Consider the operator $L[u]\equiv(p(x)u')'+q(x)u$ $(p(x)>0)$ defined in the interval $a\leqslant x\leqslant b$, with boundary conditions of the form $\alpha u'+\beta u=0$ at the two endpoints. Then Green's function $g(x,\xi)$ is defined in the following way: (i) For $x\neq\xi$, $L[g(x,\xi)]=0$; (ii) $[\partial g(x,\xi)/\partial x]_{x=\xi-0}^{x=\xi+0}=1/p(\xi)$; (iii) for ξ fixed, $g(x,\xi)$ satisfies the homogeneous boundary conditions at $x=a$ and $x=b$.

Conditions (i) and (ii) mean that $L[g(x,\xi)]=\delta(x-\xi)$. We can construct $g(x,\xi)$ in the following way: Let u_1 (u_2) be the solution of $L[u]=0$ satisfying the boundary condition at $x=a$ (at $x=b$). If u_1 and u_2 are linearly independent, we can satisfy $p(u_1'u_2-u_1u_2')=1$ by choosing the constants suitably. Then Green's function $g(x,\xi)$ is defined by $g(x,\xi)=u_1(x)u_2(\xi)$ for $x\leqslant\xi$, and $g(x,\xi)=u_1(\xi)u_2(x)$ for $\xi\leqslant x$. If u_1 and u_2 are linearly dependent, there exists no Green's function that satisfies conditions (i), (ii), and (iii). However, by modifying the definition, we can get a generalized Green's function playing a similar role [2]. This method can be applied to the case of ordinary differential equations of higher order.

C. The Laplace Operator

When the domain D is the n-dimensional sphere of radius a with center at the origin, Green's function of the Laplacian relative to the boundary condition $u=0$ is obtained in the following way. Let $E(r)$ be the following fundamental solution of the Laplacian: $E(r)=(2\pi)^{-1}\log r$ for $n=2$ and $E(r)=-((n-2)\cdot\omega_n r^{n-2})^{-1}$ for $n\geqslant 3$, where $\omega_n=2\pi^{n/2}\Gamma(n/2)$ is the $(n-1)$-dimensional surface area of the n-dimensional unit sphere. Then Green's function $g(x,\xi)$ is defined by

$$g(x,\xi)=E(r)-E(\rho r'/a),$$

where $\rho=(\sum_{i=1}^n\xi_i^2)^{1/2}$, $r'=(\sum_{i=1}^n(x_i-\xi_i')^2)^{1/2}$, $\xi_i'=(a/\rho)^2\xi_i$.

D. Helmholtz's Differential Equation

Let D be an exterior domain with a smooth boundary S in $\mathbf{R}^3$. In mathematical physics, the boundary value problem of finding a solution $u(x)$ of **Helmholtz's differential equation** $(\Delta+k^2)u(x)=f(x)$ $(k>0)$ satisfying $u(x)=0$ $(x\in S)$ is of particular interest. In this case, concerning the behavior of $u(x)$ at infinity, we usually assume **Sommerfeld's radiation condition**:

When $|x|\to+\infty$, $\quad u(x)=O(|x|^{-1})$,

$$\left.\left(\frac{\partial}{\partial r}u-iku\right)\right|_{|x|=r}=o(|x|^{-1}),$$

where $\partial/\partial r$ is the derivative along the radial direction. It is known that this condition ensures the uniqueness of the solution (**Rellich's uniqueness theorem**). We can construct Green's function $G(x,\xi)$ for any $k(>0)$ so that for smooth $f(x)$ with bounded †support,

$$u(x)=\int_D G(x,\xi)f(\xi)\,d\xi$$

represents the solution satisfying $u(x)=0$ $(x\in S)$ and Sommerfeld's radiation condition. Then, with

$$G(x,\xi)=-\frac{e^{ik|x-\xi|}}{4\pi|x-\xi|}+K_c(x,\xi),$$

where

$$|x-\xi|=\left(\sum_{i=1}^3(x_i-\xi_i)^2\right)^{1/2},$$

$K_c(x, \xi)$ can be obtained by solving an integral equation of †Fredholm type [3, 4]. In this case, there exists no Green's operator in L_2 space, and $G(x, \xi)$ can be considered to be a generalized Green's function [4].

E. Stokes's Differential Equation

Let D be a bounded domain in $\mathbf{R}^3$ with smooth boundary S, and consider **Stokes's differential equation** in D

$$\mu\Delta u_i = \frac{\partial p}{\partial x_i} - \rho X_i, \quad i=1,2,3, \sum_{j=1}^{3} \frac{\partial u_j}{\partial x_j} = 0,$$

where μ, ρ are positive constants. In hydrodynamics, we consider the boundary value problem of finding solutions $(u_1(x), u_2(x), u_3(x), p(x))$ of Stokes's equation satisfying the boundary condition $u_i(x)=0$, $x \in S$ $(i=1,2,3)$. In this case, **Green's tensors** $G_{ij}(x, \xi)$, $g_i(x, \xi)$ can be constructed, and for smooth functions $X_i(x)$ $(i=1,2,3)$ the unique solution of this boundary value problem is represented by

$$u_i(x) = \rho \sum_{j=1}^{3} \int_D G_{ij}(x, \xi) X_j(\xi) d\xi,$$

$$p(x) = \rho \sum_{j=1}^{3} \int_D g_j(x, \xi) X_j(\xi) d\xi$$

[5].

F. Parabolic Equations

Consider the boundary value problem (the initial boundary value problem):

$$L[u] \equiv \frac{\partial u}{\partial t} - c^2 \frac{\partial^2 u}{\partial x^2} = f(x,t), \qquad t>0, \quad a<x<b,$$

$$u(x,0) = \varphi(x),$$

where at $x=a$ and $x=b$, $u(x,t)$ satisfies some homogeneous boundary conditions. In this case, we can construct the function $g(x, t; \xi, \tau)$ $(t \geqslant \tau)$ satisfying the following conditions: (i) $L[g]=0$ except for $x=\xi$. $t=\tau$; (ii)

$$g(x,t;\xi,\tau) = \frac{\exp(-(x-\xi)^2/4c^2(t-\tau))}{2c\sqrt{\pi(t-\tau)}} + (\text{regular function})$$

in a neighborhood of $x=\xi$, $t=\tau$; and (iii) $g(x, t; \xi, \tau)$ satisfies the given homogeneous boundary conditions at $x=a$ and $x=b$. Then

$$u(x,t) = \int_0^t \int_a^b g(x,t;\xi,\tau) f(\xi,\tau) d\xi\, d\tau + \int_a^b g(x,t;\xi,0)\varphi(\xi) d\xi$$

represents the solution of the problem stated in this section for regular functions f, φ. The function $g(x, t; \xi, \tau)$ is called Green's function relative to the boundary value problem. Detailed consideration of such elementary cases is found in [6].

G. Kernel Functions

The kernel function is closely related to Green's function of Δ (the Laplacian) relative to the first boundary value problem in a domain in $\mathbf{R}^2$.

First we explain the general definitions of the kernel function. Let E be a general set, and let $\mathfrak{F}$ be a †Hilbert space of complex-valued functions defined on E with a suitable inner product (f, g). Suppose that we are given a function $K(x, y)$ defined on $E \times E$ satisfying the following conditions: (i) For any fixed y, $K(x, y)$ regarded as a function of x belongs to $\mathfrak{F}$; and (ii) for any $f(x) \in \mathfrak{F}$, $(f(x), K(x, y))_x = f(y)$. Then $K(x, y)$ is called a **kernel function** or **reproducing kernel**. The kernel function, if it exists, is unique and is †positive definite Hermitian; that is,

$$\sum_{j,k=1}^{n} K(y_j, y_k)\xi_j\bar{\xi}_k \geqslant 0. \tag{1}$$

Conversely, any positive definite function is a reproducing kernel of some Hilbert space. A necessary and sufficient condition for the existence of the kernel function is that for any $y \in E$, the linear functional $f \to f(y)$ be bounded. In this case, the minimum of $\|f\|$ under the condition $f(y)=1$ $(f \in \mathfrak{F})$ is attained by the element $K(x, y)/K(y, y)$, and its value is $K(y, y)^{-1/2}$. When $\mathfrak{F}$ is a †separable space, then by an †orthonormal system $\{\varphi_\nu(x)\}$, we can represent $K(x, y)$ as

$$K(x,y) = \sum_{\nu=1}^{\infty} \varphi_\nu(x)\overline{\varphi_\nu(y)}. \tag{2}$$

As an example of kernel functions, the following case is of particular importance. Let E be an n-dimensional †complex manifold, φ and ψ be holomorphic †differential forms of degree n on E, and let the following inner product be given:

$$(\varphi, \psi) = \int_E \varphi \wedge \bar{\psi}.$$

Now $\mathfrak{F}$ is given by $\mathfrak{F} = \{\varphi \mid \varphi, \varphi) < +\infty\}$, and the kernel function is called the **kernel differential**. When E is a domain in $\mathbf{C}^n$, regarding the coefficients of the differential form as functions, we call the kernel function **Bergman's kernel function**. Moreover, if E can be mapped onto a bounded domain by a one-to-one holomorphic mapping, then

$$ds^2 = \sum (\partial^2 \log K(z, \bar{z})/\partial z_j \partial \bar{z}_k) dz_j d\bar{z}_k$$

is positive definite and gives a Kähler metric which is called the **Bergman metric**.

H. Kernel Functions for Domains in the Complex Plane

Let E be a domain D in the complex plane $(z=x+iy)$. Let $K(z,\zeta)$ be Bergman's kernel function of D, and let $G(z,\zeta)$ be Green's function of Δ relative to the first boundary condition with a pole at ζ. Then we have

$$K(z,\zeta)=-(2/\pi)\partial^2G(z,\zeta)/\partial z\partial\bar{\zeta}. \qquad (3)$$

Next, let $U(z,\zeta)$ be the kernel function of the Hilbert space consisting of the holomorphic differential forms whose integrals are single-valued, and let $N(z,\zeta)$ be **Neumann's function** of Δ, i.e., the function that is harmonic in $D-\{\zeta\}$, has the same singularity as G at ζ, and whose derivative in the normal direction $\partial N/\partial n$ is constant along the boundary. Then we have

$$U(z,\zeta)=(2/\pi)\partial^2N(z,\zeta)/\partial z\partial\bar{\zeta}. \qquad (4)$$

Now the kernel $H(z,\zeta)=(N(z,\zeta)-G(z,\zeta))/2\pi$ is the kernel function relative to the Hilbert space consisting of all real †harmonic functions whose integral mean value along the boundary Γ is 0 and having the inner product

$$(\varphi,\psi)=\iint_D\left(\frac{\partial\varphi}{\partial x}\frac{\partial\psi}{\partial x}+\frac{\partial\varphi}{\partial y}\frac{\partial\psi}{\partial y}\right)dx\,dy.$$

The kernel $H(z,\zeta)$ is called a **harmonic kernel function**.

Suppose that the boundary Γ is †piecewise smooth, and consider the space of all holomorphic functions in D that are continuous on the boundary of D. The inner product of such functions φ and ψ is given by $(\varphi,\psi)=\int_\Gamma\varphi\bar{\psi}\,ds$ (ds is the element of the arc length of Γ). Hence we have a Hilbert space. Then the kernel function relative to this Hilbert space is called **Szegö's kernel function**, which has a close relation with †bounded functions.

The kernel functions enable us to represent holomorphic mappings that map the domain D onto various canonical domains (→ 77 Conformal Mappings).

References

[1] R. Courant and D. Hilbert, Methods of mathematical physics, Interscience, I, 1953; II, 1962.
[2] K. Yosida, Lectures on differential and integral equations, Interscience, 1960. (Original in Japanese, 1950.)
[3] V. D. Kupradze, Randwertaufgaben der Schwingungstheorie und Integralgleichungen, Deutscher Verlag der Wiss., 1956. (Original in Russian, 1950.)
[4] S. Mizohata, Theory of partial differential equations, Cambridge Univ. Press, 1973. (Original in Japanese, 1965.)
[5] F. K. G. Odqvist, Über die Randwertaufgaben der Hydrodynamik zäher Flüssigkeiten, Math. Z., 32 (1930), 329–375.
[6] E. E. Levi, Sull' equazione del calore, Ann. Mat. Pura Appl., (3) 14 (1908), 187–264.
For kernel functions,
[7] S. Bergman, The kernel function and conformal mapping, Amer. Math. Soc. Math. Surveys, 1950.
[8] S. Bergman and M. Schiffer, Kernel functions and elliptic differential equations in mathematical physics, Academic Press, 1953.
[9] N. Aronszajn, Theory of reproducing kernels, Trans. Amer. Math. Soc., 68 (1950), 337–404.
[10] H. Meschkowski, Hilbertsche Räume mit Kernfunktion, Springer, 1962.

189 (XIII.29) Green's Operator

A. General Remarks

Consider the first and the third boundary value problems for the elliptic equation

$$A[u]\equiv-\Delta u+\sum_{i=1}^n a_i(x)\frac{\partial u}{\partial x_i}+c(x)u=f(x)$$

(→ 323 Partial Differential Equations of Elliptic Type). Let D be a bounded domain of $\mathbf{R}^n$ whose boundary S consists of a finite number of smooth hypersurfaces. By the †method of orthogonal projection, we take the domain $\mathscr{D}(A)$ as follows: (i) $\{u(x)\mid u(x)\in H^2(D)$ and $u(x)=0$ for $x\in S\}$ or (ii) $\{u(x)\mid u(x)\in H^2(D)$ and $\partial u/\partial n+\beta(x)u=0$ for $x\in S\}$ according as we are considering the first or third boundary value problem, where $H^2(D)$ is the †Sobolev space (→ 168 Function Spaces). If the operator A is a one-to-one mapping from $\mathscr{D}(A)$ onto the function space $L_2(D)$, we call the inverse operator A^{-1} **Green's operator** relative to the boundary condition, and we denote it by G. In general, the existence of A^{-1} is not guaranteed. However, if we take real t large enough, $G_t=(A+tI)^{-1}$ exists.

Consider the general case where λ is a complex parameter: $(\lambda I-A)[u]=f(x)$, $f(x)\in L_2(D)$. Letting G_t act from the left, we have $(I-(\lambda+t)G_t)[u]=-G_tf$. Conversely, if $u\in L_2(D)$ is a solution, clearly $u(x)\in\mathscr{D}(A)$, and $u(x)$ satisfies the first partial differential equation and the boundary condition. Since G_t is a †com-

pact operator in $L_2(D)$, the †Riesz-Schauder theorem can be applied (→ 68 Compact and Nuclear Operators). In particular, if $\lambda+t$ is not an †eigenvalue of G_t, $u(x)=(\lambda I-A)^{-1}f=-(I-(\lambda+t)G_t)^{-1}G_t f$ represents a unique solution.

In the equations in the first paragraph of this section, if $a_i(x)=0$ and $c(x)$ and $\beta(x)$ are real, then G_t is a †self-adjoint operator in $L_2(D)$, and therefore the †Hilbert-Schmidt expansion theorem can be applied. Namely, let $\{\lambda_i\}$ be the eigenvalues of A such that $A\omega_i(x)=\lambda_i\omega_i(x)$, where $\{\omega_i(x)\}$ is an †orthonormal system in $L_2(D)$. Then for any $f(x)\in L_2(D)$, $f(x)=\sum_{i=1}^{\infty}(f,\omega_i)\omega_i(x)$, where the right-hand side is taken in the sense of †mean convergence. Furthermore, for $f(x)\in\mathcal{D}(A)$, we have the expansion $(Af)(x)=\sum_{i=1}^{\infty}\lambda_i(f,\omega_i)\omega_i(x)$ in the same sense.

When G_t is not self-adjoint, let G_t^* be the †adjoint operator of G_t in $L_2(D)$. Then G_t^* represents Green's operator relative to the equation

$$(A^*+tI)[v]=-\Delta v-\sum_{i=1}^{n}\frac{\partial}{\partial x_i}(\overline{a_i(x)}v)$$
$$+(\overline{c(x)}+t)v$$
$$=g(x),$$

corresponding to the boundary conditions (i) $v(x)=0$, $x\in S$ (first boundary condition) and (ii) $(\partial/\partial n)v+\beta'(x)v=0$, $x\in S$, where

$$\beta'(x)=\overline{\beta(x)}+\sum_{i=1}^{n}\overline{a_i(x)}\cos nx_i,$$

with n the outer normal (third boundary condition) [2].

B. Elliptic Equations of Higher Order

Green's operator can be defined for elliptic equations of higher order. Consider the equation

$$A(x,\partial/\partial x)u(x)=f(x),\quad x\in D;$$
$$B_j(x,\partial/\partial x)u(x)=0,\quad x\in S,$$
$$j=1,2,\ldots,b(=m/2),\quad (1)$$

where A is an †elliptic operator of order m and the boundary operators $\{B_j\}$ satisfy: (i) At every point x of S, the normal direction is not †characteristic for any B_j; and (ii) the order m_j of B_j is less than m, and $m_j\neq m_k$ $(j\neq k)$. The domain $\mathcal{D}(A)$ of A is defined by

$$\mathcal{D}(A)=\{u(x)\mid u\in H^m(D)\text{ and }$$
$$B_j(x,\partial/\partial x)u(x)=0\text{ for }x\in S,$$
$$j=1,2,\ldots,b\}.$$

When A is a one-to-one mapping from $\mathcal{D}(A)$ onto $L_2(D)$, the inverse $G=A^{-1}$ is called **Green's operator**.

This general boundary value problem, especially the existence theorem, was treated under some algebraic conditions on A and $\{B_j\}$ by M. Schechter [5] who showed that $G[u]\in H^m(D)$ if $u(x)\in L_2(D)$ and that $G[u]$ depends continuously on u. In particular, if $m>n/2$, then by †Sobolev's theorem, G is a continuous mapping from $L_2(D)$ into $C^0(\bar{D})$, and G is represented by an †integral operator of Hilbert-Schmidt type (L. Gårding [6]). Namely, for any $f(x)\in L_2(D)$,

$$(Gf)(x)=\int_D G(x,\xi)f(\xi)d\xi,$$

$$\iint|G(x,\xi)|^2\,dx\,d\xi<+\infty.$$

In general, the function $G(x,\xi)$, obtained by the **kernel representation** of Green's operator G, is called **Green's function**.

On the other hand, consider, for example, the †Dirichlet problem of Δ in $\mathbf{R}^3$. Then Green's function is defined in the following way (→ 188 Green's Functions): $G(x,\xi)=-(4\pi|x-\xi|)^{-1}+u(x,\xi)$, where $u(x,\xi)$ satisfies (i) $\Delta_x u(x,\xi)=0$, and (ii) $G(x,\xi)|_{x\in S}=0$. The function defined in this manner coincides with Green's function defined as a kernel representation of Green's operator [1,2].

Suppose that in problem (1) the †elliptic operator $A(x,\partial/\partial x)$ is independent of x, and let $E(x)$ be a †fundamental solution, i.e., $E(x)$ is a distribution solution of $A(\partial/\partial x)E(x)=\delta(x)$ ($\delta(x)$ is †Dirac's δ-function; → 112 Differential Operators). Then $E(x)$ is a C^∞-function away from the origin. Moreover, in a neighborhood of the origin the following estimates hold: For $|\alpha|<m$,

$$\left|\left(\frac{\partial}{\partial x}\right)^{\alpha}E(x)\right|\leqslant\begin{cases}c|x|^{m-n-|\alpha|}, & m-n-|\alpha|<0,\\ c\log|x|^{-1}, & m-n-|\alpha|=0,\\ c, & m-n-|\alpha|>0,\end{cases}$$

where c is some positive constant. When problem (1) has Green's operator G, we can say the following, using the fundamental solution: Green's function $G(x,\xi)$ exists and can be written as $G(x,\xi)=E(x-\xi)+u(x,\xi)$, where for any fixed $\xi\in D$, $u(x,\xi)$ satisfies (i) $A(\partial/\partial x)u(x,\xi)=0$ and (ii) $B_j(x,\partial/\partial x)G(x,\xi)=0$, $x\in S$, $j=1,2,\ldots,b$.

C. Hypoelliptic Operators

Let

$$A(x,\partial/\partial x)=\sum_{|\alpha|\leqslant m}a_\alpha(x)\left(\frac{\partial}{\partial x}\right)^{\alpha},$$

$$\left(\frac{\partial}{\partial x}\right)^{\alpha}=\frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1}\ldots\partial x_n^{\alpha_n}},$$

be a general partial differential operator with C^∞-coefficients. If the kernel $E(x,\xi)$ satisfies $A(x,\partial/\partial x)E(x,\xi)=\delta(x-\xi)$, that is, if we have the relation $\langle E(x,\xi), {}^tA(x,\partial/\partial x)\varphi(x)\rangle_x=\varphi(\xi)$ for any $\varphi(x)\in\mathscr{D}$, then $E(x,\xi)$ is called a **fundamental solution** of A, where ${}^tA$ is the **transposed operator** of A:

$$ {}^tA(x,\partial/\partial x)v(x)=\sum_{|\alpha|\leqslant m}(-1)^{|\alpha|}\left(\frac{\partial}{\partial x}\right)^\alpha(a_\alpha(x)v(x)). $$

Now if there exists a fundamental solution $E'(x,\xi)$ of ${}^tA(x,\partial/\partial x)$ such that (i) $E'(x,\xi)$ defines a kernel that gives rise to two continuous mappings, one of which maps the space $\mathscr{D}_\xi$ into $\mathscr{E}_x$, and the other of which maps the space $\mathscr{D}_x$ into $\mathscr{E}_\xi$ (→ 168 Function Spaces), and (ii) for $x\neq\xi$, $E'(x,\xi)$ a C^∞-function of (x,ξ), then any distribution $u(x)$ satisfying $A(x,\partial/\partial x)u(x)=g(x)$ is a C^∞-function, where $g(x)$ is of class C^∞. In general, an operator A with the property that any solution $u(x)$ of $A(x,\partial/\partial x)u(x)=g(x)$ is of class C^∞ whenever $g(x)$ is of class C^∞, is called **hypoelliptic**. Elliptic and parabolic operators are both hypoelliptic. L. Hörmander characterized the hypoelliptic differential operators with constant coefficients [8] (→ 112 Differential Operators).

A kernel $E(x,\xi)$ such that

$$ A(x,\partial/\partial x)E(x,\xi)=\delta(x-\xi)+\varpi(x,\xi), $$

with $\varpi(x,\xi)$ a C^∞-function, is called a **parametrix** of A. To prove the hypoellipticity of A, it suffices to show the existence of a parametrix $E'(x,\xi)$ of the operator ${}^tA$ having the properties (i) and (ii) mentioned in the previous paragraph.

To explain the notion of the fundamental solution for the †evolution equations, suppose that we are given an evolution equation

$$ L[u]=\frac{\partial^m}{\partial t^m}u(x,t)+\sum_{j<m}a_{\alpha,j}(x,t)\left(\frac{\partial}{\partial x}\right)^\alpha\frac{\partial^j}{\partial t^j}u(x,t)=0, $$

$x\in\mathbf{R}^n$, $t_0\leqslant t\leqslant T$. A kernel $E(x,t;\xi,t_0)$ $(t_0\leqslant t\leqslant T)$ is called a **fundamental solution** to the evolution equation if

$$ L_{x,t}(E(x,t;\xi,t_0))=0,\quad t>t_0, $$

and

$$ \lim_{t\to t_0+0}\frac{\partial^i}{\partial t^i}E(x,t;\xi,t_0)=\begin{cases}0, & 0\leqslant j\leqslant m-2,\\ \delta(x-\xi), & j=m-1.\end{cases} $$

References

[1] H. G. Garnir, Les problèmes aux limites de la physique mathématique, Birkhäuser, 1958.
[2] S. Mizohata, Theory of partial differential equations, Cambridge Univ. Press, 1973. (Original in Japanese, 1965.)
[3] R. Courant and D. Hilbert, Methods of mathematical physics II, Interscience, 1962.
[4] L. Gårding, Dirichlet's problem for linear elliptic differential equations, Math. Scand., 1 (1953), 55–72.
[5] M. Schechter, General boundary value problems for elliptic partial differential equations, Comm. Pure Appl. Math., 12 (1959), 457–486.
[6] L. Gårding, Applications of the theory of direct integrals of Hilbert spaces to some integral and differential operators, Univ. of Maryland lecture series no. 11, 1954.
[7] L. Schwartz, Théorie des distributions, Hermann, second edition, 1966.
[8] L. Hörmander, On the theory of general partial differential operators, Acta Math., 94 (1955), 161–248.

190 (IV.1) Groups

A. Definition

Let G be a nonempty set. Suppose that for any elements a, b of G there exists a uniquely determined element c of G, which is called the **product** of a and b, written $c=ab$. We call G a **group** or **multiplicative group** if (i) the **associative law** $a(bc)=(ab)c$ holds, and (ii) for any elements $a, b\in G$ there exist uniquely determined elements $x, y\in G$ satisfying $ax=b$, $ya=b$. Then the mapping $(a,b)\to ab$ is called **multiplication** in G. Condition (ii) is equivalent to the following two conditions: (iii) There exists an element e (called the **identity element** or **unit element** of G) such that $ae=ea=a$ for any element a of G; and (iv) for any element a of G there exists an element x such that $ax=xa=e$.

The element x in condition (iv) is called the **inverse** (or **inverse element**) of a, denoted by a^{-1}. The uniqueness of the identity element e and the inverse a^{-1} follows readily from the axioms. The identity element of a multiplicative group is sometimes denoted by 1. If $ab=ba$, then we say that a and b **commute**. The **commutative law**, $ab=ba$ for any elements $a, b\in G$, is not assumed in general. A group satisfying the commutative law is called an **Abelian group** (or **commutative group**) in honor of N. H. Abel, who made use of commutative groups in his study of the theory of equations. The product in a commutative group is often written in the form $a+b$, and in this case the mapping $(a,b)\to a+b$ is called **addition**. The element $a+b$ is called the **sum** of a and b, and G is called an **additive group**. In an additive

group the identity element is usually denoted by 0 and called the **zero element**, and the inverse of a is denoted by $-a$ (→ 2 Abelian Groups; 277 Modules). To describe the †law of composition, we sometimes use notation different from multiplication or addition (→ 409 Structures).

B. Examples

A †linear space over a †field K is an additive group with respect to the usual addition of vectors (→ 256 Linear Spaces). A field is an additive group with respect to the addition, and the set of nonzero elements of a field forms a group with respect to the multiplication, which is called the **multiplicative group of the field** (→ 149 Fields).

All †invertible $n \times n$ matrices over a ring R form a group with respect to the usual multiplication of matrices. This group is called the †general linear group of degree n over R (→ 60 Classical Groups).

All one-to-one mappings from a set M onto itself (i.e., all **permutations** on M) form a group with respect to the composition defined by $(f \circ g)(x) = f(g(x))$ $(x \in M)$. (Sometimes the product $f \circ g$ is denoted by gf and $f(x)$ by xf. Then $x(gf) = (xg)f$.) The group of all permutations on M is called the **symmetric group** on M. A group G is called a **permutation group** (on M) if every element of G is a permutation on M. For instance, the general linear group of degree n over a field K may be regarded as a permutation group on the set of n-dimensional vectors, and it is also regarded as a permutation group on a †tensor space.

All †motions in a Euclidean space form a group with respect to the usual composition of motions. All invertible $n \times n$ matrices over K leaving a given †quadratic form invariant form a group with respect to the usual multiplication of matrices. This group is called the †orthogonal group belonging to the given quadratic form. If K is the †complex number field or the real number field, then these groups are Lie groups (→ 13 Algebraic Groups; 151 Finite Groups; 161 Free Groups; 249 Lie Groups; 423 Topological Groups).

C. Fundamental Concepts

If a group G consists of a finite number of elements, then G is called a **finite group**; otherwise, G is called an **infinite group**. The number of elements of G is called the **order** of G. A nonempty subset H of G is called a **subgroup** of G if H is a group with respect to the multiplication of the group G. Hence a nonempty subset H is a subgroup of G if and only if $a^{-1}b \in H$ for any $a, b \in H$. For a family $\{H_\lambda\}$ of subgroups of G, the intersection $\bigcap_\lambda H_\lambda$ is also a subgroup.

The associative law of multiplication says that elements a_1, a_2, a_3 of G determine the product $a_1 a_2 a_3$, which is the common value of $(a_1 a_2)a_3$ and $a_1(a_2 a_3)$. This law can be generalized to say that any ordered set of n elements $a_1, a_2, \ldots, a_n$ $(n > 2)$ of G determines their product $a_1 a_2 \ldots a_n$ (**general associative law**). When $a_1 = a_2 = \cdots = a_n = a$, we denote the product $aa \ldots a$ by a^n. If we define a^{-n} for $n \geqslant 0$ by $a^0 = e$ and $a^{-n} = (a^n)^{-1}$, we then have $a^n a^m = a^{n+m}$, $(a^n)^m = a^{nm}$ for any $n, m \in \mathbf{Z}$. If there exists a positive integer n such that $a^n = e$, then the smallest positive integer d with $a^d = e$ is called the **order** of the element a. If there is no such n, then a is called an element of **infinite order**. If a is of infinite order, then its powers a^0 $(=e)$, $a^{\pm 1}, a^{\pm 2}, \ldots$, are all unequal. If a is of order d, then the different powers of a are a^0 $(=e)$, $a, a^2, \ldots, a^{d-1}$. All the powers of a form a subgroup $\langle a \rangle$ of G, called a **cyclic subgroup**. The order of an element a is the same as the order of the subgroup $\langle a \rangle$. The group $\langle a \rangle$ itself is called a **cyclic group** and is an example of an Abelian group (→ 2 Abelian Groups).

Let S be a subset of a group G. Then the intersection of all subgroups of G containing S is called the subgroup **generated** by S and is denoted by $\langle S \rangle$. It is the smallest subgroup containing S, and if S is nonempty, $\langle S \rangle$ consists of all the elements of the form $a_1^{m_1} a_2^{m_2} \ldots a_r^{m_r}$ $(a_i \in S, m_i \in \mathbf{Z})$. If $\langle S \rangle = G$, the elements of S are called **generators** of G. When G has a finite set of generators, G is said to be **finitely generated**. When $S = \{a\}$, then $\langle S \rangle$ coincides with $\langle a \rangle$, and the element a is the generator of the cyclic group $\langle a \rangle$. Suppose that elements $a_1, \ldots, a_n$ of G satisfy an equation of the form $a_1^{m_1} a_2^{m_2} \ldots a_n^{m_n} = 1$. This equation is then called a **relation** among the elements $a_1, \ldots, a_n$. If we have a system of generators and all relations among the generators, then they define a group (→ 161 Free Groups). It is, however, still an open problem to find a general procedure to decide whether the group determined by a given system of generators and the relations among them contains elements other than the identity (→ 161 Free Groups B).

For a subset S and an element x of a group G, the set of all elements $x^{-1}sx$ $(s \in S)$ is denoted by $x^{-1}Sx$ or S^x, and S and S^x are called **conjugate**. We have $(ab)^x = a^x b^x$, $(a^{-1})^x = (a^x)^{-1}$. If H is a subgroup, then H^x is also a subgroup. For a subset S, the set of all elements x satisfying $S^x = S$ forms a subgroup $N(S)$, called the **normalizer** of S. The set of all elements that commute with every element of

S forms a subgroup $Z(S)$, called the **centralizer** of S. The centralizer Z of G is called the **center** of G. The set of all elements conjugate to a given element a of G is called a **conjugacy class**. A group G is the disjoint union of its conjugacy classes.

Let H be a subgroup of a group G and x an element of G. The set of elements of the form hx $(h \in H)$ is denoted by Hx and is called a **right coset** of H. A **left coset** xH is defined similarly. G is the disjoint union of left (right) cosets of H. The cardinality of the set of left cosets of H equals that of the set of right cosets of H; it is called the **index** of the subgroup H and is denoted by $(G:H)$. Given two subgroups H and K of G, the set $HxK = \{hxk \mid h \in H, k \in K\}$ is called the **double coset** of H and K, and G is the disjoint union of different double cosets of H and K. If the left cosets of a subgroup H are also the right cosets, i.e., if $Hx = xH$ for every $x \in G$, then H is called a **normal subgroup** (or **invariant subgroup**) of G. An equivalent condition is that $H = H^x$ for all $x \in G$. The center of G is always a normal subgroup of G. If H is a normal subgroup of G, the set of all products of an element of Ha and an element of Hb coincides with Hab. Thus if we define the product of two cosets Ha and Hb to be Hab, then the set of cosets of H forms a group. This group is denoted by G/H and is called the **factor group** (or **quotient group**) of G modulo H. (When G is an additive group, G/H is also denoted by $G-H$ and is called the **difference group**.) The group G itself and $\{e\}$ are normal subgroups of G. If G has no normal subgroup other than these two, then G is called a **simple group**. A subgroup of finite index contains a normal subgroup of finite index. If H is a subgroup of finite index, then we can find a common complete system of representatives of the left cosets and the right cosets of H. If G is finitely generated, then so is any subgroup of G of finite index.

Let R be an †equivalence relation defined in a group G. If xRx' and yRy' always imply $(xy)R\,(x'y')$, then we say that R is **compatible** with the multiplication. The †quotient set G/R is a group with respect to the induced multiplication. This group is called the **quotient group** of G with respect to R. The equivalence class H containing e is a normal subgroup, and xRx' if and only if $x^{-1}x' \in H$, i.e., x and x' are contained in the same coset of H. Thus G/R coincides with G/H.

If G is a finite group of order n, then the order and the index of any subgroup of G, the order of any element of G, the cardinal number of any conjugacy class of G, and the number of different conjugate subgroups of any subgroup of G are all divisors of n.

D. Isomorphisms and Homomorphisms

If there is a one-to-one mapping $a \leftrightarrow a'$ of the elements of a group G onto those of a group G' and if $a \leftrightarrow a'$ and $b \leftrightarrow b'$ imply $ab \leftrightarrow ab'$, then we say that G and G' are **isomorphic** and write $G \cong G'$. If we put $a' = f(a)$, then $f: G \to G'$ is a †bijection satisfying $f(ab) = f(a)f(b)$ $(a, b \in G)$. More generally, if a mapping $f: G \to G'$ satisfies $f(ab) = f(a)f(b)$ for all $a, b \in G$, then f is called a **homomorphism** of G to G'. An †injective (†surjective) homomorphism is also called a †monomorphism (†epimorphism). If there is a surjective homomorphism $G \to G'$, then we say that G' is **homomorphic** to G. The composite of two homomorphisms is also a homomorphism. If a homomorphism $f: G \to G'$ is a bijection, then f is called an **isomorphism**. In this case f^{-1} is also an isomorphism, and we have $G \cong G'$.

For a subgroup H of a group G, the injective homomorphism $f: H \to G$ defined by $f(a) = a$ $(a \in H)$ is called the **canonical injection** (or **natural injection**). For a factor group G/R of G, the surjective homomorphism $f: G \to G/R$ such that $a \in f(a)$ $(a \in G)$ is called the **canonical surjection** (or **natural surjection**).

Let $f: G \to G'$ be a homomorphism. Then the **image** $f(G)$ of f is a subgroup of G, and the **kernel** $H = \{a \in G \mid f(a) = e'$ (the identity of G')$\}$ of f is a normal subgroup of G. The equivalence classes of the equivalence relation given by $f(x) = f(y)$ are just the cosets of H, and f induces an isomorphism $\bar{f}: G/H \to f(G)$. The latter proposition is called the **homomorphism theorem** of groups. This theorem is extended in the following way: For simplicity let $f: G \to G'$ be a surjective homomorphism. (i) If H' is a normal subgroup of G', then the inverse image $H = f^{-1}(H')$ is a normal subgroup of G, and f induces the isomorphism $\bar{f}: G/H \to G'/H'$. (ii) if H is a subgroup and N is a normal subgroup of G, then $HN = \{hn \mid h \in H, n \in N\}$ is a subgroup of G, and the canonical injection $H \to HN$ induces an isomorphism $H/H \cap N \to HN/N$. (iii) If H and N are two normal subgroups of G such that $H \supset N$, then the canonical surjection $G \to G/N$ induces an isomorphism $G/H \to (G/N)/(H/N)$. Propositions (i), (ii), and (iii) are called the **isomorphism theorems** of groups.

A homomorphism of G to itself is called an **endomorphism** of G, and an isomorphism of G to itself is called an **automorphism** of G. The set of automorphisms of G forms a group with respect to the composition of mappings, called the **group of automorphisms** of G. Given an element a of G, the mapping $x \to a^{-1}xa$ $(x \in G)$ yields an automorphism of G which is called an **inner automorphism** of G. The set of inner automorphisms of G forms a normal subgroup of the group of automorphisms of G, called the

group of inner automorphisms of G, which is isomorphic to the factor group of G modulo its center. The factor group of the group of automorphisms of G modulo the group of inner automorphisms of G is called the **group of outer automorphisms** of G.

If a mapping $f: G \rightarrow G'$ from a group G to another group G' satisfies $f(ab)=f(b)f(a)$ $(a, b \in G)$, then f is called an **antihomomorphism**. A bijective antihomomorphism is called an **anti-isomorphism**. When $G=G'$, f is called an **anti-endomorphism** or **anti-automorphism** (e.g., $f: G \rightarrow G$ defined by $f(a)=a^{-1}$ is an anti-isomorphism).

E. Groups with Operator Domain

Let Ω be a set and G a group. Suppose that for each $\theta \in \Omega$ and $x \in G$, the product $\theta x \in G$ is defined and satisfies $\theta(xy)=\theta(x)(\theta y)$. Then Ω is called an **operator domain** of G, and G is called a group with operator domain Ω, or simply an **Ω-group**. (We sometimes write x^θ instead of θx.) The mapping $(\theta, x) \rightarrow \theta x$ from $\Omega \times G$ to G is called the †operation of Ω on G. If G is an Ω-group, then any element θ of Ω induces an endomorphism $\theta_G: x \rightarrow \theta x$ of G. Conversely, if we are given a mapping $\theta \rightarrow \theta_G$ of Ω to the set of endomorphisms of G, then we may regard G as an Ω-group. Any group may be regarded as an Ω-group with Ω equal to the empty set or to the set consisting of the identity automorphism of G. Thus the general theory of groups can be extended to the theory of groups with operator domain, and in some cases effective use of suitable operator domains can be fruitful in the investigation of the properties of groups themselves. (→ 2 Abelian Groups; 277 Modules).

A subgroup H of an Ω-group G is called an **Ω-subgroup** (or **admissible subgroup**) if $\theta x \in H$ for any $\theta \in \Omega$ and $x \in H$. In this case, H is also an Ω-group. If an equivalence relation R defined in G is compatible with the multiplication and also compatible with the operators, namely, if xRx' implies $(\theta x)R(\theta x')$ for any $\theta \in \Omega$, then the quotient group G/R is also an Ω-group. The equivalence class containing e is an **admissible normal subgroup**. Conversely, if H is an admissible normal subgroup, then the equivalence relation defined by H is compatible with the operators, and the factor group G/H is an Ω-group. A homomorphism $f: G \rightarrow G'$ of an Ω-group G to an Ω-group G' is called an **Ω-homomorphism (admissible homomorphism** or **operator homomorphism)** if $f(\theta x) = \theta f(x)$ for any $\theta \in \Omega$ and $x \in G$. If f is an isomorphism, then f is called an **Ω-isomorphism (admissible isomorphism** or **operator isomorphism)**. We have the homomorphism theorem and the isomorphism theorems of Ω-groups if we consider only admissible subgroups and admissible homomorphisms.

F. Sequences of Subgroups

Let $H_1, H_2 \ldots$ be an infinite sequence of (normal) subgroups of a group G. If $H_i \subsetneqq H_{i+1}$ $(i=1,2,\ldots)$, then the sequence is called an **ascending chain** of (normal) subgroups. If $H_i \supsetneqq H_{i+1}$ $(i=1,2,\ldots)$, then it is called a **descending chain** of (normal) subgroups. If there is no ascending (or descending) chain of (normal) subgroups of G, we say that G satisfies the **ascending** (or **descending) chain condition** for (normal) subgroups. These conditions are the same as the ascending (or descending) chain condition in the ordered set of all (normal) subgroups of G (→ 311 Ordering C). A group G satisfies the ascending chain condition for subgroups if and only if every subgroup of G is finitely generated. Also, for groups with operator domain we have similar results. The structure of Abelian groups satisfying the ascending (descending) chain condition is completely determined (→ 2 Abelian Groups). It is not known whether there is an infinite group satisfying both the ascending and descending chain conditions for subgroups. A group satisfying the descending chain condition for subgroups has no element of infinite order, but the converse is not true. There is an infinite group which is finitely generated and has no element of infinite order (→ 161 Free Groups C).

G. Normal Chains

A finite sequence $G=G_0 \supset G_1 \supset G_2 \supset \ldots \supset G_r = \{e\}$ of subgroups of a group G is called a **normal chain** if G_i is a normal subgroup of G_{i-1} for $i=1,2,\ldots,r$. We call r the **length** of the chain. The sequence $G_0/G_1, G_1/G_2, \ldots, G_{r-1}/G_r$ is called the **sequence of factor groups** of the normal chain. A normal chain $G=H_0 \supset H_1 \supset H_2 \supset \ldots \supset H_s=\{e\}$ is called a **refinement** of the chain $G=G_0 \supset G_1 \supset \ldots \supset G_r=\{e\}$ if every G_i appears in this chain. Two normal chains with the same length are called isomorphic if there is a one-to-one correspondence between their sequences of factor groups such that corresponding factor groups are isomorphic. Any two normal chains have refinements which are isomorphic to each other (Schreier's refinement theorem). A normal chain is called a **composition series** (or **Jordan-Hölder sequence**) if it consists of different subgroups of G and in

any proper refinement there appear two successive subgroups which are the same. The sequence of factor groups of a composition series is called a **composition factor series**, and the factor groups appearing in this series are called **composition factors**. Any composition factor is a simple group. As a direct consequence of the refinement theorem we see that if a group G has a composition series, then the composition factor series is unique up to isomorphism and the ordering of the factors. (This theorem is due to O. Hölder. C. Jordan proved that if G is a finite group, then the set of orders of composition factors is independent of the choice of composition series. Hence we call the theorem the **Jordan-Hölder theorem**.)

For an Ω-group G, if we consider only Ω-subgroups, we have definitions and theorems similar to those in this section. When we take the group of inner automorphisms of G as Ω, then a composition series of the Ω-group G is called a **principal series**. If we take the group of automorphisms of G as Ω, then a composition series is called a **characteristic series**. An infinite group G does not always have a composition series. Even if G has a composition series, G may have an infinite normal chain $G_1 \subset G_2 \subset \ldots \subset G$ such that each G_i is a normal subgroup of G_{i+1} and $\bigcup G_i = G$. In fact, there is a simple group which has such an infinite normal chain (P. Hall). Two groups which have isomorphic composition series are not necessarily isomorphic. A subgroup of a group G is called a **subnormal subgroup** of G if it may appear in some normal chain. The intersection of two subnormal subgroups is also subnormal, but their **join** (i.e., the subgroup generated by both of them) is not necessarily subnormal in an infinite group. The set of subgroups and the set of normal subgroups of a group form †lattices with respect to the inclusion relation (for the relationship between these lattices and the group structure → [8]).

H. Commutator Subgroups

Given two elements a and b of a group G, we call $a^{-1}b^{-1}ab = [a, b]$ the **commutator** of a and b. The subgroup C generated by all commutators in G is called the **commutator subgroup** (or **derived group**) of G. The subgroup C is a normal subgroup of G, and the factor group G/C is Abelian. On the other hand, if B is a normal subgroup of G and G/B is Abelian, then B contains C. For two subsets A, B of G, the subgroup generated by the commutators $[a, b]$ $(a \in A, b \in B)$ is called the **commutator group of** A **and** B and is denoted by $[A, B]$. If A and B are normal subgroups of G, then $C' = [A, B]$ is also a normal subgroup of G, and A/C' commutes with B/C' elementwise in the factor group G/C'. Furthermore, $[A, B]$ is the minimal normal subgroup with the property. The subgroup $[G, G]$ is the commutator subgroup of G.

If the commutator subgroup of G is Abelian, then G is called a **meta-Abelian group**. If a group G has a normal chain $G (= G_0) \supset G_1 \supset G_2 (= \{e\})$ of length 2 and the factor groups G/G_1, G_1/G_2 are Abelian, then G is meta-Abelian. Meta-Abelian groups are special cases of solvable groups, discussed in Section I.

I. Solvable Groups

Suppose that we are given a series of subgroups G_i $(i = 0, 1, 2, \ldots)$ of G such that $G = G_0$ and $[G_i, G_i] = G_{i+1}$. Then we have a normal chain $G = G_0 \supset G_1 \supset G_2 \supset \ldots$. If $G_r = \{e\}$ for some r, then G is called a **solvable group**. For the normal chain $G\ (= G_0) \supset G_1 \supset \ldots G_r\ (= \{e\})$ the factor groups G_i/G_{i+1} $(i = 0, 1, \ldots, r-1)$ are all Abelian. A finite group G is solvable if and only if G has a composition series $G = H_0 \supset H_1 \supset H_2 \supset \ldots \supset H_s = \{e\}$ such that the factor groups H_i/H_{i+1} $(i = 0, 1, \ldots, s-1)$ are all of prime order. An †irreducible algebraic equation over a field of †characteristic 0 is solvable by radicals if and only if its Galois group is solvable (→ 172 Galois Theory).

J. Nilpotent Groups

The sequence of subgroups $G = G_0 \supset G_1 \supset G_2 \supset \ldots$ defined inductively by setting $G_r = [G, G_{r-1}]$ $(r = 1, 2, \ldots)$ is called the **lower central series** of G. If $G_n = \{e\}$ for some n, then G is called a **nilpotent group**, and the least number n with $G_n = \{e\}$ is called the **class** of the nilpotent group G. A nilpotent group is solvable. Let Z_1 be the center of G, Z_2/Z_1 be the center of G/Z_1, and so on. Then we have a sequence of subgroups $Z_0 = \{e\} \subset Z_1 \subset Z_2 \subset \ldots$, called the **upper central series** of G. A group G is nilpotent if and only if $Z_m = G$ for some m, and the least number m with $Z_m = G$ is the class of G. For the subgroups G_r and Z_r $(r = 1, 2, \ldots)$, we have $[G_{i-1}, Z_i] = \{e\}$. If G is a †Lie group, then G is nilpotent if and only if the corresponding †Lie algebra $\mathfrak{g}$ is nilpotent, i.e., $\mathfrak{g}^n = 0$.

K. Infinite Solvable Groups

The concepts of solvability and nilpotency are generalized in several ways for infinite groups. For instance, a group G is called a **generalized solvable group** if any homomorphic image of G

which is unequal to $\{e\}$ contains an Abelian normal subgroup unequal to $\{e\}$, and G is called a **generalized nilpotent group** if any homomorphic image ($\neq\{e\}$) of G has center unequal to $\{e\}$. These definitions coincide with the previous ones for finite groups but not for infinite groups [7].

L. Direct Products

Let $G_1, \ldots, G_n$ be a finite number of groups. The set G of all elements $(x_1, \ldots, x_n)$ with $x_i \in G_i$ $(i=1, \ldots, n)$ is a group if we define the product of two elements $x=(x_1, \ldots, x_n)$ and $y=(y_1, \ldots, y_n)$ to be $xy=(x_1 y_1, \ldots, x_n y_n)$. We call G the **direct product** of groups $G_1, \ldots, G_n$ and write $G=G_1 \times \ldots \times G_n$. If e_j is the identity element of G_j, then $e=(e_1, \ldots, e_n)$ is the identity element of G. The mapping $(x_1, \ldots, x_n) \rightarrow x_i$ from G to G_i is a surjective homomorphism, called the **canonical surjection**. The subgroup $H_i=\{(e_1, \ldots, e_{i-1}, x_i, e_{i+1}, \ldots, e_n) \mid x_i \in G_i\}$ is isomorphic to G_i. The subgroups H_i $(i=1, \ldots, n)$ satisfy the following conditions: (i) H_i is a normal subgroup of G. (ii) H_i commutes with H_j elementwise if $i \neq j$. (iii) Any element of G can be written uniquely as the product of elements of $H_1, \ldots, H_n$. Conversely, if a group G has subgroups $H_1, \ldots, H_n$ satisfying these three conditions, then G is isomorphic to $H_1 \times \ldots \times H_n$. In this case we also write $G=H_1 \times \ldots \times H_n$, and we call this a **direct decomposition** of G. Each H_i is called a **direct factor** of G. Conditions (i), (ii), and (iii) are equivalent to condition (i), (ii') $G=H_1 H_2 \ldots H_n$, and $H_1 \ldots H_{i-1} \cap H_i=\{e\}$ $(i=2, \ldots, n)$.

A group G is called **indecomposable** if G cannot be decomposed into the direct product of two subgroups unequal to $\{e\}$, and **completely reducible** if G is the direct product of simple groups. If G satisfies the ascending or descending chain condition for normal subgroups, then G can be decomposed into the direct product of indecomposable groups. Such a decomposition is not unique in general, but if G has two direct product decompositions $G=G_1 \times \ldots \times G_m=H_1 \times \ldots \times H_n$, where G_i and H_j are indecomposable and not equal to $\{e\}$, then $m=n$ and the factors G_i are isomorphic to the factors H_j for some j; moreover, if G_1 corresponds, say, with H_1, then we have $G=H_1 \times G_2 \times \ldots \times G_m$. This fact was first stated by J. H. M. Wedderburn, and a complete proof of the theorem was given by R. Remak and O. Schmidt. Later W. Krull extended it to more general groups (with operator domain), and we call it the **Krull-Remak-Schmidt theorem**. O. Ore formulated it as a theorem on †modular lattices.

For an infinite number of groups G_λ $(\lambda \in \Lambda)$ we define the direct product $\prod_{\lambda \in \Lambda} G_\lambda$ of these groups similarly. The set of all elements $(\ldots, x_\lambda, \ldots)$ $(x_\lambda \in G_\lambda)$ such that almost all x_λ (i.e., all except a finite number of λ) are identity elements is a subgroup of the direct product, called the **direct sum** (or **restricted direct product**) of $\{G_\lambda\}$.

M. Free Products

Given a family of groups $\{G_\lambda\}_{\lambda \in \Lambda}$, we define the most general group G generated by these groups, called the free product of $\{G_\lambda\}$, together with canonical injections $f_\lambda: G_\lambda \rightarrow G$.

Let S be the disjoint union of the sets $\{G_\lambda\}_{\lambda \in \Lambda}$, and regard G_λ as a subset of S. A **word** is either void or a finite sequence $a_1, a_2, \ldots, a_n$ of elements of S, and we denote the set of all words by W. The product of two words w and w' is defined by connecting w with w' so that the †associative law holds. We write $w \succ w'$ when two words w and w' satisfy one of the following two conditions: (i) The word w has successive members a, b which belong to the same group G_λ, and the word w' is obtained from w by replacing a, b by the product ab. (ii) Some member of w is an identity element, and w' is the word obtained from w by deleting this member. For two words w and w', we write $w \equiv w'$ if there is a finite sequence of words $w=w_0, w_1, \ldots, w_n=w'$ such that for each i $(1 \leqslant i \leqslant n)$, either $w_{i-1} \succ w_i$ or $w_i \succ w_{i-1}$. This relation is an equivalence relation and is compatible with the multiplication. Thus we may define a multiplication for the quotient set G of W by this equivalence relation, and then G is a group whose identity element is the equivalence class containing the void word. Any $x \in G_\lambda$ is regarded as a word, and we have an injective homomorphism $f_\lambda: G_\lambda \rightarrow G$ by assigning the corresponding class to each element of G_λ. The group G is called the **free product** of the system of groups $\{G_\lambda\}_{\lambda \in \Lambda}$, and f_λ is called the **canonical injection**. The free product G of $\{G_\lambda\}_{\lambda \in \Lambda}$ is characterized by the following universal property: Given a group G' and homomorphisms $f'_\lambda: G_\lambda \rightarrow G'$ $(\lambda \in \Lambda)$, we can find a unique homomorphism $g: G \rightarrow G'$ such that $g \circ f_\lambda = f'_\lambda$. The free product is the dual concept of direct product and is also called the †coproduct (→ 52 Categories and Functors). If each G_λ is an infinite cyclic group generated by a_λ, then the free product of the G_λ is the †free group generated by $\{a_\lambda\}$ (→ 161 Free Groups).

The concept of free product is generalized in the following way. Let H be a fixed group. We consider the family of pairs (G, j), where G is a group and $j: H \rightarrow G$ is an injective homomorphism. A homomorphism of pairs $f:(G,j) \rightarrow$

(G',j') is defined to be a group homomorphism $f:G \to G'$ such that $f \circ j = j'$. For a given family of pairs $\{(G_\lambda, j_\lambda)\}$, we have the **amalgamated product** (G,j) of the family and the canonical homomorphism $f_\lambda:(G_\lambda, j_\lambda) \to (G,j)$, which is characterized by the following universal property: Given a pair (G',j') and homomorphisms $f'_\lambda:(G_\lambda, j_\lambda) \to (G',j')$, we have a unique homomorphism $g:(G,j) \to (G',j')$ such that $g \circ f_\lambda = f'_\lambda$. If $H = \{e\}$, then the amalgamated product is the same as the free product. Now f_λ is an injection. If we regard G_λ as a subgroup of G, then G is generated by the subgroups G_λ and $G_\lambda \cap G_\mu = j_\lambda(H) = j_\mu(H)$ $(\lambda \neq \mu)$.

The notion of the amalgamated product is useful in constructing groups with interesting properties. For instance, we have a group whose nonidentity elements are all conjugate (B. H. Neumann and G. Higman), and a group generated by a finite number of elements such that its homomorphic image $(\neq \{e\})$ is always an infinite group (Higman) so that we have an infinite simple group generated by a finite number of elements.

N. Extensions

Let N and F be groups. A group G is called an **extension** of F by N if G has a normal subgroup $\bar{N}$ isomorphic to N and $G/\bar{N} \cong F$. The problem of finding all extensions was solved by Schreier (*Monatsh. Math. Phys.*, 34 (1926); *Abh. Math. Sem. Univ. Hamburg*, 4 (1928)). Suppose that (1) to each $\sigma \in F$ there corresponds an automorphism s_σ of N; (2) there exist elements $c_{\sigma,\tau}$ $(\sigma, \tau \in F)$ of N such that $s_\sigma(s_\tau(a)) = c_{\sigma,\tau}(s_{\sigma\tau}(a))c_{\sigma,\tau}^{-1}$ $(a \in N)$; and (3) $c_{\sigma,\tau}c_{\sigma\tau,\rho} = s_\sigma(c_{\tau,\rho})c_{\sigma,\tau\rho}$. Then the set G of all symbols as_σ $(a \in N, \sigma \in F)$ is an extension of F by N if we define multiplication by $as_\sigma \cdot bs_\tau = (as_\sigma(b)c_{\sigma,\tau})s_{\sigma\tau}$. In fact, the set of all elements $\bar{a} = ac_{1,1}^{-1}s_1$ $(a \in N)$ is a normal subgroup $\bar{N}$ of G such that $G/\bar{N} \cong F$. Any extension can be obtained in this way. A system $(s_\sigma, c_{\sigma,\tau})$ satisfying (1), (2), and (3) above is called a **factor set** belonging to F. Two factor sets $(s_\sigma, c_{\sigma,\tau})$ and $(t_\sigma, d_{\sigma,\tau})$ are said to be **associated** if there exist elements a_σ $(\sigma \in F)$ of N such that $t_\sigma(a) = s_\sigma(a_\sigma a a_\sigma^{-1})$ and $d_{\sigma,\tau} = a_\sigma(s_\sigma(a_\tau))c_{\sigma,\tau}a_{\sigma\tau}^{-1}$. In this case, two extensions determined by these factor sets are isomorphic. If $(s_\sigma, c_{\sigma,\tau})$ is associated with $(t_\sigma, d_{\sigma,\tau})$ $(d_{\sigma,\tau} = 1$ for any $\sigma, \tau \in F)$, then we say that the corresponding extension is a **split extension**. In this case, the extension G contains a subgroup $\bar{F}$ isomorphic to F, and $G = \bar{F}\bar{N}$, $\bar{F} \cap \bar{N} = \{e\}$. We call such an extension a **semidirect product** of N and F.

If N is Abelian, then condition (2) is simply $s_\sigma(s_\tau(a)) = s_{\sigma\tau}(a)$, since the only inner automorphism of N is the identity mapping. The conditions of associated factor sets and split extension are also simplified. If N is contained in the center of G, then G is called a **central extension** of N.

O. Transfers

Let H be a subgroup of finite index in G and g_i $(i = 1, \ldots, h)$ be representatives of the right cosets of H. For $b \in Hg_i$ we write $g_i = \bar{b}$. Then for $x \in G$ an element $\tilde{x} = H'\prod_{i=1}^{h} g_i x (\overline{g_i x})^{-1}$ of H/H' is determined uniquely (independent of the choice of representatives), where H' is the commutator subgroup of H. The correspondence $G'x \to \tilde{x}$ yields a homomorphism of G/G' to H/H', which is called the **transfer** from G/G' to H/H'.

P. Generalizations

The concept of group can be generalized in several ways. A set S in which a multiplication $(a,b) \to ab$ satisfying $(ab)c = a(bc)$ (the associative law) is defined is called a **semigroup**. If S is a commutative semigroup in which $ax = bx$ implies $a = b$ (the **cancellation law**), then S can be embedded in a group G so that the multiplication in S is preserved in G and any $x \in G$ is the quotient of two elements of $S: x = a^{-1}b = ba^{-1}$ $(a, b \in S)$. Such a group G is determined uniquely by S. We call it the **group of quotients** of S.

The notion of semigroup is obtained by taking only associativity from the group axioms. On the other hand, if Q is a set with a law of composition $(a,b) \to ab$ which is not necessarily associative but satisfies the condition that any two among a, b, c in the equation $ab = c$ determine the third uniquely, then Q is called a **quasigroup**. A quasigroup with an identity element e such that $ea = ae = a$ for every element a is called a **loop**. For loops, we have an analog of the structure theory of groups (R. H. Bruck, *Trans. Amer. Math. Soc.*, 60 (1946)).

If we give up the possibility of forming products for all pairs of elements or the uniqueness of the product in the axioms for groups, then we have the following generalizations of groups. A set M with multiplication under which to any elements a, $b \in M$ there corresponds a nonempty subset ab of M is called a **hypergroupoid**. Moreover, if the associative law $(ab)c = a(bc)$ holds and for any elements a, $b \in M$ there exist x, $y \in M$ such that $b \in xa$, $b \in ay$, then M is called a **hypergroup**.

A set M is called a **mixed group** if (1) M can be partitioned into disjoint subsets M_0, M_1, $M_2, \ldots$; (2) for $a \in M_0$, $b \in M_i$ $(i = 0, 1, 2, \ldots)$,

elements ab, $a\backslash b$ of M_i are defined such that $a(a\backslash b)=b$; (3) for b, $c\in M_i$, an element b/c of M_0 is defined such that $(b/c)\cdot c=b$; and (4) the associative law $(ab)c=a(bc)$ $(a,b\in M_0, c\in M)$ holds (A. Loewy, 1927).

A set M is called a **groupoid** if (1) M can be partitioned into disjoint subsets M_{ij} $(i,j=1,2,\dots)$; (2) for $a\in M_{ij}$ and $b\in M_{jk}$, an element $ab\in M_{ik}$ is defined; (3) for $a\in M_{ij}$ and $b\in M_{ik}$, an element $a\backslash b\in M_{jk}$ is defined such that $a(a\backslash b)=b$; (4) for $a\in M_{ij}$ and $b\in M_{kj}$, an element $a/b\in M_{ik}$ is defined such that $(a/b)b=a$; and (5) for $a\in M_{ij}$, $b\in M_{jk}$, and $c\in M_{kl}$, the associative law $a(bc)=(ab)c$ holds (H. Brandt, 1926). These generalized concepts also have some practical applications (→ 2 Abelian Groups; 13 Algebraic Groups; 60 Classical Groups; 69 Compact Groups; 92 Crystallographic Groups; 122 Discontinuous Groups; 151 Finite Groups; 161 Free Groups; 243 Lattices; 249 Lie Groups; 277 Modules; 362 Representations; 422 Topological Abelian Groups; 423 Topological Groups; 437 Unitary Representations).

Q. History

The concept of the group was first introduced in the early 19th century, but its rudiments can be found in antiquity; in fact, it was virtually contained in the concept of motion or transformation used in ancient geometry. From the time it took explicit form in the late 19th century, it has played a fundamental role in all fields of mathematics.

In their study of algebraic equations in the late 18th century, J. L. †Lagrange, A. T. Vandermonde, and P. Ruffini saw the importance of the group of permutations of roots; using this idea N. H. Abel showed that a general equation of degree $\geqslant 5$ cannot be solved algebraically. A. L. †Cauchy studied the group of permutations of roots for its own interest, but a complete description of the relationship between groups and algebraic equations was first given by E. †Galois. C. Jordan developed a detailed exposition of the theory given by Abel and Galois in his *Traité des substitutions* (1870) [1]. Up to that time, a group meant a permutation group; the axiomatic definition of a group was given by A. Cayley (1854) and L. Kronecker (1870). F. †Klein emphasized the significance of group theory in geometry in his †Erlangen program (1872), and M. S. †Lie developed the theory of †Lie groups in the 1880s. In 1897, W. Burnside published his *Theory of groups* [3], whose second edition (1911) is one of the classics in group theory and is still valuable. Since 1896, G. Frobenius [2] and others have developed the theory of representation of groups by matrices (→ 362 Representations). By that time, the theory of finite groups had acquired all its essential features. Among the branches of abstract algebra, the theory of groups was the first to develop; it led to the progress of abstract algebra in the 1930s. Since the latter half of that decade, the theory of finite groups has been developed further; there has been increased interest in the theory, and many significant results have been obtained, especially since 1955 (→ 151 Finite Groups).

References

[1] C. Jordan, Traité des substitutions et des équations algébriques, Gauthier-Villars, 1870 (Blanchard, 1957).
[2] G. Frobenius, Gesammelte Abhandlungen I–III, Springer, 1968.
[3] W. Burnside, Theory of groups of finite order, Cambridge Univ. Press, second edition, 1911.
[4] H. Weyl, The classical groups, Princeton Univ. Press, revised edition, 1946.
[5] B. L. van der Waerden, Algebra, Springer, I, seventh edition, 1966; II, fifth edition, 1967.
[6] B. L. van der Waerden, Gruppen von linearen Transformationen, Erg. Math., Springer, 1935 (Chelsea, 1948).
[7] A. G. Kurosh, The theory of groups I, II, Chelsea, 1960. (Original in Russian, 1953.)
[8] M. Suzuki, Structure of a group and the structure of its lattice of subgroups, Erg. Math., Springer, 1956.
[9] M. Hall, The theory of groups, Macmillan, 1959.
[10] H. Zassenhaus, Lehrbuch der Gruppentheorie I, Teubner, 1937; English translation, The theory of groups, Chelsea, 1958.
[11] A. Speiser, Die Theorie der Gruppen von endlicher Ordnung, Springer, third edition, 1937.
[12] W. Specht, Gruppentheorie, Springer, 1956.
[13] R. H. Bruck, A survey of binary systems, Erg. Math., Springer, 1958.
[14] A. H. Clifford and G. B. Preston, The algebraic theory of semigroups, Amer. Math. Soc. Math. Surveys, I, 1961; II, 1967.
[15] E. S. Liapin, Semigroups, Amer. Math. Soc. Transl. of Math. Monographs, 1963.

191 (VII.8) G-Structures

Differential geometry studies differentiable manifolds and geometric objects or structures

on them. Among the geometric structures, the Riemannian and complex structures, with their contacts with other fields of mathematics and with their richness in results, occupy a central position in differential geometry. Perhaps it is impossible to say precisely what a differential geometric structure is or should be. However, the notion of **G-structure** allows a unified description of many of the interesting known geometric structures, such as Riemannian and complex structures.

A. The Notion of G-Structures

Let M be an m-dimensional C^∞-manifold, and let $\tau: T(M) \to M$ be its †tangent bundle. For $x \in M$, $T_x(M) = \tau^{-1}(x)$ is the vector space of tangent vectors at x. Let $\pi: F(M) \to M$ denote the †frame bundle of M. It is a $GL(m; \mathbf{R})$-principal bundle on M and $F_x(M) = \pi^{-1}(x)$ is the set of linear isomorphisms (called **frames** at x) of $\mathbf{R}^m$ onto $T_x(M)$. Here $GL(m; \mathbf{R})$ is the general linear group of $\mathbf{R}^m$. Hence $F_x(M)$ can be naturally identified with the set of ordered bases of $T_x(M)$. Write $\theta = (\theta^1, \ldots, \theta^m)$ for the **canonical form** of $F(M)$, which is the $\mathbf{R}^m$-valued 1-form on $F(M)$ defined by $\theta(v) = p^{-1}(\pi_*(v))$ for any $v \in T_p(F(M))$. Given a diffeomorphism $f: M \cong N$ of M onto another C^∞-manifold N, we can naturally define the bundle isomorphism $f^{(1)}: F(M) \cong F(N)$ by $f^{(1)}(p) = f_* \circ p$.

Let G be a †Lie subgroup of $GL(m; \mathbf{R})$. A principal G-subbundle $\pi_p: P \to M$ of the frame bundle $\pi: F(M) \to M$ is called a **G-structure** over M. Thus P is a regular submanifold of $F(M)$ satisfying the following three conditions:

(A1) $\pi(P) = M$.

(A2) for $p \in P$ and $\sigma \in GL(m; \mathbf{R})$, $p\sigma \in P$ if and only if $\sigma \in G$.

(A3) for any $x \in M$, there exist an open neighborhood U of x and a C^∞-mapping $s: U \to P$ such that $\pi(s(y)) = y$ for any $y \in U$.

Conversely, if a regular submanifold P of $F(M)$ satisfies the above three conditions (A1), (A2), and (A3), then $\pi_P = \pi|_P: P \to M$ is a G-structure over M. When G is closed, then condition (A3) is automatically satisfied. The restriction of θ onto P is called the **canonical form** of P and is also denoted by θ. For an open subset U of M, the restriction $P|U = \pi_P^{-1}(U)$ is a G-structure over U.

Let $\pi_P: P \to M$ and $\pi_Q: Q \to N$ be G-structures over M and N, respectively. A diffeomorphism $f: M \cong N$ is called a **G-isomorphism** of P onto Q if $f^{(1)}(P) = Q$. When such an f exists, we say that the G-structures P and Q are **equivalent**. A G-isomorphism of P onto itself is called a G-automorphism of P. Write $\mathrm{Aut}(M, P)$ for the group of G-automorphisms of P.

Let $\mathfrak{g}$ be the Lie algebra of G. Since G acts on P from the right, we have the natural Lie algebra homomorphism $\iota: \mathfrak{g} \to \Gamma(P, T(P))$. Here $\Gamma(P, T(P))$ is the Lie algebra of C^∞-vector fields on P. Actually ι is injective, and we write A^* for $\iota(A)$ $(A \in \mathfrak{g})$. We call A^* the **fundamental vector field** corresponding to A.

For $x \in \mathbf{R}^m$, define $p_0(x) \in F_x(\mathbf{R}^m)$ by $p_0(x)((v^j)) = \sum_{j=1}^m v^j (\partial/\partial x^j)_x$, where $x^1, \ldots, x^m$ are the natural coordinates of $\mathbf{R}^m$. Any open subset D of $\mathbf{R}^m$ has the natural G-structure $P(D, G)$ defined by $P(D, G) = \{p_0(x) \cdot \sigma \in F_x(\mathbf{R}^m); x \in D, \sigma \in G\}$. We say that a G-structure π_P: $P \to M$ over M is **integrable** if for each $x \in M$, there exists a local coordinate system (U, φ, D) around x such that φ is a G-isomorphism of $P|U$ onto $P(D, G)$. Equivalently, if we set $\varphi = (x^1, \ldots, x^m)$, then the frame $\{(\partial/\partial x^1)_y, \ldots, (\partial/\partial x^m)_y\}$ belongs to P for each $y \in U$.

B. Examples

The following examples (B1)–(B7) show that some classical geometric structures can be treated uniformly from the viewpoint of G-structures.

(B1) An **absolute parallelism** on M is, by definition, a system $\{X_1, \ldots, X_m\}$ of C^∞-vector fields on M such that at each point $x \in M$ $\{X_1(x), \ldots, X_m(x)\}$ is a basis of $T(M)_x$. Clearly this is equivalent to giving an $\{e\}$-structure $\pi_P: P \to M$, where e stands for the identity matrix of $GL(m; \mathbf{R})$. Then P is integrable if and only if $[X_i, X_j] = 0$, $1 \leqslant i, j \leqslant m$. In this case, $\mathrm{Aut}(M, P)$ is a finite-dimensional †Lie transformation group on M.

(B2) Let E be a k-dimensional †distribution, namely, E is a k-dimensional vector subbundle of $T(M)$. Write $\mathbf{R}^m = \mathbf{R}^k \oplus \mathbf{R}^{m-k}$. Set $GL(k, m; \mathbf{R}) = \{\sigma \in GL(m; \mathbf{R}); \sigma(\mathbf{R}^k) \subset \mathbf{R}^k\}$ and $P = \{p \in F(M); p(\mathbf{R}^k) \subset E\}$. Then P is a $GL(k, m; \mathbf{R})$-structure over M. This gives a bijective correspondence between the k-dimensional distributions on M and the $GL(k, m; \mathbf{R})$-structures on M. Moreover, P is integrable if and only if E is †involutive (Frobenius theorem).

(B3) Let g be a Riemannian metric on M. Set $O(m) = \{\sigma \in GL(m; \mathbf{R}); {}^t\sigma\sigma = e\}$. Define $P(g)$ by $P(g) = \{(v_1, \ldots, v_m) \in F(M); g(v_i, v_j) = \delta_{ij}\}$. Then $P(g)$ is an $O(m)$-structure on M. This gives a bijective correspondence between the Riemannian metrics on M and the $O(m)$-structures over M. Then $\mathrm{Aut}(M, P(g))$ is the group of †isometries of g, and a finite-dimensional Lie transformation group on M. Moreover, $P(g)$ is integrable if and only if g is flat in the sense that the Riemannian curvature tensor of g is zero.

(B4) Two Riemannian metrics g_1 and g_2 are

called **conformally equivalent** if there exists a C^∞-function $\rho>0$ with $g_1=\rho g_2$. This is an equivalence relation among the Riemannian metrics on M. An equivalence class $\{g\}$ is called a **conformal structure** on M. Set $CO(m)=\{aA; A\in O(m), a\in\mathbf{R}-\{0\}\}$. Define $P(\{g\})$ by $P(\{g\})=\{(v_1,\ldots,v_m)\in F(M); g(v_i,v_j)=a\delta_{ij}, a\in\mathbf{R}-\{0\}\}$. Then $P(\{g\})$ is a $CO(m)$-structure on M. This gives a bijective correspondence between the conformal structures on M and the $CO(m)$-structures over M. Then $\mathrm{Aut}(M,P(\{g\}))$ is the group of †conformal transformations of $\{g\}$ and is a finite-dimensional Lie transformation group. Moreover $P(\{g\})$ is integrable if and only if the conformal structure $\{g\}$ is **conformally flat** in the sense that at each point $x\in M$, there exists a local coordinate system $(x^1,\ldots,x^m)$ around x such that $g(\partial/\partial x^i,\partial/\partial x^j)=\rho\delta_{ij}$, where ρ is a positive function.

(B5) Assume that m is even, say $m=2l$. By an **almost symplectic structure** on M, we mean a differential 2-form Ω of maximal rank, i.e., $\Omega\wedge\ldots\wedge\Omega$ (l times) never vanishes. Let A be the standard skew-symmetric bilinear form on $\mathbf{R}^{2l}$ defined by $A((u^i),(v^j))=(u^1v^2-u^2v^1)+\ldots+(u^{2l-1}v^{2l}-u^{2l}v^{2l-1})$. The symplectic group $Sp(l;\mathbf{R})$ is defined by $Sp(l;\mathbf{R})=\{\sigma\in GL(2l;\mathbf{R}); A(\sigma u,\sigma v)=A(u,v),\ u,\ v\in\mathbf{R}^{2l}\}$. Set $P(\Omega)=\{p\in F(M); p^*A=\Omega_x, x=\pi_P(p)\}$. Then $P(\Omega)$ is an $Sp(l;\mathbf{R})$-structure over M. This gives a bijective correspondence between the almost symplectic structures on M and the $Sp(l;\mathbf{R})$-structures over M. Then $\mathrm{Aut}(M,P(\Omega))$ is the group of symplectic transformations of Ω, which is never a finite-dimensional Lie transformation group. Moreover $P(\Omega)$ is integrable if and only if Ω is a symplectic structure, i.e., $d\Omega=0$.

(B6) Assume that m is even, say $m=2l$. We identify $\mathbf{C}^l$ with $\mathbf{R}^{2l}$ as a real vector space. Then $GL(m;\mathbf{C})$ is a Lie subgroup of $GL(2l;\mathbf{R})$. Let J be an almost complex structure on M. Set $P(J)=\{p\in F(M); Jp(u)=p(iu), u\in\mathbf{R}^{2l}=\mathbf{C}^l\}$. Then $P(J)$ is a $GL(l;\mathbf{C})$-structure on M. This gives a bijective correspondence between the almost complex structures on M and the $GL(l;\mathbf{C})$-structures over M. Then $\mathrm{Aut}(M,P(J))$ is the group of J-analytic transformations on M. Furthermore, $\mathrm{Aut}(M,P(J))$ is a finite-dimensional Lie transformation group when M is compact. Moreover $P(J)$ is integrable if and only if J is a complex structure.

(B7) It is easy to see that there is a bijective correspondence between the $SL(m;\mathbf{R})$-structures over M and the set of volume elements on M. Any $SL(m;\mathbf{R})$-structure over M is always integrable and $\mathrm{Aut}(M,P)$ is the group of diffeomorphisms preserving Ω, which is never a finite-dimensional Lie transformation group.

C. Structure Functions

Let G be a Lie subgroup of $GL(m;\mathbf{R})$ and $\mathfrak{g}$ the Lie algebra of G. For convenience, we write V for $\mathbf{R}^m$ and V^* for the dual space of V. Then $V\otimes\bigwedge^2V^*$ can be considered as the space of skew-symmetric bilinear mappings from $V\times V$ into V, and $\mathfrak{g}\otimes V^*$ can be identified with the space of linear mappings from V into $\mathfrak{g}$. Define a linear mapping $\partial:\mathfrak{g}\otimes V^*\to V\otimes\bigwedge^2V^*$ by

$$(\partial T)(u,v)=-T(u)v+T(v)u \qquad \text{for} \quad T\in\mathfrak{g}\otimes V^*, \quad u,v\in V.$$

Put

$$H^{2,1}(\mathfrak{g})=V\otimes\bigwedge^2V^*/\partial(\mathfrak{g}\otimes V^*).$$

Let $\pi_P:P\to M$ be a G-structure over M and θ the canonical form of P. Take any frame p in P. An m-dimensional linear subspace H of $T_p(P)$ is said to be a **horizontal subspace** of P at p if $\theta_p:H\to V(=\mathbf{R}^m)$ is a linear isomorphism. Then we write f_H for $\theta_p^{-1}:V\to H$. Write $\mathfrak{H}(T_p(P))$ for the set of horizontal subspaces of P at p. For $H\in\mathfrak{H}(T_p(P))$, define $c(p,H)\in V\otimes\bigwedge^2V^*$ by

$$c(p,H)(u,v)=d\theta(f_H(u),f_H(v)) \quad \text{for} \quad u,v\in V.$$

Then we have

$$c(p,H_1)-c(p,H_2)\in\partial(\mathfrak{g}\otimes V^*) \qquad \text{for} \quad H_1,H_2\in\mathfrak{H}(T_p(P)).$$

Therefore we have a well-defined element $c(p)\in H^{2,1}(\mathfrak{g})$ by $c(p)=\{c(p,H)\}$. We call the mapping $c:P\to H^{2,1}(\mathfrak{g})$ the **structure function** of the G-structure $\pi_P:P\to M$.

Let $Q\to N$ be another G-structure over N and $f:M\cong N$ be a G-isomorphism of P onto Q. Then we have $c(f^{(1)}(p))=c(p)$ for $p\in P$. In particular, it is easily seen that if P is integrable, then $c\equiv0$. However, the converse is not true in general.

D. Prolongation of Linear Lie Algebras and Groups

Let $\mathbf{K}$ denote either $\mathbf{R}$ or $\mathbf{C}$. Let $\mathfrak{g}$ be a Lie subalgebra of $\mathfrak{gl}(m;\mathbf{K})$. For $k=0,1,2,\ldots$, let $\mathfrak{g}_k$ be the space of K-symmetric multilinear mappings

$$t:\underbrace{\mathbf{K}^m\times\ldots\times\mathbf{K}^m}_{(k+1)\text{ times}}\to\mathbf{K}^m$$

such that for each fixed $u_1,\ldots,u_k\in\mathbf{K}^m$, the linear transformation

$$v\in\mathbf{K}^m\mapsto t(v,u_1,\ldots,u_k)\in\mathbf{K}^m$$

belongs to $\mathfrak{g}$. In particular, $\mathfrak{g}_0=g$. We call $\mathfrak{g}_k$

the k**th prolongation** of $\mathfrak{g}$. If $\mathfrak{g}_k=0$, then $\mathfrak{g}_{k+1}=\mathfrak{g}_{k+2}=\ldots=0$. The first integer k such that $\mathfrak{g}_k=0$ is called the **order** of $\mathfrak{g}$. If $\mathfrak{g}_k\neq 0$ for all k, then $\mathfrak{g}$ is said to be of **infinite type**. When $\mathfrak{g}$ contains a matrix of rank 1 as an element, then $\mathfrak{g}$ is of infinite type. When $\mathbf{K}=\mathbf{C}$, the converse is also true [3]. However, when $\mathbf{K}=\mathbf{R}$, the converse is not true in general. In fact, if we consider $\mathfrak{gl}(l;\mathbf{C})$ as a real Lie subalgebra of $\mathfrak{gl}(2l;\mathbf{R})$ (example (B6)), then $\mathfrak{gl}(l;\mathbf{C})$ is of infinite type, having no matrix of rank 1 in it. In general, a Lie subalgebra $\mathfrak{g}$ of $\mathfrak{gl}(m;\mathbf{R})$ is said to be of **elliptic type** if $\mathfrak{g}$ contains no matrix of rank 1 as an element. We consider $\mathfrak{g}_1$ as an Abelian Lie subalgebra of $\mathfrak{gl}(\mathbf{K}^m\oplus\mathfrak{g})$ by the correspondence $t\in\mathfrak{g}_1\mapsto\bar{t}\in\mathfrak{gl}(\mathbf{K}^m\oplus\mathfrak{g})$, where $\bar{t}$ is defined by

$$\bar{t}(u)=t(\cdot,u)\quad\text{for}\quad u\in\mathbf{K}^m,$$

$$\bar{t}(A)=0\qquad\text{for}\quad A\in\mathfrak{g}.$$

More generally, we consider $\mathfrak{g}_{k+1}$ to be an Abelian Lie subalgebra of $\mathfrak{gl}(\mathbf{K}^m\oplus\mathfrak{g}_0\oplus\ldots\oplus\mathfrak{g}_k)$ by virtue of the correspondence $t\in\mathfrak{g}_{k+1}\mapsto\bar{t}\in\mathfrak{gl}(\mathbf{K}^m\oplus\mathfrak{g}_0\oplus\ldots\oplus\mathfrak{g}_k)$, where $\bar{t}$ is defined by

$$\bar{t}(u)=t(\cdot,\ldots,\cdot,u)\in\mathfrak{g}_k\quad\text{for}\quad u\in\mathbf{K}^m,$$

$$\bar{t}(A)=0\qquad\text{for}\quad A\in\mathfrak{g}_0\oplus\ldots\oplus\mathfrak{g}_k.$$

Identifying $\mathfrak{gl}((\mathbf{K}^m\oplus\mathfrak{g}_0\oplus\ldots\oplus\mathfrak{g}_{k-1})\oplus\mathfrak{g}_k)$ with $\mathfrak{gl}(\mathbf{K}^m\oplus\mathfrak{g}_0\oplus\ldots\oplus\mathfrak{g}_k)$, we have the important identity

(D1) $\quad(\mathfrak{g}_k)_1=\mathfrak{g}_{k+1},\qquad k=0,1,2,\ldots.$

Let G be a Lie subgroup of $GL(m;\mathbf{R})$ and $\mathfrak{g}$ the Lie subalgebra of $\mathfrak{gl}(m;\mathbf{R})$ corresponding to G. For $k=1,2,\ldots$, let G_k be the connected Abelian Lie subgroup of $GL(\mathbf{R}^m\oplus\mathfrak{g}_0\oplus\ldots\oplus\mathfrak{g}_{k-1})$ corresponding to the Abelian Lie subalgebra $\mathfrak{g}_k$; namely, G_k consists of the linear mappings $\sigma(\bar{t})$ $(t\in\mathfrak{g}_k)$ defined by

$$\sigma(\bar{t})(u)=u+t(\cdot,\ldots,\cdot,u)\quad\text{for } u\in\mathbf{R}^m,$$

$$\sigma(\bar{t})(A)=A\quad\text{for } A\in\mathfrak{g}_0\oplus\ldots\oplus\mathfrak{g}_{k-1}.$$

We call G_k the k**th prolongation** of G. We say that G is of finite (resp. infinite) type if $\mathfrak{g}$ is of finite (resp. infinite) type. We say that G is of elliptic type if $\mathfrak{g}$ is. From (D1), we have

(D2) $\quad(G_k)_1=G_{k+1}.$

We know all irreducible Lie subalgebras $\mathfrak{g}$ of $\mathfrak{gl}(m;\mathbf{K})$ with order $\geqslant 2$ [3, 6].

(D3) An irreducible Lie subalgebra $\mathfrak{g}$ of $\mathfrak{gl}(m;\mathbf{C})$ is of infinite type if and only if $\mathfrak{g}$ is one of

$$\mathfrak{gl}(m;\mathbf{C}),\quad\mathfrak{sl}(m;\mathbf{C}),\quad\mathfrak{csp}\left(\frac{m}{2};\mathbf{C}\right),\quad\mathfrak{sp}\left(\frac{m}{2};\mathbf{C}\right),$$

where $\mathfrak{csp}(m/2;\mathbf{C})=\mathbf{C}\oplus\mathfrak{sp}(m/2,\mathbf{C})$.

(D4) An irreducible Lie algebra $\mathfrak{g}$ of $\mathfrak{gl}(m;\mathbf{R})$ is of infinite type if and only if $\mathfrak{g}$ is one of

$$\mathfrak{gl}\left(\frac{m}{2};\mathbf{C}\right),\quad\mathfrak{sl}\left(\frac{m}{2};\mathbf{C}\right),\quad\mathfrak{csp}\left(\frac{m}{4};\mathbf{C}\right),$$

$$\mathfrak{sp}\left(\frac{m}{4};\mathbf{C}\right),$$

$$\mathfrak{gl}(m;\mathbf{R}),\quad\mathfrak{sl}(m;\mathbf{R}),\quad\mathfrak{csp}\left(\frac{m}{2};\mathbf{R}\right),\quad\mathfrak{sp}\left(\frac{m}{2};\mathbf{R}\right),$$

where $\mathfrak{csp}(m/2;\mathbf{R})=\mathbf{R}\oplus\mathfrak{sp}(m/2;\mathbf{R})$.

(D5) Let S be an m-dimensional irreducible Hermitian symmetric space of compact type, which is different from the complex projective space. Let $L(S)$ be the identity connected component of the group of biholomorphic transformations of S. Take any point $o\in S$ and set $L_0(S)=\{\sigma\in L(S);\sigma(o)=o\}$. Let $\mathfrak{g}(S)$ be the linear isotropy Lie subalgebra of $\mathfrak{gl}(m;\mathbf{C})$ corresponding to $L(S)$. Then $\mathfrak{g}(S)$ is an irreducible Lie subalgebra of $\mathfrak{gl}(m;\mathbf{C})$ of order 2. Conversely, every irreducible Lie subalgebra $\mathfrak{g}$ of $\mathfrak{gl}(m;\mathbf{C})$ of finite type with order $\geqslant 2$ is equal to $\mathfrak{g}(S)$ as given above.

(D6) Let S be an m-dimensional irreducible symmetric space of compact type. Assume that S admits a finite-dimensional connected Lie transformation group $L(S)$ which strictly contains the connected component of the group of isometries of S. Take any point $o\in S$, and set $L_0(S)=\{\sigma\in L(S);\sigma(o)=o\}$. Let $\mathfrak{g}(S)$ be the linear isotropy Lie subalgebra of $\mathfrak{gl}(m;\mathbf{R})$ corresponding to $L_0(S)$. If $(S,L(S))\neq$(sphere, projective transformation group) or $(S,L(S))\neq$ (complex projective space, complex projective transformation group), then $\mathfrak{g}(S)$ is an irreducible Lie subalgebra of $\mathfrak{gl}(m;\mathbf{R})$ of order 2. Conversely, every irreducible Lie subalgebra $\mathfrak{g}$ of $\mathfrak{gl}(m;\mathbf{R})$ of finite type with order $\geqslant 2$ is equal to $g(s)$ as given above. For example, if we take an m-dimensional sphere as S and the group of conformal transformations of S as $L(S)$, then $\mathfrak{g}(s)=\mathfrak{co}(m)$.

E. Prolongation of G-Structures

Let G be a Lie subgroup of $GL(m;\mathbf{R})$ and let $\mathfrak{g}$ be the Lie subalgebra of $\mathfrak{gl}(m;\mathbf{R})$ corresponding to G. We choose once and for all a linear subspace C of $V\otimes\bigwedge^2V^*$ such that

$$V\otimes\textstyle\bigwedge^2V^*=\partial(\mathfrak{g}\otimes V^*)\oplus C.$$

In general there is no natural way of choosing such a C.

Let $\pi_P:P\to M$ be a G-structure on M. For each horizoantal subspace H of P at p (i.e., $H\in\Theta(T_p(P))$), we define the frame $(p,H)\in F_p(P)$

to be a linear isomorphism $(p, H): \mathbf{R}^m \oplus \mathfrak{g} \to T_p(P)$ given by

$$(p, H)(v \oplus A) = f_H(v) \oplus A_p^* \quad \text{for} \quad v \in \mathbf{R}^m, A \in \mathfrak{g},$$

where A^* is the fundamental vector field on P induced from $A \in \mathfrak{g}$. Let $P_1 = \{(p, H) \in F(P); H \in \mathscr{H}(T_p(P)), c(p, H) \in C\}$. Then $P_1 \to P$ is a G_1-structure over P. We call $P_1 \to P$ the **first prolongation** of P. The **kth prolongation** P_k of P is defined inductively by $P_k = (P_{k-1})_1 =$ the first prolongation of P_{k-1}. From (D2), P_k is a G_k-structure over P_{k-1}.

Take any $f \in \mathrm{Aut}(M, P)$. It can be proved that $f^{(1)} \in \mathrm{Aut}(P, P_1)$. Since $P_k = (P_{k-1})_1$, we can define $f^{(k)} \in \mathrm{Aut}(P_{k-1}, P_k)$ inductively by $f^{(k)} = (f^{(k-1)})^{(1)}$. By the correspondence $f \mapsto f^{(k)}$, we can consider $\mathrm{Aut}(M, P)$ as a subgroup of $\mathrm{Aut}(P_{k-1}, P_k)$.

If G is connected, then

(E1) $\quad \mathrm{Aut}(M, P) = \mathrm{Aut}(P_{k-1}, P_k), \quad k = 1, 2, \ldots.$

Let $Q \to N$ be another G-structure. By a similar argument to that above, we see that $P \to M$ and $Q \to N$ are equivalent if and only if $P_k \to P_{k-1}$ and $Q_k \to Q_{k-1}$ are equivalent. In [1], E. Cartan studied general equivalence problems of two G-structures. In that problem the above prolongation procedure plays an important role.

F. Automorphisms of G-Structures of Finite Type

S. Kobayashi proved the following:

(F1) Theorem [4]. Let $P \to M$ be an $\{e\}$-structure over M. Then $\mathrm{Aut}(M, P)$ is a finite-dimensional Lie transformation group on M such that $\dim \mathrm{Aut}(M, P) \leqslant \dim M$. More precisely, for any point $x \in M$, the mapping $\sigma \in \mathrm{Aut}(M, P) \mapsto \sigma(x) \in M$ is injective, and its image $\{\sigma(x); \sigma \in \mathrm{Aut}(M, P)\}$ is a closed submanifold of M. The submanifold structure on this image makes $\mathrm{Aut}(M, P)$ into a finite-dimensional Lie transformation group on M.

Now let G be a Lie subgroup of $GL(m; \mathbf{R})$ of finite type, say $G_k = \{e\}$. Then $P_k \to P_{k-1}$ is an $\{e\}$-structure. From theorem (F1) above, it follows that $\mathrm{Aut}(P_{k-1}, P_k)$ is a finite-dimensional Lie transformation group. As explained in Section E, $\mathrm{Aut}(M, P)$ is a subgroup of $\mathrm{Aut}(P_{k-1}, P_k)$. Clearly $\mathrm{Aut}(M, P)$ is closed in $\mathrm{Aut}(P_{k-1}, P_k)$. Since every closed subgroup of a Lie group is again a Lie group, we have the following:

(F2) Theorem. Let $P \to M$ be a G-structure over M. If G is of finite type, then $\mathrm{Aut}(M, P)$ is a Lie transformation group of dimension $\leqslant \dim(R^m \oplus \mathfrak{g} \oplus \mathfrak{g}_1 \oplus \ldots \oplus \mathfrak{g}_{k-1})$.

G. Automorphisms of G-Structures of Elliptic Type

The following theorem of R. Palais allows us to prove that the automorphism groups of many general geometric structures are finite-dimensional Lie transformation groups.

(G1) Theorem [5]. Let L be a group of C^∞-transformations of a C^∞-manifold M. Let $\mathfrak{l}$ be the set of all vector fields X on M which generate global 1-parameter groups $\varphi_t = \exp tX$ of transformations of M such that $\varphi_t \in L$. If the set $\mathfrak{l}$ generates a finite-dimensional Lie algebra of vector fields on M, then L is a finite-dimensional Lie transformation group, and $\mathfrak{l}$ is the Lie algebra of L.

Let $P \to M$ be a G-structure over M. For any representation $\rho: G \to GL(V)$, we write $E(\rho)$ for the associated vector bundle of P with respect to ρ. Let $\rho_1: G \to GL(\mathfrak{gl}(m; \mathbf{R}))$ be the representation defined by $\rho_1(\sigma)A = \sigma A \sigma^{-1}$ for $\sigma \in G$, $A \in \mathfrak{gl}(m; \mathbf{R})$. Let $\rho_2: G \to GL(\mathfrak{g})$ be the representation defined by $\rho_2(\sigma)A = \sigma A \sigma^{-1}$ for $\sigma \in G$, $A \in \mathfrak{g}$. Then ρ_2 is a subrepresentation of ρ_1, since $\rho_1(\sigma)A = \rho_2(\sigma)A$ for $\sigma \in G$, $A \in \mathfrak{g}$. We remark that $E(\rho_1) = \mathrm{Hom}(T(M), T(M))$. We write $\mathfrak{g}(M)$ for $E(\rho_2)$. Then $\mathfrak{g}(M)$ is a vector subbundle of $\mathrm{Hom}(T(M), T(M))$. We write F for the quotient vector bundle $\mathrm{Hom}(T(M), T(M))/\mathfrak{g}(M)$. Let Φ: $\mathrm{Hom}(T(M), T(M)) \to F$ be the natural projection. We fix an affine connection ∇ on M which preserves P. We write T for the torsion tensor field of ∇. For each vector field $X \in \Gamma(M, T(M))$, we define a C^∞-section $\nabla X + T(X, \cdot)$ in $\Gamma(M, \mathrm{Hom}(T(M), T(M))$ by

$$u \in T(M)_x \mapsto \nabla_u X + T(X(x), u) \in T_x(M).$$

Then define a first-order linear differential operator

$$D: \Gamma(M, T(M)) \to \Gamma(M, F)$$

by $D(X) = \Phi(\nabla X + T(X, \cdot))$. It is easy to see that

(G2) $\quad \mathfrak{l}(M, P) \subset \ker D,$

where $\mathfrak{l}(M, P)$ is the Lie algebra of all vector fields $X \in \Gamma(M, T(M))$ which generate global 1-parameter groups $\varphi_t = \exp tX$ of transformations of M such that $\varphi_t \in \mathrm{Aut}(M, P)$. Then we can show:

(G3) D is an elliptic operator if and only if G is of elliptic type.

Now suppose M is compact. Then from the standard fact on linear elliptic operators on compact manifolds, we know the dimension of $\ker D$ is finite if D is elliptic. Thus from (G2), (G3), and theorem (G1), we have the following:

(G4) Theorem [7]. Let $P \to M$ be a G-structure on an m-dimensional compact mani-

fold M. If G is of elliptic type, then $\mathrm{Aut}(M, P)$ is a finite-dimensional Lie transformation group.

H. Local Equivalence Problem

Let $\pi_P: P \to M$ and $\pi_Q: Q \to N$ be two G-structures. We say that $P \to M$ and $Q \to N$ are **locally equivalent** if the following holds:

(H1) For arbitrarily given $(p, q) \in P \times Q$, there exists a local G-isomorphism $f: U \cong V$ of $P \mid U$ onto $Q \mid V$ such that $f^{(1)}(p) = q$, where U (resp. V) is an open neighborhood of $\pi_P(p)$ (resp. $\pi_Q(q)$).

From now on, we assume that the structure function c_P (resp. c_Q) of P (resp. Q) is constant. If such an f as in (H1) exists, we must have $c_Q(f^{(1)}(p')) = c_p(p')$ for $p' \in P \mid U$. Therefore c_P and c_Q must be the same constant. Cartan reduced the local equivalence problem between P and Q to that of certain differential systems on $P \times Q$. In fact, let $\theta_P = (\theta^1, \ldots, \theta^m)$ and $\theta_Q = (\psi^1, \ldots, \psi^m)$ be the canonical forms of P and Q respectively. Let $\alpha: P \times Q \to P$, $\beta: P \times Q \to Q$ be the canonical projections such that $\alpha(p, q) = p$, $\beta(p, q) = q$. Let Σ be the differential system on $P \times Q$ generated by $\{\alpha * \theta^1 - \beta * \psi', \ldots, \alpha * \theta^m - \beta * \psi^m\}$. If X is the m-dimensional regular submanifold of $P \times Q$ given by the graph of $f^{(1)}$ in (H1), i.e., $X = \{(p', f^{(1)}(p')) \in P \times Q; p' \in P \mid U\}$, then X is an m-dimensional integral submanifold of Σ satisfying the following conditions:

(H2) $\quad (p, q) \in X$,

(H3) $\quad \alpha * \theta^1, \ldots, \alpha * \theta^m$ are linearly independent on $T_{(p,q)}(X)$.

Conversely any m-dimensional integral submanifold of Σ satisfying (H2) and (H3) is the graph of a local G-isomorphism f required in (H1). Therefore the local equivalence problem between P and Q is equivalent to the problem of finding an m-dimensional integral submanifold X of Σ satisfying (H2) and (H3) for any $(p, q) \in P \times Q$.

We say that G is **involutive** if there exist linear subspaces $0 = V_0 \subset V_1 \subset \ldots \subset V_m = \mathbf{R}^m$ such that

(H4) $\quad \dim V_k = k \quad (k = 0, \ldots, m)$,

(H5) $\quad \dim \mathfrak{g}_1 = \sum_{k=0}^{m} \dim \mathfrak{g}(V_k)$,

where $\mathfrak{g}$ is the Lie algebra of G and $\mathfrak{g}(V_k) = \{A \in \mathfrak{g}; A(u) = 0 \text{ for } u \in V_k\}$. Now we have:

(H6) Proposition. Let $P \to M$ and $Q \to N$ be two G-structures such that their structure functions are constant and equal. The differential system Σ on $P \times Q$ defined as above is involutive at any m-dimensional integral element E of Σ on which $\alpha * \theta^1, \ldots, \alpha * \theta^m$ are linearly independent if and only if G is involutive.

Combining proposition (H6) above with the classical †Cartan-Kähler theorem, we obtain the following theorems:

(H7) Theorem ([1], also see [8]). Let $P \to M$ and $Q \to N$ be two real analytic G-structures over M and N, respectively, such that their structure functions are constant and equal. Suppose G is involutive. Then $P \to M$ and $Q \to N$ are locally equivalent.

(H8) Theorem. Let $P \to M$ be a real analytic G-structure over M. Suppose G is involutive. Then P is integrable if and only if the structure function of P is zero everywhere.

Finally, we remark that for a general Lie subgroup $G \subset GL(m, \mathbf{R})$, G_k is involutive for $k \geqslant k_0$ [8].

I. Cartan-Kähler Theorem on Differential Systems

Let U be an open subset of $\mathbf{R}^m$. We write $A^k(U)$ for the space of real analytic differential k-forms on U. Put $A^*(U) = A^0(U) \oplus \ldots \oplus A^m(U)$. Then $A^*(U)$ is a graded $\mathbf{R}$-algebra with respect to the usual exterior product. By a **differential system** on U we mean an ideal Σ of $A^*(U)$ such that

(I1) $\quad \Sigma = A^0(U) \cap \Sigma \oplus A^1(U) \cap \Sigma \oplus \ldots \oplus \Sigma \cap A^m(U)$,

(I2) $\quad d(\Sigma) \subset \Sigma$,

(I3) $\quad \Sigma$ is finitely generated as an ideal.

For convenience we write $\Sigma^{(k)} = \Sigma \cap A^k(U)$. When $\Sigma^{(0)} = \{0\}$, we call Σ a **restricted differential system**.

Let Σ be a differential system on $U \subset \mathbf{R}^m$. By a **k-dimensional integral manifold** of Σ, we mean a k-dimensional regular real analytic submanifold of U such that for any $\alpha \in \Sigma$, we have $\alpha \mid X \equiv 0$. In the above definition, it is sufficient to know that $\alpha \mid X \equiv 0$ for any $\alpha \in \Sigma^{(k)}$.

For $x \in U$, we write $G_k(T_x(U))$ for the set of all k-dimensional linear subspaces of $T_x(U)$. Set $G_k(T(U)) = \bigcup_{x \in U} G_k(T_x(U))$ (disjoint union); then $G_k(T(U))$ is naturally a real analytic manifold of dimension $m + m(m-k)$. Let Σ be a differential system on U. A k-plane $E \in G_k(T_x(U))$ is called a **k-dimensional integral element** of Σ at x if for any $\alpha \in \Sigma^{(k)}$, we have $\alpha \mid E = 0$. Thus X is a k-dimensional integral submanifold if and only if $T_x(X)$ is a k-dimensional integral element of Σ for each $x \in X$. We write $I_k(\Sigma)$ for the set of k-dimensional integral elements of Σ. In general $I_k(\Sigma)$ is a real analytic subset of $G_k(T(U))$. For any k-dimensional

integral element E at $x \in U$, we define the **polar space** $H(E)$ by

$$H(E)=\{u \in T_x(U); \alpha(u, v_1, \dots, v_k)=0 \text{ for } \alpha \in \Sigma^{(k+1)}, v_1, \dots, v_k \in E\}.$$

Then $E \subset H(E)$. Moreover $F \in G_{k+1}(T_x(U))$ with $F \supset E$ is in $I_{k+1}(\Sigma)$ if and only if $F \subset H(E)$. Let E be a k-dimensional integral element of Σ. We choose $\varphi^1, \dots, \varphi^k \in A^1(U)$ such that $\varphi^1|E, \dots, \varphi^k|E$ are linearly independent. Define $\mathscr{U}(\varphi^1, \dots, \varphi^k)$ by

$$\mathscr{U}(\varphi^1, \dots, \varphi^k)=\{F \in G_k(T(U)); \varphi^1|F, \dots, \varphi^k|F \text{ linearly independent}\}.$$

Then $\mathscr{U}(\varphi^1, \dots, \varphi^k)$ is an open neighborhood of E in $G_k(T(U))$. For any element $F \in \mathscr{U}(\varphi^1, \dots, \varphi^k)$, define $F_1, \dots, F_k \in F$ by $\varphi^i(F_k)=\delta^i_k$. Thus $\{F_1, \dots, F_k\}$ is the dual basis of $\varphi^1|F, \dots, \varphi^k|F$. Any element $\alpha \in \Sigma^{(k)}$ defines a real analytic function α_* on $\mathscr{U}(\varphi^1, \dots, \varphi^k)$ by $\alpha_*(F)=\alpha(F_1, \dots, F_k)$. We call E **regular** if there exist $\alpha^1, \dots, \alpha^l \in \Sigma^{(k)}$ and an open neighborhood $\mathscr{V}$ of E in $\mathscr{U}(\varphi^1, \dots, \varphi^k)$ such that

(I4) $I_k(\Sigma) \cap \mathscr{V}=\{F \in \mathscr{V}; \alpha^1_*(F)=\dots=\alpha^l_*(F)=0\}$,

(I5) $d\alpha^1_*, \dots, d\alpha^l_*$ are linearly independent on $\mathscr{V}$,

(I6) $\dim H(F)$ is constant for any $F \in \mathscr{V}$.

We say that $E \in I_k(\Sigma)$ is an **ordinary element** if E contains a $(k-1)$-dimensional regular integral element.

The following is the well-known theorem due to Cartan and Kähler [7].

(I7) Theorem. Let Σ be a restricted differential system on $U \subset \mathbf{R}^m$. Let X be a k-dimensional integral submanifold of Σ. Suppose that $T_x(X)$ is regular for a point $x \in X$ and that there exists a $(k+1)$-dimensional integral element F at x with $F \supset T_x(X)$. Then there exists a $(k+1)$-dimensional integral submanifold Y of Σ such that $x \in Y$, $T_x(Y)=F$, and $X \subset Y$ in a neighborhood of x.

We say that Σ is **involutive** at $E \in I_k(\Sigma)$ if there exist $0=E^0 \subset E^1 \subset \dots \subset E^{k-1} \subset E$ such that E^j is a j-dimensional regular integral element of Σ $(j=0, 1, \dots, k-1)$. Applying theorem (G7) inductively, we obtain:

Corollary. Let Σ be a restricted differential system on $U \subset \mathbf{R}^m$. If E is a k-dimensional involutive integral element of Σ at x, there exists a k-dimensional integral submanifold of Σ such that $x \in X$ and $T_x(X)=E$.

In view of this corollary, it is important to know when Σ is involutive at $E \in I_k(\Sigma)$.

(I8) Lemma [9]. Let Σ be a restricted differential system on $U \subset \mathbf{R}^m$, and E be a k-dimensional integral element of Σ. Then Σ is involutive at E if and only if there exist $0=E^0 \subset E^1 \subset \dots \subset E^{k-1} \subset E$ such that $I_k(\Sigma)$ is a submanifold of $G_k(T(U))$ near E with dimension $m+k(m-k)-\Sigma_{j=0}^{k-1} t_j$, where $t_j=m-\dim H(E^j)$.

References

[1] E. Cartan, Les problèmes d'équivalence, Oeuvres, pt. II, vol. 2, 1311–1334.
[2] V. Guillemin and S. Sternberg, An algebraic model of transitive differential geometry, Bull. Amer. Math. Soc., 70 (1964), 16–47.
[3] V. Guillemin, D. Quillen, and S. Sternberg, The classification of the irreducible complex algebras of infinite type, J. Analyse Math., 18 (1967), 107–112.
[4] S. Kobayashi, Le groupe des transformations qui laissent invariant le parallelisme, Colloque de Topologie de Strasbourg, 1954.
[5] S. Kobayashi, Transformation groups in differential geometry, Erg. Math., 70, Springer.
[6] S. Kobayashi and T. Nagano, On filtered Lie algebras and geometric structures, J. Math. Mech. I, 13 (1964), 875–908; II, 14 (1965), 513–522; III, 14 (1965), 679–706; IV, 15 (1966), 163–175; V, 15 (1966), 315–328.
[7] M. Spivak, A comprehensive introduction to differential geometry V, Publish or Perish Inc., 1975.
[8] T. Ochiai, On the automorphism group of a G-structure, J. Math. Soc. Japan, 18 (1966), 189–193.
[9] I. Singer and S. Sternberg, The infinite groups of Lie and Cartan: The transitive groups, J. Analyse Math., 15 (1965), 1–114.
[10] N. Tanaka, On the equivalence problems associated with a certain class of homogeneous spaces, J. Math. Soc. Japan, 17 (1965), 103–139.

H

192 (X.24) Harmonic Analysis

A. Fourier Transforms

Let $f(x)$ be an element of the †function space $L_1(-\infty,\infty)$ and t a real number. Then the integral

$$\hat{f}(t)=(2\pi)^{-1/2}\int_{-\infty}^{\infty} f(x)e^{-itx}dx \qquad (1)$$

converges, and the function $\hat{f}(t)$ is continuous in $(-\infty,\infty)$. We call $\hat{f}$ the †Fourier transform of f. It $f(x)\in L_2(-\infty,\infty)$, then $f(x)\in L_1(-a,a)$ for any finite interval $(-a,a)$, and if we set

$$\hat{f}(t)=(2\pi)^{-1/2}\int_{-a}^{a} f(x)e^{-itx}dx,$$

then $\hat{f}_a$ †converges in the mean of order 2 as $a\rightarrow\infty$ to a function $\hat{f}$ in L_2. In this case, we define $\hat{f}$ to be the †Fourier transform of f $(\in L_2)$. Furthermore, in this case, if we set

$$f_a(x)=(2\pi)^{-1/2}\int_{-a}^{a}\hat{f}(t)e^{itx}dt,$$

then $\mathrm{l.i.m.}_{a\rightarrow\infty}f_a(x)=f(x)$ (**Plancherel theorem**). Moreover, we have †Parseval's identity: $\int_{-\infty}^{\infty}|f(x)|^2dx=\int_{-\infty}^{\infty}|\hat{f}(t)|^2dt$ (→ 160 Fourier Transform).

Suppose that $f(x)$ is periodic with period 2π and $f\in L_2(-\pi,\pi)$. We set

$$a_n=(2\pi)^{-1}\int_{-\pi}^{\pi} f(x)e^{-inx}dx$$

(†Fourier coefficients). The nth partial sum of the †Fourier series $s_n(x)=\sum_{\nu=-n}^{n}a_\nu e^{i\nu x}$ converges in the mean of order 2 to $f(x)$, and Parseval's identity

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}|f(x)|^2dx=\sum_{n=-\infty}^{\infty}|a_n|^2$$

holds. On the other hand, if $\{a_n\}$ is a given sequence such that $\sum_{n=-\infty}^{\infty}|a_n|^2<\infty$, then $\sum_{\nu=-n}^{n}a_\nu e^{i\nu x}$ converges in the mean of order 2 to a function $f(x)$, the Fourier coefficients of $f(x)$ are $\{a_n\}$, and Parseval's identity holds (→ 159 Fourier Series).

B. Bochner's Theorem and Herglotz's Theorem

A complex-valued function $f(x)$ defined on $(-\infty,\infty)$ is said to be of **positive type** (or **positive definite**) if $\sum_{j,k=1}^{n}f(x_j-x_k)\xi_j\bar{\xi}_k\geqslant 0$ for any finite number of reals $x_1,x_2,\ldots,x_n$ and complex numbers $\xi_1,\xi_2,\ldots,\xi_n$. If $f(x)$ is measurable on $(-\infty,\infty)$ and of positive type, then there exists a †monotone increasing real-valued bounded function $\alpha(t)$ such that

$$f(x)=\int_{-\infty}^{\infty}e^{itx}d\alpha(t) \qquad (2)$$

for almost all x. If $\alpha(-\infty)=0$ and $\alpha(t)$ is right continuous, then $\alpha(t)$ is unique (**Bochner's theorem**). Conversely, if $\alpha(t)$ is nondecreasing and bounded and $f(x)$ is defined by (2), then $f(x)$ (called the **Fourier-Stieltjes transform** of $\alpha(t)$) is continuous and of positive type.

A sequence $\{a_n\}$ $(-\infty<n<\infty)$ is said to be **positive definite** (or of **positive type**) if $\sum_{j,k=1}^{n}a_{j-k}\xi_j\bar{\xi}_k\geqslant 0$ for finitely many arbitrarily chosen complex numbers $\xi_1,\xi_2,\ldots,\xi_n$. If $\{a_n\}$ is of positive type, then there exists a monotone increasing bounded function $\alpha(t)$ on $[-\pi,\pi]$ such that

$$a_n=\int_{-\pi}^{\pi}e^{int}d\alpha(t)$$

(**Herglotz's theorem**). Conversely, if $\alpha(t)$ is monotone increasing and bounded and a_n is defined by the above integral, then the sequence $\{a_n\}$ is of positive type.

C. Poisson's Summation Formula

If $f(x)\in L_1(-\infty,\infty)$ is of †bounded variation and continuous and if $\hat{f}(t)$ is its Fourier transform, then we have

$$\sqrt{a}\sum_{k=-\infty}^{\infty}f(ak)=\sqrt{b}\sum_{k=-\infty}^{\infty}\hat{f}(bk),$$

where $ab=2\pi$ $(a>0)$. This is called **Poisson's summation formula**.

D. Generalized Tauberian Theorems of Wiener

Suppose that we are given a function $f(x)\in L_1(-\infty,\infty)$ whose Fourier transform $\hat{f}(t)$ is never zero. Then the set of functions given by

$$h(x)=\int_{-\infty}^{\infty}f(x-y)g(y)dy,$$

where $g\in L_1(-\infty,\infty)$, is dense in $L_1(-\infty,\infty)$. Hence we can deduce the †generalized Tauberian theorem of Wiener: If the Fourier transform $\hat{k}_1(t)$ of $k_1(x)\in L_1(-\infty,\infty)$ does not vanish for any real t and

$$\lim_{x\rightarrow\infty}\int_{-\infty}^{\infty}k_1(x-y)f(y)dy=C\int_{-\infty}^{\infty}k_1(y)dy$$

for a function $f(x)$ that is bounded and measurable on $(-\infty,\infty)$, then for any $k_2\in L_1(-\infty,\infty)$,

$$\lim_{x\rightarrow\infty}\int_{-\infty}^{\infty}k_2(x-y)f(y)dy=C\int_{-\infty}^{\infty}k_2(y)dy$$

(→ 160 Fourier Transform G). Hence we can deduce †Tauberian theorems of the Littlewood type [3].

E. Harmonic Analysis

Let $\alpha(\lambda)$ be a complex-valued function on $(-\infty, \infty)$ that is of bounded variation and right continuous. If we have the expression

$$f(t)=\int_{-\infty}^{\infty} e^{i\lambda t}\, d\alpha(\lambda), \tag{2'}$$

then we say that the function $f(t)$ is represented by the superposition of harmonic oscillations $e^{i\lambda t}$. Conversely, when $f(t)$ is given, we have the problem of finding a function $\alpha(\lambda)$ as above such that $f(t)$ can be expressed in the form of (2'). When such a function $\alpha(\lambda)$ exists, it is also an important problem (in harmonic analysis) to find the amplitude $\alpha(\lambda)-\alpha(\lambda-0)$ of the component of a proper oscillation. Concerning this problem, we have the following three theorems:

(I) A necessary and sufficient condition for $f(x)$ to be representable in the form (2') is

$$\sup_{1\leqslant n}\int_{-\infty}^{\infty}\left|\int_{-\infty}^{\infty}\left(\frac{\sin(t/n)}{t/n}\right)^2 f(t)e^{-i\lambda t}\,dt\right|d\lambda<\infty.$$

(II) For a function $f(t)$ expressed in the form (2'), we have: (i) for any λ_0,

$$\alpha(\lambda_0)-\alpha(\lambda_0-0)=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} f(t)e^{-i\lambda_0 t}\,dt,$$

and (ii) if $\alpha(\lambda)$ is continuous at $\lambda=\lambda_0-\sigma$ and $\lambda=\lambda_0+\sigma$ $(\sigma>0)$, then

$$\alpha(\lambda_0+\sigma)-\alpha(\lambda_0-\sigma)$$
$$=\lim_{T\to\infty}\sqrt{\frac{2}{\pi}}\int_{-T}^{T}\frac{\sin\sigma t}{t}f(t)e^{-i\lambda_0 t}\,dt.$$

(III) In (2'), suppose that the discontinuity points of $\alpha(\lambda)$ are $\lambda_1, \lambda_2, \ldots$, and set $a_n=\alpha(\lambda_n)-\alpha(\lambda_n-0)$ $(n=1,2,\ldots)$; then

$$\lim_{S\to\infty}\frac{1}{S}\int_0^S f(t+s)\overline{f(s)}\,ds=\sum_{n=1}^{\infty}|a_n|^2 e^{i\lambda_n t}.$$

F. The Paley-Wiener Theorem

In formula (1), if we change the variable from t to a complex variable $\zeta=t+i\sigma$, then we have

$$F(\zeta)=(2\pi)^{-1/2}\int_{-\infty}^{\infty} f(x)e^{-i\zeta x}\,dx, \tag{3}$$

which is called the **Fourier-Laplace transform** of $f(x)$. In particular, if $f(x)$ has bounded †support, then $F(\zeta)$ is an †entire function. Concerning Fourier-Laplace transforms, Paley and Wiener proved the following theorems:

(I) A necessary and sufficient condition for an entire function $F(\zeta)$ to be the Fourier-Laplace transform of a †function of class C^∞ having its support in a finite interval $[-B, B]$ is that for any N, there exists a constant $C_N>0$ such that $|F(\zeta)|\leqslant C_N(1+|\zeta|)^{-N}e^{B|\sigma|}$ for all $\zeta=t+i\sigma$.

(II) If $g(t)\in L_2(0,\infty)$, then its one-sided †Laplace transform

$$f(z)=(2\pi)^{-1/2}\int_0^{\infty} g(t)e^{-tz}\,dt$$

satisfies: (i) $f(z)$ is holomorphic in the right half-plane $\operatorname{Re} z>0$, and (ii)

$$\sup_{x>0}\int_{-\infty}^{\infty}|f(x+iy)|^2\,dy<\infty.$$

Conversely, if $f(z)=f(x+iy)$ $(x>0)$ satisfies (i) and (ii), then the boundary function $f(iy)\in L_2(-\infty,\infty)$ exists and is such that

$$\lim_{x\downarrow 0}\int_{-\infty}^{\infty}|f(iy)-f(x+iy)|^2\,dy=0,$$

its Fourier transform in L_2

$$g(t)=\operatorname*{l.i.m.}_{N\to\infty}(2\pi)^{-1/2}\int_{-N}^{N} f(iy)e^{ity}\,dy$$

vanishes at almost all negative t, and $f(z)$ is the one-sided Laplace transform of $g(t)$.

G. Harmonic Analysis on Locally Compact Abelian Groups

The general theory of harmonic analysis on the real line was extended to a theory on locally compact Abelian groups by A. Weil, I. M. Gel'fand, D. A. Raĭkov, and others. The theory of normed rings was utilized for the development of the theory (→ 36 Banach Algebras). This theory is called harmonic **analysis on locally compact Abelian groups**, and it has been developed as described in the following sections.

H. Group Rings

Let $L_1=L_1(G)$ be the set of all integrable functions with respect to a †Haar measure on a locally compact Abelian group G. If we define the norm and multiplication in $L_1(G)$ by

$$\|f\|=\int_G |f(x)|\,dx,$$

$$f\cdot g(x)=\int_G f(xy^{-1})g(y)\,dy,$$

respectively, then L_1 has the structure of a commutative †Banach algebra. (We call $f\cdot g$ the **convolution** (or **composition product**) of f

and g.) If the topology of G is not †discrete, $L_1(G)$ does not have a unity for multiplication. Hence, adjoining a formal unity $\mathbf{1}$ to $L_1(G)$, we set $R=\{\alpha\mathbf{1}+f \mid \alpha$ is a complex number, $f\in L_1(G)\}$, and

$$\|\alpha\mathbf{1}+f\|=|\alpha|+\|f\|,$$

$$(\alpha\mathbf{1}+f)+(\beta\mathbf{1}+g)=(\alpha+\beta)\mathbf{1}+(f+g),$$

$$(\alpha\mathbf{1}+f)\cdot(\beta\mathbf{1}+g)=(\alpha\beta)\mathbf{1}+(\beta f+\alpha g+f\cdot g).$$

Then R is a commutative Banach algebra with unity. When G is discrete, $R=L_1(G)$. The Banach algebra R is called the **group algebra** of G. The group algebra R of G is †semisimple. R is algebraically isomorphic to a subalgebra of $C(\mathfrak{M})$, which is the associative algebra of all continuous functions on the compact Hausdorff space $\mathfrak{M}$ consisting of all maximal ideals in R (→ 36 Banach Algebras). By this correspondence, if $\varphi\in R$ corresponds to $\varphi(M)$, which is a function on $\mathfrak{M}$, then $\sup_{M\in\mathfrak{M}}|\varphi(M)| \leqslant \|\varphi\|$. $L_1(G)$ belongs to $\mathfrak{M}$ if and only if the group G is not discrete.

I. Fourier Transforms

According as the group G is discrete or not, we set $\mathfrak{N}=\mathfrak{M}$ or $\mathfrak{N}=\mathfrak{M}-\{L_1(G)\}$. Then there exists a one-to-one correspondence between the elements $M\in\mathfrak{N}$ and the elements χ of the †character group $\hat{G}$ of G such that the following formulas are valid:

$$f(M)=\int_G \chi(x)f(x)\,dx, \qquad f\in L_1, \tag{4}$$

$$\chi(y)=f_y(M)/f(M), \tag{5}$$

where $f_y(x)=f(xy^{-1})$ and f is a function such that $f(M)\neq 0$. This correspondence $M\leftrightarrow\chi$ gives a homeomorphism between the locally compact space $\mathfrak{N}$ and $\hat{G}$. Hence if we identify M with χ and set $f(M)=\hat{f}(\chi)$, then $\hat{f}$ is a continuous function on $\hat{G}$, called the **Fourier transform** of $f(x)$. Since $f\to f(M)$ is an algebraic isomorphism, $(fg)\hat{}(\chi)=\hat{f}(\chi)\hat{g}(\chi)$. If $\hat{f}(\chi)\equiv \hat{g}(\chi)$, then f is equal to g in $L_1(G)$. This is the **uniqueness theorem** of Fourier transforms. From it we can deduce the †maximal almost periodicity of locally compact Abelian groups.

If G is not discrete, then $L_1(G)\in\mathfrak{M}$, and $f(L_1(G))=0$ for $f\in L_1(G)$. Hence $\{\chi \mid |\hat{f}(\chi)|\geqslant\varepsilon\}$ is a compact subset of $\hat{G}$. This means that $\hat{f}$ is a continuous function vanishing at infinity on $\hat{G}$. (This is a generalization of the †Riemann-Lebesgue theorem concerning the cases $G=\mathbf{R}^1$ or $\mathbf{T}^1=\mathbf{R}/\mathbf{Z}$.) Any continuous function $u(\chi)$ on $\hat{G}$ vanishing at infinity is approximated uniformly by $\hat{f}(\chi)$, which is the Fourier transform of $f\in L_1(G)$.

J. Positive Definite Functions

A function $\varphi(x)$ defined on G is said to be **positive definite** (or of **positive type**) if the inequality $\sum_{j,k=1}^n \varphi(x_jx_k^{-1})\alpha_j\bar{\alpha}_k\geqslant 0$ holds for arbitrary elements $x_1,\ldots,x_n$ in G and arbitrary complex numbers $\alpha_1,\ldots,\alpha_n$. We denote by P_G the set of all positive definite functions on G. If $\varphi\in P_G$, then $\varphi(e)>0$ (e is the identity of G), $|\varphi(x)|\leqslant\varphi(e)$, and $\varphi(x^{-1})=\overline{\varphi(x)}$. If G is locally compact, we further assume that $\varphi\in P_G$ is measurable with respect to the Haar measure of G. Then for any $f\in L_1(G)$,

$$\iint \varphi(xy^{-1})f(y)\overline{f(y)}\,dx\,dy\geqslant 0.$$

Any $\varphi\in P_G$ is equal almost everywhere to a continuous $\varphi_1\in P_G$. (Concerning positive definite functions on locally compact groups and their relation with unitary representations → 437 Unitary Representations B.)

K. Harmonic Analysis and the Duality Theorem

If G is a locally compact Abelian group, then a function $\varphi(x)$ on G belongs to P_G if and only if there exists a nonnegative measure $\mu(\hat{G})<\infty$ on $\hat{G}$ such that $\varphi(x)=\int_{\hat{G}}\chi(x)\,d\mu(\chi)$. (When $G=\mathbf{R}^1$, this theorem is †Bochner's theorem, whereas when $G=\mathbf{Z}$, it is Herglotz's theorem.) Hence we can prove a spectral resolution of unitary representations of G: $U(x)=\int_{\hat{G}}\chi(x)\,dE(\chi)$ (a generalization of †Stone's theorem). If $f\in L_1(G)\cap P_G$, then $\hat{f}(\chi)\geqslant 0$, $\hat{f}\in L_1(\hat{G})$, and the **inversion formula** of Fourier transforms $f(x)=\int_{\hat{G}}\overline{\chi(x)}\hat{f}(\chi)\,d\chi$ holds, provided that the Haar measure on G is suitably chosen. If $f\in L_1(G)\cap L_2(G)$, then $\hat{f}\in L_2(\hat{G})$, and **Parseval's identity** $\int_G|f(x)|^2\,dx=\int_{\hat{G}}|\hat{f}(\chi)|^2\,d\chi$ holds. If we put $Uf=\hat{f}$ and $V\hat{f}=f$, then U is extendable uniquely to an isometry of $L_2(G)$ onto $L_2(G)$ and V is extendable uniquely to its inverse transformation, respectively (**Plancherel's theorem** on locally compact Abelian groups). By the inversion formula and Plancherel's theorem, we can prove that the character group $\hat{\hat{G}}$ of $\hat{G}$ is isomorphic to G as a topological group. This is called the **Pontryagin duality theorem** of locally compact Abelian groups (→ 422 Topological Abelian Groups). In particular, if G is compact, $\hat{G}$ is a discrete Abelian group. Then we can normalize the Haar measure of G and $\hat{G}$ so that the measure of G and the measure of each element of $\hat{G}$ are 1. Plancherel's theorem implies that the set of characters of G is a †complete †orthonormal set in $L_2(G)$.

L. Poisson's Summation Formula

Suppose that G is a locally compact Abelian group, H is its discrete subgroup, and G/H is compact. Then the †annihilator Γ of H is a discrete subgroup of $\hat{G}$. For any continuous function $f(x)$ on G, if $\sum_{y\in H} f(xy)$ is convergent absolutely and uniformly (hence $f\in L_1(G)$) and $\sum_{\xi\in\Gamma}\hat{f}(\xi)$ is convergent absolutely, then $\sum_{y\in H} f(y)=c\sum_{\xi\in\Gamma}\hat{f}(\xi)$, where c is a constant depending on the Haar measures of G and $\hat{G}$. This is called **Poisson's summation formula** on a locally compact Abelian group (a generalization of †Poisson's summation formula on $G=\mathbf{R}$).

M. Closed Ideals in $L_1(G)$

In the following, the group operations in G and $\hat{G}$ are denoted by $+$, and the value of the character $\gamma(x)$ $(x\in G, \gamma\in\hat{G})$ is written as (x,γ).

For a function f defined on G and any $y\in G$, the translation operator τ_y is defined by $\tau_y f(x)=f(x-y)$. A closed subspace of $L_1(G)$ is an ideal in $L_1(G)$ if and only if it is invariant under all translations (N. Wiener). A closed ideal I coincides with $L_1(G)$ if and only if the set of zeros of I, i.e., $Z(I)=\bigcap_{f\in I}\hat{f}^{-1}(0)$, is empty (Wiener's Tauberian theorem). A closed ideal I is maximal if and only if $Z(I)$ consists of a single point. If the dual $\hat{G}$ of G is discrete, then the closed ideals in $L_1(G)$ are completely characterized by the zeros; that is, †spectral synthesis is possible, but this situation does not hold generally. P. Malliavin's theorem states that if $\hat{G}$ is not discrete, then there exists a set E in $\hat{G}$ and two different closed ideals I and J such that $Z(I)=Z(J)=E$. Such a set E is called a non-S-set. For example, if $\hat{G}=\mathbf{R}^3$, the unit sphere is a non-S-set (L. Schwartz).

N. Operating Functions

Denote by $A(\hat{G})$ the set of all Fourier transforms of the functions in $L_1(G)$. Let $\hat{f}\in A(\hat{G})$ and Φ an analytic function in a neighborhood of the range of $\hat{f}$. Furthermore, assume that $\Phi(0)=0$ if G is not discrete. Then there exists a $\hat{g}\in A(\hat{G})$ such that $\hat{g}(\gamma)=\Phi(\hat{f}(\gamma))$ for $\gamma\in\hat{G}$ (Wiener-Lévy theorem). In general a function Φ defined on a set D in the complex plane is said to **operate** in a function algebra R or to be an **operating function** on R if $\Phi(f)\in R$ for all $f\in R$ whose range lies in D. The converse of the Wiener-Lévy theorem holds in the following form. Let G be an infinite Abelian group and Φ a function on the interval $[-1,1]$. If Φ operates in $A(G)$, then Φ is analytic in a neighborhood of the origin if G is compact, and analytic in a neighborhood of $[-1,1]$ if G is not compact [14, 15]. Let E be a subset of $\hat{G}$, and I_E the set of all $\hat{f}\in A(\hat{G})$ such that $\hat{f}=0$ on E. Let $A(E)=A(\hat{G})/I_E$ be the quotient algebra. The set E is called a **set of analyticity** if every operating function on $A(E)$ is analytic. For a characterization of such a set → [11].

O. Measure Algebras

For a locally compact Abelian group G, let $M(G)$ be the set of all †regular bounded complex measures. For λ and μ in $M(G)$, the convolution $\lambda*\mu$ is defined by $(\lambda*\mu)(E)=\int_G\lambda(E-y)\cdot d\mu(y)$, where E is a Borel set in G. Then $M(G)$ is a semisimple commutative Banach algebra whose product is defined to be the convolution. The Fourier-Stieltjes transform of $\mu\in M(G)$ is defined by

$$\hat{\mu}(\gamma)=\int_G (x,\gamma)\,d\mu(x),\qquad \gamma\in\hat{G}.$$

A continuous function on $\hat{G}$ is positive definite if and only if it is the Fourier-Stieltjes transform of a positive measure in $M(G)$ (Bochner's theorem). Assume that G is not discrete. A function on the interval $[-1,1]$ that operates in the Fourier-Stieltjes transforms of measures in $M(G)$ can be extended to an entire function, and a function on the whole complex plane that operates in the †Gel'fand representation of $M(G)$ is an entire function [15]. From this fact, it follows that $M(G)$ is asymmetric and nonregular. Furthermore, there exists a measure $\mu\in M(G)$ such that $\hat{\mu}(\gamma)\geqslant 1$ but $1/\hat{\mu}$ is not a Fourier-Stieltjes transform of $M(G)$. (See also Wiener and Pitt [16], Shreĭder [17], and Hewitt and Kakutani [18]; for the general description of measure algebra, see Rudin [11] and Hewitt and Ross [12].)

P. Idempotent Measures

A measure $\mu\in M(G)$ is called **idempotent** if $\mu*\mu=\mu$, that is, $\hat{\mu}(\gamma)=0$ or 1 for all $\gamma\in\hat{G}$. Then $\hat{\mu}$ is the characteristic function of the set $\{\gamma\in\hat{G}\mid\hat{\mu}(\gamma)=1\}$. The smallest ring of subsets of $\hat{G}$ that contains all open cosets of subgroups of $\hat{G}$ is called the coset ring of $\hat{G}$. The characteristic function of a set E in $\hat{G}$ is the Fourier-Stieltjes transform of an idempotent measure in $M(G)$ if and only if E belongs to the coset ring of $\hat{G}$ [19]. A simple proof of this theorem is given by T. Ito and I. Amemiya (*Bull. Amer. Math. Soc.*, 70 (1964)). When G is the unit circle, the coset ring consists of sequences periodic except at a finite number of points, and for this case the theorem was obtained by H. Helson. Let

$n_1, n_2, \ldots, n_k$ be distinct integers and $d\mu(x) = \sum_{j=1}^{k} e^{in_jx} \cdot dx$. Then μ is an idempotent measure on the unit circle. J. E. Littlewood conjectured that the norm of μ exceeds $c\log k$, where c is a positive constant not depending on the choice of $\{n_j\}$. A partial answer was given by P. J. Cohen [19] for compact connected Abelian groups and was improved by H. Davenport (*Mathematika*, 7 (1960)) and E. Hewitt and H. S. Zuckerman (*Proc. Amer. Math. Soc.*, 14 (1963)).

Q. Mappings of Group Algebras

Let G and H be two locally compact Abelian groups and φ a nontrivial homomorphism of $L_1(G)$ into $M(H)$. Associated with φ there is a mapping φ^* of a subset Y of $\hat{H}$ into $\hat{G}$ such that $\varphi(f)(\gamma) = \hat{f}(\varphi^*(\gamma))$ for $\gamma \in Y$ and $=0$ for $\gamma \notin Y$, or symbolically $\varphi(f) = \hat{f}(\varphi^*)$. A continuous mapping α of Y into $\hat{G}$ is said to be piecewise affine if there exist a finite number of mutually disjoint sets $S_j, j=1, \ldots, n$, in the coset ring of $\hat{H}$ and mappings α_j such that (i) $Y = \bigcup_{j=1}^{n} S_j$; (ii) α_j is defined on an open coset K_j of $\hat{H}$, where $K_j \supset S_j$; (iii) $\alpha_j = \alpha$ on S_j; and (iv) $\alpha_j(\gamma + \gamma' - \gamma'') = \alpha_j(\gamma) + \alpha_j(\gamma') - \alpha_j(\gamma'')$ for all γ, $\gamma', \gamma'' \in K_j, j=1, \ldots, n$. P. J. Cohen's theorem is: If φ is a homomorphism of $L_1(G)$ into $M(H)$, then Y belongs to the coset ring of $\hat{H}$ and φ^* is a piecewise affine mapping of Y into $\hat{G}$. Conversely, for any piecewise affine mapping α, there is a homomorphism φ of $L_1(G)$ into $M(H)$ such that $\varphi^* = \alpha$. Related theorems have been studied by A. Beurling, H. Helson, J.-P. Kahane, Z. L. Leibenson, and W. Rudin [11].

R. Exceptional Sets

Let G be a locally compact Abelian group. A subset E is said to be independent if $n_1x_1 + \ldots + n_kx_k = 0$, where the n_j are integers and $x_j \in E$ implies $n_jx_j = 0, j=1, \ldots, k$. A set E in G is called a **Kronecker set** if for every continuous function φ on E of absolute value 1 and $\varepsilon > 0$ there exists a $\gamma \in \hat{G}$ such that $|\varphi(x) - (x, \gamma)| < \varepsilon$, $x \in E$. Every Kronecker set is independent and of infinite order, but independent sets are not necessarily Kronecker sets. For a group G whose elements are of finite order p, a set E is called of type K_p if for every continuous function φ on E with values $\exp(2\pi ik/p)$, $k=0, \ldots, p-1$, there is a $\gamma \in \hat{G}$ such that $\varphi = \gamma$ on E. If E is a compact Kronecker set in G and μ is a measure with support in E, i.e., $\mu \in M(E)$, then $\|\mu\| = \|\hat{\mu}\|_\infty$. A compact set E is called a **Helson set** if there is a constant C such that $\|\mu\| \leqslant C\|\hat{\mu}\|_\infty$ for $\mu \in M(E)$. Every K_p set is also a Helson set. For a Helson set E, $C(E) = A(E)$. A discrete analog of a Helson set is a Sidon set. A subset F of a discrete group $\hat{G}$ is called a **Sidon set** if there is a constant C such that $\sum_{\gamma \in F} |a_\gamma| \leqslant C\sup_x |\sum_{\gamma \in F} a_\gamma(x, \gamma)|$ for every polynomial $\sum a_\gamma(x, \gamma)$. For example, a †lacunary sequence $\{n_k\}$, $n_{k+1}/n_k > q > 1$, of integers is a Sidon set. These sets are deeply connected with harmonic analysis on groups, and measures concentrated on these sets have some unexpected pathological properties (→ e.g., [11, 13]).

S. Tensor Algebras and Group Algebras

Let X and Y be compact Hausdorff spaces, and denote by $V(X, Y)$ the projective tensor product $C(X) \hat{\otimes} C(Y)$ of continuous function spaces $C(X)$ and $C(Y)$. The norm of $\varphi(x, y) = \sum_{j=1}^{\infty} f_j(x)g_j(y)$ is defined by $\|\varphi\| = \inf \sum_{j=1}^{\infty} \|f_j\|_\infty \|g_j\|_\infty$, where the infimum is taken for all expressions of φ. If $\hat{G}$ is an infinite compact group, then there exist two subsets K_1 and K_2 such that (i) K_1 and K_2 are homeomorphic to the †Cantor ternary set; (ii) the expression $\gamma_1 + \gamma_2$ of an element of $E = K_1 + K_2$ is unique, where $\gamma_1 \in K_1$ and $\gamma_2 \in K_2$; (iii) $K_1 \cap K_2 \neq \varnothing$; and (iv) $K_1 \cup K_2$ is a Kronecker set or a set of type K_p for some p. Varopoulos' theorem states that the algebra $V(K_1, K_2)$ is isomorphic to $A(E)$ which denotes the algebra of restriction of functions in $A(\hat{G})$ on the set E. By this theorem, the problems of spectral synthesis and operating functions of group algebras are transformed into problems of tensor algebras. For a more precise discussion → [20].

References

[1] E. C. Titchmarsh, Introduction to the theory of Fourier integrals, Clarendon Press, 1937.
[2] A. Zygmund, Trigonometric series I, Cambridge Univ. Press, second edition, 1959.
[3] N. Wiener, The Fourier integrals and certain of its applications, Cambridge Univ. Press, 1933.
[4] R. E. A. C. Paley and N. Wiener, Fourier transforms in the complex domain, Amer. Math. Soc. Colloq. Publ., 1934.
[5] K. Yosida, Functional analysis, Springer, 1965, sixth edition, 1980.
[6] M. A. Naĭmark, Normed rings, Noordhoff, 1959. (Original in Russian, 1956.)
[7] L. H. Loomis, An introduction to abstract harmonic analysis, Van Nostrand, 1953.
[8] H. Cartan and R. Godement, Théorie de la dualité et analyse harmonique dans les groupes abéliens localement compacts, Ann. Sci. Ecole Norm. Sup., (3) 64 (1947), 79–99.

[9] R. Godement, Théorèmes taubériens et théorie spectrale, Ann. Sci. Ecole Norm. Sup., (3) 64 (1947), 118–138.
[10] A. Weil, L'intégration dans les groupes topologiques et ses applications, Actualités Sci. Ind., Hermann, 1940, second edition, 1951.
[11] W. Rudin, Fourier analysis on groups, Interscience, 1962.
[12] E. Hewitt and K. A. Ross, Abstract harmonic analysis, Springer, I, 1963; II, 1970.
[13] J.-P. Kahane and R. Salem, Ensembles parfaits et séries trigonométriques, Actualités Sci. Ind., Hermann, 1963.
[14] Y. Katznelson, Sur le calcul symbolique dans quelques algèbres de Banach, Ann. Sci. Ecole Norm. Sup., (3) 76 (1959), 83–123.
[15] H. Helson, J.-P. Kahane, Y. Katznelson, and W. Rudin, The functions which operate on Fourier transforms, Acta Math., 102 (1959), 135–157.
[16] N. Wiener and H. R. Pitt, On absolutely convergent Fourier-Stieltjes transforms, Duke Math. J., 4 (1938), 420–436.
[17] Yu. A. Shreĭder (Šreĭder), The structure of maximal ideals in rings of measures with convolution, Amer. Math. Soc. Transl., 81 (1953). (Original in Russian, 1950.)
[18] E. Hewitt and S. Kakutani, Some multiplicative linear functionals on $M(G)$, Ann. Math., (2) 79 (1964), 489–505.
[19] P. J. Cohen, On a conjecture of Littlewood and idempotent measures, Amer. J. Math., 82 (1960), 191–212.
[20] N. T. Varopoulos, Tensor algebras and harmonic analysis, Acta Math., 119 (1967), 51–112.

193 (X.29) Harmonic Functions and Subharmonic Functions

A. General Remarks

A real-valued function u of †class C^2 defined in a domain D in the n-dimensional Euclidean space $\mathbf{R}^n$ is called **harmonic** if it satisfies the Laplace equation

$$\Delta u(P)=\frac{\partial^2 u}{\partial x_1^2}+\ldots+\frac{\partial^2 u}{\partial x_n^2}=0 \quad (P=(x_1,\ldots,x_n))$$

in D. A harmonic function is, by definition, twice continuously differentiable, but turns out to be real analytic. It is not true, however, that the solutions of the Laplace equation are real analytic. For example, for the function $u(x,y)=\operatorname{Re}\exp(-z^{-4})$ $(z=x+iy\neq 0)$, $u(0,0)=0$, u_{xx} and u_{yy} exist everywhere, and u satisfies the Laplace equation, but is not continuous at the origin.

The fundamental properties of harmonic functions do not depend essentially on n.

A real-valued function u of class C^2 satisfying the inequality $\Delta u\geqslant 0$ is called **subharmonic**. For a more general definition of subharmonic functions and their properties → Sections P–U.

B. Invariance of Harmonicity

Harmonicity in $\mathbf{R}^2$ is invariant under any †conformal transformation. Namely, when there exists a conformal bijection sending a domain D in the xy-plane onto a domain D' in the $\xi\eta$-plane, every harmonic function $u(x,y)$ on D is transformed into a harmonic function of (ξ,η) on D'. In $\mathbf{R}^n$ for $n\geqslant 3$, harmonicity is not generally preserved under conformal transformations. However, harmonicity is preserved in the following special case: Let D be a domain in $\mathbf{R}^n$ $(n\geqslant 3)$, and consider the **inversion** $f:D\to D'$ defined by $f(x_1,\ldots,x_n)=(x_1',\ldots,x_n')=(a^2x_1/r^2,\ldots,a^2x_n/r^2)$, $r=(x_1^2+\ldots+x_n^2)^{1/2}$. Let $u(x_1,\ldots,x_n)$ be a harmonic function on D, and let $v(x_1',\ldots,x_n')$ be the function on D' obtained by applying the **Kelvin transformation** to u. Namely, $v(x_1',\ldots,x_n')=(a/r')^{n-2}u(a^2x_1'/r'^2,\ldots,a^2x_n'/r'^2)$, where $r'^2=x_1'^2+\ldots+x_n'^2$. Then the function v is harmonic on D'. A function u that is harmonic outside a compact set is called **regular at the point at infinity** if any Kelvin transform of u is harmonic in a neighborhood of the origin, in which case $u(P)\to 0$ as $\overline{OP}\to\infty$. Now let T: $x_k=x_k(x_1',\ldots,x_n')$, $1\leqslant k\leqslant n$, be a one-to-one analytic transformation of a domain D' onto another domain D. If there exists a positive function $\varphi(x_1',\ldots,x_n')$ in D' such that $\varphi(x_1',\ldots,x_n')u(x_1(x_1',\ldots,x_n'),\ldots,x_n(x_1',\ldots,x_n'))$ is harmonic for any harmonic function $u(x_1,\ldots,x_n)$ in D, then T is conformal. A conformal transformation as it is known in differential geometry is either (i) a †similarity transformation, (ii) an inversion with respect to a sphere or a plane, or (iii) a finite combination of transformations of types (i) and (ii).

C. Examples of Harmonic Functions

(1) If u is a polynomial in $x_1,\ldots,x_n$ and harmonic in $\mathbf{R}^n$, then the terms of degree k in u form a harmonic function for each $k\geqslant 0$. A harmonic homogeneous polynomial is said to be a **spherical harmonic**. (2) $\log r$ in $\mathbf{R}^2$ and r^{2-n} in $\mathbf{R}^n$ $(n\geqslant 3)$ are harmonic except at $r=0$. (3) Every †logarithmic potential in $\mathbf{R}^2$ and every †Newtonian potential in $\mathbf{R}^n$ $(n\geqslant 3)$ is harmonic outside the †support of the measure. Con-

versely, any harmonic function defined on a domain D is represented in an arbitrary relatively compact domain D' in D as the sum of a logarithmic ($n=2$) or Newtonian ($n \geqslant 3$) potential of a measure on $\partial D'$ and the potential of a †double layer. (4) Both the real part u and the imaginary part v of an analytic function of a complex variable are harmonic. We call v a **conjugate harmonic function** of u. If u is harmonic on a †simply connected domain D, then the conjugate v of u is given by

$$v(x, y)=\int_{(a,b)}^{(x,y)}\left(-\frac{\partial u}{\partial y}dx+\frac{\partial u}{\partial x}dy\right)+\text{constant},$$

where (a, b) is a fixed point in D and the path of integration is contained in D. When D is a †multiply connected domain, v may take many values in accordance with the †homology classes of the paths of integration.

D. Green's Formulas

In the following one should substitute "curve" for the term "surface" when $n=2$. Let D be a bounded domain whose boundary S consists of a finite number of closed surfaces that are piecewise of class C^1. Let u and v be harmonic in D, and suppose that all the first-order partial derivatives of u and v have finite limits at every boundary point. We call D the inside of S. Let n be a normal on S toward the outside of D. Then the relation

$$\int_S\left(u\frac{\partial v}{\partial n}-v\frac{\partial u}{\partial n}\right)d\sigma=0 \qquad (1)$$

follows immediately from †Gauss's formula, where $d\sigma$ is the †surface element on S. In particular, when v is identically equal to 1, formula (1) gives

$$\int_S\frac{\partial u}{\partial n}d\sigma=0. \qquad (2)$$

Equations (1) and (2) are called **Green's** and **Gauss's formulas**, respectively. Conversely, u is harmonic in D if u is a function of class C^2 in D and at every point $p\in D$, there is a sequence $\{r_k\}$ decreasing to zero and such that $\int_{S(P,r_k)}(\partial u/\partial n)\,d\sigma=0$ $(k=1,2,\ldots)$, where $S(P,r_k)$ is the spherical surface with center at P and radius r_k. Another sufficient condition for u to be harmonic is that u is of class C^1 and, at every point $P\in D$, there be an $r_P>0$ such that $\int_{S(P,r)}(\partial u/\partial n)\,d\sigma=0$ for every r, $0<r<r_P$ (Koebe, Bocher).

E. Mean Value Theorems

We assume that u is a harmonic function, D the domain of definition of u, and S the boundary of D. The mean value of u on the surface or the interior of any ball in D is equal to the value of u at the center of the ball. Namely,

$$u(P)=\frac{1}{\tau_n r^n}\int_{B(P,r)}u\,d\tau=\frac{1}{\sigma_n r^{n-1}}\int_{S(P,r)}u\,d\sigma,$$

where τ_n and σ_n are the volume and surface area of a unit ball in $\mathbf{R}^n$, respectively, $B(P,r)$ is the open ball with center at P and radius r, and $d\tau$ is the volume element. These relations are called **mean value theorems**. Conversely, if v is continuous in D and at every point $P\in D$, there is a sequence $\{r_k\}$ decreasing to zero and such that the mean value of v over $B(P,r_k)$ or $S(P,r_k)$ is equal to $v(P)$ for each k, then v is harmonic in D. This result is called **Koebe's theorem**. From the mean value theorems the **maximum principle** for harmonic functions follows: Any nonconstant u assumes neither maximum nor minimum in D. If both u and v are harmonic in D and have the same finite boundary value at every point on S, then $u\equiv v$ in D by the maximum principle. This is called the **uniqueness theorem**.

F. Boundary Value Problems

The **first boundary value problem** (or **Dirichlet problem**) is the problem of finding a harmonic function defined on D that assumes boundary values prescribed on S (→ 120 Dirichlet Problem). The **second boundary value problem** (or **Neumann problem**) is the problem of finding a harmonic function u whose normal derivative $\partial u/\partial n$ is equal to a function f prescribed on the piecewise smooth boundary S. The solution, if it exists, is uniquely determined up to an additive constant. In order for the solution to exist, f should satisfy the condition $\int_S f\,d\sigma=0$. The **third boundary value problem** is the problem of finding a harmonic function u on D that satisfies $\partial u/\partial n=hu+f$ on S, where h and f are functions prescribed on S. All these problems can be reduced to certain †Fredholm integral equations. There is also the boundary value problem of mixed type, in which the boundary values are prescribed in a part of S and the normal derivatives are prescribed on the rest.

G. The Poisson Integral

Let D be a bounded domain with smooth boundary S and u a function harmonic in D and continuous on $D\cup S$. Let $G(P,Q)$ be Green's function in D. Then (1) yields

$$u(P)=-\frac{1}{\sigma_n}\int_S u(Q)\frac{\partial G(P,Q)}{\partial n_Q}d\sigma(Q).$$

In particular, if $D=B(0,r)$, then

$$u(P)=\frac{r^2-\overline{OP}^2}{\sigma_n r}\int_{S(0,r)}\frac{u(Q)}{\overline{PQ}^n}d\sigma(Q).$$

Conversely, given an integrable function f on $S(0,r)$, we set

$$u(P)=\frac{r^2-\overline{OP}^2}{\sigma_n r}\int_{S(0,r)}\frac{f(Q)}{\overline{PQ}^n}d\sigma(Q).$$

Then $u(p)$ is harmonic in $B(0,r)$ and converges to $f(Q)$ as P tends to any point Q on $S(0,r)$ where f is continuous. We call u a **Poisson integral**. Sometimes it is possible to represent a harmonic function u in $B(0,r)$ in the following form, which is more general than the Poisson integral:

$$u(P)=\frac{r^2-\overline{OP}^2}{\sigma_n r}\int_{S(0,r)}\frac{1}{\overline{PQ}^n}d\alpha(Q), \qquad (3)$$

where α is a signed †Radon measure on $S(0,r)$. In order for u to admit such a representation, it is necessary and sufficient that $\int_{S(0,r')}|u|\,d\sigma$ be a bounded function of r' for $0<r'<r$, or equivalently, that the †subharmonic function $|u|$ have a †harmonic majorant. Furthermore, if α is absolutely continuous, then the Poisson integral representation of u is possible, and vice versa. A necessary and sufficient condition for the function u to admit the Poisson integral representation is that there exist a positive convex function $\varphi(t)$ on $t>0$ such that $\varphi(t)/t\to\infty$ as $t\to\infty$ and $\varphi(|u|)$ has a harmonic majorant.

When D is a general domain in which Green's function exists, every positive harmonic function $u(P)$ is represented uniquely as the integral $\int K(P,Q)\,d\mu(Q)$, where $K(P,Q)$ is a †Martin kernel and μ is a Radon measure on the Martin boundary B whose support is contained in a certain essential part of B, each point of which is called a minimal point. A similar integral representation appears in the theory of Markov processes (→ 260 Markov Chains I). The representation $\int K(P,Q)\,d\mu(Q)$ is a generalization of (3). In terms of a function similar to $\varphi(t)$, we can give a necessary and sufficient condition for $u(P)$ to be represented in the form $\int K(P,Q)f(Q)\,dv(Q)$, which corresponds to the Poisson integral representation and in which v is determined by $1=\int K(P,Q)\,dv(Q)$. This condition is equivalent to the condition that $u(P)$ be **quasibounded**, i.e., that there exist an increasing sequence of bounded harmonic functions that converges to u [7].

A positive harmonic function u is said to be **singular** if any nonnegative harmonic minorant of u vanishes identically. Every positive harmonic function can be expressed as the sum of a quasibounded harmonic function and a singular one.

H. Expansion

Let $P_0=(x_1^0,\dots,x_n^0)$ be a point in D, and denote the distance from P_0 to S by r. Then a harmonic function u is expanded uniquely into a power series

$$\sum a_{k_1,\dots,k_n}(x_1-x_1^0)^{k_1}\dots(x_n-x_n^0)^{k_n}, \qquad k_1,\dots,k_n\geqslant 0,$$

in $B(P_0,(\sqrt{2}-1)r)$. Thus u is (real) analytic in D. If u vanishes on an open set in D, then $u\equiv 0$ in D. If the power series is written as $\sum_k h_k$ with spherical harmonics h_k of degree $k=1,2,\dots$, then this series converges over all of $B(P_0,r)$.

I. Sequences of Harmonic Functions

In this section, $\{u_m\}$ is a sequence of harmonic functions in a bounded domain D. First, if each u_m is bounded and continuous on $D\cup S$ and $\{u_m\}$ converges uniformly on S, then $\{u_m\}$ converges uniformly in D, and the limiting function u is harmonic in D. Moreover, $\partial^{k_1+\dots+k_n}u_m/\partial x_1^{k_1}\dots\partial x_n^{k_n}$ converges to $\partial^{k_1+\dots+k_n}u/\partial x_1^{k_1}\dots\partial x_n^{k_n}$ uniformly on any compact subset of D (**Harnack's first theorem**). Second, if $u_1\leqslant u_2\leqslant\dots$ in D and there is a point of D at which $\{u_m\}$ is bounded, then $\{u_m\}$ converges uniformly on any compact subset of D (**Harnack's second theorem**). The following **Harnack's lemma** is useful: If u is positive harmonic in D, P_0 is a point of D, and K is a compact subset of D, then there exist positive constants c and c', depending only on P_0 and K, such that $cu(P_0)\leqslant u(P)\leqslant c'u(P_0)$ on K.

Any family of (locally) uniformly bounded harmonic functions is †normal. A family of positive harmonic functions that is bounded at a point is also normal by Harnack's lemma. If $\int_D|u_k-u_m|^p\,d\tau\to 0$ as k, $m\to\infty$ for $p\geqslant 1$, then †Hölder's inequality implies that $\{u_m\}$ converges uniformly on any compact subset of D. It follows that if $\int_D|\mathrm{grad}(u_k-u_m)|^p\,d\tau\to 0$ as k, $m\to\infty$ and $\{u_n\}$ converges at a point in D, then there exists a harmonic function u in D such that $\int_D|\mathrm{grad}(u_m-u)|^p\,d\tau\to 0$ as $m\to\infty$ and u_m converges to u uniformly on any compact subset of D, where P_0 is any point in D. Finally, if $\int_D|u_n|^p\,d\tau$ $(p\geqslant 1)$ are bounded, then $\{u_n\}$ forms a normal family.

J. Level Surfaces and Orthogonal Trajectories

The set $\{P\mid u(P)=\text{constant}\}$ is called a **level surface (niveau** or **equipotential surface)**. When a is given as the constant, the level surface is called the a-level surface. Assume that u is not a constant. A point where $\mathrm{grad}\,u$ vanishes is called **critical**. The set of critical points consists of at most countably many †real analytic

manifolds of dimension $\leqslant n-2$ ($n=\dim D$, and a manifold of dimension 0 is understood to be a point). Any compact subset of D intersects only a finite number of such manifolds; we express this fact by saying that the manifolds do not cluster in D. Each of these manifolds is contained in a certain level surface. The complement of the critical points with respect to any level surface consists of real analytic manifolds of dimension $n-1$ that do not cluster in D.

For each noncritical point there exists an analytic curve passing through it such that $\operatorname{grad} u$ is parallel to the tangent to the curve at each point on the curve. A maximal curve with this property is called an **orthogonal trajectory** (or **line of force**). Along every orthogonal trajectory, u increases strictly in one direction and hence decreases in the other, so that none of the orthogonal trajectories is a closed curve. There is exactly one orthogonal trajectory passing through any noncritical point. Therefore no two orthogonal trajectories intersect, and no orthogonal trajectory terminates at a noncritical point. Moreover, the set of limit points of any orthogonal trajectory in each direction does not contain any noncritical point. When u is a Green's function $G(P,Q)$, every orthogonal trajectory is called a **Green line**, and a Green line that originates at the pole Q and along which u decreases to 0 is called **regular**. For any sufficiently large a, the a-level surface Σ_a is an analytic closed surface homeomorphic to a spherical surface. Let E be a family of orthogonal trajectories originating at the pole. If the intersection A of E and a closed level surface Σ_a is an $(n-1)$-dimensional measurable set, then the †harmonic measure of A at Q with respect to the interior of Σ_a is called the **Green measure** of E. M. Brelot and G. Choquet proved that all orthogonal trajectories originating at the pole except those belonging to a family of Green measure zero are regular. Consider a domain D bounded by two compact sets, and denote by u the harmonic measure of one compact set with respect to D. Assume that u is not a constant. Then u changes from 0 to 1 along all orthogonal trajectories except those belonging to a family that is small with respect to a measure similar to the Green measure (see "flux" defined in Section K).

K. Harmonic Flows

Denote by Σ_a the a-level surface for a harmonic function u, and by Σ_a^0 the complement of the set of critical points with respect to Σ_a. Let σ be an $(n-1)$-dimensional domain in Σ_a^0 such that the $(n-2)$-dimensional boundary of σ is piecewise of class C^2. Suppose that u assumes the value b ($>a$) on each orthogonal trajectory passing through σ. Consider the union of orthogonal trajectories that pass through σ. The subset of this union on which u assumes values between a and b forms a set called a **regular tube**. The parts of the boundary corresponding to a and b are called the lower and upper bases of the tube; accordingly, σ is the lower base. The integral $\int(\partial u/\partial n)d\sigma$ on any section (i.e., the part of a level surface in the tube) is constant and is called the **flux** of the tube. The family of orthogonal trajectories passing through an $(n-1)$-dimensional domain (not necessarily bounded by a smooth boundary) in Σ_a^0 is called a **harmonic flow**, and a subfamily is called a harmonic subflow if its intersection A with Σ_a^0 is measurable (in the $(n-1)$-dimensional sense). The flux of a harmonic subflow is defined to be $\int_A(\partial u/\partial n)d\sigma$. Then the Green measure of family E of Green lines originating at the pole is equal to the flux of E divided by σ_n. We can compute the exact value of the †extremal length of any harmonic subflow.

L. Isolated Singularities

Let u be harmonic in an open ball except at the center O. It is expressed as the sum of a function $h(P)$ harmonic in the entire ball and $\sum_{m=0}^{\infty} H_m(P)/\overline{OP}^{2m+1}$, where H_m is a †spherical harmonic of degree m. If $\overline{OP}^{\alpha}u(P)\to 0$ as $P\to O$ for $\alpha>0$, then $u(P)$ is equal to $h(P)+c/\overline{OP}+\ldots+H_m(P)/\overline{OP}^{2m+1}$ with $m<\alpha-1$. In particular, if u is bounded in a neighborhood of O, then O is a **removable singularity** for u. If u is bounded above (below), then $u(P)=h(P)+c/\overline{OP}^{n-2}$, where $c\leqslant 0$ ($c\geqslant 0$). (When $n=2$, we have $u(p)=h(p)+c\log 1/\overline{OP}$. For the removability of a set of capacity zero → 169 Function-Theoretic Null Sets).

If u is harmonic near the point at infinity, i.e., outside some closed ball, then

$$u(P)=\sum_{m=0}^{\infty}\frac{H_m(P)}{\overline{OP}^{2m+1}}+\sum_{m=0}^{\infty}U_m(P),$$

where the first sum is regular at the point at infinity and U_m is a spherical harmonic of degree m. If $\overline{OP}^{-\alpha}u(P)\to 0$ as $\overline{OP}\to\infty$ with $\alpha\geqslant 0$, then $U_m\equiv 0$ for all $m>\alpha$. If u is bounded above or below, then $U_m\equiv 0$ for all $m\geqslant 1$. If u is harmonic in $\mathbf{R}^n$ and $\overline{OP}^{-\alpha}u(P)\to 0$ as $\overline{OP}\to\infty$ with $\alpha\geqslant 0$, then u is a polynomial of degree m ($<\alpha$). If u is harmonic and bounded above or below in $\mathbf{R}^n$, then u is constant. Brelot called a function u harmonic at the point at infinity if

$$u(P)=\text{constant}+\sum_{m=1}^{\infty}\frac{H_m(P)}{\overline{OP}^{2m+1}}$$

(note that $m\geqslant 1$) near the point at infinity [1].

M. Harmonic Continuation

If u vanishes in a subdomain of D, then $u \equiv 0$ in D. If u_k is harmonic in D_k $(k=1,2)$, $D_1 \cap D_2 \neq \varnothing$, and $u_1 \equiv u_2$ in $D_1 \cap D_2$, then u_1 and u_2 define a harmonic function in $D_1 \cup D_2$. If the boundaries of mutually disjoint domains D_1 and D_2 have a surface S_0 of class C^1 in common, u_k is harmonic in D_k $(k=1,2)$, $u_1=u_2$ on S_0, and $\partial u_1/\partial n$ and $-\partial u_2/\partial n$ exist and coincide on S_0, then u_1 and u_2 define a harmonic function in the domain $D_1 \cup S_0 \cup D_2$. We express this fact by saying that one of u_1 and u_2 is a **harmonic continuation** of the other. It follows that $u \equiv 0$ in D if the boundary of D contains a surface S_0 of class C^1 and $u = \partial u/\partial n = 0$ on S_0. Consider the case $n=2$. If the boundary of a Jordan domain contains an analytic arc C and u (or $\partial u/\partial n$) vanishes on C, then a harmonic continuation of u into a certain domain beyond C is possible. If $n=3$, however, nothing is known except in the case where S_0 is a part of a spherical surface or a plane and $u=0$ (or $\partial u/\partial n=0$) on S_0.

Boundary values of u do not always exist, but in some special cases, u has limits. For instance, a positive harmonic function in a ball has a finite limit at almost every boundary point Q if the variable is restricted to any angular domain with vertex at Q.

N. Green Spaces

As a generalization of †Riemann surfaces, Brelot and Choquet introduced $\mathscr{E}$**-spaces** [3]. It is required that $\mathscr{E}$ be a separable connected topological space and satisfy the following two conditions: (i) At each point P there exists a neighborhood V_P of P and a homeomorphism between V_P and an open set V'_P in the †Alexandrov compactification $\mathbf{R}^n \cup \{\infty\}$; (ii) if $A = V_{P_1} \cap V_{P_2} \neq \varnothing$ and A'_k is the part of V'_{P_k} that corresponds to A $(k=1,2)$, then the correspondence between A'_1 and A'_2 via A is a conformal (possibly with the sense of angles reversed) transformation when $n=2$ and an †isometric transformation (which keeps ∞ invariant) when $n \geqslant 3$. If a †Green's function exists on $\mathscr{E}$, then $\mathscr{E}$ is called a **Green space**. Harmonic functions and the †Dirichlet problem on a Green space have been discussed from various points of view.

O. Biharmonic Functions

A function v is called **polyharmonic** if $\Delta^k v=0$ $(k \geqslant 2)$ and **biharmonic** if $\Delta\Delta v=0$; sometimes, polyharmonic functions are also called biharmonic. A biharmonic function in a plane domain D is written as $\operatorname{Re}(\bar{z}f(z)+g(z))$, where f and g are complex analytic in D (Goursat's representation). Biharmonic functions are used in the theory of elasticity and hydrodynamics.

P. Subharmonic Functions

Let D be a †domain in the n-dimensional Euclidean space $\mathbf{R}^n$ $(n \geqslant 2)$. A real-valued function $u(P)$ in D is called **subharmonic** if (1) $-\infty \leqslant u < +\infty$, $u \not\equiv -\infty$; (2) u is †upper semicontinuous; and (3) at every point P_0 of D, the mean value of u over the surface of any closed †ball in D with center at P_0 is not smaller than $u(P_0)$, i.e.,

$$u(P_0) \leqslant \frac{1}{\sigma_n r^{n-1}} \int u \, d\sigma \equiv L(P_0, r),$$

where σ_n is the area of the surface of a unit ball in $\mathbf{R}^n$. Condition (3) can be replaced by: (3') The mean value $A(P_0, r)$ of u over the closed ball is $\geqslant u(P_0)$. In order that an upper semicontinuous function u be subharmonic it is necessary and sufficient that, for any subdomain D' of D and for any harmonic function h in D', the maximum principle hold for $u-h$.

We call $-u$ **superharmonic** when u is subharmonic. A harmonic function is subharmonic and superharmonic. The converse is also true (→ Section E).

When u is of †class C^2, then u is subharmonic if and only if

$$\Delta u = \frac{\partial^2 u}{\partial x_1^2} + \ldots + \frac{\partial^2 u}{\partial x_n^2} \geqslant 0 \qquad (P=(x_1, \ldots, x_n)).$$

When u is an upper semicontinuous function that is not necessarily differentiable, u is subharmonic if and only if Δu interpreted as a †distribution is a positive †measure.

If $u_1, \ldots, u_k$ are subharmonic and $a_1, \ldots, a_k$ are positive constants, then $a_1 u_1 + \ldots + a_k u_k$ and $\max(u_1(P), u_2(P), \ldots, u_k(P))$ are subharmonic. If a subharmonic function u is replaced by the †Poisson integral for the boundary function u inside a closed ball in D, then the resulting function in D is subharmonic. If $f(t)$ is a monotone increasing convex function, then $f(u)$ is subharmonic. If $v \geqslant 0$ and $\log v$ is subharmonic, then v is subharmonic. If $f(z)$ is a holomorphic function of the complex variable z and $\lambda > 0$, then $\lambda \log|f(z)|$ and hence $|f(z)|^\lambda$ are subharmonic. If h is harmonic, then $|h|$ is subharmonic. Any †logarithmic potential $(n=2)$ or †Newtonian potential $(n \geqslant 3)$ is superharmonic in $\mathbf{R}^n$.

Q. Properties of Mean Values

Condition (3) (resp. (3')) can be replaced by the condition that there exists an $r(P_0)>0$ at any

P_0 such that $u(P_0) \leqslant L(P_0, r)(u(P_0) \leqslant A(P_0, r))$ for every r, $0 < r < r(P_0)$. The relation $-\infty < A(P_0, r) \leqslant L(P_0, r)$ always holds, and both $A(P_0, r)$ and $L(P_0, r)$ decrease to $u(P_0)$ as $r \downarrow 0$. On any †compact subset of D, u is †integrable. Both $A(P_0, r)$ and $L(P_0, r)$ increase with r and are convex functions of $-\log r$ ($n=2$) and r^{2-n} ($n \geqslant 3$); hence they are continuous functions of r. If D' is relatively compact in D and r_0 is the distance between $\partial D'$ and ∂D, then $A(P, r)$ is a continuous subharmonic function of P in D', where r is fixed in the interval $(0, r_0)$. By taking the average of $A(P, r)$ k times, a subharmonic function of class C^k is obtained that decreases to u as $r \downarrow 0$. If $\varphi_r((x_1^2 + \ldots + x_n^2)^{1/2})$ is suitably chosen, then the †convolution $u * \varphi_r$ is a subharmonic function of class C^∞ and decreases to u as $r \downarrow 0$.

R. Sequences of Subharmonic Functions

The limit of a decreasing sequence or a downward-directed net of subharmonic functions is subharmonic or equal constantly to $-\infty$. The limit of a uniformly convergent sequence of subharmonic functions is subharmonic. If $u_1, u_2, \ldots$ are subharmonic, then $\max(u_1, \ldots, u_k)$ is subharmonic for every k, but $\sup(u_1, u_2, \ldots)$ may not be subharmonic. Let U be a family of subharmonic functions in D that are uniformly bounded above on every compact subset of D. Then the **upper envelope** of U, i.e., the function defined by $\sup_{u \in U} u$ in D, coincides with a subharmonic function except on a set of †capacity zero.

S. Harmonic Majorants and Riesz Decompositions

Suppose that we are given a subharmonic function u in D. If there is a harmonic function h satisfying $h \geqslant u$ in D, then h is called a **harmonic majorant** of u. When there is a harmonic majorant of u, there exists a least one among them, denoted by h_D. For any relatively compact subdomain D' of D, $h_{D'}$ always exists and equals the †Perron-Brelot solution in D' for the boundary function u. As D' increases to D, $h_{D'}$ increases to a function h that is either harmonic or equal constantly to ∞. If h_D exists, it coincides with h, and hence h is harmonic. Conversely, if h is harmonic, then h_D exists and equals h. Generally, there is a unique †Radon measure μ in D with the following property: Let δ be any subdomain of D such that h_δ and the †Green's function G_δ exist in δ (δ may coincide with D). Then $h_\delta - u$ is equal to the potential $\int_\delta G_\delta \, d\mu$, and $u = h_\delta - \int_\delta G_\delta \, d\mu$. In general, a representation of a superharmonic (subharmonic) function as the sum (difference) of a harmonic function and a potential is called a **Riesz decomposition**.

T. Boundary Values

Let D be a domain in which a Green's function exists, and consider the †fine topology on the †Martin compactification of D. Any negative subharmonic function u in D has a finite limit with respect to the fine topology at every point of the †Martin boundary Δ except at the points of a subset of Δ of harmonic measure zero (J. L. Doob). When D is a ball, u has a limit in the ordinary sense along almost every radius. However, it may happen that even if u is bounded there exists no angular limit at any point of the boundary. If the mean value of $|u|$ on every concentric smaller ball in D is bounded, then u can be decomposed into the sum of a nonnegative harmonic function and a negative subharmonic function by the Riesz decomposition, and hence u has a limit both radially and with respect to the fine topology.

Let D be a domain and K be a compact subset of D of †capacity zero. If a function u is subharmonic and bounded above in $D - K$, then u can be extended to be a subharmonic function in D. A function that is equal to a subharmonic function almost everywhere is called **almost subharmonic**, and an almost subharmonic function satisfying condition (3′) is called **submedian**.

Subharmonic functions can be discussed in a space more general than $\mathbf{R}^n$, e.g., a †Riemann surface ($n=2$), or more generally, an †$\mathscr{E}$-space of dimension n ($\geqslant 2$) in the sense of Brelot and Choquet.

U. The Axiomatic Treatment

†Newtonian potentials were the main object of interest in the early stages of †potential theory. A major part of potential theory can be discussed on the basis of the theory of superharmonic functions [8]. For example, a †polar set is defined as a set on which some superharmonic function assumes the value ∞, and a set X is †thin at a point $P_0 \notin X$ if and only if P_0 has a positive distance from X or there exists a superharmonic function $v(P)$ in a neighborhood of P_0 such that $\limsup v(P) > v(P_0)$ as $P \in X$ tends to P_0. Moreover, we can discuss †balayage, define potentials, and obtain Riesz decompositions. Generalizing results of Doob (1954) and starting from a family of harmonic functions defined axiomatically in a locally compact Hausdorff space, M. Brelot defined superharmonic functions and potentials and discussed balayage, Riesz decompositions, and

the †Dirichlet problem (1957). Further progress in axiomatic potential theory has been made by Brelot, H. Bauer, C. Constantinescu, A. Cornea, and others [11, 14].

References

[1] M. Brelot, Sur le rôle du point à l'infini dans la théorie des fonctions harmoniques, Ann. Sci. Ecole Norm. Sup., (3) 61 (1944), 301–332.
[2] M. Brelot, Eléments de la théorie classique du potentiel, Centre de Documentation Universitaire, Paris, third edition, 1965.
[3] M. Brelot and G. Choquet, Espaces et lignes de Green, Ann. Inst. Fourier, 3 (1951), 199–263.
[4] O. D. Kellogg, Foundations of potential theory, Springer, 1929.
[5] M. Nicolescu, Les fonctions polyharmoniques, Actualités Sci. Ind., Hermann, 1936.
[6] M. Ohtsuka, Extremal length of level surfaces and orthogonal trajectories, J. Sci. Hiroshima Univ., 28 (1964), 259–270.
[7] M. Parreau, Sur les moyennes des fonctions harmoniques et analytiques et la classification des surfaces de Riemann, Ann. Inst. Fourier, 3 (1952), 103–197.
[8] M. Brelot, Sur la théorie autonome des fonctions sous-harmoniques, Bull. Sci. Math., (2) 65 (1941), 72–98.
[9] M. Brelot, Lectures on potential theory, Tata Inst., 1960.
[10] T. Radó, Subharmonic functions, Erg. Math., Springer, 1937 (Chelsea, 1970).
[11] Colloque International sur la Théorie du Potentiel, Paris-Orsay, 1964, C.N.R.S. (Ann. Inst. Fourier, 15 (1965), fasc. 1.)
[12] L. Helms, Introduction to potential theory, Wiley Interscience, 1969.
[13] W. K. Hayman and P. B. Kennedy, Subharmonic functions I, Academic Press, 1976.
[14] C. Constantinescu and A. Cornea, Potential theory on harmonic spaces, Springer, 1972.
Also → references to 120 Dirichlet Problem and 338 Potential Theory.

194 (VII.11) Harmonic Integrals

A. Introduction

†De Rham's theorem shows that the cohomology group with real coefficients of a †differentiable manifold of class C^∞ is isomorphic to the cohomology group of the cochain complex of †differential forms with respect to the exterior derivative d. Thus every element of the cohomology group can be represented by a class of †closed differential forms. Harmonic forms enable us to choose one definite differential form in each cohomology class. The theory of harmonic forms, called the theory of **harmonic integrals**, is modeled after the theory of holomorphic differentials and their integrals (Abelian integrals) in function theory [2, 4, 5, 8].

B. Definitions

Let X be an oriented n-dimensional differentiable manifold of class C^∞ with a †Riemannian metric ds^2 of class C^∞ (→ 105 Differentiable Manifolds, 364 Riemannian Manifolds A). For every (C^∞) p-form φ on X we define an $(n-p)$-form $*\varphi$ on X as follows: First denote the †volume element of X by dv. If we choose a basis $\{\omega_1, \dots, \omega_n\}$ of the space of 1-forms on an open set U of X such that $ds^2 = \sum_i \omega_i^2$ and $dv = \omega_1 \wedge \dots \wedge \omega_n$, then φ can be expressed on U in the form $\varphi = (1/p!) \sum \varphi_{i_1, \dots, i_p} \omega_{i_1} \wedge \dots \wedge \omega_{i_p}$. If we let $*\varphi = (1/(n-p)!) \sum (*\varphi)_{j_1, \dots, j_{n-p}} \omega_{j_1} \wedge \dots \wedge \omega_{j_{n-p}}$, where $(*\varphi)_{j_1 \dots j_{n-p}} = (1/p!) \sum \delta^{1\,2 \dots n}_{i_1 \dots i_p j_1 \dots j_{n-p}} \varphi_{i_1 \dots i_p}$, then $*\varphi$ is an $(n-p)$-form on U that does not depend on the choice of $(\omega_1, \dots, \omega_n)$ and is determined only by φ. Since X is covered by open sets as above, $*$ defines a linear mapping that transforms p-forms to $(n-p)$-forms. If we let $ds^2 = \sum g_{jk} dx^j dx^k$ in terms of the local coordinate system $(x^1, \dots, x^n)$ and $\varphi = (1/p!) \varphi_{i_1, \dots, i_p} dx^{i_1} \wedge \dots \wedge dx^{i_p}$, then, in the notation of tensor calculus, we have

$$*\varphi = (1/(n-p)!)(*\varphi)_{j_1 \dots j_{n-p}} dx^{j_1} \wedge \dots \wedge dx^{j_{n-p}},$$

$$(*\varphi)_{j_1 \dots j_{n-p}} = \sqrt{g} \cdot \delta^{1\,2 \dots n}_{k_1 \dots k_p j_1 \dots j_{n-p}} \varphi^{k_1 \dots k_p} \qquad (g = \det(g_{ij})).$$

For two p-forms φ and ψ, we define the inner product by $(\varphi, \psi) = \int_X \varphi \wedge *\psi$ if the right-hand side converges. In order for the inner product (φ, ψ) to be defined, it suffices that either φ or ψ has compact †support. Then (φ, ψ) is a symmetric, positive definite bilinear form.

If we let $\delta = (-1)^{np+n+1} * d *$ operate on p-forms, where d is the †exterior derivative, then d and δ are **adjoint** to each other with respect to the inner product. That is, if either φ or ψ has compact support, we have $(d\varphi, \psi) = (\varphi, \delta\psi)$ (†Stokes's theorem). We call $\Delta = d\delta + \delta d$ the **Laplace-Beltrami operator**, which is a self-adjoint †elliptic differential operator. These operators satisfy relations such as $** = (-1)^{p(n-p)}$, $dd = 0$, $\delta\delta = 0$, $*\Delta = \Delta *$,

$*\delta=(-1)^p d*$, and $\delta *=(-1)^{n-p+1}*d$ (when they operate on p-forms).

A differential form φ is said to be **harmonic** if $d\varphi=0$ and $\delta\varphi=0$. Then $\Delta\varphi=0$. Since Δ is an elliptic operator, a †weak solution φ of the equation $\Delta\varphi=\mu$ is an ordinary solution of class $\mathbf{C}^\infty$ on the domain where μ is of class $\mathbf{C}^\infty$. Therefore, if φ is harmonic (as a weak solution), φ is of class $\mathbf{C}^\infty$.

C. Harmonic Forms on Compact Manifolds

On a compact manifold X, any φ with $\Delta\varphi=0$ is harmonic, since $(\varphi,\Delta\varphi)=(d\varphi,d\varphi)+(\delta\varphi,\delta\varphi)$. Let $L_p(X)$ be the linear space of p-forms of class $\mathbf{C}^\infty$ on X, and denote by $\mathfrak{L}_p(X)$ the completion of $L_p(X)$ with respect to the inner product (φ,ψ). Then $\mathfrak{L}_p(X)$ is the Hilbert space of square integrable measurable p-forms. Then $\mathfrak{H}_p(X)=\{\varphi\in\mathfrak{L}_p(X)\mid\Delta\varphi=0$ (in the weak sense)$\}$ is a finite-dimensional subspace of $\mathfrak{L}_p(X)$ and is contained in $L_p(X)$, as we have seen before. Also, $\mathfrak{H}_p(X)$ is closed in $\mathfrak{L}_p(X)$, and the †projection operator $H:\mathfrak{L}_p(X)\to\mathfrak{H}_p(X)$ is an †integral operator with kernel of class $\mathbf{C}^\infty$. The orthogonal complement of $\mathfrak{H}_p(X)$ in $\mathfrak{L}_p(X)$ is mapped onto itself by Δ and has the inverse operator G of Δ, which is a continuous operator of the Hilbert space. By letting $G=0$ on $\mathfrak{H}_p(X)$, we can extend G to an operator from $\mathfrak{L}_p(X)$ to $\mathfrak{L}_p(X)$ that is called **Green's operator**. It is also denoted by G, maps L_p into itself, commutes with d and δ, and satisfies $GH=HG=0$, $\mathrm{H}+\Delta\mathrm{G}=1$ (= identity mapping). Therefore, for $\varphi\in L_p(X)$ we have $\varphi=H\varphi+G\delta d\varphi+d\delta G\varphi$, which shows that H is †homotopic to the identity mapping of the †cochain complex $(\sum_p L_p(X), d)$. From this we infer that every cohomology class of de Rham cohomology contains a unique harmonic form that represents the cohomology class. However, since products of harmonic forms are not always harmonic, it is not appropriate to use harmonic forms to study the ring structure of cohomology. G is also an integral operator with kernel of class $\mathbf{C}^\infty$ outside the diagonal subset in $X\times X$.

D. Harmonic Forms on Noncompact Manifolds

If X is a noncompact manifold, let $L_p(X)$ be the space of p-forms of class $\mathbf{C}^\infty$ with compact support, and let $\mathfrak{L}_p(X)$ be its completion. Let $\mathfrak{B}_p(X)$ and $\mathfrak{B}_p^*(X)$ be the respective closures of $dL_{p-1}(X)$ and $\delta L_{p+1}(X)$ in $\mathfrak{L}_p(X)$, and let $\mathfrak{Z}_p(X)$ and $\mathfrak{Z}_p^*(X)$ be the respective orthogonal complements of $\mathfrak{B}_p(X)$ and $\mathfrak{B}_p^*(X)$ in $\mathfrak{L}_p(X)$. Then $\mathfrak{Z}_p(X)\cap\mathfrak{Z}_p^*(X)=\mathfrak{H}_p(X)$ is a subspace of the square integrable harmonic forms, and we have the direct sum decomposition $\mathfrak{L}_p(X)=\mathfrak{B}_p(X)+\mathfrak{B}_p^*(X)+\mathfrak{H}_p(X)$. In this decomposition any component of a form of class $\mathbf{C}^\infty$ is also of class $\mathbf{C}^\infty$.

If X is an open submanifold of another manifold Y, $\bar{X}$ is compact, and $\partial X=\bar{X}-X$ is a closed submanifold of Y, then the theory in this section is just a generalized potential theory with boundary condition $\varphi=0$ on ∂X. We sometimes treat decompositions of other Hilbert spaces that correspond to other boundary conditions.

E. Generalization to Complex Manifolds

If X is a complex manifold, we consider complex-valued differential forms (→ 72 Complex Manifolds C). Then the space $L_p(X)$ of p-forms is the direct sum of the spaces $L_{r,s}(X)$ of forms of type (r,s), and the exterior derivative d has the expression $d=d'+d''$, where d' is of type $(1,0)$ (i.e., $L_{r,s}(X)\to L_{r+1,s}(X)$) and d'' is of type $(0,1)$. If we are given a holomorphic vector bundle E on X, we can define an operator d'' on differential forms with values in E, and we have the generalized †Dolbeault theorem. If X is compact and has a †Hermitian metric, we can define a Hermitian inner product on E as follows: There is an open covering $\{U_j\}$ such that over each U_j the vector bundle E is isomorphic to $U_j\times\mathbf{C}^q$. A point of E over U_j is represented by (x,ξ_j), where $x\in U_j$ and $\xi_j\in\mathbf{C}^q$. For $x\in U_j\cap U_k$ we have $(x,\xi_j)=(x,\xi_k)$ (the sides are the respective expressions over U_j and U_k) if and only if $\xi_j=g_{jk}(x)\xi_k$, where $g_{jk}(x)$ is a holomorphic mapping from $U_j\cap U_k$ to $GL(q,\mathbf{C})$ satisfying $g_{jk}g_{kl}=g_{jl}$ on $U_j\cap U_k\cap U_l$. A differential form φ with values in E is expressed as a family $\{\varphi_j\}$ of differential forms on U_j with values in $\mathbf{C}^q$ such that $\varphi_j(x)=g_{jk}(x)\varphi_k(x)$ on $U_j\cap U_k$. If we take a positive definite Hermitian matrix h_j whose components are C^∞-functions on U_j such that ${}^t g_{jk}h_j\bar{g}_{jk}=h_k$ on $U_j\cap U_k$, then $\{{}^t\xi_j h_j\bar{\xi}_j\}$ determines a Hermitian inner product on each fiber of E. We can also endow the space $L_{r,s}(E,X)$ (of forms of type (r,s) of class C^∞ with values in E) with a Hermitian inner product by setting $(\varphi,\psi)=\int_X\sum_{\alpha,\beta}h_{j\alpha\beta}\varphi_j^\alpha\wedge\overline{*\psi_j^\beta}$ for $\varphi,\psi\in L_{r,s}(E,X)$ (where the φ_j^α $(\alpha=1,\ldots,q)$ are the components of φ_j). If we denote by $\mathfrak{d}$ the adjoint operator of d'' with respect to this inner product and let $A=d''\mathfrak{d}+\mathfrak{d}d''$, then A is a self-adjoint elliptic differential operator, and results similar to those for Δ mentioned above hold for A. For example, the space $\mathfrak{H}_{r,s}(E,X)$ of harmonic forms of type (r,s) is of finite dimension, and there is a continuous linear operator G on $\mathfrak{L}_{r,s}(E,X)$, the completion of $L_{r,s}(E,X)$, that satisfies $1=H+AG$, $HG=GH=0$, $d''G=Gd''$, and $\mathfrak{d}G=G\mathfrak{d}$. Here H

denotes the projection $\mathfrak{L}\to\mathfrak{H}$, which is an integral operator with kernel of class C^∞. Also, G maps $L_{r,s}(E,X)$ into itself. Therefore H is homotopic to the identity on the cochain complex $(\sum_s L_{r,s}(E,X), d'')$, and any element of †Dolbeault's cohomology groups (d''-cohomology groups) is represented by a unique harmonic form (→ 232 Kähler Manifolds B).

F. Other Generalizations

Even if a manifold X is not of class C^∞, if X is a manifold of class C_1^1, we can develop the theory of harmonic forms [6]. We say that X is of class C_1^1 if it is of class C^1 and has a set of local coordinate systems whose transition functions have derivatives satisfying the †Lipschitz condition.

If X is a real analytic manifold with a real analytic Riemannian metric, then harmonic forms are also real analytic. Using this fact, we can embed real analytically a compact manifold with a real analytic Riemannian metric into a Euclidean space (P. Bidal and G. de Rham; this result is now included in the theorems of C. B. Morrey and H. Grauert).

We can consider the theory of harmonic forms with singularities [4, 5], a generalization of the theory of differential forms of the second and third kinds. Here the notion of †current is very useful.

G. Cohomology Vanishing Theorems

Since the operator Δ is closely related to the Riemannian metric, some metrics may admit no harmonic forms of certain degrees except zero. This is important since it means that the corresponding cohomology group of the manifold vanishes. The condition for this phenomenon to occur can be described in terms of the curvature of the metric. This study has its origin in S. Bochner's results [2].

Here is an example of a **vanishing theorem**: Let B be a holomorphic line bundle on a compact complex manifold X of dimension n. If the †Chern class of B is expressed by a real closed differential form of type $(1,1)$ as $\omega=\sqrt{-1}\sum a_{\alpha\beta}dz^\alpha\wedge d\bar{z}^\beta$, where the Hermitian matrix $(a_{\alpha\beta})$ is positive definite at every point of X, then $H^q(X,\Omega^p(B))=0$ for $p+q>n$. In this case, $ds^2=2\sum a_{\alpha\beta}dz^\alpha\wedge d\bar{z}^\beta$ is a †Hodge metric on X (→ 232 Kähler Manifolds D).

References

[1] W. L. Baily, The decomposition theorem for V-manifolds, Amer. J. Math., 78 (1956), 862–888.

[2] S. Bochner and K. Yano, Curvature and Betti numbers, Ann. Math. Studies 22, Princeton Univ. Press, 1953.

[3] S. I. Goldberg, Curvature and homology, Academic Press, 1962.

[4] W. V. D. Hodge, The theory and applications of harmonic integrals, Cambridge Univ. Press, second edition, 1952.

[5] K. Kodaira, Harmonic fields in Riemannian manifolds (generalized potential theory), Ann. Math., (2) 50 (1949), 587–665.

[6] C. B. Morrey and J. Eells, A variational method in the theory of harmonic integrals I, Ann. Math., (2) 63 (1956), 91–128.

[7] G. de Rham, Variétés différentiables, Actualités Sci. Ind., Hermann, second edition, 1960.

[8] G. de Rham and K. Kodaira, Harmonic integrals, Lecture notes, Inst. for Advanced Study, Princeton, 1950.

195 (VII.15) Harmonic Mappings

A. General Remarks

The theory of harmonic mappings between Riemannian manifolds has its origin in the study of †Plateau's problem. The basic problem in the theory is to deform a given mapping into a harmonic one, which is a problem of the †calculus of variations and †global analysis (→ 46 Calculus of Variations, 183 Global Analysis). Recently, the theory of harmonic mappings has been applied to problems in various branches of geometry [7–9, 11].

B. Definitions and Examples

Let (M,g) and (N,h) be †Riemannian manifolds with metrics $g=\sum g_{ij}dx^i dx^j$ and $h=\sum h_{\alpha\beta}dy^\alpha dy^\beta$, respectively. We define the **energy** of a C^1-mapping $f: M\to N$ by

$$E(f)=\frac{1}{2}\int_M |df(x)|^2\,dx,$$

where $|df(x)|$ is the †Hilbert-Schmidt norm of the differential $df_x: T_x(M)\to T_{f(x)}(N)$ of f at $x\in M$ and dx is the canonical Lebesgue measure defined by g on M (assumed compact). Thus $E(f)$ can be considered to be a generalization of the classical †Dirichlet integral for functions. The integrand $e(f)(x)=|df(x)|^2$ is called the **energy density** of f; it measures the sum of the squares of elements of length stretched on a complete set of mutually perpendicular directions.

The †Euler-Lagrange differential equations of the energy functional $E(f)$ yield a vector field $\tau(f)$ along f, i.e., a section of the bundle $f^{\#}T(N)$ induced from the tangent bundle $T(N)$ of N by f. In fact, given a family f_t of mappings depending differentiably on t with $f_0=f$, we have

$$\frac{d}{dt}E(f_t)|_{t=0}=-\int_M\left\langle\tau(f),\frac{\partial f_t}{\partial t}\bigg|_{t=0}\right\rangle dx,$$

where $\langle\ ,\ \rangle$ denotes the inner product of tangent vectors along f. The vector field $\tau(f)$ is called the **tension field** of the mapping f; it indicates the direction in which the energy of f decreases most rapidly.

The Euler-Lagrange differential equations $\tau(f)=0$ are a system of †quasilinear elliptic partial differential equations of the second order. In local coordinates, these can be written in the form

$$\Delta f^\alpha+\sum g^{ij}\Gamma'^{\alpha}_{\beta\gamma}(f)\frac{\partial f^\beta}{\partial x^i}\frac{\partial f^\gamma}{\partial x^j}=0,$$

where Δ is the †Laplace-Beltrami operator on M and the $\Gamma'^{\alpha}_{\beta\gamma}(f)(x)$ are the †Christoffel symbols on N at $f(x)$. (The f^α are local coordinates of the point $f(x)$, and (g^{ij}) is the inverse matrix of (g_{ij}).) A C^2-mapping $f:M\rightarrow N$ is said to be **harmonic** if its tension field $\tau(f)$ vanishes. Thus, if M is compact, a C^2-mapping $f:M\rightarrow N$ is harmonic if and only if it is an extremal of the energy functional $E(f)$.

Examples of harmonic mappings appear in various contexts of differential geometry. For instance:

(1) If $N=\mathbf{R}$, then the harmonic mappings $M\rightarrow\mathbf{R}$ are the †harmonic functions on M.

(2) If M is the circle S^1, then a harmonic mapping $S^1\rightarrow N$ is a closed †geodesic of N parametrized by arc length.

(3) Let $f:M\rightarrow N$ be an †isometric immersion of M into N. Then f is harmonic if and only if it is a †minimal immersion.

(4) If M and N are †Kähler manifolds, then every †holomorphic or **antiholomorphic** mapping $M\rightarrow N$ is harmonic, where by an antiholomorphic mapping is meant a mapping whose differential mapping carries a differential form of type (1,0) into that of type (0,1). We note that each (anti-)holomorphic mapping is an absolute minimum for the energy in its homotopy class. There are also examples of nonholomorphic (and nonantiholomorphic) harmonic mappings between Kähler manifolds (→ Section D).

C. Fundamental Properties

(1) **Regularity**. Since a harmonic mapping is a solution of a second-order quasilinear elliptic system of partial differential equations $\tau(f)=0$ (→ Section B), it is a smooth (i.e., of class C^∞) mapping. More generally, it is known that a continuous mapping that satisfies $\tau(f)=0$ in a weak sense is smooth [4, 6].

(2) **Unique continuation property**. The following unique continuation theorem is valid for harmonic mappings: If two harmonic mappings of M into N agree up to infinitely high order at some point of M, then they are identical (M being assumed connected). In particular, a harmonic mapping that is constant on an open set is a constant mapping.

The global natures of harmonic mappings are closely related to the curvatures of the manifolds under consideration. For instance, suppose that M and N are compact and that the sectional curvatures of N are nonpositive everywhere. Then we have:

(3) **Uniqueness**. Let $f:M\rightarrow N$ be a harmonic mapping, and assume that there is a point of $f(M)$ where the sectional curvatures of N are negative. Then f is unique in its homotopy class unless $f(M)$ is a closed geodesic γ of N; and in this case we have uniqueness up to rotation of γ, i.e., an isometry of γ which moves each point of γ a fixed oriented distance along γ.

(4) **Degeneracy**. Suppose further that the †Ricci tensor (R_{ij}) of M is positive semidefinite everywhere. Then the energy density $e(f)$ is a †subharmonic function for every harmonic mapping. This implies that any harmonic mapping $f:M\rightarrow N$ is †totally geodesic. Moreover, if N is of negative sectional curvature, then f is either constant or maps M onto a closed geodesic of N; if (R_{ij}) is positive definite at some point, then f is constant.

(5) **Finiteness**. Assume now that N is of negative sectional curvature. Then, for each $K\geqslant 1$, there are only finitely many nonconstant harmonic mappings $f:M\rightarrow N$ of dilatation bounded by K. Here, we say that the dilatation of f is bounded by K if and only if at each point of M we have $df=0$ or $(\lambda_1/\lambda_2)^{1/2}\leqslant K$, $\lambda_1\geqslant\lambda_2\geqslant\ldots>0$ being the positive eigenvalues of the pullback quadratic form $f^*h(x)$ on $T_x(M)$ induced from the metric h of N by f.

D. Harmonic Mappings of a Surface

Let M be a compact surface. Then the energy of a mapping $M\rightarrow N$ is the †Dirichlet-Douglas functional, and harmonic mappings are closely connected with solutions of Plateau's problem (→ 334 Plateau's Problem). In fact, if a †conformal mapping $M\rightarrow N$ minimizes the †area functional, then it also minimizes the energy functional.

Now let M and N be compact orientable

surfaces whose genera are denoted by p and q, respectively. Then the problem of existence (or nonexistence) of harmonic mappings is well understood. In fact:

(1) When $q \neq 0$, for any metrics g and h on M and N, every homotopy class of mappings $M \to N$ contains a harmonic mapping.

(2) When $q=0$ (i.e., N is the 2-sphere S^2), every harmonic mapping whose †degree d satisfies $|d| \geqslant p$ is holomorphic or antiholomorphic with respect to the complex structures associated with g and h. For example, consider the homotopy classes of mappings from the 2-torus T^2 to S^2 with any metrics. Then all classes with degree $|d| \geqslant 2$ have harmonic representatives, and any such is holomorphic or antiholomorphic; and the classes with $d = \pm 1$ have no harmonic representatives.

(3) When $q=0$ and $|d| \leqslant p-1$, we have, for every such p and d, a surface M of genus p and a metric h on S^2 such that there exists a harmonic nonholomorphic (and nonantiholomorphic) mapping of degree d from M to S^2.

E. Existence Theorems

The basic problem in the study of harmonic mappings is to prove their existence in general geometric contexts.

(1) In regard to this problem, translating the problem of the elliptic system $\tau(f)=0$ into the †initial value problem of the corresponding nonlinear †parabolic system $\partial f/\partial t = \tau(f)$, J. Eells and J. H. Sampson [1] proved that if M and N are compact and if N has nonpositive sectional curvature everywhere, then every homotopy class of mappings $M \to N$ contains a harmonic mapping that minimizes the energy in that class. Subsequently, the uniqueness of these harmonic mappings was established by P. Hartman [2] in the form stated in Section C.

(2) For harmonic mappings of surfaces, more general existence results have been known.

First, by the †direct method of the calculus of variations, L. Lemaire (1977) and others proved that if M and N are compact and if M is 2-dimensional, then every conjugacy class of homomorphisms $\pi_1(M) \to \pi_1(N)$ of the fundamental groups is induced by a minimizing harmonic mapping. It follows that if, in particular, the second homotopy group $\pi_2(N)$ of N is zero, then every homotopy class of mappings of a compact surface M to N contains a harmonic representative realizing the minimum of the energy in that class.

On the other hand, by making use of the generalized †Morse theory for a perturbed energy functional, J. Sacks and K. Uhlenbeck [5] succeeded in giving a satisfactory answer to the structure of $\pi_2(N)$, which is a $\pi_1(N)$-module, in terms of harmonic mappings. They proved that there exists a generating set for $\pi_2(N)$ consisting of harmonic mappings of spheres that minimize energy and area in their homotopy classes. We note that these harmonic mappings are minimal immersions with †branch points.

(3) Next, we mention the case of harmonic mappings of manifolds with boundary. In this case, we can naturally formulate the †Dirichlet and the †Neumann boundary value problem for harmonic mappings.

In his study of Plateau's problem on Riemannian manifolds, C. Morrey (1948) discussed the Dirichlet problem for harmonic surfaces with boundary.

The problem in arbitrary dimensions has been studied by R. S. Hamilton [3], who extended the result of Eells and Sampson mentioned above to the case where M and N have boundaries. In fact, let M and N be compact Riemannian manifolds with boundary, and assume that N has nonpositive sectional curvature and that the boundary ∂N of N is †convex (or empty). Then there exists a unique minimizing harmonic mapping in each †relative homotopy class determined by the prescribed Dirichlet boundary value. We note that if ∂N is not convex, then it is easy to formulate Dirichlet problems with no solutions. Hamilton also treated the Neumann problem.

Subsequently, S. Hildebrandt, H. Kaul, and K.-O. Widman [4] gave another existence proof of solutions of the Dirichlet problem that covers the case where N admits positive sectional curvature.

References

[1] J. Eells and J. H. Sampson, Harmonic mappings of Riemannian manifolds, Amer. J. Math., 86 (1964), 109–160.
[2] P. Hartman, On homotopic harmonic maps, Canad. J. Math., 19 (1967), 673–687.
[3] R. S. Hamilton, Harmonic maps of manifolds with boundary, Lecture notes in math. 471, Springer, 1975.
[4] S. Hildebrandt, H. Kaul, and K.-O. Widman, An existence theorem for harmonic mappings of Riemannian manifolds, Acta Math., 138 (1977), 1–16.
[5] J. Sacks and K. Uhlenbeck, The existence of minimal immersions of two-spheres, Ann. Math., (2) 113 (1981), 1–24.
[6] J. Eells and L. Lemaire, A report on harmonic maps, Bull. London Math. Soc., 10 (1978), 1–68.

[7] Y.-T. Siu and S.-T. Yau, Compact Kähler manifolds of positive bisectional curvature, Inventiones Math., 59 (1980), 189–204.
[8] Y.-T. Siu, The complex-analyticity of harmonic maps and the strong ridigity of compact Kähler manifolds, Ann. Math., (2) 112 (1980), 73–111.
[9] S. Nishikawa and K. Shiga, On the holomorphic equivalence of bounded domains in complete Kähler manifolds of nonpositive curvature, J. Math. Soc. Japan, 35 (1983) 273–278.
[10] T. Ishihara, The index of a holomorphic mapping and the index theorem, Proc. Amer. Math. Soc., 66 (1977), 169–174.
[11] T. Sunada, Rigidity of certain harmonic mappings, Inventiones Math., 51 (1979), 297–307.

196 (XXI.28) Hilbert, David

David Hilbert (January 23, 1862–February 14, 1943) was born in Königsberg, Germany. He attended the University of Königsberg from 1882 to 1885, when he received his doctoral degree with a thesis on the theory of invariants. It was there that he established a lifelong friendship with H. Minkowski. In 1892 he became a professor at the University, and in 1895 he was appointed to a professorship at the University of Göttingen, a position he held until his death. He obtained his basic theorem on invariants between 1890 and 1893, and next began research on the foundations of geometry (→ 155 Foundations of Geometry) and the theory of †algebraic number fields. Concerning the former, he published *Grundlagen der Geometrie* (first edition, 1899), in which he gave the complete axioms of Euclidean geometry and a logical examination of them. Concerning the latter, he systematized all the important known results of algebraic number theory in his monumental *Zahlbericht* (1897). In number theory, he enunciated his significant conjecture on †class field theory. At the international congress of mathematicians held in Paris in 1900, he put forth 23 problems as targets for mathematics of the 20th century (Table 1). Between 1904 and 1906 he conducted research on the †Dirichlet principle of †potential theory and on the direct method in the †calculus of variations. Around 1909 he established the foundations of the theory of †Hilbert spaces. After 1910 he was chiefly involved in research on the †foundations of mathematics, and he advocated the standpoint of †formalism. He is one of the greatest mathematicians of the first half of the 20th century.

Table 1. The 23 Problems of Hilbert

(1) To prove the continuum hypothesis (→ 33 Axiomatic Set Theory D).

(2) To investigate the consistency of the axioms of arithmetic (→ 156 Foundations of Mathematics E).

(3) To show that it is impossible to prove the following fact utilizing only congruence axioms: Two tetrahedra having the same altitude and base area have the same volume. Solved by M. Dehn (1900).

(4) To investigate geometries in which the line segment between any pair of points gives the shortest path between the pair (→ 155 Foundations of Geometry).

(5) To obtain the conditions under which a topological group has the structure of a Lie group (→ 423 Topological Groups M). Solved by A. M. Gleason and D. Montgomery and L. Zippin (1952) and H. Yamabe (1953).

(6) To axiomatize those physical sciences in which mathematics plays an important role.

(7) To establish the transcendence of certain numbers (→ 430 Transcendental Numbers B). The transcendence of $2^{\sqrt{2}}$, which was one of the numbers put forth by Hilbert, was shown by A. Gel'fond (1934) and T. Schneider (1935).

(8) To investigate problems concerning the distribution of prime numbers; in particular, to show the correctness of the Riemann hypothesis (→ 450 Zeta Functions). Unsolved.

(9) To establish a general law of reciprocity (→ 59 Class Field Theory A). Solved by T. Takagi (1921) and E. Artin (1927).

(10) To establish effective methods to determine the solvability of Diophantine equations (→ 97 Decision Problem; 182 Geometry of Numbers). Solved affirmatively for equations of two unknowns by A. Baker, *Philos. Trans. Roy. Soc. London*, (A) 263 (1968); solved negatively for the general case by Yu. V. Matiyasevich (1970).

(11) To investigate the theory of quadratic forms over an arbitrary algebraic number field of finite degree (→ 348 Quadratic Forms).

(12) To construct class fields of algebraic number fields (→ 73 Complex Multiplication).

(13) To show the impossibility of the solution of the general algebraic equation of the seventh degree by compositions of continuous functions of two variables. Solved negatively. In general, V. I. Arnold proved that every real, continuous function $f(x_1, x_2, x_3)$ on $[0, 1]$ can be represented in the form $\sum_{i=1}^{9} h_i(g_i(x_1, x_2), x_3)$, where h_i and g_i are real, continuous functions, and A. N. Kolmogorov proved that $f(x_1, x_2, x_3)$ can be represented

in the form $\sum_{i=1}^{7} h_i(g_{i1}(x_1)+g_{i2}(x_2)+g_{i3}(x_3))$, where h_i and g_{ij} are real, continuous functions and g_{ij} can be chosen once for all independently of f (*Dokl. Akad. Nauk SSSR*, 114 (1957), *Amer. Math. Soc. Transl.*, 28 (1963)).

(14) Let k be a field, $x_1, \ldots, x_n$ be variables, and $f_i(x_1, \ldots, x_n)$ be given polynomials in $k[x_1, \ldots, x_n]$ $(i=1, \ldots, m)$. Furthermore, let R be the ring formed by rational functions $F(X_1, \ldots, X_m)$ in $k(X_1, \ldots, X_m)$ such that $F(f_1, \ldots, f_m) \in k[x_1, \ldots, x_n]$. The problem is to determine whether the ring R has a finite set of generators. Solved negatively by M. Nagata, *Amer. J. Math.*, 81 (1959).

(15) To establish the foundations of algebraic geometry (→ 12 Algebraic Geometry). Solved by B. L. van der Waerden (1938–1940), A. Weil (1950), and others.

(16) To conduct topological studies of algebraic curves and surfaces.

(17) Let $f(x_1, \ldots, x_n)$ be a rational function with real coefficients that takes a positive value for any real n-tuple $(x_1, \ldots, x_n)$. The problem is to determine whether the function f can be written as the sum of squares of rational functions (→ 149 Fields O). Solved in the affirmative by E. Artin (1927).

(18) To express Euclidean n-space as a disjoint union $\bigcup_\lambda P_\lambda$, where each P_λ is congruent to one of a set of given polyhedra.

(19) To determine whether the solutions of regular problems in the calculus of variations are necessarily analytic (→ 323 Partial Differential Equations of Elliptic Type). Solved by S. N. Bernshteĭn, I. G. Petrovskiĭ, and others.

(20) To investigate the general boundary value problem (→ 120 Dirichlet Problem; 323 Partial Differential Equations of Elliptic Type).

(21) To show that there always exists a linear differential equation of the Fuchsian class with given singular points and monodromic group (→ 253 Linear Ordinary Differential Equations (Global Theory)). Solved by H. Röhrl and others (1957).

(22) To uniformize complex analytic functions by means of automorphic functions (→ 367 Riemann Surfaces). Solved for the case of one variable by P. Koebe (1907).

(23) To develop the methodology of the calculus of variations (→ 46 Calculus of Variations).

References

[1] D. Hilbert, Gesammelte Abhandlungen I–III, Springer, 1932–1935 (Chelsea, 1967).
[2] D. Hilbert, Grundlagen der Geometrie, Teubner, seventh edition, 1930.
[3] D. Hilbert, Grundzüge der allgemeinen Theorie der linearen Integralgleichungen, Teubner, second edition, 1924 (Chelsea, 1953).
[4] D. Hilbert and W. Ackermann, Grundzüge der theoretischen Logik, Springer, third edition, 1949; English translation, Principles of mathematical logic, Chelsea, 1950.
[5] D. Hilbert and P. Bernays, Grundlagen der Mathematik, Springer, second edition, I, 1968; II, 1970.
[6] F. Klein, Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert I, Springer, 1926 (Chelsea, 1956).
[7] H. Weyl, David Hilbert and his mathematical work, Bull. Amer. Math. Soc., 50 (1944), 612–654.
[8] C. Reid, Hilbert. With an appreciation of Hilbert's mathematical work by H. Weyl, Springer, 1970.

197 (XII.2) Hilbert Spaces

A. General Remarks

The theory of Hilbert spaces arose from problems in the theory of †integral equations. D. Hilbert noticed that a linear integral equation can be transformed into an infinite system of linear equations for the †Fourier coefficients of the unknown function. He considered the linear space l_2 consisting of all sequences of numbers $\{x_n\}$ for which $\sum_{n=1}^{\infty} |x_n|^2$ is finite, and defined for each pair of elements $x=\{x_n\}$, $y=\{y_n\} \in l_2$ their inner product as $(x, y) = \sum_{n=1}^{\infty} x_n \bar{y}_n$. The space l_2 can be regarded as an infinite-dimensional extension of the notion of a Euclidean space. F. Riesz considered the space of functions now termed L_2-space and succeeded in giving a satisfactory answer to the Fourier expansion problem. In his book [3], J. von Neumann established a rigorous foundation of quantum mechanics employing Hilbert spaces and the spectral expansion of self-adjoint operators. The following axiomatic definition (→ Section B) of Hilbert spaces is due to von Neumann. H. Weyl later justified the †Dirichlet principle of Riemann by the method of orthogonal projection in a Hilbert space, and thus paved the way for the function-analytic study of differential equations.

B. Definition

Let K be the field of complex or real numbers, the elements of which we denote by $\alpha, \beta, \ldots$.

Let H be a †linear space over K, and to any pair of elements $x, y \in H$ let there correspond a number $(x, y) \in K$ satisfying the following five conditions: (i) $(x_1 + x_2, y) = (x_1, y) + (x_2, y)$; (ii) $(\alpha x, y) = \alpha(x, y)$; (iii) $(x, y) = \overline{(y, x)}$; (iv) $(x, x) \geqslant 0$; and (v) $(x, x) = 0 \Leftrightarrow x = 0$. Then we call H a **pre-Hilbert space** and (x, y) the **inner product** of x and y.

With the norm $\|x\| = \sqrt{(x, x)}$, H is a †normed linear space. If H is †complete with respect to the distance $\|x - y\|$ (i.e., $\|x_n - x_m\| \rightarrow 0$ $(m, n \rightarrow \infty)$ implies the existence of $\lim x_n = x$), then we call H a **Hilbert space**. According as K is complex or real, we call H a **complex** or **real Hilbert space**. A Hilbert space is a †Banach space.

A normed linear space with norm $\|x\|$ can be made a pre-Hilbert space, by defining an inner product (x, y) so that $\|x\| = \sqrt{(x, x)}$, if and only if the equality $\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2)$ holds for any x, y.

C. Orthonormal Sets

Two elements $x, y \in H$ are said to be mutually **orthogonal** if $(x, y) = 0$. A subset Σ of H is called an **orthogonal set** (or **system**) if $0 \notin \Sigma$ and every distinct pair $x, y \in \Sigma$ is mutually orthogonal. If every element of an orthogonal set Σ is of norm 1, then Σ is called an **orthonormal set**. Any orthogonal set $\Sigma = \{x_i\}$ can be **normalized** into an orthonormal set $\{x_i / \|x_i\|\}$. A maximal orthonormal set is called a **complete orthonormal set** or an **orthonormal basis**. All the complete orthonormal sets of a given H have the same cardinal number, which we call the **dimension** of H. Two Hilbert spaces are isomorphic if and only if they have the same dimension.

Let $\Sigma = \{x_i\}$ be an orthonormal set. Then for every $x \in H$, its **Fourier coefficients** (x, x_i) vanish for all but a countable number of i, and the **Bessel inequality** $\|x\|^2 \geqslant \sum_i |(x, x_i)|^2$ holds. The following three statements are equivalent in a Hilbert space: (i) Σ is complete; (ii) **Parseval's equality** $\|x\|^2 = \sum_i |(x, x_i)|^2$ holds for every x; (iii) every x can be expanded in a **Fourier series** $x = \sum_i (x, x_i) x_i$ (→ 317 Orthogonal Functions).

D. Examples of Hilbert Spaces

The space l_2 (→ Section A) is a Hilbert space of dimension $\aleph_0$. The †function space L_2 on a measure space (X, μ) is a Hilbert space if the inner product of $f, g \in L_2$ is defined by $(f, g) = \int_X f\bar{g}\, d\mu$. In the case of the †Lebesgue measure in a Euclidean space, L_2 is of dimension $\aleph_0$, so that it is a Hilbert space isomorphic to l_2. Further examples of Hilbert spaces are $A_2(\Omega)$, $W_2^l(\Omega)$ $(= H^l(\Omega))$, and $H_0^l(\Omega)$ (→ 168 Function Spaces).

E. Closed Linear Subspaces and Projections

Let M be a **closed linear subspace** of a Hilbert space H, i.e., a linear subspace that is closed in the norm topology of H. It is a Hilbert space with respect to the restriction of the inner product in H. For a given M the set of all $x \in H$ such that $(x, y) = 0$ for every $y \in M$ forms a closed linear subspace $M^\perp$ called the **orthogonal complement** of M. The orthogonal complement of $M^\perp$ is M (i.e., $M^{\perp\perp} = M$), and H is the **direct sum** of M and $M^\perp$ (i.e., every $x \in H$ can be uniquely represented as $x = y + z$, $y \in M$, $z \in M^\perp$, and $\|x\|^2 = \|y\|^2 + \|z\|^2$). Thus the quotient space H/M is isomorphic to $M^\perp$ and is also a Hilbert space. The operator P_M that maps x to y is called the **projection** (or **orthogonal projection** or **projection operator**) to M. A bounded linear operator P is a projection if and only if it is idempotent ($P^2 = P$) and self-adjoint ($(Px, y) = (x, Py)$ for any $x, y \in H$) (→ 251 Linear Operators).

F. Conjugate Spaces

A linear operator from H to K is called a **linear functional**. The set H' of all continuous linear functionals f on H forms a Hilbert space with norm $\|f\| = \sup\{|f(x)| \mid \|x\| = 1\}$. For every $f \in H'$ there exists a unique $y \in H$ such that $f(x) = (x, y)$ for all $x \in H$ (**Riesz's theorem**), and the correspondence $f \rightarrow y$ gives an †antilinear isometric operator from H' onto H (for †linear operators on Hilbert spaces → 68 Compact and Nuclear Operators; 251 Linear Operators; 390 Spectral Analysis of Operators).

References

[1] D. Hilbert, Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen, Teubner, second edition, 1924 (Chelsea, 1953).
[2] E. Hellinger and O. Toeplitz, Integralgleichungen und Gleichungen mit unendlichvielen Unbekannten, Enzykl. Math., Teubner, 1928 (Chelsea, 1953).
[3] J. von Neumann, Mathematische Grundlagen der Quantenmechanik, Springer, 1932.
[4] J. von Neumann, Collected works II, III, Pergamon, 1961.
[5] M. H. Stone, Linear transformations in Hilbert space and their applications to analysis, Amer. Math. Soc. Colloq. Publ., 1932.

[6] B. Sz.-Nagy, Spektraldarstellung linearer Transformationen des Hilbertschen Raumes, Erg. Math., Springer, 1942.
[7] F. Riesz and B. Sz.-Nagy, Leçons d'analyse fonctionelle, Akademiai Kiadó, third edition, 1955; English translation, Functional analysis, Ungar, 1955.
[8] N. I. Akhiezer and I. M. Glazman, Theory of linear operators in Hilbert space, Ungar, I, 1961; II, 1963. (Original in Russian, 1950.)
[9] K. Yosida, Functional analysis, Springer, 1965.
[10] R. Courant and D. Hilbert, Methods of mathematical physics I, Interscience, 1953.
[11] P. R. Halmos, Introduction to Hilbert space and the theory of spectral multiplicity, Chelsea, second edition, 1957.
[12] P. R. Halmos, A Hilbert space problem book, Van Nostrand, 1967.

198 (XI.1) Holomorphic Functions

A. Differentiation of Complex Functions

Let $f(z)$ be a †complex-valued function defined in an open set D in the †complex plane $\mathbf{C}$. We say that $f(z)$ is **differentiable** at z if the limit

$$\lim_{h\to 0}(f(z+h)-f(z))/h=f'(z) \qquad (1)$$

exists and is finite as the complex number h tends to zero. We call $f'(z)$ the **derivative** of $f(z)$ at z. This definition is a formal extension of the definition of differentiability of a function of a real variable to that of a complex variable (→ 106 Differential Calculus), but it is a much stronger condition than the differentiability of a real function, since $z+h$ in (1) may be an arbitrary point in a 2-dimensional neighborhood of z. Hence many results essentially different from those for functions of a real variable follow from it.

If a function $f(z)$ is differentiable at each point of an open set D, it is said to be **holomorphic** (or **regular**) in D, or $f(z)$ is a **holomorphic function** on D. (For the definition of holomorphy of a complex-valued function of several complex variables → 21 Analytic Functions of Several Complex Variables C.)

Let E be an arbitrary nonempty subset of $\mathbf{C}$. We say that $f(z)$ is holomorphic on E if it is defined in an open set D containing E and is holomorphic on D. Some results valid for differentiable real functions also hold for holomorphic functions. For instance, the derivative of a sum, product, or quotient is given by the usual rules. The derivative of a composite function is determined by the chain rule. The set of all functions holomorphic in a †domain D forms a †ring.

Suppose $f(z)$ is holomorphic on D and $f'(z_0)\neq 0$, $z_0\in D$. Then two curves that form an angle at z_0 are mapped by f to two curves forming the same angle at $f(z_0)$. Because of this property, the mapping f is said to be **conformal** at all points z with $f'(z)\neq 0$.

The following four conditions are equivalent for a function $f=u+iv$ defined on an open set D. (1) f is holomorphic in D. (2) $u=u(x,y)$ and $v=v(x,y)$ are †totally differentiable at each point $z=x+iy$ and satisfy the **Cauchy-Riemann differential equations**

$$\partial u/\partial x=\partial v/\partial y, \qquad \partial u/\partial y=-\partial v/\partial x.$$

(3) f is represented by a †power series $\sum_{n=0}^{\infty}c_n(z-a)^n$ in a neighborhood of each point a of D; that is, $f(z)$ is **analytic** in D. (4) f is continuous and $\int_C f(z)\,dz=0$ for every rectifiable Jordan closed curve C whose interior is contained, together with C, in D. The proposition that (1) implies (4) is called **Cauchy's integral theorem**, and the proposition that (4) implies (1) is called **Morera's theorem**.

The hypothesis of Morera's theorem can be weakened as follows: Let $f(z)$ be continuous in a domain D. If $\int_C f(z)\,dz=0$ for every rectangle C in D with sides parallel to the axes and whose interior consists of only points of D, then $f(z)$ is holomorphic in D. In the statement of this theorem, if we let C be an arbitrary circle, we get the same conclusion.

The following complex differential operators are often useful:

$$\frac{\partial}{\partial z}=\frac{1}{2}\left(\frac{\partial}{\partial x}-i\frac{\partial}{\partial y}\right), \qquad \frac{\partial}{\partial \bar z}=\frac{1}{2}\left(\frac{\partial}{\partial x}+i\frac{\partial}{\partial y}\right).$$

Generally, $\overline{(\partial\varphi/\partial z)}=\partial\bar\varphi/\partial\bar z$. The Cauchy-Riemann equations above can be expressed in a single equation : $\partial f/\partial\bar z=0$. If f is holomorphic, $\partial f/\partial z=f'(z)$.

In order to show that $f=u+iv$ is holomorphic in D, assumption (2) can be weakened. Actually, we have the **Looman-Men'shov theorem**: Suppose that u and v are continuous in D, $\partial u/\partial x$, $\partial u/\partial y$, $\partial v/\partial x$, and $\partial v/\partial y$ exist at every point of D except for at most a countable number of points, and the Cauchy-Riemann equations hold in D except for a set of 2-dimensional †measure zero; then $f=u+iv$ is holomorphic in D. D. E. Men'shov extended this theorem and obtained various conditions for holomorphy. For example, he proved the following theorem: If f is a topological mapping of D and f is conformal in D (i.e.,

$$\lim_{h\to 0}\arg(f(z+h)-f(z))/h$$

exists) except for at most a countable number of points, then f is holomorphic in D.

As another type of sufficient condition for holomorphy, we have the proposition : If f is locally †Lebesgue integrable in D and satisfies the Cauchy-Riemann equation $f_{\bar{z}}=0$ in the sense of †distribution, then there exists a holomorphic function g in D such that $g=f$ †almost everywhere.

B. Cauchy's Integral Theorem

Cauchy's integral theorem can be stated as follows: If $f(z)$ is a holomorphic function in a †simply connected domain D in the complex plane, the equality $\int_C f(z)dz=0$ holds for every (rectifiable) closed curve C in D. In particular, the integral $\int_\alpha^\beta f(z)dz$ $(\alpha,\beta\in D)$ is uniquely determined by α and β provided that its path of integration lies in D. The function $F(z)=\int_\alpha^z f(\zeta)d\zeta$ $(z\in D)$ is called the **indefinite integral** of f. $F(z)$ is holomorphic in D and $F'(z)=f(z)$.

In the proof of his integral theorem, Cauchy assumed the existence and continuity of the derivative $f'(z)$ in D. However, E. Goursat proved the theorem utilizing only the existence of $f'(z)$. Actually, by virtue of the integral formula (2) in this section, the existence of $f'(z)$ in D implies the continuity of $f'(z)$. This is sometimes called **Goursat's theorem**.

Let $C, C_1, C_2, \ldots, C_n$ be rectifiable Jordan curves. Suppose that $C_1, C_2, \ldots, C_n$ are in the interior of C and that each one lies in the exterior of the others. If $f(z)$ is holomorphic in the region D bounded by these $n+1$ curves and continuous on $D\cup C\cup C_1\cup\ldots\cup C_n=\bar{D}$, then we have

$$\int_C f(z)dz=\sum_{k=1}^{n}\int_{C_k} f(z)dz.$$

Here the curvilinear integrals are taken in the **positive direction** (i.e., we take the direction such that $\int_C(z-a)^{-1}dz=\int_{C_k}(z-a)^{-1}dz=2\pi i$ for a point a in the interior of C or C_k, respectively). Henceforth, an integral along a closed curve is taken in the positive direction unless otherwise noted. Cauchy's integral theorem under the assumption that f is holomorphic in D and is continuous on $\bar{D}$ is sometimes called the **stronger form of Cauchy's integral theorem**.

Under the same assumptions as in the stronger form, we have **Cauchy's integral formula** for $z\in D$:

$$f(z)=\frac{1}{2\pi i}\int_C\frac{f(\zeta)}{\zeta-z}d\zeta-\frac{1}{2\pi i}\int_{C_1}\frac{f(\zeta)}{\zeta-z}d\zeta\ldots-\frac{1}{2\pi i}\int_{C_n}\frac{f(\zeta)}{\zeta-z}d\zeta. \tag{2}$$

This integral formula expresses the value of $f(z)$ at a point z in the domain D in terms of the values of f on the boundary of D. In particular, when $n=0$, the integral formula reads as

$$f(z)=\frac{1}{2\pi i}\int_C\frac{f(\zeta)}{\zeta-z}d\zeta. \tag{3}$$

Furthermore, if C is a circle $|z|=R$ (i.e., D is the disk $|z|<R$), we obtain **Poisson's integral formula**:

$$f(z)=\frac{1}{2\pi}\int_0^{2\pi}f(Re^{i\varphi})\frac{R^2-r^2}{R^2+r^2-2Rr\cos(\theta-\varphi)}d\varphi,$$

$$f(z)=\frac{1}{2\pi}\int_0^{2\pi}\operatorname{Re}f(Re^{i\varphi})\frac{Re^{i\varphi}+z}{Re^{i\varphi}-z}d\varphi+i\operatorname{Im}f(0),$$

$$z=re^{i\theta},\quad 0\leqslant r<R. \tag{3$'$}$$

Formula (3)′ is valid for a †harmonic function.

Let C be a rectifiable curve and $f(\zeta)$ be a continuous function defined on C; then the **integral of Cauchy type**

$$F(z)=\frac{1}{2\pi i}\int_C\frac{f(\zeta)}{\zeta-z}d\zeta$$

is holomorphic outside C. The nth derivative $F^{(n)}$ of F is given by $(n!/2\pi i)\int_C f(\zeta)/(\zeta-z)^{n+1}d\zeta$; moreover, F can be expanded in a Taylor series about every point a outside C:

$$F(z)=\sum_{n=0}^{\infty}a_n(z-a)^n,\qquad a_n=\frac{F^{(n)}(a)}{n!},$$

which converges in $|z-a|<\rho$, ρ being the distance from a to C. In particular, formula (3) implies that a holomorphic function f is infinitely differentiable and is expanded in a Taylor series about every point of D as above.

Let C be a closed curve not passing through a point a. Then the integral $(1/2\pi i)\int_C dz/(z-a)$ is an integer. It is called the **winding number** of C about a and is denoted by $n(C;a)$. A cycle γ (a finite sum of oriented closed curves) in an open set D is said to be **homologous to zero** in D if $n(\gamma;a)=0$ for all points a in the complement of $\bar{D}$. The general form of Cauchy's integral theorem is stated as follows. If f is holomorphic in D, then $\int_\gamma f(z)dz=0$ for every cycle γ which is homologous to zero in D (E. Artin). From this we have the general form of Cauchy's integral formula: if f is holomorphic in a domain D, then

$$n(\gamma;z)\cdot f(z)=\frac{1}{2\pi i}\int_\gamma\frac{f(\zeta)}{\zeta-z}dz,\qquad z\in D-\gamma,$$

for every cycle γ which is homologous to zero in D.

C. Zero Points

Let $f(z)$ be a holomorphic function not identically equal to zero. If $f(a)=0$, we call a a **zero point** of f. Every zero point of f is an isolated point, and there exists a unique positive integer k and a function h holomorphic at a such that

$$f(z)=(z-a)^k g(z), \qquad g(a)\neq 0. \tag{4}$$

We call k the **order of the zero point** a and a a **zero point of the** k**th order**. The equality (4) implies that the †Taylor series of $f(z)$ at a begins with the term $c_k(z-a)^k$. Suppose that a is a zero point of $f(z)-\gamma$ of the kth order; then we call a a γ**-point of the** k**th order**.

For a function $f(z)$ defined in a neighborhood of the †point at infinity, we set $f(1/w)=g(w)$ $(f(\infty)=g(0))$ and call f holomorphic at ∞ if g is holomorphic at 0; f is said to have a zero of **order** k at ∞ if g has a zero of order k at 0. If two functions f and g are holomorphic in D and $f(z)=g(z)$ on a subset E that has an †accumulation point in D, then f is identically equal to g in D (**theorem of identity** or **uniqueness theorem**) since the zeros of holomorphic functions must be isolated.

D. Isolated Singularities

Let $f(z)$ be holomorphic in an annulus $D=\{z \mid R_1<|z-a|<R_2, 0\leqslant R_1<R_2\leqslant+\infty\}$. Then $f(z)$ is expanded in the †Laurent series

$$f(z)=\sum_{n=-\infty}^{+\infty} c_n(z-a)^n. \tag{5}$$

This is called the **Laurent expansion** of f about a. The coefficients c_n are given by $c_n=(1/2\pi i)\int_C f(\zeta)d\zeta/(\zeta-a)^{n+1}$ with $C=\{z \mid |z-a|=r\}$, $R_1<r<R_2$. In particular, if $f(z)$ is holomorphic in $D=\{z \mid 0<|z-a|<R\}$ (or, if $a=\infty$, in $D=\{z \mid R<|z|<+\infty\}$) but not holomorphic in $D\cup\{a\}$, we call a an **isolated singular point** (or **isolated singularity**) of f. By utilizing the †local canonical parameter $t=z-a$ (or $t=1/z$ for $a=\infty$), the Laurent expansion (5) of f is then written as $f(z)=\sum_{n=-\infty}^{-1} c_n t^n+\sum_{n=0}^{+\infty} c_n t^n$. The second sum is an ordinary power series, called the **holomorphic part** of $f(z)$. The first sum is a power series of $1/t$ with no constant term, called the **singular part** of $f(z)$ at a or the **principal part** of the singularity (or of the Laurent expansion at a).

If we have $\lim_{t\to 0} tf(z)=0$, the Laurent expansion (5) of $f(z)$ lacks its singular part, and the limit of $f(z)$ exists as $t\to 0$ $(z\to a)$ and is equal to c_0. If we set $f(a)=c_0$, then the function $f(z)$ is holomorphic in $D\cup\{a\}$. In this case, the point a is called a **removable singularity**. If $f(z)$ is bounded in a neighborhood of a singularity a, then a is removable (**Riemann's theorem**). Usually, we assume that the removable singularities of a function have already been removed in this way.

When the singular part of $f(z)$ at a exists and consists of a finite number of terms, the point a is called a **pole**; when it consists of an infinite number of terms, the point is called an **essential singularity**. If a is a pole, $f(z)$ is represented by the Laurent series $\sum_{n=-k}^{\infty} c_n t^n$ $(c_{-k}\neq 0)$ and $f(z)\to\infty$ as $z\to a$. In this case, the index k is called the **order of the pole** a. Then a relation such as (4) holds, where the index k is replaced by $-k$. Hence the point a is sometimes called a **zero point of the** $-k$**th order**. If a is an essential singularity, then for an arbitrary number c there exists a sequence z_n converging to a such that $\lim_{n\to\infty} f(z_n)=c$ (the **Casorati-Weierstrass theorem** or simply **Weierstrass's theorem**). Related to Weierstrass's theorem, we have †Picard's theorem, which gives a detailed description of the behavior of a function around its singularities.

E. Residues

Let $a(\neq\infty)$ be an isolated singularity of $f(z)$. Then the coefficient c_{-1} of $(z-a)^{-1}$ in the Laurent expansion (5) of $f(z)$ is called the **residue** of $f(z)$ at a and is denoted by $\mathrm{Res}[f]_a$, $R(a;f)$, or $R(a)$ if we need not indicate f. We have

$$R(a)=c_{-1}=\frac{1}{2\pi i}\int_{|\zeta-a|=r} f(\zeta)\,d\zeta,$$

where the integral is taken in the positive direction along a path for $0<r<R$. If $f(z)$ is holomorphic at $z=a$, then $R(a)=0$. If $f(z)$ has a pole of the first order at a,

$$R(a)=\lim_{z\to a}(z-a)f(z).$$

The residue at the point at infinity is defined to be $-a_{-1}$, where a_{-1} is the coefficient of $1/z$ of the Laurent expansion of $f(z)$ at $\infty: f(z)=\sum_{n=-\infty}^{\infty} a_n z^n$, and we have

$$-a_{-1}=\frac{-1}{2\pi i}\int_{|\zeta|=r} f(\zeta)\,d\zeta, \qquad R^{-1}<r<+\infty.$$

Thus the notion of the residue of $f(z)$ is actually related to the differential form $f(z)dz$ and not to $f(z)$ itself.

From the first formula in this section and the formula for $-a_{-1}$, the **residue theorem** follows (Cauchy, 1825): Let C be a rectifiable Jordan curve in the complex plane. Let $a_1,\ldots,a_m$ be a finite number of points inside C, and let D be a domain containing C and its

interior. If $f(z)$ is a function holomorphic in $D - \{a_1, \ldots, a_m\}$, we have

$$\frac{1}{2\pi i}\int_C f(z)\,dz = \sum_{n=1}^{m} R(a_n).$$

Furthermore, if $f(z)$ is holomorphic in the extended complex plane (including the point at infinity) except for a finite number of poles, the sum of all residues is equal to zero.

F. Calculus of Residues

The **calculus of residues** is a field of calculus based on application of the notion of residues. For example, we have methods for the calculation of definite integrals. Actually, one of the reasons why Cauchy studied the theory of complex functions was that he believed that the theory would provide a unified method of computing definite integrals. For example, if $\varphi(z)$ is a rational function without poles on the real axis and with a zero point at infinity whose order is at least 2, then we have

$$\int_{-\infty}^{\infty} \varphi(x)\,dx = 2\pi i \sum_{\operatorname{Im}\alpha>0} R(\alpha; \varphi(z)), \tag{6}$$

$$\int_{-\infty}^{\infty} e^{ix}\varphi(x)\,dx = 2\pi i \sum_{\operatorname{Im}\alpha>0} R(\alpha; e^{iz}\varphi(z)). \tag{7}$$

Here the sums are taken over all the poles in the upper half-plane. Formula (7) is valid also for a rational function $\varphi(z)$ with a simple zero at infinity. If $\varphi(z)$ has simple poles at a_k $(k = 1, \ldots, n)$ on the real axis, then we take the principal values of the integrals at those poles and add $\pi i R(a_k)$ $(k = 1, \ldots, n)$ to the terms on the right-hand side of (6) and (7). Sometimes we use the residue theorem to obtain the value of the sum of a series (e.g., the †Gaussian sum) by expressing it as an integral.

Let $f(z)$ be a single-valued function that is †meromorphic and not identically equal to zero in a domain D, and let $\varphi(z)$ be a function holomorphic in D. Draw a rectifiable Jordan curve C in D such that the interior of C is contained in D and $f(z)$ has neither zeros nor poles on C. Let $\alpha_1, \ldots, \alpha_N$ and $\beta_1, \ldots, \beta_P$ be the zeros and poles inside C, respectively (where each of them is repeated as often as its order). Then we have

$$\sum_{n=1}^{N} \varphi(\alpha_n) - \sum_{p=1}^{P} \varphi(\beta_p) = \frac{1}{2\pi i}\int_C \varphi(z)\frac{f'(z)}{f(z)}\,dz.$$

If $\varphi(z) = 1$, we get

$$\frac{1}{2\pi}\int_C d\arg f(z) = N - P.$$

This is called the **argument principle**. Next, let $f(z)$ be a function meromorphic for $|z| < R \leqslant +\infty$, $f(0) \neq 0$, $\neq \infty$, and set $\varphi(z) = \log z$. Take C as a closed curve consisting of the boundary of an annulus $0 < \rho < |z| < r < R$ (where ρ is sufficiently small) and two sides of a suitable †crosscut joining a point of $|z| = \rho$ and $|z| = r$. Then we obtain **Jensen's formula**:

$$\log\left|\frac{\alpha_1, \alpha_2, \ldots, \alpha_N}{\beta_1, \beta_2, \ldots, \beta_P}\right| = \log|f(0)| + (N-P)\log r$$
$$-\frac{1}{2\pi}\int_0^{2\pi} \log|f(re^{i\psi})|\,d\psi.$$

The argument principle can be utilized to prove **Rouché's theorem**: Let $f(z)$ and $g(z)$ be functions holomorphic in a domain D that contains a rectifiable Jordan curve C and its interior. Suppose that $f(z) + \lambda g(z)$ never vanishes on C for any λ with $0 \leqslant \lambda \leqslant 1$. Then the number of zeros of $f(z)$ in the interior of C is equal to that of $f(z) + g(z)$. If $|f(z)| > |g(z)|$ on C or $\arg f(z) - \arg g(z) \neq (2n+1)\pi$ (n is an integer), the hypothesis of Rouché's theorem is satisfied. This theorem is useful in proving the existence of a zero of a complex function (for example, a polynomial) and in finding its position.

G. Analytic Continuations

Let $f(z)$ be a holomorphic function in a domain D of the complex plane $\mathbf{C}$ and D^* be a domain containing D as a proper subset. If there exists a function $F(z)$ holomorphic in D^* that coincides with $f(z)$ in D, then $F(z)$ is called an **analytic continuation** (or **analytic prolongation**) of $f(z)$ from D to D^*. By the theorem of identity an analytic continuation of $F(z)$ is uniquely determined if it exists.

The function $f_1(z)$ defined by the power series $P(z;a) = \sum_{n=0}^{\infty} a_n(z-a)^n$ with the radius of convergence $r_1 > 0$ is holomorphic in the domain $D_1 : |z-a| < r_1$, and at a point b of D_1 it can be expanded into a power series $P(z;b) = \sum_{n=0}^{\infty} b_n(z-b)^n$ with the radius of convergence r_2 $(\geqslant r_1 - |b-a|)$. If $r_2 > r_1 - |b-a|$, the domain $D_2 : |z-b| < r_2$ is not entirely contained in D_1. Let $f_2(z)$ be the function defined in D_2 by $P(z;b)$. Then the function $F(z)$ that is equal to $f_1(z)$ in D_1 and to $f_2(z)$ in D_2 is an analytic continuation of $f_1(z)$ from D_1 to $D_1 \cup D_2$ (a **direct analytic continuation** by power series).

We have the following classical theorems about analytic continuations:

Let D_1 and D_2 be two disjoint domains, and suppose that their respective boundaries C_1 and C_2 are †rectifiable simple closed curves and that the intersection of C_1 and C_2 contains an open arc Γ. If two holomorphic functions $f_1(z)$, $f_2(z)$ defined in D_1 and D_2, respectively, have finite common †boundary values at every

point of Γ, then there exists an analytic continuation $F(z)$ of $f_1(z)$ and $f_2(z)$ to $D_1 \cup \Gamma \cup D_2$ (**Painlevé's theorem**). We sometimes call $f_2(z)$ a continuation of $f_1(z)$ beyond Γ. If Γ is not rectifiable, the continuation beyond Γ does not exist, in general.

Let $f(z)$ be holomorphic in a †Jordan domain D lying in the half-plane on one side of the real axis and containing an open interval I of the real axis in its boundary. If $f(z)$ has finite real boundary values at every point of I, then it can be continued analytically beyond I to the other side of the real axis; there the continued function is given by $\overline{f(\bar{z})}$ (**Schwarz's principle of reflection**). This theorem can be generalized to the case where the real interval is replaced by an †analytic curve.

A **harmonic continuation** of †harmonic functions is defined analogously to analytic continuation. Let D be a Jordan domain lying in the half-plane on one side of the real axis and having an open interval I on the real axis as a part of its boundary. If $u(z)$ is harmonic in D and has the boundary value 0 at every point of I, then $u(z)$ has a harmonic continuation beyond I.

For other properties of holomorphic functions → 43 Bounded Functions; 429 Transcendental Entire Functions.

H. Analytic Functions

A real-valued function $f(t)$ of a real variable t is said to be **analytic** at $t=t_0$ if it can be represented by a †power series in $t-t_0$ in a neighborhood of t_0 in **R**. If $f(t)$ is defined on an open set of **R** at every point of which it is analytic, then $f(t)$ is called an **analytic function**, or, more precisely, a **real analytic function**.

Analogously, a complex-valued function $f(z)$ of a complex variable z defined on a †domain D of the complex plane **C** is said to be **analytic** at $z=z_0$ ($\in D$) if it can be represented by a power series in $z-z_0$ in a neighborhood of z_0 in **C**, and $f(z)$ is an **analytic function** in D if it is analytic at every point of D. In the remainder of this article, we are concerned with analytic functions in this sense. To distinguish them from the real case, they are also called **complex analytic functions**. A complex analytic function $f(z)$ defined on D is †differentiable in D; therefore it is †holomorphic in D. The converse is also true. Thus the term "analytic function" is synonymous with "holomorphic function" insofar as it concerns a complex function (i.e., a complex-valued function of a complex variable) on a domain, but in the theory of functions it takes on an additional meaning that is explained in the following section.

I. Analytic Functions in the Sense of Weierstrass

Let a be a point of the †z-sphere and t the †local canonical parameter at a; i.e., $t=z-a$ if $a\neq\infty$ and $t=z^{-1}$ if $a=\infty$. If a power series $P(z;a)=\sum_{n=0}^{\infty}c_nt^n$ has a positive radius r_a of convergence, we call $P(z;a)$ a **function element** with center a on the z-sphere, after K. Weierstrass. $P(z;a)=\sum_{n=0}^{\infty}c_n(z-a)^n$ if $a\neq\infty$, and $P(z;a)=\sum_{n=0}^{\infty}c_nz^{-n}$ if $a=\infty$. These represent a holomorphic function in $|z-a|<r_a$ or in $r_a^{-1}<|z|\leqslant\infty$, respectively. If b is a point inside the circle of convergence of the function element $P(z;a)$, by the †Taylor expansion of $P(z;a)$ at $z=b$, we obtain the power series $P(z;b)$ in $z-b$, which is a direct analytic continuation of $P(z;a)$. Let a and b be two points on the z-sphere, and let $C: z=z(s)$ $(0\leqslant s\leqslant 1, z(0)=a, z(1)=b)$ be a curve joining a and b. We say that $P(z;a)$ is **analytically continuable** along C and that we obtain $P(z;b)$ at the end point b by the **analytic continuation of** $P(z;a)$ **along** C if the following two conditions are satisfied: (i) To every $s\in[0,1]$ there corresponds a function element $P(z;z(s))$ with center $z(s)$; (ii) for every $s_0\in[0,1]$, we can take a suitable subarc $z=z(s)$ $(|s-s_0|\leqslant\varepsilon, \varepsilon>0)$ of C contained inside the circle of convergence of $P(z;z(s_0))$ such that every function element $P(z;z(s))$ with $|s-s_0|\leqslant\varepsilon$ is a direct analytic continuation of $P(z;z(s_0))$. When $P(z;a)$ and the curve C are given, the analytic continuation along C is uniquely determined (**uniqueness theorem of the analytic continuation**).

Given a function element $P(z;a)$ with center a, the set of all function elements obtained by every possible analytic continuation along every curve starting from a is called an **analytic function in the sense of Weierstrass** determined by $P(z;a)$. In this definition, we can restrict the curves to polygonal lines. An analytic function in this sense is completely determined by a single arbitrary function element belonging to it, so two analytic functions are identically equal if they have a common function element.

A †germ of a holomorphic function is identical to a function element, and the set of all germs has the natural structure of a †sheaf $\mathcal{O}$. In the terminology of sheaves, an analytic function is a connected component of $\mathcal{O}$, and an **analytic continuation along a curve** C is a continuous curve Γ in $\mathcal{O}$ whose projection is C.

J. Values and Branches of Analytic Functions

The value of an analytic function at a point b is, by definition, the value at b of its function elements with center b (whose existence is

assumed; there may be several such elements). An analytic function is, in general, a multiple-valued function because analytic continuations along different curves with the same end points may lead to different function elements. For a given analytic function $f(z)$, if the maximal number of its function elements with the same center is n, we say it is **n-valued**, and if $n \geqslant 2$ we say it is **multiple-valued** (or **many-valued**). The number of function elements of $f(z)$ with the same center is at most †countably infinite, so the value of $f(z)$ at a point is a countable set (**Poincaré-Volterra theorem**). By introducing a †Riemann surface instead of the complex plane as the domain of definition of an analytic function, we can regard multiple-valued analytic functions as single-valued functions defined on a suitable Riemann surface (→ 367 Riemann Surfaces).

Let $f(z)$ be an analytic function and $P(z;a)$ be a function element belonging to $f(z)$, where a is a point of a domain D. The set of all function elements obtained from $P(z;a)$ by every possible analytic continuation along all curves in D is called a **branch** of $f(z)$ in D determined by $P(z;a)$. When D coincides with the whole complex plane, the branch of $f(z)$ in D is the function $f(z)$ itself. A function holomorphic in a domain D can be expanded in a power series with any point of D as its center, and the set of these power series (function elements) constitutes a branch of an analytic function.

If analytic continuations of a function element are possible along all curves in D, then the analytic continuations along two †homotopic curves in D lead to the same result (**monodromy theorem**). In particular, if D is †simply connected and if analytic continuations of $P(z;a)$ are possible along all curves in D starting from a, then the branch of $f(z)$ in D determined by $P(z;a)$ is single-valued.

K. Invariance Theorem of Analytic Relations

Suppose the following four conditions hold: (1) $F(z,w)$ is a holomorphic function of two variables for $z \in \Delta_1$ and $w \in \Delta_2$, where Δ_1, Δ_2 are domains in the complex plane. (2) A curve C: $z=z(s)$ $(0 \leqslant s \leqslant 1, z(0)=a, z(1)=b)$ and two sets of function elements $P(z;z(s))$ and $Q(z;z(s))$ defined for every s $(0 \leqslant s \leqslant 1)$ are given. (3) $P(z;a)$ and $Q(z;a)$ can be continued analytically along C using $P(z;z(s))$ and $Q(z;z(s))$, respectively. (4) There exists a positive number $R(s)$ for every s $(0 \leqslant s \leqslant 1)$ such that, if $|z-z(s)| < R(s)$, the values of $P(z;z(s))$ and $Q(z;z(s))$ belong to Δ_1 and Δ_2, respectively. Under these conditions, if $F(P(z;a), Q(z;a))=0$ holds for $|z-a|<R(0)$, then $F(P(z;b), Q(z;b))=0$ holds for $|z-b|<R(1)$. In other words, an analytic relation between function elements belonging to two analytic functions that holds in a neighborhood of the starting point of a curve C is conserved for function elements with center at the terminal point b of C. This is called the **invariance theorem of analytic relations**. The same statement is valid for relations among more than two analytic functions and their derivatives (differential equations).

L. Inverse Functions

Suppose that $P(z;a)$ $(a \neq \infty)$ belongs to an analytic function $f(z)$ and $P'(a;a) \neq 0$. We consider the inverse function of $P(z;a)$ in a neighborhood of a and let $\mathfrak{B}(w;\alpha)$ $(\alpha = P(a;a))$ be its expansion as the power series in $w-\alpha$. We call $\mathfrak{B}(w;\alpha)$ the **inverse function element** (or simply **inverse element**) of $P(z;a)$ and the analytic function determined by $\mathfrak{B}(w;\alpha)$ the **inverse analytic function** (or simply **inverse function**) of $f(z)$. The inverse function is completely determined by $f(z)$ and is independent of the choice of $P(z;a)$. For example, analytic functions represented by $\sqrt{w}$ or $\log w$ are defined as the inverse function of z^2 or e^z, respectively.

M. Singularities of Analytic Functions

Hereafter, when we speak of a curve $C: z=z(s)$ $(0 \leqslant s \leqslant 1)$, it is always supposed that C is a curve in the complex plane starting at a and ending at ω. Let K_r be the open disk $|z-\omega|<r$; we denote by C_r the connected component of $C \cap K_r$ that contains ω. If analytic continuations of $P(z;a)$ are possible along any subarc of C with a terminal point arbitrarily near ω but impossible along the whole C, we say that the analytic continuation of $P(z;a)$ along C defines a **singularity** Ω of the coordinate ω, and that Ω lies over ω. For example, if $P(z;a)$ has a finite radius of convergence, for a suitable point ω on the circle of convergence the analytic continuation of $P(z;a)$ along the radius $a\omega$ defines a singularity over ω. Now take a point z_r on C_r and denote by $F_r(z)$ the branch of an analytic function determined by $P(z;z_r)$ in K_r. Let Ω be a singularity determined by C and $P(z;a)$, and suppose that we are given another singularity Ω^* over ω determined by C^* and $P^*(z;a^*)$. If they define the same branch $F_r(z)$ for every K_r, by definition, we put $\Omega = \Omega^*$. Thus $F_r(z)$ defines an †unramified covering surface W_r *of* $K_r - \{\omega\}$, and it is single-valued on W_r.

Singularities are classified according to the geometric structure of W_r and the value distribution of $F_r(z)$ on it. First, if W_r has no †relative boundary over $0<|z-\omega|<r$ for a suitable r, then Ω is called an **isolated singularity** of the

analytic function. In this case, the number k of points of W_r lying over a point z in $K_r - \{\omega\}$ is constant. If $k=\infty$, W_r has a †logarithmic branch point over ω, and Ω is called a **logarithmic singularity**. If $k<\infty$, $F_r(z)$ can be represented as a single-valued holomorphic function in $0<|t|<r^{1/k}$ by putting $z=\omega+t^k$. In this case, if we introduce an additional point P_0 corresponding to $z=\omega$, then $W_r \cup \{P_0\}$ has only an †algebraic branch point over ω. Now, taking into account the value of $w=F_r(z)$, we call Ω an **algebraic singularity** if $\lim w$ exists. In this case, we have $F_r(z)=\sum_{n=l}^{\infty} c_n t^n$, and if we admit analytic continuations in the wider sense (→ Section O), $P(z;a)$ is analytically continuable along the whole C.

N. The Natural Boundary

Given a domain D and an analytic function $f(z)$ holomorphic in D, if all boundary points of D are singularities of $f(z)$ and $f(z)$ is not continuable to the exterior of D, the boundary of D is called the **natural boundary** of $f(z)$. This phenomenon was first discovered for †elliptic modular functions. Many results are known about power series for which the circle of convergence is the natural boundary (→ 339 Power Series). For any given domain D in $\mathbf{C}$, there exists an analytic function whose natural boundary is the boundary of D. The original proof of this fact, given by Weierstrass, contained a defect that was corrected by J. Besse.

O. Analytic Continuation in the Wider Sense

Let two †Laurent series (with parameter t) $z=P(t)=\sum_{n=k}^{\infty} a_n t^n$ and $w=Q(t)=\sum_{n=l}^{\infty} b_n t^n$ (k and l are integers, and $a_k b_l \neq 0$) converge in $0<|t|<r$, and let $(P(t_1), Q(t_1)) \neq (P(t_2), Q(t_2))$ if $t_1 \neq t_2$; then we say that the pair (P, Q) defines a **function element in the wider sense**. If a change of parameter $\tau = r_1 t + r_2 t^2 + \ldots$ ($r_1 \neq 0$ and the radius of convergence >0) gives $P(t)=\Pi(\tau)$, $Q(t)=K(\tau)$, we say that (Π, K) and (P, Q) define the same function element. By a suitable choice of parameter, any function element can be given in the form $z=t^k+a$ (or $a=t^{-k}$), $w=\sum_{n=l}^{\infty} b_n t^n$, and the elimination of t gives the representation of w as a †Puiseux series of z. So if $k=1$ and $l \geqslant 0$, it reduces to a holomorphic function element. When $k=1$, with $l<0$ not excluded, the above element is called a **rational element**. If $k>1$ it is called a **ramified element**, and if $l<0$ it is called a **polar element**.

If P', Q' are the direct analytic continuations of P and Q at t_0 ($0<|t_0|<r$), i.e., their Taylor expansions at t_0, the function element (P', Q') is called a direct analytic continuation of (P, Q), which is also considered its own direct analytic continuation. For a fixed r, the set of all direct continuations of (P, Q) thus obtained is called an **analytic neighborhood** of (P, Q), and these neighborhoods define a topology in the set of all function elements. A curve in this topological space is called an **analytic continuation in the wider sense**, and a †connected component of this space is called an **analytic function in the wider sense**. An analytic function in the wider sense is a set of function elements in the wider sense, but it can also be regarded as a function $w=f(z)$ (with an independent variable z and a dependent variable w) defined by each function element $p(z,w): z=P(t)$, $w=Q(t)$.
An analytic continuation in the wider sense,

$$p(s)=p(z,w;s);$$

$$z=z(s)+t^{k(s)}, \qquad w=\sum_{n=l(s)}^{\infty} c_n(s) t^n, \qquad 0 \leqslant s \leqslant 1,$$

is sometimes called an analytic continuation along the curve $C: z=z(s)$ $(0 \leqslant s \leqslant 1)$ in the complex plane. If all $p(s)$ are holomorphic function elements, this coincides with the analytic continuation along C in the original sense, but if this is not the case, $p(0)$ and C do not necessarily determine $p(1)$ uniquely. Actually, an analytic function in the wider sense is just an analytic function in the original sense with at most a countable number of ramified or polar elements added.

P. Singularities of Analytic Functions in the Wider Sense

Suppose the following three conditions hold: (1) For every point on C except ω, that is, for $z(s)$ $(0 \leqslant s < 1)$, a function element in the wider sense $p(z,w;s)$ is given. (2) For every λ (<1), $p(z,w;s)$ $(0 \leqslant s \leqslant \lambda)$ constitutes an analytic continuation in the wider sense. (3) It is impossible to find a function element $p(z,w;1)$ such that $p(z,w;s)$ $(0 \leqslant s \leqslant 1)$ is an analytic continuation in the wider sense. When these three conditions are satisfied, we say that $p(z,w;s)$ $(0 \leqslant s < 1)$ defines a **transcendental singularity** Ω with ω as its coordinate. The method of determining a branch $w=F_r(z)$ in an open disk with center ω is completely parallel to the case of holomorphic analytic functions. Because of the appearance of function elements in the wider sense in $F_r(z)$, the covering surface W_r of K_r defined by $F_r(z)$ may have algebraic branch points. If W_r has a logarithmic branch point over ω, Ω is called a **logarithmic singularity**. If W_r has no point over ω for suitable r, Ω is called a **direct transcendental singularity**; otherwise, it is called an **indirect transcendental singularity**. The logarithmic singularities are

direct singularities. The inverse function of $z = w\sin w$ has a direct singularity over $z=\infty$ that is not logarithmic, and the inverse function of $z=(\sin w)/w$ has an indirect singularity over $z=0$. Taking into account the value of $w=F_r(z)$, if the †cluster set of F_r at $\Omega: S_\Omega = \bigcap_{r>0}\overline{\{F_r(z)\}}$ consists of only one point, it is an **ordinary singularity**; if not, it is an **essential singularity**.

Q. History

A function of a complex variable is **monogenic** in the sense of A. L. Cauchy if it is differentiable at every point of its domain of definition. It was B. Riemann who succeeded in developing Cauchy's concept. Riemann considered an analytic function as a function defined on a †Riemann surface, that is, a 1-dimensional complex analytic manifold. On the other hand, Weierstrass constructed the theory of analytic functions starting from power series. When we speak of single-valued functions defined in a domain of the complex plane, the monogenic functions of Cauchy and the analytic functions of Weierstrass are identical. Although the analytic functions are very special functions, the study of complex analytic functions is usually called the theory of functions of a complex variable, or simply the theory of functions.

By considering the following point set C, which is more general than a domain, E. Borel showed that a monogenic function on C is not necessarily holomorphic in the ordinary sense. Take a countable dense subset in a subdomain D' of a domain D and a double sequence of positive numbers $\{r_n^{(h)}\}$. Put $S_n^{(h)}=\{z \mid |z-z_n|< r_n^{(h)}\}$ and $C^{(h)}=D-\bigcup_{n=1}^{\infty} S_n^{(h)}$. By a suitable choice of $r_n^{(h)}$, we can suppose that the $\mathbf{C}^{(h)}$ are connected and monotone increasing with respect to h. Put $C=\bigcup_{h=1}^{\infty} C^{(h)}$. A function defined in C is by definition **monogenic** if it is differentiable in $C^{(h)}$ for every h. For such a monogenic function, Cauchy's †integral formula in a generalized form holds, and the function is infinitely differentiable. If $f(z)$ and $g(z)$ are monogenic in C and coincide on a curve in C, then they are identical in C. Let D be the set $\{z \mid 0<\operatorname{Re} z<1, 0<\operatorname{Im} z<1\}$ and $\{z_n\}$ be all rational points in D $(z_n=(p+iq)/m)$. For a natural number h, we define $C^{(h)}$ to be the set D minus the union of open disks with radius $\exp(-e^{m^2})/h$ and center $(p+iq)/m$. The function

$$f(z)=\sum_{m=1}^{\infty}\sum_{p=0}^{m}\sum_{q=0}^{m}\frac{\exp(-e^{m^4})}{z-(p+iq)/m}$$

is monogenic in C in the above-mentioned sense, but not holomorphic in C. The study of these functions developed into the theory of †quasi-analytic functions.

The concept of †analytic functions of several complex variables can also be defined analogously to the case of one variable. Then nonuniformizable singularities appear that lead to a generalization of the concept of †manifolds (→ 23 Analytic Spaces).

References

[1] E. Borel, Leçons sur les fonctions monogènes uniformes d'une variable complexe, Gauthier-Villars, 1917.
[2] A. Hurwitz and R. Courant, Volesungen über allgemeine Funktionentheorie und elliptische Funktionen, Springer, fourth edition, 1964.
[3] L. Bieberbach, Lehrbuch der Funktionentheorie, Teubner, I, 1921; II, 1927 (Johnson Reprint Co., 1969).
[4] E. C. Titchmarsh, The theory of functions, Oxford Univ. Press, second edition, 1939.
[5] C. Carathéodory, Funktionentheorie I, II, Birkhäuser, 1950; English translation, Theory of functions, Chelsea, I, 1958; II, 1960.
[6] S. Saks and A. Zygmund, Analytic functions, Warsaw, 1952.
[7] L. V. Ahlfors, Complex analysis, McGraw-Hill, 1953; third edition, 1979.
[8] H. Behnke and F. Sommer, Theorie der analytischen Funktionen einer komplexen Veränderlichen, Springer, second edition, 1962.
[9] H. Kneser, Funktionentheorie, Vandenhoeck & Ruprecht, 1958.
[10] E. Hille, Analytic function theory, Ginn, I, 1959; II, 1962.
[11] H. P. Cartan, Théorie élémentaire des fonctions analytiques d'une ou plusieurs variables complexes, Hermann, 1961; English translation, Elementary theory of analytic functions of one or several complex variables, Addison-Wesley, 1963.
[12] A. Dinghas, Vorlesungen über Funktionentheorie, Springer, 1961.
[13] R. Nevanlinna and V. Paatero, Einführung in die Funktionentheorie, Birkhäuser, 1964; English translation, Introduction to complex analysis, Addison-Wesley, 1968.
[14] M. Heins, Complex function theory, Academic Press, 1968.
[15] B. A. Fuks and V. I. Levin, Functions of a complex variable and some of their applications I, II, Pergamon, 1961. (Original in Russian, 1951.)
[16] W. Rudin, Real and complex analysis, McGraw-Hill, 1966.
[17] S. Lang, Complex analysis, Addison-Wesley, 1977.
As to the Looman-Men'shov theorem,

[18] D. Men'shov (Menchoff), Les conditions de monogénéité, Actualités Sci. Ind., Hermann, 1936.
[19] D. Men'shov (Menchoff), Sur la généralisation des conditions de Cauchy-Riemann, Fund. Math., 25 (1935), 59–97.
[20] S. Saks, Theory of the integral, Warsaw, 1937, 188–201.
For a new proof of Cauchy's integral theorem,
[21] E. Artin, On the theory of complex functions, Notre Dame mathematical lectures, 1944, 57–70. Also → [7, 17].
For problems,
[22] L. Volkovyskiĭ, G. Lunts, and I. Aramovich, A collection of problems on complex analysis, Addison-Wesley, 1965.
For analytic continuation,
[23] H. Weyl, Die Idee der Riemannschen Fläche, Teubner, 1913; third edition, 1955; English translation, The concept of a Riemann surface, Addison-Wesley, 1964. Also → [1–17].
For the history and concepts of analytic functions,
[24] G. Julia, Essai sur le développement de la théorie des fonctions de variables complexes, Gauthier-Villars, 1933.

199 (IV.12) Homogeneous Spaces

A. General Remarks

Let M be a †differentiable manifold. If a †Lie group G acts †transitively on M as a †Lie transformation group, the manifold M is said to be a **homogeneous space** having G as its transformation group (→ 431 Transformation Groups). The †stabilizer (isotropy subgroup) H_x of G at a point x of M is a closed subgroup of G, and a one-to-one correspondence between G/H_x and M preserving the action of G is defined by associating the element sH_x $(s \in G)$ of G/H_x with the point $s(x)$ of M. This correspondence is a †diffeomorphism between the manifold M and the quotient manifold G/H_x if the number of connected components of G is at most countable. Under this condition we may therefore identify a homogeneous space M with the quotient manifold G/H of a Lie group G by a closed Lie subgroup H (→ 249 Lie Groups). However, H is not uniquely determined by M, and it may be replaced by $H_{s(x)} = sH_xs^{-1} (s \in G)$. Each element h of the stabilizer H_x at a point x induces a linear transformation $\tilde{h}$ on the †tangent space V_x of M at the point x. The set $\tilde{H}_x$ of all $\tilde{h}$ is called the **linear isotropy group** at the point x.

If we represent the homogeneous space M as G/H, we obtain the canonical map $\pi: s \rightarrow sH$ of G onto M, which we call the **projection** of G onto M. Let $\mathfrak{g}$ be the †Lie algebra of G, and $\mathfrak{h}$ the Lie subalgebra corresponding to the closed subgroup H. When we identify $\mathfrak{g}$ with the tangent space at the identity element e of G and $\mathfrak{h}$ with its subspace, the projection π induces a linear isomorphism of $\mathfrak{g}/\mathfrak{h}$ with the tangent space V_x of M at the point $x = \pi(e)$. The †adjoint representation of G gives rise to a linear representation $h \rightarrow \mathrm{Ad}(h)$ modulo $\mathfrak{h}$ of the group H on the linear space $\mathfrak{g}/\mathfrak{h}$. Through the linear isomorphism between $\mathfrak{g}/\mathfrak{h}$ and the tangent space V_x defined by the projection π, this representation of H is equivalent to the one which associates with h the linear transformation $\tilde{h}$ defined by h on the tangent space V_x.

The homogeneous space G/H is said to be **reductive** if there exists a linear subspace $\mathfrak{m}$ of $\mathfrak{g}$ such that $\mathfrak{g} = \mathfrak{h} + \mathfrak{m}$ (direct sum as linear spaces) and $(\mathrm{Ad}\, H)\mathfrak{m} \subset \mathfrak{m}$. H is said to be **reductive** in $\mathfrak{g}$ if the representation $h \rightarrow \mathrm{Ad}(h)$ of H in $\mathfrak{g}$ is †completely reducible.

If a †tensor field P on the homogeneous space $M = G/H$ is G-invariant (namely, invariant under the transformations defined by the elements of G), then the value of P at the point $x = \pi(e)$ is a †tensor over the tangent space V_x at x which is invariant under the linear isotropy group $\tilde{H}$. Conversely, such a tensor over V_x is uniquely extended to a G-invariant tensor field on M. If G/H is reductive, then G-invariant tensor fields over M are in one-to-one correspondence with $\tilde{H}$-invariant tensors over $\mathfrak{m}$. For instance, if H is compact, then H is reductive in $\mathfrak{g}$ and an $\tilde{H}$-invariant positive definite quadratic form on $\mathfrak{m}$ defines a G-invariant †Riemannian metric on G/H.

We say that the homogeneous space $M = G/H$ is a **Riemannian (linearly connected, complex Hermitian, Kähler) homogeneous space** if there exists on M a G-invariant Riemannian metric (†linear connection, †Hermitian metric, †Kähler metric). Concerning such homogeneous spaces, there are various results on their structures and geometric properties [1–5] (→ 412 Symmetric Riemannian Spaces and Real Forms; 427 Topology of Lie Groups and Homogeneous Spaces).

B. Examples

Stiefel Manifold. A **k-frame** $(1 \leqslant k \leqslant n)$ in a real n-dimensional Euclidean vector space $\mathbf{R}^n$ is an ordered system consisting of k linearly independent vectors. If we regard the real †general linear group of degree n, $GL(n, \mathbf{R})$, as the regular linear transformation group of $\mathbf{R}^n$, $GL(n, \mathbf{R})$ acts transitively on the set $V'_{n,k}(\mathbf{R})$ of

all k-frames in $\mathbf{R}^n$. Therefore, if H denotes the subgroup of $GL(n, \mathbf{R})$ consisting of the elements which leave fixed a given k-frame v_0^k, we may identify the set $V'_{n,k}$ and the quotient set $GL(n, \mathbf{R})/H$. Transferring the differentiable manifold structure of $GL(n, \mathbf{R})/H$ to $V'_{n,k}$ through this identification, we see that $V'_{n,k}(\mathbf{R}) = GL(n, \mathbf{R})/H$ becomes a homogeneous space (the differentiable manifold structure of $V'_{n,k}$ is defined independently of the choice of v_0^k). The space $V'_{n,k}(\mathbf{R})$ is called the **(real) Stiefel manifold of k-frames** in $\mathbf{R}^n$.

A k-frame is called an **orthogonal k-frame** if the vectors belonging to the frame are of length 1 and are orthogonal to each other. The set $V_{n,k}(\mathbf{R})$ of all orthogonal k-frames is a submanifold of $V'_{n,k}(\mathbf{R})$. The †orthogonal group $O(n)$ acts transitively on $V'_{n,k}(\mathbf{R})$, which is a homogeneous space represented as $V_{n,k}(\mathbf{R}) = O(n)/I_k \times O(n-k)$. The manifold $V_{n,1}(\mathbf{R})$ is actually the $(n-1)$-dimensional sphere. We call $V_{n,k}(\mathbf{R})$ the **(real) Stiefel manifold of orthogonal k-frames** (or simply **Stiefel manifold**). The **complex Stiefel manifold** $V_{n,k}(\mathbf{C}) = U(n)/I_k \times U(n-k)$ is defined analogously.

Grassmann Manifold. Let $M_{n,k}(\mathbf{R})$ $(1 \leqslant k \leqslant n)$ be the set of all k-dimensional linear subspaces of $\mathbf{R}^n$. The group $O(n)$ acts transitively on $M_{n,k}(\mathbf{R})$, so that we may put $M_{n,k}(\mathbf{R}) = O(n)/O(k) \times O(n-k)$. Here $O(k)$ and $O(n-k)$ are identified with the subgroups of $O(n)$ consisting of all elements leaving fixed every point of a fixed $(n-k)$-dimensional subspace and of its orthogonal complement, respectively. In this way, $M_{n,k}(\mathbf{R})$ is a homogeneous space, which we call the **(real) Grassmann manifold**. The †proper orthogonal group $SO(n)$ acts transitively on $M_{n,k}(\mathbf{R})$, and $M_{n,k}(\mathbf{R})$ may be represented as a homogeneous space having $SO(n)$ as its transformation group. It follows that $M_{n,k}(\mathbf{R})$ is connected. The homogeneous space $\tilde{M}_{n,k}(\mathbf{R}) = SO(n)/SO(k) \times SO(n-k)$ is called the **Grassmann manifold formed by oriented subspaces.** $M_{n,1}(\mathbf{R})$ and $\tilde{M}_{n,1}(\mathbf{R})$ may be identified with the $(n-1)$-dimensional real projective space and the $(n-1)$-dimensional sphere, respectively.

Applying the above process for real Grassmann manifolds to the complex Euclidean vector space $\mathbf{C}^n$ instead of $\mathbf{R}^n$, we see that the set $M_{n,k}(\mathbf{C})$ of all k-dimensional linear subspaces in $\mathbf{C}^n$ is a homogeneous space with the †unitary group $U(n)$ of degree n as its transformation group, and we represent it as $U(n)/U(k) \times U(n-k)$. This space is called the **complex Grassmann manifold.** The manifold $M_{n,k}(\mathbf{C})$ is a simply connected complex manifold and has a cellular decomposition as a †CW complex whose cells are †Schubert varieties (→ 56 Characteristic Classes E). On the other hand, $M_{n,k}(\mathbf{C})$ may be regarded as the set of all $(k-1)$-dimensional linear subspaces in the $(n-1)$-dimensional complex projective space. Then, by using the †Plücker coordinates of these subspaces, $M_{n,k}(\mathbf{C})$ can be realized as an †algebraic variety without singularity in the projective space of dimension $\binom{n}{k}-1$ (→ 90 Coordinates B). Sometimes $M_{n,k}(\mathbf{R})$ is denoted by $G_{n,k}(\mathbf{R})$ or $G(n,k)$. In the same way, the homogeneous space represented as $Sp(n)/Sp(k) \times Sp(n-k)$ is called the †quaternion Grassmann manifold and is denoted by $M_{n,k}(\mathbf{H})$.

Flag Manifold. Let $k_1, \ldots, k_r$ be a sequence of integers such that $n > k_1 > \ldots > k_r > 0$, and let $F(k_1, \ldots, k_r)$ be the set of all monotone sequences $V_1 \supset \ldots \supset V_r$, where $V_i (i = 1, \ldots, r)$ is a k_i-dimensional linear subspace in $\mathbf{R}^n$. For the two sequences $V_1 \supset \ldots \supset V_r$ and $V'_1 \supset \ldots \supset V'_r$ belonging to $F(k_1, \ldots, k_r)$, there exists an element $s \in GL(n, \mathbf{R})$ such that $s(V_i) = V'_i$ $(i = 1, \ldots, r)$. Therefore $F(k_1, \ldots, k_r)$ is a homogeneous space with $GL(n, \mathbf{R})$ as its transformation group, and is called the **proper flag manifold.** Since the unitary group $U(n)$ of degree n acts transitively on it, $F(k_1, \ldots, k_r)$ is also regarded as a homogeneous space admitting $U(n)$ as its transformation group. In this case, putting $F(k_1, \ldots, k_r) = U(n)/H$, H is isomorphic to the direct product $U(k_1 - k_2) \times U(k_2 - k_3) \times \ldots \times U(k_r)$. In particular, when $r = n-1$, $k_i = n - i$, the homogeneous space is the quotient space of the compact Lie group $U(n)$ by a maximal †torus T. In general, the quotient space G/T of a compact connected Lie group G by a maximal torus of G is called a **flag manifold.** If G acts effectively on G/T, G is a †semisimple compact Lie group. The complex Lie group G^C is a Lie transformation group of †biregular transformations which acts transitively on the flag manifold G/T, a simply connected Kähler homogeneous space. Here G^C is a complex Lie group having G as a maximal compact subgroup. If B is a maximal †solvable Lie subgroup (†Borel subgroup) of G^C, G/T is represented as G^C/B.

References

[1] A. Borel and R. Remmert, Über kompakte homogene Kählersche Mannigfaltigkeiten, Math. Ann., 145 (1961–1962), 429–439.

[2] Y. Matsushima, Espaces homogènes de Stein des groupes de Lie complexes, Nagoya Math. J., 16 (1960), 205–218.

[3] K. Nomizu, Invariant affine connections on homogeneous spaces, Amer. J. Math., 76 (1954), 33–65.

[4] H. C. Wang, Closed manifolds with homogeneous complex structure, Amer. J. Math., 76 (1954), 1–32.
[5] N. R. Wallach, Harmonic analysis on homogeneous spaces, Dekker, 1973.
[6] S. Helgason, Differential geometry, Lie groups, and symmetric spaces, Academic Press, 1978.

200 (III.24) Homological Algebra

A. General Remarks

Homological algebra is a new branch of mathematics that developed rapidly after World War II. The introduction of the theory was motivated by the observation that some algebraic ideas and mechanisms that arose in the development of †algebraic topology, in particular, †homology theory, can provide powerful tools for treating from a unified viewpoint various problems in algebra that previously were treated differently. One of its characteristic features lies in emphasizing, from the standpoint of categories and functors (→ 52 Categories and Functors), the functional structure of the objects to be investigated rather than their inner structure. Thus the theory of derived functors constitutes the main theme of homological algebra. This new theory turned out to have wide applications in other areas of mathematics, and the philosophy embodied in the theory has been influential in the general progress of mathematics. For general references → [2, 5, 6]; for †sheaf cohomology → [3, 4, 8].

B. Graded Modules and Graded Objects

Let A be a †ring with unity element and X be a †unitary A-module. If we are given a sequence of A-submodules X_n ($n \in \mathbf{Z}$) such that $X = \sum_{n\in\mathbf{Z}} X_n$ (†direct sum), we call X a **graded A-module** and X_n the **component of degree** n of X. Each element x of a graded A-module X has a unique representation $x = \sum_{n\in\mathbf{Z}} x_n$ ($x_n \in X_n$); we call x_n the **component** of degree n of x. An A-submodule Y of a graded A-module X is called **homogeneous** if $x \in Y$ implies $x_n \in Y$ ($n \in \mathbf{Z}$). In this case, $Y = \sum_n Y_n$ and the quotient module $X/Y = \sum_n X_n/Y_n$ are graded A-modules, where $Y_n = Y \cap X_n$. Let $X = \sum_n X_n$ and $Y = \sum_n Y_n$ be graded A-modules and $f: X \rightarrow Y$ be an A-homomorphism. If there is a fixed integer p such that $f(X_n) \subset Y_{n+p}$ for any $n \in \mathbf{Z}$, f is called an **A-homomorphism of degree** p. In this case, $\operatorname{Ker} f = \sum_n \operatorname{Ker} f_n$ and $\operatorname{Im} f = \sum_n \operatorname{Im} f_{n-p}$ are homogeneous A-submodules of X and Y, respectively, where $f_n: X_n \rightarrow Y_{n+p}$ is the restriction of f on X_n.

Sometimes, by a graded module we mean only a sequence $\{X_n\}$ of A-modules X_n, without considering the direct sum $\sum_n X_n$. Similarly, we have the notion of a **graded object** $\{C_n\}$ in any category $\mathscr{C}$.

C. Chain Complexes and Homology Modules

By a **chain complex** (X, ∂) over A we mean a graded A-module $X = \sum_n X_n$ together with an A-homomorphism $\partial: X \rightarrow X$ of degree -1 such that $\partial \circ \partial = 0$. Hence a chain complex over A is completely determined by a sequence

$$\ldots \rightarrow X_{n+1} \xrightarrow{\partial_{n+1}} X_n \xrightarrow{\partial_n} X_{n-1} \rightarrow \ldots$$

of A-modules and A-homomorphisms such that $\partial_n \circ \partial_{n+1} = 0$ for all n. We call ∂ the **boundary operator**. For a chain complex (X, ∂), we write $\operatorname{Ker} \partial = Z(X)$, $\operatorname{Ker} \partial_n = Z_n(X)$, $\operatorname{Im} \partial = B(X)$, $\operatorname{Im} \partial_{n+1} = B_n(X)$. Then $Z(X) = \sum_n Z_n(X)$, $B(X) = \sum_n B_n(X)$ are homogeneous submodules of X, called the **module of cycles** and the **module of boundaries**, respectively. $B(X)$ is a homogeneous submodule of $Z(X)$, and the quotient modules $Z(X)/B(X)$, $Z_n(X)/B_n(X)$ are denoted by $H(X)$, $H_n(X)$, respectively. We call $H(X) = \sum_n H_n(X)$ the **homology module** of the chain complex (X, ∂).

If (X, ∂), (Y, ∂') are chain complexes over A, an A-homomorphism $f: X \rightarrow Y$ of degree 0 satisfying $\partial' \circ f = f \circ \partial$ (i.e., $\partial'_n f_n = f_{n-1} \partial_n$) is called a **chain mapping** of X to Y. For a chain mapping f, we have $f(Z_n(X)) \subset Z_n(Y)$, $f(B_n(X)) \subset B_n(Y)$, and hence f induces an A-homomorphism $f_*: H(X) \rightarrow H(Y)$ of degree 0, which is called the **homological mapping** induced by f. We have $(1_X)_* = 1_{H(X)}$, and $(g \circ f)_* = g_* \circ f_*$ for chain mappings $f: X \rightarrow Y$ and $g: Y \rightarrow Z$.

Let $f, g: X \rightarrow Y$ be two chain mappings. If there is an A-homomorphism $D: X \rightarrow Y$ of degree $+1$ such that $f - g = D \circ \partial + \partial' \circ D$, we say that f is **chain homotopic** to g and write $f \simeq g$; D is called a **chain homotopy** of f to g. If f is chain homotopic to g, we have $f_* = g_*: H(X) \rightarrow H(Y)$. For chain complexes X and Y, if there are chain mappings $f: X \rightarrow Y$ and $g: Y \rightarrow X$ such that $f \circ g \simeq 1_Y$ and $g \circ f \simeq 1_X$, we say that X is **chain equivalent** to Y. In this case $f_*: H(X) \rightarrow H(Y)$ is an isomorphism and $g_*: H(Y) \rightarrow H(X)$ is its inverse.

Let (X, ∂) be a chain complex over A and $Y = \sum_n Y_n$ be a homogeneous A-submodule of X such that $\partial Y \subset Y$. Then Y and X/Y are chain complexes over A with the boundary operators induced by ∂. Y is called a **chain**

subcomplex of X, and X/Y is called the **quotient chain complex** of X by Y or the **relative chain complex** of X mod Y. For a chain complex X and its subcomplex Y we have an †exact sequence $0 \to Y \xrightarrow{i} X \xrightarrow{j} X/Y \to 0$, where i is the †canonical injection and j the †canonical surjection.

Let (W, ∂'), (X, ∂), (Y, ∂'') be chain complexes over A, and $f: W \to X$, $g: X \to Y$ be chain mappings such that $0 \to W \xrightarrow{f} X \xrightarrow{g} Y \to 0$ is exact. Then an A-homomorphism $\partial_*: H(Y) \to H(W)$ of degree -1, called the **connecting homomorphism**, is defined by $\partial_*(y + B(Y)) = f^{-1} \circ \partial \circ g^{-1}(y) + B(W)$ $(y \in Z(Y))$, and we have the **exact sequence of homology**:

$$\ldots \xrightarrow{\partial_*} H_n(W) \xrightarrow{f_*} H_n(X) \xrightarrow{g_*} H_n(Y) \xrightarrow{\partial_*} H_{n-1}(W) \xrightarrow{f_*} H_{n-1}(X) \xrightarrow{g_*} H_{n-1}(Y) \xrightarrow{\partial_*} \ldots.$$

For a †commutative diagram

$$\begin{array}{ccccccc} 0 \to & W & \to & X & \to & Y & \to 0 \\ & \downarrow\varphi & & \downarrow & & \downarrow\psi & \\ 0 \to & W' & \to & X' & \to & Y' & \to 0 \end{array} \qquad (1)$$

consisting of chain complexes and chain mappings in which each row is exact, we have $\partial_* \circ \psi_* = \varphi_* \circ \partial_*: H(Y) \to H(W')$.

For the †inductive limit $\varinjlim X_\lambda$ of chain complexes X_λ over A, we have

$$H(\varinjlim X_\lambda) = \varinjlim H(X_\lambda).$$

A chain complex X is said to be **positive** if $X_n = 0$ for all $n < 0$. If X is a positive chain complex over A and M is an A-module, then we mean by an **augmentation of** X **over** M an A-homomorphism $\varepsilon: X_0 \to M$ such that the composition $X_1 \xrightarrow{\partial_1} X_0 \xrightarrow{\varepsilon} M$ is trivial: $\varepsilon \circ \partial_1 = 0$. A positive chain complex X together with an augmentation ε of X over M is called an **augmented chain complex** over M. It is said to be **acyclic** if the sequence

$$\ldots \to X_n \xrightarrow{\partial_n} X_{n-1} \to \ldots \to X_1 \xrightarrow{\partial_1} X_0 \xrightarrow{\varepsilon} M \to 0$$

is exact, namely, if $H_n(X) = 0$ $(n \neq 0)$ and ε induces an A-isomorphism $H_0(X) \cong M$. In this case X is also called a **left resolution** of M. Moreover, if each X_n is a †projective A-module, X is called a **left projective resolution**. For any A-module M, there exists a left projective resolution of M.

Let $\alpha: M \to M'$ be an A-homomorphism of A-modules, and X, X' be augmented chain complexes over M, M' having augmentations ε, ε', respectively. Then a chain mapping $f: X \to X'$ satisfying $\varepsilon' \circ f_0 = \alpha \circ \varepsilon$ is called a **chain mapping over** α. If X, X' are left projective resolutions of M, M', respectively, then there exist chain mappings of X to X' over α, and any two such mappings are chain homotopic. In particular, a left projective resolution of an A-module M is uniquely determined up to chain homotopy.

D. Tor

Given a right A-module M and a left A-module N, **Z**-modules $\mathrm{Tor}_n^A(M, N)$ $(n = 0, 1, 2, \ldots)$, called the **torsion products** (or **Tor groups**), are defined as follows: Let

$$Y: \ldots \to Y_n \xrightarrow{\partial_n} Y_{n-1} \to \ldots \xrightarrow{\partial_1} Y_0 \xrightarrow{\varepsilon} N \to 0$$

be a projective resolution of N, and consider the chain complex

$$(M \otimes_A Y, 1 \otimes \partial): \ldots \to M \otimes_A Y_n \xrightarrow{1 \otimes \partial_n} M \otimes_A Y_{n-1} \to \ldots \xrightarrow{1 \otimes \partial_1} M \otimes_A Y_0 \to 0$$

obtained by forming the †tensor product of M and Y. Then we see that the homology module $H_n(M \otimes_A Y)$ is uniquely determined for any choice of left projective resolution of N. We define $\mathrm{Tor}_n^A(M, N) = H_n(M \otimes_A Y)$. In particular, we have $\mathrm{Tor}_0^A(M, N) \cong M \otimes_A N$.

Properties of Tor. (1) If M is a †flat A-module, we have $\mathrm{Tor}_n^A(M, N) = 0$ $(n = 1, 2, \ldots)$.

(2) An A-homomorphism $f: M_1 \to M_2$ induces a homomorphism $f_*: \mathrm{Tor}_n^A(M_1, N) \to \mathrm{Tor}_n^A(M_2, N)$. We have $(1_M)_* = 1$, and $(g \circ f)_* = g_* \circ f_*$ for $f: M_1 \to M_2$, $g: M_2 \to M_3$.

(3) For an exact sequence $0 \to M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3 \to 0$, we have the following **exact sequence of Tor**:

$$\ldots \xrightarrow{\partial_*} \mathrm{Tor}_n^A(M_1, N) \xrightarrow{f_*} \mathrm{Tor}_n^A(M_2, N) \xrightarrow{g_*} \mathrm{Tor}_n^A(M_3, N) \xrightarrow{\partial_*} \mathrm{Tor}_{n-1}^A(M_1, N) \to \ldots \to \mathrm{Tor}_1^A(M_3, N) \xrightarrow{\partial_*} M_1 \otimes_A N \to M_2 \otimes_A N \to M_3 \otimes_A N \to 0,$$

where ∂_* are the connecting homomorphisms.

(4) For a commutative diagram

$$\begin{array}{ccccccc} 0 \to & M_1 & \to & M_2 & \to & M_3 & \to 0 \\ & \downarrow\varphi & & \downarrow & & \downarrow\psi & \\ 0 \to & M_1' & \to & M_2' & \to & M_3' & \to 0 \end{array}$$

of A-modules and A-homomorphisms with exact rows, we have $\partial_* \circ \psi_* = \varphi_* \circ \partial_*$.

(5) $\mathrm{Tor}_n^A(\Sigma_\lambda M_\lambda, N) \cong \Sigma_\lambda \mathrm{Tor}_n^A(M_\lambda, N)$

(6) $\mathrm{Tor}_n^A(\varinjlim M_\lambda, N) \cong \varinjlim \mathrm{Tor}_n^A(M_\lambda, N)$.

On the other hand, take a left projective resolution X of M and consider the chain complex $X \otimes_A N$. Then we have $H_n(X \otimes_A N) \cong \mathrm{Tor}_n^A(M, N)$ for $n = 0, 1, \ldots$. Therefore properties similar to (1)–(6) hold with respect to the second variable N of $\mathrm{Tor}_n^A(M, N)$.

(7) If A° is a ring †anti-isomorphic to A, then $\mathrm{Tor}_n^A(M, N) \cong \mathrm{Tor}_n^{A^\circ}(N, M)$. In particular,

if A is commutative, then $\operatorname{Tor}_n^A(M,N)$ is an A-module and we have $\operatorname{Tor}_n^A(M,N)\cong\operatorname{Tor}_n^A(N,M)$.

(8) Let A be a †principal ideal ring. Then $\operatorname{Tor}_n^A(M,N)=0$ $(n=2,3,\dots)$ and $\operatorname{Tor}_1^A(M,N)$ is also denoted by $M *_A N$. For an exact sequence $0\to M_1\to M_2\to M_3\to 0$, we have the exact sequence $0\to M_1 *_A N\to M_2 *_A N\to M_3 *_A N\to M_1\otimes_A N\to M_2\otimes_A N\to M_3\otimes_A N\to 0$. In particular, $\mathbf{Z} *_{\mathbf{Z}} N=0$ and $(\mathbf{Z}/n\mathbf{Z}) *_{\mathbf{Z}} N\cong {}_nN$ $(=\{x\in N\mid nx=0\})$.

Universal Coefficient Theorem for Homology. If (X,∂) is a chain complex over A and N is a left A-module, then $(X\otimes_A N, \partial\otimes 1)$ is a chain complex. If A is a principal ideal ring and each X_n is a †torsion-free A-module, then we have a formula

$$H_n(X\otimes_A N)$$
$$\cong H_n(X)\otimes_A N + H_{n-1}(X) *_A N,$$

called the **universal coefficient theorem**.

E. Double Chain Complexes

By a **double chain complex** $(X_{p,q},\partial',\partial'')$ over A we mean a family of left A-modules $X_{p,q}$ $(p,q\in\mathbf{Z})$ together with A-homomorphisms $\partial'_{p,q}:X_{p,q}\to X_{p-1,q}$ and $\partial''_{p,q}:X_{p,q}\to X_{p,q-1}$ such that $\partial'_{p-1,q}\circ\partial'_{p,q}=\partial''_{p,q-1}\circ\partial''_{p,q}=\partial'_{p,q-1}\circ\partial''_{p,q}+\partial''_{p-1,q}\circ\partial'_{p,q}=0$. We define the associated chain complex (X_n,∂) by setting $X_n=\sum_{p+q=n}X_{p,q}$, $\partial_n=\sum_{p+q=n}\partial'_{p,q}+\partial''_{p,q}$. We call ∂ the **total boundary operator**, and ∂', ∂'' the **partial boundary operators**.

Given a chain complex X consisting of right A-modules and a chain complex Y consisting of left A-modules, a double chain complex $(Z_{p,q},\partial',\partial'')$ is defined by setting $Z_{p,q}=X_p\otimes_A Y_q$, $\partial'_{p,q}=\partial_p\otimes 1$, $\partial''_{p,q}=(-1)^p 1\otimes\partial_q$, where ∂_p, ∂_q are the boundary operators of X, Y, respectively. It is called the **product double chain complex** of X and Y, and the homology module of its associated chain complex is denoted by $H(X\otimes_A Y)$. With respect to this homology module, the following facts hold. If X is a left projective resolution of a right A-module M and Y is that of a left A-module N, then $H_n(X\otimes_A Y)=\operatorname{Tor}_n^A(M,N)$. If A is a principal ideal ring and each X_n is a torsion-free A-module, then we have the formula

$$H_n(X\otimes_A Y)\cong\sum_{p+q=n}H_p(X)\otimes_A H_q(Y)$$
$$+\sum_{p+q=n-1}H_p(X) *_A H_q(Y),$$

the **Künneth theorem**.

F. Cochain Complexes

By a **cochain complex** (X,d) over A we mean a graded A-module X together with an A-homomorphism $d:X\to X$ of degree $+1$ such that $d\circ d=0$; d is called the **coboundary operator** or **differential**. For a cochain complex (X,d), we denote by X^n the component of degree n of X, and by d^n, $X^n\to X^{n+1}$ the restriction of d on X^n. Then a chain complex (Y,∂) is defined by $Y_n=X^{-n}$ and $\partial_n: Y_n\to Y_{n-1}$ is equal to $d^{-n}:X^{-n}\to X^{-n+1}$.

For a cochain complex (X,d), we write $\operatorname{Ker} d^n=Z^n(X)$, $\operatorname{Ker} d=Z(X)$ $(Z(X)=\sum Z^n(X))$, $\operatorname{Im} d^{n-1}=B^n(X)$, $\operatorname{Im} d=B(X)$ $(B(X)=\sum B^n(X))$, and $Z^n(X)/B^n(X)=H^n(X))$, $Z(X)/B(X)=H(X)$ $(H(X)=\sum H^n(X))$. These modules $Z(X)$ $(Z^n(X))$, $B(X)$ $(B^n(X))$, and $H(X)$ $(H^n(X))$ are called the **module of cocycles**, the **module of coboundaries**, and the **cohomology module** of X, respectively. If we consider the associated chain complex (Y,∂) of (X,d), then $H_{-n}(Y)$ corresponds to $H^n(X)$. In this way, results on chain complexes give results on cochain complexes. Thus the concepts of **cochain mapping, cochain homotopy, cochain equivalence, cochain subcomplex**, and **relative cochain complex** can be defined as in the case of chain complexes in B, and we have corresponding results. In particular, given an exact sequence $0\to W\xrightarrow{f}X\xrightarrow{g}Y\to 0$ of cochain complexes and cochain mappings, the **connecting homomorphism** $d_*:H^n(Y)\to H^{n+1}(W)$ is defined, and the **exact sequence of cohomology**

$$\dots\xrightarrow{d_*}H^n(W)\xrightarrow{f_*}H^n(X)\xrightarrow{g_*}H^n(Y)$$
$$\xrightarrow{d_*}H^{n+1}(W)\xrightarrow{f_*}H^{n+1}(X)\xrightarrow{g_*}H^{n+1}(Y)\xrightarrow{d_*}\dots$$

exists. For a commutative diagram

$$\begin{array}{ccccccccc} 0 & \to & W & \to & X & \to & Y & \to & 0 \\ & & \downarrow\varphi & & \downarrow & & \downarrow\psi & & \\ 0 & \to & W' & \to & X' & \to & Y' & \to & 0 \end{array}$$

of cochain complexes and cochain mappings with exact rows, we have $d_*\circ\psi_*=\varphi_*\circ d_*$.

A cochain complex X is said to be **positive** if $X^n=0$ for $n<0$. If X is a positive cochain complex over A and M is an A-module, we mean by an **augmentation** of X over M an A-homomorphism $\varepsilon: M\to X^0$ such that the composition $M\xrightarrow{\varepsilon}X^0\xrightarrow{d^0}X^1$ is trivial. If the sequence

$$0\to M\xrightarrow{\varepsilon}X^0\xrightarrow{d^0}\dots\to X^n\xrightarrow{d^n}X^{n+1}\to\dots$$

is exact, X is called a **right resolution** of M. Moreover, if each X^n is an †injective A-module, X is called a **right injective resolution** of M. For any A-module M, there exists a right injective resolution of M, and any two such resolutions are cochain homotopic.

G. Ext

Given left A-modules M and N, $\mathbf{Z}$-modules $\operatorname{Ext}_A^n(M,N)$ $(n=0,1,2,\dots)$, called the **Ext**

groups, are defined as follows: Let $X: \dots \to X_n \to X_{n-1} \to \dots \to X_0 \to M \to 0$ be a projective resolution of M, and consider the cochain complex $\mathrm{Hom}_A(X, N)$:

$$0 \to \mathrm{Hom}_A(X_0, N) \to \dots \to \mathrm{Hom}_A(X_{n-1}, N) \to \mathrm{Hom}_A(X_n, N) \to \dots$$

obtained by forming the †module of A-homomorphisms. Then we can show that the cohomology module $H^n(\mathrm{Hom}_A(X, N))$ is uniquely determined for any choice of projective resolution of M. We define $\mathbf{Ext}_A^n(M, N) = H^n(\mathrm{Hom}_A(X, N))$. This can also be defined as the cohomology module $H^n(\mathrm{Hom}_A(M, Y))$ of the cochain complex $\mathrm{Hom}_A(M, Y): 0 \to \mathrm{Hom}_A(M, Y^0) \to \dots \to \mathrm{Hom}_A(M, Y^{n-1}) \to \mathrm{Hom}_A(M, Y^n) \to \dots$, where $Y: 0 \to N \to Y^0 \to \dots \to Y^{n-1} \to Y^n \to \dots$ is a right injective resolution of N. Furthermore, for a left projective resolution X of M and a right injective resolution Y of N, we see that $\mathrm{Ext}_A^n(M, N)$ is isomorphic to the cohomology module $H^n(\mathrm{Hom}_A(X, Y))$ of the associated cochain complex of the double cochain complex $\mathrm{Hom}_A(X, Y) = (\mathrm{Hom}_A(X_p, Y^q), d', d'')$, where $d'_{p,q}: \mathrm{Hom}_A(X_p, Y^q) \to \mathrm{Hom}_A(X_{p+1}, Y^q)$ and $d''_{p,q}: \mathrm{Hom}_A(X_p, Y^q) \to \mathrm{Hom}_A(X_p, Y^{q+1})$ are given by $d'_{p,q}(u) = u \circ \partial_{p+1}$, $d''_{p,q}(u) = (-1)^{p+q+1} d^q \circ u$ $(u \in \mathrm{Hom}_A(X_p, Y^q))$ by using the boundary operator ∂ of X and the coboundary operator d of Y.

Properties of Ext. (1) We have $\mathrm{Ext}_A^0(M, N) \cong \mathrm{Hom}_A(M, N)$.

(2) If M is a projective A-module or N is an injective A-module, then $\mathrm{Ext}_A^n(M, N) = 0$ $(n = 1, 2, \dots)$.

(3) An A-homomorphism $f: M_1 \to M_2$ (resp. $f: N_1 \to N_2$) induces a homomorphism $f^*: \mathrm{Ext}_A^n(M_2, N) \to \mathrm{Ext}_A^n(M_1, N)$ (resp. $f_*: \mathrm{Ext}_A^n(M, N_1) \to \mathrm{Ext}_A^n(M, N_2)$). We have $1_M^* = 1$ and $(g \circ f)^* = f^* \circ g^*$ for $f: M_1 \to M_2$, $g: M_2 \to M_3$ (resp. $1_{N*} = 1$ and $(g \circ f)_* = g_* \circ f_*$ for $f: N_1 \to N_2$, $g: N_2 \to N_3$).

(4) For an exact sequence $0 \to M_1 \to M_2 \to M_3 \to 0$ (resp. $0 \to N_1 \to N_2 \to N_3 \to 0$), we have the **exact sequence of Ext**:

$$0 \to \mathrm{Hom}_A(M_3, N) \to \mathrm{Hom}_A(M_2, N) \to \mathrm{Hom}_A(M_1, N) \to \mathrm{Ext}_A^1(M_3, N) \to \mathrm{Ext}_A^1(M_2, N) \to \dots$$

$$(\text{resp. } 0 \to \mathrm{Hom}_A(M, N_1) \to \mathrm{Hom}(M, N_2) \to \mathrm{Hom}(M, N_3) \to \mathrm{Ext}_A^1(M, N_1) \to \mathrm{Ext}_A^1(M, N_2) \to \dots)$$

(5)

$$\mathrm{Ext}_A^n\Big(\sum_\alpha M_\alpha, \prod_\beta N_\beta\Big) = \prod_{\alpha,\beta} \mathrm{Ext}_A^n(M_\alpha, N_\beta).$$

(6) If A is a principal ideal ring, then $\mathrm{Ext}_A^n(M, N) = 0$ $(n = 2, 3, \dots)$, and $\mathrm{Ext}_A^1(M, N)$ is also denoted by $\mathrm{Ext}_A(M, N)$. In particular, $\mathrm{Ext}_{\mathbf{Z}}(\mathbf{Z}, N) = 0$, $\mathrm{Ext}_{\mathbf{Z}}(\mathbf{Z}/n\mathbf{Z}, N) \cong N/nN$, $\mathrm{Ext}_{\mathbf{Z}}(M, \mathbf{Q}/\mathbf{Z}) = 0$, $\mathrm{Ext}_{\mathbf{Z}}(M, \mathbf{Z}/n\mathbf{Z}) = \hat{M}/n\hat{M}$, where $\hat{M} = \mathrm{Hom}_{\mathbf{Z}}(M, \mathbf{Q}/\mathbf{Z})$.

Universal Coefficient Theorem for Cohomology. If X is a chain complex over a principal ideal ring A such that each X_n is a free A-module, then for any A-module N we have the formula

$$H^n(\mathrm{Hom}_A(X, N)) \cong \mathrm{Hom}_A(H_n(X), N) + \mathrm{Ext}_A(H_{n-1}(X), N),$$

the **universal coefficient theorem**. This is generalized as follows: Let X be a chain complex and Y a cochain complex, both over a principal ideal ring A. Assume that each X_n is a free A-module or that each Y^n is an injective A-module. Then we have the formula

$$H^n(\mathrm{Hom}_A(X, Y)) \cong \sum_{p+q=n} \mathrm{Hom}_A(H_p(X), H^q(Y)) + \sum_{p+q=n-1} \mathrm{Ext}_A(H_p(X), H^q(Y))$$

(→ 201 Homology Theory).

H. Complexes in Abelian Categories

We mainly consider general †Abelian categories $\mathscr{C}$. Consideration may, however, be restricted to the †category (Ab) of Abelian groups (whose †objects are Abelian groups and whose †morphisms are homomorphisms) or the †category ${}_R\mathscr{M}$ of R-modules.

A **(cochain) complex** C in an Abelian category $\mathscr{C}$ is a graded object $\{C^n\}$ in $\mathscr{C}$ with **differentials** $d^n: C^n \to C^{n+1}$ subject to the condition that $d^{n+1} \circ d^n = 0$ $(n \in \mathbf{Z})$. The nth **cohomology** $H^n(C)$ of C is defined by the †exact sequence $0 \to B^n(C) \to Z^n(C) \to H^n(C) \to 0$, where $B^n(C)$ and $Z^n(C)$ are objects representing $\mathrm{Im}\, d^{n-1}$ and $\mathrm{Ker}\, d^n$, respectively. The complex C is called **positive (negative)** if $C^n = 0$ for $n < 0$ $(n > 0)$. We sometimes interchange positive superscripts and negative subscripts and write C_{-n} instead of C^n. Then the differentials become $d_n: C_n \to C_{n-1}$, and C is then called a **chain complex**. The quotient of $\mathrm{Ker}\, d_n = Z_n$ by $\mathrm{Im}\, d_{n+1} = B_n$ is called the nth **homology** $H_n(C)$. Negative complexes are usually described in this manner. When C^n, Z^n, B^n, and H^n are sets, as in the category ${}_R\mathscr{M}$ of R-modules, their elements are called **cochains, cocycles, coboundaries**, and **cohomology classes**, respectively. Similarly, in the group C_n of **chains**, residue classes of **cycles** $(\in Z_n)$ modulo **boundaries** $(\in B_n)$ are called **homology classes** $(\in H_n)$.

A **morphism** (or **chain transformation**) $f: C \to C'$ is a †natural transformation of the complexes considered as †functors $\mathbf{Z} \to \mathscr{C}$; i.e., f is a family of morphisms $f^n: C^n \to C'^n$ ($n \in \mathbf{Z}$) satisfying $f^{n+1} \circ d^n = d^n \circ f^n$. It induces a morphism of cohomology $H^n(C) \to H^n(C')$. A **subcomplex** of C is an equivalence class of †monomorphisms $D \to C$, usually denoted by any representative D of the class. A **(chain) homotopy** between two chain transformations $f, g: C \to C'$ is a family of morphisms $h^n: C^n \to C'^{n-1}$ ($n \in \mathbf{Z}$) satisfying $f^n - g^n = h^{n+1} \circ d^n + d^{n-1} \circ h^n$. If there exists a homotopy between f and g, then f and g induce the same morphism of cohomology. A morphism $f: C \to C'$ is called a **(chain) equivalence** if there exists a morphism $f': C' \to C$ such that $f' \circ f$ and $f \circ f'$ are homotopic to the identities of C and C', respectively. In this case we have $H^n(C) \cong H^n(C')$. An exact sequence of complexes $0 \to C' \to C \to C'' \to 0$ gives rise to the **connecting morphisms** $H^n(C'') \to H^{n+1}(C')$ ($n \in \mathbf{Z}$), and the resulting sequence $\ldots \to H^{n-1}(C'') \to H^n(C') \to H^n(C) \to H^n(C'') \to H^{n+1}(C') \to \ldots$ is exact (the **exact cohomology sequence**), and similarly for homology instead of cohomology. An object $A \in \mathscr{C}$ defines a complex (also denoted by A) such that $A^0 = A$, $d^0 = 0$. A positive complex C together with a morphism $\varepsilon: A \to C$ is called a **complex over** A, and ε is the **augmentation**. A complex C over A is **acyclic** if $0 \to A \xrightarrow{\varepsilon} C^0 \to C^1 \to \ldots$ is exact. An acyclic positive complex over A is called a right **resolution** of A. Let $\{C, \varepsilon\}$, $\{C', \varepsilon'\}$ be complexes over A and A', respectively, and α a morphism $A \to A'$. Then a morphism $f: C \to C'$ satisfying $f \circ \varepsilon = \varepsilon' \circ \alpha$ is called a morphism over α. For a negative complex C, we define similarly augmentations $\varepsilon: C \to A$, acyclicity, left resolutions, etc.

A **bicomplex** (or **double complex**) C in $\mathscr{C}$ consists of objects $C^{p,q}$ ($p, q \in \mathbf{Z}$) and two differentials $d_{\mathrm{I}}: C^{p,q} \to C^{p+1,q}$, $d_{\mathrm{II}}: C^{p,q} \to C^{p,q+1}$ subject to $d_{\mathrm{I}}^2 = d_{\mathrm{II}}^2 = 0$ and $d_{\mathrm{I}} d_{\mathrm{II}} = d_{\mathrm{II}} d_{\mathrm{I}}$ (sometimes replaced by anticommutativity, $d_{\mathrm{I}} d_{\mathrm{II}} + d_{\mathrm{II}} d_{\mathrm{I}} = 0$). Morphisms of bicomplexes are defined as for single complexes. A bicomplex C becomes a (single) complex if we put $C^n = \sum_{p+q=n} C^{p,q}$ (when the sum exists) and define the differential d to be $d_{\mathrm{I}} + (-1)^p d_{\mathrm{II}}$ on $C^{p,q}$. Then d is called the **total differential** and d_{I}, d_{II} the **partial differentials**. On the other hand, $C_{\mathrm{I}}^q = \{C^{p,q} (p \in \mathbf{Z}), d_{\mathrm{I}}\}$ constitutes a complex for each q, whose cohomology $H^p(C_{\mathrm{I}}^q)$ is denoted by $H_{\mathrm{I}}^p(C^q)$. Then d_{II} induces morphisms $H_{\mathrm{I}}^p(C^q) \to H_{\mathrm{I}}^p(C^{q+1})$, so that we obtain a complex $H_{\mathrm{I}}^p(C)$. The cohomology of $H_{\mathrm{I}}^p(C)$ is denoted by $H_{\mathrm{II}}^q(H_{\mathrm{I}}^p(C))$. We define $H_{\mathrm{I}}^p(H_{\mathrm{II}}^q(C))$ similarly. The cohomology of C with respect to the total differential is denoted simply by $H^n(C)$. Similar constructions are applied to double chain complexes $\{C_{p,q}\}$ and further to **multiple complexes**, as we shall show in the case of bicomplexes.

Let T be a †bifunctor $\mathscr{C}_1 \times \mathscr{C}_2 \to \mathscr{C}'$ and C_i be complexes in $\mathscr{C}_i$ ($i = 1, 2$). Then $T(C_1, C_2)$ is a bicomplex in $\mathscr{C}'$. For instance, $\mathrm{Hom}(C', C)$ is a positive (bipositive) complex if $C(C')$ is a positive (negative) complex in $\mathscr{C}$. If C, C' are complexes in $\mathscr{M}_R$, ${}_R\mathscr{M}$, respectively, the †tensor product $C \otimes_R C'$ is a complex in (Ab) (the **product complex**). There is a canonical morphism $H_p(C) \otimes H_q(C') \to H_{p+q}(C \otimes C')$. If C_n and B_n are †flat for all $n \in \mathbf{Z}$, we have the following exact sequence (**Künneth's formula**):

$$0 \to \sum_{p+q=n} H_p(C) \otimes H_q(C') \to H_n(C \otimes C')$$

$$\to \sum_{p+q=n-1} \mathrm{Tor}_1(H_p(C), H_q(C')) \to 0$$

(for the definition of Tor → Section D). For $C' = A \in {}_R\mathscr{M}$, Künneth's formula reduces to the exact sequence $0 \to H_n(C) \otimes A \to H_n(C \otimes A) \to \mathrm{Tor}_1(H_{n-1}(C), A) \to 0$ (**universal coefficient theorem**). The corresponding exact sequence for cohomology is

$$0 \to \mathrm{Ext}^1(H_{n-1}(C), A) \to H^n(C, A)$$

$$\to \mathrm{Hom}(H_n(C), A) \to 0$$

(→ Section G; 201 Homology Theory).

I. Satellites and Derived Functors

Let $\mathscr{C}$ and $\mathscr{C}'$ be Abelian categories. All functors in this section are †additive. A †covariant functor $T: \mathscr{C} \to \mathscr{C}'$ is called **exact** if T maps every exact sequence in $\mathscr{C}$ to an exact sequence in $\mathscr{C}'$. T is called **half-exact**, **left exact**, or **right exact** if for every **short exact sequence** $0 \to A \to B \to C \to 0$, the sequence $T(A) \to T(B) \to T(C)$, $0 \to T(A) \to T(B) \to T(C)$, or $T(A) \to T(B) \to T(C) \to 0$, respectively, is exact. Similar definitions apply for †contravariant functors. The functor $\mathrm{Hom}: \mathscr{C} \times \mathscr{C} \to (\mathrm{Ab})$ (which defines the category $\mathscr{C}$) is left exact in both factors. An object P is **projective** if $h_P(\cdot) = \mathrm{Hom}(P, \cdot)$ is exact, while Q is **injective** if $h^Q(\cdot) = \mathrm{Hom}(\cdot, Q)$ is exact. If every object A admits an †epimorphism from a projective object $P \to A$ (resp. †monomorphism into an injective object $A \to Q$), $\mathscr{C}$ is said to have enough projectives (injectives). An object G is called a **generator** (**cogenerator**) if the natural mapping $\mathrm{Hom}(A, B) \to \mathrm{Hom}(h_G(A), h_G(B))$ ($\mathrm{Hom}(h^G(B), h^G(A))$) is one-to-one.

An Abelian category $\mathscr{C}$ is called a **Grothendieck category** if (1) $\mathscr{C}$ has a generator, (2) †direct sums always exist, and (3) the identity $(\bigcup A_i) \cap B = \bigcup (A_i \cap B)$ holds for any object A, its †subobject B, and a †totally ordered family $\{A_i\}$ of subobjects. A Grothendieck cate-

gory has enough injectives (R. Baer, 1940, for (Ab); A. Grothendieck, 1957, for general $\mathscr{C}$). A monomorphism into an injective object $f: A \to Q$ is called an **injective envelope** if $\operatorname{Im} f \cap \operatorname{Im} g \neq 0$ for any nonzero monomorphism $g: B \to Q$. Every object A in a Grothendieck category admits an injective envelope, which is unique up to isomorphism (B. Eckmann and A. Schopf, 1953, for ${}_R\mathscr{M}$; B. Mitchell, 1960, for general $\mathscr{C}$).

We say that a covariant **∂-functor** $\mathscr{C} \to \mathscr{C}'$ is given if we have a sequence of covariant functors $T = \{T^i: \mathscr{C} \to \mathscr{C}'\}$ and the **connecting morphisms** $\partial: T^i(A'') \to T^{i+1}(A')$ for an arbitrary short exact sequence $0 \to A' \to A \to A'' \to 0$ satisfying the following conditions: (i) $\partial \circ T^i(f'') = T^{i+1}(f') \circ \partial$ for a morphism f of short exact sequences; and (ii) the sequence $\ldots \to T^{i-1}(A'') \xrightarrow{\partial} T^i(A') \to T^i(A) \to T^i(A'') \xrightarrow{\partial} T^{i+1}(A') \to \ldots$ constitutes a complex. $T = \{T^i\}$ is called a covariant **∂^*-functor** if instead of ∂ there are given $\partial^*: T^i(A'') \to T^{i-1}(A')$ satisfying similar conditions (i*) and (ii*). By taking duals, we define the notion of contravariant ∂- and ∂^*-functors. They are also called **connected sequences of functors**. A ∂-(∂^*-)functor defined for $-\infty < i < +\infty$ is called a **cohomological functor (homological functor)** if the sequence in condition (ii) (resp. (ii*)) is always exact. A morphism of ∂-functors $f: S \to T$ consists of natural transformations $f^i: S^i \to T^i$ that commute with the connecting morphisms. A ∂-functor S defined for $a \leqslant i < b$ is called **universal** if for any ∂-functor T defined in the same interval and any natural transformation $\varphi: S^a \to T^a$, there exists one and only one morphism $f: S \to T$ such that $f^a = \varphi$. Let $F: \mathscr{C} \to \mathscr{C}'$ be a covariant functor and b any positive integer. A universal covariant ∂-functor S defined for $0 \leqslant i < b$ is called a **right satellite** of F if $S^0 = F$ (S is then denoted by $\{S^i F\}$). If such an S exists, then it is unique and satisfies $S^{i+1}(F) = S^1(S^i F)$. If $\mathscr{C}$ has enough injective objects, the right satellites always exist, and if F is left exact, then $\{S^i F\}$ is a cohomological functor. The universality of ∂^*-functors is defined by reversing the arrows; the satellites $\{S_i F\}$ are then written as $\{S^{-i} F\}$ and called the **left satellites**.

Let $\mathscr{C}$ be an Abelian category with enough injectives. An **injective resolution** of an object A is a right resolution $Q = \{Q^i\}$ such that all Q^i are injective. Every A admits an injective resolution, which is unique up to chain equivalence (H. Cartan, 1950). For a covariant functor $F: \mathscr{C} \to \mathscr{C}'$, the functor $A \to H^i(F(Q))$, called the ith **right derived functor** $R^i F$ of F, is independent of Q. $\{R^i F\}$ is a cohomological functor. By the universality of satellites, there exists a morphism of ∂-functors $\{S^i F\} \to \{R^i F\}$ which is an isomorphism if and only if F is left exact. The **left derived functors** $L_i F$ of a contravariant functor F are defined similarly and are isomorphic to the left satellites when F is right exact. If $\mathscr{C}$ has enough projectives (instead of injectives), we define left (right) derived functors of covariant (contravariant) functors via **projective resolutions**. For a multifunctor, we define **partial derived functors** as well as (total) derived functors of the functor viewed as a functor defined in the †product category. For instance, let $T(A, B): \mathscr{C}_1 \times \mathscr{C}_2 \to \mathscr{C}'$ be contravariant in A and covariant in B. When $\mathscr{C}_2$ has enough injectives, we obtain $R_2^i T(A, B) = H^i(T(A, Q))$ using an injective resolution Q of B. Suppose that T satisfies condition (i) $A \to T(A, B)$ is exact for any injective B. Then for a fixed injective B, $R_2^i T(A, B)$ is a cohomological functor in A. When A has a projective resolution P in $\mathscr{C}_1$, we obtain $R_1^i T(A, B) = H^i(T(P, B))$ as well as the equation for the total derived functor $R^i T(A, B) = H^i(T(P, Q))$. We say that a functor T is **right balanced** if it satisfies (i) and also (ii) $B \to T(A, B)$ is exact for any projective A. In this case, the three derived functors are isomorphic. The **left balanced** functors are defined similarly. When the right derived functors of the functor Hom (which defines the category) exist, they are denoted by $\operatorname{Ext}^i(A, B)$.

J. Spectral Sequences

In this section, we deal with cohomology in the category ${}_R\mathscr{M}$ of R-modules. A similar theory for homology is obtained by modifying the theory in a natural way. Similar constructions are also possible for general Abelian categories [3, 8].

A **filtration** F of a module A is a family of submodules $\{F^p(A) \mid p \in \mathbf{Z}\}$ such that $F^p(A) \supset F^{p+1}(A)$. We say that the filtration F is **convergent from above** (or **exhaustive**) if $\bigcup_p F^p(A) = A$, and F is **bounded from below** (or **discrete**) if $F^p(A) = 0$ for some p. The †graded module $G(A) = \{G^p(A) = F^p(A)/F^{p+1}(A) \mid p \in \mathbf{Z}\}$ is said to be **associated** with A. A **morphism** of filtered modules $f: A \to A'$ is a module homomorphism such that $f(F^p(A)) \subset F^p(A')$. It induces a homomorphism of the graded modules $G(A) \to G(A')$. A filtration of a complex $C = \{C^n, d\}$ consists of subcomplexes $F^p(C) = \{F^p(C^n)\}$ such that $F^p(C) \supset F^{p+1}(C)$. We assume that the complex C satisfies $\bigcup_p F^p(C) = C$, and is bounded from below; i.e., for every n there exists some p such that $F^p(C^n) = 0$. In particular, if $F^0(C) = C$, $F^{p+1}(C^p) = 0$, the complex C is called **canonically bounded**. Writing $C^{p,q} = G^p(C^{p+q})$, we obtain a †bigraded module $\{C^{p,q}\}$, in which p, q, and $p+q$ are called the

filtration degree, the **complementary degree**, and the **total degree**, respectively.

A **spectral sequence** $\{E_r\}$ with a graded module $D=\{D^n\}$ as its **limit** (denoted by $E_2^{p,q}\Rightarrow {}_pD^n$) consists of a family of doubly graded modules $E_r=\{E_r^{p,q}\mid p,q\in\mathbf{Z}\}$ ($r\geqslant 2$ or sometimes $r\geqslant 1$) and differentiations $d_r: E_r^{p,q}\rightarrow E_r^{p+r,q-r+1}$ ($p,q\in\mathbf{Z}$) of degree $(r,1-r)$ satisfying $d_r^2=0$ and satisfying the following two conditions: (i) $H(E_r)$ (with respect to d_r) is isomorphic to E_{r+1} (hence there exists a sequence of graded submodules of $E_2: 0=B_2\subset B_3\subset\ldots\subset Z_3\subset Z_2=E_2$ such that $Z_r/B_r\cong E_r$); and (ii) there are submodules Z_∞ and B_∞ such that $\bigcup_k B_k\subset B_\infty\subset Z_\infty\subset\bigcap_k Z_k$, and $E_\infty=Z_\infty/B_\infty$ is isomorphic to the doubly graded module associated with a certain filtration F of D (i.e., $E_\infty^{p,q}\cong G^p(D^{p+q})$). We assume that $Z_\infty=\bigcap_k Z_k$ and $B_\infty=\bigcup_k B_k$ (**weak convergence**). Suppose that F is convergent from above and bounded from below and that $Z_k(E_2^{p,q})$ is stationary for every p,q. Then $\{E_r\}$ is called **regular**. $\{E_r\}$ is **bounded from below** if for every n there exists a p_0 such that $E_2^{p,n-p}=0$ for $p<p_0$. In particular, if $E_2^{p,q}=0$ ($p<0,q<0$), then $\{E_r\}$ is called the **first quadrant** (or **cohomology spectral sequence**). In the latter case, the **edge homomorphisms** $E_2^{p,0}\rightarrow E_\infty^{p,0}$, $E_\infty^{0,q}\rightarrow E_2^{0,q}$ are defined through **base terms** $E_k^{p,0}$ and **fiber terms** $E_k^{0,q}$, respectively. A morphism of spectral sequences $f:\{E_r,D\}\rightarrow\{E_r',D'\}$ consists of $f_r: E_r\rightarrow E_r'$ of degree $(0,0)$ and $f: D\rightarrow D'$ of degree 0 which preserve the mechanism of spectral sequences. When the spectral sequences are regular, a morphism f is an isomorphism if one of the f_r is an isomorphism. Addition is naturally introduced in the set of morphisms so that spectral sequences form an additive category. An additive functor from an Abelian category $\mathscr{C}$ to this category is called a **spectral functor**. A filtered complex $\{C,F\}$ gives rise to a spectral sequence $E_2^{p,q}\Rightarrow G(H(C))$ if we put $Z_r^p=\{a\in F^p(C)\mid da\in F^{p+r}(C)\}$, $B_r^p=dZ_r^{p-r}$, $E_r^p=Z_r^p/(Z_{r-1}^{p+1}+B_{r-1}^p)$, $E_r=\sum_p E_r^p$. A double complex $C=\{C^{p,q},d_{\mathrm{I}},d_{\mathrm{II}}\}$ admits two natural filtrations $F_{\mathrm{I}}: F_{\mathrm{I}}^p(C)=\sum_{s\geqslant p}\sum_q C^{s,q}$ and $F_{\mathrm{II}}: F_{\mathrm{II}}^q(C)=\sum_{t\geqslant q}\sum_p C^{p,t}$. By the procedure above, these filtrations give rise to spectral sequences $H_{\mathrm{I}}^p(H_{\mathrm{II}}^q(C))\Rightarrow {}_pH^n(C)$ and $H_{\mathrm{II}}^q(H_{\mathrm{I}}^p(C))\Rightarrow {}_qH^n(C)$, respectively. Comparison of these sequences yields many useful results. Let T be an additive covariant functor from an Abelian category $\mathscr{C}$ to ${}_R\mathscr{M}$, C be a complex in $\mathscr{C}$, and $Q=\{Q^{p,q}\}$ be an injective resolution of C. The double complex Q gives rise to spectral sequences $H^p(R^qT(C))\Rightarrow H(T(Q))$ and $R^pT(H^q(C))\Rightarrow H(T(Q))$. The limit $H(T(Q))$ is independent of Q and is called the **hypercohomology** of T with respect to C [2, 8]. We can similarly define hypercohomology of multifunctors. The theory of spectral sequences was initiated by J. Leray (1946), and suitable algebraic formulations were given by J. L. Koszul (1950).

K. Categories of Modules

The category ${}_R\mathscr{M}$ (resp. $\mathscr{M}_R$) of left (right) R-modules over a †unitary ring R is not only an Abelian category but also a Grothendieck category (→ 277 Modules). The †full embedding theorem permits us to deduce many propositions about general Abelian categories from the consideration of ${}_R\mathscr{M}$. An object P of ${}_R\mathscr{M}$ is projective if and only if it is isomorphic to a direct summand of a †free module. Any projective module is the direct sum of countably generated projective modules (I. Kaplansky, 1958). It follows that any projective module over a †local ring is free. Finitely generated projective modules P_1 and P_2 are said to be equivalent if there exist finitely generated free modules F_1, F_2 such that $P_1\oplus F_1\cong P_2\oplus F_2$. The equivalence classes then form an Abelian group with respect to the direct sum construction called the **projective class group** of a ring R. The category of complex †vector bundles over a compact space X is equivalent to the category of projective modules over $\mathbf{C}(X)$, the ring of complex-valued continuous functions on X, and similarly for other types of spaces and bundles. Many investigations have been made involving the problem of whether every projective module over a polynomial ring is free (J.-P. Serre, 1955). This problem was settled affirmatively by D. Quillen [13] and independently by A. Suslin. It has been observed that "big" projective modules are often free: for example, nonfinitely generated projective modules over an †indecomposable weakly Noetherian ring are free (Y. Hinohara, 1963).

The nth right derived functor of $\mathrm{Hom}_R(A,B)$ is denoted by $\mathbf{Ext}_R^n(A,B)$ (→ Section G). This is a bifunctor ${}_R\mathscr{M}\times{}_R\mathscr{M}\rightarrow(\mathrm{Ab})$, contravariant in A and covariant in B. Ext_R^0 is isomorphic to and identified with Hom_R. An exact sequence $0\rightarrow A'\rightarrow A\rightarrow A''\rightarrow 0$ gives rise to the connecting homomorphisms $\Delta^n:\mathrm{Ext}_R^n(A',B)\rightarrow\mathrm{Ext}_R^{n+1}(A'',B)$, and the following sequence is exact: $\ldots\rightarrow\mathrm{Ext}_R^{n-1}(A',B)\xrightarrow{\Delta}\mathrm{Ext}_R^n(A'',B)\rightarrow\mathrm{Ext}_R^n(A,B)\rightarrow\mathrm{Ext}_R^n(A',B)\xrightarrow{\Delta}\mathrm{Ext}_R^{n+1}(A'',B)\rightarrow\ldots$ (the exact sequence of Ext). Similarly, an exact sequence $0\rightarrow B'\rightarrow B\rightarrow B''\rightarrow 0$ gives rise to $\Delta^n:\mathrm{Ext}_R^n(A,B'')\rightarrow\mathrm{Ext}_R^{n+1}(A,B')$ and to an exact sequence of Ext. An **extension** of A by B (or of B by A) is an exact sequence $(E): 0\rightarrow B\rightarrow X\rightarrow A\rightarrow 0$. The set of equivalence classes of extensions of A by B is in one-to-one correspondence with $\mathrm{Ext}_R^1(A,B)$ by assigning to (E) its **characteristic class** $\chi_E=\Delta^0(1)\in\mathrm{Ext}_R^1(A,B)$,

where 1 denotes the identity of $\mathrm{Hom}_R(B,B)$. In this correspondence, the sum of two extensions is obtained by a construction called **Baer's sum** of extensions. Similarly, $\mathrm{Ext}_R^n(A,B)$ is interpreted as the set of the equivalence classes of n-fold extensions $0\to B\to X_{n-1}\to\ldots\to X_0\to A\to 0$ (exact). This point of view permits us to establish a theory of Ext, etc., in more general (additive) categories lacking enough projectives or injectives (N. Yoneda, 1954, 1960).

The tensor product $A\otimes_R B$ is a right exact covariant bifunctor $\mathcal{M}_R\otimes_R\mathcal{M}\to(\mathrm{Ab})$. If the functor $t_P(\cdot)=\cdot\otimes P$ is exact, P is called a †flat module. A projective module is flat. In general, a flat module is the †inductive limit of finitely generated free modules (M. Lazard, 1964). A flat module P is called †faithfully flat if $P\neq\mathfrak{m}P$ for every maximal ideal $\mathfrak{m}$ of R. The functors $\otimes$ and Hom are related by †adjointness (→ 52 Categories and Functors). From this viewpoint $\otimes$ can be introduced in more general categories. Left-derived functors of $A\otimes_R B$ are denoted by $\mathbf{Tor}_n^R(A,B)$ and are called nth **torsion products** of A and B. $\mathrm{Tor}_1^R(A,B)$ is often denoted by $A*_R B$. The functor $\otimes_R$ is left balanced, hence Tor is calculated by using projective resolutions of A, B, or both A and B. We have $\mathrm{Tor}_0^R=\otimes_R$. An exact sequence $0\to A'\to A\to A''\to 0$ gives rise to $\Delta_n:\mathrm{Tor}_{n+1}^R(A'',B)\to\mathrm{Tor}_n^R(A',B)$ and the infinite exact sequence of Tor, and similarly for the second variables.

From the various relations between Hom and $\otimes$ follow the corresponding relations between their derived functors. When Λ and Γ are algebras over K and $\Omega=\Lambda\otimes\Gamma$, we can define the **external product** ($\top$ **-product**), which is a mapping

$$\top:\mathrm{Tor}_p^\Lambda(A,B)\otimes\mathrm{Tor}_q^\Gamma(A',B')\to\mathrm{Tor}_{p+q}^\Omega(A\otimes A',B\otimes B').$$

In particular, if Λ and Γ are K-projective and $\mathrm{Tor}_n^k(A,A')=0$ $(n>0)$, then we can define the **wedge product** ($\vee$**-product**) $\vee:\mathrm{Ext}_\Lambda^p(A,B)\otimes\mathrm{Ext}_\Gamma^q(A',B')\to\mathrm{Ext}_\Omega^{p+q}(A\otimes A',B\otimes B')$. The latter is described in terms of the composition of module extensions. When $K=\Lambda=\Gamma=\Omega$, the $\top$-product reduces to the **internal product**, called the $\pitchfork$**-product**. If Λ is a †Hopf algebra over K, the †comultiplication $\Lambda\to\Lambda\otimes\Lambda$ induces $\mathrm{Ext}_{\Lambda\otimes\Lambda}\to\mathrm{Ext}_\Lambda$. This, combined with the $\vee$-product, yields the **cup product** ($\smile$**-product**)$\smile:\mathrm{Ext}_\Lambda^p(A,B)\otimes\mathrm{Ext}_\Lambda^q(A',B')\to\mathrm{Ext}_\Lambda^{p+q}(A\otimes A',B\otimes B')$. We define similarly $\perp$**-product**, $\wedge$**-product**, $\uplus$**-product**, and $\frown$**-product (cap product)** [2]. Let Λ, Γ, and Σ be algebras over K, with Λ K-projective; let $A\in\mathcal{M}_{\Lambda\otimes\Gamma}$, $B\in{}_\Lambda\mathcal{M}_\Sigma$, $C\in\mathcal{M}_{\Gamma\otimes\Sigma}$, and assume $\mathrm{Tor}_n^\Lambda(A,B)=0$ $(n>0)$. The natural isomorphism $\mathrm{Hom}_{\Lambda\otimes\Gamma}(A,\mathrm{Hom}_\Sigma(B,C))\cong\mathrm{Hom}_{\Gamma\otimes\Sigma}(A\otimes_\Lambda B,C)$ then yields a spectral sequence $\mathrm{Ext}_{\Lambda\otimes\Gamma}^p(A,\mathrm{Ext}_\Sigma^q(B,C))\Rightarrow_p\mathrm{Ext}_{\Gamma\otimes\Sigma}^n(A\otimes_\Lambda B,C)$ by the double complex argument in Section D.

The **homological dimension** $\mathrm{h\,dim}_R A$, $\mathrm{dh}_R A$, or **projective dimension** $\mathrm{proj\,dim}_R A$ of $A\in{}_R\mathcal{M}$ is the supremum $(\leqslant\infty)$ of n such that $\mathrm{Ext}_R^n(A,B)\neq 0$ for some B. The relation $\mathrm{h\,dim}_R A\leqslant 0$ means that A is projective. The **injective dimension** $\mathrm{inj\,dim}_R B$ of $B\in{}_R\mathcal{M}$ is defined similarly by means of the functor $\mathrm{Ext}_R^n(\cdot,B)$, and the **weak dimension** $\mathrm{w\,dim}_R C$ of $C\in{}_R\mathcal{M}$ by the functor $\mathrm{Tor}_n^R(\cdot,C)$. The common value $\sup\{\mathrm{proj\,dim}_R A\mid A\in{}_R\mathcal{M}\}=\sup\{\mathrm{inj\,dim}_R B\mid B\in{}_R\mathcal{M}\}$ is called the **left global dimension** $\mathrm{l\ gl\ dim}\,R$ of R. It is identical with the supremum of homological dimensions of †cyclic modules (M. Auslander, 1955). The **right global dimension** $\mathrm{r\ gl\ dim}\,R$ is defined similarly. The common value $\sup\{\mathrm{w\,dim}_R A\mid A\in\mathcal{M}_R\}=\sup\{\mathrm{w\,dim}_R C\mid C\in{}_R\mathcal{M}\}$ is called the **weak global dimension** $\mathrm{w\ gl\ dim}\,R$ of R. We have $\mathrm{w\ gl\ dim}\,R\leqslant\mathrm{l\ gl\ dim}\,R$, $\mathrm{r\ gl\ dim}\,R$. The equality may fail to hold (Kaplansky, 1958). If R is †Noetherian, the three global dimensions coincide (Auslander, 1955) and are called simply the **global dimension** of R: $\mathrm{gl\,dim}\,R$. The condition $\mathrm{l\ gl\ dim}\,R=0$ (or $\mathrm{r\ gl\ dim}\,R=0$) holds if and only if R is an †Artinian semisimple ring, while $\mathrm{w\ gl\ dim}\,R=0$ if and only if R is a †regular ring in the sense of J. von Neumann (M. Harada, 1956). A ring R is called **left (right) hereditary** if $\mathrm{l\ gl\ dim}\,R\leqslant 1$ ($\mathrm{r\ gl\ dim}\,R\leqslant 1$), and **left (right) semihereditary** if every finitely generated left (right) ideal is projective. A left and right (semi)hereditary ring is called a **(semi)hereditary ring**. Since projectivity and †invertibility of an ideal of a (commutative) integral domain R are equivalent, R is hereditary if and only if R is a †Dedekind ring. In this case, the projective class group of R reduces to the †ideal class group. An integral domain R is semihereditary if and only if $\mathrm{w\ gl\ dim}\,R\leqslant 1$ (A. Hattori, 1957), and in that case R is called a **Prüfer ring**. A †maximal order over a Dedekind ring is hereditary. A commutative semihereditary ring R is characterized by the property that flatness of R-modules is equivalent to torsion-freeness (S. Endo, 1961). A Noetherian ring R is left self-injective if and only if R is †quasi-Frobenius (M. Ikeda, 1952), and the global dimension of a quasi-Frobenius ring is 0 or ∞ (S. Eilenberg and T. Nakayama, 1955). A polynomial ring $R=K[X_1,\ldots,X_n]$ over a commutative ring K satisfies $\mathrm{gl\ dim}\,R=\mathrm{gl\ dim}\,K+n$. When K is a field, this is a reformulation of Hilbert's theory of syzygy sequences (→ 369 Rings of Polynomials). In this sense, the study of the global dimension of rings and categories is sometimes called **syzygy theory** (Eilenberg, 1956). The homological algebra of commutative Noetherian rings has been studied extensively and

is useful in algebraic geometry. Since $\operatorname{gl\,dim} R = \sup_{\mathfrak{m}} \operatorname{gl\,dim} R_{\mathfrak{m}}$ ($R_{\mathfrak{m}}$ is the †ring of quotients relative to $\mathfrak{m}$), with $\mathfrak{m}$ running over the maximal ideals of R, the problem of determining $\operatorname{gl\,dim} R$ reduces to the case of †local rings. A finitely generated flat module over a local ring R is free. If K denotes the residue field $R/\mathfrak{m}$, where $\mathfrak{m}$ is the maximal ideal of the local ring R, $\operatorname{Tor}^R(K,K)$ has the structure of a Hopf algebra (E. F. Assmus, Jr., 1959). Detailed results concerning the **Betti numbers** $\dim \operatorname{Tor}_i^R(K,K)$ of R have been obtained (J. Tate, 1957, and others). In particular, R is †regular if and only if $\operatorname{gl\,dim} R < \infty$ (Serre, 1955). A local ring R is called a **Gorenstein ring** if the injective dimension of R-module R is finite. This is a notion intermediate between regular rings and †Macaulay rings (→ 284 Noetherian Rings).

Consideration of a ring R in relation to a subring S leads to **relative homological algebra**. Foundations for this theory were established by Hochschild (1956). An exact sequence of R-modules that †splits as a sequence of S-modules is called an **(R,S)-exact sequence**. An R-module P is called an **(R,S)-projective module** if $\operatorname{Hom}_R(P,\cdot)$ maps any (R,S)-exact sequence to an exact sequence. **(R,S)-injective modules** are defined similarly. Based on these notions, $\operatorname{Ext}_{(R,S)}$ and $\operatorname{Tor}^{(R,S)}$ are defined as the **relative derived functors** of Hom_R and $\otimes_R$, respectively. We also have a relative theory from a different viewpoint (S. Takasu, 1957). Relative theory is extended to general categories from various viewpoints [6, 14] (→ Section Q).

L. Cohomology Theory for Associative Algebras

Let Λ be an †algebra over a commutative ring K and A a †two-sided Λ-module. Let C^n be the module of all n-linear mappings of Λ into A called **n-cochains** ($C^0=A$). Define the coboundary operator $\delta^n: C^n \to C^{n+1}$ by $(\delta^n f)(\lambda_1,\ldots,\lambda_{n+1}) = \lambda_1 f(\lambda_2,\ldots,\lambda_{n+1}) + \sum_{i=1}^n (-1)^i f(\lambda_1,\ldots,\lambda_i\lambda_{i+1},\ldots,\lambda_{n+1}) + (-1)^{n+1} f(\lambda_1,\ldots,\lambda_n)\lambda_{n+1}$.
We thus obtain a complex whose cohomology is denoted by $H^n(\Lambda,A)$ and is called the nth **Hochschild's cohomology group** of Λ relative to the **coefficient module** A (Hochschild, 1945). A cochain f is called **normalized** if $f(\lambda_1,\ldots,\lambda_n)=0$ whenever one of the λ_i is 1. We obtain the same cohomology group $H^n(\Lambda,A)$ from the subcomplex of normalized cochains. $\{H^n(\Lambda,\cdot)\}$ is a cohomological functor from the category ${}_\Lambda\mathcal{M}_\Lambda$ of two-sided Λ-modules to the category ${}_K\mathcal{M}$. Using the **enveloping algebra** $\Lambda^e = \Lambda \otimes_K \Lambda^0$, where Λ^0 is an anti-isomorphic copy of Λ, ${}_\Lambda\mathcal{M}_\Lambda$ can be identified with ${}_{\Lambda^e}\mathcal{M}$ (and $\mathcal{M}_{\Lambda^e}$). If Λ is K-projective, $\{H^n(\Lambda,\cdot)\}$ is isomorphic to $\{\operatorname{Ext}^n_{\Lambda^e}(\Lambda,\cdot)\}$. (In [2], $H^n(\Lambda,A)$ is defined as $\operatorname{Ext}^n_{\Lambda^e}(\Lambda,A)$ in general.) We have $H^0(\Lambda,A)=\{a\in A \mid \lambda a = a\lambda, \forall\lambda\in\Lambda\}$. We call 1-cocycles **derivations** (or **crossed homomorphisms**) of Λ in A and 1-coboundaries **inner derivations**. Thus $H^1(\Lambda,A)$ is the derivation class group and is related to the †ramification. When K is a field, $H^1(\Lambda,\cdot)=0$ if and only if Λ is a †separable algebra. In general, an algebra Λ over a commutative ring K is called a **separable algebra** if Λ is Λ^e-projective, i.e., if $\operatorname{Ext}^1_{\Lambda^e}(\Lambda,\cdot)=0$ (Auslander and O. Goldman, 1960). This is a generalization of the notion of †maximally central algebras (Nakayama and G. Azumaya, 1948). We have a one-to-one correspondence of $H^2(\Lambda,A)$ to the family of **algebra extensions** of Λ with kernel A (i.e., K-algebras Γ containing A as a two-sided ideal such that $\Gamma/A=\Lambda$) satisfying $A^2=0$. Any extension of an algebra Λ over a field K such that $H^2(\Lambda,\cdot)=0$ splits over a nilpotent kernel (J. H. C. Whitehead and G. Hochschild). This holds in particular for a separable algebra, and we obtain the †Wedderburn-Mal'tsev theorem. There are some interpretations of $H^3(\Lambda,A)$ in terms of extensions.

The supremum ($\leqslant\infty$) of n such that $H^n(\Lambda,A)\neq 0$ for some A is called the **cohomological dimension** of Λ and written $\dim\Lambda$. For a finite-dimensional algebra Λ over a field K, $\dim\Lambda<\infty$ if and only if Λ/N is separable and $\operatorname{gl\,dim}\Lambda<\infty$, where N is the †radical of Λ (N. Ikeda, H. Nagao, and Nakayama, 1954).

The homology groups $H_n(\Lambda,A)$ of Λ relative to a coefficient module A are defined similarly. If Λ is K-projective, $\{H_n(\Lambda,\cdot)\}$ is isomorphic to $\{\operatorname{Tor}_n^{\Lambda^e}(\cdot,\Lambda)\}$.

M. Cohomology of Groups

The pair consisting of an algebra Λ over K and an algebra homomorphism $\varepsilon:\Lambda\to K$ is called a **supplemented algebra** [2] (or **augmented algebra** [6]), of which ε is the **augmentation**. The †group algebra $\mathbf{Z}[G]$ of a group G over the ring of rational integers is a supplemented algebra, in which the augmentation is defined by $\varepsilon(x)=1$ ($x\in G$). The category of left G-modules is identified with the category of left $\mathbf{Z}[G]$-modules. For a finite group G, a finitely generated projective G-module is not necessarily free (D. S. Rim, 1959) and is isomorphic to the direct sum of a free module and a left ideal of $\mathbf{Z}[G]$. It follows that the projective class group of $\mathbf{Z}[G]$ is a finite group (R. G. Swan, 1960). The **cohomology groups** and **homology groups** of G relative to $A\in{}_G\mathcal{M}$ (Eilenberg and S. MacLane, 1943) are defined

by $H^n(G, A) = \mathrm{Ext}^n_{\mathbf{Z}[G]}(\mathbf{Z}, A)$ and $H_n(G, A) = \mathrm{Tor}_n^{\mathbf{Z}[G]}(\mathbf{Z}, A)$, respectively. Their concrete description is given usually via the $\mathbf{Z}[G]$-**standard resolution** of $\mathbf{Z}$.

(1) Homogeneous formulation. The group of **homogeneous** n**-chains** is the free Abelian group with basis $G \times \ldots \times G$ ($n+1$ times), on which G operates by $x(x_0, \ldots, x_n) = (xx_0, \ldots, xx_n)$, and the boundary operator is defined by $d(x_0, \ldots, x_n) = \sum_{i=0}^n (-1)^i (x_0, \ldots, \hat{x}_i, \ldots, x_n)$.

(2) Nonhomogeneous formulation. The group of **nonhomogeneous** n**-chains** is the $\mathbf{Z}[G]$-free module with basis $G \times \ldots \times G$ (n times), and the boundary operator is defined by $d(x_1, \ldots, x_n) = x_1(x_2, \ldots, x_n) + \sum_{i=1}^{n-1} (-1)^i (x_1, \ldots, x_i x_{i+1}, \ldots, x_n) + (-1)^n (x_1, \ldots, x_{n-1})$. A nonhomogeneous 2-cocycle is sometimes called a **factor set**. $H^0(G, A)$ is the submodule A^G of A consisting of the G-invariant elements, while $H_0(G, A)$ is the largest residue class module A_G of A on which G acts trivially. Given two groups G and K, an exact sequence of group homomorphisms $1 \to K \to E \to G \to 1$ is called a **group extension** of G over the kernel K. When K is Abelian, the extension canonically induces a G-module structure on K, and the deviation of K from being a semidirect factor of E is measured by a factor set. The group $H^2(G, A)$ is thus in one-to-one correspondence with the set of equivalence classes of the †group extensions of G over A which induce the originally given G-module structure on A (→ 190 Groups N). This point of view is essential in the proof of the †Schur-Zassenhaus theorem (→ 151 Finite Groups). $H^3(G, A)$ is interpreted as the set of obstructions for extensions (Eilenberg and MacLane, 1947). For a †free group F, $H^n(F, A) = 0$ $(n > 1)$. If a group G is represented as a factor group F/R of a free group F, we have a group extension $1 \to K \to E \to G \to 1$, where $K = R/[R, R]$ and $E = F/[R, R]$. Let $\xi \in H^2(G, K)$ correspond to this extension. Then for any G-module A, the cup product $\chi \to \chi \smile \xi$ followed by the pairing $\mathrm{Hom}(K, A) \otimes K \to A$ provides isomorphisms $H^n(G, \mathrm{Hom}(K, A)) \cong H^{n+2}(G, A)$ $(n > 0)$ (the **cup product reduction theorem** of Eilenberg and MacLane, 1947); similarly, we have the reduction theorem for the homology. The $\mathbf{Z}$-algebra $H(G, \mathbf{Z}) = \sum_{n=0}^{\infty} H^n(G, \mathbf{Z})$ under the multiplication defined by the cup product is finitely generated if G is a finite group (B. B. Venkov, 1959; L. Evens, 1961).

The following are mappings relative to a subgroup H.

(1) The inner automorphism by $x \in G$ induces an isomorphism of $H^n(H, A)$ and $H^n(xHx^{-1}, A)$ which reduces to the identity of $H^n(G, A)$ if $H = G$. Hence if H is a normal subgroup, $H^n(H, A)$ has the structure of a G/H-module, and similarly for $H_n(H, A)$.

(2) When H is a normal subgroup of G, the mapping $(x_1, \ldots, x_n) \to (x_1 H, \ldots, x_n H)$ of nonhomogeneous chains induces the **inflation** (or **lift**) $\mathrm{Inf}: H^n(G/H, A^H) \to H^n(G, A)$ and the **deflation** $\mathrm{Def}: H_n(G, A) \to H_n(G/H, A_H)$.

(3) The embedding of nonhomogeneous chains induces the **restriction** $\mathrm{Res}: H^n(G, A) \to H^n(H, A)$ and the **injection** Inj (or **corestriction** Cor)$: H_n(H, A) \to H_n(G, A)$. The theory of †induced representation gives another construction of these mappings; that is, if we put $\iota^G(A) = \mathrm{Hom}_{\mathbf{Z}[H]}(\mathbf{Z}[G], A)$, Res is obtained by the isomorphism $H^n(G, \iota^G(A)) \cong H^n(H, A)$ combined with the homomorphism induced by $A \to \iota^G(A)$; while if we put $\iota_G(A) = \mathbf{Z}[G] \otimes_{\mathbf{Z}[H]} A$, Inj is obtained by the isomorphism $H_n(H, A) \cong H_n(G, \iota_G(A))$ followed by the homomorphism induced by $\iota_G(A) \to A$.

(4) If $(G:H) < \infty$, we have $\iota^G(A) \cong \iota_G(A)$. The composition of $H^n(H, A) \to H^n(G, \iota_G(A)) \to H^n(G, A)$ defines on the cohomology groups the $\mathrm{Inj}: H^n(H, A) \to H^n(G, A)$, while the composition $H_n(G, A) \to H_n(G, \iota^G(A)) \to H_n(H, A)$ gives the $\mathrm{Res}: H_n(G, A) \to H_n(H, A)$. In particular, $\mathrm{Res}: H_1(G, \mathbf{Z}) \to H_1(H, \mathbf{Z})$ coincides with the †transfer $G/[G, G] \to H/[H, H]$.

(5) Let H be a normal subgroup of G. Consider the additive relation ρ (the correspondence) between $h \in Z^n(H, A)^G$ and $f \in Z^{n+1}(G/H, A^H)$ determined by $\rho(h, f)$ if and only if there exists a $g \in C^n(G, A)$ such that $h = \mathrm{Res}\, g$ and $\mathrm{Inf}\, f = \delta g$. If the relation induces a homomorphism $H^n(H, A)^G \to H^{n+1}(G/H, A^H)$, then it is called the **transgression**. If $H^i(H, A) = 0$ $(0 < i < n)$, the sequence $0 \to H^n(G/H, A^H) \to H^n(G, A) \to H^n(H, A)^G \to H^{n+1}(G/H, A^H) \to H^{n+1}(G, A)$ composed of inflation, restriction, and transgression mappings is exact (Hochschild and Serre, 1953) and is called the **fundamental exact sequence**. This exact sequence can be derived from a certain spectral sequence $H^p(G/H, H^q(H, A)) \Rightarrow_p H^n(G, A)$ (R. C. Lyndon, 1948; Hochschild and Serre, 1953).

The relative (co)homology theory relative to a subgroup (I. T. Adamson, 1954) can be dealt with in terms of the relative Ext and the relative Tor (Hochschild, 1956). Many results in the absolute case are generalized to the relative case: for example, the fundamental exact sequence (Nakayama and Hattori, 1958). The relative theory is further generalized to the cohomology theory of †permutation representations of G (E. Snapper, 1964).

Non-Abelian Cohomology. For a non-Abelian G-group A, the cohomology "set" $H^1(G, A)$ (and $H^0(G, A)$) is defined as in the Abelian case by means of the nonhomogeneous cochains (→ e.g., [9]). Some efforts are being made toward the construction of a more general non-

Abelian theory (J. Giraud, *Cohomologie non abeliénne*, Springer, 1971).

N. Finite Groups

Let G be a finite group and A a G-module. Define the norm $N: A \to A$ by $N(a)=\sum_{x \in G} xa$, and denote $\operatorname{Ker} N$ by ${}_N A$. The kernel of the augmentation $\varepsilon: \mathbf{Z}[G] \to \mathbf{Z}$ is denoted by I. Put $\hat{H}^n(G, A)=H^n(G, A)$ $(n>0)$, $\hat{H}^0(G, A)=A^G/NA$, $\hat{H}^{-1}(G, A)={}_N A/IA$, and $\hat{H}^{-n}(G, A)=H_{n-1}(G, A)$ $(n>1)$. Then $\{\hat{H}^n(G, \cdot)\}$ forms a cohomological functor, called the **Tate cohomology** (E. Artin and J. T. Tate), that can be described as the set of cohomology groups concerning a certain complex called a **complete free resolution** of **Z**. (Similar arguments are valid more generally for quasi-Frobenius rings (Nakayama, 1957), and a theory of this kind is called **complete cohomology theory**.) We have $\hat{H}^n(G, A)=0$ $(n \in \mathbf{Z})$ if and only if $\operatorname{hdim}_{Z[G]} A \leqslant 1$ (Nakayama, 1957). If A satisfies the conditions (i) $\hat{H}^1(G_p, A)=0$ for any Sylow p-subgroup G_p of G, and (ii) there exists a $\xi \in \hat{H}^2(G, A)$ such that $\operatorname{Res} \xi \in \hat{H}^2(G_p, A)$ has the same order as G_p and generates all of $\hat{H}^2(G_p, A)$, then the homomorphisms $\hat{H}^n(H, B) \to \hat{H}^{n+2}(H, A \otimes B)$ $(n \in \mathbf{Z})$ defined by the cup product with $\operatorname{Res} \xi$ are isomorphisms for every subgroup H and every G-module B such that $\operatorname{Tor}(A, B)=0$ (Nakayama, 1957; for $B=\mathbf{Z}$, Tate, 1952). If G is cyclic, the mappings $\hat{H}^n(A) \to \hat{H}^{n+2}(A)$ $(n \in \mathbf{Z})$ defined by the cup product with a generator of $\hat{H}^2(\mathbf{Z})$ are isomorphisms. (The notation is abbreviated by omitting G.) If the orders of $\hat{H}^0(A)$ and $\hat{H}^1(A)$ are finite, their ratio is called the **Herbrand quotient** $h(A)$ of A. If $0 \to A' \to A \to A'' \to 0$ is exact, then $h(A)=h(A')h(A'')$. If A is finite, then $h(A)=1$. By combining these two facts we obtain **Herbrand's lemma**: If A' is a G-submodule of A of finite index and $h(A')$ exists, then $h(A)$ also exists and $h(A)=h(A')$. The periodicity $\hat{H}^n(A)=\hat{H}^{n+p}(A)$ $(n \in \mathbf{Z}, A \in {}_G\mathscr{M})$ holds if and only if every Sylow subgroup is cyclic or a †generalized quaternion group (Artin and Tate; [2]).

Let L/K be a finite †Galois extension with the †Galois group G. The cohomology groups of various types of G-modules related to L/K are called the **Galois cohomology** groups (→ 172 Galois Theory). Using **continuous cocycles**, a cohomology theory (**Tate cohomology**) is developed for infinite Galois extensions as well [9, 10]. By means of Galois cohomology (→ 59 Class Field Theory), the cohomology theory of finite groups and of †totally disconnected compact groups (which are †profinite groups) plays an important role in class field theory and its related areas.

O. Cohomology Theory of Lie Algebras

Let $\mathfrak{g}$ be a †Lie algebra over a commutative ring K, and assume that $\mathfrak{g}$ is K-free. The †enveloping algebra $U=U(\mathfrak{g})$ is a †supplemented algebra over K. For a $\mathfrak{g}$-module ($=U$-module) A, $\operatorname{Ext}_U^n(K, A)$ and $\operatorname{Tor}_n^U(K, A)$ are called the **cohomology groups** $H^n(\mathfrak{g}, A)$ and **homology groups** $H_n(\mathfrak{g}, A)$, respectively, of $\mathfrak{g}$ relative to the coefficient module A. They are usually described by means of the U-free resolution $U \otimes \bigwedge_K(\mathfrak{g})$ of K (called the **standard complex** of $\mathfrak{g}$) constructed by C. Chevalley and Eilenberg (1948), where $\bigwedge_K(\mathfrak{g})$ is the exterior algebra of the K-module $\mathfrak{g}$ and (denoting $1 \otimes (x_1 \wedge \ldots \wedge x_n)$ by $(x_1, \ldots, x_n)$) the differentiation is given by

$$d(x_1, \ldots, x_n)=\sum_{i=1}^{n}(-1)^{i+1} x_i(x_1, \ldots, \hat{x}_i, \ldots, x_n) + \sum_{1 \leqslant i<j \leqslant n}(-1)^{i+j}([x_i, x_j], x_1, \ldots, \hat{x}_i, \ldots, \hat{x}_j, \ldots, x_n).$$

For $n>[\mathfrak{g}:K]$, $H^n(g, A)=H_n(\mathfrak{g}, A)=0$. If $\mathfrak{g}$ is a †semisimple Lie algebra over a field K of characteristic 0, we have $H^1(\mathfrak{g}, A)=0$, $H^2(\mathfrak{g}, A)=0$, while $H^3(\mathfrak{g}, K) \neq 0$. $H^1(\mathfrak{g}, A)=0$ is equivalent to Weyl's theorem, which asserts the complete reducibility of finite-dimensional representations (→ 248 Lie Algebras E). $H^2(\mathfrak{g}, A)$ and $H^3(\mathfrak{g}, A)$ are interpreted by means of Lie algebra extensions as in the cohomology of groups. The theorem on †Levi decomposition is derived from $H^2(\mathfrak{g}, A)=0$. Chevalley and Eilenberg constructed this cohomology by algebraization of the cohomology of compact †Lie groups. They also introduced the notion of relative cohomology groups $H^n(\mathfrak{g}, \mathfrak{h}, A)$ relative to a Lie subalgebra $\mathfrak{h}$ of $\mathfrak{g}$, which correspond to the cohomology of homogeneous spaces. $H^n(\mathfrak{g}, \mathfrak{h}, A)$ does not always coincide with $\operatorname{Ext}^n_{U(\mathfrak{g}), U(\mathfrak{h})}(K, A)$ (Hochschild, 1956), but does so in an important case where K is a field of characteristic 0 and $\mathfrak{h}$ is †reductive in $\mathfrak{g}$ (→ 248 Lie Algebras).

For †transformation spaces of †linear algebraic groups G over a field K, the **rational cohomology groups** are introduced using the notion of **rational injectivity** (Hochschild, 1961). In particular, if G is a †unipotent algebraic group over a field K of characteristic 0, then $H(G, A)$ is isomorphic to $H(\mathfrak{g}, A)$, where $\mathfrak{g}$ is the Lie algebra of G. There is also a relative theory.

P. Amitsur Cohomology

Let R be a commutative ring, and F a covariant functor from the category $\mathscr{C}_R$ of com-

mutative R-algebras to the category of Abelian groups. For $S \in \mathscr{C}_R$ and $n = 0, 1, 2, \ldots$, we write $S^{(n)} = S \otimes \ldots \otimes S$ (n-fold tensor product over R). Let $\varepsilon_i : S^{(n+1)} \to S^{(n+2)}$ ($i = 0, 1, \ldots, n+1$) be $\mathscr{C}_R$-morphisms defined by $\varepsilon_i(x_0 \otimes \ldots \otimes x_n) = x_0 \otimes \ldots \otimes x_{i-1} \otimes 1 \otimes x_i \otimes \ldots \otimes x_n$. Defining $d^n : F(S^{(n+1)}) \to F(S^{(n+2)})$ by $d^n = \sum_{i=0}^{n+1} (-1)^i F(\varepsilon_i)$, we obtain a cochain complex $\{F(S^{(n+1)}), d^n\}$. This complex and its cohomology groups are called the **Amitsur complex** and the **Amitsur cohomology groups**, and are usually denoted by $C(S/R, F)$ and $H^n(S/R, F)$, respectively.

If S/R is a finite Galois extension with Galois group G, then the group $H^n(S/R, U)$ of the unit group functor U is naturally isomorphic to $H^n(G, U(S))$. If S/R is a finite purely inseparable extension, then $H^n(S/R, U) = 0$ for $n \geqslant 3$. The group $H^2(S/R, U)$ is related to the †Brauer group $B(S/R)$ (→ 29 Associative Algebras K).

Q. Relative Theory

In the course of the development of homological algebra, it has been recognized that the notion of projective (resp. injective) resolutions should be generalized ([14, 4]; Hochschild, 1956). In the meantime, a method has been introduced that utilizes simplicial objects in order to define the derived functor of an arbitrary functor with Abelian category as its range ([15, 17]; J. Beck, 1967). As a consequence of these developments, there emerged a viewpoint, which we describe below, making it possible to unify various known definitions of (co-)homology theories that has been designed for particular cases.

Let $\mathscr{A}$ be a category and $\mathscr{P}$ a class of objects in $\mathscr{A}$. In this section, we denote the set $\mathrm{Hom}_{\mathscr{A}}(A, B)$ of morphisms by $\mathscr{A}(A, B)$. A morphism $f : A \to B$ in $\mathscr{A}$ is called a **$\mathscr{P}$-epimorphism** if the induced mapping $\mathscr{A}(P, f)$ $(= \mathrm{Hom}(P, f)) : \mathscr{A}(P, A) \to \mathscr{A}(P, B)$ is surjective for any $P \in \mathscr{P}$. The class $\mathscr{P}$ is called a **projective class** in $\mathscr{A}$ if there exist an object $P \in \mathscr{P}$ and a $\mathscr{P}$-epimorphism $f : P \to A$ for each object $A \in \mathscr{A}$. To any category $\mathscr{A}$, we associate a preadditive category $\mathbf{Z}\mathscr{A}$, adding a zero object to $\mathscr{A}$ if necessary: Put $\mathrm{Ob}\mathbf{Z}\mathscr{A} = \mathrm{Ob}\mathscr{A}$, $\mathbf{Z}\mathscr{A}(A, B) =$ free Abelian group generated by the set $\mathscr{A}(A, B)$. $\mathscr{A}$ is regarded as a subcategory of $\mathbf{Z}\mathscr{A}$ by the natural inclusion $J : \mathscr{A} \to \mathbf{Z}\mathscr{A}$. Any functor T from $\mathscr{A}$ into an Abelian category $\mathscr{B}$ has a unique additive extension $\bar{T} : \mathbf{Z}\mathscr{A} \to \mathscr{B}$ such that $T = \bar{T}J$. If $\mathscr{A}$ is an additive category, there is a canonical projection $\theta : \mathbf{Z}\mathscr{A} \to \mathscr{A}$ such that $\theta J = Id$. If furthermore T is additive, then $\bar{T} = T\theta$.

Now suppose that a projective class $\mathscr{P}$ in $\mathscr{A}$ is given. For $A \in \mathscr{A}$, an augmented chain complex $X. \to A$ in $\mathbf{Z}\mathscr{A}$,

$$\to X_n \xrightarrow{d_n} X_{n-1} \to \ldots \xrightarrow{d_1} X_0 \xrightarrow{d_0} A, \quad d_i d_{i+1} = 0,$$

is called a **$\mathscr{P}$-projective resolution** of A if (i) $X_n \in \mathscr{P}$ $(n \geqslant 0)$, and (ii) it is **$\mathscr{P}$-acyclic** (i.e., the sequence

$$\to \mathbf{Z}\mathscr{A}(P, X_n) \xrightarrow{d_n} \mathbf{Z}\mathscr{A}(P, X_{n-1}) \to \ldots \xrightarrow{d_1} \mathbf{Z}\mathscr{A}(P, X_0) \xrightarrow{d_0} \mathbf{Z}\mathscr{A}(P, A) \to 0$$

is exact for any $P \in \mathscr{P}$ [16]). Note that the $\mathscr{P}$-acyclicity in this case implies the existence of a contracting homotopy

$$h_n : \mathbf{Z}\mathscr{A}(P, X_n) \to \mathbf{Z}\mathscr{A}(P, X_{n+1}), \qquad n \geqslant -1, X_{-1} = A,$$

satisfying $d_{n+1} h_n + h_{n-1} d_n = 1$.

Comparison theorem: Given two $\mathscr{P}$-projective resolutions $X. \to A$ and $Y. \to A$, we have that the chain complexes $X.$ and $Y.$ are chain equivalent in $\mathbf{Z}\mathscr{A}$.

Let $T : \mathscr{A} \to \mathscr{B}$ be a (covariant) functor with $\mathscr{B}$ an Abelian category. The nth **left derived functor** $L_n T : \mathscr{A} \to \mathscr{B}$ $(n \geqslant 0)$ of T, with respect to $\mathscr{P}$, is defined by $L_n T(A) = H_n(\bar{T}X.)$ for a $\mathscr{P}$-projective resolution $X. \to A$. The derived functors $L_n T$ will remain unaffected if we replace the projective class $\mathscr{P}$ by its enlargement $\bar{\mathscr{P}} =$ {objects in $\mathscr{P}$ together with their retracts}.

We can easily verify that (i) $L_0 T(P) = T(P)$ and $L_n T(P) = 0$ $(n > 0)$ for $P \in \mathscr{P}$, and (ii) a short exact sequence $0 \to T' \to T \to T'' \to 0$ of functors: $\mathscr{A} \to \mathscr{B}$ induces a long exact sequence of derived functors

$$\to L_n T' \to L_n T \to L_n T'' \xrightarrow{\partial} L_{n-1} T' \to \ldots \to L_0 T'' \to 0.$$

If $\mathscr{A}$ is preadditive and has a zero object and kernels, then it is routine to give a $\mathscr{P}$-projective resolution of any object. If $\mathscr{A}$ has finite †limits (i.e., †finite products and †equalizers), it can be proved that there exists a $\mathscr{P}$-projective resolution for any object [17].

There is a standard functorial construction which provides canonically a projective class $\mathscr{P}$ in a category $\mathscr{A}$ and a $\mathscr{P}$-projective resolution of any object in $\mathscr{A}$. Let (G, ε, δ) be a **cotriple** (or **comonad** [18] or functor coalgebra) in $\mathscr{A}$. Here $G : \mathscr{A} \to \mathscr{A}$ is an endofunctor, $\varepsilon : G \to Id$ and $\delta : G \to G^2 = GG$ are natural transformations such that $G\varepsilon \circ \delta = \varepsilon G \circ \delta = 1_G$ and $G\delta \circ \delta = \delta G \circ \delta$. A cotriple comes usually from a pair (F, U) of †adjoint functors $U : \mathscr{A} \to \mathscr{C}$, $F : \mathscr{C} \to \mathscr{A}$ with natural bijection $\lambda : \mathscr{A}(F(C), A) \cong \mathscr{C}(C, U(A))$. Putting $G = FU : \mathscr{A} \to \mathscr{A}$, $\varepsilon(A) = \lambda^{-1}(1_{U(A)}) : FU(A) \to A$, $\eta(C) = \lambda(1_{F(C)}) : C \to UF(C)$, we have a cotriple $(G = FU, \varepsilon, \delta = F\eta U)$ in $\mathscr{A}$. Conversely, it is known that any cotriple in $\mathscr{A}$ is induced from a suitable pair of adjoint functors.

Given a cotriple (G, ε, δ) in $\mathscr{A}$, we define a projective class $\mathscr{P}_G = \{G(A) \mid A \in \mathscr{A}\}$ (or its enlargement $\bar{\mathscr{P}}_G$) in $\mathscr{A}$, and an augmented simplicial object $G_*(A) \rightarrow A$ for any object A as follows. Put $G_n(A) = G^{n+1}(A)$ $(n \geqslant 0)$, $\partial_i = \partial_i^n = G^i \varepsilon G^{n-i}(A): G_n(A) \rightarrow G_{n-1}(A)$ (face operator) and $\delta_i = \delta_i^n = G^i \delta G^{n-i}(A): G_n(A) \rightarrow G_{n+1}(A)$ (degeneracy operator) for $0 \leqslant i \leqslant n$, $G_{-1}(A) = A$. Then $G_*(A) \rightarrow A$ gives rise to a $\mathscr{P}_G$ (or equivalently $\bar{\mathscr{P}}_G$)-projective resolution of A in $\mathbf{Z}\mathscr{A}$ with differentials $d_n = \sum_{i=0}^{n} (-1)^i \partial_i$. This is the bar resolution (or standard resolution) in a generalized sense, and there are many (co-)homology theories defined by means of such constructions.

Most of the above definitions and constructions can be dualized so as to give **injective classes** $\mathscr{I}$, **$\mathscr{I}$-injective resolutions**, and **right derived functors** with respect to $\mathscr{I}$, **triples** (or **monads**), etc. (See also [19] for generalized (co-)homology).

References

[1] Sém. H. Cartan, 1950–1951, Paris, 1951.
[2] H. P. Cartan and S. Eilenberg, Homological algebra, Princeton Univ. Press, 1956.
[3] A. Grothendieck, Sur quelques points d'algèbre homologique, Tôhoku Math. J., (2) 9 (1957), 119–221.
[4] R. Godement, Topologie algébrique et théorie des faisceaux, Actualités Sci. Ind., Hermann, 1958.
[5] D. G. Northcott, An introduction to homological algebra, Cambridge Univ. Press, 1960.
[6] S. MacLane, Homology, Springer, 1963.
[7] P. Freyd, Abelian categories, Harper, 1964.
[8] A. Grothendieck (and J. Dieudonné), Eléments de géometrie algébrique I–III, Publ. Math. Inst. HES, 1960–1963.
[9] J.-P. Serre, Cohomologie galoisienne, Lecture notes in math. 5, Springer, 1964.
[10] S. Lang, Rapport sur la cohomologie des groupes, Benjamin, 1966.
[11] E. Weiss, Cohomology of groups, Academic Press, 1969.
[12] H. Bass, Algebraic K-theory, Benjamin, 1968.
[13] D. Quillen, Projective modules over polynomial rings, Inventiones Math., 36 (1976), 167–171.
[14] S. Eilenberg and J. C. Moore, Foundation of relative homological algebra, Mem. Amer. Math. Soc., 55 (1965).
[15] A. Dold and D. Puppe, Homologie Nicht-Additiver Funktoren, Anwendungen, Ann. Inst. Fourier, 11 (1961), 201–312.
[16] M. Barr and J. Beck, Homology and standard constructions, Lecture notes in math. 80, Springer, 1969.
[17] M. Tierney and W. Vogel, Simplicial derived functors, Lecture notes in math. 86, Springer, 1969.
[18] S. MacLane, Categories for the working mathematician, Springer, 1971.
[19] J. F. Adams, Stable homotopy and generalised homology, Lecture notes, Univ. of Chicago, 1971.

201 (IX.6) Homology Theory

A. History

Homology theory is the oldest and most extensively developed portion of †algebraic topology. Historically, it started with measuring the higher-dimensional **connectivity** of a space in the sense that the 0-dimensional connectivity is the number of †connected components of the space. For example, take a †2-sphere S^2 and a †2-torus T^2. Then T^2 is distinguished from S^2 by the fact that on T^2 a closed curve can be drawn without forming a boundary, while this is not true for S^2. In fact, a curve (c_1 or c_1' in Fig. 1) can be drawn on T^2 so that it does not form a boundary of an embedded 2-disk. On more complicated †surfaces there are many kinds of such closed curves. The maximum number of such closed curves is the 1-dimensional connectivity of the surface; this is a topological property of the surface. For example, the 1-dimensional connectivity of S^2 is 0, that of T^2 is 2, and that of the surface in Fig. 2 is 6. A more general consideration of the bounding properties of q-dimensional †closed submanifolds of a manifold led E. Betti (*Ann. Mat. Pure Appl.*, 4 (1871)) to introduce the notion of the q-dimensional connectivity of the manifold, which was a precursor of homology theory.

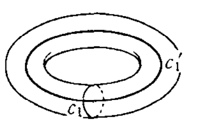


Fig. 1

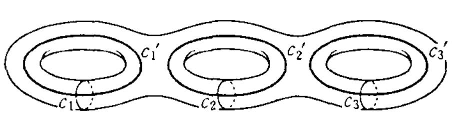


Fig. 2

The foundation of homology theory was laid by †H. Poincaré [1]. He started his study of homology with analytic treatment of manifolds, which led to a series of complications.

Poincaré then introduced a new method for the study, now called †combinatorial topology: He decomposed the manifold into elementary pieces or †cells, which adjoin one another in a regular fashion; he then substituted the algebraic notions of †cycles and †boundary operators for the geometrical notions of closed submanifolds and boundaries. Thereby the notion of homology groups acquired an exact logical meaning, and the fundamental formulas, which are now called Poincaré formulas, were proved.

After Poincaré, much of the development of homology theory centered around the question of the **topological invariance of homology groups**, that is, the independence of the homology groups on the choice of †cellular decomposition. Through the development of †simplicial complexes and their techniques, J. W. Alexander (*Trans. Amer. Math. Soc.*, 28 (1926)) gave the first fully satisfactory proof for topological invariance of the homology groups of †polyhedra. In those days, the homology groups themselves were barely recognized; instead, one dealt with numerical invariants such as the †Betti numbers and the †torsion coefficients [2, 3].

During the period 1925–1935 there was a gradual shift of interest from the numerical invariants to the homology groups themselves, and homology theory developed intensively [4, 5]. S. Lefschetz (*Trans. Amer. Math. Soc.*, 28 (1926)) added the theory of the †intersection products to the homology of manifolds. He also invented †relative homology theory (*Proc. Nat. Acad. Sci. US*, 13 (1927)) and generalized the †duality theorems of Poincaré and Alexander (*ibid.*, 15 (1929)). G. de Rham (*J. Math. Pures Appl.*, 10 (1931)) obtained a duality theorem that relates the †exterior differential forms in a manifold to the homology groups of the manifold. L. S. Pontryagin (*Ann. Math.*, (2) 35 (1934)) proved the complete group-invariant form of the Alexander duality theorem. These duality theorems seemed to reflect the existence of a theory dual to homology theory, and the genesis of this dual theory, now called †cohomology, occurred in 1935, in the work of Alexander and A. N. Kolmogorov. It was discovered subsequently by Alexander (*Ann. Math.*, (2) 37 (1936)), E. Čech (*ibid.*), and H. Whitney (*ibid.*, 39 (1938)) that the cohomology of a polyhedron can be made into a ring.

On the other hand, after L. Vietoris (*Math. Ann.*, 97 (1927)) and P. S. Aleksandrov (*Ann. Math.*, (2) 30 (1928)), many devices were invented to extend the homology theory of polyhedra to general †topological spaces, and numerous variants of homology theory appeared at the hands of Čech (1932), Lefschetz (1933), Alexander (1935), and Kolmogorov (1936). This development served to clarify the relations between combinatorial and set-theoretic methods in topology (→ 426 Topology), whereas it produced complexity and confusion in homology theory [6]. S. Eilenberg and N. E. Steenrod (*Proc. Nat. Acad. Sci. US*, 31 (1945) and [7]) cleared the air by treating homology axiomatically.

Roughly speaking, a homology theory assigns †Abelian groups to †topological spaces and †homomorphisms to †continuous mappings of one space to another. In this way, a homology theory is an algebraic image of topology; it converts topological problems to algebraic problems. Starting from this viewpoint, Eilenberg and Steenrod selected some fundamental properties as axioms to characterize homology theory. This unified homology theories and allowed systematic treatment of homological problems which had previously been done separately "by hand" in each case. Moreover, it motivated the birth of a new branch of mathematics, called †homological algebra.

B. Homology of Chain Complexes

A **chain complex** $C=\{C_q, \partial_q\}$ is a collection of (additive) Abelian groups C_q, one for each integer q, and of homomorphisms $\partial_q: C_q \to C_{q-1}$ such that $\partial_q \circ \partial_{q+1}=0$ for each q. Elements of C_q are called **q-chains** of C, and ∂_q is called the **boundary operator**. A **subcomplex** $C'=\{C'_q, \partial'_q\}$ of C is a chain complex such that $C'_q \subset C_q$ and $\partial'_q=\partial_q | C'_q$ for each q.

The †kernel of ∂_q is denoted by $Z_q(C)$, and its element is called a **q-cycle** of C. The †image of ∂_{q+1} is denoted by $B_q(C)$, and its element is called a **q-boundary** of C. The relation $\partial_q \circ \partial_{q+1}=0$ implies $B_q(C) \subset Z_q(C)$. The †quotient group $Z_q(C)/B_q(C)$ is denoted by $H_q(C)$, the **qth homology group** of C. Elements of $H_q(C)$ are called **q-dimensional homology classes** of C. Two cycles representing the same homology class are said to be **homologous**. The †direct sum $\sum_q H_q(C)$ is denoted by $H_*(C)$ and is called the **homology group** of C.

If $C=\{C_q, \partial_q\}$ and $C'=\{C'_q, \partial'_q\}$ are chain complexes, a **chain mapping (chain map)** $\varphi: C \to C'$ is a sequence of homomorphisms $\varphi_q: C_q \to C'_q$ such that $\partial'_q \circ \varphi_q=\varphi_{q-1} \circ \partial_q$ for each q. If $\varphi: C \to C'$ is a chain mapping, then φ_q sends $Z_q(C)$ to $Z_q(C')$ and $B_q(C)$ to $B_q(C')$, and hence φ induces a homomorphism of $H_q(C)$ to $H_q(C')$.

If $H_q(C)$ is †finitely generated, it can be decomposed into the direct sum of a free Abelian group $B_q(C)$ and a finite Abelian group $T_q(C)$. $B_q(C)$ and $T_q(C)$ are called the qth **Betti group**

of C and the qth **torsion group** of C, respectively. The †rank ρ_q of $B_q(C)$ is called the qth **Betti number** of C. $T_q(C)$ is isomorphic to the direct sum of $\tau(q)$ finite cyclic groups of orders $\theta_1^q, \theta_2^q, \ldots, \theta_{\tau(q)}^q$, where $\theta_1^q > 1$ and θ_i^q divides θ_{i+1}^q for $i = 1, \ldots, \tau(q)-1$ (→ 2 Abelian Groups). The numbers $\theta_1^q, \theta_2^q, \ldots, \theta_{\tau(q)}^q$ are called the qth **torsion coefficients** of C.

If $H_*(C)$ is finitely generated, then the number $\chi(C) = \sum_q (-1)^q \rho_q$ is called the **Euler number**, the **Euler characteristic**, or the **Euler-Poincaré characteristic** of C. In this case a †polynomial $\sum_q \rho_q t^q$ with variable t is called the **Poincaré polynomial** of C.

Let C be a chain complex such that, for each q, C_q is a †free Abelian group of finite †rank. Then the qth Betti number and the qth torsion coefficients are well defined for each q. Denote the ranks of C_q and $B_q(C)$ by α_q and β_q, respectively. Then it holds that $\rho_q = \alpha_q - \beta_q - \beta_{q-1}$, and hence $\chi(C) = \sum_q (-1)^q \alpha_q$. (**Euler-Poincaré formula**). Moreover, there exists a set of bases, one for each C_q, with the following properties: For each q, the base for C_q is composed of five types of elements, a_i^q $(1 \leqslant i \leqslant \beta_q - \tau_q)$, b_i^q $(1 \leqslant i \leqslant \tau_q)$, c_i^q $(1 \leqslant i \leqslant \rho_q)$, d_i^q $(1 \leqslant i \leqslant \tau_{q-1})$, and e_i^q $(1 \leqslant i \leqslant \beta_{q-1} - \tau_{q-1})$; ∂_q satisfies $\partial a_i^q = 0$, $\partial b_i^q = 0$, $\partial c_i^q = 0$, $\partial d_i^q = \theta_i^{q-1} b_i^{q-1}$, and $\partial e_i^q = a_i^{q-1}$. Such a set of bases is called the **canonical basis** of C.

Let C be a chain complex such that each C_q is a free Abelian group with a given base $\{\sigma_i^q\}$. Then the **incidence number** $[\sigma_i^q : \sigma_j^{q-1}] \in \mathbf{Z}$ is defined by $\partial_q(\sigma_i^q) = \sum_j [\sigma_i^q : \sigma_j^{q-1}] \sigma_j^{q-1}$. This notion was commonly used in the early days of topology.

C. Homology of Simplicial Complexes

Let K be a †simplicial complex. An **oriented q-simplex** σ of K is a q-simplex $s \in K$ together with an equivalence class of †total orderings of the vertices of s, two orderings being equivalent if they differ by an even permutation of the vertices. If $a_0, \ldots, a_q$ are the vertices of s, then $[a_0, a_1, \ldots, a_q]$ denotes the oriented q-simplex of K consisting of the simplex s together with the equivalence class of the ordering $a_0 < a_1 < \ldots < a_q$ of its vertices. For every vertex a of K there is a unique oriented 0-simplex $[a]$, and to every q-simplex with $q \geqslant 1$ there correspond exactly two oriented q-simplexes, which are said to be **opposites** of one another.

Let $C_q(K)$ denote the Abelian group generated by the oriented q-simplexes of K with the relations $\sigma + \sigma' = 0$ if σ' is the opposite of σ. If we choose an oriented q-simplex σ_i^q for each q-simplex s_i^q of K, then each element of $C_q(K)$ is written uniquely as a finite sum $\sum_i g_i \sigma_i^q$ with integers $g_i \neq 0$, and $C_q(K)$ is a free Abelian group generated by the set $\{\sigma_i^q\}$. We define a homomorphism $\partial_q : C_q(K) \to C_{q-1}(K)$ by $\partial_q[a_0, a_1, \ldots, a_q] = \sum_{i=0}^q (-1)^i [a_0, \ldots, a_{i-1}, a_{i+1}, \ldots, a_q]$. Then $\partial_q \circ \partial_{q+1} = 0$ holds, and we have a chain complex $C(K) = \{C_q(K), \partial_q\}$ $(C_q(K) = \{0\}$ if $q < 0)$, called the **(oriented) simplicial chain complex**. The homology group $H_*(C(K))$ is denoted by $H_*(K)$ and is called the **(integral) homology group of the simplicial complex** K.

Let K_1 and K_2 be simplicial complexes, and let $f: K_1 \to K_2$ be a †simplicial mapping. Then for each q, a homomorphism $f_{\#q} : C_q(K_1) \to C_q(K_2)$ can be defined by $f_{\#q}([a_0, a_1, \ldots, a_q]) = [f(a_0), f(a_1), \ldots, f(a_q)]$, where the right-hand side is understood to be 0 if $f(a_0), f(a_1), \ldots, f(a_q)$ are not distinct. The sequence $f_\# = \{f_{\#q}\}$ is a chain mapping of $C(K_1)$ to $C(K_2)$, and it induces a homomorphism $f_* : H_*(K_1) \to H_*(K_2)$.

If K is a †finite simplicial complex, then the Betti numbers, the torsion coefficients, and the Euler characteristic of K are defined to be those for the chain complex $C(K)$.

Let K_1 and K_2 be a subcomplexes of a simplicial complex K. Then we have the following †exact sequence which relates the homology group of $K_1 \cup K_2$ to the homology groups of K_1, K_2, and $K_1 \cap K_2$: $\ldots \xrightarrow{\partial_*} H_q(K_1 \cap K_2) \xrightarrow{\alpha} H_q(K_1) + H_q(K_2) \xrightarrow{\beta} H_q(K_1 \cup K_2) \xrightarrow{\partial_*} H_{q-1}(K_1 \cap K_2) \xrightarrow{\alpha} \ldots$, where α, β, and ∂_* are defined as follows. Let $i_l : K_1 \cap K_2 \to K_l$ and $j_l : K_l \to K_1 \cup K_2$ $(l = 1, 2)$ be inclusion mappings; then $\alpha(a) = (i_{1*}(a), -i_{2*}(a))$ and $\beta(a_1, a_2) = j_{1*}(a_1) + j_{2*}(a_2)$. If $z = c_1 + c_2$ $(c_l \in C(K_l))$ is a cycle of $C(K_1 \cup K_2)$ then $\partial(c_1) = -\partial(c_2)$ is a cycle of $C(K_1 \cap K_2)$; ∂_* sends the homology class of z to the homology class of $\partial(c_1)$. The sequence is called the **Mayer-Vietoris exact sequence** of the couple $\{K_1, K_2\}$, and ∂_* is referred to as the **connecting homomorphism**. The prototype of the Mayer-Vietoris exact sequence was obtained by W. Mayer (*Monatsh. Math. Phys.*, 36 (1929)) and L. Vietoris (*ibid.*, 37 (1930)). The present form is due to Eilenberg and Steenrod [7].

D. Homology of Polyhedra

If K is a †simplicial complex and K' is a †subdivision of K, then there exists a canonical isomorphism $H_*(K) \cong H_*(K')$. This proves that if K_1 and K_2 are †simplicial decompositions of a †polyhedron then $H_*(K_1)$ and $H_*(K_2)$ are isomorphic, because there exists a common subdivision of K_1 and K_2. Thus we may define the **(integral) homology groups $H_*(X)$ of a polyhedron** X to be the homology group $H_*(K)$ of a simplicial decomposition K of X.

Let X and Y be polyhedra, and let $f: X \to Y$ be a continuous mapping. Take a †simplicial approximation $\varphi: K \to L$ of f. Then a homo-

morphism of $H_*(X)$ to $H_*(Y)$ given by the induced homomorphism $\varphi_*: H_*(K) \to H_*(L)$ is independent of the choice of φ, and is denoted by f_*. The following properties hold: (i) $1_* = 1: H_*(X) \to H_*(X)$, where 1 is the identity; (ii) If $f: X \to Y$ and $g: Y \to Z$ are continuous mappings, then $(g \circ f)_* = g_* \circ f_*: H_*(X) \to H_*(Z)$; (iii) If $f, f': X \to Y$ are homotopic, then $f_* = f'_*: H_*(X) \to H_*(Y)$. These imply the **homotopy invariance of the homology group** stated as follows: If X and Y are polyhedra which are †homotopy equivalent, then $H_*(X)$ and $H_*(Y)$ are isomorphic. Specifically, the homology group is a topological invariant. Thus if X is a †triangulable space (for example, if X is a †differentiable manifold), then its homology group $H_*(X)$ can be defined to be the homology group $H_*(K)$, where (K, t) is a †triangulation of X. This homology group is referred to as the **simplicial homology group of** X. Similarly, if X is a compact triangulable space, the Betti numbers $\rho_q(X)$ of X, etc. can be defined to be those for the chain complex $C(K)$.

If we denote by pt a single point, then $H_0(pt) = \mathbf{Z}$ (the group of integers) and $H_q(pt) = 0$ if $q \neq 0$. For an n-sphere S^n, a 2-torus T^2, and a †real projective plane P^2, the homology groups can be computed as follows by means of their triangulations: (1) $H_0(S^n) \cong H_n(S^n) \cong \mathbf{Z}$, and $H_q(S^n) = 0$ if $q \neq 0, n$; (2) $H_0(T^2) \cong H_2(T^2) \cong \mathbf{Z}$, $H_1(T^2) \cong \mathbf{Z} + \mathbf{Z}$, and $H_q(T^2) = 0$ if $q \neq 0, 1, 2$; (3) $H_0(P^2) \cong \mathbf{Z}$, $H_1(P^2) \cong \mathbf{Z}_2$, and $H_q(P^2) = 0$ if $q \neq 0, 1$, where $\mathbf{Z}_2 = \mathbf{Z}/2\mathbf{Z}$.

Two surfaces are homeomorphic if and only if their integral homology groups are isomorphic (→ 410 Surfaces).

E. Singular Homology

There are various devices for defining homology groups of general topological spaces. A familiar one is the **singular homology theory** initiated by S. Lefschetz (*Bull. Amer. Math. Soc.*, 39 (1933)) and improved by S. Eilenberg (*Ann. Math.*, (2) 45 (1944)).

The **standard** q**-simplex** is the convex set $\Delta^q \subset \mathbf{R}^{q+1}$ consisting of all $(q+1)$-tuples $(t_0, t_1, \ldots, t_q)$ of real numbers with $t_i \geqslant 0$, $t_0 + t_1 + \ldots + t_q = 1$. Any continuous mapping of Δ^q to a topological space X is called a **singular** q**-simplex** in X. The i**th face** of a singular q-simplex $\sigma: \Delta^q \to X$ is the singular $(q-1)$-simplex $\sigma \circ \varepsilon_i: \Delta^{q-1} \to X$, where the linear embedding $\varepsilon_i: \Delta^{q-1} \to \Delta^q$ is defined by $\varepsilon_i(t_0, \ldots, t_{i-1}, t_{i+1}, \ldots, t_q) = (t_0, \ldots, t_{i-1}, 0, t_{i+1}, \ldots, t_q)$.

For each integer q, let $S_q(X)$ denote the free Abelian group generated by the singular q-simplexes in X ($S_q(X) = 0$ if $q < 0$), and define a homomorphism $\partial_q: S_q(X) \to S_{q-1}(X)$ by $\partial_q(\sigma) = \sum_{i=0}^{q}(-1)^i \sigma \circ \varepsilon_i$. Then we have a chain complex $S(X) = \{S_q(X), \partial_q\}$, called the **singular chain complex** of X. The homology group $H_*(S(X))$ is denoted by $H_*(X)$ and is called the **integral singular homology group** of the topological space X.

Given a continuous mapping $f: X \to Y$, a chain mapping $f_\#: S(X) \to S(Y)$ is defined by sending each singular simplex $\sigma: \Delta^q \to X$ to the singular simplex $f \circ \sigma: \Delta^q \to Y$, and it induces the homomorphism $f_*: H_*(X) \to H_*(Y)$. The properties (i), (ii), (iii) of f_* in Section D hold for continuous mappings of topological spaces, and hence the singular homology group is a homotopy invariant.

The homology group $H_*(K)$ of a simplicial complex K is isomorphic to the singular homology group $H_*(|K|)$ of the polyhedron $|K|$. Therefore the simplicial homology group of a triangulable space is isomorphic to the singular homology group of the space.

If $\{X_i\}$ is the set of †arcwise connected components of a topological space X, then $H_*(X) \cong \sum_i H_*(X_i)$. If X is arcwise connected, then $H_0(X) \cong \mathbf{Z}$. If $\{A_\lambda\}$ is the collection of all the compact subsets of X directed by inclusion, then $H_*(X)$ is isomorphic to the †inductive limit $\varinjlim H_*(A_\lambda)$. It is not true that there is a Mayer-Vietoris sequence in singular homology for any couple $\{X_1, X_2\}$ of subsets of X. However, for certain couples $\{X_1, X_2\}$, there is a Mayer-Vietoris exact sequence of $\{X_1, X_2\}$: $\ldots \xrightarrow{\partial_*} H_q(X_1 \cap X_2) \xrightarrow{\alpha} H_q(X_1) + H_q(X_2) \xrightarrow{\beta} H_q(X_1 \cup X_2) \xrightarrow{\partial_*} H_{q-1}(X_1 \cap X_2) \xrightarrow{\alpha} \ldots$. For example, this holds if $X = \operatorname{Int} X_1 \cup \operatorname{Int} X_2$, where Int denotes the †interior.

Let $c: X \to pt$ be the mapping of a topological space to a single point. Then the kernel of $c_*: H_*(X) \to H_*(pt)$ is denoted by $\tilde{H}_*(X)$ and is called the **reduced homology group** of X. It holds that $H_*(X) \cong \tilde{H}_*(X) + H_*(pt)$. Regard the †suspension SX as the union of two copies of the †cone over X. Then the connecting homomorphism in the Meyer-Vietoris sequence gives an isomorphism $\tilde{H}_q(SX) \cong \tilde{H}_{q-1}(X)$ for any q. The inverse of this isomorphism is called the **suspension isomorphism for homology**.

Let M be a †C^∞-manifold. A C^∞-singular q-simplex in M is a singular q-simplex $\sigma: \Delta^q \to M$ such that σ extends to a C^∞-mapping from an open neighborhood of Δ^q in $\{(t_0, t_1, \ldots, t_q) \in \mathbf{R}^{q+1} \mid t_0 + t_1 + \ldots + t_q = 1\}$ to M. The totality of C^∞-singular simplexes in M generates a subcomplex of $S(M)$, denoted by $S^\infty(M)$. The inclusion $S^\infty(M) \subset S(M)$ induces an isomorphism $H_*(S^\infty(M)) \cong H_*(M)$.

Let M be an n-dimensional †topological manifold. Then $H_q(M) = 0$ unless $0 \leqslant q \leqslant n$, and $H_*(M)$ is finitely generated if M is compact.

F. Homology of CW Complexes

Homology theory is tractable in the category of †CW complexes by virtue of the facts stated below.

Let X be a topological space and A its subset. Then we denote by X/A the quotient space obtained from X by shrinking A to a point, understanding $X/\varnothing$ to be the disjoint union $X \cup pt$, If X_1, X_2 are †subcomplexes of a CW complex, then the Mayer-Vietoris exact sequence of $\{X_1, X_2\}$ and the **excision isomorphism** $i_*: H_*(X_1/(X_1 \cap X_2)) \cong H_*((X_1 \cup X_2)/X_2)$ (i: inclusion) are valid. If A is a subcomplex of a CW complex X, then we have the following **reduced homology exact sequence** of (X, A): $\ldots \xrightarrow{\partial_*} \tilde{H}_q(A) \xrightarrow{i_*} \tilde{H}_q(X) \xrightarrow{j_*} \tilde{H}_q(X/A) \xrightarrow{\partial_*} \tilde{H}_{q-1}(A) \xrightarrow{i_*} \ldots$, where i_* and j_* are induced by the inclusion $i: A \to X$ and the collapsing $j: X \to X/A$, and ∂_* is given by a commutative diagram

$$\begin{array}{ccc} \tilde{H}_q(X/A) & \xrightarrow{\partial_*} & \tilde{H}_{q-1}(A) \\ \cong \uparrow h^* & & \cong \downarrow S \\ \tilde{H}_q(X \cup CA) & \xrightarrow{h'_*} & \tilde{H}_q(SA). \end{array}$$

Here CA is the cone over A, S is the suspension isomorphism, and $h: X \cup CA \to (X \cup CA)/CA = X/A$ and $h': X \cup CA \to (X \cup CA)/X = SA$ are collapsings. More generally, if A, B are subcomplexes of a CW complex X and $A \supset B$, then we have the following **reduced homology exact sequence** of (X, A, B): $\ldots \xrightarrow{\partial_*} \tilde{H}_q(A/B) \xrightarrow{i_*} \tilde{H}_q(X/B) \xrightarrow{j_*} \tilde{H}_q(X/A) \xrightarrow{\partial_*} \tilde{H}_{q-1}(A/B) \xrightarrow{i_*} \ldots$. Furthermore, the homology group $H_*(X)$ of a CW complex X can be computed in the following manner.

Let X^q denote the q-skeleton of X, i.e., the union of all cells of dimensions $\leqslant q$. Put $C_q(X) = \tilde{H}_q(X^q/X^{q-1})$, and let $\partial_q: C_q(X) \to C_{q-1}(X)$ be the connecting homomorphism $\partial_*: \tilde{H}_q(X^q/X^{q-1}) \to \tilde{H}_{q-1}(X^{q-1}/X^{q-2})$ in the reduced homology exact sequence of (X^q, X^{q-1}, X^{q-2}). Then $C(X) = \{C_q(X), \partial_q\}$ is a chain complex such that $C_q(X)$ is a free Abelian group with one generator for each q-cell of X. If X is a polyhedron $|K|$, then $C(X)$ coincides with the simplicial chain complex $C(K)$. The homology group $H_*(C(K))$ is called the **cellular homology group** of the CW complex X. This is isomorphic to the singular homology group $H_*(X)$.

Since a CW decomposition frequently requires fewer cells than a simplicial decomposition, the cellular homology groups are useful in calculating the homology groups. For example, the †complex n-dimensional projective space $\mathbf{C}P^n$ has a CW decomposition with a single $2i$-cell for each $i = 0, 1, \ldots, n$, and hence we see immediately that $H_q(\mathbf{C}P^n) \cong \mathbf{Z}$ if $q = 2i$ $(0 \leqslant i \leqslant n)$ and $= 0$ otherwise.

If X is a finite CW complex and α_q denotes the number of q-cells of X, then we have the **Euler-Poincaré formula** $\chi(X) = \sum_q (-1)^q \alpha_q$. In particular, if X is homeomorphic to S^2 then we have the **Euler theorem on polyhedra**: $\alpha_0 - \alpha_1 + \alpha_2 = 2$. This was the first important result in topology (L. Euler, 1752).

G. Homology with Coefficients in Abelian Groups

Given a chain complex C and an Abelian group G, we have a new chain complex $C \otimes G$ given by $(C \otimes G)_q = C_q \otimes G$ and $\partial_q(c \otimes g) = \partial_q c \otimes g$ $(c \in C_q, g \in G)$, where $\otimes$ is the †tensor product of Abelian groups. For a topological space X, the homology group of the chain complex $S(X) \otimes G$ is denoted by $H_*(X; G)$ and is called the **singular homology group of** X **with coefficients in** G. The **homology group** $H_*(K; G)$ **of a simplicial complex** K **with coefficients in** G and the **cellular homology group** $H_*(C(X); G)$ **of a CW complex** X **with coefficients in** G are similarly defined. The homology group with coefficients in $\mathbf{Z}$ is the integral homology group.

The previous results for the integral homology groups generalize in a straightforward fashion to homology groups with coefficients in G.

We have the homomorphism $\kappa: H_q(X) \otimes G \to H_q(X; G)$ sending $a \otimes g \in H_q(X) \otimes G$ to the homology class of $z \otimes g \in Z_q(S(X) \otimes G)$, where z is a representative cycle of a. The following theorem is known as the **universal coefficient theorem for homology**, since it expresses $H_q(X; G)$ in terms of $H_q(X)$, $H_{q-1}(X)$, and G: There is an exact sequence $0 \to H_q(X) \otimes G \xrightarrow{\kappa} H_q(X; G) \to \operatorname{Tor}(H_{q-1}(X), G) \to 0$, and this sequence is split (→ 200 Homological Algebra). Universal coefficient theorems of this type were first shown by S. Eilenberg and S. MacLane (*Ann. Math.*, (2) 43 (1942)).

Let Λ be a †ring with a unit 1. Then a **chain complex over** Λ is a chain complex C such that each C_q is a †Λ-module and each ∂_q is a †Λ-homomorphism. The homology groups $H_q(C)$ of a chain complex C over Λ are Λ-modules. If C is a chain complex, $C \otimes \Lambda$ forms naturally a chain complex over Λ. In particular, if X is a topological space, then $S(X) \otimes \Lambda$ is a chain complex over Λ, and $H_q(X; \Lambda)$ are Λ-modules. In this case, the induced homomorphisms $f_*: H_q(X; \Lambda) \to H_q(Y; \Lambda)$ are Λ-homomorphisms. The homology groups with coefficients in a field $\mathbf{k}$ are †vector spaces over $\mathbf{k}$ and are useful in applications. If $H_*(X)$ is finitely generated, then $\chi(X) = \sum_q (-1)^q \dim_{\mathbf{k}} H_q(X; k)$ holds for any field $\mathbf{k}$.

H. Cohomology

A **cochain complex** $C=\{C^q,\delta^q\}$ is a collection of Abelian groups C^q, one for each integer q, and of homomorphisms $\delta^q: C^q \to C^{q+1}$ such that $\delta^{q+1}\circ\delta^q=0$. Elements of C^q are called q-**cochains**, and δ^q is called the **coboundary operator**. The notions of **subcomplex of a cochain complex**, **cocycle**, **coboundary**, **cohomology group**, and **cochain mapping** are defined as in chain complex. Let $\mathrm{Hom}(A,B)$ denote the †group of homomorphisms from an Abelian group A to an Abelian group B. Given a chain complex C and an Abelian group G, a cochain complex $C^*=\mathrm{Hom}(C,G)$ is defined by $C^q=\mathrm{Hom}(C_q,G)$ and $(\delta^q u)(c)=u(\partial_{q+1}c)$ $(u\in C^q, c\in C_{q+1})$.

For a topological space X, the cochain complex $\mathrm{Hom}(S(X),G)$ is called the **singular cochain complex** of X with coefficients in G, and elements of $\mathrm{Hom}(S_q(X),G)$ are called **singular** q-**cochains** of X. The cohomology group of $\mathrm{Hom}(S(X),G)$ is denoted by $H^*(X;G)$ and is called the **singular cohomology group of** X **with coefficients in** G. We write $H^*(X)$ for $H^*(X;\mathbf{Z})$; this is called the **integral cohomology group** of X. Similarly, the **cohomology group** $H^*(K;G)$ **of a simplicial complex** K **with coefficients in** G and the **cellular cohomology group** $H^*(C(X);G)$ **of a CW complex** X **with coefficients** in G are defined. There are isomorphisms $H^*(K;G)\cong H^*(|K|;G)$ and $H^*(C(X);G)\cong H^*(X;G)$.

If $f: X\to Y$ is a continuous mapping, then a cochain map $f^\#: \mathrm{Hom}(S(Y),G)\to\mathrm{Hom}(S(X),G)$ is defined by $(f^\# u)(c)=u(f_\# c)$ with $u\in \mathrm{Hom}(S_q(Y),G)$ and $c\in S_q(X)$. Therefore f induces the homomorphism $f^*: H^*(Y;G)\to H^*(X;G)$. The following properties hold: (i) $1^*=1$; (ii) $(g\circ f)^*=f^*\circ g^*$; (iii) If f and f' are homotopic, then $f^*=f'^*$. In particular, the singular cohomology groups are homotopy invariants. The †cokernel of $c^*: H^*(pt;G)\to H^*(X;G)$ induced by the mapping $c: X\to pt$ is denoted by $\tilde{H}^*(X;G)$ and is called the **reduced cohomology group** of X.

For $\xi\in H^q(X;G)$ and $a\in H_q(X)$, the **Kronecker index** $\langle\xi,a\rangle\in G$ is defined naturally in terms of representatives of ξ and a. We have the following **universal coefficient theorem for cohomology**: There is an exact sequence $0\to \mathrm{Ext}(H_{q-1}(X),G)\to H^q(X;G)\xrightarrow{\kappa}\mathrm{Hom}(H_q(X),G)\to 0$, and this sequence is split, where κ is given by the Kronecker index (→ 200 Homological Algebra).

If Λ is a ring with 1, then a **cochain complex over** Λ is defined analogously to a chain complex over Λ. The singular cochain complex $\mathrm{Hom}(S(X),\Lambda)$ forms naturally a cochain complex over Λ, and $H^q(X;\Lambda)$ are Λ-modules. For $\xi\in H^q(X;\Lambda)$ and $a\in H_q(X;\Lambda)$, the **Kronecker index** $\langle\xi,a\rangle\in\Lambda$ is defined naturally. If $\mathbf{k}$ is a field, then the vector spaces $H^q(X;\mathbf{k})$ are identified with the dual space of $H_q(X,\mathbf{k})$ by means of the Kronecker index.

If M is a C^∞-manifold, then there is a cochain complex $\mathfrak{D}(M)=\{\mathfrak{D}^q(M),d\}$ over the field $\mathbf{R}$ of real numbers, where $\mathfrak{D}^q(M)$ is the real vector space consisting of the †differential forms of degree q on M, and $d:\mathfrak{D}^q(M)\to\mathfrak{D}^{q+1}(M)$ is the †exterior differentiation. The cochain complex $\mathfrak{D}(M)$ is called the **de Rham complex** of M, and its cohomology group $H^*(\mathfrak{D}(M))$ is called the **de Rham cohomology group** of M. A cochain mapping $\mathscr{I}:\mathfrak{D}(M)\to\mathrm{Hom}(S^\infty(M),\mathbf{R})$ is defined by

$$(\mathscr{I}(\omega))(\sigma)=\int_{\Delta^q}\sigma^*\omega,$$

where $\omega\in D^q(M)$, $\sigma:\Delta^q\to M$ is a C^∞ singular q-simplex in M, and $\sigma^*\omega$ denotes the †pullback of ω by σ. We have isomorphisms

$$H^*(\mathfrak{D}(M))\xrightarrow[\cong]{\mathscr{I}_*}H^*(\mathrm{Hom}(S^\infty(M),\mathbf{R}))\xleftarrow[\cong]{i^*}H^*(M;\mathbf{R}),$$

where i^* is induced by the inclusion $S^\infty(M)\subset S(M)$. This result is called the **de Rham theorem** on the cohomology of manifolds (→ 105 Differentiable Manifolds).

I. Cohomology Rings

Given a topological space X and a ring Λ, the **cup product** $u\smile v\in\mathrm{Hom}(S_{p+q}(X),\Lambda)$ of cochains $u\in\mathrm{Hom}(S_p(X),\Lambda)$ and $v\in\mathrm{Hom}(S_q(X),\Lambda)$ is defined by $(u\smile v)(\sigma)=u(\sigma\circ\varepsilon)v(\sigma\circ\varepsilon')$, where $\sigma:\Delta^{p+q}\to X$ is a singular $(p+q)$-simplex, $\varepsilon:\Delta^p\to\Delta^{p+q}$ and $\varepsilon':\Delta^q\to\Delta^{p+q}$ are given by $\varepsilon(t_0,t_1,\dots,t_p)=(t_0,t_1,\dots,t_p,0,\dots,0)$, and $\varepsilon'(t_p,t_{p+1},\dots,t_{p+q})=(0,\dots,0,t_p,t_{p+1},\dots,t_{p+q})$. The product operation is bilinear, and the formula $\delta(u\smile v)=\delta u\smile v+(-1)^p u\smile\delta v$ holds. Therefore it gives rise to the **cup product** $\xi\smile\eta\in H^{p+q}(X;\Lambda)$ of cohomology classes $\xi\in H^p(X;\Lambda)$ and $\eta\in H^q(X;\Lambda)$. This cup product operation makes $H^*(X;\Lambda)$ into a ring, which is called the **singular cohomology ring** of X with coefficients in Λ. If Λ has 1, the cohomology class represented by the 0-cocycle taking the value 1 on each singular 0-simplex serves as 1 of $H^*(X;\Lambda)$. If Λ is commutative, then $\xi\eta=(-1)^{pq}\eta\xi$ holds. The induced homomorphism $f^*: H^*(Y;\Lambda)\to H^*(X;\Lambda)$ preserves the product, and hence the cohomology ring is a homotopy invariant.

If K is a simplicial complex, the **cup product** operation $\smile: H^p(K;\Lambda)\otimes H^q(K;\Lambda)\to H^{p+q}(K;\Lambda)$ is induced from the operation $\smile: \mathrm{Hom}(C_p(K),\Lambda)\otimes\mathrm{Hom}(C_q(K),\Lambda)\to\mathrm{Hom}(C_{p+q}(K),\Lambda)$ defined as follows. Adopting a †linear ordering of vertices of K, we write all oriented simplexes in this ordering. Then,

for $u \in \mathrm{Hom}(C_p(K), \Lambda)$ and $v \in \mathrm{Hom}(C_q(K), \Lambda)$, we define $u \smile v \in \mathrm{Hom}(C_{p+q}(K), \Lambda)$ by $(u \smile v)([a_0, a_1, \ldots, a_{p+q}]) = u([a_0, a_1, \ldots, a_p])v([a_p, a_{p+1}, \ldots, a_{p+q}])$. The canonical isomorphism from the cohomology of K to the singular cohomology of $|K|$ preserves the cup product.

On the de Rham complex $\mathfrak{D}(M)$ of a C^∞-manifold M, we have the †exterior product $\omega \wedge \theta \in \mathfrak{D}^{p+q}(M)$ of $\omega \in \mathfrak{D}^p(M)$ and $\theta \in \mathfrak{D}^q(M)$. This makes $H^*(\mathfrak{D}(M))$ into a ring, which is called the **de Rham cohomology ring**. The canonical isomorphism of $H^*(\mathfrak{D}(M))$ to $H^*(M; R)$ preserves the product (→ 105 Differentiable Manifolds).

Examples. (1) Let $T^n = S^1 \times \ldots \times S^1$ denote the n-dimensional torus, and let $\pi_i: T^n \to S^1$ denote the projection to the ith factor ($1 \leqslant i \leqslant n$). Take a generator ξ of the Λ-module $H^1(S^1; \Lambda)$, and put $\xi_i = \pi_i^*(\xi) \in H^1(T^n; \Lambda)$. Then $H^*(T^n; \Lambda)$ is the †exterior algebra over Λ generated by $\xi_1, \ldots, \xi_n$. (2) If we denote by $\mathbf{C}P^n$ the complex n-dimensional projective space, then $H^q(\mathbf{C}P^n; \Lambda)$ is Λ if $q = 2i$ ($0 \leqslant i \leqslant n$) and 0 otherwise. If ξ is a generator of the Λ-module $H^2(\mathbf{C}P^n; \Lambda)$, then ξ^i generates the Λ-module $H^{2i}(\mathbf{C}P^n; \Lambda)$ ($0 \leqslant i \leqslant n$). Thus $H^*(\mathbf{C}P^n; \Lambda)$ is the quotient ring $\Lambda[\xi]/(\xi^{n+1})$ of the †polynomial ring $\Lambda[\xi]$ by the †ideal (ξ^{n+1}). (3) If P^n denote the real n-dimensional projective space, then $H^*(P^n; \mathbf{Z}_2) \cong \mathbf{Z}_2[\xi]/(\xi^{n+1})$, where ξ is the generator of $H^1(P^n; \mathbf{Z}_2)$.

J. Homology of Product Spaces

If C and D are chain complexes, their **tensor product** $C \otimes D$ is a chain complex given by $(C \otimes D)_n = \sum_{p+q=n} C_p \otimes D_q$, and $\partial_n(c \otimes d) = \partial_p(c) \otimes d + (-1)^p c \otimes \partial_q(d)$ ($c \in C_p, d \in D_q$). The following **Eilenberg-Zilber theorem** (*Amer. J. Math.*, 75 (1953)) is the link between the algebra of tensor products and the geometry of product spaces: For the product space $X \times Y$ of topological spaces X and Y, there is an isomorphism $\rho_*: H_*(X \times Y; G) \cong H_*(S(X) \otimes S(Y) \otimes G)$ induced from a chain mapping ρ: $S(X \times Y) \to S(X) \otimes S(Y)$ defined as follows: Given a singular n-simplex $\sigma: \Delta^n \to X \times Y$, we define for each p ($0 \leqslant p \leqslant n$) a singular p-simplex σ'_p in X to be the composite $\Delta^p \xrightarrow{\varepsilon} \Delta^n \xrightarrow{\sigma} X \times Y \xrightarrow{\pi_1} X$, where $\varepsilon(t_0, t_1, \ldots, t_p) = (t_0, t_1, \ldots, t_p, 0, \ldots, 0)$ and π_1 is the projection to the first factor. Similarly, we define for each q ($0 \leqslant q \leqslant n$) a singular q-simplex σ''_q in Y to be the composite $\Delta^q \xrightarrow{\varepsilon'} \Delta^n \xrightarrow{\sigma} X \times Y \xrightarrow{\pi_2} Y$, where $\varepsilon'(t_{n-q}, \ldots, t_n) = (0, \ldots, 0, t_{n-q}, \ldots, t_n)$ and π_2 is the projection to the second factor. Then ρ is defined by $\rho(\sigma) = \sum_{p+q=n} \sigma'_p \otimes \sigma''_q$ and is called the **Alexander-Whitney mapping (Alexander-Whitney map)**.

Given a ring Λ, a chain mapping $\mu: (S(X) \otimes \Lambda) \otimes (S(Y) \otimes \Lambda) \to S(X) \otimes S(Y) \otimes \Lambda$ is defined by $\mu((c \otimes \lambda') \otimes (d \otimes \lambda'')) = c \otimes d \otimes \lambda'\lambda''$; it induces homomorphisms $\mu_*: H_p(X; \Lambda) \otimes H_q(Y; \Lambda) \to H_{p+q}(S(X) \otimes S(Y) \otimes \Lambda)$. The **cross product** $a \times b \in H_{p+q}(X \times Y; \Lambda)$ of $a \in H_p(X; \Lambda)$ and $b \in H_q(Y; \Lambda)$ is defined to be $\rho_*^{-1}(\mu_*(a \times b))$. If Λ is a commutative ring with 1, then the cross product defines a Λ-homomorphism $\times$: $H_p(X; \Lambda) \otimes_\Lambda H_q(Y; \Lambda) \to H_{p+q}(X \times Y; \Lambda)$, and satisfies $(a \times b) \times c = a \times (b \times c)$, $T_*(a \times b) = (-1)^{pq} b \times a$, $(f \times g)_*(a \times b) = f_*(a) \times g_*(b)$, where $T: X \times Y \to Y \times X$ is the mapping interchanging factors, and $f: X \to X'$, $g: Y \to Y'$ are continuous mappings. If Λ is a †principal ideal domain, there is an exact sequence

$$0 \to \sum_{p+q=n} H_p(X; \Lambda) \otimes_\Lambda H_q(Y; \Lambda) \xrightarrow{\times} H_n(X \times Y; \Lambda)$$
$$\to \sum_{p+q=n-1} \mathrm{Tor}_\Lambda(H_p(X; \Lambda), H_q(Y; \Lambda)) \to 0,$$

and this sequence is split (→ 200 Homological Algebra). In particular, if $\mathbf{k}$ is a field we have the following isomorphism of vector spaces:

$$\times: \sum_{p+q=n} H_p(X; \mathbf{k}) \underset{\mathbf{k}}{\otimes} H_q(Y; \mathbf{k}) \cong H_n(X \times Y; \mathbf{k}).$$

This is called the **Künneth theorem**, since the prototype was proved by D. Künneth (*Math. Ann.*, 90 (1923); *ibid.*, 91 (1924)). The present form was given by H. Cartan and S. Eilenberg [8].

The **tensor product** $C \otimes D$ of cochain complexes C and D is defined analogously to that of chain complexes. Given topological spaces X, Y and a ring Λ, a cochain mapping μ: $\mathrm{Hom}(S(X), \Lambda) \otimes \mathrm{Hom}(S(Y), \Lambda) \to \mathrm{Hom}(S(X) \otimes S(Y), \Lambda)$ is defined by $(\mu(u \otimes v))(c \otimes d) = u(c)v(d)$, where $u \in \mathrm{Hom}(S_p(X), \Lambda)$, $v \in \mathrm{Hom}(S_q(Y), \Lambda)$, $c \in S_s(X)$, $d \in S_t(Y)$, and $u(c)v(d)$ is understood to be 0 if $(p, q) \neq (s, t)$. We then have the composite $H^p(X; \Lambda) \otimes H^q(Y; \Lambda) \xrightarrow{\mu_*} H^{p+q}(\mathrm{Hom}(S(X) \otimes S(Y), \Lambda)) \xrightarrow{\rho^*} H^{p+q}(X \times Y; \Lambda)$, where ρ is the Alexander-Whitney mapping. For $\xi \in H^p(X; \Lambda)$ and $\eta \in H^q(Y; \Lambda)$, the **cross product** $\xi \times \eta \in H^{p+q}(X \times Y; \Lambda)$ is defined to be $\rho^* \mu_*(\xi \otimes \eta)$. The cohomology cross product satisfies the properties analogous to the homology cross product.

The cup product and the cohomology cross product are given in terms of each other: $\xi \smile \eta = d^*(\xi \times \eta)$, $\xi \times \eta = \pi_1^*(\xi) \smile \pi_2^*(\eta)$, where $d: X \to X \times X$ is given by $d(x) = (x, x)$, and $\pi_1: X \times Y \to X$, $\pi_2: X \times Y \to Y$ are projections.

We have the following **Künneth theorem** for cohomology: If Λ is a principal ideal domain and each $H_q(X; \Lambda)$ is finitely generated over Λ, then there is an exact sequence

$$0 \to \sum_{p+q=n} H^p(X; \Lambda) \otimes_\Lambda H^q(Y; \Lambda) \xrightarrow{\kappa} H^n(X \times Y; \Lambda)$$
$$\to \sum_{p+q=n+1} \mathrm{Tor}_\Lambda(H^p(X; \Lambda), H^q(Y; \Lambda)) \to 0,$$

and this sequence is split (→ 200 Homological Algebra). For $\xi \in H^p(X;\Lambda)$, $\eta \in H^q(Y;\Lambda)$, $\xi' \in H^s(X;\Lambda)$, $\eta' \in H^t(Y;\Lambda)$, the formula $(\xi \times \eta) \smile (\xi' \times \eta') = (-1)^{qs}(\xi \smile \xi') \times (\eta \smile \eta')$ holds. Therefore, if $\mathbf{k}$ is a field and $\dim_{\mathbf{k}} H_q(X;\mathbf{k}) < \infty$ for each q, then the cohomology ring $H^*(X \times Y;\mathbf{k})$ is determined by the cohomology rings $H^*(X;\mathbf{k})$ and $H^*(Y;\mathbf{k})$.

†Fiber bundles can be considered as generalized product spaces. Let E be the total space of a fiber bundle with base B and fiber F. The following **Leray-Hirsch theorem** (*J. Math. Pures Appl.*, 29 (1950)) asserts that, under certain conditions, the cohomology of E is additively isomorphic to that of $B \times F$: Let Λ be a principal ideal domain, and assume that $H_*(F;\Lambda)$ is free and finitely generated over Λ. Furthermore, assume that there is a homomorphism $\theta: H^*(F;\Lambda) \to H^*(E;\Lambda)$ such that the composite $H^*(F;\Lambda) \xrightarrow{\theta} H^*(E;\Lambda) \xrightarrow{i_b^*} H^*(p^{-1}(b);\Lambda)$ is an isomorphism for each $b \in B$, where $p: E \to B$ is the projection and $i_b: p^{-1}(b) \subset E$. Then an isomorphism $\Phi: H^*(B;\Lambda) \otimes_\Lambda H^*(F;\Lambda) \cong H^*(E;\Lambda)$ is given by $\Phi(\xi \otimes \eta) = p^*\xi \smile \theta(\eta)$, where $\xi \in H^*(B;\Lambda)$, $\eta \in H^*(F;\Lambda)$.

A general connection between (co)homology of E and $B \times F$ is given by means of spectral sequences (→ 148 Fiber Spaces).

K. Cap and Slant Products

There are other products closely related to the cup product or the cross product that involve cohomology and homology together.

Given a topological space X and a ring Λ, the **cap product** $v \frown c \in S_p(X) \otimes \Lambda$ **of a cochain** $v \in \mathrm{Hom}(S_q(X), \Lambda)$ **and a chain** $c = \sum_i \sigma_i \otimes \lambda_i \in S_{p+q}(X) \otimes \Lambda$ is defined by $v \frown c = \sum_i \sigma_i \circ \varepsilon \otimes v(\sigma_i \circ \varepsilon')\lambda_i$, where σ_i are singular $(p+q)$-simplexes in X, $\lambda_i \in \Lambda$, and $\varepsilon: \Delta^p \to \Delta^{p+q}$, $\varepsilon': \Delta^q \to \Delta^{p+q}$ are the mappings used in the definition of cup product. For any $u \in \mathrm{Hom}(S_p(X), \Lambda)$, the formula $\langle u \smile v, c\rangle = \langle u, v \frown c\rangle$ holds. The cap product satisfies $\partial(v \frown c) = (-1)^p \delta v \frown c + v \frown \partial c$, and hence it induces the **cap product** $\xi \frown a \in H_p(X;\Lambda)$ **of a cohomology class** $\xi \in H^q(X;\Lambda)$ **and a homology class** $a \in H_{p+q}(X;\Lambda)$. If Λ is a commutative ring with 1, then the cap product operation is bilinear and satisfies the following properties: $(\xi \smile \xi') \frown a = \xi \frown (\xi' \frown a)$, $f_*(f^*\eta \frown a) = \eta \frown f_*(a)$, $1 \frown a = a$, $(\xi \times \eta) \frown (a \times b) = (-1)^{p(t-q)}(\xi \frown a) \times (\eta \frown b)$, where $\xi \in H^p(X;\Lambda)$, $\xi' \in H^{p'}(X;\Lambda)$, $\eta \in H^q(Y;\Lambda)$, $a \in H_s(X;\Lambda)$, $b \in H_t(Y;\Lambda)$, and $f: X \to Y$ is a continuous mapping.

Given topological spaces X, Y and a ring Λ, the **slant product** $w/d \in \mathrm{Hom}(S_p(X), \Lambda)$ **of a cochain** $w \in \mathrm{Hom}(S(X) \otimes S(Y))_{p+q}, \Lambda)$ **and a chain** $d = \sum_i \tau_i \otimes \lambda_i \in S_q(Y) \otimes \Lambda$ is defined by $(w/d)(\sigma) = \sum_i (w(\sigma \otimes \tau_i))\lambda_i$, where σ is a singular p-simplex in X, τ_i are singular q-simplexes in Y, and $\lambda_i \in \Lambda$. The slant operation satisfies $\delta(w/d) = (\delta w)/d - (-1)^p w/\partial d$. Therefore, under the identification $H^*(X \times Y;\Lambda) = H^*(\mathrm{Hom}(S(X) \otimes S(Y), \Lambda))$, it induces the **slant product** $\zeta/b \in H^p(X;\Lambda)$ **of a cohomology class** $\zeta \in H^{p+q}(X \times Y;\Lambda)$ **and a homology class** $b \in H_q(Y;\Lambda)$. For any $a \in H_p(X;\Lambda)$, it holds that $\langle \zeta/b, a\rangle = \langle \zeta, a \times b\rangle$.

Let G, G' and G'' be Abelian groups. Given a homomorphism $G' \otimes G'' \to G$, we write $g'g''$ for the image of $g' \otimes g'' \in G' \otimes G''$ in G. Then the cap product $\frown: H^q(X;G') \otimes H_{p+q}(X;G'') \to H_p(X;G)$ can be defined in the same way as before. Similar definitions are valid for the cup product, the cross products for homology and cohomology, and the slant product.

L. Relative Homology

If C' is a subcomplex of a chain complex C, then we have a chain complex $C/C' = \{C_q/C'_q, \bar{\partial}_q\}$, where C_q/C'_q denotes the quotient group and $\bar{\partial}_q$ is induced from ∂_q by passing to the quotient. C/C' is called the **quotient complex** of C by C'.

A **topological pair** (X, A) is composed of a topological space X and its subset A. Given a topological pair (X, A) and an Abelian group G, we have the chain complex $(S(X)/S(A)) \otimes G$ and the cochain complex $\mathrm{Hom}(S(X)/S(A), G)$. The homology group of $(S(X)/S(A)) \otimes G$ is denoted by $H_*(X, A; G)$ and is called the **relative singular homology group of** X **modulo** A **with coefficients in** G or the **singular homology group of** (X, A) **with coefficients in** G. The homology group $H_*(X;G) = H_*(X, \varnothing; G)$ is sometimes called the **absolute homology group**. Similar definitions are made for the cohomology group $H^*(X, A; G)$ of the cochain complex $\mathrm{Hom}(S(X)/S(A), G)$.

A **simplicial pair** (K, L) is composed of a simplicial complex K and its subcomplex L, and a CW **pair** (X, A) is composed of a CW complex X and its subcomplex A. The relative homology group $H_*(K, L; G) = H_*((C(K)/C(L)) \otimes G)$ of a simplicial pair (K, L) is isomorphic to $H_*(|K|, |L|; G)$. For a CW pair (X, A), there is an isomorphism $H_*(X, A; G) \cong \tilde{H}_*(X/A; G)$. Similar statements hold for the cohomology groups.

A continuous mapping $f: (X, A) \to (Y, B)$ of topological pairs is a continuous mapping $f: X \to Y$ such that $f(A) \subset B$. If $f: (X, A) \to (Y, B)$ is a continuous mapping, then $f_\#: S(X) \to S(Y)$ sends $S(A)$ to $S(B)$, and hence f induces homomorphisms $f_*: H_*(X, A; G) \to H_*(Y, B; G)$ and

$f^*: H^*(Y, B; G) \to H^*(X, A; G)$. Since the properties are analogous, we state them below only for the case of relative homology.

The following six properties are fundamental. (i) $1_* = 1$. (ii) $(g \circ f)_* = g_* \circ f_*$. (iii) Homotopy property: If $f, f' : (X, A) \to (Y, B)$ are †homotopic, then $f_* = f'_*$. (iv) Exactness property: There exists a **homology exact sequence of** (X, A): $\ldots \xrightarrow{\partial_*} H_q(A; G) \xrightarrow{i_*} H_q(X; G) \xrightarrow{j_*} H_q(X, A; G) \xrightarrow{\partial_*} H_{q-1}(A; G) \xrightarrow{i_*} \ldots$, where $i: A \subset X$, $j: (X, \varnothing) \subset (X, A)$, and ∂_* sends the homology class of a cycle of $(S(X)/S(A)) \otimes G$ represented by a chain $c \in S(X) \otimes G$ to the homology class of ∂c which is a cycle of $S(A) \otimes G$. ∂_* is called the **boundary homomorphism** or the **connecting homomorphism**. (v) Naturality of ∂_*: For any continuous mapping $f: (X, A) \to (Y, B)$, it holds that $\partial_* \circ f_* = (f | A)_* \circ \partial_*$. (vi) Excision property: If U is a subset of X such that the closure $\bar{U}$ is in $\operatorname{Int} A$, then the **excision isomorphism** $H_*(X - U, A - U; G) \cong H_*(X, A; G)$ is induced by inclusion.

The exactness property extends to the **homology exact sequence of a triple** (X, A, B): $\ldots \xrightarrow{\partial_*} H_q(A, B; G) \xrightarrow{i_*} H_q(X, B; G) \xrightarrow{j_*} H_q(X, A; G) \xrightarrow{\partial_*} H_{q-1}(A, B; G) \xrightarrow{i_*} \ldots$. A couple $\{X_1, X_2\}$ of subsets of X is said to be **excisive** if $H_*(X_1, X_1 \cap X_2) \cong H_*(X_1 \cup X_2, X_2)$ is induced by inclusion. If $\{X_1, X_2\}$ is excisive, so is $\{X_2, X_1\}$. For example, if $X = \operatorname{Int} X_1 \cup \operatorname{Int} X_2$ or if X_1 and X_2 are subcomplexes of a CW complex, then $\{X_1, X_2\}$ is excisive. If $\{X_1, X_2\}$ and $\{A_1, A_2\}$ are excisive couples such that $A_1 \subset X_1$ and $A_2 \subset X_2$, then we have the **relative Mayer-Vietoris exact sequence**: $\ldots \to H_q(X_1 \cap X_2, A_1 \cap A_2; G) \to H_q(X_1, A_1; G) + H_q(X_2, A_2; G) \to H_q(X_1 \cup X_2, A_1 \cup A_2; G) \to H_{q-1}(X_1 \cap X_2, A_1 \cap A_2; G) \to \ldots$. For the case of relative cohomology, we use terms such as **cohomology exact sequence** and **coboundary homomorphism**, correspondingly.

The universal coefficient theorems are valid for the relative (co)homology groups. Given a homomorphism $G' \otimes G'' \to G$, if $\{A, B\}$ is excisive in X, then the cup product $\smile: H^p(X, A; G') \otimes H^q(X, B; G'') \to H^{p+q}(X, A \cup B; G)$ and the cap products $\frown: H^q(X, B; G') \otimes H_{p+q}(X, A \cup B; G'') \to H_p(X, A; G)$ can be defined. The product $(X, A) \times (Y, B)$ is defined to be the pair $(X \times Y, A \times Y \cup X \times B)$. Given a homomorphism $G' \otimes G'' \to G$, the homology cross products $\times: H_p(X, A; G') \otimes H_q(Y, B; G'') \to H_{p+q}((X, A) \times (Y, B); G)$ and the slant products $/: H^{p+q}((X, A) \times (Y, B); G') \otimes H_q(Y, B; G'') \to H^p(X, A; G)$ can be defined. If $\{A \times Y, X \times B\}$ is excisive, then the cohomology cross products $\times: H^p(X, A; G') \otimes H^q(Y, B; G'') \to H^{p+q}((X, A) \times (Y, B); G)$ can also be defined. If $\{A \times Y, X \times B\}$ is excisive, the Künneth theorems are valid for the relative (co)homology groups [9, 10, 11].

M. Čech Homology Theory

Another homology theory is commonly used along with the singular theory. The theory was originated by E. Čech (*Fund. Math.*, 19 (1932)) and was modified by C. H. Dowker (*Ann. Math.*, (2) 51 (1950)).

Given a topological space X and an Abelian group G, the **Čech homology group** $\check{H}_*(X; G)$ and the **Čech cohomology group** $\check{H}^*(X; G)$ are defined as follows. We take the family of all †open coverings of X directed by †refinement, and we consider the †nerve $K(\mathfrak{U})$ of each open covering $\mathfrak{U}$, that is, the simplicial complex whose simplexes are finite nonempty subsets of $\mathfrak{U}$ with nonempty intersection. If $\mathfrak{U}'$ is a refinement of $\mathfrak{U}$, then a simplicial mapping $\pi(\mathfrak{U}, \mathfrak{U}'): K(\mathfrak{U}') \to K(\mathfrak{U})$ is obtained by assigning to each $U' \in \mathfrak{U}'$ an element $U \in \mathfrak{U}$ such that $U' \subset U$. The induced homomorphisms $\pi(\mathfrak{U}, \mathfrak{U}')_*: H_*(K(\mathfrak{U}'); G) \to H_*(K(\mathfrak{U}); G)$ and $\pi(\mathfrak{U}, \mathfrak{U}')^*: H^*(K(\mathfrak{U}); G) \to H^*(K(\mathfrak{U}'); G)$ are independent of the choice of $\pi(\mathfrak{U}, \mathfrak{U}')$, and we have the †inverse system $\{H_*(K(\mathfrak{U}); G), \pi(\mathfrak{U}, \mathfrak{U}')_*\}$ and the †direct system $\{H^*(K(\mathfrak{U}); G), \pi(\mathfrak{U}, \mathfrak{U}')^*\}$. We now define $\check{H}_*(X; G) = \varprojlim H_*(K(\mathfrak{U}); G)$ and $\check{H}^*(X; G) = \varinjlim H^*(K(\mathfrak{U}); G)$.

A continuous mapping $f: X \to Y$ induces homomorphisms $f_*: \check{H}_*(X; G) \to \check{H}_*(Y; G)$ and $f^*: \check{H}^*(Y; G) \to \check{H}^*(X; G)$ as follows. If $\mathfrak{V}$ is an open covering of Y then a simplicial mapping $f_{\mathfrak{V}}: K(f^{-1}(\mathfrak{V})) \to K(\mathfrak{V})$ is defined by $f_{\mathfrak{V}}(f^{-1}(V)) = V$ $(V \in \mathfrak{V})$. The induced homomorphisms $f_{\mathfrak{V}*}: H_*(K(f^{-1}(\mathfrak{V})); G) \to H_*(K(\mathfrak{V}); G)$ for all open coverings $\mathfrak{V}$ of Y gives rise to $f_*: \check{H}_*(X; G) \to \check{H}_*(Y; G)$. Similarly f induces f^*.

Another approach to Čech cohomology theory is called the Alexander-Kolmogorov construction (*Proc. Nat. Acad. Sci.*, 21 (1935) and *C. R. Acad. Sci. Paris*, 202 (1936)). The approach was improved by E. H. Spanier (*Ann. Math.*, (2) 49 (1948)), and the theory is now called the **Alexander** (or **Alexander-Kolmogorov-Spanier**) **cohomology theory**.

The **Alexander cohomology group** $\bar{H}^*(X; G)$ is defined as follows. Let $\Phi^q(X; G)$ be the Abelian group of all functions from the $(q+1)$-fold product space X^{q+1} to G with addition defined pointwise. An element $\varphi \in \Phi^q(X; G)$ is said to be locally zero if there is an open covering $\mathfrak{U}$ of X such that $\varphi(x_0, \ldots, x_q)$ vanishes if $x_0, \ldots, x_q$ are contained simultaneously in some $U \in \mathfrak{U}$. The subgroup of $\Phi^q(X; G)$ consisting of locally zero functions is denoted by $\Phi^q_0(X; G)$. We define a homomorphism $\delta^q: \Phi^q(X; G) \to \Phi^{q+1}(X; G)$ by $(\delta^q \varphi)(x_0, x_1, \ldots, x_{q+1}) = \sum_{i=0}^{q+1} (-1)^i \varphi(x_0, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{q+1})$. Then $\Phi(X; G) = \{\Phi^q(X; G), \delta^q\}$ is a cochain complex, and $\Phi_0(X; G) = \{\Phi^q_0(X; G), \delta^q\}$ is its subcomplex. We now define $\bar{H}^*(X; G)$ to be

the cohomology group of the quotient complex $\bar{\Phi}(X;G)=\Phi(X;G)/\Phi_0(X;G)$.

If $f: X\rightarrow Y$ is a continuous mapping, then a cochain map $f^{\#}:\Phi(Y;G)\rightarrow\Phi(X;G)$ is defined by $(f^{\#}\varphi)(x_0,x_1,\ldots,x_q)=\varphi(f(x_0),f(x_1),\ldots,f(x_q))$, and it induces the homomorphism f^*: $\bar{H}^*(Y;G)\rightarrow\bar{H}^*(X;G)$. There is a natural isomorphism $\check{H}^*(X;G)\cong\bar{H}^*(X;G)$.

The (co)homology group of a simplicial complex K is isomorphic to the Čech (co)homology group of $|K|$. If X is a manifold or a CW complex, then its singular (co)homology group and its Čech (co)homology group are isomorphic. However, even for compact metric spaces X, the singular (co)homology group of X is not necessarily isomorphic with the Čech (co)homology group of X.

If $\{X_\lambda\}$ is an †inverse system of compact Hausdorff spaces and $X=\varprojlim X_\lambda$, then there are isomorphisms $\varprojlim\check{H}_*(X_\lambda;G)\cong\check{H}_*(X;G)$ and $\varinjlim\check{H}^*(X_\lambda;G)\cong\check{H}^*(X;G)$. This is called the **continuity property for Čech theory**. If A is any closed subset of a manifold M, then there is an isomorphism $\varinjlim H^*(W;G)\cong\check{H}^*(A;G)$, where W varies over neighborhoods of A in M directed downward by inclusion. If the †covering dimension of X is n, $\check{H}^q(X;G)=0$ for $q>n$.

The cup product in the Čech cohomology is introduced simply by passing to the limit with cup products in simplicial complexes, and the cup product in the Alexander cohomology is induced from the operation $\smile:\Phi^p(X;G')\otimes\Phi^q(X;G'')\rightarrow\Phi^{p+q}(X;G)$ defined by $(\varphi\smile\psi)(x_0,x_1,\ldots,x_{p+q})=\varphi(x_0,x_1,\ldots,x_p)\psi(x_p,x_{p+1},\ldots,x_{p+q})$.

The **relative Čech homology group** $\check{H}_*(X,A;G)$ and the **relative Čech cohomology group** $\check{H}^*(X,A;G)$ are defined as follows: An open covering of (X,A) is a pair $(\mathfrak{U},\mathfrak{N})$ of an open covering $\mathfrak{U}$ of X and an open covering $\mathfrak{N}$ of A such that $\mathfrak{N}\subset\mathfrak{U}$. To such a pair $(\mathfrak{U},\mathfrak{N})$ we assign a simplicial pair $(K(\mathfrak{U}),K'(\mathfrak{N}))$, where $K'(\mathfrak{N})$ is the nerve of $\mathfrak{N}\cap A=\{N\cap A\mid N\in\mathfrak{N}\}$. Considering the family of all open coverings of (X,A), we define now $\check{H}_*(X,A;G)=\varprojlim H_*(K(\mathfrak{U}),K'(\mathfrak{N});G)$ and $\check{H}^*(X,A;G)=\varinjlim H^*(K(\mathfrak{U}),K'(\mathfrak{N});G)$.

The **relative Alexander cohomology group** $\bar{H}^*(X,A;G)$ is defined to be the cohomology group of the cochain complex which is the kernel of the cochain mapping $i^{\#}:\bar{\Phi}(X;G)\rightarrow\bar{\Phi}(A;G)$ induced by inclusion. There is a natural isomorphism $\check{H}^*(X,A;G)\cong\bar{H}^*(X,A;G)$.

The relative Čech (co)homology groups satisfy the properties analogous to the relative singular (co)homology groups except the exactness property for homology (→ Section Q). In certain cases, the excision property is strengthened for Čech (co)homology. For example, we have the following theorem: Assume that X and Y are compact Hausdorff spaces, A and B are closed subsets of X and Y, respectively, and $f:(X,A)\rightarrow(Y,B)$ is a continuous mapping which maps $X-A$ onto $Y-B$ homeomorphically. Then $f_*:\check{H}_*(X,A;G)\cong\check{H}_*(Y,B;G)$ and $f^*:\check{H}^*(Y,B;G)\cong\check{H}^*(X,A;G)$ hold.

N. Fundamental Classes of Manifolds

For a topological space X and a point x of X, the **local homology groups** $H_*(X,X-x)$ represent a topological property of X around x.

The notion of †orientation for differentiable manifolds and triangulable manifolds generalizes to †topological manifolds by using local homology groups as follows. Let M be an n-dimensional (topological) manifold with †boundary ∂M. If x is a point of the interior $M_0=M-\partial M$, then $H_q(M,M-x)\cong H_q(\mathbf{R}^n,\mathbf{R}^n-0)$ is $\mathbf{Z}$ for $q=n$ and is 0 for $q\neq n$. We define a **local orientation** o_x for M at $x\in M_0$ to be a choice of one of the two possible generators for $H_n(M,M-x)$, and we then define an **orientation** for M to be a function which assigns to each $x\in M_0$ a local orientation o_x which varies continuously with x in the following sense: For each x there should exist a compact neighborhood $N\subset M_0$ and an element $o_N\in H_n(M,M-N)$ such that $i_{y*}(o_N)=o_y$ for each $y\in N$, where $i_y:(M,M-N)\subset(M,M-y)$. If there is an orientation for M, then M is said to be **orientable**, and the pair of M and an orientation is called an **oriented manifold**. If M is a nonorientable manifold without boundary, the set of local orientations for M forms an orientable manifold doubly covering M, called the **orientation manifold** of M.

If M is an oriented n-dimensional manifold, then for any compact subset K of M there is a unique element $o_K\in H_n(M,(M-K)\cup\partial M)$ such that $i_{x*}(o_K)=o_x$ for each $x\in K\cap M_0$, where $i_x:(M,(M-K)\cup\partial M)\subset(M,M-x)$. The element o_K is called the **fundamental homology class around** K. In particular, if M is itself compact, $o_M\in H_n(M,\partial M)$ is usually denoted by $[M]$ and is called the **fundamental homology class** of M. A connected compact n-dimensional manifold M is orientable if and only if $H_n(M,\partial M)\neq 0$, and in this case $H_n(M,\partial M)$ is a free cyclic group generated by a fundamental class $[M]$. If M is an orientable compact n-dimensional manifold, then ∂M is an orientable compact $(n-1)$-dimensional manifold without boundary, and the boundary homomorphism $\partial_*:H_n(M,\partial M)\rightarrow H_{n-1}(\partial M)$ sends a fundamental class $[M]$ to a fundamental class $[\partial M]$.

An n-dimensional manifold M is orientable if and only if there exists an element $U\in H^n(M\times M,M\times M-dM)$ such that, for each $x\in M_0$,

$j_x^*(U)$ is a generator of $H^n(M, M-x)$, where dM is the diagonal in $M\times M$, and $j_x:(M, M-x)\to(M\times M, M\times M-dM)$ is given by $j_x(y)=(x,y)$ $(y\in M)$. In fact, U corresponds to an orientation which assigns to each $x\in M_0$ a local orientation o_x such that $\langle j_x^*(U), o_x\rangle=1$. The element U is called the **orientation cohomology class** of M. If M is a compact manifold without boundary, it holds that $\langle d^*(U),[M]\rangle=\chi(M)$, where $d^*:H^n(M\times M, M\times M-dM)\to H^n(M)$ is induced by the diagonal mapping. The element $d^*(U)\in H^n(M)$ is called the **Euler class** of M.

If we work with the (co)homology groups with coefficients in $\mathbf{Z}_2$, the fundamental classes are defined for an arbitrary manifold without making any assumption of orientability. If M is connected and compact, then $H_n(M, \partial M;\mathbf{Z}_2)\cong\mathbf{Z}_2$ is generated by $[M]$.

O. Duality in Manifolds

Let M be a compact n-dimensional manifold, and let M_1, M_2 be compact $(n-1)$-dimensional manifolds such that $M_1\cup M_2=\partial M$ and $M_1\cap M_2=\partial M_1=\partial M_2$. Assume either that M is oriented or that $G=\mathbf{Z}_2$. Then for each q, an isomorphism $D:H^q(M, M_1;G)\cong H_{n-q}(M, M_2;G)$ is defined by $D(\xi)=\xi\frown[M]$. In particular, there are the isomorphisms $D:H^q(M, \partial M;G)\cong H_{n-q}(M;G)$ and $H^q(M;G)\cong H_{n-q}(M, \partial M;G)$, where the cap product is taken with respect to the homomorphism $G\otimes\mathbf{Z}\to G$ defined by multiplication. This theorem is called the **Poincaré-Lefschetz duality theorem**, and the special case for $\partial M=\varnothing$ is often referred to as **Poincaré duality**.

Poincaré duality implies the following consequences for a compact n-dimensional manifold M without boundary. If M is orientable, then the qth Betti number is equal to the $(n-q)$th Betti number, and the qth torsion coefficients are equal to the $(n-q-1)$th torsion coefficients. If n is odd, then $\chi(M)=0$, and if M is orientable and $n\equiv 2 \bmod 4$, then $\chi(M)$ is even.

Poincaré duality generalizes to the following duality theorem. Let M be an n-dimensional manifold without boundary, and let K be a compact subset of M. Assume either that M is oriented or that $G=\mathbf{Z}_2$. Then there is an isomorphism $D:\check{H}^q(K;G)\cong H_{n-q}(M, M-K;G)$ for any q, which is given as follows: For each open neighborhood W of K, define $D_W:H^q(W;G)\to H_{n-q}(M, M-K;G)$ by $D_W(\xi)=k_*(\xi\frown k_*^{-1}(o_K))$, where k_* is the excision isomorphism induced by $k:(W, W-K)\subset(M, M-K)$. Now D is defined to be the limit of D_W, where W varies over open neighborhoods of K. The inverse of D up to sign is given in terms of the slant product with the fundamental cohomology class U of M as follows: For each open neighborhood W of K, we define a homomorphism $\gamma_W:H_{n-q}(M, M-W;G)\to H^q(W;G)$ by $\gamma_W(a)=j^*(U)/a$, where $j^*:H^n(M\times M, M\times M-dM)\to H^n(W\times(M, M-W))$ is induced by inclusion. Then, passing to the limit, these γ_W define the desired one.

If we take an n-sphere S^n as M in the above duality theorem and use the homology exact sequence of (S^n, S^n-K), then we have the following **Alexander duality theorem** (*Trans. Amer. Math. Soc.*, 23 (1922)): If K is a closed subset of S^n, then the qth reduced Čech cohomology group of K is isomorphic to the $(n-q-1)$th reduced singular homology group of S^n-K for any coefficient group G and any q. In particular, if K is a †neighborhood retract, $\tilde{H}^q(K;G)\cong\tilde{H}_{n-q-1}(S^n-K;G)$ holds. This shows that $H_*(S^n-K)$ depends only on K and not on the way K is embedded in S^n. The Alexander duality theorem for $n=2$ and $K=S^1$ gives the classical †Jordan curve theorem.

In view of the duality theorems, certain classical definitions in the homology of manifolds can be given in terms of cohomology. For example, if $f:M\to M'$ is a continuous mapping of oriented closed manifolds, then the **Umkehr homomorphism** or **Gysin homomorphism** $f^!:H_q(M';G)\to H_{q+d}(M;G)$ $(d=\dim M-\dim M')$ (W. Gysin, *Comment. Math. Helv.*, 14 (1941)) can be defined by $D\circ f^!=f^*\circ D$. In cohomology we have $f_!:H^q(M;G)\to H^{q-d}(M';G)$. Similarly, if M is an oriented n-dimensional closed manifold and $a\in H_p(M)$, $b\in H_q(M)$, then the **intersection product** $a\cdot b\in H_{p+q-n}(M)$ of Lefschetz can be defined by $a\cdot b=D^{-1}a\frown b=D(D^{-1}a\smile D^{-1}b)$. If $p+q=n$, the number $a\cdot b\in H_0(M)\cong\mathbf{Z}$ is called the **intersection number** of a and b. The classical definitions are still meaningful today, since they are closer to geometric intuition and therefore possess considerable heuristic value. For example, the following fact serves to compute cup products in manifolds. If M is an oriented closed differentiable manifold and a, $b\in H_*(M)$ are represented by closed submanifolds N_1, N_2 which intersect †transversally, then $\pm a\cdot b$ is represented by $N_1\cap N_2$ [11]. See [12] for a rigorous discussion of classical intersection theory.

P. Cohomology with Compact Supports

Let X be a topological space. A subset V of X is said to be **cobounded** if $\overline{X-V}$ is compact. A singular q-cochain $u\in\operatorname{Hom}(S_q(X), G)$ is said to have **compact support** if there exists a cobounded set V such that $u(\sigma)=0$ for every singular q-simplex σ in V. The singular cochains with compact support form a subcom-

plex of the cochain complex $\mathrm{Hom}(S(X), G)$. The cohomology group of this subcomplex is denoted by $H_c^*(X; G)$ and is called the **singular cohomology group of X with compact supports**. There is an isomorphism $H_c^*(X; G) \cong \varinjlim H^*(X, V; G)$, where V varies over cobounded subsets of X.

Let K be a simplicial complex. A q-cochain $u \in \mathrm{Hom}(C_q(K), G)$ is called a **finite cochain** of K if $u(\sigma) = 0$ except for a finite number of oriented q-simplexes σ of K. If K is a †locally finite simplicial complex, then finite cochains of K form a subcomplex of the cochain complex $\mathrm{Hom}(C(K), G)$ whose cohomology group is isomorphic to $H_c^*(|K|; G)$.

Let X be a †locally compact Hausdorff space, and let $X \cup \{\infty\}$ denote the †one-point compactification of X. Then the **Čech cohomology group of X with compact supports**, denoted by $\check{H}_c^*(X; G)$ is defined to be the reduced Čech cohomology group of $X \cup \{\infty\}$ with coefficients in G. There is an isomorphism $\check{H}_c^*(X; G) \cong \varinjlim \check{H}^*(X, V; G)$, where V varies over cobounded subsets of X. If X is a manifold or a CW complex, then $H_c^*(X; G) \cong \check{H}_c^*(X; G)$. If X is a compact Hausdorff space and A is closed in X, then $\check{H}_c^*(X - A; G) \cong H^*(X, A; G)$. The Alexander-Kolmogorov construction gives a direct approach to $\check{H}_c^*(X; G)$ [10, 13].

A †proper continuous mapping $f: X \to Y$ of locally compact Hausdorff spaces induces homomorphisms $f^*: H_c^*(Y; G) \to H_c^*(X; G)$ and $f^*: \check{H}_c^*(Y; G) \to \check{H}_c^*(X; G)$, and if $f, f': X \to Y$ are properly homotopic, then they induce the same homomorphisms.

The cohomology with compact supports is useful in order to extend results in the cohomology of compact spaces to noncompact spaces. For example, the conclusion of the duality theorem on a compact set $K \subset M$ in Section O generalizes to the case of a closed set $K \subset M$ as follows: There is an isomorphism $\check{H}_c^q(K; G) \cong H_{n-q}(M, M - K; G)$ for any q. This implies the following generalization of Poincaré duality: $H_c^q(M; G) \cong H_{n-q}(M; G)$ holds for an orientable n-dimensional manifold M without boundary.

There are homology theories associated with the cohomology theories with compact supports [13].

Q. Eilenberg-Steenrod Axioms

Let H^* be a collection of the following three functions: (1) A function assigning to each topological pair (X, A) and each integer q an Abelian group $H^q(X, A)$. (2) A function assigning to each continuous mapping $f: (X, A) \to (Y, B)$ and each integer q a homomorphism $f^*: H^q(Y, B) \to H^q(X, A)$. (3) A function assigning to each topological pair (X, A) and each integer q a homomorphism $\delta^*: H^q(A) \to H^{q+1}(X, A)$. Then H^* is called a **cohomology theory on the category of topological pairs** if the following seven axioms are satisfied [7]. (i) $1_* = 1$, where 1 is identity. (ii) $(g \circ f)^* = f^* \circ g^*: H^q(Z, C) \to H^q(X, A)$ for continuous mappings $f: (X, A) \to (Y, B)$ and $g: (Y, B) \to (Z, C)$. (iii) Homotopy axiom: If $f, f': (X, A) \to (Y, B)$ are homotopic, then $f_* = f'_*: H^q(Y, B) \to H^q(X, A)$. (iv) Exactness axiom: The sequence $\ldots \xrightarrow{\delta^*} H^q(X, A) \xrightarrow{j^*} H^q(X) \xrightarrow{i^*} H^q(A) \xrightarrow{\delta^*} H^{q+1}(X, A) \xrightarrow{j^*} \ldots$ is exact, where $i: A \subset X$ and $j: (X, \varnothing) \subset (X, A)$. (v) $f^* \circ \delta^* = \delta^* \circ (f | A)^*: H^q(B) \to H^{q+1}(X, A)$ for a continuous mapping $f: (X, A) \to (Y, B)$. (vi) Excision axiom: If U is an open set of X such that $\bar{U} \subset \mathrm{Int}\, A$, then $i^*: H^q(X, A) \cong H^q(X - U, A - U)$, where i is the inclusion. (vii) Dimension axiom: $H^q(pt) = 0$ if $q \neq 0$. Axioms (i)–(vii) are called the **Eilenberg-Steenrod axioms**, and the group $H^0(pt)$ is called the **coefficient group** of the cohomology theory H^*.

A cohomology theory on the category of pairs of compact Hausdorff spaces is defined similarly. A cohomology theory on the category of CW pairs (or finite CW pairs) is defined similarly except that axiom (vi) is replaced by the following excision axiom: If $\{X_1, X_2\}$ is a couple of subcomplexes of a CW complex, then $i^*: H^q(X_1 \cup X_2, X_2) \cong H^q(X_1, X_1 \cap X_2)$, where i is the inclusion. Two cohomology theories H^* and H'^* on the same category are **isomorphic** if there is an isomorphism $h_q: H^q(X, A) \cong H'^q(X, A)$ for each (X, A) and each q, and they commute with f^* and δ^*. A homology theory on various categories is defined similarly by dualization.

A singular (co)homology theory with coefficients in G is an example of a (co)homology theory on the category of topological pairs. The Čech cohomology groups with coefficients in G can be made into a cohomology theory on the category of topological pairs. However, the Čech homology groups do not constitute a homology theory on the category of topological pairs; the homology sequence of any pair (X, A) is defined, but it can be proved only that the composite of any two successive homomorphisms is zero. The Čech homology groups with coefficients in a field constitute a homology theory on the category of compact Hausdorff pairs. The Alexander cohomology groups constitute a cohomology theory on the same category of topological pairs, and it is isomorphic to the Čech cohomology theory if their coefficient groups are isomorphic [10].

The Čech (co)homology constitutes a (co)homology theory on the category of CW pairs, and it is isomorphic to the singular (co)homology theory on the same category if

the coefficient groups are isomorphic. (Co)-homology theories on the category of finite CW pairs are determined, up to isomorphisms, by their coefficient groups. This fact is called the **uniqueness theorem of homology theory** on the category of finite CW pairs. Cohomology theories on the category of pairs of compact Hausdorff spaces which satisfy the following continuity axiom are determined, up to isomorphisms, by their coefficient groups: If $\{(X_\lambda, A_\lambda)\}$ is an inverse system of pairs of compact Hausdorff spaces, then $H^q(\varprojlim X_\lambda, \varprojlim A_\lambda) \cong \varinjlim H^q(X_\lambda, A_\lambda)$. The Čech cohomology theory satisfies this axiom.

During recent years, many (co)homology theories have been developed which satisfy the first six Eilenberg-Steenrod axioms but fail to satisfy the dimension axiom. These are called **generalized (co)homology theories**, and include various †K-theories, †bordism theories, and †stable homotopy theories (→ 202 Homotopy Theory).

R. Homology with Coefficients in Local Systems

N. E. Steenrod (*Ann. Math.*, (2) 44 (1943)) introduced the (co)homology group with coefficients in a local system of Abelian groups, which is useful in †obstruction theory and in the homology theory of †fiber spaces.

A **local system** $\mathfrak{G}$ of Abelian groups on a topological space X is a set of Abelian groups G_x, one for each $x \in X$, together with an isomorphism $l^*: G_{l(0)} \to G_{l(1)}$ for each †path $l:[0,1] \to X$ subject to the following conditions: (1) If two paths l and l' are homotopic with endpoints fixed, then $l^* = l'^*$. (2) If l and m are paths such that $l(1) = m(0)$, then $(l \cdot m)^* = m^* \circ l^*$, where $l \cdot m$ denotes the †product of l and m. An example is provided by the †homotopy groups $\pi_n(X, x)$ for $n \geqslant 2$. Let M be an n-dimensional topological manifold. Then $x \to H_n(M, M - x)$ is a local system of infinite cyclic groups. It is called the **orientation sheaf** of M. A local system $\mathfrak{G}$ is said to be trivial if $l^* = l'^*$ for any paths l, l' with the same initial and final points.

Given a local system $\mathfrak{G}$ of Abelian groups on a topological space X, a chain complex $S(X; \mathfrak{G}) = \{S_q(X; \mathfrak{G}), \partial_q\}$ is defined as follows: If $q < 0$, then $S_q(X; \mathfrak{G}) = 0$, and if $q \geqslant 0$, then $S_q(X; \mathfrak{G})$ is the Abelian group of formal finite sums $\sum g_\sigma \sigma$, where $\sigma: \Delta^q \to X$ are singular q-simplexes in X and $g_\sigma \in G_{\sigma(1,0,\ldots,0)}$; the boundary operator $\partial_q: S_q(X; \mathfrak{G}) \to S_{q-1}(X; \mathfrak{G})$ is given by $\partial_q(g_\sigma \sigma) = l_\sigma^*(g_\sigma)\sigma \circ \varepsilon_0 + \sum_{i=1}^{q}(-1)^i g_\sigma \sigma \circ \varepsilon_i$, where $\sigma \circ \varepsilon_i$ is the ith face of σ, and $l_\sigma: [0,1] \to X$ is given by $l_\sigma(t) = \sigma(1-t, t, 0, \ldots, 0)$. The homology group of the chain complex $S(X; \mathfrak{G})$ is denoted by $H_*(X; \mathfrak{G})$ and is called the **singular homology group of X with coefficients in $\mathfrak{G}$**.

Similarly, a cochain complex $S^*(X; \mathfrak{G}) = \{S^q(X; \mathfrak{G}), \delta^q\}$ is defined as follows: If $q < 0$, then $S^q(X; \mathfrak{G}) = 0$, and if $q \geqslant 0$, then $S^q(X; \mathfrak{G})$ is the Abelian group of functions u assigning to every singular q-simplex σ in X an element $u(\sigma) \in G_{\sigma(1,0,\ldots,0)}$; the coboundary operator $\delta^q: S^q(X; \mathfrak{G}) \to S^{q+1}(X; \mathfrak{G})$ for $q \geqslant 0$ is given by $(\delta^q u)(\sigma) = l_\sigma^{*-1} u(\sigma \circ \varepsilon_0) + \sum_{i=1}^{q+1}(-1)^i u(\sigma \circ \varepsilon_i)$, where σ is a singular $(q+1)$-simplex in X. The cohomology group of the cochain complex $S^*(X; \mathfrak{G})$ is denoted by $H^*(X; \mathfrak{G})$ and is called the **singular cohomology group of X with coefficients in $\mathfrak{G}$**.

If $\mathfrak{G}$ is trivial, then the (co)homology group with coefficients in $\mathfrak{G}$ coincides with the (co)-homology group with coefficients in $G \cong G_x$. The various notions and theorems on the ordinary (co)homology can be extended to (co)homology with coefficients in $\mathfrak{G}$. The Čech (co)homology group with coefficients in $\mathfrak{G}$ is also defined [10]. The cohomology groups with coefficients in $\mathfrak{G}$ are generalized to the cohomology groups with coefficients in a sheaf [10, 14] (→ 383 Sheaves).

References

[1] H. Poincaré, Analysis situs, J. École Polytech., 1 (1985), 1–121; Rend. Circ. Mat. Palermo, 13 (1899), 285–343.

[2] O. Veblen, Analysis situs, Amer. Math. Soc. Colloq. Publ., 1922.

[3] S. Lefschetz, Topology, Amer. Math. Soc. Colloq. Publ., 1930.

[4] H. Seifert and W. Threlfall, Lehrbuch der Topologie, Teubner, 1934 (Academic Press, 1980).

[5] P. S. Aleksandrov and H. Hopf, Topologie I, Springer, 1935 (Chelsea, 1965).

[6] S. Lefschetz, Algebraic topology, Amer. Math. Soc. Colloq. Publ., 1942.

[7] S. Eilenberg and N. E. Steenrod: Foundations of algebraic topology, Princeton Univ. Press, 1952.

[8] H. Cartan and S. Eilenberg, Homological algebra, Princeton Univ. Press, 1956.

[9] P. J. Hilton and S. Wylie, Homology theory, Cambridge Univ. Press, 1960.

[10] E. H. Spanier, Algebraic topology, McGraw-Hill, 1966.

[11] A. Dold, Lectures on algebraic topology, Springer, 1972.

[12] M. Glezerman and L. S. Pontryagin, Intersections in manifolds, Amer. Math. Soc. Transl., 50 (1951). (Original in Russian, 1947.)

[13] W. S. Massey, Homology and cohomology theory, Dekker, 1978.

[14] R. Godement, Topologie algébrique et théorie des faisceaux, Hermann, 1958.
[15] S. Lefschetz, Introduction to topology, Princeton Univ. Press, 1949.
[16] J. G. Hocking and G. Young, Topology, Addison Wesley, 1961.
[17] D. G. Bourgin, Modern algebraic topology, Macmillan, 1963.
[18] H. Schubert, Topologie, Teubner, 1964 (Allyn and Bacon, 1968).
[19] S. T. Hu, Homology theory, Holden-Day, 1966.
[20] M. J. Greenberg, Lectures on algebraic topology, Benjamin, 1967.
[21] G. E. Cooke and R. L. Finney, Homology of cell complexes, Princeton Univ. Press, 1967.
[22] C. R. F. Maunder, Algebraic topology, Van Nostrand, 1970.
[23] A. H. Wallace, Algebraic topology, Benjamin, 1970.
[24] C. Godbillon, Elements de topologie algébrique, Hermann, 1971.
[25] J. Vick, Homology theory, Academic Press, 1973.
[26] J. W. Milnor and J. D. Stasheff, Characteristic classes, Princeton Univ. Press, 1974.
[27] W. S. Massey, Singular homology theory, Springer, 1980.

202 (IX.9) Homotopy Theory

A. General Remarks

Given a topological space X, we utilize the concept of homotopy to define the †fundamental group, homotopy groups, and cohomotopy groups of X. These groups, together with (co)homology groups, are useful tools in topology.

Since the research of H. Hopf, W. Hurewicz, and H. Freudenthal in the 1930s, homotopy theory has made rapid progress and now plays an important role in topology.

B. Homotopy

If a family $f_t: X \to Y$ ($t \in I = \{t \mid 0 \leqslant t \leqslant 1\}$) of †continuous mappings from a †topological space X into a topological space Y is also continuous with respect to t, that is, if the mapping F from the product space $X \times I$ into Y defined by $F(x,t) = f_t(x)$ ($x \in X, t \in I$) is continuous, then $\{f_t\}$ or F is called a **homotopy**. In this case, f_0 and f_1 are said to be **homotopic.** This relation between f_0 and f_1 is indicated by $f_0 \simeq f_1: X \to Y$, or simply $f_0 \simeq f_1$, and is called the relation of homotopy. Denote by Y^X the set of all continuous mappings from X into Y. The homotopy relation is an †equivalence relation on Y^X, and the equivalence class $[f]$ of a mapping $f: X \to Y$ is called the **homotopy class** (or **mapping class**) of f. The set of all homotopy classes of mappings of X into Y is called the **homotopy set** and is denoted by $\pi(X; Y)$ or $[X, Y]$. A function γ of continuous mappings $f \in Y^X$ is called a **homotopy invariant** if $f \simeq g$ implies $\gamma(f) = \gamma(g)$. When X consists of a point $*$ we write $\pi(*; Y) = \pi_0(Y)$. If all continuous mappings in Y^X are homotopic to each other, we write $\pi(X; Y) = 0$; $\pi_0(Y) = 0$ means that Y is †arcwise connected. A mapping f from a compact space into an n-dimensional sphere S^n is called essential if any mapping g homotopic to f satisfies $g(X) = S^n$. A mapping is inessential if and only if it is homotopic to the constant mapping.

These concepts are generalized as follows: Let A_i and B_i ($i = 1, 2, \ldots$) be subspaces of X and Y, respectively, and denote by $Y^X(A_1, A_2, \ldots; B_1, B_2, \ldots)$ the set of continuous mappings $f \in Y^X$ satisfying $f(A_i) \subset B_i$. If a homotopy $\{f_t\}$ is such that $f_t \in Y^X(A_i; B_i)$, then $\{f_t\}$ is called a **restricted homotopy** with respect to A_i, B_i or a homotopy from a system of spaces $(X, A_1, A_2, \ldots)$ into a system of spaces $(Y, B_1, B_2, \ldots)$. The notation $f_0 \simeq f_1: (X, A_1, A_2, \ldots) \to (Y, B_1, B_2, \ldots)$ and the homotopy set $\pi(X, A_1, A_2, \ldots; Y, B_1, B_2, \ldots)$ are defined accordingly.

For the composite $g \circ f \in Z^X(A_i; C_i)$ of $f \in Y^X(A_i; B_i)$ and $g \in Z^Y(B_i; C_i)$, $f \simeq f'$ and $g \simeq g'$ imply $g \circ f \simeq g' \circ f'$. Thus the **composite** $\beta \circ \alpha = [g \circ f] \in \pi(X, A_i; Z, C_i)$ of $[f] = \alpha \in \pi(X, A_i; Y, B_i)$ and $[g] = \beta \in \pi(Y, B_i; Z, C_i)$ is defined. By putting $g_*[f] = [g \circ f] = f^*[g]$ we induce two mappings,

$$g_*: \pi(X, A_i; Y, B_i) \to \pi(X, A_i; Z, C_i),$$

$$f^*: \pi(Y, B_i; Z, C_i) \to \pi(X, A_i; Z, C_i).$$

Then $f \simeq f'$ implies $f^* = f'^*$ and $g \simeq g'$ implies $g_* = g'_*$. Also $(g \circ f)_* = g_* \circ f_*$, $(g \circ f)^* = f^* \circ g^*$, and $h^* \circ g_* = g_* \circ h^*$, where $h \in X^W(D_i; A_i)$.

The **category of pointed topological spaces** is defined to be the †category in which each object X, which is a topological space, has a point fixed as a **base point** and each mapping $X \to Y$ carries the base point of X to the base point of Y. In this category, we define a homotopy set, denoted by $\pi(X; Y)_0$ or $[X; Y]_0$, as follows: Denoting the base points by $*$, we have $\pi(X, A_i, *; Y, B_i, *) = \pi(X, A_i; Y, B_i)_0$. A continuous mapping f homotopic to the constant mapping $X \to * \in Y$ is said to be **homotopic to zero** (or **null-homotopic**). This is indicated by $f \simeq 0$, and $\pi(X; Y)_0 = 0$ means that all continuous mappings are homotopic to zero. Let S^0 be a set of two points; then

$\pi(X; S^0)_0 = 0$ means that X is †connected. In contrast to these specific homotopies, the usual homotopy is sometimes called a **free homotopy**.

Suppose that a homotopy $\{f_t\}$ $(f_t: X \to Y)$ is such that the restriction of f_t to a subspace A of X is stationary, that is, $f_t(a) = f_0(a)$ $(a \in A, t \in I)$. Then f_0, f_1 are said to be **homotopic relative to** A, indicated by $f_0 \simeq f_1$ (rel. A). If a homotopy $\{f_t\}$ $(f_t: X \to Y)$ is such that each f_t is a homeomorphism into Y, then $\{f_t\}$ is called an **isotopy** and f_0 is called **isotopic** to f_1 (→ 235 Knot Theory).

Research done by L. E. Brouwer, H. Hopf, W. Hurewicz, K. Borsuk, L. S. Pontryagin, and S. Eilenberg has contributed to the theory of homotopy, an important field of topology still in the process of development.

C. Mapping Spaces

We endow the set Y^X of all continuous mappings $f: X \to Y$, with the †compact-open topology. The topological space Y^X is called a **mapping space**. In particular we denote Y^I $(0, 1; *, *)$ $(* \in Y)$ by $\Omega(Y) = \Omega(Y, *)$ and call it the **space of closed paths** (or **loop space**) of Y. Two points f, g of Y^X are connected by a †path in Y^X if and only if $f \simeq g: X \to Y$. Thus $\pi_0(Y^X) = \pi(X; Y)$ and $\pi_0(Y^X(A_i; B_i)) = \pi(X, A_i; Y, B_i)$.

D. Retracts

Let A be a subspace of a topological space X. If there exists an $f \in A^X$ such that the restriction $f|A$ is the identity mapping of A, then A is called a **retract** of X, and f a **retraction**. If A is a retract of X, any continuous mapping of A into any topological space can be extended to a continuous mapping of X. If A is a retract of some neighborhood $U(A)$, A is called a **neighborhood retract** or **NR** of X. If for any †homeomorphism of a metric space A onto a closed subspace A_0 of any metric space X, A_0 is a retract (neighborhood retract) of X, then A is called an **absolute retract** or **AR (absolute neighborhood retract** or **ANR)**. For example, an n-dimensional simplex or an n-dimensional Euclidean space is an AR. If a retraction f is homotopic to the identity mapping of X (resp. $U(A)$), we call A a **deformation retract (neighborhood deformation retract)** of X. Moreover, if $f \simeq 1_X$ (rel. A), then A is called a **strong deformation retract**. In particular, if a point x_0 is a (strong) deformation retract of X, we say that X is **contractible** to the point x_0. For example, any †polyhedron P and any compact n-dimensional †topological manifold are ANRs; any polyhedron P_0 contained in P in a strong deformation retract of some neighborhood in P. X is called **locally contractible** if each point x of X has a contractible neighborhood U of x.

E. The Extension Property

Let X, Y be topological spaces, $A \subset X$, $f_0, f_1 \in Y^X$, and $\{g_t: A \to Y\}$ a homotopy such that $g_i = f_i | A$ $(i = 0, 1)$. We can extend $\{g_t\}$ to a homotopy $\{f_t\}$ of X if and only if the mapping $F: (X \times 0) \cup (A \times I) \cup (X \times 1) \to Y$ defined by $F(x, i) = f_i(x)$, $F(a, t) = g_t(a)$ can be extended to a continuous mapping sending $X \times I$ into Y. Therefore the problem of whether $f_0 \simeq f_1$ can be reduced to the problem of whether a continuous mapping defined on a subspace can be extended to the whole space. If for any homotopy $\{g_t: A \to Y\}$ and any continuous mapping $f_0: X \to Y$ into any topological space Y satisfying $f_0 | A = g_0$ there exists a homotopy $\{f_t: X \to Y\}$ satisfying $f_t | A = g_t$, then we say that (X, A) has the **homotopy extension property**. This occurs if and only if $(X \times 0) \cup (A \times I)$ is a retract of $X \times I$. A pair (X, A) of ANRs, where A is closed in X, and a pair (P, P_0) with P a †CW complex and P_0 a subcomplex of P have this property. Given a continuous mapping $h: B \to A$ of a subspace B of a topological space Y into a topological space A, we identify $b \in B$ with $h(b) \in A$ in the †direct sum $A \cup Y$ and obtain the †identification space denoted by $A \cup_h Y$, which is called an **attaching space** under h. If (Y, B) has the homotopy extension property, then $(Y \times X, B \times X)$ and $(A \cup_h Y, A)$ also have the same property. When A consists of a point $*$, we write $* \cup_h Y = Y/B$ and call the space Y/B **a space smashing (shrinking** or **pinching)** B **to a point**. If $Y = B \times I$, $B = B \times 0$, then we call $A \cup_h (B \times I)$ a **mapping cylinder** of h, $(A \cup_h (B \times I))/(B \times 1)$ a **mapping cone** of h, and the mapping cylinder and mapping cone of $h: B \to *$ the **cone** over B and **suspension** of B, respectively.

F. Homotopy Type

For systems (X, A_i), (Y, B_i) of topological spaces, if there exist $f \in Y^X(A_i; B_i)$, $g \in X^Y(B_i; A_i)$ such that $g \circ f$ and $f \circ g$ are homotopic to the identity mappings of (X, A_i) and (Y, B_i), respectively, then we say that (X, A_i) and (Y, B_i) have the same **homotopy type** or are **homotopy equivalent**. Such mappings f and g are called **homotopy equivalences**. For a homotopy equivalence f, the induced mappings f_* and f^* are bijective. Therefore, in homotopy theory, systems of spaces having the same homotopy type are considered equivalent.
If A is a deformation retract of X, then A and X have the same homotopy type, and the

injection of A into X and the retraction of X onto A are homotopy equivalences. A contractible space has the same homotopy type as a point. Spaces having the same homotopy type have isomorphic homotopy groups and †(co)homology groups. Since the mapping cylinder $Z_f = Y \cup_f (X \times I)$ of $f \in Y^X$ contains Y as its deformation retract, it has the same homotopy type as Y. By this homotopy equivalence, f can be replaced by the injection of $X \times 1$ into Z_f. If to each topological space there corresponds a value (which may be some element of **R** or some algebraic structure) and the values are the same for homotopy equivalent spaces, then the value is called a **homotopy type invariant**. A homotopy type invariant is a †topological invariant; for example, $\pi(X; Y)$ is a homotopy type invariant of X. If a continuous mapping $f: X \rightarrow Y$ induces isomorphisms of the homotopy groups of each †arcwise connected component, then f is called a **weak homotopy equivalence**. Conversely, if X and Y are CW complexes, then a weak homotopy equivalence is a homotopy equivalence (J. H. C. Whitehead).

Now we consider the category of pointed topological spaces. Let A and B be pointed topological spaces. Then the †direct sum in this category is the **one-point union** (or **bouquet**) $A \vee B$ obtained from the disjoint union $A \cup B$ by identifying two base points $*_A$ and $*_B$. $A \vee B$ is identified with the subspace $(A \times *_B) \cup (*_A \times B)$ in $A \times B$. The **reduced join** (or **smash product**) of A, B is the space obtained from $A \times B$ by smashing its subspace $A \vee B$ to a point and is denoted by $A \wedge B$. We call $A \wedge S^1$ the **(reduced) suspension** of A and denote it by SA. Repeating the suspension n times, we have the **n-fold reduced suspension** of A. We call $CA = A \wedge I$ $(I = [0, 1])$ the **reduced cone** of A (I has the base point 1). For a continuous mapping $f: X \rightarrow Y$, the space obtained by identifying each point $(x, 0)$ of the base of CX with $f(x) \in Y$ is called the **reduced mapping cone** and is denoted by $C_f = Y \cup_f CX$. The **reduced join** (or **smash product**) of mappings $f: Y \rightarrow X$ and $f': Y' \rightarrow X'$ is the mapping $f \wedge f'$: $Y \wedge Y' \rightarrow X \wedge X'$ induced from the product mapping $f \times f': Y \times Y' \rightarrow X \times X'$. The reduced join of $f: Y \rightarrow X$ and $1: S^1 \rightarrow S^1$ (identity mapping) is written as $Sf = f \wedge 1$ and is called the **suspension** of f.

G. Puppe Exact Sequences

For $f: X \rightarrow Y$ and $g: Y \rightarrow Z$, we have $g \circ f \simeq 0$ if and only if g can be extended to a continuous mapping from C_f into Z. In other words, the sequence

$$\pi(C_f; Z)_0 \xrightarrow{i^*} \pi(Y; Z)_0 \xrightarrow{f^*} \pi(X; Z)_0$$

is exact (i.e., $\operatorname{Im} i^* = \operatorname{Ker} f^* = f^{*-1}(0)$, where $i: Y \rightarrow C_f$ is the canonical inclusion and 0 is the class of the constant mapping). The inclusion $i: Y \rightarrow C_f$ gives rise to the reduced mapping cone C_i. We also have the canonical inclusion $i': C_f \rightarrow C_i$. Adding the term $\pi(C_i: Z)_0 \xrightarrow{i'^*}$ to the left-hand side of the sequence above, we have a new exact sequence. Continuing this process, we obtain an exact sequence of infinite length. If X, Y satisfy a suitable condition (e.g., X, Y are CW complexes), then C_i has the same homotopy type as the reduced suspension SX of X; i'^* is equivalent to $p^*: \pi(SX; Z)_0 \rightarrow \pi(C_f; Z)_0$ induced by a mapping $p: C_f \rightarrow SX$ smashing Y to a point; and furthermore C_p has the same homotopy type as SY, and the inclusion $i_0: SX \rightarrow C_p$ is equivalent to the suspension $Sf: SX \rightarrow SY$ of f. Thus the following **Puppe exact sequence** is obtained:

$$\ldots \xrightarrow{Sp^*} \pi(SC_f; Z)_0 \xrightarrow{Si^*} \pi(SY; Z)_0 \xrightarrow{Sf^*} \pi(SX; Z)_0$$
$$\xrightarrow{p^*} \pi(C_f; Z)_0 \xrightarrow{i^*} \pi(Y; Z)_0 \xrightarrow{f^*} \pi(X; Z)_0.$$

In this exact sequence, if Y is a CW complex, X is a subcomplex of Y, and f is the inclusion $i: X \rightarrow Y$, then $C_i = C_f = Y \cup C_X$ is homotopy equivalent to the space $C_i/CX = Y/X$ obtained by smashing CX to a point, and an exact sequence of the following type is obtained:

$$\ldots \xrightarrow{Si^*} \pi(SX; Z)_0 \xrightarrow{\delta^*} \pi(Y/X; Z)_0 \xrightarrow{p^*} \pi(Y; Z)_0$$
$$\xrightarrow{i^*} \pi(X; Z)_0.$$

A sequence equivalent to the sequence $X \xrightarrow{f} Y \xrightarrow{i} C_f$ is called a **cofibering**, for which a similar exact sequence is obtained. For a continuous mapping $f: X \rightarrow Y$, consider the subspace $E_f = \{(x, \varphi) \mid f(x) = \varphi(0)\}$ of the product space $X \times Y^I$. By identifying X with $\{(x, \varphi_x) \mid \varphi_x(I) = f(x)\}$, we can regard X as a deformation retract of E_f. By putting $p_1(x, \varphi) = \varphi(1)$, we obtain a †fiber space (E_f, p_1, Y). The fiber $T_f = p_1^{-1}(*)$ is called a **mapping track** of f. Using the †covering homotopy property, we see that the sequence

$$\pi(W; T_f)_0 \xrightarrow{p_*} \pi(W; X)_0 \xrightarrow{f_*} \pi(W; Y)_0$$

is exact, where $p(x, \varphi) = x$. This sequence is also extended infinitely to the left as

$$\ldots \rightarrow \pi(W; \Omega X)_0 \xrightarrow{(\Omega f)_*} \pi(W; \Omega Y)_0 \xrightarrow{i_*} (W; T_f)_0 \xrightarrow{p_*},$$

where i is the inclusion of the loop space ΩY into T_f and $\Omega f: \Omega X \rightarrow \Omega Y$ is the correspondence of the loops induced from f.

H. Homotopy Sets that Form Groups

If $X = SX'$ or $Y = \Omega Y'$ (or, generally, if Y is a †homotopy associative †H-space having a †homotopy inverse), then $\pi(X; Y)_0$ forms a group. In the general case the product of the

loops induces the product of $\pi(X;\Omega Y')_0$. We represent a point of SX by (x,t) $(x\in X, t\in I)$ and define the mapping $\Omega_0 g: X\to\Omega Y$ for each $g: SX\to Y$ by $\Omega_0 g(x)(t)=g(x,t)$. Hence an isomorphism $\Omega_0:\pi(SX;Y)_0\cong\pi(X;\Omega Y)_0$ is obtained. Each of the following pairs of homomorphisms is equivalent: $f_*:\pi(SX;Y)_0\to\pi(SX;Y')_0$ and $\Omega f_*:\pi(X;\Omega Y)_0\to\pi(X;\Omega Y')_0$; and $h^*:\pi(X';\Omega Y)_0\to\pi(X;\Omega Y)_0$ and $Sh^*:\pi(SX';Y)_0\to\pi(SX;Y)_0$.

If S^n is an n-dimensional sphere, then $\pi_n(X)=\pi(S^n;X)_0$ is the n-dimensional homotopy group (→ Section J). Let $\pi^n(X)=\pi(X;S^n)_0$. If X is a CW complex of dimension less than $2n-1$, $\pi^n(X)$ is the cohomotopy group isomorphic to $\pi(X;\Omega S^{n+1})_0$ (→ Section I). Let K_n be an †Eilenberg-MacLane space of type (Π,n). Then we have $K_n=\Omega K_{n+1}$, and if (X,A) is a pair of CW complexes, then $\pi(X/A;K_n)_0$ coincides with the cohomology group $H^n(X,A;\Pi)$. For the †classifying space $B_{\mathrm{O}}(B_{\mathrm{U}})$ of the infinite orthogonal group $\mathbf{O}$ (infinite unitary group $\mathbf{U}$) (→ Section V), $\pi(X/A;B_{\mathrm{O}})$ $(\pi(X/A;B_{\mathrm{U}}))$ may be considered the KO-group $KO(X,A)$ (K-group $K(X,A)$) (→ 237 K-Theory).

I. Cohomotopy Groups

K. Borsuk defined a sum of mapping classes of X into S^n (1936), which was named Borsuk's **cohomotopy group** by E. Spanier. Spanier also studied the duality of the cohomotopy group with the homotopy group and its relations to the usual cohomology groups. A cohomotopy group of (X,A) is defined to be $\pi^n(X,A)=\pi(X,A;S^n,*)$, which forms a group if $\dim X/A<2n-1$. A mapping $F:X/A\to S^n\times S^n$ given by $F(x)=(f(x),g(x))$ with $f,g:X/A\to S^n$ is homotopic to a mapping into $S^n\vee S^n$. If we compose F with a folding mapping of $S^n\vee S^n$ onto S^n, we obtain a mapping that represents the sum $[f]+[g]$. With each homotopy class of a continuous mapping f of an n-dimensional †polyhedron K^n into an n-dimensional sphere S^n, we associate the image $f^*(u)$ of the fundamental class $u\in H^n(S^n;\mathbf{Z})$ under the induced homomorphism $f^*: H^n(S^n;\mathbf{Z})\to H^n(K^n;\mathbf{Z})$. We then obtain a bijective relation $\pi^n(K^n)\to H^n(K^n;\mathbf{Z})$, called **Hopf's classification theorem**.

J. Homotopy Groups

Let X be a topological space with a base point $*$, $I^n=\{t=(t_1,t_2,\dots,t_n)\mid 0\leqslant t_1,t_2,\dots,t_n\leqslant 1\}$ be the unit n-cube, and $\dot{I}^n$ its boundary. Write $\Omega^n(X,*)=X^{I^n}(\dot{I}^n,*)$ (in particular, $\Omega^1(X,*)$ is the loop space), and denote by $\pi_n(X,*)$ or simply $\pi_n(X)$ the set of arcwise connected components of $\Omega^n(X,*)$, i.e., the homotopy classes $[f]$. Using the notation of homotopy sets, we have $\pi_n(X,*)=\pi(I^n,\dot{I}^n;X,*)$. If we choose the constant mapping as the base point $*$ of $\Omega^n(X,*)$, then $\Omega^m(\Omega^n(X,*),*)=\Omega^{m+n}(X,*)$. Thus $\pi_m(\Omega^n(X,*),*)=\pi_{m+n}(X,*)$. Since π_1 is the †fundamental group, $\pi_n(X,*)=\pi_1(\Omega^{n-1}(X),*)$ is also a group, called the n-dimensional **homotopy group** of X with base point $*$. "Multiplication" in homotopy groups is defined as follows: Given $f_1,f_2\in\Omega^n(X,*)$ we define $f_1+f_2\in\Omega^n(X,*)$ by

$$(f_1+f_2)(t)=\begin{cases} f_1(2t_1,t_2,\dots,t_n), & 0\leqslant t_1\leqslant\frac{1}{2},\\ f_2(2t_1-1,t_2,\dots,t_n), & \frac{1}{2}\leqslant t_1\leqslant 1\end{cases}$$

(Fig. 1). Then the product or sum of $[f_1]$ and $[f_2]$ is given by $[f_1+f_2]$. The identity is the class of the constant mapping (denoted by 0), and the inverse of $[f]$ is $[\bar{f}]$, represented by $\bar{f}(t)=f(1-t_1,t_2,\dots,t_n)$. The space $\Omega^n(X,*)$ is an †H-space, where multiplication is given by the correspondence $(f_1,f_2)\to f_1+f_2$. Since the fundamental group of an H-space is commutative, $\pi_n(X,*)$ is an Abelian group for $n\geqslant 2$.

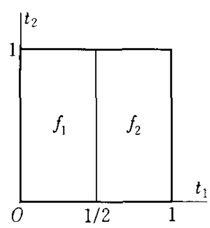


Fig. 1

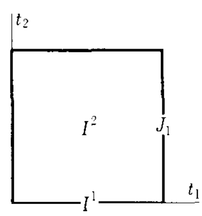


Fig. 2

Let $S^n=\{t=(t_1,\dots,t_{n+1})\mid\sum t_i^2=1\}$ be the n-sphere, and take $*=(1,0,\dots,0)$ as its base point. Suppose that we are given a continuous mapping $\psi_n:(I^n,\dot{I}^n)\to(S^n,*)$ such that $\psi_n:I^n-\dot{I}^n\to S^n-*$ is homeomorphic. Then the correspondence $\psi_n^*:\pi(S^n;X)_0\to\pi_n(X,*)$ determined by $\psi_n^*[g]=[g\circ\psi_n]$ is bijective. Thus we can identify the homotopy group $\pi_n(X,*)$ with $\pi(S^n;X)_0$.

K. Relative Homotopy Groups

Suppose that we are given a topological space X and a subspace A of X sharing the same base point $*$. Identify I^{n-1} with the face $t_n=0$ of I^n, and let J^{n-1} be the closure of $\dot{I}^n-I^{n-1}$ (Fig. 2). Denote by $\pi_n(X,A,*)$ the set of homotopy classes of continuous mappings $f:(I^n,\dot{I}^n,J^{n-1})\to(X,A,*)$. Let $\Omega^n(X,A,*)$ be the mapping space consisting of such mappings f, and let $\pi_n(X,A,*)=\pi_0(\Omega^n(X,A,*))$. Since $\Omega^m(\Omega^n(X,A,*),*)$ is homeomorphic to $\Omega^{m+n}(X,A,*)$, we have $\pi_m(\Omega^n(X,A,*),*)\cong\pi_{m+n}(X,A,*)$. Thus $\pi_n(X,A,*)$ is a group for $n\geqslant 2$ and an Abelian group for $n\geqslant 3$. This group is called the n-dimensional **relative**

homotopy group of (X, A) with respect to the base point $*$, or simply the n-dimensional homotopy group of (X, A). In the same manner as in Section J multiplication in this group can be defined using f_1+f_2. Since $\Omega^n(X,*,*)$ and $\Omega^n(X,*)$ are identical, we have $\pi_n(X,*,*)=\pi_n(X,*)$. Hence homotopy groups are special cases of relative homotopy groups.

Let $g:(X,A,*)\to(Y,B,*)$ be a continuous mapping. Then a correspondence g_*: $\pi_n(X,A,*)\to\pi_n(Y,B,*)$ is obtained by $g_*[f]=[g\circ f]$, with g_* a homomorphism of homotopy groups for $n\geqslant 2$ and for $n=1$, $A=*$. We call g_* the **homomorphism induced by** g. Let $E^n=\{t=(t_1,\ldots,t_n)\mid\sum t_i^2=1\}$ be the unit n-cell with boundary S^{n-1}. Utilizing a suitable relative homeomorphism $\psi_n':(I^n,J^{n-1})\to(E^n,*)$, $\psi_n'(\dot{I}^n)=S^{n-1}$, we obtain a one-to-one correspondence $\psi_n'^*:\pi(E^n,S^{n-1};X,A)_0\to\pi_n(X,A,*)$, and $\Omega^n(X,A,*)$ is homeomorphic (via ψ_n') to the mapping space $X^{E^n}(S^{n-1},*;A,*)$.

L. Homotopy Exact Sequences

Given an element $\alpha=[f]\in\pi_n(X,A,*)$, and letting $\partial\alpha=[f\mid I^{n-1}]\in\pi_{n-1}(A,*)$, we obtain a homomorphism $(n\geqslant 2)$ $\partial:\pi_n(X,A,*)\to\pi_{n-1}(A,*)$, which is called the **boundary homomorphism**. Furthermore, we have the following exact sequence involving homomorphisms i_*, j_* induced by two inclusions $i:(A,*)\to(X,*)$, $j:(X,*,*)\to(X,A,*)$:

$$\ldots\xrightarrow{\partial}\pi_n(A,*)\xrightarrow{i_*}\pi_n(X,*)\xrightarrow{j_*}\pi_n(X,A,*)$$
$$\xrightarrow{\partial}\ldots\to\pi_1(X,*)\xrightarrow{j_*}\pi_1(X,A,*)\xrightarrow{\partial}\pi_0(A)\xrightarrow{i_*}\pi_0(X).$$

This sequence is called the **homotopy exact sequence** of the pair (X, A). A system of topological spaces $X\supset A\supset B\ni *$ is called a **triple**. In this homotopy exact sequence, if we replace $(A,*)$, $(X,*)$ by $(A,B,*)$, $(X,B,*)$, respectively, we obtain an exact sequence, called the **homotopy exact sequence of the triple** (X,A,B).

The homotopy group $\pi_n(A\times B)$ of the product space is isomorphic to the direct sum $\pi_n(A)+\pi_n(B)$, and the projections $p(p'):A\times B\to A(B)$ of the product space induce the projections from $\pi_n(A\times B)$ onto the direct summands $\pi_n(A)$, $\pi_n(B)$. This is a special case of the †Hurewicz-Steenrod isomorphism theorem in fiber spaces (→ 148 Fiber Spaces). Setting $A\vee B=(A\times *)\cup(*\times B)$, we obtain a direct sum decomposition $\pi_n(A\vee B)\cong\pi_n(A)+\pi_n(B)+\pi_{n+1}(A\times B,A\vee B)$. Next we consider a fixed pair (X, A) and move the base point $*$ to investigate its effect on the elements of the homotopy group. Suppose that we are given a path $h:I\to A$ with terminal point $*=h(1)$ and an element $\alpha\in\pi_n(X,A,*)$ $(\alpha=[f], f:(I^n,\dot{I}^n,J^{n-1})\to(X,A,*))$. By the homotopy extension property, we can construct a homotopy $f_\theta:(I^n,\dot{I}^n)\to(X,A)$ satisfying $f_\theta(J^{n-1})=h(\theta)$ and $f_1=f$. Then the homotopy class $[f_0]$ of f_0 with respect to the base point $*'=h(0)$ is determined only by α and the homotopy class ω of the path h. We denote the homotopy class $[f_0]$ by $\alpha^\omega\in\pi_n(X,A,*')$. The correspondence $\alpha\to\alpha^\omega$ is a group isomorphism, and $(\alpha^\omega)^{\omega'}=\alpha^{\omega\omega'}$. Thus if A is arcwise connected, $\pi_n(X,A,*)$ is isomorphic to $\pi_n(X,A,*')$. Hence, in this case, we may simply write $\pi_n(X,A)$ instead of $\pi_n(X,A,*)$. When $*=*'$, the correspondence $\alpha\to\alpha^\omega$ determines the action of the group $\pi_1(A,*)$ on $\pi_n(X,A,*)$. Given an element $\alpha\in\pi_n(X,*)$ and a class ω of paths in X, we define $\alpha^\omega\in\pi_n(X,*')$ as for relative homotopy. Specifically, if $\omega\in\pi_1(X,*)$, then $\alpha^\omega-\alpha$ coincides with the Whitehead product $[\omega,\alpha]$ (when $n=1$, we have $\alpha^\omega\cdot\alpha^{-1}=[\omega,\alpha]=\omega\alpha\omega^{-1}\alpha^{-1}$) (→ Section P).

A pair (X, A) consisting of a topological space X and an arcwise connected subspace A of X is said to be **n-simple** if the operation of $\pi_1(A)$ on $\pi_n(X,A)$ is trivial. Similarly, an arcwise connected space X is called **n-simple** if the operation of $\pi_1(X)$ on $\pi_n(X)$ is trivial. For example, a pair (X, A) consisting of an H-space X and an H-subspace A is **simple**, i.e., n-simple for each n. If a topological space X satisfies $\pi_i(X)=0$ $(0\leqslant i\leqslant n)$, then X is said to be **n-connected**. 0-connectedness coincides with arcwise connectedness and 1-connectedness means †simple connectedness. S^n is $(n-1)$-connected. A pair (X, A) is said to be **n-connected** if $\pi_0(A)=\pi_0(X)=\pi_i(X,A)=0$ $(1\leqslant i\leqslant n)$, and (E^n,S^{n-1}) is $(n-1)$-connected.

M. Homotopy Groups of Triads

Let $(X;A,B,*)$ be a system, called a **triad**, of a topological space X and its subspaces A, B satisfying $A\cap B\ni *$ (base point). Let $\pi_n(X;A,B,*)=\pi_{n-1}(\Omega^1(X,B),\Omega^1(A,A\cap B),*)$ $(n\geqslant 2)$; $\pi_n(X;A,B,*)$ is a group for $n\geqslant 3$ and an Abelian group for $n\geqslant 4$. We call $\pi_n(X;A,B,*)$ the **homotopy group of the triad**. From the homotopy exact sequence of the pair, we obtain the following **homotopy exact sequence of the triad**:

$$\ldots\xrightarrow{\partial}\pi_j(A,A\cap B,*)\xrightarrow{i_*}\pi_j(X,B,*)$$
$$\xrightarrow{j_*}\pi_j(X;A,B,*)\xrightarrow{\partial}\pi_{j-1}(A,A\cap B,*)\xrightarrow{i_*}\ldots.$$

Assume for simplicity that $A\cap B$ is simply connected, $X=\operatorname{Int}A\cup\operatorname{Int}B$ ($\operatorname{Int}A$ is the †interior of A), $(A,A\cap B)$ is m-connected, and $(B,A\cap B)$ is n-connected. Then $(X;A,B)$ is $(m+n)$-connected, i.e., $\pi_j(X;A,B,*)=0$ $(2\leqslant j\leqslant m+n)$ **(Blakers-Massey theorem)**.

Furthermore, in this case we have a replica of the †excision isomorphism in homology theory for $j<m+n$; that is, we have the isomorphism $i_*:\pi_j(A,A\cap B,*)\cong\pi_j(X,B,*)$ in-

duced by the inclusion $i:(A, A\cap B)\to(X, B)$. On the other hand, $\pi_{m+n+1}(X; A, B, *)$ is isomorphic to $\pi_{m+1}(A, A\cap B, *)\otimes\pi_{n+1}(B, A\cap B, *)$. This shows that the excision isomorphism does not always hold for homotopy groups, an important difference from homology theory. However, if we replace the excision axiom by the Hurewicz-Steenrod isomorphism theorem, which is valid for fiber spaces (→ 148 Fiber Spaces), then we can construct homotopy theory axiomatically in the same manner as homology theory (→ 201 Homology Theory).

N. The Hurewicz Isomorphism Theorem

The **Hurewicz homomorphism** τ of $\pi_n(X, A)$ into the n-dimensional integral homology group $H_n(X, A)$ is defined by $\tau([f])=f_*(\varepsilon_n)$ (where ε_n is a generator of $H_n(I^n, \dot{I}^n)$). Then we have the **Hurewicz isomorphism theorem**: Suppose that the pair (X, A) is n-simple (e.g., $A = *$) and $(n-1)$-connected. Then we have $H_i(X, A)=0$ $(i<n)$ and the isomorphism τ: $\pi_n(X, A)\cong H_n(X, A)$ (for $n=1$ → 170 Fundamental Groups). Let X, Y be simply connected topological spaces, and let $f: X\to Y$ be a continuous mapping. Then the following two conditions are equivalent: (1) $f_*:\pi_i(X)\to\pi_i(Y)$ is injective for $i<n$ and surjective for $i\leqslant n$. (2) $f_*: H_i(X)\to H_i(Y)$ is injective for $i<n$ and surjective for $i\leqslant n$ (**J. H. C. Whitehead's theorem**).

J.-P. Serre generalized these theorems as follows: A family $\mathscr{C}$ of Abelian groups satisfying condition (i) is called a **class of Abelian groups**: (i) If a sequence $F\to G\to H$ of Abelian groups is exact and F, $H\in\mathscr{C}$, then $G\in\mathscr{C}$. Furthermore, we consider the following conditions: (ii) The tensor product $G\otimes F$ of an arbitrary Abelian group F with an element $G\in\mathscr{C}$ also belongs to $\mathscr{C}$. (ii') If both F, $G\in\mathscr{C}$, then $F\otimes G$, $\mathrm{Tor}(F, G)\in\mathscr{C}$. (iii) If $G\in\mathscr{C}$, then its †homology group $H_i(G)\in\mathscr{C}$ $(i>0)$. Condition (ii') is implied by (ii). A homomorphism f: $F\to G$ is called $\mathscr{C}$-injective if $\mathrm{Ker}\, f\in\mathscr{C}$, $\mathscr{C}$-surjective if $\mathrm{Coker}\, f= G/\mathrm{Im}\, f\in\mathscr{C}$, and a $\mathscr{C}$-isomorphism if f is $\mathscr{C}$-injective and $\mathscr{C}$-surjective. Two Abelian groups G and G' are called $\mathscr{C}$-isomorphic if there exist $\mathscr{C}$-isomorphisms $f: F\to G$ and $f': F\to G'$. In particular, if the class $\mathscr{C}_0$ consists of only the trivial group 0, then concepts such as $\mathscr{C}_0$-isomorphism coincide with the usual concepts of isomorphism, and so on. Let $\mathscr{C}_p$ be the class of finite Abelian groups whose orders are relatively prime to a fixed prime number p. Here, instead of the terms $\mathscr{C}_p$-isomorphism and so on, we use the terms **mod p isomorphism** and so on. Let $\mathscr{D}$ be a class of finitely generated Abelian groups. Then $\mathscr{C}_p$ satisfies conditions (ii) and (iii), and $\mathscr{D}$ satisfies (ii') and (iii).

We have the following **generalized Hurewicz theorem**: (A) Suppose that a class $\mathscr{C}$ satisfies (ii) and (iii) and we are given a 2-connected pair (X, A) of simply connected spaces X, A. If $\pi_i(X, A)\in\mathscr{C}$ $(i<n)$, then $H_i(X, A)\in\mathscr{C}$, and $\tau:\pi_n(X, A)\to H_n(X, A)$ is a $\mathscr{C}$-isomorphism. (B) Suppose that $A = *$, $\mathscr{C}$ satisfies conditions (ii') and (iii), and X is simply connected. Then an assertion similar to (A) holds. In particular, a simply connected space X having finitely generated homology groups (e.g., a simply connected finite polyhedron) has finitely generated homotopy groups. As a corollary to theorem (A), we obtain a **generalized Whitehead theorem**. In particular, applying the theorem to the class $\mathscr{D}\cap\mathscr{C}_p$, we obtain the following frequently used theorem: Suppose that we are given simply connected spaces X, Y whose homology groups are finitely generated and $f: X\to Y$ satisfies $f_*\pi_2(X)=\pi_2(Y)$. Then the following two conditions are equivalent: (1) $f_*:\pi_i(X)\to\pi_i(Y)$ is a mod p isomorphism for $i<n$ and a mod p surjection for $i=n$. (2) $f_*: H_i(X, \mathbf{Z}_p)\to H_i(Y, \mathbf{Z}_p)$ is an isomorphism for $i<n$ and a surjection for $i=n$ (where $\mathbf{Z}_p=\mathbf{Z}/p\mathbf{Z}$). The theory above, which makes use of the notion of class $\mathscr{C}$, is an example of **Serre's $\mathscr{C}$-theory**. Concepts such as †spectral sequences for fiber spaces and †n-connective fiber spaces are important tools in Serre's $\mathscr{C}$-theory (→ 148 Fiber Spaces).

To calculate homotopy groups, we use notions such as exact sequences, fiber spaces, (co)homology groups of n-connective fiber spaces, and †Postnikov systems. Given an arbitrary group (more generally, a Postnikov system), there exists a CW complex having the given group (system) as its homotopy group (Postnikov system) (**realization theorem of homotopy groups**). For an arbitrary arcwise connected topological space X there exist topological spaces (X, n) and continuous mappings $p_n:(X, n+1)\to(X, n)$ $(n=1, 2, \dots)$ satisfying the following two conditions: (i) $((X, n+1), p_n, (X, n))$ is a fiber space whose fiber is an †Eilenberg-MacLane space. (ii) $(X, 1)=X$, and $((X, n+1), p_1\circ\dots\circ p_n, X)$ is an n-connective fiber space. The method of obtaining the homotopy group $\pi_n(X)\cong H_n((X, n))$ by computing (co)homology groups of (X, n) is called a **killing method**.

O. Homotopy Operations

Let X, Y, X', Y' be topological spaces. If to each continuous mapping $f\in Y^X$ there corresponds a homotopy class $\Phi(f)\in\pi(X'; Y')$ that is a homotopy invariant of f (satisfying a certain naturality condition), then Φ is called a **homotopy operation**. More generally, we

may consider the case where Φ is a mapping from $\pi(X_1; Y_1)\times\ldots\times\pi(X_r; Y_r)$ into $\pi(X'; Y')$. The **naturality** of Φ is defined as follows: Consider the †category $\mathscr{C}$ of topological spaces (or its subcategory). Let $Y=Y'$ be an arbitrary †object of $\mathscr{C}$, and fix X and X'. In this case, the naturality of $\Phi_Y:\pi(X; Y)\rightarrow\pi(X'; Y)$ is defined to be the commutativity of the diagram:

$$\begin{array}{ccc}\pi(X; Y) & \xrightarrow{\Phi_Y} & \pi(X'; Y)\\ \downarrow g_* & & \downarrow g_*\\ \pi(X; Z) & \xrightarrow{\Phi_Z} & \pi(X'; Z)\end{array}$$

i.e., $g_*\circ\Phi_Y=\Phi_Z\circ g_*$ for an arbitrary †morphism (i.e., continuous mapping) $g: Y\rightarrow Z$ of the category $\mathscr{C}$. Similarly, when objects Y, Y' of the category $\mathscr{C}$ are fixed and $X=X'$ is an arbitrary object of $\mathscr{C}$, to say that a homotopy operation $\Phi_X:\pi(X; Y)\rightarrow\pi(X; Y')$ is **natural** means that $h^*\circ\Phi_X=\Phi_W\circ h^*$ for an arbitrary morphism $h: W\rightarrow X$.

We have the following theorem: In the category of topological spaces and continuous mappings, the homotopy operations Φ_Y: $\pi(X; Y)\rightarrow\pi(X'; Y)$ and the elements of $\pi(X'; X)$ are in one-to-one correspondence. The correspondence is obtained by associating a homotopy operation $\Phi(\beta)=\beta\circ\alpha$ ($\beta\in\pi(X; Y)$) with each $\alpha\in\pi(X'; X)$. Similarly, the homotopy operations $\Phi_X:\pi(X; Y)\rightarrow\pi(X; Y')$ and the elements of $\pi(Y; Y')$ are in one-to-one correspondence. This theorem holds also for the case involving several variables if we consider $\pi(X'; X_1\vee X_2\vee\ldots)$ or $\pi(Y_1\times Y_2\times\ldots; Y')$ instead of $\pi(X'; Y)$ or $\pi(X; Y')$. The theorem remains valid if we replace the spaces X, Y by systems of spaces.

P. Homotopy Operations in Homotopy Groups

(1) If X, X' are spheres S^n, S^p with base points and Y, Y' are topological spaces with base points, a homotopy operation $\Phi_Y:\pi_n(Y)\rightarrow\pi_p(Y)$ is said to be of type (n,p). By the theorem in Section O, the homotopy operations of type (n,p) are in one-to-one correspondence with the elements of the homotopy group of the sphere $\pi_p(S^n)$.

(2) As an example of the 2-variable homotopy operations $\Phi_Y:\pi_m(Y)\times\pi_n(Y)\rightarrow\pi_p(Y)$ of type $(m,n;p)$ we have the Whitehead product defined as follows: Suppose that $\alpha\in\pi_m(Y)$, $\beta\in\pi_n(Y)$ are elements represented by $f:(I^m,\dot{I}^m)\rightarrow(Y,*)$ and $g:(I^n,\dot{I}^n)\rightarrow(Y,*)$, respectively. Define a continuous mapping F from the boundary $\dot{I}^{m+n}=(I^m\times\dot{I}^n)\cup(\dot{I}^m\times I^n)$ of $I^{m+n}=I^m\times I^n$ into Y by $F(x,y)=f(x)$ for $(x,y)\in I^m\times\dot{I}^n$ and $F(x,y)=g(y)$ for $(x,y)\in\dot{I}^m\times I^n$. Since $\dot{I}^{m+n}$ is homeomorphic to S^{m+n-1}, we can identify them. The homotopy class represented by F is an element of $\pi_{m+n-1}(Y)$ determined by α and β, denoted by $[\alpha,\beta]$ and called the **Whitehead product** of α and β (J. H. C. Whitehead, *Ann. Math.*, (2) 42 (1941)). The Whitehead product is a homotopy operation of type $(m,n;m+n-1)$. Let $\psi_m:(I^m,\dot{I}^m)\rightarrow(S^m,*)$ be a mapping that smashes $\dot{I}^m$ to a point. The product of ψ_m and ψ_n defines a mapping $\psi_{m,n}$: $S^{m+n-1}\rightarrow S^m\vee S^n=(S^m\times *)\cup(*\times S^n)$. Let $\iota\in\pi_m(S^m\vee S^n)$, $\iota'\in\pi_n(S^m\vee S^n)$ be the homotopy classes of the natural inclusions of S^m, S^n into $S^m\vee S^n$; then the homotopy class of $\psi_{m,n}$ is $[\iota,\iota']$. G. W. Whitehead showed that a direct sum decomposition $\pi_p(S^m\vee S^n)=\iota_*\pi_p(S^m)+\iota'_*\pi_p(S^n)+[\iota,\iota']_*\pi_p(S^{m+n-1})$ ($\iota_*,\iota'_*,[\iota,\iota']_*$ are injective) holds for $1<p<m+n+\min(m,n)-3$. Furthermore, P. J. Hilton showed that for general $p>1$, $\pi_p(S^m\vee S^n)$ is the direct sum of the images of injections ι_*, ι'_*, $[\iota,\iota']_*$, $[[\iota,\iota'],\iota]_*$, $[[\iota,\iota'],\iota']_*$, etc. The homotopy operations of type $(m,n;p)$ are in one-to-one correspondence with the elements of $\pi_p(S^m\vee S^n)$; hence such operations can be constructed by means of composition and the Whitehead product. The last proposition is also valid for homotopy operations of type $(m_1,\ldots,m_r;p)$. The Whitehead product $[a,\beta]$ ($\alpha\in\pi_m(X)$, $\beta\in\pi_n(X)$) is distributive with respect to α (resp. β) for $m>1$ ($n>1$), and we have $[\beta,\alpha]=(-1)^{mn}[\alpha,\beta]$ and $f_*[\alpha,\beta]=[f_*\alpha,f_*\beta]$ for $f:X\rightarrow Y$. Moreover, for $\gamma\in\pi_r(X)$ the **Jacobi identity** holds: $(-1)^{mr}[[\alpha,\beta],\gamma]+(-1)^{mn}[[\beta,\gamma],\alpha]+(-1)^{rn}[[\gamma,\alpha],\beta]=0$ (M. Nakaoka and H. Toda; H. Uehara and W. S. Massey; Hilton).

Q. Suspensions and Generalized Hopf Invariants

We denote by $\alpha\wedge\beta\in\pi(X\wedge X'; Y\wedge Y')_0$ the class of the reduced join of f, g, where f represents $\alpha\in\pi(X; Y)_0$ and g represents $\beta\in\pi(X'; Y')_0$. We call $\alpha\wedge\beta$ the **reduced join** of α and β. In particular, if $Y=Y'=S^1$, β is the identity mapping of S^1, and α is represented by f, then $\alpha\wedge\beta$ is called the **suspension** of α and is denoted by $S\alpha$. $S\alpha$ is the class of the suspension Sf of f and belongs to $\pi(SX; SY)_0$, where SX indicates the reduced suspension of X. The suspension $S\alpha$ is often denoted by $E\alpha$ in reference to the German term *Einhängung*. The identity mapping 1 of SY gives rise to an injection $i=\Omega_0 1$ sending Y into the loop space $\Omega(SY)$ determined by the formula $i(y)(t)=(y,t)$. Then we have

$$i_*=\Omega_0\circ S:\pi(X; Y)_0\rightarrow\pi(SX; SY)_0 \cong\pi(X;\Omega SY)_0,$$

and S and i_* are equivalent. Let Y_k be the identifying space $Y^k/\sim$, where Y^k is the product space $Y\times\ldots\times Y$ of k copies of Y and $\sim$

is the equivalence relation determined by

$$(*, y_1, y_2, \ldots, y_{k-1}) \sim (y_1, *, y_2, \ldots, y_{k-1}) \sim \cdots \sim (y_1, \ldots, y_{k-1}, *).$$

Denote by $Y_\infty = \bigcup_k Y_k$ the limit space with respect to the injection $Y_{k-1} \rightarrow Y_k$ given by $(y_1, \ldots, y_{k-1}) \rightarrow (y_1, \ldots, y_{k-1}, *)$ and call it the **reduced product space** of Y. Let Y be a CW complex of 0-†section $*$. The mapping $i: Y = Y_1 \rightarrow \Omega S Y$ can then be extended to $\bar{i}: Y_\infty \rightarrow \Omega S Y$, where $\bar{i}$ is a weak homotopy equivalence. If X is also a CW complex, then $\Omega_0^{-1} \circ \bar{i}_*: \pi(X; Y_\infty)_0 \rightarrow \pi(SX; SY)_0$ is bijective. By smashing the subset Y of Y_2, we have $Y \wedge Y = Y_2/Y$. This smashing mapping can be extended to $h: Y_\infty \rightarrow (Y \wedge Y)_\infty$ (I. M. James). Utilizing $h_*: \pi(X; Y_\infty)_0 \rightarrow \pi(X; (Y \wedge Y)_\infty)_0$ and the bijection $\Omega_0^{-1} \circ \bar{i}_*$, we obtain a correspondence $H: \pi(SX; SY)_0 \rightarrow \pi(SX; S(Y \wedge Y))_0$. We call $H(\alpha)$ the **generalized Hopf invariant** of α. When $X = S^{2n-2}$, $Y = S^{n-1}$, H is equivalent to the Hopf invariant $\gamma: \pi_{2n-1}(S^n) \rightarrow \mathbf{Z}$ (→ Section U). In general, we have $H \circ S = 0$, and the exactness of $\xrightarrow{S} \xrightarrow{H}$ holds under various conditions. Denote also by $\circ$ the composition of homotopy classes; then we have $S(\alpha \circ \beta) = S\alpha \circ S\beta$ and $H(\alpha \circ S\beta) = H\alpha \circ S\beta$. Also, $H(S\alpha \circ \beta) = S(\alpha \wedge \alpha) \circ H\beta$. Under the condition $i < 3n-3$, we have $(\alpha_1 + \alpha_2) \circ \beta = \alpha_1 \circ \beta + \alpha_2 \circ \beta + [\alpha_1, \alpha_2] \circ H(\beta)$ for $\alpha_1, \alpha_2 \in \pi_n(X)$ and $\beta \in \pi_i(S^n)$ (G. W. Whitehead). Thus the composition $\alpha \circ \beta$ is not always left distributive but is always right distributive, and $\alpha \circ \beta$ is left distributive if $\beta = S\beta'$. The composition is defined over the stable homotopy groups G_r of spheres (→ Section U): $\alpha \circ \beta \in G_{p+q}$ $(\alpha \in G_p, \beta \in G_q)$. It is distributive and satisfies $\beta \circ \alpha = (-1)^{pq}\, \alpha \circ \beta$.

When Y and Y' are Eilenberg-MacLane spaces, B_{O}, and B_{U}, we have †cohomology operations on cohomology groups $H^n(\ ;\Pi)$, KO groups, and K groups, respectively. As typical examples there are †Steenrod square operations $Sq^i: H^n(X; \mathbf{Z}_2) \rightarrow H^{n+i}(X; \mathbf{Z}_2)$, †Steenrod pth power operations $\mathscr{P}^i: H^n(X; \mathbf{Z}_p) \rightarrow H^{n+2i(p-1)}(X; \mathbf{Z}_p)$, †Chern characters $ch^n: K(X) \rightarrow H^{2n}(X; \mathbf{Q})$ ($\mathbf{Q}$: rational field), †Adams operations $\Phi_i: KO(X) \rightarrow KO(X)$ $(K(X) \rightarrow K(X))$. They are all homomorphisms (→ 64 Cohomology Operations; 237 K-Theory).

R. Secondary Compositions

Suppose that $\alpha \circ \beta = 0$, $\beta \circ \gamma = 0$ for $\gamma \in \pi(W; X)_0$, $\beta \in \pi(X; Y)_0$, $\alpha \in \pi(Y; Z)_0$. In the commutative diagram of Puppe exact sequences

$$\begin{array}{ccccccc}
\xrightarrow{S\gamma^*} \pi(SW; Y)_0 & \xrightarrow{p^*} & \pi(C_\gamma; Y)_0 & \xrightarrow{i^*} & \pi(X; Y)_0 & \xrightarrow{\gamma^*} \\
\downarrow \alpha_* & & \downarrow \alpha_* & & \downarrow \alpha_* & \\
\xrightarrow{S\gamma^*} \pi(SW; Z)_0 & \xrightarrow{p^*} & \pi(C_\gamma; Z)_0 & \xrightarrow{i^*} & \pi(X; Z)_0 & \xrightarrow{\gamma^*}
\end{array}$$

the set of elements $\bar{\beta}$ in $\pi(SW; Z)$ such that $p^*(\bar{\beta}) \in \alpha_* i^{*-1}(\beta)$ is denoted by $\{\alpha, \beta, \gamma\}$ and is called a **secondary composition** or **Toda bracket**. If θ, η are elements of $\pi(SW; Y)_0$, $\pi(SX; Z)_0$, respectively, then we have $\{\alpha, \beta, \gamma\} + \alpha_*\theta = \{\alpha, \beta, \gamma\}$, $\{\alpha, \beta, \gamma\} + S\gamma^*\eta = \{\alpha, \beta, \gamma\}$. Hence we may consider the set $\{\alpha, \beta, \gamma\}$ to be a residue class modulo a submodule generated by $\alpha_*\pi(SW; Y)_0$ and $S\gamma^*\pi(SX; Z)_0$.

The secondary composition $\{\alpha, \beta, \gamma\}$ has the following properties: (i) $\{\alpha, \beta, \gamma\}$ is linear with respect to α, β, γ (if the sum is defined); (ii) $\alpha \circ \{\beta, \gamma, \delta\} = \{\alpha, \beta, \gamma\} \circ (-S\delta)$; (iii) $S\{\alpha, \beta, \gamma\} \equiv -\{S\alpha, S\beta, S\gamma\}$; (iv) $\alpha \circ \{\beta, \gamma, \delta\} \equiv \{\alpha \circ \beta, \gamma, \delta\}$, $\{\alpha \circ \beta, \gamma, \delta\} \equiv \{\alpha, \beta \circ \gamma, \delta\}, \ldots$; (v) $\{\{\alpha, \beta, \gamma\}, S\delta, S\varepsilon\} + \{\alpha, \{\beta, \gamma, \delta\}, S\varepsilon\} + \{\alpha, \beta, \{\gamma, \delta, \varepsilon\}\} \equiv 0$. Suppose that the spaces X, Y, Z, W are spheres. Then by (iii) the secondary composition $\{\alpha, \beta, \gamma\} \in G_{p+q+r+1}/(\alpha \circ G_{q+r+1} + \gamma \circ G_{p+q+1})$ is defined in the stable homotopy groups $G_r = \lim_{n \to \infty} \pi_{n+r}(S^n)$ of spheres. From this we obtain (vi) $\{\gamma, \beta, \alpha\} = (-1)^{pq+qr+rp+1}\{\alpha, \beta, \gamma\}$ and (vii) $(-1)^{pr}\{\alpha, \beta, \gamma\} + (-1)^{qp}\{\beta, \gamma, \alpha\} + (-1)^{rq}\{\gamma, \alpha, \beta\} \equiv 0$.

S. Functional Operations

Let Φ be an operation corresponding to α and γ be the class of f. We put $\Phi_f(\beta) = \{\alpha, \beta, \gamma\}$ and call Φ_f a **functional Φ-operation**. When Φ is a cohomology operation, Φ_f is called a **functional cohomology operation**. Then $\Phi_f(\beta)$ is defined for β satisfying $f^*(\beta) = \Phi(\beta) = 0$, and $\Phi_f(\beta)$ is determined modulo $\operatorname{Im} Sf^* + \operatorname{Im} \Phi$. For $f: S^{n+k} \rightarrow S^n$, $k = 2i(p-1)-1$, we denote by $H_p(f) \cdot \varepsilon_{n+k+1} \in H^{n+k+1}(S^{n+k+1}; \mathbf{Z}_p)$ the image of a generator ε_n of $H^n(S^n; \mathbf{Z}_p)$ under the functional $\mathscr{P}^i_f$ operation. Then the **Hopf invariant modulo** p (or **mod** p **Hopf invariant**) $H_p: \pi_{n+k}(S^n) \rightarrow \mathbf{Z}_p$ is obtained (we use Sq^{2i} for $p = 2$). The following statements are equivalent: (i) The mod 2 Hopf invariant is not trivial ($H_2 \neq 0$); (ii) there exists a mapping: $S^{2k+1} \rightarrow S^{k+1}$ of Hopf invariant 1; (iii) S^k is an H-space; (iv) the Whitehead product $[\iota, \iota]$ of a generator ι of $\pi_k(S^k)$ vanishes. Also, $H_2 \neq 0$ if and only if $k = 2$, 4, 8 (J. Adams), and for an odd prime p, $H_p \neq 0$ if and only if $k = 2p-3$ (A. L. Liulevicius; N. Shimada and T. Yamanoshita).

T. Stable Homotopy Groups and Spectra

The homotopy set $\pi(S^nX, S^nY)_0$ for n-fold iterated suspensions $S^nX = X \wedge S^n = SS^{n-1}X$ and S^nY, forms a group (an Abelian group) if $n \geqslant 1$ $(n \geqslant 2)$. The limit $\pi^s(X; Y) = \lim \pi(S^nX; S^nY)_0$ with respect to the suspension homomorphisms $S: \pi(S^nX; S^nY)_0 \rightarrow \pi(S^{n+1}X; S^{n+1}Y)_0$ is called a **stable homotopy group** of X and

Y. For an r-connected space Y and a CW-complex X, $S:\pi(X;Y)_0\to\pi(SX;SY)_0$ is bijective if $\dim X\leqslant 2r$ and surjective if $\dim X\leqslant 2r+1$ (**generalized suspension theorem**). Thus, if X is a finite-dimensional CW-complex, $\pi^s(X,Y)$ is isomorphic to $\pi(S^nX,S^nY)_0$ for sufficiently large n. To discuss stable homotopy groups more generally, the following concept of spectra is used. A system $\mathbf{E}=\{E_k,\varepsilon_k\}$ which consists of CW-complexes E_k and continuous mappings $\varepsilon_k:SE_k\to E_{k+1}$ is called a **spectrum**. When $E_k=S^k$ and $\varepsilon_k=1_{k+1}:SS^k\to S^{k+1}$, $\mathbf{S}=\{S^k,1_{k+1}\}$ is called a **sphere spectrum**. When $E_k=K(G,k)$ (†Eilenberg-MacLane complex) and ε_k induces a homotopy equivalence $K(G,k)\simeq\Omega K(G,k+1)$, $\mathbf{HG}=\{K(G,k),\varepsilon_k\}$ is called an **Eilenberg-MacLane spectrum**. As in the latter, a spectrum $\mathbf{E}$ in which ε_k induces a homotopy equivalence $E_k\simeq\Omega E_{k+1}$ is called an Ω-spectrum. By Bott's periodicity, Ω-spectra $\mathbf{KU}=\{\mathbf{Z}\times\mathbf{B_u},\mathbf{U},\mathbf{Z}\times\mathbf{B_u},\mathbf{U},\dots\}$ and $\mathbf{KO}=\{\mathbf{Z}\times\mathbf{B_o},\mathbf{U/O},\mathbf{Sp/U},\mathbf{Sp},\mathbf{Z}\times\mathbf{B_{Sp}},\mathbf{U/Sp},\mathbf{SO/U},\mathbf{O},\mathbf{Z}\times\mathbf{B_o},\dots\}$ are obtained (→ Section V). Also, using †Thom complexes, **Thom spectra MU, MO**, etc. are obtained. Given a spectrum $\mathbf{E}$, by putting $\mathbf{E}^n(X,A)=\lim\pi(S^k(X/A);E_{k+1})_0$ for each pair (X,A) of CW-complexes, we obtain a **generalized cohomology theory with E-coefficient**; and by putting $\mathbf{E}_n(X,A)=\lim\pi_k(E_{k-n}\wedge(X/A))$, we obtain a **generalized homology theory with E-coefficient**. †Generalized (co)homology theory on (finite) CW-complexes can be represented by a suitable spectrum (G. W. Whitehead, E. H. Brown, Adams). Corresponding to $\mathbf{E}=\mathbf{S},\mathbf{HG},\mathbf{KU},\mathbf{MU}$, etc., we have stable (co)homotopy groups, G-coefficient (co)homology groups, K-groups, †(co)bordism groups, etc., respectively (→ 201 Homology Theory).

U. Homotopy Groups of Spheres

The spheres S^n and their homotopy groups are basic objects in homotopy theory. Although much research has been done concerning these objects, there are still open problems.

S^n is $(n-1)$-connected: $\pi_i(S^n)=0$ $(i<n)$. The fact that $\pi_n(S^n)\cong\mathbf{Z}$ (infinite cyclic group) was obtained from the †Brouwer mapping theorem. Also, $\pi_i(S^1)=0$ $(i>1)$ follows from the fact that the †universal covering space of S^1 is contractible. Suppose that we are given a continuous mapping $f:S^{2n-1}\to S^n$. We approximate it by a †simplicial mapping φ. Then the inverse image $\varphi^{-1}(*)$ of a point $*$ in the interior of an n-simplex of S^n is an $(n-1)$-dimensional †pseudomanifold which is orientable by means of a suitable generator $\varepsilon\in H_{n-1}(\varphi^{-1}(*))$. The boundary isomorphism $\partial:H_n(S^{2n-1},\varphi^{-1}(*))\cong H_{n-1}(\varphi^{-1}(*))$ and the homomorphism $\varphi_*:H_n(S^{2n-1},\varphi^{-1}(*))\to H_n(S^n,*)$ give rise to an integer $\gamma(\varphi)$ determined by the relation $\varphi_*\partial^{-1}(\varepsilon)=\gamma(\varphi)\varepsilon_n$ (ε_n is an orientation of S^n). This integer is independent of the choice of φ, so we can set $\gamma(f)=\gamma(\varphi)$. Then $f\simeq g$ implies that $\gamma(f)=\gamma(g)$. We call $\gamma(f)$ the **Hopf invariant** of f. H. Hopf defined γ and showed $\gamma:\pi_3(S^2)\cong\mathbf{Z}$ (1931); $\gamma(\pi_{2n-1}(S^n))=0$ for odd n; $\gamma(\pi_{2n-1}(S^n))\supset 2\mathbf{Z}$ for even n; and $\gamma(\pi_{2n-1}(S^n))=\mathbf{Z}$ for $n=4,8$ (1935). H. Freudenthal defined a homomorphism $E:\pi_i(S^n)\to\pi_{i+1}(S^{n+1})$, $E[f]=[Sf]$, and proved the **Freudenthal theorem**: (1) E is an isomorphism for $i<2n-1$; (2) E is a surjection for $i=2n-1$; and (3) the image of E coincides with the kernel of γ for $i=2n$. Furthermore he obtained $\pi_{n+1}(S^n)\cong\mathbf{Z}_2$ $(n\geqslant 3)$ (1937). For $n=2,4,8$, a mapping $f:S^{2n-1}\to S^n$ (†Hopf mapping) such that $\gamma(f)=1$ (given by Hopf) is the projection of a †fiber bundle S^{2n-1} over the base space S^n, and the correspondence $(\alpha,\beta)\to E\alpha+f_*\beta$ gives an isomorphism $\pi_{i-1}(S^{n-1})+\pi_i(S^{2n-1})$ (direct sum) $\cong\pi_i(S^n)$. Hence we obtain $\pi_4(S^2)=\mathbf{Z}_2$. It was shown by G. W. Whitehead and L. S. Pontryagin that $\pi_{n+2}(S^n)$ $(n\geqslant 3)$ is isomorphic to $\mathbf{Z}_2$ (1949). Whitehead also defined a **generalized Hopf homomorphism** $H:\pi_i(S^n)\to\pi_i(S^{2n-1})$ for a range of $i<3n-3$, and this restriction on the dimension was removed by P. J. Hilton and I. M. James. Using H, many nontrivial results concerning $\pi_i(S^n)$ have been obtained. Serre obtained the following (1951–1953): $\pi_i(S^n)$ is finite except when $i=n$ or $i=4m-1$ and $n=2m$. Furthermore, $\pi_{4m-1}(S^{2m})$ is the direct sum of $\mathbf{Z}$ and a finite group. Let p be an odd prime and n be even. Then $\pi_i(S^n)$ is $\mathscr{C}_p$-isomorphic to $\pi_{i-1}(S^{n-1})+\pi_i(S^{2n-1})$. Let n be odd. Then $\pi_{n+k}(S^n)\in\mathscr{C}_p$ $(k<2p-3)$, and $\pi_{n+2p-3}(S^n)$ is $\mathscr{C}_p$-isomorphic to $\mathbf{Z}_p$. Serre and H. Toda determined $\pi_{n+k}(S^n)$ for $k=3,4,5$, and Serre further determined it for $k=6,7,8$. Utilizing the reduced product space of S^n, James gave the sequence

$$\dots\to\pi_i(S^n)\xrightarrow{E}\pi_{i+1}(S^{n+1})\xrightarrow{H}\pi_{i+1}(S^{2n+1})\to\pi_{i-1}(S^n)\to\dots$$

and showed that it is an exact sequence if n is odd and an exact sequence mod 2 if n is even (1953). Using this exact sequence and the secondary composition, Toda determined $\pi_{n+k}(S^n)$ for $k\leqslant 19$ (→ Appendix A, Table 6.VI).

By the Freudenthal theorem (1), the $\pi_{n+k}(S^n)$ $(n>k+1)$ for a fixed k are isomorphic to each other. We call $\pi_{n+k}(S^n)$ $(n>k+1)$ the **stable homotopy group of the k-stem** of the sphere and denote it by G_k. For $k=0,1,2,\dots,15,\dots$, $G_k\cong\mathbf{Z},\mathbf{Z}_2,\mathbf{Z}_2,\mathbf{Z}_{24},0,0,\mathbf{Z}_2,\mathbf{Z}_{240},\mathbf{Z}_2+\mathbf{Z}_2,\mathbf{Z}_2+\mathbf{Z}_2+\mathbf{Z}_2,\mathbf{Z}_6,\mathbf{Z}_{504},0,\mathbf{Z}_3,\mathbf{Z}_2+\mathbf{Z}_2,\mathbf{Z}_{480}+\mathbf{Z}_2,\dots$. For the computation of G_k, the notion of n-connective fiber spaces is important. By

utilizing the Adams spectral sequence, we can show that G_k is closely related to the cohomology of the †Steenrod algebra. Let p be an odd prime. There exist the following sequences of elements of order p: $\{\alpha_i \in G_k\ (k=2i(p-1)-1)\}$, $\{\beta_i \in G_k\ (k=2(ip+i-1)(p-1)-2)\}$ and for $p>3\ \{\gamma_i \in G_k\ (k=2(ip^2+(i-1)p+i-2)(p-1)-3)\}$. The p-component of G_k is determined for $k<2p^2(p-1)-3$ by using Steenrod algebra. To compute G_k for higher k, relations such as $\alpha_1\beta_1^p=0$, $\beta_i\beta_1^p=0\ (i>1)$ are necessary. In general, each element of $G_k\ (k\neq 0)$ is nilpotent (G. Nishida). Let $\pi_i(S^n:p)$ be the p-component of $\pi_i(S^n)$. To survey this group for the nonstable case $(i\geqslant 2n-1)$, we utilize Serre's mod p direct sum decomposition (for n even), and we have the following two exact sequences for the case of odd n:

$$\ldots\to\pi_i(S^n)\xrightarrow{E^2}\pi_{i+2}(S^{n+2})\to\pi_i(\Omega^2(S^{n+2}),S^n)\xrightarrow{\partial}\ldots,$$

$$\ldots\to\pi_{i+3}(S^{pn+p+1}:p)\xrightarrow{\Delta}\pi_{i+1}(S^{pn+p-1}:p)$$

$$\to\pi_i(\Omega^2(S^{n+2}),S^n:p)\to\pi_{i+2}(S^{pn+p+1}:p)\xrightarrow{\Delta}\ldots,$$

where $E^2=E\circ E$ and $\Delta E^2(\alpha)=p\alpha$ (→ Appendix A, Table 6.VI).

V. Homotopy Groups of Classical Groups

Consider the classical group $U(n,\Lambda)$, which is either the orthogonal group $O(n)$ $(\Lambda=\mathbf{R})$; the unitary group $U(n)$ $(\Lambda=\mathbf{C})$; or the symplectic group $Sp(n)$ $(\Lambda=\mathbf{H})$. The **infinite classical group** $U(\infty,\Lambda)$ is defined to be the inductive limit group of $\{U(n,\Lambda)\mid n=1,2,\ldots\}$ with respect to the natural injection $U(n,\Lambda)\subset U(n+1,\Lambda)$. We call $U(\infty,\Lambda)$ the **infinite orthogonal group**, **infinite unitary group**, and **infinite symplectic group** for $\Lambda=\mathbf{R}$, $\mathbf{C}$, and $\mathbf{H}$, respectively. The dimensions of the cells of $U(\infty,\Lambda)-U(n,\Lambda)$ are $\geqslant\lambda(n+1)-1$, where $\lambda=\dim_{\mathbf{R}}\Lambda$ $(=1\ (\Lambda=\mathbf{R}),\ =2\ (\Lambda=\mathbf{C}),\ =4\ (\Lambda=\mathbf{H}))$. It follows that $\pi_k(U(n,\Lambda))$ is isomorphic to $\pi_k(U(\infty,\Lambda))$ for $k<\lambda(n+1)-2$, which is called the kth **stable homotopy group of the classical group**. Let $\mathbf{O}=U(\infty,\mathbf{R})$, $\mathbf{U}=U(\infty,\mathbf{C})$, $\mathbf{Sp}=U(\infty,\mathbf{H})$. The homotopy groups of the classical groups are periodic $(k\geqslant 0)$:

$$\pi_k(\mathbf{U})\cong\pi_{k+2}(\mathbf{U})\cong\mathbf{Z},\qquad k \text{ odd},$$

$$\cong 0,\qquad k \text{ even},$$

$$\pi_k(\mathbf{O})\cong\pi_{k+4}(\mathbf{Sp})\cong\pi_{k+8}(\mathbf{O}),$$

$$\cong\mathbf{Z},\qquad k\equiv 3,7\ (\mathrm{mod}\ 8),$$

$$\cong\mathbf{Z}_2,\qquad k\equiv 0,1\ (\mathrm{mod}\ 8),$$

$$\cong 0,\qquad k\not\equiv 0,1,3,7\ (\mathrm{mod}\ 8).$$

This is called the **Bott periodicity theorem**. The relations are deduced from weak homotopy equivalences $\mathbf{U}\to\Omega(B_{\mathbf{U}})$, $B_{\mathbf{U}}\times\mathbf{Z}\to\Omega(\mathbf{U})$, $B_{\mathbf{O}}\times\mathbf{Z}\to\Omega(\mathbf{U}/\mathbf{O})$, $\mathbf{U}/\mathbf{O}\to\Omega(\mathbf{Sp}/\mathbf{U})$, $\mathbf{Sp}/\mathbf{U}\to\Omega(\mathbf{Sp})$, $\mathbf{Sp}\to\Omega(B_{\mathbf{Sp}})$, $B_{\mathbf{Sp}}\times\mathbf{Z}\to\Omega(\mathbf{U}/\mathbf{Sp})$, $\mathbf{U}/\mathbf{Sp}\to\Omega(\mathbf{O}/\mathbf{U})$, $\mathbf{O}/\mathbf{U}\to\Omega(\mathbf{O})$. This result is applied to nonstable cases; for example, $\pi_{2n}(U(n))$ is a cyclic group of order $n!$ (→ Appendix A, Table 6.VI). The 2-dimensional homotopy group $\pi_2(G)$ of any Lie group is trivial.

Let $\alpha\in\pi_k(O(n))$, where $\alpha=[f]$, $f:S^k\to O(n)$. We define $\bar{f}:S^k\times S^{n-1}\to S^{n-1}$ by $\bar{f}(x,y)=f(x)\cdot y$ and identify S^{k+n} with the boundary $(E^{k+1}\times S^{n-1})\cup(S^k\times E^n)$ of $E^{k+1}\times E^n$. We extend $\bar{f}$ to $f:S^{k+n}\to S^n$ so that it maps $E^{k+1}\times S^{n-1}$, $S^k\times E^n$ into the upper and lower hemisphere of S^n $(S^{n-1}=$ the equator), respectively. Let $J(\alpha)\in\pi_{n+k}(S^n)$ be the class of the mapping thus obtained. This homomorphism $J:\pi_k(O(n))\to\pi_{n+k}(S^n)$ is called a **J-homomorphism** of Hopf and Whitehead. For the stable case, $J:\pi_k(\mathbf{O})\to G_k$ is injective for $k\equiv 0,1\ (\mathrm{mod}\ 8)$, and the order of the image of J is the denominator of $B_{2t}/4t$ (B_{2t} is a †Bernoulli number) or its double for $k=4t-1$ (Adams).

References

[1] H. Hopf, Über die Abbildungen von Sphären auf Sphären niedriger Dimension, Fund. Math., 25 (1935), 427–440.
[2] W. Hurewicz, Beiträge zur Topologie der Deformationen I–IV, Proc. Acad. Amsterdam, 38 (1935), 112–229, 521–528; 39 (1936), 117–125, 215–224.
[3] H. Freudenthal, Uber die Klassen der Sphären-abbildungen, Compositio Math., 5 (1937), 299–314.
[4] J.-P. Serre, Groupes d'homotopie et classes des groups abeliéns, Ann. Math., (2) 58 (1953), 258–294.
[5] R. Bott, The stable homotopy of the classical groups, Ann. Math., (2) 70 (1959), 313–337.
[6] H. Toda, Composition methods in homotopy groups of spheres, Princeton Univ. Press, 1962.
[7] E. H. Spanier, Algebraic topology, McGraw-Hill, 1966.
[8] R. M. Switzer, Algebraic topology, Springer, 1975.
[9] G. W. Whitehead, Elements of homotopy theory, Springer, 1978.

203 (III.25) Hopf Algebras

A. General Remarks

The concept of Hopf algebras arose from two directions. First in the field of algebraic topol-

ogy the notion of Hopf algebras arose from the study of homology and cohomology of Lie groups or, more generally, H-spaces. It was introduced by H. Hopf [1], whose basic structure theorem was generalized and applied to several problems by A. Borel [2]. These Hopf algebras are graded as algebras and coalgebras, and they are now used as standard tools in algebraic topology. On the other hand, Hopf algebras without grading were studied in connection with affine algebraic groups and formal groups. The study of nongraded Hopf algebras as an algebraic system was initiated by M. E. Sweedler [3], and many results on Hopf algebras of this type have been applied, not only to the theory of algebraic groups but also to the Galois theory of field extension and to combinatorial theory.

Although these two types of Hopf algebras have similar structures, and the same terminology is used to describe their properties, they are somewhat different from each other. So to avoid confusion in this article, we distinguish between graded Hopf algebras and Hopf algebras.

B. Graded Algebras

A †graded module $A=\sum_{n\geqslant 0}A_n$ over a field k is said to be of **finite type** when each A_n is finite-dimensional, A is **connected** when an isomorphism $\eta: k\cong A_\sigma$ is given. The †tensor product of two graded modules A and B is a graded module with $A\otimes B=\sum_n (A\otimes B)_n$, $(A\otimes B)_n=\sum_p A_p\otimes B_{n-p}$. We call $A^*=\sum A_n^*$ (where A_n^* is the †dual module of A_n) the **dual graded module** of A. When A and B are of finite type, $A\otimes B$ and A^* are also of finite type, and we have $(A\otimes B)^*=A^*\otimes B^*$ and $A^{**}=A$. When A and B are connected, A^* and $A\otimes B$ are also connected.

Let A be a graded module. If there exists a degree-preserving linear mapping $\varphi: A\otimes A\rightarrow A$, we call (A,φ) a **graded algebra**, whereas if there exists a degree-preserving linear mapping $\psi: A\rightarrow A\otimes A$, we call (A,ψ) a **graded coalgebra**. We call φ a **multiplication**, and ψ a **comultiplication** (or **diagonal mapping**). Usually we write $\varphi(a\otimes b)=ab$ (the **product** of $a,b\in A$), and call $\psi(a)$ the **coproduct** of a. Multiplication and comultiplication are dual operations. If A is of finite type and (A,φ) is a graded algebra, then (A^*,φ^*) (where φ^* is the dual mapping of φ) is a graded coalgebra, and vice versa. A multiplication φ is called **associative (commutative)** if $\varphi(1\otimes\varphi)=\varphi(\varphi\otimes 1)$ $(\varphi\circ T=\varphi)$, where $T: A\otimes A\rightarrow A\otimes A$ is the mapping defined by $T(a\otimes b)=(-1)^{pq}b\otimes a$ for $a\in A_p$ and $b\in A_q$. Associativity and commutativity of a comultiplication are defined dually. Let (A,ψ) be a graded connected coalgebra of finite type, and identify k with A_0 via η. Then the graded algebra (A^*,ψ^*) has the unity of k as unity if and only if ψ satisfies $\psi(1)=1\otimes 1$ and $\psi(x)=1\otimes x+x\otimes 1+\sum_i x_i'\otimes x_i''$ $(0<\deg x_i'<\deg x)$ for $\deg x>0$. In this case we say that ψ has the unity of k as **counity**. For graded algebras (A,φ) and (B,φ'), if $\varphi''=(\varphi\otimes\varphi')\circ(1\otimes T\otimes 1)$: $A\otimes B\otimes A\otimes B\rightarrow A\otimes B$, then $(A\otimes B,\varphi'')$ is also a graded algebra, which we denote by $(A\otimes B,\varphi'')=(A,\varphi)\otimes(B,\varphi')$. The tensor product of graded coalgebras is defined as the dual notion of $(A,\varphi)\otimes(B,\varphi')$.

C. Graded Hopf Algebras

For simplicity we assume that graded modules are defined over a field, connected, and of finite type. Let a graded module A be equipped with a multiplication φ and a comultiplication ψ. If φ and ψ have the unity of k as unity and $\psi:(A,\varphi)\rightarrow(A,\varphi)\otimes(A,\varphi)$ is an algebra homomorphism, then we call (A,φ,ψ) a **graded Hopf algebra**. The last condition for a graded Hopf algebra is satisfied if and only if $\varphi:(A,\psi)\otimes(A,\psi)\rightarrow(A,\psi)$ is a homomorphism of graded coalgebras. The dual (A^*,ψ^*,φ^*) is also a graded Hopf algebra, called the **dual Hopf algebra** of (A,φ,ψ).

D. H-Spaces

Let X be a topological space. The †cohomology group $H^*(X)$ (†homology group $H_*(X)$) considered over a field k has a multiplication d^* (comultiplication d_*), which is induced by the diagonal mapping $d: X\rightarrow X\times X$ and becomes a commutative and associative graded algebra (coalgebra). The groups $H^*(X)$ and $H_*(X)$ are dual to each other (→ 201 Homology Theory I, J). When X is equipped with a base point x_0 and a base point–preserving continuous mapping $h: X\times X\rightarrow X$ such that $h\circ\iota_i\simeq 1_X$ (†homotopic) for $i=1$ and 2 (where $\iota_1(x)=(x,x_0)$ and $\iota_2(x)=(x_0,x)$), we call (X,h) an H-**space**, h a **multiplication**, and x_0 a **homotopy identity** of X. Then h induces, through a †Künneth isomorphism, a comultiplication $h^*: H^*(X)\rightarrow H^*(X)\otimes H^*(X)$ (**Hopf comultiplication**) and a multiplication $h_*: H_*(X)\otimes H_*(X)\rightarrow H_*(X)$ (**Pontryagin multiplication**). Then $h^*(\alpha)$ $(\alpha\in H^*(X))$ is called the **Hopf coproduct** of α, and $h_*(\beta\otimes\gamma)$ $(\beta,\gamma\in H_*(X))$ is called the **Pontryagin product** of β and γ. When X is †arcwise connected and $H_*(X)$ is of finite type, $(H^*(X),d^*,h^*)$ and $(H_*(X),h_*,d_*)$ are graded Hopf algebras dual to each other. In particular, when h is **homotopy associative**, i.e., $h\circ(h\times 1_X)\simeq h\circ(1_X\times h)$ (**homotopy commutative**, i.e., $h\simeq h\circ T$, where

$T(x_1, x_2) = (x_2, x_1)$ for $x_i \in X$), then h^* and h_* are associative (commutative). †Topological groups and †loop spaces are homotopy associative H-spaces. If a continuous mapping $g: X \to X$ satisfies $h \circ (1_X \times g) \simeq h \circ (g \times 1_X) \simeq c$ (constant mapping $X \to \{x_0\}$), then g is called a **homotopy inverse** for X, h.

Suppose that a graded Hopf algebra A is defined over a field k of characteristic p and equipped with associative and commutative multiplication, and A is generated by a single element $a \in A_n$. Then A is a †polynomial ring $k[a]$ (n is even when $p \neq 2$) or a †quotient ring $k[a]/(a^2)$ (n is odd when $p \neq 2$) or $k[a]/(a^{p^f})$ (only when $p \neq 0$; n is even when $p \neq 2$). These are called **elementary Hopf algebras**. Every graded Hopf algebra over a †perfect field k with associative and commutative multiplication is isomorphic (as a graded algebra) to a tensor product of elementary Hopf algebras (Borel's theorem) [2]. In particular, the cohomology algebra over a field of characteristic 0 of a †compact connected Lie group is isomorphic to a †Grassmann algebra generated by elements of odd degrees [1].

E. Steenrod Algebras

The †Steenrod algebra $\mathscr{A}_p$ over $\mathbf{Z}_p$ is generated by †Steenrod operations Sq^i ($p = 2$), $\mathscr{P}^i$ ($p > 2$), and the †Bockstein operation Δ_p ($p > 2$), with composition of operations defined as multiplication. Then $\mathscr{A}_p$ is a connected associative graded algebra of finite type (not commutative). Defining a comultiplication ψ of $\mathscr{A}_p$ by $\psi(Sq^n) = \sum Sq^i \otimes Sq^{n-i}$, $\psi(\mathscr{P}^n) = \sum \mathscr{P}^i \otimes \mathscr{P}^{n-i}$, and $\psi(\Delta_p) = 1 \otimes \Delta_p + \Delta_p \otimes 1$, $\mathscr{A}_p$ becomes a graded Hopf algebra with an associative and commutative comultiplication. Thus its dual $\mathscr{A}_p^*$ is a graded Hopf algebra with an associative and commutative multiplication, and we can apply Borel's theorem to $\mathscr{A}_p^*$ in order to investigate the structure of $\mathscr{A}_p$ [4].

Let (A, φ, ψ) be a graded Hopf algebra with associative multiplication and comultiplication. Putting $c(1) = 1$ and $c(a) = -a - \sum a_i' \cdot c(a_i'')$ for $\deg a > 0$ (where $\psi(a) = 1 \otimes a + a \otimes 1 + \sum a_i' \otimes a_i''$), we obtain a linear mapping $c: A \to A$ satisfying $c\varphi = \varphi(c \otimes c)T$. We call c the **conjugation mapping** of A. When the multiplication or comultiplication is commutative, we obtain the relation $c^2 = 1$, and c is a bijection. The conjugation mapping is utilized in studying Steenrod algebras [4, 5].

F. Coalgebras

Now we turn to nongraded cases. Let A be a vector space over a field k, and let $\mu: A \otimes_k A \to A$ and $\eta: k \to A$ be linear mappings. Then the triple (A, μ, η) is said to be an **algebra** over k if $\mu \circ (\mu \otimes 1_A) = \mu \circ (1_A \otimes \mu)$ and $\mu \circ (1_A \otimes \eta) = \mu \circ (\eta \otimes 1_A) = 1_A$, where 1_A is the identity mapping of A, and $A \otimes_k k$ and $k \otimes_k A$ are identified naturally with A. We call μ the **multiplication** and η the **unit mapping** of the algebra. Dually a triple (C, Δ, ε) with C a vector space over k, linear mappings $\Delta: C \to C \otimes_k C$, and $\varepsilon: C \to k$ is said to be a **coalgebra** over k if $(1_C \otimes \Delta) \circ \Delta = (\Delta \otimes 1_C) \circ \Delta$ and $(1_C \otimes \varepsilon) \circ \Delta = (\varepsilon \otimes 1_C) \circ \Delta = 1_C$. We call Δ the **comultiplication** or the **diagonal mapping** and ε the **augmentation** or the **counit** of the coalgebra.

An algebra (A, μ, η) is called **commutative** if $\mu \circ T = \mu$, where T is the twist mapping $a \otimes b \mapsto b \otimes a$. A **cocommutative coalgebra** is defined dually. Definitions of the tensor product of two algebras or coalgebras are similar to the definitions in the graded case. If D is a subspace of a coalgebra (C, Δ, ε) over k satisfying $\Delta(D) \subset D \otimes_k D$, then $(D, \Delta|_D, \varepsilon|_D)$ is a coalgebra and is said to be a **subcoalgebra** of C. A subspace I of a coalgebra (C, Δ, ε) over k is called a **coideal** of C if $\Delta(I) \subset C \otimes_k I + I \otimes_k C$ and $\varepsilon(I) = 0$. Then the quotient space C/I has a coalgebra structure induced naturally from (C, Δ, ε) and is said to be a **quotient coalgebra** of C. The intersection and the sum of subcoalgebras of C are again subcoalgebras of C. If S is a subset of C, then the intersection of all subcoalgebras containing S is said to be the subcoalgebra generated by S. The subcoalgebra generated by any finite set or finite-dimensional subspace of C is finite-dimensional.

If (C, Δ, ε) is a coalgebra over k, then the dual space C^* of C has an algebra structure over k with multiplication μ and unit mapping η defined naturally from the †dual mappings of Δ and ε respectively. (C^*, μ, η) is called the **dual algebra** of (C, Δ, ε). Suppose that (A, μ, η) is an algebra over k, and let A° be the subset of the dual space A^* of A consisting of elements f whose kernel contains an ideal I such that A/I is finite-dimensional. Then $(A^\circ, \Delta, \varepsilon)$ is a coalgebra over k, where Δ and ε are the linear mappings induced from the dual ones of μ and η, respectively. $(A^\circ, \Delta, \varepsilon)$ is called the **dual coalgebra** of (A, μ, η). The †functors $(\ )^*$ and $(\ )^\circ$ are adjoint to one another in the sense that there is a natural bijective correspondence between the set of algebra homomorphisms of A to C^* and that of coalgebra homomorphisms of C to A° for any coalgebra C and algebra A, where **coalgebra homomorphisms** are defined as the dual notion of †algebra homomorphisms.

A nonzero subcoalgebra D of a coalgebra C is called **simple** if D has no nonzero proper subcoalgebras, and the sum of all simple subcoalgebras of C is called the **coradical** of C. If

C coincides with its coradical, then C is said to be **cosemisimple**. If C has only one simple subcoalgebra, then C is called **irreducible**. C is called **pointed** if all simple subcoalgebras of C are 1-dimensional. An element g of a coalgebra (C, Δ, ε) is called **grouplike** if $\Delta(g)=g \otimes g$ and $\varepsilon(g)=1$. The set $G(C)$ of grouplike elements in C is linearly independent over k.

G. Bialgebras

A system $(H, \mu, \eta, \Delta, \varepsilon)$ with an algebra structure (H, μ, η) and a coalgebra structure (H, Δ, ε) is said to be a **bialgebra** over k if Δ and ε are algebra homomorphisms. This last condition is equivalent to saying that μ and η are coalgebra homomorphisms. If K is a subspace of a bialgebra $(H, \mu, \eta, \Delta, \varepsilon)$ which is simultaneously a subalgebra and a subcoalgebra of H, then we call K a **subbialgebra** of H. An ideal I of (H, μ, η) which is also a coideal of (H, Δ, ε) is called a **biideal** of H and the quotient space H/I has a bialgebra structure which is said to be a **quotient bialgebra** of H. A linear mapping between bialgebras is a **bialgebra homomorphism** if it is simultaneously an algebra homomorphism and a coalgebra homomorphism. A bialgebra $(H, \mu, \eta, \Delta, \varepsilon)$ is called commutative (cocommutative) if (H, μ, η) $((H, \Delta, \varepsilon))$ is commutative (cocommutative).

Examples. Let kG be the vector space with a set G as free basis over a field k. If we define linear mappings $\Delta: kG \to G \otimes_k kG$ by $\Delta(x)= x \otimes x$ and $\varepsilon: kG \to k$ by $\varepsilon(x)=1$ for x in G, then $(kG, \Delta, \varepsilon)$ is a cocommutative coalgebra over k such that the set $G(kG)$ of grouplike elements is equal to G. Moreover if G is a †semigroup with unit element, then $(kG, \mu, \eta, \Delta, \varepsilon)$ is a cocommutative bialgebra over k, where $\mu(\eta)$ is the multiplication (unit mapping) of the †semigroup algebra kG. This bialgebra is called a **semigroup bialgebra** over k. If L is a †Lie algebra over k and $U(L)$ the †universal enveloping algebra of L with multiplication μ and unit mapping η, then Lie algebra homomorphisms $x \to x \oplus x$ $(L \to L \oplus L)$ and $x \to 0$ $(L \to \{0\})$ induce algebra homomorphisms $\Delta: U(L) \to U(L \oplus L) \cong U(L) \otimes U(L)$ and $\varepsilon: U(L) \to k$, respectively. Then $(U(L), \mu, \eta, \Delta, \varepsilon)$ is a cocommutative bialgebra over k, called the **universal enveloping bialgebra** of L.

H. Hopf Algebras

Let (A, μ, η) be an algebra over k and (C, Δ, ε) a coalgebra over k. If f and g are in $R = \mathrm{Hom}_k(C, A)$, then $f * g = \mu \circ (f \otimes g) \circ \Delta$ is called the **convolution** of f and g. Defining $\mu': R \otimes_k R \to R$ and $\eta': k \to R$ by $\mu'(f \otimes g) = f * g$ and $\eta'(\alpha) = \alpha\eta \circ \varepsilon$, then (R, μ', η') is an algebra over k. If H is a bialgebra over k with underlying coalgebra H^C and algebra H^A, then $R_H = \mathrm{Hom}_k(H^C, H^A)$ is an algebra over k in the same manner as above. If the identity mapping 1_H of H has the inverse S in R_H under the multiplication μ', then H is said to be a **Hopf algebra** over k with **antipode** S. Then S satisfies the following: $S(gh)=S(h)S(g)$ for g, h in H, $S \circ \eta = \eta$, $\varepsilon \circ S = \varepsilon$ and $T \circ (S \otimes S) \circ \Delta = \Delta \circ S$, where T is the twist mapping $a \otimes b \to b \otimes a$. If H is commutative or cocommutative, then $S \circ S = 1_H$.

If H' is another Hopf algebra over k with antipode S', then a bialgebra homomorphism f of H to H' such that $S'f = fS$ is called a **Hopf algebra homomorphism.** If H and H' are both commutative (cocommutative), then any bialgebra homomorphism of H to H' is a Hopf algebra homomorphism. The †category whose objects are commutative and cocommutative Hopf algebras over a field k and whose morphisms are Hopf algebra homomorphisms is an †Abelian category (A. Grothendieck). If H is a commutative Hopf algebra over a field of characteristic zero, then the underlying algebra H^A has no †nilpotent elements (P. Cartier).

If G is a †group, then the group bialgebra kG given above has an antipode S defined by $S(x)=x^{-1}$ for x in G, and hence kG is a cocommutative Hopf algebra over k. Another example of Hopf algebras is the †coordinate ring of an †algebraic group defined over k. More generally, let $X = \mathrm{Spec}(A)$ be an †affine group scheme over k. Then algebra homomorphisms $\Delta: A \to A \otimes_k A$, $\varepsilon: A \to k$, and $S: A \to A$ are naturally induced from the group structure of X, and $(A, \mu, \eta, \Delta, \varepsilon, S)$ is a commutative Hopf algebra over k, where μ and η are the multiplication and unit mapping of A, respectively. Conversely, if $(A, \mu, \eta, \Delta, \varepsilon, S)$ is a commutative Hopf algebra over k, then $X = \mathrm{Spec}(A)$ is an affine group scheme over k with the group structure induced from Δ, ε, and S. Hence a commutative Hopf algebra over k is nothing but a cogroup object of the category of commutative algebras over k (i.e., a †group object of the †dual category). Dually a cocommutative Hopf algebra over k is nothing but a group object of the category of cocommutative coalgebras over k.

I. Hyperalgebras

If (C, Δ, ε) is a pointed irreducible coalgebra over k, then C contains a unique grouplike element g, and kg is the unique simple subcoalgebra of C. An element a of C satisfying $\Delta(a) = a \otimes g + g \otimes a$ is called **primitive**. The set $P(C)$ of primitive elements in C is a vector

subspace of C. A cocommutative coalgebra C is called **colocal** if C is irreducible, i.e., if the dual algebra C^* of C is a †quasilocal ring. The †dimension of $P(C)$ of a colocal pointed coalgebra C is finite if and only if the dual algebra C^* of C is a †Noetherian complete local ring. A bialgebra $(H, \mu, \eta, \Delta, \varepsilon)$ over k is said to be a **hyperalgebra** over k if the underlying coalgebra is colocal. Then the unique simple subcoalgebra (grouplike element) of H is $\eta(k)$ ($\eta(1)$), and $P(H)$ has a Lie algebra structure defined by $[x, y] = xy - yx$ for x, y in $P(H)$.

The universal enveloping bialgebra $U(L)$ of a Lie algebra L over k is a hyperalgebra over k such that the set $P(U(L))$ of primitive elements in $U(L)$ is equal to L. Conversely, if the characteristic of k is zero, any hyperalgebra H over k is isomorphic to the universal enveloping algebra $U(P(H))$ of the Lie algebra $P(H)$. But in positive-characteristic cases $U(P(H))$ is generally a proper subbialgebra of H. Another important example of hyperalgebras is the dual coalgebra $hy(X) = A^\circ$ of the †stalk A at the neutral point of the †structure sheaf of an †algebraic group scheme X over k. In addition, $hy(X)$ has an algebra structure defined from the group structure of X and is a hyperalgebra over k such that $P(hy(X))$ is equal to the Lie algebra $L(X)$ of X. Although $L(X)$ plays an important role in the infinitesimal theory of algebraic groups over a field of characteristic zero, it does not give any information on infinitesimals of orders higher than p in the case of positive characteristic p. In the case of characteristic zero we see $hy(X) = U(L(X))$ and $L(X) = P(hy(X))$, and so $hy(X)$ is a natural substitute for $L(X)$ in positive-characteristic cases. From this viewpoint many interesting results on $hy(X)$ of an algebraic group scheme X which are parallel to those on $L(X)$ in the case of characteristic zero have been obtained [6–8].

References

[1] H. Hopf, Über die Topologie der Gruppen-Mannigfaltigkeiten und ihre Verallgemeinerungen, Ann. Math., (2) 42 (1941), 22–52.
[2] A. Borel, Sur la cohomologie des espaces fibrés principaux et des espaces homogènes des groupes de Lie compacts, Ann. Math., (2) 57 (1953), 115–207.
[3] M. E. Sweedler, Hopf algebras, Benjamin, 1969.
[4] J. W. Milnor, The Steenrod algebra and its dual, Ann. Math., (2) 67 (1958), 150–171.
[5] J. W. Milnor and J. C. Moore, On the structure of Hopf algebras, Ann. Math., (2) 81 (1965), 211–264.
[6] J. Dieudonné, Introduction to the theory of formal groups, Dekker, 1973.
[7] M. Takeuchi, Tangent coalgebras and hyperalgebras I, Japan. J. Math., 42 (1974), 1–143.
[8] H. Yanagihara, Theory of Hopf algebras attached to group schemes, Lecture notes in math. 614, Springer, 1977.
[9] E. Abe, Hopf algebras, Cambridge Univ. Press, 1980. (Original in Japanese, 1977.)
[10] G.-C. Rota, Coalgebras and bialgebras in combinatorics, Lecture notes at the umbral calculus conference, University of Oklahoma, 1978.

204 (XX.10) Hydrodynamical Equations

A. General Remarks

Mathematical analysis of the motion of fluids (→ 205 Hydrodynamics) gives rise to various kinds of mathematical problems; the equations that govern flows are amongst the most important and most extensively studied of the nonlinear partial differential equations. Here we review only basic and noteworthy results. Sections B–E are concerned with incompressible fluids, while Sections F and G deal with compressible fluids.

B. Nonstationary Solutions of the Navier-Stokes Equation

Let Ω be a bounded domain in $\mathbf{R}^m$ ($m = 2$ or 3) occupied by a fluid, with smooth boundary $\partial\Omega$. If the fluid is viscous and incompressible, its motion can be described by means of the velocity $u = u(t, x)$ and the pressure $p = p(t, x)$ with $t \geqslant 0$ and $x \in \Omega$ the time variable and the space variable, respectively. For simplicity, we assume that external forces are absent. Then u and p satisfy the **Navier-Stokes equation**

$$\frac{\partial u}{\partial t} + (u\nabla)u = \nu\Delta u - \nabla p, \tag{1}$$

and the **equation of continuity**

$$\operatorname{div} u = 0, \tag{2}$$

where the positive constant ν stands for the (kinetic) viscosity. On $\partial\Omega$, u is subject to the boundary condition

$$u|_{\partial\Omega} = \beta(t, x). \tag{3}$$

The initial value of u is also prescribed:

$$u|_{t=0} = u_0(x). \tag{4}$$

If Ω is unbounded, e.g., if Ω is an exterior domain outside compact surfaces, then

$$u(t,x)\to U_0(t) \quad (|x|\to\infty) \tag{3'}$$

is imposed in addition as the boundary condition at infinity.

The problem of finding u and p that satisfy (1)–(4) for given β and u_0 is called the **Navier-Stokes initial value problem** (abbreviation, NS initial value problem). Pioneering mathematical studies of this problem were initiated by J. Leray, and since the 1950s various contributions have been made by many authors, including E. Hopf, O. A. Ladyzhenskaya, H. Fujita, T. Kato, and S. Ito (→ [13, 17, 21]). We state here the result in terms of regular (classical) solutions under the simplifying assumption that Ω is bounded, $\beta\equiv 0$, and u_0 is solenoidal ($\operatorname{div} u_0=0$) and smooth. The situation depends nontrivially upon the dimension m. Namely, if $m=2$, then a regular solution of the NS initial value problem exists uniquely and globally, i.e., for all time. If $m=3$, we can prove only a local existence theorem (i.e., one holding in a finite interval) of a regular solution. This solution is unique in the internal of its existence; however, it can be extended over the whole interval when the Reynolds number is sufficiently small. In other words, the question of well-posedness of the 3-dimensional NS initial value problem is open at present.

C. Weak and Strong Solutions of the Navier-Stokes Equation

In 1951 Hopf introduced the notion of **weak solutions** of the NS initial value problem and succeeded in proving their global existence (without uniqueness). We here give the definition of **Hopf's weak solution**: Let $C_{0,\sigma}^{\infty}$ be the set of vector functions $u\in C_0^{\infty}(\Omega)$ with $\operatorname{div} u=0$. By H we denote the closure of $C_{0,\sigma}^{\infty}$ under the L^2-norm, and by V the closure of $C_{0,\sigma}^{\infty}$ under the $W_2^1(\Omega)$-norm (or equivalently, the Dirichlet norm for bounded Ω). Then an H-valued function $u=u(t)$ is a weak solution of the NS initial value problem with $\beta\equiv 0$ in $[0,T)$ $(0<T\leqslant +\infty)$ if

(i) $u\in L^{\infty}(0,T;H)\cap L^2(0,\infty;V)$,

(ii) u is weakly continuous from $[0,T)$ to H, and

(iii) u satisfies the weak equation

$$\int_0^T\Big\{(u,\varphi_t)_{L^2(\Omega)}-\nu(\nabla u,\nabla\varphi)_{L^2(\Omega)} + ((u\cdot\nabla)\varphi,u)_{L^2(\Omega)}\Big\}dt=-(u_0,\varphi(0))_{L^2(\Omega)} \tag{5}$$

for all $\varphi\in C_0^{\infty}([0,T)\times\Omega)$ with $\operatorname{div}\varphi=0$.

Note that the pressure p has been eliminated in the weak equation. In general, the uniqueness of Hopf's weak solution is not known except for $m=2$. However, when a regular solution does exist, then the weak solution coincides with it (a.e.), and is unique. Solutions of the NS initial value problem which are stronger than the weak solution to the extent that the uniqueness can be proved have been introduced, for instance, by A. A. Kiselev and Ladyzhenskaya [12] and Fujita and Kato [5]. Actually the existence and uniqueness theorems of the aforementioned regular solutions are proved using existence theorems of such **strong solutions**. Recently, Y. Giga and T. Miyakawa succeeded in generalizing the Fujita-Kato theory from L^2 to L^p, and thus they have shown that if $u_0\in L^m(\Omega)$ and is solenoidal, a unique strong solution exists at least locally for any m.

Those strong solutions that satisfy uniqueness, and hence are regular solutions of the NS initial value problem, turn out to be smooth for $t>0$; they are analytic in $t>0$ and $x\in\Omega$.

D. Stationary Solutions of the Navier-Stokes Equation

If the flow is steady, u and p are solutions of the boundary value problem consisting of (1) with $\partial u/\partial t$ omitted, (2) and (3) with β independent of t (and, in addition, (3′) with a constant U_0 if Ω is an exterior domain). The existence of solutions of this boundary value problem for the case of bounded Ω was established by Leray as one of the earliest applications of his fixed-point theorem. For the case of unbounded Ω, Leray's study was completed and extended by R. Finn, H. Fujita, and others to yield theorems on existence, regularity, and asymptotic behavior in the wakes of solutions (→ [4, 13, 21]). These stationary solutions are unique if the Reynolds number is sufficiently small. On the other hand, under certain circumstances that involve Quette flow, nonuniqueness or bifurcation of stationary solutions for large Reynolds numbers has been positively proved.

E. Euler's Equation

If the fluid is inviscid and incompressible, the Navier-Stokes equation is reduced to Euler's equation

$$\frac{\partial u}{\partial t}+(u\cdot\nabla)u=-\nabla p. \tag{6}$$

Then the boundary condition is replaced by the frictionless boundary condition, which, in

the homogeneous case, takes the form

$$u_n|_{\partial\Omega}=0, \qquad (7)$$

where u_n is the normal component of u. Regular solutions of the initial value problem consisting of (6) with (2), (4), and (7) have been proved to exist for all t if $m=2$ and in a finite time if $m=3$ [1, 11].

F. The General Navier-Stokes Equations

If the fluid is compressible, viscous, and heat-conductive, its motion is described in terms of the density ρ, the velocity u, and some thermodynamic quantity, say the absolute temperature θ, and is governed by the following system of equations, sometimes called the **general Navier-Stokes equations**:

$$\frac{\partial\rho}{\partial t}+\operatorname{div}(\rho u)=0, \qquad (8)$$

$$\frac{\partial u}{\partial t}+(u\cdot\nabla)u=\frac{\mu}{\rho}\left(\Delta+\frac{1}{3}\nabla\operatorname{div}\right)u-\frac{1}{\rho}\nabla p, \qquad (9)$$

$$\frac{\partial\theta}{\partial t}+(u\cdot\nabla)\theta=\frac{\kappa}{c_V\rho}\Delta\theta+\frac{\theta}{c_V\rho}(\operatorname{div}u)p_\theta+\frac{\Psi}{c_V\rho}. \qquad (10)$$

Here the pressure p is regarded as a function of ρ and θ through the equation of state. The viscosity coefficient μ, the coefficient of heat conduction κ, and c_V, the specific heat at constant volume, are positive constants. We have for simplicity assumed that the external force is absent, and that the Stokes condition for viscosity is satisfied. Finally, Ψ is the dissipation function:

$$\Psi=-\frac{2}{3}\mu(\operatorname{div}u)^2+\frac{\mu}{2}\sum_{i,j=1}^{m}\left(\frac{\partial}{\partial x_i}u_j+\frac{\partial}{\partial x_j}u_i\right)^2.$$

If Ω is the whole space, then the initial values ρ_0, u_0, θ_0 of ρ, u, θ are given at $t=0$, and we obtain the Cauchy problem for the general Navier-Stokes equation. Mathematical study of this Cauchy problem has become active since J. Nash [19] and N. Itaya [8] proved the existence of unique regular solutions local in time. Following Itaya's argument, A. Tani constructed a unique regular solution local in time for the initial boundary value problem consisting of (8)–(10) and boundary conditions imposed on u and θ. When it comes to the global existence of solutions, only restricted results are known. The global existence of a regular solution in the 1-dimensional version of the Cauchy problem has been established by Ya. Kanel' [10] and Itaya under certain simplifying assumptions, such that the fluid is a barotropic gas, i.e., one obeying $p=C\rho^\gamma$, where C and $\gamma\geqslant 1$ are positive constants. The 1-dimensional initial boundary value problem can also be solved globally if the gas is ideal and polytropic, i.e., one for which $p=R\rho\theta$, and with the internal energy proportional to θ and [9]. For the 3-dimensional case, we can only refer to [18], where existence of global solutions of the Cauchy problem has been proved for initial data close to constants under the assumption that the gas is ideal and polytropic.

G. Equations for Inviscid Ideal Gases

When the gas under consideration is inviscid, ideal, and barotropic, we put $\mu=0$ and $p=C\rho^\gamma$ in (9). The equation thus obtained is combined with (8) to yield the following quasilinear hyperbolic system, which admits conservation laws:

$$\frac{\partial\rho}{\partial t}+\operatorname{div}(\rho u)=0,$$
$$\frac{\partial u}{\partial t}+(u\cdot\nabla)u+\frac{C}{\rho}\nabla\rho^\gamma=0. \qquad (11)$$

It the initial data ρ_0 and u_0 are smooth to some extent, then the Cauchy problem for (11) has a regular solution local in time. Generally, we cannot expect existence of regular solutions global in time, namely, discontinuity is likely to take place in a finite time, which corresponds to the occurrence of shock waves. Therefore we have to introduce weak solutions that admit discontinuity. By definition a piecewise continuous function $\{\rho, u\}$ is a weak solution of (11) if it satisfies (11) in the distribution sense and if its discontinuity is subject to a certain jump condition, called the **Rankine-Hugoniot relation** as well as another condition, called the **entropy condition**, which two conditions allow us to distinguish a physically realizable solution among many possible discontinuous solutions [3, 14]. Global existence of the weak solution has been proved so far only for the 1-dimensional problem with initial data close to constants in the sense that their oscillations and total variations are sufficiently small. Actually in this case we can apply J. Glimm's method [6] to construct weak solutions by means of a difference approximation which involves random numbers. Little is known regarding the uniqueness of weak solutions. Finally, if we are concerned with steady-state solutions of an inviscid compressible fluid and assume that the flow is irrotational, then we are led to a quasilinear partial differential equation of mixed type for the velocity potential Φ, which is elliptic in the subsonic region and hyperbolic in the supersonic region. Classical results concerning these equations may be found in [2, 3].

References

[1] C. Bardos, Euler equation and Berger equation—Relation to turbulence, Lecture notes in math. 648, Springer, 1978, 1–46.
[2] L. Bers, Mathematical aspects of subsonic and transonic gas dynamics, Wiley, 1958.
[3] R. Courant and K. O. Friedrichs, Supersonic flow and shock waves, Interscience, 1948.
[4] R. Finn, Stationary solutions of the Navier-Stokes equation, Amer. Math. Soc. Proc. Symp. Appl. Math., 17 (1965), 121–153.
[5] H. Fujita and T. Kato, On the Navier-Stokes initial value problem I, Arch. Rational Mech. Anal., 16 (1964), 269–315.
[6] J. Glimm, Solutions in the large for nonlinear hyperbolic systems of equations, Comm. Pure Appl. Math., 18 (1965), 697–715.
[7] E. Hopf, Über die Anfangswertaufgabe für die hydrodynamischen Grundgleichungen, Math. Nachr., 4 (1950–1951), 213–231.
[8] N. Itaya, On the Cauchy problem for the system of fundamental equations describing the movement of compressible viscous fluid, Kōdai Math. Sem. Rep., 23 (1971), 60–120; corrigenda and addenda, J. Math. Kyoto Univ., 16 (1976), 239–240.
[9] A. V. Kazhikhov and V. V. Shelukhin, Overall solvability of single-valued functions for the initial boundary problem for one-dimensional equations of viscous gases (in Russian), Prikl. Mat. Mekh., 41 (1977), 282–291.
[10] Ya. I. Kanel', A model system of equations for the movement of gases in one dimension (in Russian), Differentsial'nye Uravneniya, 4 (1968), 721–734.
[11] T. Kato, On classical solutions of two-dimensional nonstationary Euler equation, Arch. Rational Mech. Anal., 25 (1967), 188–200.
[12] A. A. Kiselev and O. A. Ladyzhenskaya, On the existence and uniqueness of solutions of nonstationary problems for viscous, incompressible fluids (in Russian), Izv. Akad. Nauk SSSR, 21 (1957), 655–680.
[13] O. A. Ladyzhenskaya, The mathematical theory of viscous incompressible flows, Gordon & Breach, 1963. (Original in Russian, 1961.)
[14] P. D. Lax, Hyperbolic systems of conservation laws and the mathematical theory of shock waves, SIAM Regional Conf. Ser. Appl. Math., 11 (1973).
[15] J. Leray, Etude de diverses équations intégrales non linéaires et de quelques problèmes que pose l'hydrodynamique, J. Math. Pures Appl., (9) 12 (1933), 1–82.
[16] J. Leray, Sur le mouvement d'un liquid visqueux emplissant l'espace, Acta Math., 63 (1934), 193–248.
[17] J. L. Lions, Quelques methodes de résolution des problemes aux limites non linéaires, Dunod, 1969.
[18] A. Matsumura and T. Nishida, The initial value problem for the equations of motion of viscous and heat-conductive gases, J. Math. Kyoto Univ., 20 (1980), 67–104.
[19] J. Nash, Le problème de Cauchy pour les équations differentielles d'un fluide général, Bull. Soc. Math. France, 90 (1962), 487–497.
[20] B. L. Rozhdestvenskiĭ and N. N. Yanenko, Systems of quasilinear equations and their application to gas dynamics (in Russian), Nauka, 1978.
[21] R. Temam, Navier-Stokes equations, North-Holland, 1977.

205 (XX.9) Hydrodynamics

A. General Remarks

Gases and liquids are easily deformed, and they share many kinetic properties. They are examples of fluids. By definition, a **fluid** is a continuous substance having the property that when it is not moving, any part of the substance separated from the rest by a surface exerts an outward force that is perpendicular to the given surface.

Hydrodynamics (or **fluid dynamics**) is concerned with the equilibrium and the motion of gases and liquids without considering their molecular structure. In particular, the branch of the theory concerning fluids in equilibrium is called **hydrostatics**, and **hydrodynamics** sometimes refers to the branch concerning fluids in motion.

There are two methods of describing the motion of a fluid. One regards a fluid as a system consisting of an infinite number of particles and discusses the motion of each particle as a function of time. This is **Lagrange's method**. For example, suppose that a fluid particle with the coordinates $(x, y, z) = (a, b, c)$ at the moment $t = 0$ has coordinates $x = f_1(a, b, c, t)$, $y = f_2(a, b, c, t)$, $z = f_3(a, b, c, t)$ at an arbitrary time t. Then the motion of the fluid is perfectly determined by the functions f_1, f_2, and f_3.

The other is **Euler's method**, which discusses the values of the velocity $\mathbf{v}(u, v, w)$, the density ρ, the pressure p, etc., of the fluid at arbitrary times and positions. From this standpoint each quantity of the fluid is regarded as a function of a space-time point (x, y, z, t).

The rate at which any physical quantity F varies while moving with the fluid particle is

the **Lagrangian derivative** DF/Dt, which is related to the ordinary partial derivatives by

$$\frac{DF}{Dt}=\frac{\partial F}{\partial t}+u\frac{\partial F}{\partial x}+v\frac{\partial F}{\partial y}+w\frac{\partial F}{\partial z}.$$

The three components (u, v, w) and the two state quantities (p, ρ) (in general, other state quantities, for example, the temperature T and the †entropy S, are assumed to be determined by equations of state such as $T=T(p,\rho)$, $S=S(p,\rho)$) are determined by five $(=1+3+1)$ relations derived from the conservation laws of mass, momentum, and energy, namely, the **equation of continuity**, which corresponds to the conservation of mass,

$$\partial\rho/\partial t+\operatorname{div}(\rho\mathbf{v})=0; \tag{1}$$

the **equation of motion**, which corresponds to the conservation of momentum,

$$\partial(\rho\mathbf{v})/\partial t+\operatorname{div}(\rho\mathbf{v}\otimes\mathbf{v}-\mathbf{p})=\rho\mathbf{K}, \tag{2}$$

where K is the external force per unit mass, $\mathbf{p}$ is the stress tensor, and $\otimes$ denotes the †tensor product, while †divergence is applied to each row vector, and by virtue of (1), the equation (2) can be expressed component-wise as

$$\rho\frac{Du}{Dt}=\frac{\partial p_{xx}}{\partial x}+\frac{\partial p_{xy}}{\partial y}+\frac{\partial p_{xz}}{\partial z}+\rho K_x,$$

$$\rho\frac{Dv}{Dt}=\frac{\partial p_{yx}}{\partial x}+\frac{\partial p_{yy}}{\partial y}+\frac{\partial p_{yz}}{\partial z}+\rho K_y, \tag{2'}$$

$$\rho\frac{Dw}{Dt}=\frac{\partial p_{zx}}{\partial x}+\frac{\partial p_{zy}}{\partial y}+\frac{\partial p_{zz}}{\partial z}+\rho K_z;$$

and the **energy equation**, which corresponds to the conservation of energy,

$$\partial(\rho v^2/2+\rho E)/\partial t$$
$$+\operatorname{div}(\rho\mathbf{v}(v^2/2+E)-\mathbf{v}\cdot\mathbf{p}+\mathbf{h})=\rho\mathbf{v}\cdot\mathbf{K}, \tag{3}$$

or the **equation of entropy production**, which is another expression of (3),

$$\rho T DS/Dt=-\operatorname{div}\mathbf{h}+Q, \tag{3'}$$

where E is the internal energy per unit mass, Q the heat generated per unit time and volume, and $\mathbf{h}$ the heat flux. Here $\mathbf{K}$, p_{ik}, $\mathbf{h}$, and Q or their relations with other quantities (e.g., $\mathbf{h}=-\kappa\operatorname{grad}T$, where κ is the thermal conductivity) are assumed to be known.

B. Perfect Fluids

When there is a velocity gradient in the flow, a tangential stress appears which tends to make the velocity uniform, so that $\mathbf{p}$ is not a diagonal tensor ($-p\delta_{ik}$, i.e., pressure). This property is called fluid **viscosity**. Generally, Q and $\mathbf{h}$ do not vanish in this case. However, in order to simplify the problem we consider a nonviscous (sometimes also adiabatic) fluid, which is called a **perfect fluid** and is a good approximation in the large of actual fluids. The motion of a perfect fluid is determined by **Euler's equation of motion**

$$\rho D\mathbf{v}/Dt=-\operatorname{grad}p+\rho\mathbf{K}, \tag{4}$$

which is obtained from (1) and (2) by replacing p_{ik} by the pressure only, and also by the thermodynamic relation $DS/Dt=0$ obtained from (3′) by putting $Q=0$ and $\mathbf{h}=0$ or its integral $S=$constant in **homentropic flow**, which is governed by the **adiabatic law** $p\propto\rho^\gamma$, where γ denotes the ratio of specific heat at constant pressure to that at constant volume. In particular, for a liquid, the density variation can be neglected. Putting $\rho=$constant in (1), we have

$$\operatorname{div}\mathbf{v}=0, \tag{5}$$

which, in conjunction with (4), determines four unknowns (u, v, w, p) as functions of (x, y, z, t).

A fluid of constant density is called an **incompressible fluid**, and one of variable density a **compressible fluid**. Even though it might seem natural to consider gases as examples of compressible fluids, they can be treated as incompressible fluids if the speed of the flow of the gas $q=|\mathbf{v}|$ is small compared with the velocity $c=\sqrt{dp/d\rho}$ of sound propagating in the gas. We call $q/c=M$ the **Mach number**.

The vector $\omega(\xi,\eta,\zeta)$, which is derived from the velocity vector $\mathbf{v}$ as $\omega=\operatorname{rot}\mathbf{v}$, is called the **vorticity**. A small part of the fluid rotates with angular velocity $\omega/2$. If $\omega=0$, the flow is called **irrotational**, otherwise **rotational**. The curves $dx:dy:dz=u:v:w$ and $dx:dy:dz=\xi:\eta:\zeta$ are called, respectively, **stream lines** and **vortex lines**. The line integral $\oint_C v_s ds$ along a closed circuit C is called the **circulation** around C.

In irrotational flow, the velocity is expressed as $\mathbf{v}=\operatorname{grad}\Phi$, where Φ is called a **velocity potential**. When the external force K has a potential Ω ($K=-\operatorname{grad}\Omega$) and p is a definite function of ρ, we have the **pressure equation**

$$\frac{\partial\Phi}{\partial t}+\frac{1}{2}q^2+\int\frac{dp}{\rho}+\Omega=\text{constant},$$

which is valid everywhere in the flow. In a steady flow,

$$\frac{1}{2}q^2+\int\frac{dp}{\rho}+\Omega=\text{constant} \tag{6}$$

is valid along each stream line or each vortex line; this is called the **Bernoulli theorem**. These two equations correspond to †energy integrals of the equation of motion. Furthermore, corresponding to the conservation of †angular momentum, we have **Helmholtz's vorticity theorem**: When $\mathbf{K}=-\operatorname{grad}\Omega$ and $p=f(\rho)$, vorticity is neither created nor annihilated in the fluid.

For the irrotational motion of an incompressible fluid, †Laplace's equation $\Delta\Phi=0$ is derived from (5). Hence the problem reduces to the determination of a †harmonic function Φ under appropriate boundary conditions (e.g., for a fixed wall, normal velocity $v_n=\partial\Phi/\partial n=0$). For the 2-dimensional problem a **stream function** Ψ is introduced to satisfy (5) by the relation $u=\partial\Psi/\partial y$, $v=-\partial\Psi/\partial x$. Since the †Cauchy-Riemann equations $\partial\Phi/\partial x=\partial\Psi/\partial y$, $\partial\Phi/\partial y=-\partial\Psi/\partial x$ are valid in this case, $f=\Phi+i\Psi$ is an †analytic function of $z=x+iy$. Therefore the theory of 2-dimensional irrotational motion is essentially equivalent to the theory of complex †analytic functions, and consequently the theory of †comformal mapping is a powerful method in the theory of such fluid motion.

For irrotational steady flow of a compressible fluid in which $\Omega=0$, c is determined from (6) as a function of q. Then (1) and (4) yield a †nonlinear partial differential equation for Φ:

$$\left(1-\frac{u^2}{c^2}\right)\frac{\partial^2\Phi}{\partial x^2}+\left(1-\frac{v^2}{c^2}\right)\frac{\partial^2\Phi}{\partial y^2}+\left(1-\frac{w^2}{c^2}\right)\frac{\partial^2\Phi}{\partial z^2}$$
$$-2\frac{vw}{c^2}\frac{\partial^2\Phi}{\partial y\partial z}-2\frac{wu}{c^2}\frac{\partial^2\Phi}{\partial z\partial x}-2\frac{uv}{c^2}\frac{\partial^2\Phi}{\partial x\partial y}=0. \qquad (7)$$

This equation is †elliptic or †hyperbolic (→ 326 Partial Differential Equations of Mixed Type) according as M is less than 1 (**subsonic**) or greater than 1 (**supersonic**).

For 2-dimensional flow, we can introduce a stream function Ψ from (1) by $u=\partial\Phi/\partial x=(1/\rho)(\partial\Psi/\partial y)$, $v=\partial\Phi/\partial y=-(1/\rho)(\partial\Psi/\partial x)$. By utilizing the idea of †Legendre transformation, this system of nonlinear equations for Φ and Ψ can be reduced to a system of linear equations in the **hodograph plane** (q,θ):

$$\frac{\partial\Phi}{\partial q}=q\frac{d}{dq}\left(\frac{1}{\rho q}\right)\frac{\partial\Psi}{\partial\theta}=-\frac{1-M^2}{\rho q}\frac{\partial\Psi}{\partial\theta},$$
$$\frac{\partial\Phi}{\partial\theta}=\frac{q}{\rho}\frac{\partial\Psi}{\partial q}$$

($d(\rho q)/dq=\rho\,(1-M^2)$), where the independent variables q and θ are the magnitude and the inclination of the velocity, respectively. The treatment of 2-dimensional compressible flow on the basis of this system is called the **hodograph method**. For a flow of small M, there is a method of successive approximation (**M^2-expansion method**) which starts from Laplace's equation, neglecting the terms of $O(M^2)$ in (7). For uniform flow (velocity U in the x-direction) past a thin wing or slender body where v and w are small, we have **thin wing theory** or **slender body theory**, whose first approximation is

$$(1-M^2)\frac{\partial^2\Phi}{\partial x^2}+\frac{\partial^2\Phi}{\partial y^2}+\frac{\partial^2\Phi}{\partial z^2}=0. \qquad (8)$$

If $M<1$ or $M>1$ (although not too large) a linearization (**Prandtl-Glauert approximation**) is possible by replacing M by the Mach number at infinity $M_\infty=U/c_\infty$. For $M>1$, (8) has a †characteristic surface, which is the **Mach cone** whose central axis makes an angle $\arcsin c/q=\arcsin 1/M$ with the flow. This can be interpreted also as an envelope produced by spherical sound waves with velocity c from a source drifting with velocity q. For $M\sim1$, we put $\Phi=c_*x+\varphi$ (c_* is the fluid velocity when $q=c$). Then for an adiabatic gas, (8) is approximated by a partial differential equation of †mixed type:

$$\frac{\partial^2\varphi}{\partial y^2}+\frac{\partial^2\varphi}{\partial z^2}=\frac{\gamma+1}{c_*}\frac{\partial\varphi}{\partial x}\frac{\partial^2\varphi}{\partial x^2}. \qquad (9)$$

Such a flow in which both domains $M\gtrless 1$ coexist is called the **transonic flow**, and exact solutions by the hodograph method are known. However, continuous deceleration from $M>1$ to $M<1$ generally tends to be unstable or impossible, and the appearance of a **shock wave**, i.e., a discontinuous surface of state quantities, is not unusual. This can be considered as the †weak solution of (1), (2) (3) for a perfect fluid. In particular, in the coordinate system fixed to the surface, its integrated form can be obtained as follows: $[\rho v_n]=0$, $[p\delta_{in}+\rho v_i v_n]=0$, $[q^2/2+E+p/\rho]=0$ ([ ] is the jump of the quantity at the surface, and n is the normal component). Supplemented by the entropy increase, these formulas give relations between the fluid velocity and the state variables at the front and back of the shock. In an ideal gas they are called the **Rankine-Hugoniot relation**. Entropy is not uniform behind a curved shock, and the flow is not irrotational. For a weak shock starting from the tip of a pointed slender body, however, the discontinuity is small and approaches the †characteristic surface of (8), i.e., the **Mach wave** (compressive wave, in this case). Rarefactive Mach waves are found in the supersonic flow of acceleration around a convex surface. Such waves contribute to the drag on an obstacle placed in supersonic flow.

C. Viscous Fluids

A body moving uniformly in a fluid at rest (with velocity less than that of sound) suffers no drag as long as the viscosity of the fluid is negligible and the flow is continuous (**d'Alembert's paradox**). Hence we must take the viscosity into account in order to discuss the creation and annihilation of vortices, the generation and structure of shock waves, and the drag acting on obstacles. For this purpose, we extend **Newton's law** stating that frictional stress is proportional to the velocity gradient

and assume that the stress tensor **p** is a linear function of the rate-of-strain tensor **e**:

$$p_{xx} = -p + \mu' \operatorname{div} \mathbf{v} + \mu e_{xx} - \frac{2}{3}\mu \operatorname{div} \mathbf{v}$$

$$= -p + 2\mu\frac{\partial u}{\partial x} + \left(\mu' - \frac{2}{3}\mu\right)\operatorname{div}\mathbf{v}, \ldots,$$

$$p_{yz} = \mu e_{yz} = \mu\left(\frac{\partial w}{\partial y} + \frac{\partial v}{\partial z}\right), \ldots.$$

The proportionality constants μ and μ' are called, respectively, the **coefficients of (shear) viscosity** and **bulk viscosity**. The bulk viscosity is sometimes neglected (**Stokes's assumption**) in the usual hydrodynamics, and then the mean value of the normal stress components equals the pressure. When a fluid satisfies this linear relation between **p** and **e**, it is called a **Newtonian fluid**. Otherwise, it is called a **non-Newtonian fluid**. Except for a few cases, such as colloid solutions, fluids can be regarded as Newtonian.

If we take the viscosity into account, the equation of motion of an incompressible fluid becomes

$$\rho\, D\mathbf{v}/Dt = \rho\mathbf{K} - \operatorname{grad} p + \mu\Delta\mathbf{v}. \tag{10}$$

This is called the **Navier-Stokes equation**. A nondimensional quantity $R = \rho UL/\mu$ formed by representative length L, velocity U, density ρ, and viscosity μ of a flow is called the **Reynolds number**. In order for two flows with geometrically similar boundaries to share similar kinetic properties, their Reynolds numbers must be equal. This is called the **Reynolds law of similarity**.

For small R, we can approximate the equation of motion (10) by replacing the acceleration $D\mathbf{v}/Dt$ by $\partial\mathbf{v}/\partial t$ (**Stokes approximation**) or by $\partial\mathbf{v}/\partial t + U\partial\mathbf{v}/\partial x$ (**Oseen approximation**) for a body placed in the uniform flow of velocity U in the x-direction.

For large R, the flow can be regarded as that of a perfect fluid, since we can neglect $\mu\Delta\mathbf{v}$ as long as the velocity gradient is not too large. In the vicinity of a fixed wall, however, the velocity gradient becomes large, because in a very thin layer the velocity decreases rapidly from the value U of a perfect fluid to zero at the wall. This layer is called the **boundary layer**. For the boundary layer, **Prandtl's boundary layer equation**

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = \frac{\partial U}{\partial t} + U\frac{\partial U}{\partial x} + \frac{\mu}{\rho}\frac{\partial^2 u}{\partial y^2},$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \tag{11}$$

is valid, where x and y are the coordinates parallel and perpendicular to the wall, respectively, and U is the velocity outside the boundary layer.

If $\partial U/\partial x < 0$, it sometimes happens that the boundary layer separates from the surface of the body. In this case a vortex is generated in the flow, as large vorticities in the boundary layer are carried into the flow. For a body without separation of the boundary layer, the d'Alembert paradox holds and is no longer a "paradox," and the drag is small. Such bodies are called **streamlined**.

In compressible flow the new problem arises of the interaction of shock waves with the boundary layer. A rapid increase in pressure due to the shock wave formed on the surface of a body invalidates the assumption of a boundary layer and causes its separation. If the Mach number becomes sufficiently large ($M \gtrsim 5$, **hypersonic flow**), the bow shock approaches the body and interferes with the boundary layer. The generation of heat at the boundary layer (e.g., viscous dissipation in Q) requires the consideration of heat transfer as well as viscosity. In this manner, it becomes necessary to treat a complete system of equations which take into account the energy equation (3) as well as the temperature dependence of κ, μ, and μ'.

D. Laws of Similarity

For such complicated systems, †dimensional analysis is often useful (→ 116 Dimensional Analysis). As laws of similarity, we can consider not only those like the Reynolds law but also others for bodies which transform similarly by †affine transformations. Corresponding to equation (8), the **Prandtl-Glauert law of similarity** for subsonic flow is famous: The pressure coefficient (nondimensional pressure change) for a thin wing of chord (i.e., the length in the direction of flow) 1, span L, and thickness τ is

$$C_p(L,\tau) = \lambda C_{p_0}(\sqrt{1-M_\infty^2}\,L, \tau/\sqrt{1-M_\infty^2}\,\lambda),$$

when λ is an arbitrary constant and C_{p_0} is C_p for a body of scaled length and thickness placed in an incompressible flow. Corresponding to (9), an extension of the famous **von Kármán transonic similarity** is possible:

$$C_p(L,\tau) = \tau^{2/3}(\gamma+1)^{-1/3} \times f(\sqrt{|1-M_\infty^2|}\,L, (\gamma+1)\tau/|1-M_\infty^2|^{3/2}).$$

E. Turbulence

For low Reynolds numbers, the flow generally has smooth streamlines. For high Reynolds numbers, however, extremely irregular motion in space and time appears. The former is called

laminar flow, and the latter **turbulent flow** (→ 433 Turbulence and Chaos). The transition from laminar to turbulent flow is considered to be due to the instability of the laminar flow, and this transition has been studied by the method of small oscillations. Recently, nonlinear effects have also been examined. Regarding the internal structure of turbulence, statistical theories originated by T. von Kármán and G. I. Taylor (*Proc. Roy. Soc. London*, 151 (1935)) and A. N. Kolmogorov (*Dokl. Akad. Nauk. SSSR*, 30 (1941)) are of central importance.

F. Water Waves

†Surface waves that occur on the free surface of water (or other liquids) are called **water waves**; their restoring forces are gravity and surface tension. If we consider waves generated on still water (assumed to be inviscid and incompressible) in equilibrium, we can regard the flow field associated with the wave motion as †irrotational by virtue of †Helmholtz's vorticity theorem. Hence the flow velocity can be derived from the †velocity potential Φ which satisfies †Laplace's equation $\Delta\Phi=0$, together with the boundary conditions

$$\frac{D(H+z)}{Dt}=\frac{\partial\Phi}{\partial x}\frac{\partial H}{\partial x}+\frac{\partial\Phi}{\partial y}\frac{\partial H}{\partial y}+\frac{\partial\Phi}{\partial z}=0 \quad \text{at } z=-H(x,y), \qquad (12)$$

$$\frac{D(h-z)}{Dt}=\frac{\partial h}{\partial t}+\frac{\partial\Phi}{\partial x}\frac{\partial h}{\partial x}+\frac{\partial\Phi}{\partial y}\frac{\partial h}{\partial y}-\frac{\partial\Phi}{\partial z}=0 \quad \text{at } z=z(x,y,t), \qquad (13)$$

$$\frac{\partial\Phi}{\partial t}+\frac{1}{2}(\operatorname{grad}\Phi)^2+gz=\frac{p_0-p}{\rho}=\frac{\sigma}{\rho R_m} \quad \text{at } z=h(x,y,t), \qquad (14)$$

where the Cartesian coordinates x and y are taken in the undisturbed horizontal free surface, while the positive z-axis is vertically upward. The equations $z=-H(x,y)$ and $z=h(x,y,t)$ denote, respectively, the bottom surface (assumed to be known) and the elevation of the free surface measured from the undisturbed level $z=0$, while g stands for the gravitational acceleration, σ the surface tension, p_0 the atmospheric pressure, and $1/R_m$ the †mean curvature of the disturbed free surface expressed as

$$\frac{1}{R_m}=\left[\left\{1+\left(\frac{\partial h}{\partial x}\right)^2\right\}\frac{\partial^2 h}{\partial y^2}+\left\{1+\left(\frac{\partial h}{\partial y}\right)^2\right\}\frac{\partial^2 h}{\partial x^2}-2\frac{\partial h}{\partial x}\frac{\partial h}{\partial y}\frac{\partial^2 h}{\partial x\partial y}\right]\Bigg/\left\{1+\left(\frac{\partial h}{\partial x}\right)^2+\left(\frac{\partial h}{\partial y}\right)^2\right\}^{3/2}$$

The conditions (12) and (13) imply, respectively, that the fluid does not cross the bottom and the free surface, while (14) expresses the fact that the difference between atmospheric and fluid pressures at the free surface is equal to the normal force (per unit area) due to the surface tension. Thus the problem is formulated as a nonlinear boundary value problem for Laplace's equation including the unknown boundary $z=h(x,y,t)$.

Let us first consider linear waves for which the wave amplitude of the surface elevation is much smaller than any other characteristic linear dimension such as the wavelength or the water depth H (for simplicity, we assume hereafter $H=$constant, i.e., a flat horizontal bottom). Linearizing the boundary conditions (12)–(14) with respect to h and $\operatorname{grad}\Phi$, and assuming a sinusoidal wave proportional to $\exp[i(\mathbf{k}\cdot\mathbf{r}-\omega t)]$, $\mathbf{k}(k_x,k_y)$ and ω being respectively the (2-dimensional) †wave number vector and the †angular frequency and $\mathbf{r}=(x,y)$, we obtain the dispersion relation

$$\omega^2=\left(1+\frac{\sigma k^2}{\rho g}\right)gk\tanh(kH). \qquad (15)$$

In a layer of still water the waves are isotropic in the horizontal plane and the dispersion relation involves only the magnitude k of the wave number vector. It is readily seen from (15) that the water waves are typical †dispersive waves in which the †phase velocity $c_p(=\omega/k)$ depends on the wave number k or the wavelength $\lambda(=2\pi/k)$. It is also evident from (15) that the quantity $\sigma k^2/(\rho g)$ measures the relative importance of surface tension and gravity. Hence for waves with wavelengths much larger than $\lambda_m=2\pi\sqrt{\sigma/(\rho g)}$ (~ 1.7 cm for water), the effect of surface tension is negligible, and we have **gravity waves**. Conversely, when $\lambda \ll \lambda_m$, the effect of surface tension becomes dominant, and we have **capillary waves** or **ripples**. When the water depth H is much larger than the wavelength λ, we can approximate (15) by $\omega^2=gk+\sigma k^3/\rho$, since $\tanh(kH)\sim 1$. We call such waves **deep water waves**. On the other hand, if H is much smaller than the wavelength ($kH \ll 2\pi$), we have **shallow** (or **long**) **water waves** for which (15) can be approximated by $\omega^2=gHk^2[1+\{\sigma/(\rho gH^2)-1/3\}(kH)^2+\dots]$ if $\sigma/(\rho gH^2)=O(1)$. In particular, if we neglect $O(kH)^2$, we recover the well-known dispersionless **long gravity waves** whose phase velocity is simply $\sqrt{gH}$. In all the cases mentioned above, the amplitude function of the velocity potential Φ is proportional to $-i\omega\cosh\{k(z+H)\}/\{k\sinh(kH)\}$, so that the flow velocity due to deep water waves decreases exponentially as one proceeds vertically downward from the free surface. In the

limiting case of long gravity waves, however, the fluid motion is nearly horizontal throughout the fluid layer.

When the wave amplitude becomes larger, nonlinear effects are no longer negligible. For such waves, various †singular perturbation methods provide powerful tools. For example, the basic system of equations for weakly nonlinear (1-dimensional) shallow water waves can be reduced to a simple solvable nonlinear equation called the †Korteweg-de Vries equation, whose solitary wave solution is known as a prototype of solitons (→ 387 Solitons). Another classical example is a (1-dimensional) deep gravity wave called the **Stokes wave**, which can be obtained as a power series in the **wave steepness** (amplitude × wave number). The first term of the series is of the form of a linear sinusoidal wave, and the higher-order terms correspond to the higher harmonics, while the angular frequency is shifted from the linear case and depends not only on the wave number but also on the amplitude. Similar singular perturbation methods have also been applied to various kinds of resonant interactions such as nonlinear self-modulation, higher harmonic resonances, and multiwave interactions. Finally it should be mentioned that an exact solution representing a (1-dimensional) deep capillary wave was obtained by G. D. Crapper (*J. Fluid Mech.*, 2 (1957) 532–540; extended later to the case of finite depth by W. Kinnersley, *J. Fluid Mech.*, 77 (1976), 229–241). This is one of the few realistic exact solutions obtained so far; a famous exact solution of Gerstner's trochoidal wave does not satisfy the irrotationality condition.

References

[1] H. Lamb, Hydrodynamics, Cambridge Univ. Press, sixth edition, 1932.
[2] S. Goldstein, Modern developments in fluid dynamics, Clarendon Press, 1938.
[3] R. Courant and K. O. Friedrichs, Supersonic flow and shock waves, Interscience, 1948.
[4] L. Howarth, Modern developments in fluid dynamics, High speed flow I, II, Oxford Univ. Press, 1953.
[5] K. Oswatitsch, Gas dynamics, Academic Press, 1956.
[6] H. W. Liepmann and A. Roshko, Elements of gasdynamics, Wiley, 1956.
[7] J. J. Stoker, Water waves, Interscience, 1957.
[8] L. D. Landau and E. M. Lifshitz, Fluid mechanics. Pergamon, 1959. (Original in Russian, 1954.)
[9] K. G. Guderley, The theory of transonic flow, Pergamon, 1962.
[10] L. Rosenhead, Laminar boundary layers, Clarendon Press, 1963.
[11] G. K. Batchelor, An introduction to fluid dynamics, Cambridge Univ. Press, 1967.
[12] G. B. Whitham, Linear and nonlinear waves, Wiley, 1974.
[13] M. J. Lighthill, Waves in fluids, Cambridge Univ. Press, 1978.
[14] H. Schlichting, Boundary-layer theory, McGraw-Hill, 1979.

206 (XIV.5) Hypergeometric Functions

A. Hypergeometric Functions

The †power series

$$F(\alpha,\beta,\gamma;z)=\frac{\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\beta)}\sum_{n=1}^{\infty}\frac{\Gamma(\alpha+n)\Gamma(\beta+n)}{n!\Gamma(\gamma+n)}z^n$$

in the complex variable z is called the **hypergeometric series** or **Gauss's series**.

It is convergent for any α, β, and γ if $|z|<1$, and is convergent for $\mathrm{Re}(\alpha+\beta-\gamma)<0$ if $|z|=1$. If $z=1$, its sum is equal to $\Gamma(\gamma)\Gamma(\gamma-\alpha-\beta)/\Gamma(\gamma-\alpha)\Gamma(\gamma-\beta)$ (except when γ is a nonpositive integer). The **hypergeometric functions** are obtained as analytic continuations of the functions determined by hypergeometric series that are single-valued analytic functions defined on the domain obtained from the complex plane by deleting a line connecting branch points $z=1$ and $z=\infty$ (→ Appendix A, Table 18.1).

A hypergeometric function is a solution of the differential equation

$$z(1-z)\frac{d^2w}{dz^2}+(\gamma-(\alpha+\beta+1)z)\frac{dw}{dz}-\alpha\beta w=0, \tag{1}$$

which is called the **hypergeometric differential equation** or **Gaussian differential equation**. This equation is a differential equation of †Fuchsian type with †regular singular points at 0, 1, and ∞, whose solutions are expressed, in terms of the †P-function of Riemann, by

$$w=P\left\{\begin{matrix}0 & \infty & 1 \\ 0 & \alpha & 0 \\ 1-\gamma & \beta & \gamma-\alpha-\beta\end{matrix}\right\}z.$$

If any one of the values of γ, $\gamma-\alpha-\beta$, or $\alpha-\beta$ is integral, there exists a series containing $\log z$, representing a solution of the differential equation (1) in a neighborhood of the corresponding singular points. When none of the γ, $\gamma-\alpha-\beta$, or $\alpha-\beta$ values is integral, since the linear transformations $z'=z$, $z'=1/z$, $z'=1-z$,

$z'=z/(z-1)$, $z'=(z-1)/z$, $z'=1/(1-z)$ permute singular points, there exist 24 particular solutions around the singular points. The latter fact was first proved by E. E. Kummer (1836).

There exist various curves C for which the integral

$$w=\int_C u^{\alpha-1}(1-u)^{\gamma-\alpha-1}(1-zu)^{-\beta}du$$

is a solution of (1). Among them we can take the segment $[0,1]$ when $\operatorname{Re}\alpha>0$, $\operatorname{Re}(\gamma-\alpha)>0$. Then the corresponding solution is holomorphic in the interior of the unit circle, and

$$F(\alpha,\beta,\gamma;z)=\frac{\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\gamma-\alpha)}\times\int_C u^{\alpha-1}(1-u)^{\gamma-\alpha-1}(1-zu)^{-\beta}du.$$

Since the integrand has branch points at 0, 1, and $1/z$, we have the following expression when γ is not an integer:

$$F(\alpha,\beta,\gamma;z)=\frac{1}{(1-e^{2\pi i(\gamma-\alpha)})(1-e^{2\pi i\alpha})}\frac{\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\gamma-\alpha)}\times\oint^{(1+,0+,1-,0-)}u^{\alpha}(1-u)^{\gamma-\alpha-1}(1-zu)^{-\beta}du,$$

where $\operatorname{Re}\alpha>0$, $\operatorname{Re}(\gamma-\alpha)>0$; whereas if γ is an integer, then

$$F(\alpha,\beta,\gamma;z)=\frac{1}{(1-e^{-2\pi i\alpha})}\frac{\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\gamma-\alpha)}\times\oint^{(1+,0+)}u^{\alpha}(1-u)^{\gamma-\alpha-1}(1-zu)^{-\beta}du,$$

where the contour in the first expression encircles successively each of 1, 0, 1, and 0 once with indicated directions. These expressions can be adopted as a definition of the hypergeometric functions for the general value of z. Other integral expressions are also known (→ Appendix A, Table 18.I).

B. The Ladder Method

A linear ordinary differential equation of the second order having three regular singular points on the complex sphere is easily transformed into an equation of the form (1). To solve such an equation with a parameter, it is often useful to decompose, in two different ways, the main part of the equation into two factors of the first order, and find a recurrence formula involving the parameter, as we shall see in the following example. This method is called the **ladder method** or **factorization method**.

For example, †Legendre's differential equation

$$L_n[w]\equiv(1-z^2)(((1-z^2)w')'+n(n+1)w)=0$$

is decomposed as follows:

$$L_n=S_n\cdot T_n+n^2=T_{n+1}\cdot S_{n+1}+(n+1)^2,$$

$$T_n=(1-z^2)\frac{d}{dz}+nz,\qquad S_n=(1-z^2)\frac{d}{dz}-nz.$$

If w_n is a solution of $L_n[w]=0$, then multiplying both sides of $S_n\cdot T_n[w_n]+n^2w_n=0$ by T_n, we find that $T_n\cdot S_n(T_n[w_n])+n^2(T_n[w_n])=0$, that is, $T_n[w_n]$ is a solution of $L_{n-1}[w]=0$. Similarly, we see that $S_{n+1}[w_n]$ is a solution of $L_{n+1}[w]=0$. In this sense, S_n and T_n are called, respectively, the **step-up operator**, or **up-ladder**, and the **step-down operator**, or **down-ladder**, with respect to the parameter n.

The above relation constitutes a recurrence formula for Legendre functions (→ Appendix A, Table 18.II).

C. Extensions of Hypergeometric Functions

J. Thomas (1870) proposed the series

$$1+\sum_{n=1}^{\infty}\frac{(\alpha_1)_n(\alpha_2)_n\dots(\alpha_h)_n}{(\beta_1)_n(\beta_2)_n\dots(\beta_h)_n}z^n,$$

$$(\lambda)_n=\lambda(\lambda+1)\dots(\lambda+n-1)$$

as an extension of the hypergeometric series. The sum of this series satisfies the hth-order differential equation

$$(1-z)\frac{d^hw}{dt^h}+(A_1-B_1z)\frac{d^{h-1}w}{dt^{h-1}}+(A_2-B_2z)\frac{d^{h-2}w}{dt^{h-2}}+\dots+(A_h-B_hz)w=0,$$

$t=\log z$.

When $h=2$ and $\beta_1=1$, it reduces to the ordinary hypergeometric series. The notation

$$ {}_pF_q(\alpha_1,\alpha_2,\dots,\alpha_p;\beta_1,\beta_2,\dots,\beta_q;z)=\sum_{n=0}^{\infty}\frac{(\alpha_1)_n(\alpha_2)_n\dots(\alpha_p)_n}{n!(\beta_1)_n(\beta_2)_n\dots(\beta_q)_n}z^n, \qquad (2)$$

which is due to L. Pochhammer and modified by E. W. Barnes, is used to denote the extended hypergeometric series, and the function defined by (2) is often called **Barnes's extended hypergeometric function**. For example, Gauss's series in this notation is ${}_2F_1(\alpha,\beta,\gamma;z)$.

Corresponding to Barnes's integral expression for hypergeometric functions, it is known that the integral

$$W(z)=\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}K(\zeta)H(\zeta)z^{-\zeta}d\zeta,$$

where

$$K(\zeta)=K(\zeta+1)$$

and

$$H(\zeta)=\frac{\Gamma(\zeta+\alpha_1)\Gamma(\zeta+\alpha_2)\ldots\Gamma(\zeta+\alpha_h)}{\Gamma(\zeta+1+\beta_1)\Gamma(\zeta+1+\beta_2)\ldots\Gamma(\zeta+1+\beta_h)},$$

is a solution of the hth-order differential equation at the beginning of this section. The hypergeometric function expressed by the definite integral

$$\int\zeta^{a}(\zeta-1)^{b}(\zeta-z)^{c}\,d\zeta$$

has an obvious formal extension

$$\int(\zeta-a_1)^{b_1}(\zeta-a_2)^{b_2}\ldots(\zeta-a_m)^{b_m}(\zeta-z)^{c}\,d\zeta.$$

On the other hand, the equation

$$\sum_{\nu=0}^{m}\varphi_\nu(z)\frac{d^\nu w}{dz^\nu}=0,$$

where

$$\varphi_\nu(z)=\frac{(-1)^{m-1-\nu}}{(h+m-2)\ldots(h+1)h}\times\left(\binom{h+m-\nu-2}{m-\nu-1}P_1^{(m-1-\nu)}(z)+\binom{h+m-\nu-2}{m-\nu}P_0^{(m-\nu)}(z)\right),$$

$$P_0(z)=(z-a_1)(z-a_2)\ldots(z-a_m),$$

$$P_1(z)=P_0(z)\left(\frac{\beta_1}{z-a_1}+\frac{\beta_2}{z-a_2}+\ldots+\frac{\beta_m}{z-a_m}\right),$$

called the **Tissot-Pochhammer differential equation**, has a solution

$$w(z)=\int_C(\zeta-a_1)^{\beta_1-1}(\zeta-a_2)^{\beta_2-1}\ldots\times(\zeta-a_m)^{\beta_m-1}(\zeta-z)^{h+m-2}\,d\zeta.$$

After Pochhammer (1870), this is sometimes called Pochhammer's generalized hypergeometric function.

As another extension of Gauss's series, H. E. Heine (1846) introduced **Heine's series**:

$$\varphi(a,b,c;q;z)=1+\frac{(1-q^a)(1-q^b)}{(1-q)(1-q^c)}q^z+\frac{(1-q^a)(1-q^{a+1})(1-q^b)(1-q^{b+1})}{(1-q)(1-q^2)(1-q^c)(1-q^{c+1})}q^{2z}+\ldots.$$

Setting $q=1+\varepsilon$, $z=(1/\varepsilon)\log x$, and letting $\varepsilon\to0$, we obtain Gauss's series as the limit of Heine's series.

D. Hypergeometric Functions of Several Variables

P. Appell (1880) formally extended Gauss's series to the case of two variables and defined four kinds of functions [3]:

$$F_1(\alpha;\beta,\beta';\gamma;x,y)=\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{(\alpha)_{m+n}(\beta)_m(\beta')_n}{m!n!(\gamma)_{m+n}}x^my^n,$$

$$F_2(\alpha;\beta,\beta';\gamma,\gamma';x,y)=\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{(\alpha)_{m+n}(\beta)_m(\beta')_n}{m!n!(\gamma)_m(\gamma')_n}x^my^n,$$

$$F_3(\alpha,\alpha';\beta,\beta';\gamma;x,y)=\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{(\alpha)_m(\alpha')_n(\beta)_m(\beta')_n}{m!n!(\gamma)_{m+n}}x^my^n,$$

$$F_4(\alpha;\beta;\gamma,\gamma';x,y)=\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{(\alpha)_{m+n}(\beta)_{m+n}}{m!n!(\gamma)_m(\gamma')_n}x^my^n.$$

They are called **Appell's hypergeometric functions of two variables**. Each satisfies a corresponding system of partial differential equations:

$$F_1\begin{cases}x(1-x)r+y(1-x)s+(\gamma-cx)p-\beta yq-\alpha\beta z=0,\\ y(1-y)t+x(1-y)s+(\gamma-c'x)q-\beta'xp-\alpha\beta'z=0,\\ (x-y)s-\beta'p+\beta q=0,\end{cases}$$

$$F_2\begin{cases}x(1-x)r-xys+(\gamma-cx)p-\beta yq-\alpha\beta z=0,\\ y(1-y)t-xys+(\gamma'-c'y)q-\beta'xp-\alpha\beta'z=0,\end{cases}$$

$$F_3\begin{cases}x(1-x)r+ys+(\gamma-cx)p-\alpha\beta z=0,\\ y(1-y)t+xs+(\gamma-c''y)q-\alpha'\beta'z=0,\end{cases}$$

$$F_4\begin{cases}x(1-x)r-y^2t-2xys+(\gamma-cx)p-cyq-\alpha\beta z=0,\\ y(1-y)t-x^2r-2xys+(\gamma'-cy)q-cxp-\alpha\beta z=0,\end{cases}$$

where

$$c=\alpha+\beta+1,\quad c'=\alpha+\beta'+1,\quad c''=\alpha'+\beta'+1,$$

$$p=\partial z/\partial x,\quad q=\partial z/\partial y,\quad r=\partial^2z/\partial x^2,$$

$$s=\partial^2z/\partial x\partial y,\quad t=\partial^2z/\partial y^2.$$

Appell's hypergeometric functions can also be represented by integrals: for example,

$$F_1=\frac{\Gamma(\gamma)}{\Gamma(\beta)\Gamma(\beta')\Gamma(\gamma-\beta-\beta')}\iint u^{\beta-1}v^{\beta'-1}\times(1-u-v)^{\gamma-\beta-\beta'-1}(1-ux-vy)^{-\alpha}\,du\,dv,$$

$$F_2=\frac{\Gamma(\gamma)\Gamma(\gamma')}{\Gamma(\beta)\Gamma(\beta')\Gamma(\gamma-\beta)\Gamma(\gamma'-\beta')}\int_0^1\int_0^1u^{\beta-1}v^{\beta'-1}\times(1-u)^{\gamma-\beta-1}(1-v)^{\gamma'-\beta'-1}(1-ux-vy)^{-\alpha}\,du\,dv,$$

$$F_3=\frac{\Gamma(\gamma)}{\Gamma(\beta)\Gamma(\beta')\Gamma(\gamma-\beta-\beta')}\iint u^{\beta-1}v^{\beta'-1}$$
$$\times(1-u-v)^{\gamma-\beta-\beta'-1}(1-ux)^{-\alpha}(1-vy)^{-\alpha'}\,du\,dv,$$

where, for F_1 and F_3, the domain of integration is $u\geqslant 0$, $v\geqslant 0$, $1-u-v\geqslant 0$. E. Picard (1881) showed that F_1 can also be expressed by a single integral:

$$F_1=\frac{\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\alpha-\gamma)}\int_0^1 u^{\alpha-1}(1-u)^{\gamma-\alpha-1}$$
$$\times(1-ux)^{-\beta}(1-uy)^{-\beta'}\,du.$$

G. Lauricella (1893) extended the foregoing functions to the case of more than two variables. More general hypergeometric series of several variables were defined by R. Mellin, J. Horn, and J. Kampé de Fériet [3,4]. Every algebraic equation can be solved analytically in terms of the foregoing functions (Mellin, R. Birkeland) [4]. Also there are studies concerning †Riemann's problem and †automorphic functions derived from F_1 (Picard; T. Terada [5]).

E. Hypergeometric Functions with Matrix Argument

For symmetric matrices Z of degree m, C. S. Herz defined **hypergeometric functions with matrix argument** as follows [5]: Denoting by $\operatorname{etr} Z$ the exponential $\exp(\operatorname{tr} Z)$ of the †trace of Z, let

$${}_0F_0(Z)=\operatorname{etr} Z,$$

$${}_{p+1}F_q(\alpha_1,\ldots,\alpha_p;\beta_1,\ldots,\beta_q;\gamma;Z)$$
$$=\frac{1}{\Gamma_m(\gamma)}\int_{\Lambda>0}\operatorname{etr}(-\Lambda)_pF_q(\alpha_1,\ldots,\alpha_p;\beta_1,\ldots,\beta_q;$$
$$\Lambda Z)(\det\Lambda)^{\gamma-p}\,d\lambda_{11}\,d\lambda_{22}\ldots d\lambda_{mm},\quad(3)$$

$${}_pF_{q+1}(\alpha_1,\ldots,\alpha_p;\beta_1,\ldots,\beta_q;\gamma;\Lambda)$$
$$=\frac{\Gamma_m(\gamma)}{(2\pi i)^{m(m+1)/2}}\int_{\operatorname{Re}Z=X_0>0}\operatorname{etr} Z_pF_q(\alpha_1,\ldots,\alpha_p;$$
$$\beta_1,\ldots,\beta_q;\Lambda Z^{-1})(\det Z)^{-\gamma}\,dz_{11}\,dz_{22}\ldots dz_{mm},\quad(4)$$

where

$$\Lambda=(\lambda_{ij})_{i,j=1,\ldots,m},$$

$$Z=((1+\delta_{ij})z_{ij}/2)_{i,j=1,\ldots,m},$$

$$\Gamma_m(\gamma)=\pi^{m(m-1)/4}\Gamma(\gamma)\Gamma(\gamma-1/2)\ldots$$
$$\times\Gamma(\gamma-(m-1)/2),$$

and $\Lambda>0$ means that Λ is †positive definite. The integral (3) converges for $-Z>0$ if $\operatorname{Re}\gamma>(m-1)/2$. If $\operatorname{Re}\gamma$ is sufficiently large, then for suitably chosen X_0, (4) converges in a domain of the space of Λ and represents an analytic function of its argument. In particular, we have

$${}_1F_0(\alpha;Z)=(\det(E-Z))^{-\alpha}.$$

Based on this definition, many special functions and formulas are extended to the case of a matrix argument. For example,

$$A_\delta(Z)={}_0F_1(\delta+(m+1)/2;\ -Z)/\Gamma_m(\delta+(m+1)/2)\quad(5)$$

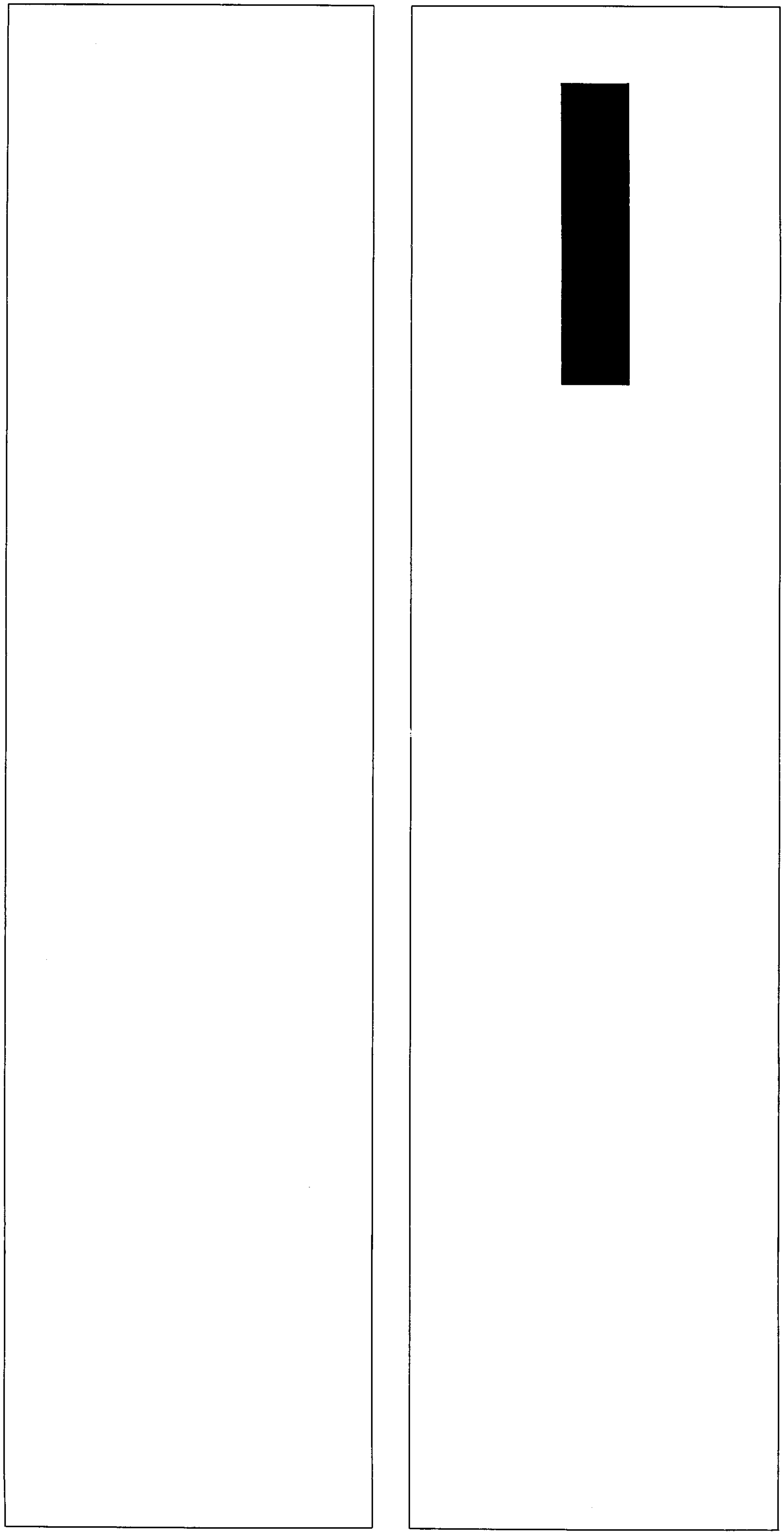
is an extension of the †Bessel function, and this reduces to

$$(t/2)^{-\delta}J_\delta(t)=A_\delta((t/2)^2)$$

when $m=1$. Formula (5) is applied to the †noncentral Wishart distribution in mathematical statistics.

References

[1] F. Klein, Vorlesungen über die hypergeometrische Funktionen, Springer, 1933.
Also → references to 389 Special Functions.
For applications of the ladder method,
[2] L. Infeld and T. E. Hull, The factorization method, Rev. Mod. Phys., 23 (1951), 21–68.
For hypergeometric functions of several variables,
[3] P. Appell, Sur les fonctions hypergéométriques de plusieurs variables, Mémor. Sci. Math., Gauthier-Villars, 1925.
[4] G. Belardinelli, Fonctions hypergéométriques de plusieurs variables et résolution analytique des équations algébriques générales, Mémor, Sci. Math., Gauthier-Villars, 1960.
[5] T. Terada, Problème de Riemann et fonctions automorphes provenant des fonctions hypergéométriques de plusieurs variables, J. Math. Kyoto Univ. 13 (1973), 557–578.
For the case of a matrix variable,
[6] C. S. Herz, Bessel functions of matrix argument, Ann. Math., (2) 61 (1955), 474–523.
[7] L. J. Slater, Generalized hypergeometric functions, Cambridge Univ. Press, 1966.



207 (XI.13) Ideal Boundaries

A. Ideal Boundaries

For a given Hausdorff space R, a †compact Hausdorff space R^* that contains R as its dense subspace is called a **compactification** of R, and $\Delta = R^* - R$ is called an **ideal boundary** of R. In the present article, we deal mainly with properties (in particular, function-theoretic properties) of ideal boundaries of †Riemann surfaces R.

B. Harmonic Boundaries

By R we mean an open Riemann surface. The set Γ of points p^* in Δ such that $\liminf_{R \ni p \to p^*} P(p) = 0$ for every †potential P (i.e., a positive †superharmonic function P for which the class of nonnegative †harmonic functions smaller than P consists only of the constant function 0) is a compact subset of R^*. The set Γ is called the **harmonic boundary** of R with respect to R^*. For an arbitrary compact subset K in $\Delta - \Gamma$, there exists a finite-valued potential P_K with $\lim_{R \ni p \to p^*} P_K(p) = \infty$ $(p^* \in K)$. From this, various kinds of †maximum principles are derived. For instance, if u is a harmonic function bounded above for which $\limsup_{p \to \Gamma} u(p) \leqslant M$ holds, then $u \leqslant M$ on R.

There are infinitely many compactifications of R. For two compactifications R_i^* $(i = 1, 2)$ of R, we say that R_1^* is **greater than** R_2^* or, equivalently, **lies over** R_2^*, if the identity mapping of R can be extended to a continuous mapping of R_1^* onto R_2^*. In order that deep function-theoretic studies of R^* can be carried out, various conditions must be imposed on R^*. A compactification R^* is said to be of **Stoïlow type** (or of **type** S) if for every †connected open subset G^* in R^* whose boundary in R^* is contained in R, $G^* - \Delta$ is also connected. Next suppose that $R \notin O_G$ (→ 367 Riemann Surfaces E). For a given real-valued function f on Δ, let $\bar{\mathfrak{U}}_f^{R,R^*}$ ($\underline{\mathfrak{U}}_f^{R,R^*}$) be the class of †superharmonic functions s bounded from below (†subharmonic functions s bounded from above) such that $\liminf_{R \ni p \to p^*} s(p) \geqslant f(p^*)$ ($\limsup_{R \ni p \to p^*} s(p) \leqslant f(p^*)$) for every $p^* \in \Delta$. If these classes are nonempty, then $\bar{H}_f^{R,R^*}(p) = \inf\{s(p) \mid s \in \bar{\mathfrak{U}}_f^{R,R^*}\}$ and $\underline{H}_f^{R,R^*}(p) = \sup\{s(p) \mid s \in \underline{\mathfrak{U}}_f^{R,R^*}\}$ are harmonic on R, and $\underline{H}_f^{R,R^*} \leqslant \bar{H}_f^{R,R^*}$. In particular, if $\underline{H}_f^{R,R^*} \equiv \bar{H}_f^{R,R^*}$, then the common function is denoted by H_f^{R,R^*}, and the function f is said to be **resolutive** with respect to R^*. A compactification such that every bounded continuous function on Δ is resolutive is called a **resolutive compactification**. In such a case, a point p^* in Δ is said to be **regular** with respect to the †Dirichlet problem if $\lim_{R \ni p \to p^*} H_f^{R,R^*}(p) = f(p^*)$ for every bounded continuous function f on Δ (→ 120 Dirichlet Problem). The set Δ_r of regular points in Δ is contained in Γ. If R^* is a resolutive compactification, then there exists a unique positive †Borel measure μ_p such that $H_f^{R,R^*}(p) = \int_\Delta f(p^*) d\mu_p(p^*)$ for every bounded continuous function f on Δ. This measure is called the **harmonic measure** with respect to $p \in R$. There exists a function $P(p, p^*)$ on $R \times \Delta$ with $d\mu_p(p^*) = P(p, p^*) d\mu_o(p^*)$ for an arbitrary fixed point o in R satisfying the following three conditions: (i) $P(p, p^*)$ is harmonic on R as a function of p; (ii) $P(p, p^*)$ is Borel measurable as a function of p^*; (iii) $k(o,p)^{-1} \leqslant P(p, p^*) \leqslant k(o,p)$, with the Harnack constant $k(o,p)$ of $\{o, p\}$ relative to R [9].

C. Compactifications Determined by Function Families

A family F of real-valued continuous functions on R admitting infinite values is called a **separating family** on R if there exists an f in F such that $f(p) \neq f(q)$ for any pair of given distinct points p and q in R. A compactification R^* is called an **F-compactification**, denoted by R_F^*, if every function in F can be continuously extended to R^* and the family of extended functions again constitutes a separating family on R^*. The correspondence $\varphi: F \to R_F^*$ defines a single-valued mapping of all separating families F on R onto all F-compactifications of R. If $F_1 \supset F_2$, then $\varphi(F_1)$ lies over $\varphi(F_2)$. For any R^*, $\varphi^{-1}(R^*)$ contains infinitely many separating families, among which the separating families constituting †associative algebras are important. The following are typical examples of compactifications determined by function families:

(1) The **Aleksandrov compactification** is the $\mathfrak{U}$-compactification $R_{\mathfrak{U}}^*$ with the family $\mathfrak{U}$ of bounded continuous functions on R with compact support. It is the smallest compactification of R and is often used in function theory in discussing Dirichlet problems for relatively noncompact subregions in reference to relative boundaries.

(2) The **Stone-Čech compactification** is the $\mathfrak{C}$-compactification $R_{\mathfrak{C}}^*$ with the family $\mathfrak{C}$ of bounded continuous functions on R. It is the largest compactification of R. It is rarely used in function theory, but an application is found in the work of M. Nakai [8].

(3) The **Kerékjártó-Stoïlow compactification** is the $\mathfrak{S}$-compactification $R_{\mathfrak{S}}^*$ with the family $\mathfrak{S}$ of bounded continuous functions f on R

such that there exist compact sets K_f with the property that the f are constants on each connected component of $R-K_f$. This is the smallest compactification of Stoïlow type. Many applications of this compactification can be found in function theory, among which the investigation done by M. Ohtsuka on the Dirichlet problem and the theory of conformal mappings is typical.

(4) The **Royden compactification** is the $\mathfrak{N}$-compactification $R^*_{\mathfrak{N}}$ with the family $\mathfrak{N}$ of bounded C^∞ functions f on R with finite Dirichlet integrals $\iint_R df \wedge {*df}$. It was introduced by H. L. Royden and developed further by S. Mori, M. Ôta, Y. Kusunoki, Nakai, and others. This compactification has been used effectively in the study of HD-functions and the classification problem of Riemann surfaces (→ 367 Riemann Surfaces).

(5) The **Wiener compactification** is the $\mathfrak{W}$-compactification $R^*_{\mathfrak{W}}$ with the family $\mathfrak{W}$ of bounded continuous functions f on R such that $\{H_f^{G_n}\}$ converges to a unique harmonic function independent of the choice of exhaustions $\{G_n\}$ of an arbitrary fixed subregion $G \notin O_G$, where the G_n are relatively compact subregions of G. It is the largest resolutive compactification, and compactifications smaller than $R^*_{\mathfrak{W}}$ are always resolutive. This compactification was introduced independently by Mori, K. Hayashi, Kusunoki, and C. Constantinescu and A. Cornea and is useful for the study of HB-functions and the classification of Riemann surfaces.

(6) The **Martin compactification** is the $\mathfrak{M}$-compactification $R^*_{\mathfrak{M}}$ with the family $\mathfrak{M}$ of bounded continuous functions f on R such that there exist relatively compact regions R_f with the property that $f=H_{f^*}^{R-R_f,R^*_{\mathfrak{A}}-R_f}/H_{1^*}^{R-R_f,R^*_{\mathfrak{A}}-R_f}$ on $R-R_f$. Here f^* coincides with f on R_f and equals 0 on $R^*_{\mathfrak{A}}-R$, and 1^* is similarly defined. The set $R^*_{\mathfrak{M}}-R$ is called the **Martin boundary** of R. If †Green's function g exists on R, then the function $m(p,q)=g(p,q)/g(o,q)$ for an arbitrary fixed $o \in R$ can be extended continuously to $R \times R^*_{\mathfrak{M}}$, which is called the **Martin kernel**. By the metric $d_{\mathfrak{M}}(q,r)=\sup_{p\in R_0}|m(p,q)/(1+m(p,q))-m(p,r)/(1+m(p,r))|$ with a parametric disk R_0 in R, $R^*_{\mathfrak{M}}$ is †metrizable. This compactification was introduced by R. S. Martin, and many applications of it to the study of HP-functions, potential theory, Markov chains, and cluster sets were obtained by M. H. Heins, Z. Kuramochi, J. L. Doob, Constantinescu and Cornea, and others.

(7) For a function f on R, $(R)\partial f/\partial n=0$ means that there exists a relatively compact subregion R_f such that f is of class C^∞ on R outside R_f and the Dirichlet integral of f over $R-R_f$ is not greater than those of functions on $R-R_f$ that coincide with f on the boundary of R_f. The **Kuramochi compactification** is the $\mathfrak{K}$-compactification $R^*_{\mathfrak{K}}$ with the family $\mathfrak{K}$ of bounded continuous functions f on R satisfying $(R)\partial f/\partial n=0$. The continuous function $k(p,q)$ on R such that $(R)\partial k/\partial n=0$ vanishes in a fixed parametric disk R_0 in R and is harmonic in $R-\bar{R}_0$ except for a positive †logarithmic singularity at a point q can be extended continuously to $R \times R^*_{\mathfrak{K}}$, which is called the **Kuramochi kernel**. By the use of this kernel, $R^*_{\mathfrak{K}}$ is metrizable, as in the case of Martin compactification. This compactification was introduced by Kuramochi, and its important applications to the study of HD-functions, potential theory, and cluster sets were made by Kuramochi, Constantinescu and Cornea, and others.

Among compactifications (1)–(7), no boundary point in (2), (4), or (5) satisfies the †first countability axiom (hence they are not metrizable), while the others are all metrizable. In (4) and (5), $\Delta_r=\Gamma$. Fig. 1 shows the relationship among the seven examples. Here $A \to B$ means that A lies over B, and $A \neq B$ means that in general neither $A \to B$ nor $B \to A$.

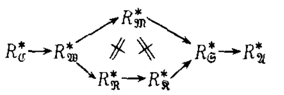


Fig. 1

D. Remarks

In contrast to topological compactifications (1)–(3), (4)–(7) can be regarded as potential-theoretic and have enough ideal boundary points so that one can solve the Dirichlet problem and introduce various measures there. Utilizing Green's functions, R. S. Martin [6] deduced the first important compactification and gave the integral representation of positive harmonic functions (the extension of the †Poisson integral). Z. Kuramochi [3] obtained his compactification similarly by using N-Green functions introduced by himself instead of the usual Green's functions. In this compactification, the ideal boundary points can be considered to be interior points of the surfaces in a potential-theoretic sense. In case of finitely connected domains with smooth boundaries, both the Martin and Kuramochi boundaries coincide with the usual boundaries.

The Royden and Wiener compactifications were introduced as the maximal ideal spaces of respective function algebras. Their ideal boundaries contain extremely many points. However, these compactifications have elegant

properties and many applications. An analytic mapping φ of a Riemann surface R into another R' is called a Dirichlet (Fatou) mapping if φ can be extended to a continuous mapping of $R^*_{\mathfrak{N}}(R^*_{\mathfrak{W}})$ to $R'^*_{\mathfrak{N}}(R'^*_{\mathfrak{W}})$. For example, a Lindelöfian mapping (AD-function) is a Fatou (Dirichlet) mapping. The Dirichlet and Fatou mappings were investigated by Constantinescu, Cornea, and others.

By using the Martin boundary, Z. Kuramochi and M. Nakai proved the extension of the Evans-Selberg theorem to parabolic Riemann surfaces [3, 8]. The normal derivatives of HD-functions on the ideal boundaries and their applications were studied by Constantinescu, Cornea, F. Maeda, Y. Kusunoki, and others. The theory of compactification can be generalized to domains in $\mathbf{R}^n$, †Green spaces, and harmonic spaces.

References

[1] C. Constantinescu and A. Cornea, Ideale Ränder Riemannscher Flächen, Springer, 1963.
[2] K. Hayashi, Sur une frontière des surfaces de Riemann, Proc. Japan Acad., 37 (1961), 469–472.
[3] Z. Kuramochi, Mass distributions on the ideal boundaries of abstract Riemann surfaces I, II, Osaka Math. J., 8 (1956), 119–137, 145–186.
[4] Y. Kusunoki, On a compactification of Green spaces, Dirichlet problem and theorems of Riesz type, J. Math. Kyoto Univ., 1 (1962), 385–402.
[5] Y. Kusunoki and S. Mori, On the harmonic boundary of an open Riemann surface I, Japan. J. Math., 29 (1959), 52–56.
[6] R. S. Martin, Minimal positive harmonic functions, Trans. Amer. Math. Soc., 49 (1941), 137–172.
[7] S. Mori, On a compactification of an open Riemann surface and its application, J. Math. Kyoto Univ., 1 (1961), 21–42.
[8] M. Nakai, On Evans potential, Proc. Japan Acad., 38 (1962), 624–629.
[9] M. Nakai, Radon-Nikodym densities between harmonic measures on the ideal boundary of an open Riemann surface, Nagoya Math. J., 27 (1966), 71–76.
[10] H. L. Royden, On the ideal boundary of a Riemann surface, Contributions to the Theory of Riemann Surfaces, Princeton Univ. Press, 1953, 107–109.
[11] F. Maeda, M. Ohtsuka, et al., Kuramochi boundaries of Riemann surfaces, Lecture notes in math. 58, Springer, 1968.
[12] L. Sario and M. Nakai, Classification theory of Riemann surfaces, Springer, 1970.

208 (X.7) Implicit Functions

A. General Remarks

Historically, a function y of x was called an implicit function of x if there was given a functional relation $f(x,y)=0$ between x and y, but no explicit representation of y in terms of x (→ 165 Functions). Nowadays, however, the notion of implicit function is rigorously defined as follows: Suppose that a function $f(x_1,\dots,x_n,y)$ is of †class C^1 in a domain G in the real $(n+1)$-dimensional Euclidean space $\mathbf{R}^{n+1}$ and that $f(x_1^0,\dots,x_n^0,y^0)=0$, $f_y(x_1^0,\dots,x_n^0,y^0)\neq 0$ at a point $(x_1^0,\dots,x_n^0,y^0)$ in G. Then there is a unique function $g(x_1,\dots,x_n)$ of class C^1 in a neighborhood of the point $(x_1^0,\dots,x_n^0)$ that satisfies $f(x_1,\dots,x_n,g(x_1,\dots,x_n))=0$, $y^0=g(x_1^0,\dots,x_n^0)$ (**implicit function theorem**). The function g is called the **implicit function** determined by $f=0$. The partial derivatives of g are given by the relation

$$\partial g/\partial x_j=-(\partial f/\partial x_j)/(\partial f/\partial y),$$

where $y=g(x_1,\dots,x_n)$. If the function f is of class C^r $(1\leqslant r\leqslant\infty$ or $r=\omega)$, then the function g is also of class C^r. In particular, when $n=1$, letting x_1 be x, we have $dg/dx=-f_x/f_y$.

B. Jacobian Matrices and Jacobian Determinants

A mapping u from a domain G in $\mathbf{R}^n$ into $\mathbf{R}^m$

$$u(x)=(u_1(x_1,\dots,x_n),\dots,u_m(x_1,\dots,x_n)),$$

$$x=(x_1,\dots,x_n),$$

is called a **mapping of class** C^r if each component $u_1,\dots,u_m$ is of class C^r $(0\leqslant r\leqslant\infty$ or $r=\omega)$ in G. Given a mapping u of class C^1 from G into $\mathbf{R}^m$, we consider the following matrix, which gives rise to the differential du_x of the mapping u (→ 105 Differentiable Manifolds I):

$$\partial(u)/\partial(x)=(\partial u_j/\partial x_k)_{1\leqslant j\leqslant m,\,1\leqslant k\leqslant n}. \qquad (1)$$

This matrix is called the **Jacobian matrix** of the mapping u at x. If there is another mapping v of class C^1 from a domain containing the †range U of u into $\mathbf{R}^l$, then we have the law of composition:

$$(\partial(v)/\partial(u))(\partial(u)/\partial(x))=\partial(v)/\partial(x).$$

When $n=m$, the †determinant of the matrix (1) is called the **Jacobian determinant** (or simply **Jacobian**), and is denoted by $D(u)/D(x)$, $D(u_1,\dots,u_n)/D(x_1,\dots,x_n)$ or

$$\frac{D(u_1,\dots,u_n)}{D(x_1,\dots,x_n)}.$$

Sometimes the notation ∂ is used instead of D; but in the present article we distinguish the matrix from the determinant, using ∂ for the matrix and D for the determinant.

If $m=n$ and $D(u)/D(x)$ never vanishes at any point of the domain G, then u is called a **regular** (or **nonsingular**) **mapping of class** C^1. If the Jacobian $D(u)/D(x)$ is 0 at x, we say that u is **singular** at x. A mapping that is singular at every point in a set $S \subset G$ is said to be **degenerate** on S. For a regular mapping u, the sign of the Jacobian is constant in a connected domain G. If it is positive, the mapping u preserves the orientation of the coordinate system at each point in G, while if it is negative, the mapping changes the orientation. A point where u is degenerate is called a **critical point** of the mapping u, and its image under u is called a **critical value**. In general, the image of the mapping is "folded" along the set of critical points. The set of critical values of a mapping u of class C^1 (sending a domain in $\mathbf{R}^n$ into $\mathbf{R}^n$) is of Lebesgue measure 0 in $\mathbf{R}^n$ (**Sard's theorem**). If u is a regular mapping, then each point in the domain G of u has a neighborhood V such that the restriction of u on V is a †topological mapping. Its inverse mapping $x(u)$ is also a regular mapping of class C^1 and satisfies the relation

$$\frac{D(u)}{D(x)}\frac{D(x)}{D(u)}=1$$

(**inverse mapping theorem**). If u is of class C^r $(1 \leqslant r \leqslant \infty$ or $r=\omega)$, then so is its inverse mapping.

C. Functional Relations

A function $F(u_1, \ldots, u_n)$ defined on a domain B in $\mathbf{R}^n$ is called a **function with scattered zeros** if F has a zero point (i.e., there exists a point u for which $F(u)=0$) and if every open subset of B contains a point u such that $F(u) \neq 0$. Every †analytic function $\not\equiv 0$ has scattered zeros. Let $u(x)$ be a mapping from a domain G in $\mathbf{R}^n$ into $B \subset \mathbf{R}^n$. Suppose that there exists a function $F(u)$ defined in B, of class C^r, with scattered zeros. If $F(u(x))=0$ for every x in G, then we say that the components $u_1, \ldots, u_n$ of the mapping u have a **functional relation of class** C^r or are **functionally dependent of class** C^r. In such a case, we sometimes say simply that $u_1, \ldots, u_n$ are **functionally dependent** or that they have a **functional relation**. If the components $u_1, \ldots, u_n$ of a mapping u of class C^1 are functionally dependent of class C^0, then the Jacobian $D(u)/D(x)$ of u must vanish. Conversely, if the Jacobian $D(u)/D(x)$ of a mapping u of class C^1 is identically 0 in the domain G, the components $u_1, \ldots, u_n$ are functionally dependent of class C^∞ on every compact set in G (**Knopp-Schmidt theorem**) [1].

D. Implicit Functions Determined by Systems of Functions

Suppose that the †rank of the Jacobian matrix of (1) is $r<m$ everywhere in G. Suppose that $u(x)$ is a mapping of class C^1 from a domain G in $\mathbf{R}^n$ into $\mathbf{R}^m$, the Jacobian determinant $D(u_1, \ldots, u_r)/D(x_1, \ldots, x_r)$ never vanishes in G, and further $D(u_1, \ldots, u_r, u_\rho)/D(x_1, \ldots, x_r, x_\sigma)$ is identically 0 in G for each ρ, σ with $r<\rho \leqslant m$, $r<\sigma \leqslant n$. Then the values $u_{r+1}(x), \ldots, u_m(x)$ are determined by the values $u_1(x), \ldots, u_r(x)$, and each u_ρ is represented as a function of class C^1 of $u_1, \ldots, u_r$.

Let $u(x)$ be a mapping of class C^1 from a domain G in $\mathbf{R}^n$ into $\mathbf{R}^m$ and V the †inverse image of a point u^0. To study the properties of the set V, we assume, for simplicity, that u^0 is the origin. Suppose that the rank of the matrix $\partial(u)/\partial(x)$ is r for every point x in G and each of $u_{r+1}, \ldots, u_m$ is functionally dependent on $u_1, \ldots, u_r$. Then each u_ρ $(r<\rho \leqslant m)$ is a function $u_\rho(u_1, \ldots, u_r)$ of $u_1, \ldots, u_r$. The set V is empty if there is a ρ such that $u_\rho(0, \ldots, 0) \neq 0$. On the other hand, if $u_\rho(0, \ldots, 0)=0$ for all ρ $(r<\rho \leqslant m)$, then V is the set of common zero points of the functions $u_1(x), \ldots, u_r(x)$. Therefore, to study the set V, we can assume that $r=m \leqslant n$. If $m=n$, V consists of isolated points only. If $m<n$, then, changing the order of the variables $x_1, \ldots, x_n$ if necessary, we can assume that $D(u_1, \ldots, u_m)/D(x_1, \ldots, x_m) \neq 0$ at a point (x^0) in V. In this case, there is a unique function $\xi_\mu(x_{m+1}, \ldots, x_n)$ of class C^1 $(1 \leqslant \mu \leqslant m)$ in a neighborhood of (x_i^0) satisfying the following two conditions: (i) $x_\mu^0=\xi_\mu(x_{m+1}^0, \ldots, x_n^0)$; (ii) if the point $(x_{m+1}, \ldots, x_n)$ is in a neighborhood of $(x_{m+1}^0, \ldots, x_n^0)$, then the point

$$(\xi_1(x_{m+1}, \ldots, x_n), \ldots, \xi_m(x_{m+1}, \ldots, x_n), \\ x_{m+1}, \ldots, x_n) \in V.$$

Each function ξ_μ is called an **implicit function** of $x_{m+1}, \ldots, x_n$ determined by the relations $u_1=\ldots=u_m=0$. The †total derivatives of the ξ_μ are determined from the system of linear equations

$$\sum_{k=1}^{m}\frac{\partial u_j}{\partial x_k}d\xi_k+\sum_{l=m+1}^{n}\frac{\partial u_j}{\partial x_l}dx_l=0, \qquad j=1, \ldots, m.$$

The foregoing implicit function theorem is a local one. Among the **global implicit function theorems**, the following one, due to Hadamard, is useful: Let $x \mapsto y(x)$ be a mapping of class C^1 from $\mathbf{R}^n$ into $\mathbf{R}^n$ such that the inverse $\Phi^{-1}(x)$

of its Jacobian matrix $\Phi(x)$ is bounded in $\mathbf{R}^n$; then the mapping is a †diffeomorphism of class C^1 from $\mathbf{R}^n$ onto $\mathbf{R}^n$.

The implicit function theorem holds also in complex spaces. Let $f_i(x_1, \ldots, x_n, y_1, \ldots, y_p)$, $1 \leqslant i \leqslant p$, be a system of †holomorphic functions in a domain in the complex $(n+p)$-dimensional space $\mathbf{C}^{n+p}$. If (i) $f_i(x_1^0, \ldots, x_n^0, y_1^0, \ldots, y_p^0) = 0$, $1 \leqslant i \leqslant p$, and (ii) $D(f_1, \ldots, f_p)/D(y_1, \ldots, y_p)_{(x,y)=(x^0,y^0)} \neq 0$, then there exists a unique holomorphic solution $y_i = y_i(x_1, \ldots, x_n)$ $(1 \leqslant i \leqslant p)$ in a neighborhood of the point $x^0 = (x_1^0, \ldots, x_n^0)$ that satisfies $f_i(x_1, \ldots, x_n, y_1(x_1, \ldots, x_n), \ldots, y_p(x_1, \ldots, x_n)) = 0$ $(1 \leqslant i \leqslant p)$.

E. Linear Relations

Suppose that $u(x)$ is a mapping of class C^{m-1} from $\mathbf{R}^1$ into $\mathbf{R}^m$. Its components $(u_1, \ldots, u_m)$ are functions of class C^{m-1}. Then the determinant

$$\begin{vmatrix} u_1 & u_2 & \ldots & u_m \\ u_1' & u_2' & \ldots & u_m' \\ \ldots & \ldots & \ldots & \ldots \\ u_1^{(m-1)} & u_2^{(m-1)} & \ldots & u_m^{(m-1)} \end{vmatrix}$$

is called the **Wronskian determinant** (or simply the **Wronskian**) of the functions $u_1, \ldots, u_m$ and is denoted by $W(u_1, u_2, \ldots, u_m)$. If the functions $u_1, \ldots, u_m$ are †linearly dependent, i.e., if there exist constants c_j not all zero satisfying $\sum_{j=1}^m c_j u_j(x) = 0$ identically, then the Wronskian vanishes identically. Therefore, if $W(u_1, \ldots, u_m) \neq 0$, then the functions $u_1, \ldots, u_m$ are linearly independent. Conversely, if $W(u_1, \ldots, u_m) = 0$ identically, and further if there is at least one nonvanishing Wronskian for $u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_m$ $(1 \leqslant i \leqslant m)$, then the functions $u_1, \ldots, u_m$ are linearly dependent. The necessity of the additional condition is shown by the following example: $u_1 = x^3$ and $u_2 = |x|^3$ are of class C^1 in the interval $[-1, 1]$ and linearly independent, but they satisfy $W(x^3, |x|^3) = 0$ identically. However, the additional condition is unnecessary if the functions $u_1, \ldots, u_m$ are analytic. Similar theorems are valid in a domain in a complex plane.

Furthermore, if $u(x)$ is a continuous mapping from an interval $[a, b]$ in $\mathbf{R}^1$ into $\mathbf{R}^m$, the determinant

$$G(u_1, \ldots, u_m) = \begin{vmatrix} (1,1) & \ldots & (1,m) \\ (2,1) & \ldots & (2,m) \\ \ldots & \ldots & \ldots \\ (m,1) & \ldots & (m,m) \end{vmatrix},$$

$$(j,k) = \int_a^b u_j(x) u_k(x)\, dx$$

is called the **Gramian determinant** (or simply **Gramian**). The Gramian is the †discriminant of the quadratic form

$$\int_a^b \left(\sum_{j=1}^m \xi_j u_j(x) \right)^2 dx$$

of $\xi_1, \ldots, \xi_m$ and is equal to

$$\frac{1}{m!} \int_a^b \ldots \int_a^b (\det(u_j(x_k)))^2\, dx_1 \ldots dx_m.$$

We always have $G(u_1, \ldots, u_m) \geqslant 0$, and $G(u_1, \ldots, u_m) = 0$ if and only if $u_1, \ldots, u_m$ are linearly dependent. The Gramian is defined if the functions $u_1, \ldots, u_m$ are †square integrable in the sense of Lebesgue. In that case, the condition $G(u_1, \ldots, u_m) = 0$ holds if and only if $u_1, \ldots, u_m$ are linearly dependent †almost everywhere, i.e., there are constants $c_1, \ldots, c_m$ not all zero such that the relation $c_1 u_1(x) + \ldots + c_m u_m(x) = 0$ holds except on a set of Lebesgue measure 0.

References

[1] K. Knopp and R. Schmidt, Funktionaldeterminanten und Abhängigkeit von Funktionen, Math. Z., 25 (1926), 373–381.

[2] G. Doetsch, Die Funktionaldeterminante als Deformationsmass einer Abbildung und als Kriterium der Abhängigkeit von Funktionen, Math. Ann., 99 (1928), 590–601.

[3] W. Rudin, Principles of mathematical analysis, McGraw-Hill, second edition, 1964.

[4] H. Cartan, Calcul différentiel, Hermann, 1967.

[5] J. T. Schwartz, Nonlinear functional analysis, Gordon & Breach, 1969.

209 (XXI.5) Indian Mathematics

India was one of the earliest civilizations, but because it has no precise chronological record of ancient times, it is said to possess no history. Indian mathematics seems to have developed under the influence of the cult of Brahma, as did the calendar. It may also have some relation to the mathematics of the Near East and China, but this is difficult to trace. The word *ganita* (computation) appears in early religious writings; after the beginning of the Christian Era, it was classified into *pāṭī-gaṇita* (arithmetic), *bīja-gaṇita* (algebra), and *kṣetra-gaṇita* (geometry), thus showing some systematization. The Buddhists (notably Nagarjuna) had a kind of logic, but it had no re-

lation to mathematics. Unlike the Greeks, the Indians had no demonstrational geometry, but they had symbolic algebra and a position system of numeration.

Indian geometry was computational; Āryabhaṭa (c. 476–c. 550) computed the value of π as 3.1416; Brahmagupta (598–c. 660) had a formula to compute the area of quadrangles inscribed in a circle; and Bhāskara (1114–1185) gave a proof of the Pythagorean theorem. In trigonometry, Āryabhaṭa made a table of sines of angles between 0° and 90° for every 3.75° interval. The name "sine" is related to the Sanskrit *jya*, which referred to half of the chord of the double arc.

The Indians had a remarkable system of algebra. At the beginning they had no operational symbols and described in words the rules for solving equations. Brahmagupta worked on the †Pell equation $ax^2+1=y^2$. Bhâskara knew that a quadratic equation can have two roots that can be positive and negative, but did not assign any meaning to the negative root in such cases. Bhāskara also introduced algebraic symbols.

The symbol 0 was used in India from about 200 B.C. to denote the void place in the position system of numeration; 0 as a number is found in a book by Bakhshāli published in the 3rd century A.D. The number 0 is defined as $a-a=0$ in our notation, and the rules $a\pm 0=a$, $0\times a=0$, $\sqrt{0}=0$, $0\div a=0$ are mentioned. Brahmagupta prohibited division by 0 in arithmetic, but in algebra he called the "quantity" $a\div 0$ *taccheda*. Bhāskara called it *khahara* and made it play a role similar to that of our infinity.

Some historians assert that the Indians had the ideas of infinity and infinitesimal. Some hold that the Indian position system of numeration arose from the circumstance that the names of numbers differed according to their positions. The Indian numeration system was exported to Europe through Arabia, and had great influence on the development of mathematics.

References

[1] M. Cantor, Vorlesungen über Geschichte der Mathematik I, Teubner, third edition, 1907.

[2] G. Sarton, Introduction to the history of science I, From Homer to Omar Khayyam Carnegie Institute of Washington, 1927.

[3] B. Datta and A. N. Singh, History of Hindu mathematics I, II, Lahore, 1935–1938. (Asia Publishing House, 1962).

210 (II.25) Inductive Limits and Projective Limits

A. General Remarks

Inductive and projective limits can be defined over any †preordered set I and in any †category. We first explain the definition of these limits in the special case where I is a †directed set and the category is that of sets, of groups, or of topological spaces. The simplest case is when I is the ordered set $\mathbf{N}$ of the natural numbers.

B. The Limit of Sets

Let I be a directed set. Suppose that we are given a set X_i for each $i\in I$ and a mapping $\varphi_{ji}:X_i\rightarrow X_j$ for each pair (i,j) of elements of I with $i\leqslant j$, such that $\varphi_{ii}=1_{X_i}$ (the identity mapping on X_i) and $\varphi_{ki}=\varphi_{kj}\circ\varphi_{ji}$ $(i\leqslant j\leqslant k)$. Then we denote the system by (X_i, φ_{ji}) and call it an **inductive system** (or **direct system**) of sets over I. Let S be the †direct sum $\coprod_i X_i$ of the sets X_i $(i\in I)$, and define an equivalence relation in S as follows: $x\in X_i$ and $y\in X_j$ are equivalent if and only if there exists a $k\in I$ such that $i\leqslant k, j\leqslant k$, $\varphi_{ki}(x)=\varphi_{kj}(y)$. Let D be the †quotient set of S by this equivalence relation, and let $f_i:X_i\rightarrow D$ $(i\in I)$ be the canonical mappings. Then we have I(1) $f_j\circ\varphi_{ji}=f_i$ $(i\leqslant j)$; I(2) for any set X, and for any system of mappings $g_i:X_i\rightarrow X$ $(i\in I)$ satisfying $g_j\circ\varphi_{ji}=g_i$ $(i\leqslant j)$, there exists a unique mapping $f:D\rightarrow X$ such that $f\circ f_i=g_i$ $(i\in I)$. We call (D,f_i) the **inductive limit** (or **direct limit**) of the inductive system (X_i,φ_{ji}) over I, and denote it by $\varinjlim X_i$ or $\operatorname{ind}\lim X_i$ (more precisely, by $\varinjlim_{i\in I} X_i$ or $\operatorname{ind}\lim_{i\in I} X_i$).

Suppose, dually, that we are given a set X_i for $i\in I$ and a mapping $\psi_{ij}:X_j\rightarrow X_i$ for each $i\leqslant j$, such that $\psi_{ii}=1_{X_i}$ and $\psi_{ik}=\psi_{ij}\circ\psi_{jk}$ $(i\leqslant j\leqslant k)$. Then we denote the system by (X_i, ψ_{ij}) and call it a **projective system** (or **inverse system**) of sets over I. Let P be the subset of the Cartesian product $\prod X_i$ defined by $P=\{(x_i)\mid\psi_{ij}(x_j)=x_i\ (i\leqslant j)\}$, and let $p_i:P\rightarrow X_i$ be the canonical mappings. Then we have P(1) $\psi_{ij}\circ p_j=p_i$ $(i\leqslant j)$; P(2) for any set X, and for any system of mappings $q_i:X\rightarrow X_i$ satisfying $\psi_{ij}\circ q_jq_i$ $(i\leqslant j)$, there exists a unique mapping $p:X\rightarrow P$ such that $p_i\circ p=q_i$ $(i\in I)$. We call (P,p_i) the **projective limit** (or **inverse limit**) of the projective system (X_i,ψ_{ij}) over I and denote it by $\varprojlim X_i$ or $\operatorname{proj}\lim X_i$.

Note that we may replace I by any cofinal subset of I without changing the limits.

C. The Limit of Groups and of Topological Spaces

If, in the notation of Section B, X_i is a group and φ_{ji} (ψ_{ij}) is a homomorphism, then we say that (X_i, φ_{ji}) ((X_i, ψ_{ij})) is an **inductive (projective) system of groups**. The inductive limit (as a set) $D = \varinjlim X_i$ has the structure of a group for which the canonical mappings f_i are homomorphisms. With this group structure, D is called the **inductive limit (group)** of the inductive system of groups. It satisfies properties I(1) and I(2) with group X and homomorphisms g_i and f. Similarly, the projective limit (as a set) $P = \varprojlim X_i$ has a unique group structure such that each $p_i: P \to X_i$ is a homomorphism, namely, that of a subgroup of the direct product group $\prod X_i$. The group P is called the **projective limit (group)** of the projective system of groups. When each X_i is a module over a fixed ring A, we get entirely similar results by considering A-homomorphisms instead of group homomorphisms.

Next, let X_i be a topological space and φ_{ji} and ψ_{ij} be continuous mappings. Then (X_i, φ_{ji}) ((X_i, ψ_{ij})) is called an **inductive (projective) system of topological spaces**. If we introduce in $D = \varinjlim X_i$ the topology of a quotient space of the †topological direct sum of the spaces X_i ($i \in I$), then the f_i are continuous, and I(1) and I(2) hold with sets and mappings replaced by topological spaces and continuous mappings. Similarly, if we view $P = \varprojlim X_i$ as a subspace of the †product space ΠX_i, then the p_i are continuous and P(1) and P(2) hold with the same modification as before. The spaces D and P are called the **inductive limit (space)** and the **projective limit (space)** of the system of topological spaces, respectively. The projective limit of Hausdorff (compact) spaces is also Hausdorff (compact).

Furthermore, if the X_i ($i \in I$) form a topological group and φ_{ji}, ψ_{ij} are continuous homomorphisms, then $\varinjlim X_i$ and $\varprojlim X_i$ are topological groups, and properties I(1), I(2), P(1), and P(2) are satisfied for topological groups and continuous homomorphisms (→ 423 Topological Groups). In particular, projective limits of finite groups are †totally disconnected compact groups and are called **profinite groups**; they occur, e.g., as the ring of †p-adic integers and as the †Galois group of an infinite †Galois extension. Conversely, the †germs of continuous functions at a point x in a topological space X, and other kinds of germs (→ 383 Sheaves), are important examples of inductive limits of groups.

D. Limits in a Category

Let I be a preordered set and $\mathscr{C}$ a category. If we are given an object X_i of a category $\mathscr{C}$ for each $i \in I$ and a †morphism $\varphi_{ji}: X_i \to X_j$ of $\mathscr{C}$ for each pair (i,j) of elements of I with $i \leqslant j$, and if the conditions $\varphi_{ii} = 1_{X_i}$, $\varphi_{ki} = \varphi_{kj} \circ \varphi_{ji}$ ($i \leqslant j \leqslant k$) are satisfied, then we call the system (X_i, φ_{ji}) an **inductive system** over I in the category $\mathscr{C}$. A **projective system** in $\mathscr{C}$ is defined dually: It is an inductive system in the †dual category $\mathscr{C}^\circ$. If we view I as a category (→ 52 Categories and Functors B), then an inductive (projective) system in $\mathscr{C}$ over the index set I is a †covariant (†contravariant) functor from I to $\mathscr{C}$. Now if an object $D \in \mathscr{C}$ and morphisms $f_i: X_i \to D$ ($i \in I$) satisfy conditions I(1) and I(2) with the modification that X is an object and g_i, f are morphisms in $\mathscr{C}$, then the system (D, f_i) is called the **inductive limit** of (X_i, φ_{ji}) and is denoted by $\varinjlim X_i$. Similarly, if an object $P \in \mathscr{C}$ and morphisms $p_i: P \to X_i$ ($i \in I$) satisfy P(1) and P(2) with a similar modification, then (P, p_i) is called the **projective limit** of (X_i, ψ_{ij}) and is written $\varprojlim X_i$. By I(2) and P(2), these limits are unique if they exist.

In the categories of sets, of groups, of modules, and of topological spaces, inductive and projective limits always exist. Note that if the ordering of I is such that $i \leqslant j$ implies $i = j$, i.e., if there is no ordering between two distinct elements of I, then the inductive (projective) limit is the †direct sum (†direct product) (→ 52 Categories and Functors E).

Let (X_i, φ_{ji}), (X'_i, φ'_{ji}) be two inductive systems over the same index set I, and let φ_i: $X_i \to X'_i$ ($i \in I$) be morphisms satisfying $\varphi'_{ji} \circ \varphi_i = \varphi_j \circ \varphi_{ji}$ ($i \leqslant j$). Then the system (φ_i) is called a **morphism** between the inductive systems. Such a morphism is a †natural transformation between the inductive systems viewed as functors $I \to \mathscr{C}$. If $\varinjlim X_i$ and $\varinjlim X'_i$ exist, then (φ_i) induces a morphism $\varinjlim \varphi_i: \varinjlim X_i \to \varinjlim X'_i$ in a natural way, and similarly for projective limits.

For the more abstract notion of limit of a functor → [5]. For the theory of procategories → [4].

References

[1] S. Lefschetz, Algebraic topology, Amer. Math. Soc. Colloq. Publ., 1942.
[2] S. Eilenberg and N. Steenrod, Foundations of algebraic topology, Princeton Univ. Press, 1952.
[3] H. Cartan and S. Eilenberg, Homological algebra, Princeton Univ. Press, 1956.
[4] A. Grothendieck, Technique de descente et théorèmes d'existence en géométrie algébrique II. Le théorème d'existence en théorie formelle

des modules, Sém. Bourbaki, Exposé 195, 1959/1960 (Benjamin, 1966).
[5] M. Artin, Grothendieck topology, Lecture notes, Harvard Univ. 1962.

211 (X.3) Inequalities

A. General Remarks

In this article we consider various properties of inequalities between real numbers. An inequality that holds for every real number (e.g., $x^2 \geqslant 0$) is called an **absolute inequality**; otherwise it is called a **conditional inequality**. When we are given a conditional inequality (e.g., $x(x-1)<0$), the set of all real numbers that satisfy it is called the **solution** of the inequality. The process of obtaining a solution is called **solving** the conditional inequality.

B. Solution of a Conditional Inequality

Suppose that a conditional inequality is given by $f(x)>0$ (or $f(x)\geqslant 0$), where f is a continuous function defined for every real number. If the equation $f(x)=0$ has no solution, then we have either $f(x)>0$ or $f(x)<0$ for all x. On the other hand, if α and β ($\alpha<\beta$) are adjacent roots of the equation $f(x)=0$, the sign of $f(x)$ is unchanged in the open interval (α, β). Therefore the solution of the given inequality depends essentially on the solution of the equation $f(x)=0$. If inequalities involve two variables x, y and are given by $f(x,y)>0$, $g(x,y)>0$ for continuous functions f and g, the solution is, in general, a domain in the xy-plane bounded by the curves $f(x,y)=0$ and $g(x,y)=0$. Similar results hold for the case of inequalities involving more than two variables.

C. Famous Absolute Inequalities

(1) Inequalities concerning **means** (or **averages**): Suppose that we are given an n-tuple $a=(a_1, \ldots, a_n)$, $a_r \geqslant 0$. We set

$$M_r = M_r(a) = \left(\frac{1}{n}\sum_{v=1}^{n} a_v^r\right)^{1/r}.$$

If at least one a_r is 0 and $r<0$, we put $M_r=0$. In particular, we put

$$A = M_1,$$

$$G = \lim_{r\to\infty} M_r = \left(\prod_{v=1}^{n} a_v\right)^{1/n},$$

$$H = M_{-1};$$

these are called the **arithmetic mean**, **geometric mean**, and **harmonic mean** of a_v ($v=1, \ldots, n$), respectively. Except when either all a_v are identical or some a_v is 0 and $r\leqslant 0$, the function M_r increases †strictly monotonically as r increases, and $M_r \to \min a_v$ ($r\to -\infty$), $M_r \to \max a_v$ ($r \to +\infty$). Therefore we always have $\min a_v \leqslant M_r \leqslant \max a_v$. In particular, we have $H<G<A$ if the a_v are all positive and not all equal.

Let $p(x)$ (>0), $f(x)$ $(\geqslant 0)$ be †integrable functions on a †measurable set E. We put

$$M_r(f) = \left(\int_E pf^r dx \Big/ \int_E p\,dx\right)^{1/r}, \qquad r\neq 0.$$

Furthermore, if $M_r(f)$ is strictly positive for some $r>0$, we put

$$M_0(f) = \lim_{r\to+0} M_r(f)$$

$$= \exp\left(\int_E p \log f dx \Big/ \int_E p\,dx\right).$$

We call $M_r(f)$ the **mean of degree** r of the function $f(x)$ with respect to the **weight function** $p(x)$. It has properties similar to those of $M_r(a)$. In particular, when the weight function $p=1$, the means $M_1(f)$, $M_0(f)$, $M_{-1}(f)$ are called the **arithmetic mean**, **geometric mean**, and **harmonic mean** of f, respectively.

(2) The Hölder inequality: Suppose that $p\neq 0, 1$ and $(p-1)(q-1)=1$; that is, $1/p+1/q=1$, and $a_v>0$, $b_v>0$. Then, in general, we have the **Hölder inequality**:

$$\sum_r a_v b_v \gtrless \left(\sum_v a_v^p\right)^{1/p}\left(\sum_v b_v^q\right)^{1/q}, \qquad p\lessgtr 1,$$

where the inequality signs in the first inequality are taken in accordance with $p<1$ or $p>1$. The summation may be infinite if the sums are convergent. The inequality sign is replaced by the equality sign if and only if there exist constant factors λ and μ such that $\lambda a_v^p = \mu b_v^q$ for all v. The Hölder inequality for $p=q=2$ is called the **Cauchy inequality** (or **Cauchy-Schwarz inequality**).

For two measurable positive functions $f(x)$, $g(x)$, we have the **Hölder integral inequality**:

$$\int_E fg\,dx \gtrless \left(\int_E f^p dx\right)^{1/p}\left(\int_E g^q dx\right)^{1/q}, \qquad p\lessgtr 1,$$

except when there exist two constant factors λ and μ ($(\lambda,\mu)\neq(0,0)$) such that $\lambda f^p = \mu g^q$ holds †almost everywhere. The above inequality is replaced by equality if and only if we are in this exceptional case. The case where $p=q=2$ is called the **Schwarz inequality** (or **Bunyakovskiĭ inequality**).

(3) The Minkowski inequality: Suppose that $p\neq 0, 1$ and $a_v>0$, $b_v>0$. Then we have the

Minkowski inequality:

$$\left(\sum_\nu (a_\nu+b_\nu)^p\right)^{1/p} \gtrless \left(\sum_\nu a_\nu^p\right)^{1/p} + \left(\sum_\nu b_\nu^p\right)^{1/p}, \qquad p \lessgtr 1,$$

except when $\{a_\nu\}$ and $\{b_\nu\}$ are proportional. The inequality is replaced by equality if and only if we are in the exceptional case.

The corresponding integral inequality for positive functions $f(x)$, $g(x)$ is

$$\left(\int_E (f+g)^p dx\right)^{1/p} \gtrless \left(\int_E f^p dx\right)^{1/p} + \left(\int_E g^p dx\right)^{1/p}, \qquad p \lessgtr 1,$$

except when $f(x)/g(x)$ is constant almost everywhere. The inequality is replaced by equality if and only if we are in the exceptional case.

D. Related Topics

Absolute inequalities are important in analysis, especially in connection with techniques to prove convergence or for error estimates. However, there seldom are general principles for deriving such inequalities, except for a few elementary theorems.

For other famous inequalities → Appendix A, Table 8. For related topics → 88 Convex Analysis; for convex functions and their applications → 255 Linear Programming; for linear inequalities → 212 Inequalities in Physics.

References

[1] G. H. Hardy, J. E. Littlewood, and G. Pólya, Inequalities, Cambridge Univ. Press, first edition, 1934; revised edition, 1952.

[2] V. I. Levin and S. B. Stechkin (Stečkin), Inequalities, Amer. Math. Soc. Transl., (2) 14 (1960), 1–29 (English translation of the appendix to the Russian translation of [1]).

[3] E. F. Beckenbach and R. Bellman, Inequalities, Erg. Math., Springer, second revised printing, 1965.

[4] E. F. Beckenbach and R. Bellman, An introduction to inequalities, Random House, 1961.

[5] N. D. Kazarinoff, Geometric inequalities, Random House, 1961.

[6] N. D. Kazarinoff, Analytic inequalities, Holt, Rinehart and Winston, 1961.

[7] W. Walter, Differential- und Integral-Ungleichungen, Springer, 1964.

[8] O. Shisha (ed.), Inequalities, Academic Press, I, 1967; II, 1970; III, 1972.

212 (XX.36) Inequalities in Physics

A. Correlation Inequalities

Let μ be a probability measure on a space X (with σ-field $\mathscr{B}$) and f_j be measurable functions on X, and write

$$\langle f_1 \ldots f_n \rangle = \int_X f_1(x) \ldots f_n(x)\, d\mu(x).$$

A number of inequalities among such expectations under a variety of conditions on the measure μ and functions f_j are known as correlation inequalities, after their original occurrence for correlation functions in the statistical mechanics of lattice gases.

The earliest results were obtained by Griffiths (*J. Math. Phys.*, 8 (1967)), for the case where X is the product $\prod_{i=1}^N I_i$ of two point sets $I_i = \{1, -1\}$, the f are kth coordinate functions $\sigma_k = \sigma_k(x)$ $(=\pm 1)$ of $x \in X$, $d\mu(x)$ is the counting measure multiplied by a Gibbs factor $Z^{-1}e^{-\beta H(x)}$ with a ferromagnetic Hamiltonian

$$H(x) = -\sum_{i<j}^N J_{ij}\sigma_i(x)\sigma_j(x), \qquad J_{ij} = J_{ji} \geq 0,$$

$\beta \geq 0$, and the normalization factor $Z = \sum_x e^{-\beta H(x)}$. His conclusions are

$\langle \sigma_k \sigma_l \rangle \geq 0$ (Griffiths's first inequality),

$\langle \sigma_k \sigma_l \sigma_m \sigma_n \rangle \geq \langle \sigma_k \sigma_l \rangle \langle \sigma_m \sigma_n \rangle$ (Griffiths's second inequality).

These were extended by Kelly and Sherman (*J. Math. Phys.*, 9 (1968)) as

$\langle \sigma_A \rangle \geq 0$ (GKS first inequality),

$\langle \sigma_A \sigma_B \rangle \geq \langle \sigma_A \rangle \langle \sigma_B \rangle$ (GKS second inequality),

where A and B are subsets of $\{1, \ldots, N\}$, $\sigma_A = \prod_{i \in A} \sigma_i$, and $H = -\sum_{A \subset N} J_A \sigma_A$ with $J_A \geq 0$. A further generalization (for example, $I_i = \mathbf{R}$) can be found in [1, 2].

Under the same situation with

$$H = -\sum_{i<j}^N J_{ij}\sigma_i\sigma_j - \sum_{i=1}^N h_i\sigma_i, \quad J_{ij} \geq 0,\ h_i \geq 0,$$

the following GHS inequality by Griffiths, Hurst, and Sherman (*J. Math. Phys.*, 11 (1970)) holds:

$$\langle \sigma_i\sigma_j\sigma_k \rangle - \langle \sigma_i \rangle \langle \sigma_j\sigma_k \rangle - \langle \sigma_j \rangle \langle \sigma_i\sigma_k \rangle - \langle \sigma_k \rangle \langle \sigma_i\sigma_j \rangle + 2\langle \sigma_i \rangle \langle \sigma_j \rangle \langle \sigma_k \rangle \leq 0.$$

If $h_i = 0$ for all i, then

$$\langle \sigma_i\sigma_j\sigma_k\sigma_l \rangle - \langle \sigma_i\sigma_l \rangle \langle \sigma_j\sigma_k \rangle - \langle \sigma_j\sigma_l \rangle \langle \sigma_i\sigma_k \rangle - \langle \sigma_k\sigma_l \rangle \langle \sigma_i\sigma_j \rangle + 2\langle \sigma_i\sigma_l \rangle \langle \sigma_j\sigma_l \rangle \langle \sigma_k\sigma_l \rangle \leq 0.$$

Further generalizations can be found in [2, 3], and the references quoted therein.

Let X be a finite distributive lattice, μ be a positive measure satisfying the condition $\mu(x\wedge y)\mu(x\vee y)\geqslant\mu(x)\mu(y)$, and f and g be both increasing or both decreasing functions on X. Then the following FKG inequality by Fortuin, Kasteleyn, and Ginibre (*Comm. Math. Phys.*, 22 (1971)) holds:

$$\langle fg\rangle\geqslant\langle f\rangle\langle g\rangle.$$

B. Inequalities Involving Traces of Matrices

Let $\operatorname{tr}A$ denote the trace of a matrix A. Some of the earlier results applied to statistical mechanics are as follows, where $\rho\geqslant 0$, $\sigma\geqslant 0$, $A^*=A$, $B^*=B$:

$$\operatorname{tr}(\rho\log\rho-\rho\log\sigma-\rho+\sigma)\geqslant 0$$

(Klein inequality),

$$\operatorname{tr}(e^{A+B})/\operatorname{tr}e^A\geqslant\exp(\operatorname{tr}e^A B/\operatorname{tr}e^A)$$

(Peierls-Bogolyubov inequality),

$$\operatorname{tr}(e^A e^B)\geqslant\operatorname{tr}(e^{A+B})$$

(Golden-Thompson inequality).

The following inequality is related to the last inequality and was proved for a general $m\geqslant 0$ by Lieb and Thirring (*Studies in Mathematical Physics*, Lieb, Simon, and Wightman (eds.), Princeton Univ. Press, 1976):

$$\operatorname{tr}(\rho\sigma)^m\leqslant\operatorname{tr}(\rho^m\sigma^m),\quad \rho\geqslant 0,\ \sigma\geqslant 0.$$

For the Hilbert-Schmidt norm $\|A\|_{\mathrm{H.S.}}\equiv(\operatorname{tr}A^*A)^{1/2}$ and the trace-class norm $\|A\|_{\mathrm{tr}}\equiv\operatorname{tr}|A|$, where $|A|=(A^*A)^{1/2}$, the Powers-Størmer inequality (*Comm. Math. Phys.*, 16 (1970)) holds:

$$\|\rho^2-\sigma^2\|_{\mathrm{tr}}\geqslant(\|\rho-\sigma\|_{\mathrm{H.S.}})^2,\quad \rho\geqslant 0,\ \sigma\geqslant 0.$$

Araki and Yamagami (*Comm. Math. Phys.*, 81 (1981)) give

$$\||C|-|D|\|_{\mathrm{H.S.}}\leqslant\sqrt{2}\|C-D\|_{\mathrm{H.S.}}.$$

If $C^*=C$ and $D^*=D$, then $\sqrt{2}$ can be removed.

The entropy $S(\rho)=-\operatorname{tr}\rho\log\rho$ is a concave function of $\rho\geqslant 0$ satisfying the triangular inequality (Araki and Lieb, *Comm. Math. Phys.*, 18 (1970)):

$$S(\rho_1)+S(\rho_2)\geqslant S(\rho_{12})\geqslant|S(\rho_1)-S(\rho_2)|$$

and strong subadditivity (Lieb and Ruskai, *J. Math. Phys.*, 14 (1973)):

$$S(\rho_{123})-S(\rho_{12})-S(\rho_{13})+S(\rho_1)\leqslant 0,$$

where ρ_{123} is a matrix on the tensor product space $H_1\otimes H_2\otimes H_3$, $\rho_{jk}=\operatorname{tr}_i\rho_{123}$, $\rho_i=\operatorname{tr}_j\operatorname{tr}_k\rho_{123}$ ($\{i,j,k\}=\{1,2,3\}$) and tr_j is a (partial) trace taken only on the space H_j. The last inequality is based on the concavity of the function $f(\rho)=\operatorname{tr}\exp(A+\log\rho)$ in $\rho\geqslant 0$ ($A^*=A$) proved by Lieb [4] (also see Epstein, *Comm. Math. Phys.*, 31 (1973)), who also proved the joint concavity of $\operatorname{tr}(C^*\rho^s C\sigma^r)$ ($r\geqslant 0, s\geqslant 0, r+s\leqslant 1$) in $\rho\geqslant 0$ and $\sigma\geqslant 0$. This concavity for the case $r=s=1/2$ was previously proved by Wigner and Yanase (*Proc. Nat. Acad. Sci. US*, 49 (1963)), and the general case has been conjectured by Wigner, Yanase, and Dyson. It leads to the joint concavity of the relative entropy $S(\rho,\sigma)=\operatorname{tr}\rho(\log\rho-\log\sigma)$ (defined to be $+\infty$ if $\sigma\psi=0$ and $\rho\psi\neq 0$ for some vector ψ) in $\rho\geqslant 0$ and $\sigma\geqslant 0$ as well as its monotonicity $S(\alpha(\rho),\alpha(\sigma))\leqslant S(\rho,\sigma)$ for any trace-preserving expectation mapping α. (For entropy, see also [5].)

The above results have generalizations in the context of von Neumann algebras (Araki, *Publ. Res. Inst. Math. Sci.*, 11 (1975); 13 (1977); Uhlmann, *Comm. Math. Phys.*, 54 (1977); [6]).

C. Operator Monotone and Operator Convex Functions

A real-valued function $f(x)$ defined on an interval I (finite or infinite; open, half-open, or closed) is called matrix monotone increasing (decreasing) of order m if $f(A)\geqslant f(B)$ whenever $m\times m$ Hermitian matrices A and B with their eigenvalues contained in I satisfy $A\geqslant B$ ($A\leqslant B$) and is called operator monotone if it is matrix monotone of order n for all positive integers n. $f(x)$ is called matrix convex of order m if $f[(1-t)A+tB]\leqslant(1-t)f(A)+tf(B)$ for $0\leqslant t\leqslant 1$ and all $m\times m$ Hermitian matrices A and B with their eigenvalues in I and is called operator convex if it is matrix convex of order n for any positive integer n. The functions x^α ($0\leqslant\alpha\leqslant 1$), $\log x$, and $-(x+a)^{-1}$ ($a\geqslant 0$) are all operator monotone increasing in the half-line interval $x>0$. The functions $(x+a)^{-1}$ and $x\log x$ are operator convex in the same interval.

A function $f(x)$ is operator monotone increasing in an open interval (a,b) if and only if it is analytic in (a,b) and has an analytic continuation to the whole upper half-plane where the function has a nonnegative imaginary part [7]. Another necessary and sufficient condition is

$$\sum_{i,k=0}^{N} f^{(i+k+1)}(x)\zeta_i\zeta_k/(i+k+1)!\geqslant 0$$

for all $x\in(a,b)$, for all real ζ, and for all positive integers N. An operator monotone function $f(x)$ on an open interval $(-R,R)$ has the integral representation

$$f(x)=f(0)+f'(0)\int x/(1-tx)\,d\mu(t),$$

where μ is a probability measure with its support contained in $[-R^{-1}, R^{-1}]$ and, if f is not a constant, is uniquely determined by f. The set of all extremal points of the set of all operator monotone increasing functions on $(-1,1)$ satisfying $f(0)=0$ and $f'(0)=1$ is exactly the set of functions $x/(1-tx)$, $|t|\leqslant 1$, which appear as integrands in the above formula. An operator monotone function on **R** must be linear.

An operator convex function in an open interval $(-R, R)$ has the integral representation

$$f(x)=f(0)+f'(0)x+(f''(0)/2)\int x^2/(1-tx)\,d\mu(t),$$

where μ is a probability measure with its support contained in $[-R^{-1}, R^{-1}]$ and, if f is not linear, is uniquely determined by f. An operator convex function in **R** must be at most a quadratic polynomial satisfying $f''(0)\geqslant 0$.

For a continuous real function on an interval $[0,\alpha)$ $(0\leqslant\alpha\leqslant\infty)$, the following four conditions are mutually equivalent: (i) f is operator convex and $f(0)\leqslant 0$, (ii) $f(a^*Aa)\leqslant a^*f(A)a$ for any self-adjoint operator A with its spectrum in $[0,\alpha)$ and any operator a with its norm not exceeding 1, (iii) the preceding condition with a limited to projections, (iv) $x^{-1}f(x)$ is operator monotone increasing in an open interval $(0,\alpha)$. (See [8–11] and Hansen and Pedersen, Jensen's inequality for operators and Löwner's theorem, *Math. Ann.*, 258 (1982), 229–241.

References

[1] J. Ginibre, General formulation of Griffiths' inequality, Comm. Math. Phys., 16 (1970), 310–328.
[2] B. Simon, Functional integration and quantum physics, Academic Press, 1979, ch. 4.
[3] C. Newman, Gaussian correlation inequalities, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 33 (1975), 75–93.
[4] E. H. Lieb, Convex trace functions and the Wigner-Yanase-Dyson conjecture, Advances in Math., 11 (1973), 267–288.
[5] A. Wehrl, General properties of entropy, Rev. Mod. Phys., 50 (1978), 221–260.
[6] H. Araki, Inequalities in von Neumann algebras, CNRS Publication R.C.P., 25 vols., 22 (1975), 1–25.
[7] K. Löwner, Über monotone Matrixfunktionen, Math. Z., 38 (1934), 177–216.
[8] J. Bendat and S. Sherman, Monotone and convex operator functions, Trans. Amer. Math. Soc., 79 (1955), 58–71.
[9] W. Donoghue, Monotone matrix functions and analytic continuation, Springer, 1974.
[10] C. Davis, Notions generalizing convexity for functions defined on spaces of matrices, Amer. Math. Soc. Proc. Symposia in Pure Math., 7 (1963), 187–201.
[11] F. Kraus, Über konvexe Matrixfunktionen, Math. Z., 41 (1936), 18–42.

213 (XIX.12) Information Theory

A. General Remarks

The mathematical theory of information transmission in communication systems, first developed by C. E. Shannon [1] and now called **information theory**, is one of the most important fields of mathematical science. It consists of two main themes, **channel coding theory** and **source coding theory**. The purpose of channel coding theory is to ascertain the existence of schemes for transmitting information or data reliably over a **noisy channel** at a fixed transmission rate. The purpose of source coding theory is to show the existence of schemes for compressing data emitted from an **information source** and reproducing them within tolerable limits of distortion. Both theories are based profoundly on the theory of probability, statistics, and the theory of stochastic processes.

B. Entropy

In information theory it is customary to call a finite set $A=\{\alpha_1,\ldots,\alpha_n\}$ an **alphabet** and to call its elements the **letters** of the alphabet. The simplest model of information sources consisting of an alphabet A and a †probability distribution p over A is denoted by $X=[A,p]$, which can be regarded as a †finite probability space, where the probability of outcome of a letter $\alpha_i\in A$ from the source is denoted by $p_i=p(\alpha_i)$. A real-valued function defined by $I(\alpha_i)=-\log p(\alpha_i)$ is called the **self-information** of the event $\alpha_i\in A$. The †mean value of the self-information, i.e.,

$$H(X)=\sum_{i=1}^{n}-p(\alpha_i)\log p(\alpha_i),$$

is called the **entropy** of the source. This can be interpreted as a measure of the average a priori uncertainty as to which letter will emanate from the source, or a measure of the average amount of information one obtains upon receiving a single letter from the source. The unit of information or entropy for base e logarithms is called a **nat**, while that for base 2 logarithms is called a **bit**.

Let $A=\{\alpha_1,\ldots,\alpha_n\}$ and $B=\{\beta_1,\ldots,\beta_m\}$ be two alphabets. Let $r(\alpha_i,\beta_j)$ be a †joint probability distribution defined on the product AB, and denote the †probability space by $XY=[AB,r]$. Then the joint distribution gives rise to †marginal distributions $p(\alpha_i)=\sum_{j=1}^{m} r(\alpha_i,\beta_j)$ and $q(\beta_j)=\sum_{i=1}^{n} r(\alpha_i,\beta_j)$ and to †conditional distributions $p(\alpha_i\,|\,\beta_j)=r(\alpha_i,\beta_j)/q(\beta_j)$ and $q(\beta_j\,|\,\alpha_i)=r(\alpha_i,\beta_j)/p(\alpha_i)$. Probability spaces $X=[A,p]$ and $Y=[B,q]$ are subspaces of the probability space $XY=[AB,r]$. The entropy of $XY=[AB,r]$ is defined by

$$H(XY)=\sum_{i=1}^{n}\sum_{j=1}^{m} -r(\alpha_i,\beta_j)\log r(\alpha_i,\beta_j)$$

as well. A real-valued function $I(\alpha_i\,|\,\beta_j)=-\log p(\alpha_i\,|\,\beta_j)$ is called the **conditional self-information**. The average of $I(\alpha_i\,|\,\beta_j)$ over α_i defined by

$$H(X\,|\,\beta_j)=\sum_{i=1}^{n} p(\alpha_i)I(\alpha_i\,|\,\beta_j)$$

is called the **conditional entropy** of X for given $\beta_j\in B$. The average of $H(X\,|\,\beta_j)$ over β_j defined by

$$H(X\,|\,Y)=\sum_{j=1}^{m} q(\beta_j)H(X\,|\,\beta_j)$$

is called the conditional entropy of X for given Y. The conditional entropy of Y for given X, denoted by $H(Y|X)$, is defined similarly. Then it can be shown that

$$H(XY)=H(Y)+H(X\,|\,Y)=H(X)+H(Y|X),$$

$$H(XY)\leqslant H(X)+H(Y),$$

$$H(X\,|\,Y)\leqslant H(X),$$

$$H(Y|X)\leqslant H(Y),$$

where equalities in the last three expressions hold if and only if $r(\alpha_i,\beta_j)=p(\alpha_i)q(\beta_j)$ for all $\alpha_i\in A$ and $\beta_j\in B$.

C. Information Sources

Given a finite alphabet A, we consider the finite product set $A^N=A\times\ldots\times A$ and the doubly infinite product $A^Z=\prod_{k=-\infty}^{\infty} A_k$, where $A_k=A$, $k=0,\pm1,\pm2,\ldots$. Let $\mathcal{F}_A$ be the σ-algebra generated by all †cylinder sets in A^Z. Given a †probability measure P over $\mathcal{F}_A$, an information source is defined as a probability space $[A^Z,P]$ or as a †random process $\mathbf{X}=\ldots X_{-1}X_0X_1X_2\ldots$. When P is invariant under the †shift transformation T on A^Z, i.e., $P(E)=P(TE)$ for any $E\in\mathcal{F}_A$, then $[A^Z,P]$ is said to be a **stationary source**. In particular, if $P(E)=0$ or 1 whenever $TE=E\in\mathcal{F}_A$, then the information source is said to be **ergodic**. If $\mathbf{X}$ is an †independently and identically distributed random process, the source $\mathbf{X}=[A^Z,P]$ is said to be **memoryless**.

For a given $\mathbf{X}$, denote the subsequence $X_1\ldots X_N$ of $\mathbf{X}$ by X^N. Then a probability measure P on A^N and a finite probability space $X^N=[A^N,P^N]$ are naturally induced. The entropy of the stationary information source $\mathbf{X}=[A^Z,P]$ is defined as $H(\mathbf{X})=\lim_{N\to\infty}H(X^N)/N$ or as $H(\mathbf{X})=\lim_{N\to\infty}H(X_N\,|\,X^{N-1})$, because both limits exist and are identical. If $\mathbf{X}$ is memoryless, the entropy of $\mathbf{X}$ is equivalent to that of $X_1=[A^1,P^1]$, i.e., $H(\mathbf{X})=\sum_{i=1}^{n}-p_i\log p_i$, where $p_i=p^1(X_1=\alpha_i)$ and $H(X^N)=NH(X^1)=NH(\mathbf{X})$.

D. Source Coding Theorem

Let $x^N=x_1\ldots x_N$ be a sequence of N consecutive letters from the source $[A^Z,P]$. Suppose that we wish to encode such sequences into fixed-length code words $u^L=u_1\ldots u_L$ consisting of L letters from a code alphabet U of size v. The number $R=(\log v^L)/N$ is called the **coding rate** per source letter. A mapping φ: $A^N\to U^L$ is called an encoder and $\varphi: U^L\to A^N$ a decoder. The set U^L with specified encoder-decorder pair $[\varphi,\psi]$ is called a **code** with rate $R=(L\log v)/N$. The **error probability** of the code is defined by

$$P_e[\varphi,\psi]=P[\psi\{\varphi(X^N)\neq X^N\}].$$

The fixed-length **source coding theorem** [1, 2] states: Let $\mathbf{X}=[A^Z,P]$ be a stationary ergodic source. Then for any $\delta>0$, if $R\geqslant H(\mathbf{X})+\delta$, there exists a code $[\varphi,\psi]$ with rate R such that the error probability $P_e[\varphi,\psi]$ can be made arbitrarily small by making N sufficiently large. Conversely, if $R\leqslant H(\mathbf{X})-\delta$ then for any code with rate R, $P_e[\varphi,\psi]$ must become arbitrarily close to 1 as $N\to\infty$.

This theorem is trivial if $R\geqslant\log n$. For memoryless sources, the theorem follows immediately from the †weak law of large numbers, which implies that for arbitrary $\varepsilon>0$ and $\delta>0$ there exists an integer $N_o(\varepsilon,\delta)$ such that for all $N>N_o(\varepsilon,\delta)$

$$P\left[\left|\frac{-\log P^N(X^N)}{N}-H(\mathbf{X})\right|>\delta\right]<\varepsilon.$$

The validity of this property for stationary ergodic sources was proved by B. McMillan [2]. In case of memoryless sources, the exact asymptotic form of the error probability for an optimal code with rate R was given by F. Jelinek [3], I. Csiszár and G. Longo [4], Blahut [5], and Longo and A. Sgarro [6] as

$$\lim_{N\to\infty}-\frac{1}{N}\log P_e[\varphi,\psi]=\min_{q:H(q)\geqslant R} D(q\,\|\,p),$$

where p is the source distribution, q denotes a

probability distribution on A, and

$$D(q\|p)=\sum_{i=1}^{n} q_i \log\frac{q_i}{p_i},$$

which is called **Kullback's discrimination information** [7] or the **divergence**.

Source sequences x^N can be encoded into variable-length code words consisting of letters from alphabet U of size v. A set of n^N code words is called a **prefix condition code** if there is no code word which is equivalent to the prefix of any other code word. Denote the length of the code word corresponding to a source sequence x^N by $L(x^N)$ and the average length of the code words per source letter by

$$\frac{1}{N}\bar{L}=\sum_{x^N}\frac{1}{N}P^N(x^N)L(x^N).$$

The variable length source coding theorem states: Given a memoryless source $\mathbf{X}$ with entropy $H(\mathbf{X})$, there is a prefix condition code such that the average length of the code words per source letter satisfies

$$H(\mathbf{X})\leqslant \bar{L}\log v\leqslant H(\mathbf{X})+\frac{1}{N}\log v.$$

This theorem is valid for stationary ergodic sources if $(\log v)/N$ in the last term is replaced by $\varepsilon(N)$, where $\varepsilon(N)\to 0$ as $N\to\infty$.

These two theorems are referred to as **noiseless source coding theorems.**

E. Source Coding Theorem with Fidelity Criterion

The noiseless source coding theorem implies that the average number of code letters per source letter can be reduced to the source entropy $H(\mathbf{X})$ under the requirement that the source sequence be exactly reproduced from the encoded sequence. If an approximate reproduction of the source sequence to within a given fidelity criterion is required, the coding rate per source letter must be reduced further to a certain value below the source entropy.

Suppose that a **distortion measure** $d(\alpha_i,\alpha_j)$ is defined for $\alpha_i,\alpha_j\in A$, where it is assumed that $d(\alpha_i,\alpha_j)\geqslant 0$ and $d(\alpha_i,\alpha_i)=0$. For blocks $x^N=x_1\ldots x_N$ and $y^N=y_1\ldots y_N$, define

$$d_N(x^N,y^N)=\frac{1}{N}\sum_{k=1}^{N} d(x_k,y_k).$$

Any set $\mathscr{C}=\{y_1^N,\ldots,y_M^N\}$, $y_m^N\in A^N$, of reproducing words is called a source code of block length N. Each source sequence x^N of the source $\mathbf{X}=[A^Z,P]$ is mapped into whichever code word $y_m^N\in\mathscr{C}$ minimizes $d_N(x^N,y^N)$, i.e.,

$$d_N(x^N,y_m^N)=\min_{y^N\in\mathscr{C}} d_N(x^N,y^N),$$

and the suffix m is transmitted. Hence the coding rate per source letter is $R=(\log M)/N$, and the average distortion is

$$\bar{d}_N(\mathscr{C})=\sum_{x^N\in A^N} P^N(x^N)\min_{y^N\in\mathscr{C}} d_N(x^N,y^N).$$

The problem is how far we can reduce the rate under the condition that the average distortion keeps satisfying a given fidelity criterion, which is specified as a maximum tolerable value d for the average distortion.

The **source coding theorem with a fidelity criterion** states: Let $\mathbf{X}=[A^Z,P]$ be a memoryless source. For any specified $d\geqslant 0$, any $\varepsilon>0$, and $\delta>0$, there exists a source code $\mathscr{C}$ with rate $R\geqslant R(d)+\delta$ and with sufficiently large block length N for which the average distortion satisfies

$$\bar{d}_N(\mathscr{C})\leqslant d+\varepsilon,$$

where $R(d)$ is the **rate distortion function** defined by

$$R(d)=\min_{W\in\mathscr{W}(d)} I(p;W),$$

$$I(p;W)=\sum_{i=1}^{n}\sum_{j=1}^{n} p_i W(j|i)\log\frac{W(j|i)}{\sum_{i=1}^{n} p_i W(j|i)},$$

$$\mathscr{W}(d)=\left\{W:\sum_{i=1}^{n}\sum_{j=1}^{n} p_i W(j|i)d(\alpha_i,\alpha_j)\leqslant d\right\},$$

where $W(j|i)$ denotes a conditional probability distribution referred to as the test channel, and where $I(p;W)$ is called the **mutual information**. It should be noted that $R(0)=H(\mathbf{X})$. The rate distortion function $R(d)$, which was first defined by Shannon [1, 8], is closely related to the **ε-entropy** introduced by A. N. Kolmogorov [9]. $R(d)$ is a monotonically decreasing and convex function.

The theorem was first proved by Shannon [8], and was extended by R. G. Gallager [10] to the case of stationary ergodic sources with discrete alphabets and by T. Berger [11] to stationary ergodic sources with abstract alphabets. More recently, R. G. Gray and L. D. Davisson [12] have proved source coding theorems without the ergodic assumption for stationary sources subject to a fidelity criterion. The proof was based on Rokhlin's ergodic decomposition theorem [13]. Other important source coding theorems have been obtained by F. Jelinek [14] for tree codes, A. J. Viterbi and J. K. Omura [15] for trellis codes, and Gray, D. L. Neuhoff, and D. S. Ornstein [16] for sliding block codes.

The rate distortion function for memoryless Gaussian sources subject to squared error distortion was given by Shannon [8], and that for autoregressive Gaussian sources was determined by Kolmogorov [9], M. S. Pinsker

[17], Berger [11], Gray [18], and T. Hashimoto and S. Arimoto [19].

F. Channel Coding Theory

A mathematical model for a channel over which information is transmitted is specified in terms of the set of possible inputs, the set of outputs, and a probability measure on the output events conditional on each input. The simplest channels are the noiseless ones, for which there is a one-to-one correspondence between input and output and no loss of information in transmission through the channel. The second simplest channels are **discrete memoryless channels** (DMCs) which are defined as follows: The input and the output are sequences of letters from finite alphabets (say, $\alpha_i \in A$ and $\beta_i \in B$), and the output letter at a given time depends statistically only on the corresponding input letter. That is, a DMC is characterized by a fixed conditional probability distribution $W(j|i) = W(\beta_j|\alpha_i)$ because the probability measure on the input and output sequences satisfies

$$W^N(y^N|x^N) = \prod_{k=1}^{N} W(y_k|x_k), \qquad x_k \in A, \quad y_k \in B.$$

By $U = \{1, \ldots, M\}$ we denote the set of integers each of which is assigned to each corresponding possible message from the source. A mapping $\varphi: U \to A^N$ induces a collection $\mathscr{C} = \{x_1^N, \ldots, x_M^N\}$, called the **block code** with rate $R = (\log M)/N$; each element is called a code word. Only code words are transmitted over the channel. A mapping $\psi: B^N \to U$ is called the decoding. Thus, given an encoding and decoding pair $[\varphi, \psi]$, the error probability is defined by

$$P_e[\varphi, \psi] = \frac{1}{N} \sum_{m=1}^{M} \sum{}^{*} W^N(y^N|\varphi(m)),$$

where the summation $\sum^*$ is taken with respect to all y^N such that $\psi(y^N) \neq m$. The **capacity** for a DMC is defined by

$$C = \max_p I(p; W)$$

$$= \max_p \sum_{i=1}^{n} \sum_{j=1}^{m} p_i W(j|i) \log \frac{W(j|i)}{\sum_{i=1}^{n} p_i W(j|i)}.$$

The fundamental theorem of channel coding theory states: Given a DMC with capacity $C > 0$, there exist a block code with rate R below capacity C, a pair $[\varphi, \psi]$, and a function $E(R) > 0$ for $0 \leqslant R < C$ such that the error probability satisfies

$$P_e[\varphi, \psi] < \exp\{-NE(R)\}.$$

This was first discovered by Shannon [1], and its precise proof was first given by A. Feinstein [20]. The precise expression of $E(R)$ was subsequently given by P. Elias [21], R. M. Fano [22], and Gallager [23]. The best expression of $E(R)$, due to Gallager [23], is

$$E(R) = \max\{E_r(R), E_{ex}(R + O(N^{-1}))\},$$

where

$$E_r(R) = \max_{0 \leqslant \rho \leqslant 1} \max_p \left[-\rho R - \log \sum_j \left\{ \sum_i p_i W(j|i)^{1/(1+\rho)} \right\}^{1+\rho} \right],$$

$$E_{ex}(R) = \sup_{\rho \geqslant 1} \max_p \left[-\rho R - \log \sum_{i,k} p_i p_k \left\{ \sum_j \sqrt{W(j|i) W(j|k)} \right\}^{1/\rho} \right]$$

The converse of the fundamental theorem states: Given a DMC with capacity C, for any block code with rate R above C and any pair $[\varphi, \psi]$, the error probability satisfies

$$P_e[\varphi, \psi] > 1 - \exp\{-NE(R)\},$$

where $E(R)$ is a function positive for $R > C$. This was first proved by J. Wolfowitz [24], and the precise expression, found by S. Arimoto [25], is described by

$$E(R) = \max_{-1 \leqslant \rho \leqslant 0} \min_p \left[-\rho R - \log \sum_j \left\{ \sum_i p_i W(j|i)^{1/(1+\rho)} \right\}^{1+\rho} \right].$$

The fundamental theorem of coding theory and its converse imply that the capacity is a critical rate; above the capacity, information cannot be transmitted reliably through the channel. Unfortunately, there is, in general, no direct method for computing the capacity, and therefore an iterative method was proposed by Arimoto [26]. Another iterative method for computing the rate distortion function was given by R. E. Blahut [27].

A discrete channel with memory is, in general, defined by a list of probability measure $\{W_{\mathbf{x}}, \mathbf{x} \in A^Z\}$ on a †measurable space $\{B^Z, \mathscr{F}_B\}$ such that for each $F \in \mathscr{F}_B$, $W_{\mathbf{x}}(F)$ is a †measurable function of $\mathbf{x}$. A channel W is called stationary if $W_{T\mathbf{x}}(TF) = W_{\mathbf{x}}(F)$ for all $\mathbf{x} \in A^Z$ and all $F \in \mathscr{F}_B$. Given an input source $[A^Z, P]$ and a channel W, connecting the input to the channel induces a joint process of input and output, denoted by $[A^Z \times B^Z, PW]$ where PW is a measure on $\{A^Z \times B^Z, \mathscr{F}_{A \times B}\}$. If a joint process $[A^Z \times B^Z, PW]$ is stationary the mutual information between input and output is defined by

$$I(\mathbf{X}; \mathbf{Y}) = H(\mathbf{X}) + H(\mathbf{Y}) - H(\mathbf{XY}).$$

A channel W is called ergodic if for every

ergodic input $[A^Z, P]$ the induced joint process $[A^Z \times B^Z, PW]$ is ergodic. The channel capacity for a discrete stationary channel is defined in various ways. The important ones are the **ergodic capacity** defined by $C_e = \sup.$ of $I(\mathbf{X}; \mathbf{Y})$ with respect to all stationary ergodic sources $\mathbf{X} = [A^Z, P]$, and the **stationary capacity** defined by $C_s = \sup.$ of $I(\mathbf{X}; \mathbf{Y})$ w.r.t. all stationary sources. K. R. Parthasarathy [28] proved $C_e = C_s$ for all discrete stationary channels. In another way J. Nedoma [29] introduced the operational source/channel block coding capacity C_{scb}, which is defined as the supremum of the entropies of all admissible stationary ergodic sources in the sense that there exist source/channel block codes such that the error probability $P_e \to 0$ as block length $N \to \infty$. Nedoma [29] also pointed out an example of a stationary channel where $C_s > C_{scb}$. Hence **block coding theorems** have been proved for various channels: for **finite memory channels** by A. I. Khinchin [30], K. Takano [31], Nedoma [29], and Feinstein [32], for **$\bar{d}$-continuous channels** by Gray and Ornstein [33], and for **almost finite memory channels** by D. L. Neuhoff and P. C. Shields [34].

References

[1] C. E. Shannon, A mathematical theory of communication, Bell System Tech. J., 27 (1948), 379–423, 623–656.
[2] B. McMillan, The basic theorems of information theory, Ann. Math. Statist., 24 (1953), 196–219.
[3] F. Jelinek, Probabilistic information theory, McGraw-Hill, 1968.
[4] I. Csiszár and G. Longo, On the error exponent for source coding, Studia Math. Acad. Hung., 6 (1971), 181–191.
[5] R. E. Blahut, Hypothesis testing and information theory, IEEE Trans. Information Theory, 20 (1974), 405–417.
[6] G. Longo and A. Sgarro, The source coding theorem revisited: A combinatorial approach, IEEE Trans. Information Theory, 25 (1979), 544–548.
[7] S. Kullback, Information theory and statistics, Wiley, 1957.
[8] C. E. Shannon, Coding theorems for a discrete source with a fidelity criterion, IRE National Convention Record, pt. 4 (1959), 142–163.
[9] A. N. Kolmogorov, On the Shannon theory of information transmission in the case of continuous signals, IRE Trans. Information Theory, 2 (1956), 102–108.
[10] R. G. Gallager, Information theory and reliable communication theory, Wiley, 1968.
[11] T. Berger, Rate distortion theory for sources with abstract alphabets and memory, Information and Control, 13 (1968), 254–273.
[12] R. G. Gray and L. D. Davisson, Source coding theorems without the ergodic assumption, IEEE Trans. Information Theory, 20 (1974), 502–516.
[13] V. A. Rokhlin, On the fundamental ideas of measure theory, Amer. Math. Soc. Transl., 71 (1952).
[14] F. Jelinek, Tree-encoding of memoryless time-discrete sources with a fidelity criterion, IEEE Trans. Information Theory, 15 (1969), 584–590.
[15] A. J. Viterbi and J. K. Omura, Trellis encoding of memoryless discrete-time sources with a fidelity criterion, IEEE Trans. Information Theory, 20 (1974), 325–332.
[16] R. M. Gray, D. L. Neuhoff, and D. S. Ornstein, Nonblock source coding with a fidelity criterion, Ann. Probability, 3 (1975), 478–491.
[17] M. S. Pinsker, Information and information stability of random variables and processes, Holden-Day, 1965.
[18] R. M. Gray, Information rates of autoregressive processes, IEEE Trans. Information Theory, 16 (1970), 412–421.
[19] T. Hashimoto and S. Arimoto, On the rate-distortion function for the nonstationary Gaussian AR process, IEEE Trans. Information Theory, 26 (1980), 478–480.
[20] A. Feinstein, A new basic theorem of information theory, IRE Trans. Information Theory, 4 (1954), 2–22.
[21] P. Elias, Coding for noisy channels, IRE National Convention Record, pt. 4 (1955), 37–46.
[22] R. M. Fano, Transmission of information, MIT Press and Wiley, 1961.
[23] R. G. Gallager, A simple derivation of the coding theorem and some applications, IEEE Trans. Information Theory, 11 (1965), 3–18.
[24] J. Wolfowitz, The coding of messages subject to chance errors, Illinois J. Math., 1 (1957), 591–606.
[25] S. Arimoto, On the converse to the coding theorem for discrete memoryless channels, IEEE Trans. Information Theory, 19 (1973), 357–359.
[26] S. Arimoto, An algorithm for computing the capacity of arbitrary discrete memoryless channels, IEEE Trans. Information Theory, 18 (1972), 14–20.
[27] R. E. Blahut, Computation of channel capacity and rate-distortion functions, IEEE Trans. Information Theory, 18 (1972), 460–473.
[28] K. R. Parthasarathy, On the integral representation of the rate of transmission of a stationary channel, Illinois J. Math., 2 (1961), 299–305.

[29] J. Nedoma, The capacity of a discrete channel, Trans. 1st Prague Conf. Information Theory (1957), 143–181.
[30] A. Ya. Khinchin, Mathematical foundations of information theory, Dover, 1957.
[31] K. Takano, On the basic theorems of information theory, Ann. Inst. Statist. Math., 9 (1957), 53–77.
[32] A. Feinstein, On the coding theorem and its converse for finite-memory channels, Information and Control, 2 (1959), 25–44.
[33] R. M. Gray and D. S. Ornstein, Block coding for discrete stationary $\bar{d}$-continuous noisy channels, IEEE Trans. Information Theory, 25 (1979), 292–306.
[34] D. L. Neuhoff and P. C. Shields, Channels with almost finite memory, IEEE Trans. Information Theory, 25 (1979), 440–447.

214 (XVIII.18) Insurance Mathematics

A. General Remarks

Insurance is a system in which a large number of people contribute a small precalculated amount of money (called a **premium**) to fill the economic need that arises when a person meets adversity. The amount of economic need filled by this system is called the **amount of insurance** (or **amount insured**). The **insurer** is the one who implements the system. **Actuarial mathematics** is the branch of applied mathematics that studies the mathematical basis of insurance, one of the first cases in which mathematics was successfully applied to a social question. Actuarial mathematics can be divided into two branches according to its application. The first includes the calculation of various values of each individual policy, such as premiums or reserves. The second is mainly connected with management of an insurance business and includes the study of reinsurance systems, of the maximum amount of insurance, of the contingency fund, or the analysis of profits. There is only one basic principle in actuarial mathematics, called the **principle of equivalence**. It determines the premium and reserve in each year so that the present value of future premium income of the insurer is equal to the present value of future benefits for each policy.

The basic factors of actuarial calculations are (1) probabilities of contingencies, (2) an expected rate of interest in the future (often referred to as the **assumed rate of interest**), and (3) cost of administration of the system. Premiums are calculated using these factors and the principle of equivalence. The following is an example of the classical method of calculation for a life insurance policy with the use of "commutation symbols," which is an old device for the convenience of calculations.

We write P for the net premium (in which the cost of administration is disregarded), P' for the gross premium, T_t for the amount of death benefits payable in the tth year after the policy is issued, E_t for the amount of survival benefits payable at the beginning of the tth year, n for the period for which the insurance is effective, and m for the period for which premiums are to be paid. Let α, β, and γ stand for three positive constants determining the initial expenses $=\alpha(T_1$ or $E_1)$, the premium collection expenses $=\beta P'$, and the general expenses for maintenance $=\gamma(T_t$ or $E_t)$. The factor that comes into consideration next is a **model of human death and survival** (measurement). Assume that l_x is the number of lives attaining age x, and write q_x for the probability that a life of x years will end within one year. Then d_x, the number of lives ending within one year out of l_x, is $l_x q_x$, and l_{x+1}, the number of lives remaining after one year at age $x+1$, is $l_x - d_x = l_x(1-q_x)$. The commutation symbols commonly employed are defined as follows: Write $v=1/(1+i)$, where i is the assumed rate of interest; then

$$D_x = l_x v^x, \qquad C_x = d_x v^{x+1},$$

$$N_x = \sum_{t=0}^{n} D_{x+t}, \qquad M_x = \sum_{t=0}^{n} C_{x+t}.$$

For a policy issued at an insured person's age x, the present value of the insurer's future income can be expressed as $P'(N_x - N_{x+m})/D_x$, and the present value of his future payments can be expressed as

$$\frac{1}{D_x}\Bigg(\sum_{t=1}^{n} T_t C_{x+t-1} + \sum_{t=1}^{n+1} E_t D_{x+t-1} + \alpha(T_1 \text{ or } E_1)D_x + \gamma \sum_{t=0}^{n} (T_t \text{ or } E_t) D_{x+t-1} + \beta P'(N_x - N_{x+m})\Bigg).$$

By assuming that the present value of the future income is equal to the present value of the future payments, the value of the gross premium P' is obtained. (The P' obtained from the assumption $\alpha=\beta=\gamma=0$ is denoted by P and is called the **net premium**. The difference $P'-P$ is called the **loading**.) For a policy in which benefits are payable on disability or contingencies other than death, we have only to obtain a model of contingencies and apply a similar calculation.

B. Liability Reserve

During the term of an insurance contract, it often happens that the present value of the future income is less than the present value of the future payments. If this is the case, the difference is to be held by the insurer as the **liability reserve**. The source of this fund is the past premium income plus interest. The net premium reserve, which disregards expenses, is calculated as

$$V_t = \frac{1}{D_{x+t}} \left(\sum_{r=t+1}^{n} T_r C_{x+r-1} + \sum_{r=t+1}^{n+1} E_r D_{x+r-1} - P(N_{x+1} - N_{x+m}) \right).$$

Between the net premium P and the net premium reserve V, we have the relation

$$P = (vV_t - V_{t-1}) + (T_t - V_t)vq_{x+t} + E_t.$$

The first term of the right-hand side of this formula is called the **savings premium**, since it is the amount left out of the premium income of the tth year and added to the reserve. The second term is called the **cost of insurance** or **risk premium** and is applied to cover the difference between the amount of insurance and that of the existing reserve in case the contingency of death arises. The third term is applied to the payment of the survival benefits (or annuities in case of an **annuity contract**). If $T_t - V_t$, the amount of risk insured by the insurer, is positive for all values of t during the period of insurance, the policy is called **death insurance**. On the other hand, if the value of $T_t - V_t$ is negative for all values of t, the policy is called **survival insurance**. If the value of $T_t - V_t$ varies between positive and negative according to the different values of t, the policy is called **mixed insurance**. If T_t is always equal to V_t, the policy constitutes mere savings. Most of the insurance policies issued today are one or another type of death insurance, while life annuity policies are a type of survival insurance. For a long time, studies have been made of the effect on premiums and reserves of changes in the three basic factors (1), (2), and (3) in Section A.

C. Risk Theory

Risk theory occupies a special position in the field of actuarial mathematics. Actuarial mathematics was first born from the theory of probability. Since the modern theory of probability based on measure theory was developed by A. N. Kolmogorov and other mathematicians (→ 342 Probability Theory), new approaches have inevitably been made to actuarial mathematics. An outstanding example is risk theory.

Risk theory can be divided into two branches. One is called **classical risk theory** (or **individual risk theory**), in which the profit or loss that may result during a certain term of an insurance contract is regarded as a †random variable. Since the insurer's profit equals the sum of these random variables over all the individual contracts, various probability functions can be obtained by applying the theory of probability. The second, called **collective risk theory**, pays no attention to each individual contract but studies changes in the insurer's balance as a whole with the lapse of time. The basis of collective risk theory was given by F. Lundberg, H. Cramér, and other mathematicians.

We explain the collective risk theory following Cramér [5]. For simplicity we consider an insurer who issues no policies other than death insurance and makes no expenditures except the policy claims. Suppose that during the time interval $(t, t+\Delta t)$ contingency occurs with probability $\lambda \Delta t + o(\Delta t)$ independently of the past. We denote by $F(x)$ the †distribution function of the amount to be paid by the insurer when a contingency occurs. Then the number of contingencies in $(0, t)$, $N(t)$, is a †Poisson process with parameter λ, and the total expenditure of the insurer during the time interval $(0, t)$, $X(t)$, is a †compound Poisson process (→ 5 Additive Processes) such that

$$E(e^{izX(t)}) = \exp\left(\lambda t \int_0^{\infty} (e^{izx} - 1) dF(x) \right).$$

If p is the premium income per unit time, then we have $p = \lambda \int_0^{\infty} x dF(x)$, because $E(X(t)) = pt$ holds by the principle of equivalence. If u denotes the initial fund of the insurer, then the fund reserved at time t, $Y(t)$, equals $u + pt - X(t)$. $Y(t)$ is a †Lévy process (→ 5 Additive Processes) such that

$$E(e^{izY(t)}) = \exp\left(izu + \lambda t \int_0^{\infty} (e^{izx} - 1 - izx) dF(x) \right).$$

The probability that $Y(t) < 0$ for some time $t < \infty$ is called the **ruin probability**, which depends on the initial fund u, denoted by $\psi(u)$. This satisfies the following †integral equation of Volterra type:

$$\frac{p}{\lambda} \psi(u) = \int_u^{\infty} Q(x) dx + \int_0^u \psi(x) Q(u-x) dx \qquad (Q(x) = 1 - F(x)),$$

from which one can derive an asymptotic relation $\psi(u) \sim Ce^{-Ru}$ as $u \to \infty$, where R and C

are positive constants depending only on λ and F. See [6] for recent developments in risk theory.

References

[1] C. W. Jordan, Society of Actuaries' textbook on life contingencies, Society of Actuaries, 1952.
[2] W. Saxer, Versicherungsmathematik, Springer, I, 1955; II, 1958.
[3] E. Zwinggi, Versicherungsmathematik, Birkhäuser, 1945.
[4] P. F. Hooker and L. H. Longley-Cook, Life and other contingencies I, Cambridge Univ. Press, 1953.
[5] H. Cramér, Collective risk theory, Nordiska Bokhandeln, 1955.
[6] H. Bühlmann, Mathematical methods in risk theory, Springer, 1970.

215 (XIX.6) Integer Programming

A. General Remarks

Integer programming, in its broadest sense, addresses itself to either minimization or maximization of a functional f over some discrete set S in $\mathbf{R}^n$, but it is usually understood as dealing with questions related to linear programming problems (→ 255 Linear Programming) with additional integrality conditions on the variables, namely, the problem P_0: Minimize $\{\mathbf{c}'\mathbf{x} \mid A\mathbf{x}=\mathbf{b}, \mathbf{x}\geqslant \mathbf{0}, x_j$ an integer, $j=1,\ldots,n_1(\leqslant n)\}$, where $A\in\mathbf{R}^{m\times n}$, $\mathbf{b}\in\mathbf{R}^m$, $\mathbf{c}\in\mathbf{R}^n$ are given data and $\mathbf{x}=(x_1,\ldots,x_n)^t\in\mathbf{R}^n$ is a vector. P_0 is called a **pure integer programming problem** or an **all-integer programming problem** if $n=n_1$, and a **mixed integer programming problem** if $n_1<n$. In particular, P_0 is called a **0-1 integer programming problem** if all the integer variables are restricted to be equal to either 0 or 1. We write

$$X^0=\{\mathbf{x}\in\mathbf{R}^n \mid A\mathbf{x}=\mathbf{b}, x_j\geqslant 0, j=1,\ldots,n\},$$

$$X^I=X^0\cap\{\mathbf{x}\in\mathbf{R}^n \mid x_j \text{ is an integer}, j=1,\ldots,n_1\},$$

and assume for simplicity that (i) $X^I\neq\varnothing$, (ii) X^0 is bounded, and (iii) all the components of A and $\mathbf{b}$ are integers. P_0 arises not only as a mathematical model for an optimization problem where some of the decision variables have indivisible minimum units but also as one for many optimization problems with some logical and/or combinatorial constraints [1].

Algorithms for solving P_0 can be classified into algebraic methods and enumerative methods. Both are based upon linear programming and **relaxation** techniques [2] in which some of the constraints of P_0 are temporarily relaxed (→ 264 Mathematical Programming).

B. Cutting Plane Methods

The origin of this class of algorithms is the **fractional cutting plane algorithm** proposed in 1958 by R. E. Gomory [3], the outline of which is described as:
(1) Let $X=X^0$.
(2) Solve a linear programming problem: Minimize $\{\mathbf{c}'\mathbf{x} \mid \mathbf{x}\in X\}$, and let $\bar{\mathbf{x}}$ be its †optimal solution. If the $\bar{x}_j$ are integer for all $j\leqslant n_1$, then stop ($\bar{\mathbf{x}}$ is optimal to P_0). Otherwise, go to (3).
(3) Generate a half-space $H=\{\mathbf{x}\in\mathbf{R}^n \mid \boldsymbol{\pi}'\mathbf{x}\geqslant\pi_0\}$ ($\boldsymbol{\pi}\in\mathbf{R}^n, \pi_0\in\mathbf{R}^1$), where H (i) contains X^I and (ii) does not contain $\bar{\mathbf{x}}$. Return to (2) by replacing X by $X\cap H$.

Here, a linear programming problem obtained by relaxing integrality conditions is solved, and as long as $\bar{\mathbf{x}}\notin X^I$, an inequality satisfying the two conditions (i), (ii) of (3) is introduced. Such an inequality $\boldsymbol{\pi}'\mathbf{x}\geqslant\pi_0$ or equality $\boldsymbol{\pi}'\mathbf{x}=\pi_0$ is called a cut or a **cutting plane**. Gomory devised a **Gomory cut** using (relaxed) integrality conditions on the variables, and showed that the algorithm above produces a point of X^I in finitely many steps. Some of the other algorithms using cutting planes are the **all-integer algorithm** (Gomory, 1963) and the primal all-integer algorithm (R. D. Young, *Operations Res.*, 16 (1968)). These algorithms, however, are generally slow and behave erratically, so that it is believed that they cannot in practice serve as general-purpose algorithms.

C. Other Algebraic Methods

Gomory [4], again in 1965, proposed a **group-theoretic approach** to P_0. This method is based upon the following observation: Let $\mathbf{x}_B$ be the vector of †basic variables (→ 255 Linear Programming) associated with a †dual feasible basis of a linear programming problem $\bar{P}_0$: Minimize $\{\mathbf{c}'\mathbf{x} \mid \mathbf{x}\in X^0\}$, and let $\tilde{X}^I$ be the set generated from X^I by relaxing the nonnegativity constraints on $\mathbf{x}_B$. Then $\tilde{X}^I$ can be shown to have a †cyclic group structure, so that a **group minimization problem** P_G: minimize $\{\mathbf{c}'\mathbf{x} \mid \mathbf{x}\in\tilde{X}^I\}$ can be solved as a †shortest path problem on a directed graph with a special structure (→ 186 Graph Theory). If the optimal solution $\tilde{\mathbf{x}}$ of P_G satisfies $\tilde{\mathbf{x}}_B\geqslant\mathbf{0}$, then it is optimal for P_0. If, on the other hand, $\tilde{\mathbf{x}}_B\ngeqslant\mathbf{0}$,

then a branch and bound algorithm described later can be applied for a systematic search of integer points near $\tilde{\mathbf{x}}$. This algorithm is reported to produce good results when the size of the associated graph is not excessively large (G. A. Gorry et al., *Management Sci.*, 17 (1971)). Some of the other important results in this area are: (i) the theory of **subadditive cuts** (R. Gomory et al., *Math. Prog.*, 3 (1972)) and **disjunctive cuts** (E. Balas, in *Nonlinear Programming* 2, O. Mangasarian et al. (eds.), Academic Press, 1975), (ii) research on facial structures of the **integer polyhedron** coX^I (convex hull of X^I) for some of the more important integer programming problems, such as that of the **corner polyhedron** co$\tilde{X}^I$ [5], knapsack polytopes (E. Balas, *Math. Prog.*, 8 (1975)), and traveling-salesman polytopes (M. Grötschel et al., *Math. Prog.*, 16 (1979)). It should be pointed out, however, that more difficulties of P_0 have been revealed rather than resolved through the intensive research in this area. Incidentally, the 0-1 integer programming problem is known to be †*NP*-complete (→ 71 Complexity of Computations).

D. Enumerative Methods

Another large class of algorithms for solving P_0 consists of the **branch and bound methods**, first proposed by Land and Doig [6] in 1960. An outline of the improved version (R. J. Dakin, *Computer J.*, 8 (1965)) is as follows.
(1) Let $\mathscr{P}=\{P_0\}$, $z^*=\infty$, $\mathbf{x}^*=$ undefined.
(2) If $\mathscr{P}=\varnothing$, then stop ($\mathbf{x}^*$ is optimal to P_0). Otherwise, choose from P the problem P_l: Minimize $\{\mathbf{c}'\mathbf{x} \mid \mathbf{x}\in X_l^I\}$.
(3) Solve $\bar{P}_l$: Minimize $\{\mathbf{c}'\mathbf{x} \mid \mathbf{x}\in X_l\}$, in which the integrality condition is relaxed from the constraints of P_l. If $\bar{P}_l$ has an optimal solution $\tilde{\mathbf{x}}$, then go to (4). Otherwise, return to (2).
(4) If $\tilde{\mathbf{x}}\in X_l^I$ and $\mathbf{c}'\mathbf{x}<z^*$, then let $\tilde{\mathbf{x}}\to\mathbf{x}^*$, $\mathbf{c}'\mathbf{x}\to z^*$, and return to (2). If $\tilde{\mathbf{x}}\in X_l^I$ and $\mathbf{c}'\mathbf{x}\geqslant z^*$, then return to (2). Otherwise, go to (5).
(5) Choose $j\leqslant n_1$, for which $\tilde{x}_j$ is not integral and generate the two **subproblems** P_j^+: minimize $\{\mathbf{c}'\mathbf{x} \mid \mathbf{x}\in X_l^I, x_j\geqslant[\tilde{x}_j]+1\}$, and P_j^-: minimize $\{\mathbf{c}'\mathbf{x} \mid \mathbf{x}\in X_l^I, x_j\leqslant[\tilde{x}_j]\}$, in both of which $[\tilde{x}_j]$ represents the largest integer not exceeding $\tilde{x}_j$. Let $\mathscr{P}\cup\{P_j^-, P_j^-\}\to\mathscr{P}$ and return to (2).

The best point of X^I identified during the preceding steps is denoted by $\mathbf{x}^*$ and called an incumbent. In summary, the branch and bound method chooses one subproblem P_l from the problem list $\mathscr{P}$ and estimates the lower bound of its optimal objective functional value. If the lower bound is worse than the current incumbent, then P_l is discarded, whereas P_l is separated into two subproblems if no conclusion can be reached. This process is continued until the problem list P becomes empty, thereby implicitly checking all points of X^I. The above method is called an LP-based branch and bound method because linear programming techniques are employed to obtain a lower bound. The branch and bound method tends to require a large amount of storage, but many engineering improvements on the method of choosing (i) a subproblem P_l and (ii) a branching variable x_j, and several improvements of the bounding techniques, in addition to the substantial progress made in linear programming codes, enable us to solve a problem of size $n\approx 100$. In particular, an improved version of the **implicit enumeration method**, proposed by E. Balas [7] for 0-1 integer programming problems which uses logical conditions for obtaining lower bounds, is known to be able to solve rather large 0-1 integer programming problems (A. Geoffrion, *Operations Res.*, 17 (1969)).

E. Other Topics

The **partitioning algorithm** [8], in which integer variables are varied parametrically, is reported to work well for 0-1 mixed integer problems with relatively few integer variables. As people begin to realize the intrinsic difficulty of P_0, they pay more attention to **heuristic algorithms** or approximate algorithms to obtain a good but not necessarily optimal solution. Among heuristic methods, the interior path method [9], which elaborates simple ideas such as rounding of the optimal solution of $\bar{P}_0$, has been reported to work well for problems in which X^0 has a nonempty interior. Also, more emphasis is being placed on special purpose algorithms for solving practical problems, such as †set partitioning and the †traveling salesman problem, etc. [1]. P_0 is a typical nonconvex programming problem, and no practically useful †duality theorem is available. Hence it is difficult to perform sensitivity and/or post-optimality analysis. Some research in this area has emerged recently (e.g., C. J. Piper et al., *Management Sci.*, 22 (1976)), but it looks as if it will be several years before a reasonably good procedure becomes available.

References

[1] R. S. Garfinkel and G. L. Nemhauser, Integer programming, Wiley, 1972.
[2] A. M. Geoffrion and R. E. Marten, Integer programming: A framework and state-of-the-art survey, Management Sci., 18 (1972), 465–491.

[3] R. E. Gomory, Outline of an algorithm for integer solutions to linear programs, Bull. Amer. Math. Soc., 64 (1958), 275–278.
[4] R. E. Gomory, On the relation between integer and non-integer solutions to linear programs, Proc. Nat. Acad. Sci. US, 53 (1965), 260–265.
[5] R. E. Gomory, Some polyhedra related to combinatorial problems, Linear Algebra and Appl., 2 (1969), 451–558.
[6] A. H. Land and A. G. Doig, An automatic method for solving discrete programming problems, Econometrica, 28 (1960), 497–520.
[7] E. Balas, An additive algorithm for solving linear programs with zero-one variables, Operations Res., 13 (1965), 517–546.
[8] J. F. Benders, Partitioning procedures for solving mixed-variables programming problems, Numer. Math., 4 (1962), 238–252.
[9] F. S. Hillier, Efficient heuristic procedures for integer linear programming with an interior, Operations Res., 17 (1969), 600–637.

216 (X.10) Integral Calculus

A. The Riemann Integral

Let $f(x)$ be a bounded real-valued function defined on an interval $[a,b]$. We shall divide this interval $I=[a,b]$ into subintervals $I_i=[x_{i-1},x_i]$ $(i=1,\ldots,n)$ by a finite number of points x_i $(a=x_0<x_1<\ldots<x_n=b)$. This division into subintervals is uniquely determined by the set $D=\{x_i\}$, called the **partition** of I. We set $M_i=\sup_{x\in I_i}f(x)$, $m_i=\inf_{x\in I_i}f(x)$, and put $\overline{\sigma}(D)=\sum_{i=1}^n M_i(x_i-x_{i-1})$, $\underline{\sigma}(D)=\sum_{i=1}^n m_i(x_i-x_{i-1})$. Considering all possible partitions D of I, we set $\overline{\int_a^b} f(x)\,dx=\inf_D\overline{\sigma}(D)$, $\underline{\int_a^b} f(x)\,dx=\sup_D\underline{\sigma}(D)$, which are called the **Riemann upper integral** and **Riemann lower integral** of f, respectively. If they coincide, then the common value is called the **Riemann integral** of f on $[a,b]$ and is denoted by $\int_a^b f(x)\,dx$. In this case, we say that f is **Riemann integrable** (or simply **integrable**) on $[a,b]$ and call f the **integrand**; a and b are called the **lower limit** and the **upper limit**, respectively. In this case, by **integrating** f from a to b we mean the process of obtaining the value $\int_a^b f(x)\,dx$.

Darboux's theorem: For each $\varepsilon>0$ there exists a positive δ such that the inequalities

$$\left|\overline{\sigma}(D)-\overline{\int_a^b} f(x)\,dx\right|<\varepsilon,$$

$$\left|\underline{\sigma}(D)-\underline{\int_a^b} f(x)\,dx\right|<\varepsilon$$

hold for any partition D with $\delta(D)=\max_i(x_i-x_{i-1})<\delta$. In other words, we have

$$\lim_{\delta(D)\to 0}\overline{\sigma}(D)=\overline{\int_a^b} f(x)\,dx,$$

$$\lim_{\delta(D)\to 0}\underline{\sigma}(D)=\underline{\int_a^b} f(x)\,dx.$$

From Darboux's theorem it follows that a necessary and sufficient condition for $f(x)$ to be integrable on $[a,b]$ is that for each positive ε there exist a positive δ such that $\delta(D)<\delta$ implies $\overline{\sigma}(D)-\underline{\sigma}(D)=\sum_{i=1}^n(M_i-m_i)(x_i-x_{i-1})<\varepsilon$.

We call M_i-m_i the **oscillation** of f on I_i and $\overline{\sigma}(D)$ and $\underline{\sigma}(D)$ the **Darboux sums**. Obviously, if f is integrable on $[a,b]$, then for each positive ε there exists a positive δ such that the following inequality holds for every partition $D=\{x_j\}$ with $\delta(D)<\delta$ and for every set of points $\xi_j\in I_j$ $(j=1,\ldots,n)$:

$$\left|\sum_{j=1}^n f(\xi_j)(x_j-x_{j-1})-\int_a^b f(x)\,dx\right|<\varepsilon.$$

The sum $\sum_{j=1}^n f(\xi_j)(x_j-x_{j-1})$ is often called a **Riemann sum** (or **sum of products**). A function that is continuous on $[a,b]$, or bounded and continuous except for a finite number of points in the interval, is integrable. Furthermore, a bounded function that is continuous on $[a,b]$ except for an infinite number of points x_λ is integrable if for an arbitrary positive number ε there exist a finite number of intervals I_i of which the total length is less than ε and if the set $\{x_\lambda\}$ of exceptional points is contained in $\bigcup I_i$. Generally, a necessary and sufficient condition for a bounded function defined on $[a,b]$ to be integrable is that the set of points where the function is not continuous be of †measure 0 (in the sense of Lebesgue). A function that is either †monotonic on $[a,b]$ (and consequently bounded) or of †bounded variation is integrable. A function that is integrable on $[a,b]$ is integrable on any subinterval of $[a,b]$, the integrand being the restriction of the given function to this subinterval.

B. Basic Properties of Integrals

Let $\mathbf{I}$ be the set of all functions integrable on $[a,b]$. If $f,g\in\mathbf{I}$, then for any numbers α,β, we have $\alpha f+\beta g\in\mathbf{I}$, $fg\in\mathbf{I}$, $\min\{f,g\}\in\mathbf{I}$, $\max\{f,g\}\in\mathbf{I}$, and $f/g\in\mathbf{I}$ provided that there exists a positive constant A such that the inequality $|g|>A$ holds. Furthermore, if $f\in\mathbf{I}$, then $|f|\in\mathbf{I}$; and if $f_n\in\mathbf{I}$ $(n=1,2,\ldots)$ and f_n converges †uniformly to f, then $f\in\mathbf{I}$. Corresponding to these properties, the following formulas hold.

(1) Linearity:

$$\int_a^b (\alpha f(x)+\beta g(x))\,dx$$

$$=\alpha\int_a^b f(x)\,dx+\beta\int_a^b f(x)\,dx,$$

where α, β are constants.

(2) Monotonicity: If $f(x)\geqslant 0$, then

$$\int_a^b f(x)\,dx\geqslant 0.$$

If, further, f is continuous at a point $x_0\in[a,b]$ and $f(x_0)>0$, then $\int_a^b f(x)\,dx>0$.

(3) Additivity with respect to intervals: If a, b, and c are points belonging to an interval on which f is integrable and $a<c<b$, then

$$\int_a^b f(x)\,dx=\int_a^c f(x)\,dx+\int_c^b f(x)\,dx.$$

Adopting the conventions that $\int_a^a f(x)\,dx=0$ and $\int_b^a f(x)\,dx=-\int_a^b f(x)\,dx$, the additivity formula holds independently of the order of a, b, and c.

It follows from (2) that $|\int_a^b f(x)\,dx|\leqslant \int_a^b |f(x)|\,dx$ if $a<b$. Further, if $f_n(x)$ converges to $f(x)$ uniformly on $[a,b]$, then

$$\lim_{n\to\infty}\int_a^b f_n(x)\,dx=\int_a^b f(x)\,dx.$$

Replacing $f_n(x)$ by partial sums of a series, we obtain the following theorem: Let $\sum a_n(x)$ be a series in which each term $a_n(x)$ is integrable on an interval $[a,b]$. If the series converges uniformly on $[a,b]$, then the sum $s(x)$ is integrable on $[a,b]$, and the series is **termwise integrable**, that is,

$$\int_a^b s(x)\,dx=\sum_{n=1}^{\infty}\int_a^b a_n(x)\,dx.$$

Also, the series $\sum_{n=1}^{\infty}\int_a^x a_n(t)\,dt$ converges uniformly on $[a,b]$ to the integral $\int_a^x s(t)\,dt$. Assume that $\sum a_n(x)$ is convergent but not uniformly convergent. If all $a_n(x)$, together with $s(x)=\sum a_n(x)$, are integrable and there is a constant M independent of n such that $|s_n(x)|\leqslant M$ $(x\in[a,b])$ for all n, where $s_n(x)$ are partial sums, then the series is termwise integrable (C. Arzelà).

The first mean value theorem: If $f(x)$ is continuous on $[a,b]$ and $\varphi(x)$ is integrable and of constant sign on $[a,b]$, then there exists θ $(0<\theta<1)$ such that

$$\int_a^b f(x)\varphi(x)\,dx=f(a+\theta(b-a))\int_a^b \varphi(x)\,dx.$$

When $\varphi(x)=1$, we have

$$\int_a^b f(x)\,dx=f(a+\theta(b-a))(b-a).$$

The second mean value theorem: If $f(x)$ is a positive, monotone decreasing function defined on $[a,b]$ and $\varphi(x)$ is an integrable function, then there exists η $(a<\eta\leqslant b)$ such that

$$\int_a^b f(x)\varphi(x)\,dx=f(a+0)\int_a^{\eta}\varphi(x)\,dx.$$

In the hypothesis of the second mean value theorem, if $f(x)$ is assumed to be monotonic but not necessarily positive, then there exists η $(a\leqslant\eta\leqslant b)$ such that

$$\int_a^b f(x)\varphi(x)\,dx$$

$$=f(a+0)\int_a^{\eta}\varphi(x)\,dx+f(b-0)\int_{\eta}^b\varphi(x)\,dx.$$

H. Okamura (1947) proved that the condition $a\leqslant\eta\leqslant b$ can be replaced by $a<\eta<b$.

In the case $f(x)\geqslant 0$ on $[a,b]$, we consider the figure F bounded by the graph of $f(x)$, the x-axis, and the lines $x=a$ and $x=b$. Then $\bar{\sigma}(D)$ and $\underline{\sigma}(D)$ are areas of polygons of which one encloses F and the other is enclosed by F, as shown in Fig. 1. Hence it can be shown that the integrability of $f(x)$ in the sense of Riemann is equivalent to the measurability of F in the sense of Jordan. The Riemann integral $\int_a^b f(x)\,dx$ is the area of F with respect to its †Jordan measure.

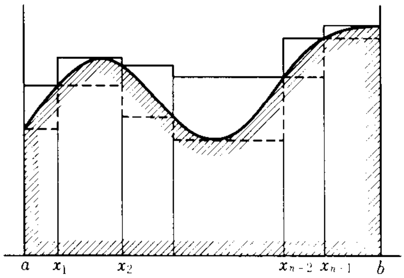


Fig. 1

C. Relation between Differentiation and Integration

Suppose that $f(x)$ is integrable on an interval I. We fix a point a of I and consider the integration $F(x)=\int_a^x f(t)\,dt$, where x varies in I. The function $F(x)$ is called the **indefinite integral** of $f(x)$. In contrast with this, the integral on a fixed interval, as considered in the previous sections, is often called the **definite integral**. The indefinite integral $F(x)$ is continuous on the interval I and of bounded variation. If $f(x)$ is continuous at a point x_0 in I, then $F(x)$ is differentiable at x_0 and $F'(x_0)=f(x_0)$. In general, if a function $G(x)$ satisfies $G'(x)=f(x)$ everywhere in I, then $G(x)$ is called a **primitive function** of $f(x)$. If $f(x)$ is continu-

ous, the indefinite integral of $f(x)$ is one of the primitive functions of $f(x)$. Furthermore, if a function $G(x)$ is a primitive function of $f(x)$, then any other primitive function can be written in the form $G(x)+C$, where C is a constant, called an **integral constant**. For a continuous function $f(x)$ on $[a,b]$ and any one of its primitive functions $G(x)$, we have

$$\int_a^b f(x)\,dx = G(b)-G(a)=[G(x)]_a^b$$

(fundamental theorem of calculus) (→ Appendix A, Table 9). From the differentiation formulas we obtain the following integration formulas:

Integration by parts: If $f(x)$ and $g(x)$ have continuous derivatives on $[a,b]$, then

$$\int_a^b f(x)g'(x)\,dx=[f(x)g(x)]_a^b-\int_a^b f'(x)g(x)\,dx.$$

More generally, if $f(x)$ and $g(x)$ are integrable on $[a,b]$, then

$$\int_a^b g(x)\left(\int_a^x f(t)\,dt\right)dx$$
$$=\int_a^b f(x)\,dx\int_a^b g(x)\,dx-\int_a^b f(x)\left(\int_a^x g(t)\,dt\right)dx.$$

Change of variables: If $f(x)$ is integrable on $[a,b]$ and $x=\varphi(t)$ and $\varphi'(t)$ are continuous on $[\alpha,\beta]$, where $a=\varphi(\alpha)$, $b=\varphi(\beta)$ $(a\leqslant\varphi(t)\leqslant b)$, then

$$\int_a^b f(x)\,dx=\int_\alpha^\beta f(\varphi(t))\varphi'(t)\,dt.$$

D. Improper Integrals

The concept of the integral can be generalized to the case where the integrand or the interval on which integration is accomplished is not bounded. Assume that $f(x)$ is not bounded on $[a,b)$ but is bounded and integrable on any interval $[a,b-\varepsilon]$ $(\subset[a,b))$. If $\int_a^{b-\varepsilon}f(x)\,dx$ has a finite limit for $\varepsilon\to 0$, the limit is denoted by $\int_a^b f(x)\,dx$ and is called the **improper Riemann integral** (or simply **improper integral**) of $f(x)$ on $[a,b)$. For example, if $f(x)$ is continuous on $[a,b)$ and $f(x)=O((b-x)^\alpha)$ for some α $(0>\alpha>-1)$, where O is the †Landau symbol, then the improper integral $\int_a^b f(x)\,dx$ exists. On the other hand, if f is integrable on $[a+\varepsilon,b]$ for each $\varepsilon>0$ but not bounded in any neighborhood of a, we can define the integral on $(a,b]$ in the same way. If f is not bounded in any neighborhood of a or b and if there exists a point c $(a<c<b)$ for which the improper integrals $\int_a^c f(x)\,dx$ and $\int_c^b f(x)\,dx$ exist, then we define $\int_a^b f(x)\,dx=\int_a^c f(x)\,dx+\int_c^b f(x)\,dx$, which is independent of the choice of the point c. Furthermore, assume that $f(x)$ is not bounded in any neighborhood of each point c_j $(j=1,\dots,n)$ $(a<c_1<\dots<c_n<b)$. Then we define

$$\int_a^b f(x)\,dx=\int_a^{c_1}f(x)\,dx+\int_{c_1}^{c_2}f(x)\,dx+\dots$$
$$+\int_{c_{n-1}}^{c_n}f(x)\,dx+\int_{c_n}^b f(x)\,dx,$$

provided that all improper integrals

$$\int_a^{c_1}f(x)\,dx,\dots,\int_{c_n}^b f(x)\,dx$$

exist. Suppose that $f(x)$ is defined on $[a,b]$ and bounded outside any neighborhood of $c\in(a,b)$ but not bounded in either $[c-\varepsilon,c]$ or $[c,c+\varepsilon]$ for any $\varepsilon>0$. It may well happen that, although neither $\lim_{\varepsilon\to0}\int_a^{c-\varepsilon}f(x)\,dx$ nor $\lim_{\varepsilon'\to0}\int_{c+\varepsilon'}^b f(x)\,dx$ exists (accordingly, the improper integral $\int_a^b f(x)\,dx$ does not exist), if we put $\varepsilon=\varepsilon'$, the limit

$$\lim_{\varepsilon\to0}\left(\int_a^{c-\varepsilon}f(x)\,dx+\int_{c+\varepsilon}^b f(x)\,dx\right)$$

does exist. This limit is called **Cauchy's principal value** and is denoted by p.v.$\int_a^b f(x)\,dx$ (v.p. in French). For example, p.v.$\int_{-1}^1(dx/x)=\lim_{\varepsilon\to0}(\int_{-1}^{-\varepsilon}(1/x)\,dx+\int_\varepsilon^1(1/x)\,dx)=0$.

E. Integrals on Infinite Intervals

Suppose that we are given a function $f(x)$ defined on an infinite interval $[a,\infty)$ and integrable on any finite interval $[a,b]$. If $\lim_{b\to\infty}\int_a^b f(x)\,dx$ exists and is finite, then this limit is called the **improper integral** of f on $[a,\infty)$ and is denoted by $\int_a^\infty f(x)\,dx$. We define similarly $\int_{-\infty}^b f(x)\,dx=\lim_{a\to-\infty}\int_a^b f(x)\,dx$, where f is defined on $(-\infty,b]$ and integrable on any interval $[a,b]$. Furthermore, $\int_{-\infty}^\infty f(x)\,dx$ is, by definition, $\int_{-\infty}^c f(x)\,dx+\int_c^\infty f(x)\,dx$, which is independent of the choice of c. Suppose that $f(x)$ is integrable on $[a,b]$ for a fixed a and an arbitrary b larger than a. If $f(x)=O(x^\alpha)$ for some $\alpha<-1$, then $\int_a^\infty f(x)\,dx$ exists. Generally, for α, β such that $-\infty\leqslant\alpha<\infty$ and $-\infty<\beta\leqslant\infty$, if the improper integral $\int_\alpha^\beta f(x)\,dx$ exists, we say that the integral is **convergent**; otherwise, it is **divergent**. Improper integrals also satisfy the three basic properties of integrals (1), (2), and (3) (→ Section B). However, the existence of an improper integral of a function f on an interval I does not imply the existence of the improper integral of $|f|$ on the same interval I. For example, let f be a function determined by $f(0)=0$, $f(x)=(1/x)\sin(1/x)$ for $0<x\leqslant\pi$. Then $\int_0^\pi f(x)\,dx$ exists, but $\int_0^\pi|f(x)|\,dx$ does not. On the other hand, if the improper integral of $|f(x)|$ exists,

then the improper integral of $f(x)$ exists, and we have

$$\left|\int_\alpha^\beta f(x)\,dx\right| \leqslant \int_\alpha^\beta |f(x)|\,dx,$$

where $-\infty \leqslant \alpha < \beta \leqslant +\infty$. In this case, we say that f is **absolutely integrable** on the interval $[\alpha, \beta]$. Assume now that $f(x)$ is defined on $(-\infty, \infty)$ and integrable on any finite interval. If $\lim_{\alpha\to\infty}\int_{-\alpha}^{\alpha} f(x)\,dx$ exists, then it is called **Cauchy's principal value** of the integral of f in $(-\infty, \infty)$.

If $f(x)$ is a monotone decreasing, positive, and continuous function defined on $[k, \infty)$ (where k is an integer), then according as $\sum_{v=k}^{\infty} f(v)$ converges or diverges, so does $\int_k^\infty f(x)\,dx$.

Suppose that a series $\sum_{n=1}^{\infty} f_n(x)$, where all the functions $f_n(x)$ $(n=1,2,\dots)$ are defined and nonnegative on an infinite interval $[a, \infty)$, satisfies $\int_a^b(\sum_{n=1}^{\infty} f_n(x))\,dx = \sum_{n=1}^{\infty}\int_a^b f_n(x)\,dx$ for arbitrary $b > a$. Then according as $\sum_{n=1}^{\infty}\int_a^\infty f_n(x)\,dx$ converges or diverges, so does $\int_a^\infty(\sum_{n=1}^{\infty} f_n(x))\,dx$. When they converge, the following equality holds: $\int_a^\infty \sum_{n=1}^{\infty} f_n(x)\,dx = \sum_{n=1}^{\infty}\int_a^\infty f_n(x)\,dx$. In this theorem, if $\int_a^\infty \sum_{n=1}^{\infty} |f_n(x)|\,dx$ or $\sum_{n=1}^{\infty}\int_a^\infty |f_n(x)|\,dx$ converges, then the same conclusion as above will follow even when the $f_n(x)$ are not necessarily positive (→ Appendix A, Table 9).

F. Multiple Integrals

Suppose that $f(x, y)$ is a function defined and bounded on an interval $I = \{(x, y) \mid a \leqslant x \leqslant b, c \leqslant y \leqslant d\}$ in the xy-plane. Partitions $\{x_j\}$ and $\{y_k\}$ of $[a, b]$ and $[c, d]$ with $a = x_0 < x_1 < \dots < x_m = b$ and $c = y_0 < y_1 < \dots < y_n = d$ determine a "partition," denoted by D, of I into subintervals of the form $I_{jk} = \{(x, y) \mid x_{j-1} \leqslant x \leqslant x_j, y_{k-1} \leqslant y \leqslant y_k\}$ $(j = 1, \dots, m;\ k = 1, \dots, n)$. Writing

$$M_{jk} = \sup_{(x,y)\in I_{jk}} f(x, y),$$

$$m_{jk} = \inf_{(x,y)\in I_{jk}} f(x, y),$$

we set

$$\bar{\sigma}(D) = \sum_{j=1}^{m}\sum_{k=1}^{n} M_{jk}(x_j - x_{j-1})(y_k - y_{k-1}),$$

$$\underline{\sigma}(D) = \sum_{j=1}^{m}\sum_{k=1}^{n} m_{jk}(x_j - x_{j-1})(y_k - y_{k-1}).$$

Then we obtain $\inf_D \bar{\sigma}(D) \geqslant \sup_D \underline{\sigma}(D)$. If $\inf_D \bar{\sigma}(D) = \sup_D \underline{\sigma}(D)$, then $f(x, y)$ is called integrable on I, and the common value is called the **double integral** of f on I and is denoted by $\iint_I f(x, y)\,dx\,dy$. Analogously, we can define **n-tuple integrals** (or **multiple integrals**) and the integrability of functions of n variables.

Let K be a bounded set in the xy-plane and I an interval containing K. Let $\varphi(x, y)$ be the characteristic function of K defined on I, that is, φ is determined by

$$\varphi(x, y) = 1 \quad \text{for} \quad (x, y) \in K,$$

$$\varphi(x, y) = 0 \quad \text{for} \quad (x, y) \in I - K.$$

Replacing $f(x, y)$ by this $\varphi(x, y)$, we consider $\inf_D \bar{\sigma}(D)$ $(\sup_D \underline{\sigma}(D))$. These values can be shown to be independent of the choice of such an interval I and are called the **outer area** (**inner area**) of K, respectively. When these two values coincide, K is said to be **of definite area**, and the common value is called the **area of** K. A necessary and sufficient condition for K to be of definite area is that the outer area of the †boundary of K be zero. Now consider a bounded function defined on a set K of definite area. Then, taking an interval I containing K, define an extension $\varphi(x, y)$ of $f(x, y)$ as follows:

$$\varphi(x, y) = f(x, y) \quad \text{for} \quad (x, y) \in K,$$

$$\varphi(x, y) = 0 \quad \text{for} \quad (x, y) \in I - K.$$

If $\varphi(x, y)$ is integrable on I, then $f(x, y)$ is called **integrable** on K, and the integral of f on K is defined by $\iint_K f(x, y)\,dx\,dy = \iint_I \varphi(x, y)\,dx\,dy$, which is independent of the special choice of I. The set K is called the **domain of integration**. Since K is of definite area, the set of boundary points of K at which $\varphi(x, y)$ is not continuous can be contained in a union of intervals whose total area can be made smaller than any preassigned positive number. Consequently, a function bounded on K and continuous at each †interior point of K is integrable on K. Like integrals of functions of a single variable, multiple integrals satisfy the three basic properties of integrals (→ Section B).

G. Multiple Integrals and Iterated Integrals

Suppose that we are given a function $f(x, y)$ that is continuous on an interval $I = \{(x, y) \mid a \leqslant x \leqslant b, c \leqslant y \leqslant d\}$. Then, for a fixed y in $[c, d]$, the function $f(x, y)$, regarded as a function of x, can be integrated with respect to x on the interval $[a, b]$, and the integral thus obtained is a continuous function of y. The integral of the function defined on $[c, d]$, namely, $\int_c^d(\int_a^b f(x, y)\,dx)\,dy$, is called the **iterated integral** (or **repeated integral**) of $f(x, y)$ and is often written as $\int_c^d dy \int_a^b f(x, y)\,dx$. The following formula gives a representation of a double integral by iterated ones:

$$\iint_I f(x, y)\,dx\,dy = \int_c^d dy \int_a^b f(x, y)\,dx$$

$$= \int_a^b dx \int_c^d f(x, y)\,dy.$$

More generally, $\varphi_1(x)$ and $\varphi_2(x)$ being continuous on $[a,b]$ and $\varphi_1(x)\leqslant\varphi_2(x)$, consider the following subset $K=\{(x,y)\mid a\leqslant x\leqslant b, \varphi_1(x)\leqslant y\leqslant\varphi_2(x)\}$ of the xy-plane. Suppose further that $f(x,y)$ is continuous on K. Then the following equality holds:

$$\iint_K f(x,y)\,dx\,dy=\int_a^b dx\int_{\varphi_1(x)}^{\varphi_2(x)} f(x,y)\,dy.$$

In the case of unbounded integrands or unbounded domains of integration, we can still define integrals under suitable restrictions. For instance, assume the following two properties: (1) There exists a sequence $\{K_n\}$ of sets, each of which is of definite area, satisfying $K_1\subset K_2\subset\ldots$ and $K=\bigcup_{n=1}^{\infty}K_n$. (2) $f(x,y)$ is bounded and integrable on each K_n ($n=1,2,\ldots$). If a finite limit $\lim_{n\to\infty}\iint_{K_n}f(x,y)\,dx\,dy$ exists and is independent of the choice of $\{K_n\}$, then $f(x,y)$ is called integrable on K. This limit is called the integral of $f(x,y)$ on K and is denoted by $\iint_K f(x,y)\,dx\,dy$:

$$\lim_{n\to\infty}\iint_{K_n} f(x,y)\,dx\,dy=\iint_K f(x,y)\,dx\,dy.$$

When the integral thus defined exists, we say that the integral is convergent. If a finite limit $\lim_{n\to\infty}\iint_{K_n}|f(x,y)|\,dx\,dy$ exists for some sequence $\{K_n\}$ with property (1), then f is integrable on K. Let $f(x,y)$ be continuous and nonnegative on $K=\{(x,y)\mid\alpha<x<\beta,\gamma<y<\delta\}$, where $-\infty\leqslant\alpha<\beta\leqslant\infty$, $-\infty\leqslant\gamma<\delta\leqslant\infty$. Furthermore, let $f(x,y)$ be integrable on K, and assume that the improper integral $F(x)=\int_\gamma^\delta f(x,y)\,dy=\lim_{c\downarrow\gamma,d\uparrow\delta}\int_c^d f(x,y)\,dy$ exists and converges uniformly with respect to x as $c\downarrow\gamma$, $d\uparrow\delta$. Then $\int_\alpha^\beta F(x)\,dx$ is well defined, and we have

$$\iint_K f(x,y)\,dx\,dy=\int_\alpha^\beta dx\int_\gamma^\delta f(x,y)\,dy.$$

In particular, if $\alpha=a$, $\gamma=b$, $\beta=\delta=\infty$, then we have

$$\int_a^\infty\int_b^\infty f(x,y)\,dx\,dy=\int_a^\infty dx\int_b^\infty f(x,y)\,dy.$$

H. Interchanging the Order of Differentiation and Integration

If both $f(x,y)$ and $\partial f(x,y)/\partial y$ are continuous on an interval $\{(x,y)\mid a\leqslant x\leqslant b, y_0-\eta\leqslant y\leqslant y_0+\eta\}$, then we can interchange the order of differentiation and integration as follows:

$$\frac{d}{dy}\int_a^b f(x,y)\,dx=\int_a^b\frac{\partial f(x,y)}{\partial y}dx\quad\text{for}\quad y=y_0.$$

Assume further that this equality holds for every $b(>a)$, the improper integral

$$\int_a^\infty f(x,y)\,dx=\lim_{b\to\infty}\int_a^b f(x,y)\,dx$$

converges, and the improper integral

$$\int_a^\infty\frac{\partial f(x,y)}{\partial y}dx=\lim_{b\to\infty}\int_a^b\frac{\partial f(x,y)}{\partial y}dx$$

converges as $b\to\infty$ uniformly for y with $|y-y_0|<\eta$. Then

$$\frac{d}{dy}\int_a^\infty f(x,y)\,dx=\int_a^\infty\frac{\partial f(x,y)}{\partial y}dx\quad\text{for}\quad y=y_0.$$

Several other similar theorems are known. Though the previous theorems are written in terms of two variables, analogous theorems hold for n variables.

I. Change of Variables in Multiple Integrals

Let G be a bounded domain of definite area in an n-dimensional Euclidean space $\mathbf{R}^n(x)$. Assume that a mapping $x\to y(x)=(y_1(x_1,\ldots,x_n),\ldots,y_n(x_1,\ldots,x_n))$ is of class C^1 from an open set containing the closure $\bar{G}$ of G into an n-dimensional Euclidean space $\mathbf{R}^n(y)$. We denote the image of G under this mapping by B. If $f(y_1,\ldots,y_n)$ is continuous on B, then the following formula on change of variables holds:

$$\int\cdots\int_B f(y_1,\ldots,y_n)\,dy_1\ldots dy_n=$$
$$\int\cdots\int_G g(x_1,\ldots,x_n)\left|\frac{D(y_1,\ldots,y_n)}{D(x_1,\ldots,x_n)}\right|dx_1\ldots dx_n,$$

where $g(x_1,\ldots,x_n)=f(y_1(x_1,\ldots,x_n),\ldots,y_n(x_1,\ldots,x_n))$ and $D(y_1,\ldots,y_n)/D(x_1,\ldots,x_n)$ is the †Jacobian determinant of the mapping $y(x)$. This formula is usually utilized in the case where $y_1,\ldots,y_n$ are †functionally independent, though otherwise both sides vanish and the formula still holds. For improper integrals, a similar formula will hold under suitable restrictions, for example, if the integrals converge absolutely.

For related topics → 94 Curvilinear Integrals and Surface Integrals, 221 Integration Theory, and 270 Measure Theory.

References

[1] T. M. Apostol, Mathematical analysis, Addison-Wesley, 1957.
[2] N. Bourbaki, Eléments de mathématique, Fonctions d'une variable réelle, Actualités Sci. Ind., 1074b, 1132a, Hermann, second edition, 1958, 1961.
[3] R. C. Buck, Advanced calculus, McGraw-Hill, second edition, 1965.

[4] R. Courant, Differential and integral calculus I, II, Nordemann, 1938.
[5] G. H. Hardy, A course of pure mathematics, Cambridge Univ. Press, seventh edition, 1938.
[6] E. Hille, Analysis I, II, Blaisdell, 1964, 1966.
[7] W. Kaplan, Advanced calculus, Addison-Wesley, 1952.
[8] E. Landau, Einführung in die Differentialrechnung und Integralrechnung, Noordhoff, 1934; English translation, Differential and integral calculus, Chelsea, 1965.
[9] J. M. H. Olmsted, Advanced calculus, Appleton-Century-Crofts, 1961.
[10] A. Ostrowski, Vorlesungen über Differential- und Integralrechnung I–III, Birkhäuser, second edition, 1960–1961.
[11] M. H. Protter and C. B. Morrey, Modern mathematical analysis, Addison-Wesley, 1964.
[12] W. Rudin, Principles of mathematical analysis, McGraw-Hill, second edition, 1964.
[13] V. I. Smirnov, A course of higher mathematics. I, Elementary calculus; II, Advanced calculus, Addison-Wesley, 1964.
[14] A. E. Taylor, Advanced calculus, Ginn, 1955.

217 (XIII.30) Integral Equations

A. General Remarks

Equations including the integrals of unknown functions are called **integral equations**. The most studied ones are the **linear integral equations**, i.e., linear in unknown functions.

Let D be a domain of n-dimensional Euclidean space and $f(x)$ and $K(x,y)$ be functions defined for $x=(x_1,x_2,\ldots,x_n)\in D$, $y=(y_1,y_2,\ldots,y_n)\in D$. **Integral equations of Fredholm type** (or **Fredholm integral equations**) [1] are those of the forms

$$\int_D K(x,y)\varphi(y)\,dy=f(x), \tag{1}$$

$$\varphi(x)-\int_D K(x,y)\varphi(y)\,dy=f(x), \tag{2}$$

$$A(x)\varphi(x)-\int_D K(x,y)\varphi(y)\,dy=f(x), \tag{3}$$

where $\varphi(x)$ is an unknown function and $\int_D dy$ means the n-fold integral $\int\ldots\int_D dy_1\ldots dy_n$. Equations of the forms (1), (2), and (3) are called equations of the **first**, **second**, and **third kind**, respectively. Equations of the second kind have been investigated in great detail. Equations of the third kind, in many cases, can be reduced formally to those of the second kind. The function $K(x,y)$ is called a **kernel** (or **integral kernel**) of the integral equation.

Integral equations of Volterra type (or **Volterra integral equations**) are those of the forms

$$\int_a^x K(x,y)\varphi(y)\,dy=f(x), \tag{1'}$$

$$\varphi(x)-\int_a^x K(x,y)\varphi(y)\,dy=f(x), \tag{2'}$$

$$A(x)\varphi(x)-\int_a^x K(x,y)\varphi(y)\,dy=f(x), \tag{3'}$$

where $\varphi(x)$ is an unknown function. Equations of the forms (1′), (2′), and (3′) are also called equations of the **first**, **second**, and **third kind**, respectively. Integral equations of Volterra type can be regarded as integral equations of Fredholm type having kernels equal to 0 for $x<y$, but these two types of equations are usually treated separately, since they have considerably different characters.

The kernels in equations (1)–(3) and (1′)–(3′) are frequently written in the form $\lambda K(x,y)$ with a parameter λ, in particular when the equations are related to eigenvalue problems, which is explained in Section F.

The theory of integral equations was originated in 1823 by N. H. Abel, who investigated the relationship between time and the path of a falling body in the field of gravitation. Let $\varphi(t)$ be a quantity varying with time, which is connected by some law with its values in some time interval of the past or the future. Then the law of variation of $\varphi(t)$ can be described mathematically by an integral equation. The situation is the same even when the variable t is not time but a coordinate of the space. In this way, various problems in physics can be reduced to solutions of integral equations.

B. Relation to Differential Equations

Many problems in differential equations can be reduced to problems related to integral equations. Such reduction often makes the problems easier to handle and clarifies the nature of the solutions. For example, consider the problem of finding a solution of the ordinary second-order linear differential equation $d^2y/dx^2+\lambda y=0$ with the boundary condition $y(0)=y(1)=0$ [4]. Let $d^2y/dx^2=u(x)$. If we integrate the equation twice, change the order of integration, and make use of the boundary condition, then we have

$$y=\int_0^x (x-\xi)u(\xi)\,d\xi-\int_0^1 x(1-\xi)u(\xi)\,d\xi,$$

from which we see that the given differential equation can be written in the form

$$u=\lambda\int_0^1 x(1-\xi)u(\xi)d\xi-\lambda\int_0^x(x-\xi)u(\xi)d\xi.$$

Decomposing the first integral on the right-hand side into the sum of an integral over $(0,x)$ and one over $(x,1)$, and combining the integral over $(0,x)$ with the second integral on the right-hand side, we obtain a Fredholm integral equation of the first kind as follows:

$$u=\lambda\int_0^1 G(x,\xi)u(\xi)d\xi,$$

$$G(x,\xi)=\begin{cases}\xi(1-x) & (0\leqslant\xi\leqslant x),\\ x(1-\xi) & (x\leqslant\xi\leqslant 1).\end{cases}$$

Clearly, the solution of this integral equation is equivalent to that of the original differential equation. The function G is called †Green's function for the boundary value condition $y(0)=y(1)=0$ in the theory of boundary value problems. Differential equations of higher orders can be treated analogously (→ 315 Ordinary Differential Equations (Boundary Value Problems)). †Initial value problems of linear ordinary differential equations can be reduced to the solution of Volterra integral equations in a similar way.

As another example [5, 6], consider the †Dirichlet problem on a plane, i.e., the problem of finding a function u satisfying the conditions (i) u is †harmonic in the interior of the region D bounded by a closed curve $C(\xi=\varphi(s),\eta=\psi(s), 0\leqslant s\leqslant l)$; (ii) $u(x,y)\rightarrow F(s)$ uniformly with respect to (x_0,y_0) as (x,y) approaches (x_0,y_0) from the inside of D, where (x_0,y_0) is an arbitrary point on C, $F(s)$ is a continuous function given on C, and s is the arc length along C. Put $f(s)=F(s)/\pi$ and

$$K(s;t)=\frac{(\psi(s)-\psi(t))\varphi'(t)-(\varphi(s)-\varphi(t))\psi'(t)}{\pi((\varphi(s)-\varphi(t))^2+(\psi(s)-\psi(t))^2)}.$$

Then it is known that a solution u of the problem can be given in the form

$$u(x,y)=\int_0^l\mu(s)\frac{\partial}{\partial n}\log\frac{1}{r}ds,$$

where $r^2=(\varphi(s)-x)^2+(\psi(s)-y)^2$, n is the inner normal of C, and $\mu(s)$ is a continuous solution of the following Fredholm integral equation of the second kind:

$$\mu(s)=f(s)-\int_0^l K(s;t)\mu(t)dt.$$

We can treat the †Neumann problem similarly, i.e., the problem in which condition (ii) is replaced by (ii') $(\partial u/\partial n)(x,y)\rightarrow F(s)$ uniformly with respect to (x_0,y_0) as (x,y) approaches (x_0,y_0) from the inside of D. In the Neumann problem, put $f(s)=F(s)/\pi$ and

$$L(s;t)=\frac{(\psi(s)-\psi(t))\varphi'(s)-(\varphi(s)-\varphi(t))\psi'(s)}{\pi((\varphi(s)-\varphi(t))^2+(\psi(s)-\psi(t))^2)}.$$

Then we have a solution u in the form

$$u(x,y)=-\int_0^l\mu(s)\log\frac{1}{r}ds+c,$$

where $\mu(s)$ is a solution of the following Fredholm integral equation of the second kind:

$$\mu(s)=f(s)-\int_0^l L(s;t)\mu(t)dt.$$

A solution of this integral equation, however, exists when and only when $\int_0^l F(s)ds=0$. In the expression of a solution $u(x,y)$ of the Neumann problem, c is an arbitrary additive constant, up to which a solution of the problem is determined uniquely. We can also treat partial differential equations of †elliptic type in an analogous way.

C. Integral Equations with Continuous Kernel

We describe some results for integral equations with m-dimensional independent variables, i.e., equations in which D is an m-dimensional closed domain. We assume that $K(x,y)$ and $f(x)$ are continuous in Sections D–H.

D. The Method of Successive Iteration

Among methods of solving Fredholm integral equations of the second kind, the simplest is the **method of successive iteration**, sometimes called the **method of successive approximation** [7]. In the method of successive iteration, we rewrite (2) in the form

$$\varphi(x)=f(x)+\int_D K(x,y)\varphi(y)dy$$

and replace the function $\varphi(y)$ on the right-hand side by the function

$$f(y)+\int_D K(y,z)\varphi(z)dz.$$

If we repeat the process successively, then we have

$$\varphi(x)=f(x)+\sum_{i=1}^n\int_D K_i(x,y)f(y)dy$$
$$+\int_D K_{n+1}(x,y)\varphi(y)dy,$$

where

$$K_1(x,y)=K(x,y),$$

$$K_i(x,y)=\int_D K_{i-1}(x,s)K(s,y)ds.$$

The functions $K_i(x,y)$ are called the **iterated kernels**. Assume that $\sum_{n=1}^{\infty} K_n(x,y)$ converges uniformly. Then, putting

$$R(x,y)=\sum_{n=1}^{\infty} K_n(x,y), \tag{4}$$

we obtain a solution of (2) in the form

$$\varphi(x)=f(x)+\int_D R(x,y)f(y)\,dy. \tag{5}$$

The series (4) is called a **Neumann series**.

For a given kernel $K(x,y)$, a function $R(x,y)$ satisfying

$$K(x,y)-R(x,y)+\int_D K(x,s)R(s,y)\,ds=0$$

and

$$K(x,y)-R(x,y)+\int_D R(x,s)K(s,y)\,ds=0$$

is called a **resolvent** of $K(x,y)$ (in some cases $-R(x,y)$ is taken as the resolvent). If a resolvent of $K(x,y)$ exists, the solution of (2) can be given uniquely by (5). If a Neumann series converges uniformly, (4) gives a resolvent of $K(x,y)$.

If we apply a similar process to Volterra integral equations of the second kind, then we have the iterated kernels defined by

$$K_{i+1}(x,y)=\int_y^x K_i(x,s)K(s,y)\,ds \quad (i=1,2,\ldots).$$

For these iterated kernels, a Neumann series defined by (4) always converges uniformly.

E. Fredholm's Method

Let D be a bounded closed domain and $K(x,y)$ a continuous kernel. A Neumann series (4) converges uniformly and gives a resolvent if $|K(x,y)|$ or the region D is sufficiently small, but otherwise it does not necessarily converge. E. I. Fredholm [1, 7] gave a method of constructing a resolvent for the more general case. Write a kernel in the form $\lambda K(x,y)$, and put

$$K\begin{pmatrix}x_1,\ldots,x_n\\ y_1,\ldots,y_n\end{pmatrix}=\begin{vmatrix}K(x_1,y_1) & \ldots & K(x_1,y_n)\\ & \ldots & \\ K(x_n,y_1) & \ldots & K(x_n,y_n)\end{vmatrix}.$$

Define $D(\lambda)$ and $D(x,y;\lambda)$ by

$$D(\lambda)=1+\sum_{n=1}^{\infty}\frac{(-\lambda)^n}{n!}\int_D\ldots\int_D K\begin{pmatrix}s_1,\ldots,s_n\\ s_1,\ldots,s_n\end{pmatrix}ds_1\ldots ds_n$$

and

$$D(x,y;\lambda)=K(x,y)+\sum_{n=1}^{\infty}\frac{(-\lambda)^n}{n!}\int_D\ldots\int_D K\begin{pmatrix}x,s_1,\ldots,s_n\\ y,s_1,\ldots,s_n\end{pmatrix}ds_1\ldots ds_n.$$

The series in these two equations both converge uniformly and hence define †entire functions of λ. The functions $D(\lambda)$ and $D(x,y;\lambda)$ are called **Fredholm's determinant** and **Fredholm's first minor** of the kernel $K(x,y)$, respectively. For small $|\lambda|$, we have

$$\frac{D'(\lambda)}{D(\lambda)}=\sum_{n=1}^{\infty}\lambda^{n-1}\int_D K_n(s,s)\,ds,$$

where the $K_n(x,y)$ are iterated kernels corresponding to $K(x,y)$. Now if $D(\lambda)\neq 0$, a resolvent $\lambda R(x,y;\lambda)$ of the kernel $\lambda K(x,y)$ can be given by

$$\frac{D(x,y;\lambda)}{D(\lambda)}=R(x,y;\lambda).$$

If $D(\lambda)=0$, some extension of the method in this section is needed. Fredholm introduced for this purpose **Fredholm's rth minor**

$$D\begin{pmatrix}x_1,\ldots,x_r\\ y_1,\ldots,y_r\end{pmatrix};\lambda\Big)$$

defined by

$$D\Big(\begin{matrix}x_1,\ldots,x_r\\ y_1,\ldots,y_r\end{matrix};\lambda\Big)=K\begin{pmatrix}x_1,\ldots,x_r\\ y_1,\ldots,y_r\end{pmatrix}+\sum_{n=1}^{\infty}\frac{(-\lambda)^n}{n!}\times\int_D\ldots\int_D K\begin{pmatrix}x_1,\ldots,x_r,s_1,\ldots,s_n\\ y_1,\ldots,y_r,s_1,\ldots,s_n\end{pmatrix}ds_1\ldots ds_n.$$

F. Eigenvalue Problems and Fredholm's Alternative Theorem

Consider a **homogeneous integral equation** of the second kind

$$\varphi(x)-\lambda\int_D K(x,y)\varphi(y)\,dy=0, \tag{6}$$

where D is a bounded closed domain and $K(x,y)$ is continuous in $D\times D$. When (6) has a nontrivial solution $\varphi(x)$ for some λ, then λ is called an **eigenvalue** corresponding to the kernel $K(x,y)$, and the corresponding nontrivial solution $\varphi(x)$ is called an **eigenfunction** corresponding to the kernel $K(x,y)$. If $D(\lambda)\neq 0$, then (6) has no nontrivial solution, from which it follows that eigenvalues must be zero points of the entire function $D(\lambda)$. For an arbitrary eigenvalue λ, there is a set of linearly independent eigenfunctions corresponding to λ such that any eigenfunction corresponding to λ can be written as a linear combination of the eigenfunctions belonging to the set under consideration. Such a set of linearly independent eigenfunctions corresponding to an eigenvalue λ is called a **fundamental system** corresponding to the eigenvalue λ. The number of elements of the fundamental system is called

the **index of the eigenvalue** λ. The index of an eigenvalue is always finite. The homogeneous integral equation of the form

$$\psi(y)-\lambda\int_D K(x,y)\psi(x)\,dx=0 \qquad (6')$$

is called an **associated** (or **transposed**) **integral equation** of (6). The associated equation has the same eigenvalues as the original equation; moreover, the index of a common eigenvalue is the same for both equations. For any eigenvalue λ, the order of the zero point λ of the entire function $D(\lambda)$ is called the **multiplicity of the eigenvalue** λ. If an eigenvalue λ is a pole of $R(x,y;\lambda)$, then we have $p+1\geqslant r+q$, where r is the order of the pole, p is the multiplicity of λ, and q is the index of λ. In particular, if λ is a simple pole of $R(x,y,\lambda)$, we have $p=q$, namely, the multiplicity is equal to the index. An example with this particular property is the integral equation with a symmetric kernel to be discussed in Section G. For the set of eigenvalues there is no finite accumulation point even if there are infinitely many eigenvalues.

If λ is not an eigenvalue, the inhomogeneous equation

$$\varphi(x)-\lambda\int_D K(x,y)\varphi(y)\,dy=f(x) \qquad (7)$$

can be solved uniquely for any continuous function $f(x)$. In this case we have $D(\lambda)\neq 0$, and the resolvent $R(x,y;\lambda)$ of the kernel $\lambda K(x,y)$ exists. If λ is an eigenvalue, we have $D(\lambda)=0$, and equation (7) has a solution if and only if

$$\int_D \psi(x)f(x)\,dx=0$$

for all solutions $\psi(y)$ of (6′) (linearly independent solutions $\psi(y)$ are finite in number). The last statement is called **Fredholm's alternative theorem** (→ 68 Compact and Nuclear Operators).

A kernel of the type

$$K(x,y)=\sum_{j=1}^{n} X_j(x)Y_j(y)$$

is called a **separated kernel**, **degenerate kernel**, or **Pincherle-Goursat kernel**. For such a kernel, we have $D(\lambda)=\det(\delta_{jk}-\lambda\int_a^b X_j(t)Y_k(t)\,dt)$, and hence we can easily obtain eigenvalues and eigenfunctions. A nondegenerate kernel can be studied using the results obtained for separated kernels, since we can regard a kernel of the general form as the limit of a sequence of separated kernels.

G. Symmetric Kernels

A kernel $K(x,y)$ is called a **symmetric kernel** if it is real and $K(x,y)=K(y,x)$. Let D be a bounded closed domain and $K(x,y)$ be a continuous symmetric kernel. In this case the associated equation (6′) clearly coincides with the original equation (6) [5, 6, 8].

Corresponding to any nontrivial symmetric kernel $K(x,y)$, there exist at least one eigenvalue and one eigenfunction. The eigenvalues are all real, and the eigenfunctions corresponding to distinct eigenvalues are mutually orthogonal. If we orthonormalize the eigenfunctions belonging to all fundamental systems and number them according to the order of increasing absolute values of the corresponding eigenvalues, then we have an orthonormal system $\{\varphi_i(x)\}$, called a **complete orthonormal system of fundamental functions** or simply a **complete orthogonal** (or **orthonormal**) **system**. If we number the eigenvalues taking their multiplicities into account and according to the order of increasing absolute values, then we have the equality

$$\sum_{i=1}^{\infty}\frac{1}{\lambda_i^2}=\int_D\int_D K^2(x,y)\,dx\,dy.$$

Corresponding to an iterated kernel $K_m(x,y)$, we have the eigenvalues $\{\lambda_i^m\}$, and we can choose the corresponding orthonormal system so that it coincides with the one corresponding to $K(x,y)$. Eigenvalues and eigenfunctions corresponding to an iterated kernel can be obtained in the following way: Put $\int_D K_n(s,s)\,ds=u_n$; then the following limit exists:

$$\lim_{n\to\infty} u_{2n}/u_{2n+2}=\lambda^2<+\infty,$$

which gives an eigenvalue of the iterated kernel $K_2(x,y)$. A function $\varphi(x,y)$ defined by

$$\lim_{n\to\infty}\lambda^{2n}K_{2n}(x,y)=\varphi(x,y)$$

(uniformly convergent) gives the corresponding eigenfunction $\varphi(x,c)$ for any constant c satisfying $\varphi(x,c)\not\equiv 0$.

Let λ^n be an eigenvalue corresponding to an iterated kernel $K_n(x,y)$ and $\varphi(x)$ be a corresponding eigenfunction. Consider the functions $\psi_j(x)$ $(j=0,1,\dots,n-1)$ defined by

$$\psi_j(x)=\varphi(x)+\sum_{k=1}^{n-1}\varepsilon^{kj}\lambda^k\int_D K_k(x,y)\varphi(y)\,dy \qquad (j=0,1,2,\dots,n-1),$$

where ε is one of the nth primitive roots of 1. Then for at least one value of j, $\varepsilon^j\lambda$ is an eigenvalue corresponding to the kernel $K(x,y)$, and ψ_j is a corresponding eigenfunction. This relationship between eigenvalues and eigenfunctions corresponding to an iterated kernel and those corresponding to the original kernel is valid even for kernels that are not necessarily symmetric [2].

Let a kernel $K^{(n)}(x,y)$ be defined by

$$K^{(n)}(x,y)=K(x,y)-\sum_{i=1}^{n}\frac{\varphi_i(x)\varphi_i(y)}{\lambda_i},$$

where the λ_i $(i=1,2,\ldots,n)$ are the eigenvalues corresponding to the kernel $K(x,y)$ and the $\varphi_i(x)$ $(i=1,\ldots,n)$ are the corresponding orthonormalized eigenfunctions. Then eigenvalues and eigenfunctions corresponding to $K^{(n)}(x,y)$ are those corresponding to $K(x,y)$, with the exception of $\lambda_1,\ldots,\lambda_n$ and $\varphi_1(x),\ldots,\varphi_n(x)$.

Let $\varphi(x)$ be any function that satisfies $\int_D(\varphi(x))^2dx=1$. Then the integral

$$J=\int_D\int_D K_2(x,y)\varphi(x)\varphi(y)\,dx\,dy$$

assumes the maximum value when $\varphi(x)$ is an eigenfunction corresponding to $K_2(x,y)$ with the smallest eigenvalue λ_1^2. Let the eigenvalues λ_n of $K(x,y)$ be numbered in order of increasing absolute values, so that $|\lambda_n|\leqslant|\lambda_{n+1}|$. Let $\varphi(x)$ be any function satisfying

$$\int_D \psi_i(x)\varphi(x)\,dx=0 \qquad (i=1,2,\ldots,n),$$

$$\int_D(\varphi(x))^2dx=1$$

for given functions $\psi_i(x)$ $(i=1,\ldots,n)$. Then the maximum value of the integral J above is the least when the set of all linear combinations of $\{\psi_1(x),\ldots,\psi_n(x)\}$ coincides with the set of all linear combinations of $\{\varphi_1(x),\ldots,\varphi_n(x)\}$, and in this case the maximum of J is attained by some eigenfunction $\varphi(x)$ corresponding to $K_2(x,y)$ with the eigenvalue λ_{n+1}^2. The results in this paragraph show that we can obtain eigenvalues by solving a variational problem concerning the integral J.

H. Expansion Theorems

Let $K(x,y)$ be a continuous symmetric kernel and $h(x)$ be a function square integrable on a bounded closed domain D. Then a function $f(x)$ such that

$$f(x)=\int_D K(x,y)h(y)\,dy$$

can be expanded in the form

$$f(x)=\sum_{n=1}^{\infty}c_n\varphi_n(x),$$

where $\{\varphi_i(x)\}$ is a complete orthonormal system of fundamental functions corresponding to $K(x,y)$ and

$$c_n=\int_D f(x)\varphi_n(x)\,dx \qquad (n=1,2,\ldots).$$

The series in the expansion of $f(x)$ converges uniformly. These facts are the content of the **Hilbert-Schmidt expansion theorem** [5, 6, 8]. By using this theorem, for a λ that is not an eigenvalue, we can obtain a solution $\varphi(x)$ of the Fredholm integral equation (7) with a symmetric kernel in the form

$$\varphi(x)=f(x)+\lambda\sum_{i=1}^{\infty}\frac{\varphi_i(x)}{\lambda_i-\lambda}\int_D f(x)\varphi_i(x)\,dx.$$

For $m\geqslant 2$, iterated kernels can be expanded in the form

$$K_m(x,y)=\sum_{i=1}^{\infty}\frac{\varphi_i(x)\varphi_i(y)}{\lambda_i^m}$$

(uniformly convergent). If $\lambda R(x,y;\lambda)$ is a resolvent of a symmetric kernel $\lambda K(x,y)$, then $R(x,y;\lambda)$ can be expanded as

$$R(x,y;\lambda)=K(x,y)+\lambda\sum_{i=1}^{\infty}\frac{\varphi_i(x)\varphi_i(y)}{\lambda_i(\lambda_i-\lambda)}.$$

If a symmetric kernel $K(x,y)$ satisfies the inequality

$$\int_D\int_D K(x,y)\varphi(x)\varphi(y)\,dx\,dy\geqslant 0$$

for all $\varphi(x)$, it is called a **positive (semidefinite) kernel**. If in this inequality the equality holds only for $\varphi(x)\equiv 0$, $K(x,y)$ is called a **positive definite kernel**. For a positive definite kernel, eigenvalues are all positive, and the kernel can be expanded in the form

$$K(x,y)=\sum_{i=1}^{\infty}\frac{\varphi_i(x)\varphi_i(y)}{\lambda_i}$$

(uniformly convergent). This result is called **Mercer's theorem**.

When a real continuous kernel $K(x,y)$ is not symmetric, we consider two positive kernels $\hat{K}'(x,y)$ and $\hat{K}''(x,y)$ defined by

$$\int_D K(x,s)K(y,s)\,ds=\hat{K}'(x,y)$$

and

$$\int_D K(s,x)K(s,y)\,ds=\hat{K}''(x,y).$$

The eigenvalues corresponding to these kernels are the same, and they are all positive. Let λ_i^2 $(i=1,2,\ldots)$ be these eigenvalues and $\{\varphi_i(x)\}$ and $\{\psi_i(x)\}$ be the corresponding complete orthonormal systems corresponding to $\hat{K}'$ and $\hat{K}''$, respectively. Then we have

$$\lambda_i\int_D K(y,x)\varphi_i(y)\,dy=\psi_i(x),$$

$$\lambda_i\int_D K(x,y)\psi_i(y)\,dy=\varphi_i(x).$$

Let $f(x)$ be an arbitrary function such that

$$f(x)=\int_D K(x,y)h(y)\,dy,$$

where $h(x)$ is a function square integrable on D. The function $f(x)$ can then be expanded in the form

$$f(x)=\sum_{i=1}^{\infty} c_i\varphi_i(x), \qquad (8)$$

where

$$c_i=\int_D f(x)\varphi_i(x)\,dx \qquad (i=1,2,\ldots).$$

The series in the expansion (8) converges uniformly.

The Fredholm integral equation (1) of the first kind with a general kernel (i.e., a kernel that is not necessarily symmetric) has a square integrable solution $\varphi(x)$ if and only if $f(x)$ has a uniformly convergent expansion (8) and $\sum_{i=1}^{\infty}(c_i\lambda_i)^2<+\infty$. When this condition is satisfied, equation (1) has a solution given by $\sum_{n=1}^{\infty}c_n\lambda_n\varphi_n(x)$ that converges in the sense of †mean convergence.

It should be noted that the theory concerning symmetric kernels can be extended to complex-valued **Hermitian kernels**, i.e., kernels such that $K(x,y)=\overline{K(y,x)}$. Also, we can obtain the theory in Section G and this section, concerning Fredholm integral equations with continuous kernels, by using the methods of functional analysis that treat $\int_D K(x,y)\varphi(y)\,dy$ as a †compact operator in the space of continuous functions (→ 68 Compact and Nuclear Operators).

I. Kernels of Hilbert-Schmidt Type

Kernels of Hilbert-Schmidt type are kernels which are square integrable in the sense of Lebesgue over $D\times D$, where D is an arbitrary domain. Most of the results mentioned in the previous section concerning integral equations with kernels continuous on bounded domains are valid also for equations with kernels of Hilbert-Schmidt type, because every operator mentioned in the previous section is also a compact operator in the space concerned, i.e., the space $L_2(D)$ [6] (→ 68 Compact and Nuclear Operators).

J. Singular Kernels

For general kernels that are not necessarily continuous, the theory described in the previous sections does not apply properly, but when an iterated kernel $K_m(x,y)$ has a resolvent $R_m(x,y)$, we can find a resolvent $R(x,y)$ of $K(x,y)$ in the form

$$R(x,y)=R_m(x,y)+H_m(x,y)+\int_D R_m(x,s)H_m(s,y)\,ds,$$

where $H_m(x,y)=\sum_{i=1}^{m-1}K^i(x,y)$. When a kernel is of the form $\lambda K(x,y)$, the relationship between eigenvalues and eigenfunctions corresponding to an iterated kernel and those corresponding to the original kernel, which was stated in Sections G and H, is still valid for general kernels. If a kernel $K(x,y)$ is continuous for $x\neq y$ and has a singularity of the form $|x-y|^{-\alpha}$ $(0<\alpha<1)$ on $x=y$, the iterated kernels $K_m(x,y)$ are continuous provided that $(1-\alpha)m\geqslant 1$. †Green's functions of partial differential equations of elliptic type have this property.

A kernel that is not square integrable is called a **singular kernel**. An integral equation whose domain of definition is unbounded or whose kernel is singular is called a **singular integral equation** [9]. Singular integral equations have some particular properties that are not seen in ordinary integral equations, i.e., integral equations with kernels continuous in a bounded closed domain. For example [10], consider the identity

$$\sqrt{\frac{2}{\pi}}\int_0^{\infty}\sin xy\left(\sqrt{\frac{\pi}{2}}\,e^{-\alpha y}+\frac{y}{\alpha^2+y^2}\right)dy=\sqrt{\frac{\pi}{2}}\,e^{-\alpha x}+\frac{x}{\alpha^2+x^2},$$

where α is an arbitrary real number. This equality shows that for the continuous kernel $\sqrt{2/\pi}\sin xy$, 1 is an eigenvalue and $\sqrt{\pi/2}\;e^{-\alpha x}+x/(\alpha^2+x^2)$ is a corresponding eigenfunction. Since α is arbitrary, the index of the eigenvalue 1 is evidently infinity. As another example, observe the equality

$$\int_{-\infty}^{\infty}e^{-|x-y|}e^{-i\alpha y}\,dy=\frac{2}{1+\alpha^2}e^{-i\alpha x},$$

where α is an arbitrary real number. From this equality, we see that for the continuous kernel $e^{-|x-y|}$ defined on $(-\infty,\infty)$, and number $\lambda=(1+\alpha^2)/2$ greater than or equal to $1/2$ is an eigenvalue. In this example, the **spectrum**, i.e., the set of eigenvalues, is a continuum. Such a spectrum is called a **continuous spectrum**.

In applications, an important role is played by integral equations with **kernels of Carleman type**:

$$K(x,y)=G(x,y)/(y-x),$$

where $G(x,y)$ is a bounded function. In integral equations with such kernels, the integral

is taken in the sense of the †Cauchy principal value [5, 9, 11, 12]. For example, the **Riemann-Hilbert problem** is the following: We are given a simple closed and smooth curve L in the complex plane and real-valued smooth functions a, b, and c defined on L, $a+ib$ never vanishing. The problem is to find a function $\varphi(z)$ which is holomorphic in the exterior of L, at most of polynomial growth at infinity, and continuous up to L with boundary value φ^+ such that $\mathrm{Re}[(a+ib)\varphi^+]=c$ on L. This problem is reduced to that of finding a function μ defined on L satisfying an integral equation

$$\mu(z)-\int_L \frac{G(z,\zeta)}{z-\zeta}\mu(\zeta)\,d\zeta=f(z) \qquad (z\in L),$$

where G is a smooth kernel determined by $a+ib$, f is a known function depending on c, and the integral is taken in the sense of the Cauchy principal value. The integer κ defined by $\kappa=(1/\pi)\int_L d(\arg(a-ib))$ is called the **index** of this problem. The full solution of the Riemann-Hilbert problem was given by I. N. Vekua [11].

A multidimensional analog of the Cauchy principal value is the **singular integral** of A. P. Calderón and A. Zygmund (→ 251 Linear Operators). A smooth function $k(x)$ defined in $\mathbf{R}^n$ except at $x=0$ is called a **kernel of Calderón-Zygmund type** if $k(x)$ is positively homogeneous of order $-n$ and if its integral mean on the unit sphere is zero. Then the operator K defined by

$$Kf(x)=\lim_{\varepsilon\to 0}\int_{|y|>\varepsilon} k(y)f(x-y)\,dy$$

is called a Calderón-Zygmund **singular integral operator** [13]. K is a bounded linear operator in $L^p(\mathbf{R}^n)$ if $1<p<+\infty$. If $n=1$ and $k(x)=1/(\pi i x)$, K is nothing but the †Hilbert transformation. The pseudodifferential operator (→ 345 Pseudodifferential Operators) is an extension in some sense of the singular integral operator.

K. Systems of Integral Equations

A system of Fredholm integral equations of the second kind can always be reduced to a single equation. In fact, as is seen easily, a system of integral equations

$$\varphi_i(x)-\lambda\sum_j\int_0^1 K_{ij}(x,y)\varphi_j(y)\,dy=f_i(x) \qquad (i=1,2,\ldots,n)$$

can be reduced to a single equation

$$\Phi(x)-\lambda\int_0^n K(x,y)\Phi(y)\,dy=F(x) \quad (0\leqslant x\leqslant n),$$

where

$$\Phi(x)=\varphi_i(x-i+1), \qquad F(x)=f_i(x-i+1),$$
$$K(x,y)=K_{ij}(x-i+1,y-j+1)$$
$$(i-1\leqslant x, j-1\leqslant y\leqslant j;\ i,j=1,2,\ldots,n).$$

A system of Volterra integral equations of the second kind can be reduced to a single equation by eliminating the unknown functions successively.

L. Integral Equations of Volterra Type

Consider a Volterra integral equation of the first kind

$$\int_a^x K(x,y)\varphi(y)\,dy=f(x)$$

such that $K(x,x)\neq 0$ and $K_x(x,y)$ and $f'(x)$ are continuous. If we differentiate both sides of the equation, then we have a Volterra integral equation of the second kind:

$$\varphi(x)+\int_a^x \frac{K_x(x,y)}{K(x,x)}\varphi(y)\,dy=\frac{f'(x)}{K(x,x)}.$$

Abel's integral equation of general form is

$$\int_0^x \frac{G(x,y)}{(x-y)^\alpha}\varphi(y)\,dy=f(x) \qquad (0<\alpha<1). \tag{9}$$

If G, G_x, and f' are continuous and $G(x,x)\neq 0$, equation (9) can be reduced to the equation

$$\int_0^u H(u,y)\varphi(y)\,dy=\int_0^u f(x)(u-x)^{\alpha-1}\,dx,$$

where

$$H(u,y)=\int_y^u \frac{G(x,y)\,dx}{(u-x)^{1-\alpha}(x-y)^\alpha}.$$

Since $H(u,u)=(\pi/\sin\alpha\pi)G(u,u)\neq 0$, it follows that

$$\varphi(u)+\int_0^u \frac{H_u(u,y)}{H(u,u)}\varphi(y)\,dy=g(u), \tag{9a}$$

where

$$g(u)=H(u,u)^{-1}\frac{d}{du}\int_0^u f(x)(u-x)^{\alpha-1}\,dx$$
$$=H(u,u)^{-1}\left(u^{\alpha-1}f(0)+\int_0^u (u-x)^{\alpha-1}f'(x)\,dx\right).$$

Clearly (9a) is a Volterra integral equation of the second kind. **Abel's problem** (→ Section A) was to find the path of a falling body for a given time of descent. The problem then can be reduced to the solution of equation (9) with $G(x,y)\equiv 1$ and $\alpha=1/2$. When $G(x,y)\equiv 1$, we

can solve equation (9) explicitly to get

$$\varphi(x)=\frac{\sin\alpha\pi}{\pi}\left(x^{\alpha-1}f(0)+\int_0^x(x-t)^{\alpha-1}f'(t)\,dt\right)$$

[5].

M. Nonlinear Integral Equations

When a **nonlinear integral equation** includes a parameter λ, it may happen that the parameter has a **bifurcation point**, i.e., a value λ_0 such that the number of real solutions is changed when λ varies through λ_0 taking real values. For example, consider the equation

$$\varphi(x)-\lambda\int_0^1\varphi^2(y)\,dy=1.$$

This equation has real solutions $\varphi(x)=(1\pm\sqrt{1-4\lambda})/2\lambda$ for $\lambda\leqslant 1/4$ but no real solution for $\lambda>1/4$. Hence $\lambda_0=1/4$ is a bifurcation point (→ 286 Nonlinear Functional Analysis).

Among nonlinear integral equations, **Hammerstein's integral equation** has been studied in detail [5, 14]. It is an equation of the form

$$\varphi(x)+\int_D K(x,y)f(y,\varphi(y))\,dy=0. \qquad (10)$$

If $K(x,y)$ and $f(y,0)$ are square integrable and $f(y,u)$ satisfies a †Lipschitz condition in u with a sufficiently small coefficient, then the integral equation (10) can be solved by successive approximations. If $K(x,y)$ is a square integrable positive kernel and $\int_D|K(x,y)|^2\,dy$ is bounded, then we can prove the existence and uniqueness of a solution of (10) under a condition on $f(y,u)$ weaker than a Lipschitz condition. We can prove similar results for equation (10) with a nonsymmetric kernel when $K(x,y)$ is **continuous in the mean**, that is,

$$\lim_{x'\to x}\int_D|K(x',y)-K(x,y)|^2\,dy=0,$$

$$\lim_{y'\to y}\int_D|K(x,y')-K(x,y)|^2\,dx=0.$$

A nonlinear Volterra integral equation of the form

$$\varphi(x)=f(x)+\int_a^x F(x,y;\varphi(y))\,dy$$

can be solved by successive approximations if $F(x,y;u)$ and $f(x)$ are square integrable, $F(x,y;u)$ satisfies a Lipschitz condition $|F(x,y;u')-F(x,y;u'')|\leqslant k(x,y)|u'-u''|$ with some square integrable function $k(x,y)$, and $\int_a^x F(x,y;f(y))\,dy$ is majorized by some square integrable function of x. When $F(x,y;u)$ and $f(x)$ are continuous, we can obtain theorems on the existence and uniqueness of continuous solutions and †comparison theorems similar to those for initial value problems in ordinary differential equations [16].

N. Numerical Solution

For the **numerical solution of integral equations**, we assume throughout that the functions appearing are all continuous and the solution of every equation is unique. Methods of numerical solution can be divided roughly into two classes. Methods of one class try to evaluate numerically the analytical solution described in the preceding sections, and methods of the other try to obtain a solution by transforming the problem to one that is numerically solvable.

(1) A Method Based on Numerical Quadrature. Consider the integral equation

$$\int_a^b F(x,y,\varphi(x),\varphi(y))\,dy=0.$$

Let $a=x_1<x_2<\ldots<x_n=b$ be points on the interval $[a,b]$ and $\varphi_1,\varphi_2,\ldots,\varphi_n$ be the values of $\varphi(x)$ at $x_1,x_2,\ldots,x_n$. By the use of numerical quadrature, we then have the following system of equations in φ_i:

$$\sum_{j=1}^n a_jF(x_i,x_j,\varphi_i,\varphi_j)=0 \qquad (i=1,2,\ldots,n).$$

The method corresponds to that of solving ordinary differential equations by their difference equation analogs. Hence the errors involved in the solutions obtained by this method can be analyzed similarly to those in the case of numerical solution of ordinary differential equations (→ 303 Numerical Solution of Ordinary Differential Equations). If the given integral equation is a Fredholm equation of the second kind, then we have a system of linear equations in φ_i. If we apply quadrature formulas to the integral appearing in the integral equation by using the values of φ_i obtained, then we have a formula by which the solution can be evaluated directly, that is, without using an interpolation formula.

(2) A Method Utilizing Recurrence Formulas. Let d_n and $d_n(x,y)$ be the respective coefficients of λ^n in the expansions of Fredholm's determinant $D(\lambda)$ and Fredholm's first minor $D(x,y;\lambda)$. They satisfy the recurrence formulas

$$d_n(x,y)=d_nK(x,y)+\int_a^b K(x,s)d_{n-1}(s,y)\,ds,$$

$$d_{n+1}=-\frac{1}{n+1}\int_a^b d_n(s,s)\,ds,$$

$$d_0=1, \qquad d_0(x,y)=K(x,y).$$

By the use of these formulas, we can compute

d_n and $d_n(x, y)$ successively and hence evaluate $D(\lambda)$ and $D(x, y; \lambda)$ approximately, and by means of these recurrence formulas we can readily obtain a solution of a Fredholm equation of the second kind.

(3) A Method Utilizing Approximate Kernels. If we replace a kernel by an approximate one in a Fredholm integral equation of the second kind, then we have an integral equation that has a solution approximately equal to the solution of the original equation. Hence if we can find an approximate kernel for which an integral equation can be solved numerically or analytically, then we can find an approximation to the desired solution by solving the modified equation. For such solutions, a method of error estimation was given by F. G. Tricomi [17].

(4) An Iterative Method. Consider the integral equation

$$\varphi(x) = \int_a^b F(x, y, \varphi(x), \varphi(y))\,dy.$$

Let $\varphi_0(x)$ be an adequate function, and define $\varphi_n(x)$ successively by

$$\varphi_{n+1}(x) = \int_a^b F(x, y, \varphi_n(x), \varphi_n(y))\,dy.$$

If the sequence $\{\varphi_n(x)\}$ converges, then the limit $\lim_{n\to\infty}\varphi_n(x) = \varphi(x)$ is a solution of the given equation, and hence we can obtain an approximation to a solution by calculating $\varphi_n(x)$ for some finite n. This method can be used effectively for Fredholm integral equations of the second kind with a parameter λ, provided that the absolute value of λ is smaller than the least absolute value of the eigenvalues.

(5) Variational Method. If some conditions are fulfilled, an integral equation of the form

$$G(x, \varphi(x)) + \int_a^b F(x, y, \varphi(x), \varphi(y))\,dy = 0$$

can be regarded as an †Euler-Lagrange equation for a variational problem

$$J[u] = \int_a^b \int_a^b E(x, y, u(x), u(y))\,dx\,dy + \int_a^b H(x, u(x))\,dx = \text{extremal}. \qquad (11)$$

In this case, we can find a solution of the given integral equation numerically by solving the variational problem (11) numerically [18].

(6) Enskog's Method. Suppose that $\{\varphi_n(x)\}$ is a complete orthonormal system for the Fredholm integral equation (7). If we put $\psi_n(x) = \varphi_n(x) - \lambda\int_a^b K(y, x)\varphi_n(y)\,dy$, then from (7) we have

$$\int_a^b \varphi(x)\psi_n(x)\,dx = \int_a^b f(x)\varphi_n(x)\,dx, \qquad (12)$$

and furthermore we see that $\{\psi_n(x)\}$ can be orthonormalized to yield a complete orthonormal system $\{\chi_n(x)\}$. The equality (12) then shows that the Fourier coefficients of a solution $\varphi(x)$ with respect to the system $\{\chi_n(x)\}$ can be obtained readily from the Fourier coefficients of $f(x)$ with respect to the system $\{\varphi_n(x)\}$. This method of obtaining a solution is called **Enskog's method**.

For Volterra integral equations, methods (1) and (4) can be used effectively. We usually transform equations of the first kind into equations of the second kind by differentiation and then apply the above numerical methods. This is done for the sake of securing stability of the numerical methods.

References

[1] I. Fredholm, Sur une classe d'équations fonctionnelles, Acta Math., 27 (1903), 365–390.
[2] G. Vivanti and F. Schwank, Elemente der Theorie der Lineare Integralgleichungen, Helwingsche Verlagsbuchhandlung, 1929.
[3] E. Hellinger and O. Toeplitz, Integralgleichungen und Gleichungen mit unendlichvielen Unbekannten, Teubner, 1928.
[4] R. Courant and D. Hilbert, Methods of mathematical physics I, II, Interscience, 1953, 1962.
[5] F. G. Tricomi, Integral equations, Interscience, 1957.
[6] K. Yosida, Lectures on differential and integral equations, Interscience, 1960. (Original in Japanese, 1950.)
[7] E. Goursat, Cours d'analyse mathématique III, Gauthier-Villars, 1928.
[8] F. Smithies, Integral equations, Cambridge, 1958.
[9] T. Carleman, Sur les équations intégrales singulières à noyau réel et symétrique, Uppsala, 1923.
[10] T. Lalesco, Introduction à la théorie des équations intégrales, Gauthier-Villars, 1912.
[11] N. I. Muskhelishvili, Singular integral equations, Noordhoff, 1953. (Original in Russian, 1946.)
[12] S. G. Mikhlin, Multidimensional singular integrals and integral equations, Pergamon, 1965. (Original in Russian, 1962.)
[13] A. P. Calderón and A. Zygmund, Singular integral operators and differential equations, Amer. J. Math., 79 (1957), 901–921.

[14] H. Schaefer, Neue Existenzsätze in der Theorie nichtlinearer Integralgleichungen, Akademie-Verlag, 1955.
[15] L. Lichtenstein, Vorlesungen über einige Klassen nichtlinearer Integralgleichungen und Integro-differentialgleichungen, Springer, 1931.
[16] T. Sato, Sur l'équation intégrale non linéaire de Volterra, Compositio Math., 11 (1953), 271–290.
[17] H. Büchner, Die praktische Behandlung von Integralgleichungen, Springer, 1952.
[18] L. Collatz, The numerical treatment of differential equations, Springer, 1960.
[19] L. V. Kantorovich and V. I. Krylov, Approximate methods of higher analysis, Interscience, 1958. (Original in Russian, 1949.)

218 (VII.19) Integral Geometry

A. General Remarks

Integral geometry, in the broad sense, is the branch of geometry concerned with integrals on manifolds, but the problems considered in **integral geometry** are usually of a more limited nature. If a †Lie group G acts on a †differentiable manifold M as a †Lie transformation group, G also acts on various figures on M, by which we mean geometric objects such as †submanifolds of M, †tangent r-frame bundles on M, etc. Let $\mathscr{F}$ be a set of such figures on M invariant under G (i.e., $gF \in \mathscr{F}$ for $g \in G$, $F \in \mathscr{F}$). Consider the following problems: (i) to know whether any G-invariant †measure μ on $\mathscr{F}$ exists, and how to determine μ if it exists; (ii) to find the integral $\int \varphi(F)\,d\mu(F)$ of functions φ on $\mathscr{F}$ with respect to the measure μ.

The term **integral geometry** was introduced by W. Blaschke, who considered the special case of problem (ii) in which $\varphi(F)$ is a function representing geometric properties of F and the integral is to be evaluated by means of the geometric invariants concerning $\mathscr{F}$[1]. Problems of so-called **geometric probability** (such as the problem of Buffon's needle) belong to this category. The measure μ is called the **kinetic measure** (or **kinetic density**), and $d\mu(F)$ is also denoted by dF. If $\mathscr{F}$ has the structure of an n-dimensional differentiable manifold and the measure μ is given by a †volume element ω (i.e., a positive †differential form of degree n), we denote ω also by dF. Problem (i) is simple: If G acts †transitively on $\mathscr{F}$, then $\mathscr{F} = G/H$, where H is the †isotropy subgroup of G. In this case $\mathscr{F}$ has the structure of a differentiable manifold, and if a G-invariant †(Radon) measure exists, it is unique up to a multiplicative constant. A condition for the existence of a G-invariant measure μ can be given by means of †Haar measures of G and H (→ 225 Invariant Measures). We now consider some examples.

B. Crofton's Formula

Let $G(p,\theta)$ be a straight line defined by the equation $x_1\cos\theta + x_2\sin\theta = p$ with respect to orthogonal coordinates in a Euclidean plane. Let $n(p,\theta)$ be the number of intersections of $G(p,\theta)$ with a curve C of length L. Then we have **Crofton's formula**,

$$\int n(p,\theta)\,dp\,d\theta = 2L, \tag{1}$$

where $dp\,d\theta$ is the †exterior product of the differential forms dp, $d\theta$ of degree 1, and the integral is extended over $p \in (-\infty, \infty)$ and $\theta \in [0, 2\pi]$.

C. Poincaré's Formula and the Principal Formula of Integral Geometry

The kinetic density dF of a figure F congruent (with the same orientation) to a fixed figure in a Euclidean plane is defined as follows: Let R be an orthogonal frame attached to F, (x_1, x_2) be the coordinates of the origin of R with respect to a fixed orthogonal frame R_0, and θ be the angle between the first axis of R and the first axis of R_0. If we put $dF = dx_1\,dx_2\,d\theta$ (exterior product), dF has the following invariance properties: (i) dF is not changed by displacements of F; (ii) dF is not changed if instead of R we take another orthogonal frame R' attached to F.

Let two plane curves C_1, C_2 of length L_1, L_2, respectively, be given, and suppose that C_1 is fixed and C_2 is mobile. If the number of intersections of C_2 in an arbitrary position with C_1 is finite and equal to n, then the integral of n extended over all possible positions of C_2 is given by

$$\int n\,dC_2 = 4L_1L_2 \tag{2}$$

(**Poincaré's formula**). L. A. Santaló applied this result to give a solution of the †isoperimetric problem (1936).

Let C_1, C_2 be two plane †Jordan curves of length L_1, L_2, respectively, and let S_1, S_2 be the areas of the domains bounded by C_1, C_2, respectively. Suppose that C_1 is fixed and C_2 mobile, and let χ be the number of connected domains common to the domains bounded by C_1, C_2 for C_2 in an arbitrary position. Then the integral of χ extended over all possible

positions of C_2 intersecting C_1 is given by

$$\int \chi \, dC_2 = L_1 L_2 + 2\pi(S_1 + S_2) \qquad (3)$$

(Blaschke [1]). This is the **principal formula of integral geometry**. Many formulas can be derived from it as special cases or limiting cases.

D. Generalization to Dimension n

The kinetic density of subspaces of dimension k in a Euclidean space and in a spherical space of dimension n was given by Blaschke, while the generalization of the principal formula (3) to a Euclidean space of dimension n was given by S. S. Chern, applying the methods of E. Cartan.

Let $(e_1, \ldots, e_n)$ be a positively oriented orthonormal frame with vertex A. The infinitesimal relative displacements are then given by

$$dA = \sum_{i=1}^{n} \omega_i e_i, \qquad de_i = \sum_{j=1}^{n} \omega_{ij} e_j,$$

where $\omega_i = (dA, e_i)$, $\omega_{ij} = (de_i, e_j) = -\omega_{ji}$ are differential forms of degree 1 in the orthogonal coordinates of A and the $n(n-1)/2$ variables that determine $e_1, \ldots, e_n$. For various positions of a figure, we take an orthogonal frame $(A, e_1, \ldots, e_n)$ fixed to this figure and form the †exterior product

$$dK = \bigwedge_i \omega_i \bigwedge_{i<j} \omega_{ij}$$

of all the ω_i and ω_{ij} $(i<j)$. This has the invariance properties (i) and (ii) of Section C and is, by definition, the kinetic density of the figure in an n-dimensional Euclidean space. Moreover, the kinetic density of dE of k-dimensional subspaces E can be obtained by considering the orthogonal frames such that the vertex A and $e_1, \ldots, e_k$ lie on E and by forming the exterior product of the corresponding ω_α, $\omega_{\alpha\lambda}$ $(\alpha = 1, \ldots, k; \lambda = k+1, \ldots, n)$,

$$dE = \bigwedge \omega_\alpha \bigwedge \omega_{\alpha\lambda}.$$

Let Σ be a compact orientable hypersurface of class C^2 in an n-dimensional Euclidean space, and let k_α $(\alpha = 1, \ldots, n-1)$ be the †principal curvatures at a point on Σ. Denote by S_i the †elementary symmetric form of degree i in k_α $(i = 1, \ldots, n-1)$, and put $S_0 = 1$. Then consider the integrals over Σ:

$$M_i = \int_\Sigma S_i \, dS \Big/ \binom{n-1}{i}, \qquad i = 0, 1, \ldots, n-1, \qquad (4)$$

where dS denotes the surface element of Σ. Let D_1, D_2 be the domains bounded by two compact orientable hypersurfaces Σ_1, Σ_2 of class C^2 with volume V_1, V_2, respectively, and let $M_i^{(1)}$, $M_i^{(2)}$ be the integrals M_i defined by (4) for Σ_1, Σ_2. If Σ_1 is fixed, Σ_2 is mobile, and the †Euler-Poincaré characteristic χ of $D_1 \cap D_2$ is finite, then the generalization of (3) has the form

$$\int \chi \, d\Sigma_2 = I_1 \ldots I_{n-1} \left(M_{n-1}^{(1)} V_2 + M_{n-1}^{(2)} V_1 + \frac{1}{n} \sum_{k=0}^{n-2} M_k^{(1)} M_{n-2-k}^{(2)} \right) \qquad (5)$$

(**Chern's formula**), where $d\Sigma_2$ is the kinetic density of Σ_2 and I_k $(k = 1, \ldots, n-1)$ is the area of the unit sphere in a Euclidean space of dimension $k+1$, with the integral extended over all positions of Σ_2 intersecting Σ_1.

Let dE be the kinetic density of the subspaces E of dimension k intersecting a compact orientable hypersurface Σ of class C^2, and let χ be the Euler characteristic of the intersection of E with the domain bounded by Σ. The integral $\int \chi \, dE$ extended over all hyperplanes of dimension k intersecting Σ is proportional to M_k relative to the hypersurface Σ defined by (4). This fact generalizes (1). Further generalizations were obtained by Chern (1966).

E. Other Generalizations

For 2-dimensional spaces of constant curvature, Santaló derived formulas analogous to those in a Euclidean plane (1942–1943) and thus solved the isoperimetric problem in these spaces. In 1952, he derived a formula corresponding to (5) in n-dimensional spaces of constant curvature, following Chern's method. He investigated further integral geometry in affine, projective, and Hermitian spaces.

Chern and others obtained the results of the previous sections by Cartan's method of general moving frames and studied integral geometry in the setting of the geometry of Lie transformation groups in the sense of F. Klein (→ 137 Erlangen Program).

Chern, P. Griffiths, and others studied the value distribution theory of holomorphic mappings in several complex variables from the point of view of integral geometry (1961) (→ 21 Analytic Functions of Several Complex Variables, 124 Distribution of Values of Functions of a Complex Variable).

F. Radon Transforms

Another important topic of integral geometry is the theory of Radon transforms. Let $\mathscr{F}$ be the set of hyperplanes $\xi(\omega, p) = \{x \in \mathbf{R}^n \mid (x, \omega) = p\}$ in the Euclidean space $\mathbf{R}^n$, where $\omega = (\lambda_1, \ldots, \lambda_n)$ is a unit vector, $(x, \omega) = \sum x_i \lambda_i$, and p is real. For a function f defined in $\mathbf{R}^n$, define

$\hat{f}(\xi)=\hat{f}(\omega,p)$ on $\mathscr{F}$ by

$$\hat{f}(\xi)=\int_{\xi} f(x)d_{\xi}x, \tag{6}$$

where $d_{\xi}x$ is the †volume element on the hyperplane ξ such that $d_{\xi}x \wedge \sum \lambda_i dx_i = dx$, with dx the volume element of $\mathbf{R}^n$. Then $\hat{f}$ is called the **Radon transform** of f. For example, if f is the †characteristic function of a bounded domain V, the value $\hat{f}(\xi)$ of the Radon transform $\hat{f}$ of f at $\xi \in \mathscr{F}$ is the volume of the section of V by ξ. Now the group G of †motions of $\mathbf{R}^n$ (the †connected component of the identity of the group of isometries) acts on $\mathscr{F}$ transitively. For every $x \in \mathbf{R}^n$, the †isotropy subgroup G_x of G with respect to x acts transitively on the set $\check{x}=\{\xi \in \mathscr{F} \mid x \in \xi\}$. Since G_x is compact, there exists a unique normalized G_x-invariant measure μ on $\check{x}$ such that $\mu(\check{x})=1$. For a function g on $\mathscr{F}$, the **conjugate Radon transform** $\check{g}$ can now be defined by

$$\check{g}(x)=\int_{x\in\xi} g(\xi)d\mu(\xi) \tag{7}$$

as a function on $\mathbf{R}^n$.

The determination of $\check{g}$ belongs to problem (ii) of integral geometry mentioned in Section A. In particular, it is important to determine $\check{g}=(\hat{f})^{\vee}$ for $g=\hat{f}$ and to find the relation between $(\hat{f})^{\vee}$ and f. These problems were solved by J. Radon for $n=2$, 3 and by F. John in the general case. The results can be formulated as follows:

In the case of odd n, let $\mathscr{S}$ be the space of †rapidly decreasing C^{∞}-functions (→ 168 Function Spaces). Let $\mathscr{S}^*(\mathscr{F})$ be the set of $g(\omega,p)\in\mathscr{S}(\mathscr{F})$ such that $\int_{-\infty}^{\infty} g(\omega,p)p^k dp=0$ for every natural number k and every ω. For every $f\in\mathscr{S}(\mathbf{R}^n)$ and every $g\in\mathscr{S}^*(\mathscr{F})$, we then have

$$f=c\Delta^{(n-1)/2}((\hat{f})^{\vee}), \qquad g=cL^{(n-1)/2}((\check{g})^{\wedge}),$$

where Δ is the †Laplacian in $\mathbf{R}^n$ and $Lg(\omega,p)=d^2g(\omega,p)/dp^2$, $c=\Gamma(n/2)^{-1}(2\pi i)^{1-n}\pi^{n/2}$.

In the case of even n, for every $f\in\mathscr{S}(\mathbf{R}^n)$, $g\in\mathscr{S}^*(\mathscr{F})$,

$$f=c_1J_1((\hat{f})^{\vee}), \qquad g=c_2J_2((\check{g})^{\wedge}),$$

where

$$J_1(f)(x)=\int_{\mathbf{R}^n} f(y)|x-y|^{1-2n}dy,$$

$$J_2(g)(\omega,p)=\int_{\mathbf{R}} g(\omega,p)|p-q|^{-n}dq.$$

These integrals are in general divergent, and they must be interpreted as regularizations defined by analytic continuation with respect to the powers of $|x-y|$ or $|p-q|$ [9].

John applied the Radon transform on $\mathbf{R}^n$ to the study of partial differential equations in $\mathbf{R}^n$ with constant coefficients [8].

A formula corresponding to †Plancherel's theorem and an analog of the †Paley-Wiener theorem for the Fourier transform are valid for the Radon transform (I. M. Gel'fand et al. [6]).

G. Horospheres

The theory of the Radon transform is also important in noncompact †symmetric Riemannian spaces M. The connected component G of the identity in the group of isometries of M is isomorphic to the †adjoint group $\mathrm{Ad}\,G$ and can be considered as a linear group. Maximal †unipotent subgroups of G are conjugate to each other. If N is such a subgroup, we call the †orbits on M of gNg^{-1} **horospheres** on M for $g\in G$. These correspond to the hyperplanes in $\mathbf{R}^n$. If M is the complex upper half-plane with the †hyperbolic non-Euclidean metric, the horospheres are precisely the circles tangent to the real axis and the straight lines parallel to the real axis.

The group G acts transitively on the set $\mathscr{F}$ of horospheres on M, and we have $\mathscr{F}=G/M_0N$. Here $G=KAN$ is an †Iwasawa decomposition of G, and M_0 is the †centralizer of A in K. For a horosphere $\xi\in\mathscr{F}$, let $d_{\xi}x$ be the volume element on ξ with respect to the †Riemannian metric on ξ induced by the Riemannian metric on M, and define the Radon transform $\hat{f}$ of a function f on M by (6) as a function on $\mathscr{F}$. For every $x\in M$, there exists a unique normalized measure on $\check{x}=\{\xi\in\mathscr{F}\mid x\in\xi\}$ invariant under the (compact) isotropy subgroup G_x of G at x ($\mu(\check{x})=1$). The conjugate Radon transform $\check{g}$ of a function g on $\mathscr{F}$ is defined by formula (7) by means of this measure μ. Then there exists an integrodifferential operator $\wedge$ such that if $\wedge^*$ is the adjoint operator, we have the inversion formula $f=(\wedge\wedge^*\hat{f})^{\vee}$ and **Plancherel's theorem**:

$$\int_M |f(x)|^2dx=\int_{\mathscr{F}} |\wedge\hat{f}(\xi)|^2d\xi,$$

where dx, $d\xi$ are G-invariant measures on M, $\mathscr{F}$, respectively, and f is an arbitrary C^{∞}-function with compact support. If the †Cartan subgroups of G are conjugate to each other, $\wedge$ is a differential operator; the inversion formula can then be written in the form $f=L((\hat{f})^{\vee})$ with some differential operator L on M [9].

S. Helgason applied the Radon transform on noncompact symmetric Riemannian spaces to solve differential equations on these spaces (1973).

Horospheres and Radon transforms can be defined not only for symmetric Riemannian spaces G/K, but also for various †homogeneous spaces G/H of noncompact semisimple

Lie groups G. The Radon transform $f \to \hat{f}$ maps a function f on G/H into a function on the space of horospheres on G/H. If a †unitary representation U of G is realized in a function space over G/H, the Radon transform $f \to \hat{f}$ helps to clarify the properties of U by going over to the function space on $\mathscr{F}$. In several examples, Gel'fand, Helgason, and others have by this method decomposed U explicitly into direct integrals of irreducible representations [6, 7, 9, 10].

Further work on the Radon transform on †symmetric Riemannian spaces (e.g., compact symmetric Riemannian spaces of rank one and †Grassmann manifolds) has been done by Helgason and others (1965). A generalization of Radon transforms to differential forms has also been given by Gel'fand and others (1969).

H. Another Generalization

Integral geometry can also be investigated in spaces admitting no displacement. Let $F(x_1, x_2, y_1, y_2)$ be positive and homogeneous of degree 1 with respect to y_1, y_2, and consider a †stationary curve of the integral $\int F(x_1, x_2, dx_1/dt, dx_2/dt)dt$. Let $p_i = \partial F/\partial y_i$ $(y_i = dx_i/dt)$ along this curve. Poincaré found that for a 2-parameter set of stationary curves, $dx_1\,dp_1 + dx_2\,dp_2$ is not changed by any displacement of the line element (x_i, p_i) along a stationary curve. Blaschke took this as the kinetic density of the stationary curve and proved a formula containing formula (1) as a special case. On the other hand, Santaló introduced kinetic density for sets of geodesics on 2-dimensional surfaces and proved a generalization of formula (2). The study has been further extended to †foliated manifolds. Some uniqueness theorems for various integral geometric problems with applications to the study of the earth's internal structure from seismological data have been given by V. G. Romanov [11].

References

[1] W. Blaschke, Vorlesungen über Integralgeometrie, Teubner, I, 1936; II, 1937.
[2] S. S. Chern, On integral geometry in Klein spaces, Ann. Math., (2) 43 (1942), 178–189.
[3] S. S. Chern, On the kinematic formula in the Euclidean space of n dimensions, Amer. J. Math., 74 (1952), 227–236.
[4] L. A. Santaló, Introduction to integral geometry, Hermann, 1953.
[5] L. A. Santaló, Integral geometry and geometric probability, Addison-Wesley, 1976.
[6] I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin, Generalized functions V, Integral geometry and representation theory, Academic Press, 1966. (Original in Russian, 1962.)
[7] I. M. Gel'fand and M. I. Graev, Geometry of homogeneous spaces, representations of groups in homogeneous spaces and related questions of integral geometry, Amer. Math. Soc. Transl., (2) 37 (1964), 351–429. (Original in Russian, 1959.)
[8] F. John, Plane waves and spherical means applied to partial differential equations, Interscience, 1955.
[9] S. Helgason, A duality in integral geometry; some generalizations of the Radon transform, Bull. Amer. Math. Soc., 70 (1964), 435–446.
[10] S. Helgason, A duality for symmetric spaces with applications to group representations, Advances in Math., 5 (1970), 1–154.
[11] V. G. Romanov, Integral geometry and inverse problems for hyperbolic equations, Springer, 1974. (Original in Russian, 1969.)
[12] R. Deltheil, Probabilités géométriques, Gauthier-Villars, 1926.
[13] D. G. Kendall and P. A. Moran, Geometrical probability, Griffin, 1963.
[14] M. I. Stoka, Géométrie integrale, Gauthier-Villars, 1968.

219 (XIII.14) Integral Invariants

A. Poincaré's Integral Invariants

Let us view a system of differential equations $dx_i/dt = X_i(x_1, \ldots, x_n, t)$ $(i = 1, 2, \ldots, n)$ as defining the motion of a point whose coordinates are $(x_1, \ldots, x_n)$ in the n-dimensional space $\mathbf{R}^n$ at time t. Let K be a p-dimensional manifold $(1 \leqslant p \leqslant n)$ in $\mathbf{R}^n$, and let K_t be the set occupied at the instant t by the points which occupy K for $t = 0$. If the integral

$$\int_{K_t} F(x_1, \ldots, x_n, t)\,dw, \qquad (1)$$

where dw is the †volume element of K_t, does not depend on t for any p-dimensional surface K, then $\int F(x_1, \ldots, x_n, t)\,dw$ is called an **integral invariant** of degree p of the original system of differential equations. (The †differential form $F\,dw$ is also called an integral invariant.) In particular, a necessary and sufficient condition for an integral $\int M(x, t)\,dx_1 \ldots dx_n$ to be an integral invariant of degree n is $\partial M/\partial t + \sum_{i=1}^{n} \partial(MX_i)/\partial x_i = 0$. Furthermore, a necessary and sufficient condition for an integral $\int \sum_{i=1}^{n} M_i(x, t)\,dx_i$ to be an integral invariant of degree 1 is $\partial M_i/\partial t + \sum_{j=1}^{n} ((\partial M_i/\partial x_j)X_j + (\partial X_i/\partial x_j)M_j) = 0$. If the integral (1) does not

depend on t for any closed p-dimensional surface K, then $\int F(x_1, \dots, x_n, t) dw$ is called a **relative integral invariant** of degree p $(1 \leqslant p \leqslant n-1)$. Corresponding to this terminology, an integral invariant is sometimes called an **absolute integral invariant**.

If a differential form θ is a relative integral invariant of degree p, then its exterior differential $d\theta$ is an absolute integral invariant of degree $p+1$.

For a Hamiltonian system

$$dp_i/dt = -\partial H/\partial q_i, \qquad dq_i/dt = \partial H/\partial p_i \tag{2}$$

$(i=1, \dots, n, H = H(p, q, t), p = (p_1, \dots, p_m), q = (q_1, \dots, q_m))$, the 1-form

$$\omega = \sum_{i=1}^{m} p_i dq_i \tag{3}$$

is a relative integral invariant. The integral of ω on a closed curve $\oint \sum_{i=1}^{m} p_i dq_i$ plays a role in classical quantum theory.

The absolute integral invariant $\Omega = d\omega = \sum_{i=1}^{m} dp_i \wedge dq_i$ has, as a differential 2-form on $\mathbf{R}^{2m}$, the properties:

$$\Omega \text{ is a closed form, that is, } d\Omega = 0. \tag{4}$$

Let ξ be a †tangent vector (at any point). If $\Omega(\xi, \eta) = 0$ for any tangent vector η (at the same point), then $\xi = 0$. (5)

Also, for a Hamiltonian system (2), $\int \dots \int dp_1 \dots dp_m dq_1 \dots dq_m$ is an integral invariant. In other words, a $2m$-dimensional figure in $(p_1, \dots, p_m, q_1, \dots, q_m)$-space may change its form according to the motion of points, but its volume remains unaltered (**Liouville's theorem**). This fact is of importance in applications to †statistical mechanics.

Poincaré developed the theory of integral invariants and applied the theory to the †three-body problem and the problem of †stability [1].

B. Cartan's Extension

Poincaré treated the position $(x_1, \dots, x_n)$ and the time t separately. E. Cartan extended Poincaré's theory by unifying the treatment of position and time.

We may consider the solution curves of a Hamiltonian system (2) through all points of a closed curve C in $(2m+1)$-dimensional space (p, q, t) together with the tube consisting of these solution curves. For any closed curve C_1 lying on and enclosing the tube, we have $\int_C \sum_{i=1}^{m} p_i dq_i - H\,dt = \int_{C_1} \sum_{i=1}^{m} p_i dq_i - H\,dt$, and $\sum_{i=1}^{m} p_i dq_i - H\,dt$ is called **Cartan's relative integral invariant**. If the curve C lies on $t =$ constant, then $\int \omega$ is a relative integral invariant, in Poincaré's terminology.

The exterior differential of this form $d(\omega - H\,dt) = \Omega - dH \wedge dt$ has the property:

Let $(p(t), q(t), t)$ be a solution curve for (2). For the velocity vector $\zeta_1 = (\dot{p}(t), \dot{q}(t), 1)$ and any vector ζ_2 in (p, q, t) space, we have $(\Omega - dH \wedge dt)(\zeta_1, \zeta_2) = 0$. (*)

Therefore the integral of the 2-form over a tube consisting of solution curves equals 0. Applying this to the tube enclosed by C and C_1, we have the former equality.

(*) also implies that a solution of (2) is derived from the †variational problem of finding the extremal of the functional

$$\int_{t_0}^{t_1} \left[\sum_{j=1}^{m} p_j(t) \dot{q}_j(t) - H(p(t), q(t), t) \right] dt$$

for the curves $(p(t), q(t), t)$ $(t_0 \leqslant t \leqslant t_1)$ satisfying $q(t_i) = q^i$ $(i = 0, 1)$ for given t_0, t_1, q^0 and q^1 (without any condition for $p = p(t)$).

If H is given by the †Legendre transformation $H = \sum_{i=1}^{m} p_i \dot{q}_i - (T - U)$, where $T = T(q, \dot{q})$ is a kinetic energy, $U = U(q)$ is a potential, and $p_i = \partial T / \partial \dot{q}_i$, this variational problem considers a wider class of curves than †Hamilton's principle $\delta \int (T - U) dt = 0$ (because there are no restrictions on $p(t)$). However both variational problems give the same extremal [3].

The vector $\xi = (\dot{p}(t), \dot{q}(t))$ is reconstructed by means of Ω and dH as follows. Put $\zeta_1 = (\xi, 1)$ and $\zeta_2 = (\eta, 0)$ (η is a vector in (p, q) space) in (*), then

$$\Omega(\xi, \eta) = -dH \cdot \eta \text{ for any } \eta. \tag{6}$$

Such a ξ is uniquely determined by virtue of (5).

C. Symplectic Structure

Let M be a $2m$-dimensional differentiable manifold. A differential form Ω of degree 2 on M is called a **symplectic structure** if it satisfies (4) and (5). And then (M, Ω) is called a **symplectic manifold**.

For $H: M \to R$, we define the vector field ξ over M by (6) (at each point $x \in M$ and $\eta \in T_x M$). H is called a **Hamiltonian function** (independent of t) and ξ is called a **Hamiltonian vector field**.

In this case, we also have an argument similar to that for the case of the usual Hamiltonian systems in Euclidian space. For example, H is invariant along the flow generated by the vector field ξ, and Ω is an integral invariant.

References

[1] H. Poincaré, Les méthodes nouvelles de la mécanique céleste III, Gauthier-Villars, 1899.

[2] E. Cartan, Leçons sur les invariants intégraux, Hermann, 1922.
[3] V. I. Arnold, Mathematical methods of classical mechanics, Springer, 1978.

220 (X.27) Integral Transforms

A. General Remarks

Given (real- or complex-valued) functions $f(y)$ and $K(x, y)$ such that their product is †integrable as a function of y in the interval $[a, b]$, we set

$$g(x)=\int_a^b K(x,y)f(y)dy. \tag{1}$$

This transformation of f to g is called the **integral transform** with the **kernel** $K(x, y)$. Now fix the kernel $K(x, y)$. When the correspondence $f \to g$ given by (1) from a set of functions f to a set of functions g is bijective, we can consider the **inverse transform** $g \to f$. The formula that describes the inverse transform $g \to f$ in terms of an integral transform is called the **inversion formula**. The kernel $K(x, y)$ can often be written as $k(x-y)$, $k(xy)$, etc., where $k(t)$ is an integrable function. Table 1 contains integral transforms that are important in applications.

Table 1

Kernel	Interval	Name
e^{ixy}	$(-\infty, \infty)$	Fourier transform
$\cos xy$	$(0, \infty)$	Fourier cosine transform
$\sin xy$	$(0, \infty)$	Fourier sine transform
e^{-xy}	$(0, \infty)$	Laplace transform
$\sqrt{xy}\ J_\nu(xy)$	$(0, \infty)$	Hankel transform
$1/(x-y)$	$(-\infty, \infty)$	Hilbert transform
x^{y-1}	$(0, \infty)$	Mellin transform
$(x+y)^{-\rho}$	$(0, \infty)$	Stieltjes transform
$e^{-(x-y)^2}$	$(-\infty, \infty)$	Gauss transform

In the Hankel transform, J_ν is the †Bessel function. In the Hilbert transform, the †principal value is to be taken in the integral. In the Stieltjes transform, ρ is assumed positive.

Since the explanations for the Fourier transform and Laplace transform are given in the corresponding articles, we deal here only with generalized Fourier, Hilbert, Mellin, and Stieltjes transforms.

B. Generalized Fourier Transform

Suppose that the kernels of the transform (1) and its inverse transform are both of the form $k(xy)$. Hence

$$f(x)=\int_a^b k(xy)g(y)dy. \tag{2}$$

In this case, we call the integral transform (1) the **generalized Fourier transform** of symmetric type or the **Watson transform**, and $k(t)$ the **Fourier kernel** of (2). The functions $\sqrt{2/\pi}\cos t$, $\sqrt{2/\pi}\sin t$, and $\sqrt{t}J_\nu(t)$ defined in the interval $(0, \infty)$ are examples of such Fourier kernels (→ 160 Fourier Transform). The last kernel gives rise to the **Hankel transform**.

Let $k(x)$ and $l(x)$ be Fourier kernels. Then the integral transform $m(y)$ of $l(x)$ with respect to the kernel $k(xy)$ is called the resultant of k and l. The resultant of two Fourier kernels is also a Fourier kernel.

Suppose that a function $K(1/2+it)$ satisfies

$$K(1/2+it)K(1/2-it)=1, \quad |K(1/2+it)|=1.$$

Then the function $k(t)=K(1/2+it)/(1/2-it)$ belongs to $L_2(-\infty, \infty)$. We set

$$k_1(x)=\frac{x}{2\pi}\operatorname*{l.i.m.}_{T\to\infty}\int_{-T}^{T} k(t)x^{-1/2-it}dt$$

(the limit is taken in the †mean convergence of order 2). Then for a function $f(x)\in L_2(0, \infty)$,

$$g(x)=\frac{d}{dx}\int_0^\infty k_1(xt)f(t)\frac{dt}{t}$$

exists almost everywhere, and $g(x)\in L_2(0, \infty)$. In this case, we have the **inversion formula**

$$f(t)=\frac{d}{dt}\int_0^\infty k_1(xt)g(x)\frac{dx}{x}$$

and the **Parseval identity**

$$\int_0^\infty (f(t))^2dt=\int_0^\infty (g(x))^2dx.$$

If a function $f(x)$ is invariant under a generalized Fourier transform, then $f(x)$ is called a **self-reciprocal function**. Such a function $f(x)$ is a solution of the homogeneous integral equation

$$f(y)=\int_a^b k(xy)f(x)dx.$$

The function $x^{-1/2}$ is an example of a function that is self-reciprocal with respect to the Fourier cosine transform. Using a function that is self-reciprocal with respect to the Hankel transform, we can derive the **lattice-point formula** of number theory: Let $r(n)$ be the number of possible ways in which a nonnegative integer n can be represented as the sum of

two square numbers. Set

$$\bar{P}(x)=\sum_{0\leqslant n\leqslant x}{}' r(n)-\pi x,$$

where $\sum'$ means that if x is an integer, we take $r(n)/2$ instead of $r(n)$. Then $f(x)=x^{-3/2}\cdot(\bar{P}(x^2/2\pi)-1)$ is self-reciprocal with respect to the Hankel transform with $\nu=2$. Utilizing this, G. H. Hardy proved (1925)

$$\bar{P}(x)=\sqrt{x}\sum_{n=1}^{\infty}\frac{r(n)}{\sqrt{n}}J_1(2\pi\sqrt{nx}).$$

A. Z. Walfisz (1926) and A. Oppenheim (1927) generalized this formula and obtained a formula concerning the number of ways in which n can be represented as the sum of p square numbers (→ 242 Lattice-Point Problems).

C. Mellin Transform

If $f(x)x^{k-1}\in L_1(0,\infty)$, then

$$F(s)=\int_0^{\infty}f(x)x^{s-1}dx,\qquad s=k+it,$$

is called the **Mellin transform** of f. If $f(x)$ is of †bounded variation in a neighborhood of x, then the **inversion formula**

$$\frac{f(x+0)+f(x-0)}{2}=\frac{1}{2\pi i}\lim_{T\to\infty}\int_{k-iT}^{k+iT}F(s)x^{-s}ds$$

holds. If $f(x)x^{k-1/2}\in L_2(0,\infty)$, the integral $\int_{1/A}^{A}f(x)x^{s-1}dx$ $(s=k+it)$ †converges in the mean of order 2 to a function $F(s)$ for a fixed k and $-\infty<t<\infty$, and the **Parseval identity**

$$\int_0^{\infty}|f(x)|^2x^{2k-1}dx=\frac{1}{2\pi}\int_{-\infty}^{\infty}|F(k+it)|^2dt$$

holds. $F(s)$ is also called the **Mellin transform** of $f(x)$. If $f(x)x^{k-1/2}$, $g(x)x^{1/2-k}\in L_2(0,\infty)$, and $F(s)$, $G(s)$ are the Mellin transforms of $f(x)$, $g(x)$, respectively, then

$$\int_0^{\infty}f(x)g(x)dx=\frac{1}{2\pi}\int_{k-i\infty}^{k+i\infty}F(s)G(1-s)ds.$$

The theory of the Mellin transform in the function space L_p is analogous to the theory of the Fourier transform [1, ch. 4].

D. Stieltjes Transform

For a function $\alpha(t)$ of bounded variation,

$$f(s)=\int_0^{\infty}\frac{d\alpha(t)}{(s+t)^{\rho}},\qquad \rho>0,$$

is called the **Stieltjes transform** of $\alpha(t)$. Usually we assume that $\rho=1$. If the Laplace transform is applied twice in succession to $\alpha(t)$, we obtain formally the Stieltjes transform of $\alpha(t)$. The Stieltjes transform has been studied systematically in connection with the theory of the Laplace transform by D. V. Widder, R. P. Boas, and others.

Assume that $\rho=1$. Let D be the domain obtained from the complex plane by removing the negative part of the real axis. If the Stieltjes transform converges at a point $s=s_0\in D$, then it converges uniformly on any compact set in D. The inversion formula is

$$\lim_{\eta\to+0}\frac{1}{2\pi i}\int_0^t(f(-\sigma-i\eta)-f(-\sigma+i\eta))d\sigma$$
$$=(\alpha(t+0)+\alpha(t-0)-(\alpha(+0)+\alpha(-0)))/2,\qquad t>0.$$

If $\alpha(t)=\int_0^t\varphi(u)du$ and $\varphi(t\pm0)$ exist, then

$$\lim_{\eta\to+0}\frac{1}{2\pi i}(f(-t-i\eta)-f(-t+i\eta))$$
$$=(\varphi(t+0)+\varphi(t-0))/2,\qquad t>0.$$

E. Hilbert Transform

Let $\varphi(z)=U(x,y)+iV(x,y)$ $(z=x+iy)$ be holomorphic in the upper half-plane and $f(x)=U(x,0)$, $g(x)=-V(x,0)$ be the respective boundary values on the real axis. Then if f, $g\in L_1(-\infty,\infty)$,

$$g(x)=\frac{1}{\pi}\text{p.v.}\int_{-\infty}^{\infty}\frac{f(x+t)}{t}dt,\tag{3}$$

$$f(x)=-\frac{1}{\pi}\text{p.v.}\int_{-\infty}^{\infty}\frac{g(x+t)}{t}dt.\tag{4}$$

Here p.v. means Cauchy's †principal value, that is,

$$\text{p.v.}\int_{-\infty}^{\infty}F(t)dt$$
$$=\lim_{A\to\infty,\varepsilon\to0}\left(\int_{-A}^{-\varepsilon}F(t)dt+\int_{\varepsilon}^{A}F(t)dt\right).$$

We call g the **Hilbert transform** of f. If $f\in L_2(-\infty,\infty)$, the **inversion formula** and the **Parseval identity** hold. More precisely, for any $f\in L_2(-\infty,\infty)$, relations (3) and (4) above hold almost everywhere, $g\in L_2(-\infty,\infty)$, and the L_2-norms of f and g are identical (→ 160 Fourier Transform). The importance of Hilbert transforms lies in the fact that they establish relations between the real and imaginary parts of an analytic function (→ 132 Elementary Particles C).

References

[1] E. C. Titchmarsh, Introduction to the theory of Fourier integrals, Oxford Univ. Press, 1937.

[2] I. I. Hirschman and D. V. Widder, The convolution transform, Princeton Univ. Press, 1955.
For formulas of integral transforms,
[3] A. Erdélyi (ed.), Tables of integral transformations I, II, McGraw-Hill, 1954.
[4] G. A. Campbell and R. M. Foster, Fourier integrals for practical applications, Bell Telephone Lab., 1931, revised edition, Van Nostrand, 1948.
[5] F. Oberhettinger, Tabellen zur Fourier Transformation, Springer, 1957.

221 (X.13) Integration Theory

A. General Remarks

The class of Riemann integrable functions is not closed under limits and infinite sums. For example, the function equal to 1 for rationals and to 0 for irrationals, the so-called **Dirichlet function**, is not Riemann integrable, though it is expressed by $\lim_{m\to\infty}\lim_{n\to\infty}(\cos(m!\,\pi x))^{2n}$. Hence the concept of Riemann integral is too narrow to be used effectively in modern analysis. To cope with this difficulty, H. Lebesgue [1] introduced a general integral which is now called the Lebesgue integral. It is not only defined for all useful functions appearing in analysis but also has nice properties, such as exchangeability with limits and infinite sums under some simple conditions which can be checked easily. The definition of the Lebesgue integral is based on the concept of Lebesgue measure (→ 270 Measure Theory F), which is a generalization of length, area, or volume. In modern analysis one discusses integrals in an abstract space endowed with a measure and defines the Lebesgue integral as a special case. Integration theory plays a basic role in modern mathematics, in particular in analysis, functional analysis, and probability theory.

Let $f(x)$ be a bounded function defined on a bounded interval $[a,b]$. $f(x)$ is †Riemann integrable if and only if it is continuous †almost everywhere with respect to the Lebesgue measure. If $f(x)$ is Riemann integrable on $[a,b]$, then $f(x)$ is Lebesgue integrable on $[a,b]$ and the integrals coincide with each other. However, the improper Riemann integral is not included in the Lebesgue integral; for example, $(\sin x)/x$ is not Lebesgue integrable on $(0,\infty)$, but $\lim_{a\to\infty}\int_0^a(\sin x)/x\,dx=\pi/2$. The theory of Denjoy integral gives insight into this situation (→ 100 Denjoy Integrals).

B. Definition of Integrals

Let X be an arbitrary set. If a †completely additive class $\mathfrak{B}$ of subsets of X and a †measure μ defined on $\mathfrak{B}$ are given, then we say that a †measure space $(X,\mathfrak{B},\mu)$ is defined. For example, the Euclidean space $\mathbf{R}^n$, the class $\mathfrak{M}_n$ of all †Lebesgue measurable sets in $\mathbf{R}^n$, and the †Lebesgue measure m_n on $\mathfrak{M}_n$ form the measure space $(\mathbf{R}^n,\mathfrak{M}_n,m_n)$ (→ 270 Measure Theory). We consider only $\mathfrak{B}$-measurable sets and $\mathfrak{B}$-measurable functions in this article, so a $\mathfrak{B}$-measurable set will be called simply a set and a $\mathfrak{B}$-measurable function simply a function. We admit $\pm\infty$ for the values of a function.

The **integral** $\int_E f(x)\,d\mu(x)$ on a set $E\subset X$ of a real-valued function $f(x)$ (we write simply $\int_E f\,d\mu$ or $\int_E f$) can be defined in steps as follows. (1) Let $f(x)\geqslant 0$ be a **simple function**, that is, a function whose range is a finite set $\{a_i\}$ $(i=1,2,\ldots,n)$. If $f(x)=a_i$ for $x\in E_i$, where $E=\bigcup_{j=1}^n E_j$, $E_j\cap E_k=\varnothing$ $(j\neq k)$, then we put $\int_E f=\sum_{a_j\neq 0}a_j\mu(E_j)$. (The value of the integral is a real number or $+\infty$. For operations concerning $\pm\infty$, → 270 Measure Theory D.) (2) For an arbitrary $f(x)\geqslant 0$, we define $\int_E f$ as the †supremum of $\int_E g$, where the supremum is taken for all simple functions g such that $0\leqslant g\leqslant f$. (3) In general, letting $f(x)=f^+(x)-f^-(x)$, where $f^+(x)=\max\{f(x),0\}$, $f^-(x)=\max\{-f(x),0\}$, we define $\int_E f=\int_E f^+-\int_E f^-$ if at least one of $\int_E f^+$ and $\int_E f^-$ is finite and say that f **has an integral** (or a **μ-integral**) **on** E. In particular, if $\int_E f$ is finite, then we say f is **integrable** (or **μ-integrable**) on E.

If the given measure space is $(\mathbf{R}^n,\mathfrak{M}_n,m_n)$, the integral defined in this section is called the **Lebesgue integral** (or simply **L-integral**), and the function that is integrable in this case is said to be **Lebesgue integrable**. The integral is often written as $\int_E f(x)\,dx$, and if E is the interval $[a,b]$, as $\int_a^b f(x)\,dx$.

C. Properties of Integrals

(1) The set of all functions integrable on E forms a †linear space over $\mathbf{R}$, that is, if f and g are integrable on E, then for any real α, β, $\alpha f+\beta g$ is also, and

$$\int_E(\alpha f+\beta g)\,d\mu=\alpha\int_E f\,d\mu+\beta\int_E g\,d\mu.$$

(2) The integrability of f, of both f^+ and f^-, and of $|f|$ are mutually equivalent. (3) If $g\leqslant f$ on E, then $\int_E g\leqslant\int_E f$. In particular, if $m\leqslant f(x)\leqslant l$ on E, then $m\mu(E)\leqslant\int_E f\,d\mu\leqslant l\mu(E)$. (4) If $f(x)$ is integrable on E (has an integral on E), then it is integrable (has an integral) on

every subset of E. (5) If $E=\bigcup_{n=1}^{\infty} E_n$, $E_j \cap E_k = \varnothing$ $(j \neq k)$, and $\int_E f d\mu$ exists, then the series $\sum_{n=1}^{\infty} \int_{E_n} f d\mu$ converges and is equal to $\int_E f d\mu$ **(complete additivity of the integral)**. (6) If $f(x) \geqslant 0$ and $\int_E f d\mu = 0$, then $f(x)=0$ †almost everywhere on E. (7) Modification of the values of $f(x)$ on a †null set influences neither the integrability nor the value of the integral. Consequently, if the function f is not defined on a null set, we can define the value of f on this set arbitrarily so that the integral has meaning. (8) If $f(x)$ is integrable on E, $E \supset E_n$, and $\mu(E_n) \to 0$, then $\lim_{n\to\infty} \int_{E_n} f d\mu = 0$. (9) If $f_n(x) \leqslant f_{n+1}(x)$ on E, and there exists a $\varphi(x)$ such that $f_n(x) \geqslant \varphi(x)$ and $\int_E \varphi > -\infty$ (for example, if $f_n(x) \geqslant 0$), then

$$\lim_{n\to\infty} \int_E f_n = \int_E \left(\lim_{n\to\infty} f_n \right).$$

(10) If $\lim_{n\to\infty} f_n(x)$ exists almost everywhere on E, and there exists a $\varphi(x)$ such that $|f_n(x)| \leqslant \varphi(x)$ and $\int_E \varphi < +\infty$ (for example, if $\mu(E) < \infty$ and $|f_n(x)| < M$), then

$$\lim_{n\to\infty} \int_E f_n = \int_E \left(\lim_{n\to\infty} f_n \right)$$

(Lebesgue's convergence theorem). (11) If there exists a $\varphi(x)$ such that $f_n(x) \geqslant \varphi(x)$ on E and $\int_E \varphi > -\infty$, then

$$\liminf_{n\to\infty} \int_E f_n \geqslant \int_E \left(\liminf_{n\to\infty} f_n \right)$$

(Fatou's theorem). (12) If $f_n(x) \geqslant 0$ on E, then $\int_E \sum_{n=1}^{\infty} f_n = \sum_{n=1}^{\infty} \int_E f_n$. If $f_n(x)$ is real on E, then this equality holds for $|f_n|$ in place of f_n. If the common value is finite, then the equality holds also for the original f_n. (13) Let φ be a mapping from $(X, \mathfrak{B}, \mu)$ to $(X', \mathfrak{B}', \mu')$ such that $B' \in \mathfrak{B}'$ implies $\varphi^{-1}(B') \in \mathfrak{B}$ and $\mu'(B') = \mu(\varphi^{-1}(B'))$. If $f(x')$ is $\mathfrak{B}'$-measurable on $E' \in \mathfrak{B}'$, then $\int_{E'} f(x') d\mu'(x') = \int_{\varphi^{-1}(E)} f(\varphi(x)) d\mu(x)$, where the equality means that if one of these integrals exists, then so does the other, and they have the same value.

D. Indefinite Integrals

If we put $F(e) = \int_e f$ for each measurable subset e of E, with $f(x)$ integrable on E, $F(e)$ is a †μ-absolutely continuous completely additive †set function (properties of integrals (4), (5), (8)). We call $F(e)$ the **indefinite integral** of $f(x)$. In the special case where X is the set $\mathbf{R}$ of real numbers and $f(x)$ is integrable on the interval $[a, b]$, the function $F(x) = \int_a^x f(t) dt$ defined for $x \in [a, b]$ is also called the indefinite integral of $f(x)$. The $F(x)$ so defined is an †absolutely continuous function. Conversely, if a function $F(x)$ is absolutely continuous on $[a, b]$, then it is differentiable almost everywhere on $[a, b]$, and we have $F(x) - F(a) = \int_a^x F'(t) dt$. (For the relationship between differentiation and integration in the case of $\mathbf{R}^n$, or more generally in the case of an arbitrary measure space, → 380 Set Functions.)

E. Fubini's Theorem

Let $(X, \mathfrak{B}_1, \mu_1)$ and $(Y, \mathfrak{B}_2, \mu_2)$ be two σ-finite measure spaces and $(X \times Y, \mathfrak{B}, \mu)$ be their †direct product measure space. Assume that $f(x, y)$ is $\mathfrak{B}$-measurable and integrable (has an integral) on $X \times Y$. Then for almost all fixed $y \in Y$, $f(x, y)$ considered as a function of x is $\mathfrak{B}_1$-measurable and integrable (has an integral), and $\int_X f(x, y) d\mu_1(x)$ is a $\mathfrak{B}_2$-measurable function of y. Moreover, in this case we have

$$\int_{X \times Y} f(x, y) d\mu(x, y)$$

$$= \int_Y \left(\int_X f(x, y) d\mu_1(x) \right) d\mu_2(y)$$

(Fubini's theorem). This fact also holds with x and y exchanged. The integral on the left-hand side of this equation is called a **multiple integral**, while that on the right-hand side is called an **iterated** (or **repeated**) **integral**.

Even if an iterated integral exists, the corresponding multiple integral need not always exist. For example, let $f(x, y)$ be defined as $(x^2 - y^2)/(x^2 + y^2)^2$ on $(0, 1) \times (0, 1)$, and otherwise 0. Then $\int_{\mathbf{R}^2} f^+ = \int_{\mathbf{R}^2} f^- = \infty$, so that $\int_{\mathbf{R}^2} f$ does not exist, but

$$\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(x, y) dx \right) dy = -\frac{\pi}{4},$$

$$\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(x, y) dy \right) dx = \frac{\pi}{4}.$$

By Fubini's theorem, if f is a nonnegative function defined on a $\mathfrak{B}$-measurable subset of a σ-finite measure space $(X, \mathfrak{B}, \mu)$, the $\mathfrak{B}$-measurability of f is equivalent to the measurability of the **ordinate set** $E_f = \{(x, y) \mid 0 \leqslant y \leqslant f(x), x \in E\}$ considered as a subset of the measure space $(X \times \mathbf{R}, \mathfrak{B}', \mu')$, which is the direct product of $(X, \mathfrak{B}, \mu)$ and $(\mathbf{R}, \mathfrak{M}_1, m_1)$. In this case we have $\int_E f d\mu = \mu'(E_f)$, which can serve as a definition of the Lebesgue integral. When $(X, \mathfrak{B}, \mu)$ coincides with $(\mathbf{R}, \mathfrak{M}_1, m_1)$, the †Jordan measure (area) of the ordinate set coincides with the Riemann integral.

References

[1] H. Lebesgue, Intégral, longueur, aire, Ann. Mat. Pura Appl., (3) 7 (1902), 231–359.

[2] S. Saks, Theory of the integral, Stechert, second edition, 1937. (Original in French, 1933, 1937.)
[3] N. Bourbaki, Éléments de mathématique, Integration, chs. 1–9, Actualités Sci. Ind., Hermann, 1965–1969.
[4] H. L. Royden, Real analysis, Macmillan, second edition, 1963.
[5] W. Rudin, Real and complex analysis, McGraw-Hill, second edition, 1974.

222 (XIII.31) Integrodifferential Equations

A. General Remarks

A †functional equation involving an operator T of the form

$$f(t, x'(t), x(t), (Tx)(t)) = 0, \quad t_0 \leqslant t \leqslant t_1, \tag{1}$$

$$x(t_0) = x_0, \tag{2}$$

reduces to a †differential equation, †differential-difference equation, †integral equation, integrodifferential equation, or other functional equation when the operator T is given a special form [1, 6]. In particular, by letting

$$f(t, x, y, z) = x - g(t, y) - z, \tag{3}$$

$$(Tx)(t) = \int_{t_0}^{\varphi(t)} K(t, s, x(s))\,ds, \tag{4}$$

we obtain an **integrodifferential equation**

$$x'(t) = g(t, x(t)) + \int_{t_0}^{\varphi(t)} K(t, s, x(s))\,ds. \tag{5}$$

If $\varphi(t) \equiv$ constant or $\varphi(t) \equiv t$, (5) is said to be an **integrodifferential equation of Fredholm type** or of **Volterra type**, respectively.

B. The Initial Value Problem

Let I be an interval $t_0 \leqslant t \leqslant t_1$, I_0 the interval $t_0 < t \leqslant t_1$, $C(I)$ a set of continuous functions on I, T an operator such that $Tx \in C(I_0)$ if $x \in C(I)$, $\mathfrak{F}$ a family of all such T, $\mathfrak{F}_+$ a subset of $\mathfrak{F}$ consisting of all T in $\mathfrak{F}$ for which $Tx \leqslant Ty$ holds at $t = s$ whenever x, y are functions in $C(I)$ satisfying $x(t) < y(t)$ for $t_0 \leqslant t < s$ for some $s \in I_0$, and Z a set of continuous functions which are differentiable on I_0. Let M be the class of all functions $f(t, x, y, z)$ defined for $t \in I_0$, $|x|$, $|y|$, $|z| < \infty$ and satisfying $f(t, x_1, y, z_1) \geqslant f(t, x_2, y, z_2)$ $(x_1 \geqslant x_2, z_1 \leqslant z_2)$.

Suppose that in (1) $f(t, x, y, z)$ is defined for $t \in I_0$, $|x|$, $|y|$, $|z| < \infty$, and $T \in \mathfrak{F}$. Suppose further that for some $\gamma > 0$ and two solutions x_1 and x_2 of (1) with (2) belonging to Z, there always exist a function $\omega \in M$ and an operator $\Omega \in \mathfrak{F}_+$ such that the inequality

$$\omega(\bar{t}, x_2' - x_1', x_2 - x_1, \Omega(x_2 - x_1)) \leqslant 0$$

holds for every $\bar{t} \in I_0$ such that $x_2(\bar{t}) - x_1(\bar{t}) = \gamma$ and $x_2(t) - x_1(t) < \gamma$ for $t_0 < t < \bar{t}$. Furthermore, suppose that there exists a function $\rho \in Z$ satisfying the following inequalities:

$$0 \leqslant \rho \leqslant \gamma, \quad t \in I_0;$$

$$\omega(t, \rho', \rho, \Omega\rho) > 0, \quad t \in I_0;$$

$$\rho(t_0 + 0) > x_2(t_0 + 0) - x_1(t_0 + 0).$$

Then equation (1) with (2) has at most one solution $x \in Z$. This result can be established by obtaining an estimate for the difference $|x_1(t) - x_2(t)|$, where $x_i(t)$ is a solution of (1) with the condition $x(t_0) = \eta_i$ $(i = 1, 2)$.

For the particular case of integrodifferential equations of Fredholm type, suppose that the following conditions are satisfied:

$$g(t, y_1) - g(t, y_2) \leqslant L(t)(y_1 - y_2), \quad t \in I_0;$$

$$\int_{t_0}^{t} (K(t, s, w_1(s)) - K(t, s, w_2(s)))\,ds \leqslant N(t) \int_{t_0}^{t} M(s)(w_1(s) - w_2(s))\,ds, \quad t \in I;$$

$$\int_{t_0}^{t} sM(s)\,ds < \infty, \quad t \in I_0;$$

$$N(t) + tL(t) \leqslant 1 + t^2 M(t), \quad t \in I_0;$$

where $y_1 \geqslant y_2$, $L \in C(I_0)$, w_1, w_2, M, $N \in C(I)$, $w_1 \geqslant w_2$, $M \geqslant 0$, and $N \geqslant 0$. Then it is possible to obtain more practical expressions for ω, Ω, and ρ:

$$\omega(t, x, y, z) = x - L(t)y - N(t)z, \quad t \in I_0,$$

$$\Omega\omega = \int_{t_0}^{t} M(s)w(s)\,ds,$$

$$\rho(t) = \beta t(1 + t)\exp \int_{t_0}^{t} s(1 + s)M(s)\,ds,$$

where $\beta > 0$ is sufficiently small.

C. Another Problem

A problem analogous to the †boundary value or †eigenvalue problems of linear ordinary differential equations is to find a solution of the linear integrodifferential equation

$$(pu')' - qu + \lambda\left(\rho u + \int_G k(x, y)u(y)\,dy\right) = 0$$

with the boundary value $u = 0$. This equation can be derived from the problem of minimizing the functional $D[\varphi]$ under the condition

that $H[\varphi]$ is constant, where

$$D[\varphi]=\int_G p\varphi'^2\,dx+\int_G q\varphi^2\,dx,$$

$$H[\varphi]=\int_G \rho\varphi^2\,dx$$

$$+\int_G\int_G k(x,y)\varphi(x)\varphi(y)\,dx\,dy.$$

The orthogonality condition for this boundary value problem is given by

$$\int_G \rho u_i(x)u_j(x)\,dx$$

$$+\int_G\int_G k(x,y)u_i(x)u_j(y)\,dx\,dy=\begin{cases}1, & i=j\\ 0, & i\neq j\end{cases}$$

[2].

Integrodifferential equations are closely related to problems of mathematical physics and engineering, and there are many investigations of such equations in the study of the equilibrium of rotating fluids [3], Prandtl's integrodifferential equation for aircraft wings in 3-dimensional space [4], the dynamics of reactors, and so on. In the second example, the circulation of the airflow $\Gamma(y)$ with constant velocity V around the profile is determined by the equation

$$\alpha(y)=\frac{\Gamma(y)}{\pi k(y)t(y)V}+\frac{1}{\pi V}\,\mathrm{p.v.}\int_{-b/2}^{b/2}\frac{d\Gamma(y')}{y-y'},$$

called **Prandtl's integrodifferential equation**, where b is the wingspan, y and y' are variables whose range is $[-b/2, b/2]$ (y is assumed fixed), t the length of the chord, α an angle of incidence from the point with buoyancy 0, $2\pi k$ the slope of the curve defined from buoyancy by the angle of incidence, and p.v. †Cauchy's principal value.

As a problem having applications in the theory of †stochastic processes, the existence of solutions that have finite limits as $t\to\infty$ has been investigated for the **Wiener-Hopf integrodifferential equation**

$$\prod_{k=1}^{n}(D+\lambda_k)f(x)=\lambda_1\ldots\lambda_n\int_0^\infty f(x+t)\,dH(t),$$

$$x\geqslant 0,\qquad D=d/dx.\quad (6)$$

For this problem, by means of the method of †semigroups of operators, equation (6) can be extended to

$$\prod_{k=1}^{n}(A+\lambda_k)f(x)=\lambda_1\ldots\lambda_n\int_0^\infty (T_tf)(x)\,dH(t),$$

$$x\in I,$$

where A is the †infinitesimal generator of a semigroup of operators $\{T_t\}$, and it has been shown that analogous results to those for (6) can be obtained for this more general equation [5].

References

[1] K. Nickel, Fehlerabschätzungs- und Eindeutigkeitssätze für Integro-Differentialgleichungen, Arch. Rational Mech. Anal., 8 (1961), 159–180.
[2] R. Courant and D. Hilbert, Methods of mathematical physics I, Interscience, 1953.
[3] L. Lichtenstein, Vorlesungen über einige Klassen nichtlinearer Integralgleichungen und Integro-Differentialgleichungen, Springer, 1931.
[4] W. F. Durand, Aerodynamic theory II, Springer, 1935.
[5] S. Karlin and G. Szegö, On certain differential-integral equations, Math. Z., 72 (1959–1960), 205–228.
[6] M. N. Oğuztöreli, Time-lag control systems, Academic Press, 1966.
[7] N. I. Muskhelishvili, Singular integral equations, Noordhoff, 1953. (Original in Russian, 1946.)
[8] T. L. Saaty, Modern nonlinear equations, McGraw-Hill, 1967.

223 (XV.2) Interpolation

A. Lagrange Interpolation

Assume that the values of a function $f(x)$ with some regularity property (e.g., differentiability up to a certain order) are given at each of $n+1$ distinct values $x_0, x_1, \ldots, x_n$. The method of finding the values $f(x)$ at $x\neq x_i$ by using the values $f_0=f(x_0), f_1=f(x_1), \ldots, f_n=f(x_n)$ is called **interpolation**. A function $L(x)$ that approximates $f(x)$ and coincides with $f(x)$ at $x=x_0, x_1, \ldots, x_n$ is called an **interpolation function** or **interpolation formula**.

We usually use a polynomial of degree n as $L(x)$. Such a polynomial is called a **Lagrange interpolation polynomial**, and the method of using such a polynomial is called **Lagrange interpolation**. If we let

$$l_i(x)=\frac{\Pi(x)}{(x-x_i)\Pi'(x_i)},\qquad \Pi(x)=\prod_{i=0}^{n}(x-x_i),$$

the Lagrange interpolation polynomial can then be expressed in the form

$$L(x)=\sum_{i=0}^{n} l_i(x)f_i.$$

$l_i(x)$ is a polynomial of degree n and is called the **Lagrange interpolation coefficient**. Sometimes, by interpolation we mean the method of finding $f(x)$ for x lying between the maximum and the minimum of x_i. When x is located elsewhere, such a method is called **extrapolation**. The Lagrange interpolation polynomial is uniquely determined, and if $f(x)$ is $(n+1)$-times differentiable, the deviation of $L(x)$ from $f(x)$ is given by $f^{(n+1)}(\xi)\Pi(x)/(n+1)!$, where ξ lies between the maximum and the minimum of $x, x_0, x_1, \ldots, x_n$. Conversely, the method of finding an approximate value of x satisfying $f(x)=f$ for a given value of f using $L(x)$ is called **inverse interpolation**.

Suppose that we restrict the interval of interpolation to $[-1,1]$ and require that the points of interpolation x_i be spaced equally, $x_i=2i/n-1$, $i=0,1,\ldots,n$. Then, if the number of points $n+1$ is increased, the Lagrange interpolation polynomial $L(x)$ may not converge to $f(x)$ even if $f(x)$ is an analytic function of x on $[-1,1]$. For example, this nonconvergence phenomenon is observed when $f(x)=1/(25+x^2)$ is interpolated; this is called the **Runge phenomenon**. On the other hand, if we choose the zeros of the †Chebyshev polynomial of degree $n+1$ as x_i, i.e., if we take $x_i=\cos\{(2i+1)\pi/2(n+1)\}$, $i=0,1,\ldots,n$, which is called **Chebyshev interpolation**, then $L(x)$ converges uniformly to $f(x)$ as $n+1$ tends to infinity, provided that $f(x)$ has a bounded derivative on $[-1,1]$.

B. Iterated Interpolation

We can compute the value of the interpolation polynomial for given x by generating a sequence of interpolants each of which involves one more point than the previous one. The following method is called the **Aitken interpolation scheme**. First we compute

$$I_{0i}=\frac{1}{x_i-x_0}\begin{vmatrix} f_0 & x_0-x \\ \vdots & \\ f_i & x_i-x \end{vmatrix}, \qquad i=1,2,\ldots,n,$$

and then successively

$$I_{01\ldots ki}=\frac{1}{x_i-x_k}\begin{vmatrix} I_{01\ldots k} & x_k-x \\ \vdots & \\ I_{01\ldots k-1\,i} & x_i-x \end{vmatrix},$$

$$i=k+1,\ldots,n.$$

If we continue successive evaluation of I_{01}, $I_{012}, \ldots$ until the successive values coincide within the desired accuracy, then we can accept the converged value $I_{01\ldots n}$ as an approximate value of the Lagrange interpolation polynomial $L(x)$ of degree n at x. In this case the x_i are not necessarily assumed to be arranged in monotonic order. It is better to arrange them in order of their distance from x rather than in ascending or descending order of magnitude.

C. Interpolation by Finite Differences

When the interpolation points are equally spaced, the values of the interpolation polynomial can be expressed in terms of **finite differences**. Suppose that for $x_i=x_0+ih$, where h is the grid size, the values $f_0, f_1, \ldots, f_n$ of the function $f(x)$ are known. The differences $\Delta f_i=f_{i+1}-f_i$ are then called (finite) differences of first order. Furthermore, we define the differences of order $k+1$ inductively by $\Delta^{k+1}f_i=\Delta^k f_{i+1}-\Delta^k f_i$. If we use the **shift operator** E defined by $Ef_i=f_{i+1}$, the difference operator is represented as $\Delta=E-1$. Sometimes the **backward difference** $\nabla=1-E^{-1}$ is used. In contrast to ∇, the operator Δ is called the **forward difference**. The **central difference** $\delta=E^{1/2}-E^{-1/2}$ which is defined by $\delta f_{i-1/2}=f_i-f_{i-1}$ is also used. Table 1 shows the relations between the finite differences and the **differentiation operator** D defined by $Df(x)=df(x)/dx$.

Table 2, in which each entry after the second column is the difference of the two entries lying immediately to its left is called the **difference table**. From the relation $\Delta=\nabla E$ or $\Delta=\delta E^{1/2}$ we can express each entry of table 2 in terms of ∇ or δ. For example, in the second

Table 1

	E	Δ	δ	∇	hD
E	E	$1+\Delta$	$1+\frac{\delta^2}{2}+\delta\mu$	$\frac{1}{1-\nabla}$	e^{hD}
Δ	$E-1$	Δ	$\delta\mu+\frac{\delta^2}{2}$	$\frac{\nabla}{1-\nabla}$	$e^{hD}-1$
δ	$E^{1/2}-E^{-1/2}$	$\frac{\Delta}{(1+\Delta)^{1/2}}$	δ	$\frac{\nabla}{(1-\nabla)^{1/2}}$	$2\sinh(hD/2)$
∇	$1-E^{-1}$	$\frac{\Delta}{1+\Delta}$	$\delta\mu-\frac{\delta^2}{2}$	∇	$1-\frac{1}{e^{hD}}$
hD	$\log E$	$\log(1+\Delta)$	$2\,\mathrm{arc\,sinh}(\delta/2)$	$-\log(1-\nabla)$	hD
μ	$\frac{E^{1/2}+E^{-1/2}}{2}$	$\frac{1+\Delta/2}{(1+\Delta)^{1/2}}$	$(1+\delta^2/4)^{1/2}$	$\frac{1-\nabla/2}{(1-\nabla)^{1/2}}$	$\cosh(hD/2)$

Table 2. Difference Table

$\vdots$				
f_{-2}				
	Δf_{-2}			
f_{-1}		$\Delta^2 f_{-2}$		
	Δf_{-1}		$\Delta^3 f_{-2}$	
f_0		$\Delta^2 f_{-1}$		$\Delta^4 f_{-2}$
	Δf_0		$\Delta^3 f_{-1}$	
f_1		$\Delta^2 f_0$		
	Δf_1			
f_2				
$\vdots$				

column, Δf_k is equal to ∇f_{k+1} and $\delta f_{k+1/2}$. If $f(x)$ is a polynomial of degree k, then $\Delta f(x)$ is a polynomial of degree $k-1$, $\Delta^k f(x)$ is a constant, and $\Delta^{k+1} f(x)$ is zero. Therefore, looking at the difference table, we can find the degree of an interpolation polynomial that can satisfactorily approximate $f(x)$. It should also be noted that, if the computation of each entry of the table is carried out with a finite number of significant figures, the error in the values $f_0, f_1, \ldots, f_n$ is multiplied by binomial coefficients corresponding to the location in the difference table.

The following are interpolation formulas for which the difference table is used. Suppose that we want to interpolate the value f_p at $x=x_0+ph$. Then from $f_p=E^p f_0=(1+\Delta)^p f_0$, we have **Newton's forward interpolation formula**:

$$f_p=f_0+p\Delta f_0+\frac{p(p-1)}{2!}\Delta^2 f_0$$
$$+\frac{p(p-1)(p-2)}{3!}\Delta^3 f_0+\ldots.$$

Similarly, from $f_p=E^p f_0=(1-\nabla)^{-p} f_0$, we have **Newton's backward interpolation formula**:

$$f_p=f_0+p\nabla f_0+\frac{p(p+1)}{2!}\nabla^2 f_0$$
$$+\frac{p(p+1)(p+2)}{3!}\nabla^3 f_0+\ldots.$$

We get these formulas by starting at f_0 and proceeding downward to the right or upward to the right. On the other hand, by proceeding in a zigzag manner, downward to the right, then upward to the right, then again downward to the right, etc., we get another formula, called **Gauss's forward interpolation formula**:

$$f_p=f_0+p\Delta f_0+\frac{p(p-1)}{2!}\Delta^2 f_{-1}$$
$$+\frac{(p+1)p(p-1)}{3!}\Delta^3 f_{-1}+\ldots.$$

Similarly, we have **Gauss's backward interpolation formula**:

$$f_p=f_0+p\Delta f_{-1}+\frac{(p+1)p}{2!}\Delta^2 f_{-1}$$
$$+\frac{(p+1)p(p-1)}{3!}\Delta^3 f_{-2}+\ldots.$$

In addition to these formulas, there are several by Everett, Bessel, Stirling, and others, which are essentially equivalent to the Lagrange interpolation polynomial that uses the same tabular points, although the representations are different.

D. Interpolation by Divided Differences

For points located at unequal intervals, the values of the interpolation polynomials can also be expressed in terms of **divided differences**, defined successively as

$$f_{ij}=\frac{f_i-f_j}{x_i-x_j}, \qquad f_{ijk}=\frac{f_{ij}-f_{jk}}{x_i-x_k}, \ldots.$$

The divided difference of order k defined as above can be expressed as

$$f_{01\ldots k}=\sum_{i=0}^{k}\frac{f_i}{\Pi'(x_i)}, \qquad \Pi(x)=\prod_{i=0}^{k}(x-x_i).$$

The Lagrange interpolation polynomial of degree n is

$$L(x)=f_0+(x-x_0)f_{01}+(x-x_0)(x-x_1)f_{012}$$
$$+\ldots+(x-x_0)(x-x_1)\ldots(x-x_{n-1})f_{01\ldots n},$$

with the error given by

$$f(x)-L(x)=(x-x_0)(x-x_1)\ldots(x-x_n)f_{012\ldots nx},$$

where $f_{012\ldots nx}$ means that the divided difference of first order is calculated from x and $f(x)$ instead of from x_j and f_j.

E. Hermite Interpolation

The polynomial $H(x)$ of degree $2n+1$ that satisfies not only $H(x)=f(x)$ but also $H'(x)=f'(x)$ at $x=x_0, x_1, \ldots, x_n$ is called the **Hermite interpolation polynomial**. The Hermite interpolation polynomial is

$$H(x)=\sum_{i=0}^{n}\{1-2l_i'(x_i)(x-x_i)\}l_i^2(x)f(x_i)$$
$$+\sum_{i=0}^{n}(x-x_i)l_i^2(x)f'(x_i),$$

where $l_i(x)$ is the Lagrange interpolation coefficient. The deviation of $H(x)$ from $f(x)$ is given by $f(x)-H(x)=f^{(2n)}(\xi)\Pi^2(x)/(2n)!$, where ξ lies between the maximum and the minimum of $x, x_0, x_1, \ldots, x_n$.

F. Spline Interpolation

Let the interval $(-\infty, \infty)$ be divided into $n+1$ subintervals not necessarily of equal length by n knots $x_0=-\infty<x_1<x_2<\ldots<x_n<x_{n+1}=\infty$. A function $S(x)$ which coincides with a polynomial of degree at most m on each subinterval (x_i, x_{i+1}), $i=1,2,\ldots,n$, and has continuous derivatives up to order $m-1$ at each knot is called a **spline**, i.e., a spline $S(x)$ is a C^{m-1} function. If $S(x_i)=f(x_i)$, $i=0,1,2,\ldots,n$, then $S(x)$ is an interpolation function. Interpolation by means of a spline is called **spline interpolation**. The term "spline" derives from the name of an instrument with which draftsmen fair a curve through points.

If a spline coincides with a polynomial of degree $2k-1$ on each subinterval and with a polynomial of degree $k-1$ on $(-\infty, x_1]$ and $[x_n, \infty)$, it is called a **natural spline**. If the data y_i is given at each knot x_i, $i=1,2,\ldots,n$, and also an integer k not larger than n is given, then the spline $S(x)$ of degree $2k-1$ that satisfies $S(x_i)=y_i$, $i=1,2,\ldots,n$, is uniquely determined. The spline that is most frequently used in practical problems is the natural cubic spline $(k=2)$.

Let $f(x)$ be an arbitrary function of C^k $(k\leqslant n)$ satisfying $f(x_i)=y_i$ at each knot x_i, $i=1,2,\ldots,n$. Then in any interval $[a,b]$ which includes $x_1, x_2, \ldots, x_n$,

$$\int_a^b [S^{(k)}(x)]^2\,dx \leqslant \int_a^b [f^{(k)}(x)]^2\,dx$$

holds, where $S(x)$ is the natural spline of degree $2k-1$ that satisfies $S(x_i)=y_i$, $i=1,2,\ldots,n$. This inequality is called the **minimum norm property**, and, in particular, the **minimum curvature property** when $k=2$, of the natural spline.

G. Polynomial Interpolation on Triangles

Polynomial interpolation on a triangular region is used in the **finite element method**. The complete polynomial of degree m,

$$P_m(x,y)=\sum_{\mu+\nu=0}^{m} a_{\mu\nu}x^\mu y^\nu,$$

can be used as an interpolation polynomial on a triangular region. The number of the coefficients $a_{\mu\nu}$ is $(m+1)(m+2)/2$. On the other hand, if we divide each side of the triangle into m equal parts and join the points of subdivision by lines parallel to the sides of the triangle, then we have m^2 congruent small triangles the number of whose vertices is $(m+1)(m+2)/2$. Therefore if we choose these vertices as the points of interpolation, we can determine the coefficients $a_{\mu\nu}$ uniquely from the data given at these points. This is the Lagrange interpolation polynomial of degree m on a triangle.

The interpolation polynomial $P_1(x,y)$ of degree $m=1$ that is uniquely determined from the data u_1, u_2, u_3 given at the vertices (x_1,y_1), (x_2,y_2), (x_3,y_3), respectively, can be expressed as

$$P_1(x,y)=u_1\xi_1(x,y)+u_2\xi_2(x,y)+u_3\xi_3(x,y),$$

where

$$\xi_i(x,y)=\frac{1}{2S}\begin{vmatrix}1 & 1 & 1\\ x & x_j & x_k\\ y & y_j & y_k\end{vmatrix}, \qquad S=\frac{1}{2}\begin{vmatrix}1 & 1 & 1\\ x_1 & x_2 & x_3\\ y_1 & y_2 & y_3\end{vmatrix}.$$

(i,j,k) is any cyclic permutation of $(1,2,3)$, and the absolute value of S is equal to the area of the triangle. $\xi_i(x,y)$ is a polynomial of degree 1 satisfying

$$\xi_i(x_j,y_j)=\begin{cases}1, & j=i,\\ 0, & j\neq i,\end{cases}$$

and is called the **shape function** of degree 1 on the triangle.

If we choose the vertices together with the midpoints of the sides as the points of interpolation, we can determine the six coefficients $a_{\mu\nu}$ of the interpolation polynomial $P_2(x,y)$ of degree 2 uniquely, and the polynomial can be written as

$$P_2(x,y)=\sum_{i=1}^{6} u_i\xi_i^{(2)}(x,y),$$

where $\xi_i^{(2)}$, $i=1,2,\ldots,6$, are the shape functions of degree 2 on the triangle; these can be expressed in terms of the shape functions of degree 1 as follows:

$$\xi_i^{(2)}(x,y)=\xi_i(2\xi_i-1), \qquad i=1,2,3,$$

$$\xi_i^{(2)}(x,y)=4\xi_1\xi_2, \quad \xi_5^{(2)}=4\xi_2\xi_3, \quad \xi_6^{(2)}=4\xi_3\xi_1.$$

The Lagrange interpolation polynomial $P_m(x,y)$ given above is defined locally on each triangle in the finite element method, and the function over the whole triangular network is constructed by connecting these Lagrange interpolation polynomials. This piecewise polynomial function of degree m over the whole network is evidently of class C^0, and has no continuity of derivatives of higher order in general. A C^1-function, for example, can be obtained from the complete polynomial $P_5(x,y)$ by determining the 21 coefficients from the values and the derivatives up to order 2 at the vertices and the normal derivatives at the midpoints of the sides. Alternatively, we can impose three conditions so that the normal derivative reduces to a cubic along each side instead of imposing that it coincide with specified data at the midpoint; then we obtain another C^1-function. A variety of interpolation functions are known for

other elementary regions, such as rectangles or tetrahedra.

References

[1] P. J. Davis, Interpolation and approximation, Blaisdell, 1963.
[2] A. Ralston and P. Rabinowitz, A first course in numerical analysis, McGraw-Hill, second edition, 1978.
[3] J. F. Steffensen, Interpolation, Chelsea, 1950.
[4] P. J. Davis and I. Polonsky, Numerical interpolation, differentiation and integration, Handbook of Mathematical Functions, M. Abramowitz and I. A. Stegun (eds.), Nat. Bur. Stand. Appl. Math. Ser., 55 (1964), 875–924.
[5] R. W. Hamming, Numerical methods for scientists and engineers, McGraw-Hill, second edition, 1973.
[6] J. H. Ahlberg, E. N. Nilson, and J. L. Walsh, The theory of splines and their applications, Academic Press, 1967.
[7] A. R. Mitchell and R. Wait, The finite element method in partial differential equations, Wiley, 1977.

224 (XII.11) Interpolation of Operators

A. General Remarks

In 1926 M. Riesz proved that if T is a bounded linear operator $L_{p_0}(\Omega)\to L_{q_0}(\Omega')$ with norm N_0 and $L_{p_1}(\Omega)\to L_{q_1}(\Omega')$ with norm N_1, $1\leqslant p_0, p_1, q_0, q_1\leqslant\infty$, simultaneously, then it is a bounded linear operator $L_p(\Omega)\to L_q(\Omega')$ with norm $\leqslant N_0^{1-\theta}N_1^{\theta}$ whenever

$$1/p=(1-\theta)/p_0+\theta/p_1, \tag{1}$$

$$1/q=(1-\theta)/q_0+\theta/q_1, \tag{2}$$

for $0<\theta<1$, provided that $q\geqslant p$. In 1939 G. O. Thorin removed the restriction $q\geqslant p$ by devising a proof based on function theory (**Riesz-Thorin theorem**). Meanwhile, J. Marcinkiewicz (1939) announced that the boundedness of $T:L_p(\Omega)\to L_q(\Omega')$ holds for a quasilinear operator T under weak type assumptions (→ Section E (2)).

From 1959 to 1964, J.-L. Lions, A. P. Calderón, J. Peetre, and others extended these results to linear operators from a couple of Banach spaces to another couple. The interpolation methods provide a powerful and often essential tool in various fields of mathematical analysis where estimates of operators play a central role, such as harmonic analysis [4, 8, 9], numerical analysis, approximation theory [7], and the theory of partial differential equations [2].

A compatible couple is a pair $\{X_0, X_1\}$ of Banach spaces (or more general topological linear spaces) that are continuously embedded in a Hausdorff topological linear space. Let $\{Y_0, Y_1\}$ be another compatible couple. A linear operator $T:X_0+X_1\to Y_0+Y_1$ is said to be continuous $\{X_0,X_1\}\to\{Y_0,Y_1\}$ with norm $\{N_0,N_1\}$ if for each $l=0, 1$, $T:X_l\to Y_l$ is continuous linear with norm N_l.

An **interpolation method** is a †functor which assigns to each compatible couple $\{X_0,X_1\}$ of Banach spaces a Banach space X with $X_0\cap X_1\subset X\subset X_0+X_1$ such that every continuous linear operator $T:\{X_0,X_1\}\to\{Y_0,Y_1\}$ induces a continuous linear operator $T:X\to Y$. X is called an **interpolation space** of $\{X_0,X_1\}$. There are two important types of interpolation methods, the complex method (due to Calderón [5], S. G. Krein, Lions) and the real method (due to E. Gagliardo, Lions, Lions and Peetre [6], Peetre). These methods generalize the classical results of Riesz and Thorin and of Marcinkiewicz, respectively.

B. The Complex Interpolation Method

In this section $\{X_0,X_1\}$ is assumed to be a compatible couple of complex Banach spaces. Let $F(X_0,X_1)$ be the Banach space of all bounded continuous functions $f(\zeta)$, $\zeta=\xi+i\eta$, on the strip $0\leqslant\xi\leqslant1$ with values in X_0+X_1, holomorphic in $0<\xi<1$, and such that for each $l=0, 1$, $f(l+i\eta)$ is a continuous and bounded X_l-valued function. $\|f\|_{F(X_0,X_1)}=\max\{\sup\|f(i\eta)\|_{X_0}, \sup\|f(1+i\eta)\|_{X_1}\}$ is the norm. The **complex interpolation space** $[X_0,X_1]_\theta$, $0<\theta<1$, is defined to be the Banach space of values $f(\theta)$ of $f\in F(X_0,X_1)$ with the norm $\|x\|_{[X_0,X_1]_\theta}=\inf\{\|f\|_{F(X_0,X_1)}\mid x=f(\theta)\}$. $X_0\cap X_1$ is dense in $[X_0,X_1]_\theta$.

Let $\{X_0,X_1\}$ and $\{Y_0,Y_1\}$ be two compatible couples, and let $T:\{X_0,X_1\}\to\{Y_0,Y_1\}$ be a bounded linear operator with norm $\{N_0,N_1\}$. Then $T:[X_0,X_1]_\theta\to[Y_0,Y_1]_\theta$ is bounded linear with norm $\leqslant N_0^{1-\theta}N_1^{\theta}$ (**interpolation theorem**).

If a holomorphic family $\{T_\zeta\}$, $0<\mathrm{Re}\,\zeta<1$, of linear operators, acting in X_0+X_1 into Y_0+Y_1, fulfills a certain boundedness requirement, then T_θ induces a bounded linear operator $[X_0,X_1]_\theta\to[Y_0,Y_1]_\theta$ (E. Stein [9]).

C. The Real Interpolation Method

The method of this section applies also to compatible couples of †quasinormed spaces (P.

Krée (1967)). If $\{X_0, X_1\}$ is a compatible couple of Banach spaces (resp. quasinormed spaces), then $X_0 \cap X_1$ and $X_0 + X_1$ are Banach spaces (resp. quasinormed spaces) under the norms $J(t,x) = \max\{\|x\|_{X_0}, t\|x\|_{X_1}\}$ and $K(t,x) = \inf\{\|x_0\|_{X_0} + t\|x_1\|_{X_1} \mid x = x_0 + x_1\}$, respectively, for any $t > 0$. Denote by L_p^*, $0 < p \leqslant \infty$, the L_p-space on $(0, \infty)$ relative to the measure dt/t. Then the **real interpolation space** $(X_0, X_1)_{\theta,p}$, $0 < \theta < 1$, $1 \leqslant p \leqslant \infty$ (resp. $0 < p \leqslant \infty$), is defined to be the Banach space (resp. quasinormed space) of all $x \in X_0 + X_1$ such that $t^{-\theta}K(t,x) \in L_p^*$ with the norm $\|x\|_{(X_0,X_1)_{\theta,p}} = \|t^{-\theta}K(t,x)\|_{L_p^*}$ (Peetre's **K-method**). $X_0 \cap X_1$ is dense in $(X_0, X_1)_{\theta,p}$ if $p < \infty$. The continuous embedding $(X_0, X_1)_{\theta,p} \subset (X_0, X_1)_{\theta,q}$ holds for $p \leqslant q \leqslant \infty$.

If $T: \{X_0, X_1\} \to \{Y_0, Y_1\}$ is a bounded linear operator with norm $\{N_0, N_1\}$, then $T: (X_0, X_1)_{\theta,p} \to (Y_0, Y_1)_{\theta,p}$, $0 < \theta < 1$, $p \leqslant \infty$, is bounded linear with norm $\leqslant N_0^{1-\theta} N_1^{\theta}$ (**interpolation theorem**).

The linearity or boundedness requirements of T can be weakened considerably (Krée, T. Holmstedt, H. Komatsu (1981)).

There are several equivalent definitions of the real interpolation spaces for compatible couples $\{X_0, X_1\}$ of Banach spaces. The **J-method** gives $(X_0, X_1)^J_{\theta,p}$, $0 < \theta < 1$, $1 \leqslant p \leqslant \infty$, which is the space of means $x = \int_0^\infty u(t)\,dt/t$ with the norm $\|x\|_{(X_0,X_1)^J_{\theta,p}} = \inf_u \|t^{-\theta}J(t, u(t))\|_{L_p^*}$, where $u(t)$ are strongly measurable functions on $(0, \infty)$ with values in $X_0 \cap X_1$ such that $t^{-\theta}J(t, u(t)) \in L_p^*$. $(X_0, X_1)^J_{\theta,p} = (X_0, X_1)_{\theta,p}$ holds with equivalent norms. A discrete version of the J-method applies also to couples $\{X_0, X_1\}$ of †quasi-Banach spaces.

D. The Reiteration Theorem

Let $\{X_0, X_1\}$ be a compatible couple of (quasi-) Banach spaces. A (quasi-) Banach space X is said to be of class $K_\theta(X_0, X_1)$, $0 \leqslant \theta \leqslant 1$, if $X_0 \cap X_1 \subset X \subset X_0 + X_1$ with continuous embeddings and if there are constants C_K and C_J such that $K(t,x) \leqslant C_K t^{\theta} \|x\|_X$, $x \in X$, and $t^{\theta}\|x\|_X \leqslant C_J J(t,x)$, $x \in X_0 \cap X_1$. The class $K_\theta(X_0, X_1)$ contains $(X_0, X_1)_{\theta,p}$, $0 < \theta < 1$, $0 < p \leqslant \infty$, and, for Banach spaces X_0, X_1, $[X_0, X_1]_\theta$. If Y_α is of class $K_\alpha(X_0, X_1)$ and Y_β of class $K_\beta(X_0, X_1)$ and if $0 \leqslant \alpha < \beta \leqslant 1$, then $(Y_\alpha, Y_\beta)_{\lambda,p} = (X_0, X_1)_{\theta(\lambda),p}$, $0 < \lambda < 1$, $0 < p \leqslant \infty$, $\theta(\lambda) = (1-\lambda)\alpha + \lambda\beta$, with equivalent norms (**reiteration theorem**).

Set $Y_\alpha = [X_0, X_1]_\alpha$ and $Y_\beta = [X_0, X_1]_\beta$ for Banach spaces X_0 and X_1. If $X_0 \cap X_1$ is dense in X_0, X_1 and $Y_\alpha \cap Y_\beta$, then $[Y_\alpha, Y_\beta]_\lambda = [X_0, X_1]_{\theta(\lambda)}$, $0 \leqslant \lambda \leqslant 1$, with equal norms (**reiteration theorem**).

E. Examples and Applications

(1) Let $L_p = L_p(\Omega, \mu)$ be the L_p-space on a σ-finite measure space (Ω, μ). Then $[L_{p_0}, L_{p_1}]_\theta = L_{p_\theta}$ holds with equal norms for $1/p_\theta = (1-\theta)/p_0 + \theta/p_1$, $0 < \theta < 1$, $1 \leqslant p_0, p_1 \leqslant \infty$. Hence follows the Riesz-Thorin theorem. In particular, the Fourier transform maps $L_p(\mathbf{R}^n)$, $1 \leqslant p \leqslant 2$, continuously into $L_{p'}(\mathbf{R}^n)$, $p' = p/(p-1)$ (the **Hausdorff-Young inequality**). The convolution operator $K*$, $K \in L_r(\mathbf{R}^n)$, $1 \leqslant r < \infty$, is bounded from $L_p(\mathbf{R}^n)$ into $L_q(\mathbf{R}^n)$ with norm $\leqslant \|K\|_{L_r}$ if $p \geqslant 1$, $1/q = 1/p + 1/r - 1 \geqslant 0$ (**Young's inequality**).

For the †Hardy space $H_1(\mathbf{R}^n)$, $[L_p(\mathbf{R}^n), H_1(\mathbf{R}^n)]_\theta = L_q(\mathbf{R}^n)$, $1/q = (1-\theta)/p + \theta$, $1 < p \leqslant \infty$, $0 < \theta < 1$ (Stein and C. Fefferman).

(2) Recall that the †Lorentz space $L_{(p,q)} = L_{(p,q)}(\Omega, \mu)$, $0 < p$, $q \leqslant \infty$, is the †quasi-Banach space of all measurable functions f with $\|f\|_{(p,q)} = \|t^{1/p} f^*(t)\|_{L_q^*} < \infty$, where $f^*(t)$ is the †rearrangement of $f(\omega)$, $\omega \in \Omega$. $L_{(p,p)} = L_p$ and $L_{(p,q)} \subset L_{(p,r)}$ for $q < r$. If $1 < p < \infty$, $1 \leqslant q \leqslant \infty$, then $L_{(p,q)}$ coincides with the Banach space $(L_\infty, L_1)_{1/p,q}$ under the equivalent norm $\|t^{1/p-1}\int_0^t f^*(s)\,ds\|_{L_q^*}$. More generally, if $0 < p_0$, $p_1 < \infty$, $p_0 \neq p_1$, $0 < \theta < 1$, $1/p_\theta = (1-\theta)/p_0 + \theta/p_1$, $0 < q_0, q_1, q \leqslant \infty$, then $(L_{(p_0,q_0)}, L_{(p_1,q_1)})_{\theta,q} = L_{(p_\theta,q)}$ with equivalent norms.

Let T be an operator which maps a space of measurable functions on (Ω, μ) into another on (Ω', μ'). The inequality $\|Tf\|_{(q,\infty)} \leqslant N\|f\|_{L_p}$ holds if and only if $\mu'\{\omega' \in \Omega' \mid |Tf(\omega')| > s\} \leqslant (N\|f\|_{L_p}/s)^q$, $s > 0$. Then T is said to be of **weak type** (p,q).

Suppose that T is **quasilinear**, i.e., that $T(f+g)$ is uniquely determined whenever Tf and Tg are defined and that $|T(f+g)(\omega')| \leqslant K\{|Tf(\omega')| + |Tg(\omega')|\}$ a.e. holds with a constant K independent of f and g. The interpolation theorem then holds. Therefore, if a quasilinear operator T is of weak type (p_0, q_0) and (p_1, q_1), $q_0 \neq q_1$, then T is of **type** (p,q), i.e., $T: L_p(\Omega) \to L_p(\Omega')$ is bounded for p, q satisfying (1), (2), and $q \geqslant p$ (**Marcinkiewicz's theorem**). When $p_0 \neq p_1$, the same conclusion is obtained if $\|Tf\|_{(q_l,\infty)} \leqslant N_l \|f\|_{(p_l,r_l)}$, $l = 0, 1$, for some r_l (R. A. Hunt).

The †Hilbert transform and the †Calderón-Zygmund singular integral operators are of weak type $(1,1)$ and of type $(2,2)$. Hence it follows from these facts together with duality that they are of type (p,p) for $1 < p < \infty$ [4, 8, 9].

The convolution operator $K*$, $K \in L_{(r,\infty)}(\mathbf{R}^n)$, $1 < r < \infty$, is of type (p,q) if $p > 1$, $1/q = 1/p + 1/r - 1 > 0$ (R. O'Neil). In particular, the convolution with $|x|^{\alpha - n}$, $0 < \alpha < n$, on $\mathbf{R}^n$ is of type (p,q) for $1 < p < n/\alpha$ and $1/q = 1/p - \alpha/n$ (the **Hardy-Littlewood-Sobolev inequality**).

For the †Hardy spaces $H_p = H_p(\mathbf{R}^n)$ and the

†John-Nirenberg space $BMO=BMO(\mathbf{R}^n)$, $(L_\infty, H_{p_1})_{\theta,q}=(BMO, H_{p_1})_{\theta,q}=H_q$ and $(H_{p_0}, H_{p_1})_{\theta,p_\theta}=H_{p_\theta}$ if $0<\theta<1$, $0<p_0, p_1<\infty$, $1/q=\theta/p_1$ and $1/p_\theta=(1-\theta)/p_0+\theta/p_1$ (S. Igari, N. Rivière, Y. Sagher, Fefferman, R. Hanks).

A †Besov space $B^s_{p,q}(\Omega)$ is understood most naturally as the real interpolation space $(L_p(\Omega), W^m_p(\Omega))_{s/m,q}$ of the Lebesgue space $L_p(\Omega)$ and the †Sobolev space $W^m_p(\Omega)$ [2, 3].

(3) Let A be a closed linear operator in a Banach space X such that the resolvent $(t+A)^{-1}$ exists for $t>0$ and satisfies $\|t(t+A)^{-1}\|\leqslant M$. If $D(A^m)$ denotes the domain of the integral mth power A^m of A equipped with the graph norm, then $D^\sigma_p(A)=(X, D(A^m))_{\sigma/m,p}$, $0<\sigma<m$, $1\leqslant p\leqslant\infty$, coincides with the space of $x\in X$ such that $\|t^\sigma(A(t+A)^{-1})^m x\|_X\in L^*_p$ and is independent of m. If $-A$ generates an †equicontinuous semigroup T_t of class (C^0) (resp. a †holomorphic semigroup T_t bounded on a sector), then $D^\sigma_p(A)$ consists also of all $x\in X$ such that $\|t^{-\sigma}(I-T_t)^m x\|_X\in L^*_p$ (resp. $\|t^{-\sigma+m}A^m T_t x\|_X\in L^*_p$). There are similar characterizations of elements in $(X, \cap D(A_i^{m_i}))_{\theta,p}$ for a commutative family $\{A_1, \ldots, A_k\}$ of such operators (Lions and Peetre, P. Grisvard, P. L. Butzer and H. Berens, Komatsu, T. Muramatu).

When $A=-\Delta$ or $A_i=\partial/\partial x_i$ in $\mathbf{R}^n$ or in a suitable domain Ω in $\mathbf{R}^n$, these results give various equivalent characterizations of the functions in Besov spaces $B^s_{p,q}(\Omega)$. The Sobolev embedding theorem for Besov spaces can also be proved from this point of view (Grisvard, Peetre, A. Yoshikawa, Komatsu).

The space $D^\sigma_p(A)$ is closely connected with the domains $D(A^\alpha)$ of †fractional powers A^α of A. $D^\sigma_1(A)\subset D(A^\alpha)\subset D^\sigma_\infty(A)$, $\sigma=\operatorname{Re}\alpha>0$. If $0<\operatorname{Re}\beta<\sigma$, then $D^\sigma_p(A)=\{x\in D(A^\beta)\mid A^\beta x\in D^{\sigma-\operatorname{Re}\beta}_p(A)\}$. If the pure imaginary powers $A^{i\eta}$ are locally uniformly bounded, then $D(A^\alpha)=[X, D(A^m)]_{\operatorname{Re}\alpha/m}$, $0<\operatorname{Re}\alpha<m$.

F. Duality

Suppose that $\{X_0, X_1\}$ is a compatible couple of Banach spaces such that $X_0\cap X_1$ is dense both in X_0 and in X_1. Denote the dual of X by X'. If one of X_0 and X_1 is reflexive, then so are $[X_0, X_1]_\theta$ and $(X_0, X_1)_{\theta,p}$, $0<\theta<1$, $1\leqslant p\leqslant\infty$. Furthermore, $[X_0, X_1]'_\theta=[X'_0, X'_1]_\theta$ and $(X_0, X_1)'_{\theta,p}=(X'_0, X'_1)_{\theta,p'}$, $p'=p/(p-1)$, with equivalent norms (Calderón, Lions and Peetre, H. Morimoto).

References

[1] J. Bergh and J. Löfström. Interpolation spaces, Springer, 1976.

[2] H. Triebel, Interpolation theory, function spaces, differential operators, North-Holland, 1978.

[3] J. Peetre, New thoughts on Besov spaces, Math. Dept., Duke Univ., 1976.

[4] A. Zygmund, Trigonometric series, Cambridge Univ. Press, second edition, 1959.

[5] A. P. Calderón, Intermediate spaces and interpolation: The complex method, Studia Math., 24 (1964) 113–190.

[6] J.-L. Lions and J. Peetre, Sur une classe d'espaces d'interpolation, Publ. Math. Inst. HES, 19 (1964), 5–68.

[7] P. Butzer and H. Berens, Semigroups of operators and approximation, Springer, 1967.

[8] E. Stein, Singular integrals and differentiability properties of functions, Princeton Univ. Press, 1970.

[9] E. Stein and G. Weiss, Introduction to Fourier analysis on Euclidean spaces, Princeton Univ. Press, 1971.

225 (X.14) Invariant Measures

A. Introduction

The †Lebesgue measure in the Euclidean space $\mathbf{R}^n$ is invariant under Euclidean motion (translation and rotation), and the measure $d\mu(\lambda)=d\lambda/\lambda$ on the half-line of positive numbers is invariant under multiplication. If we define usual †spherical coordinates (θ, φ) on the unit sphere S^2 in $\mathbf{R}^3$, the measure $d\mu(\omega)=\sin\theta\, d\theta\, d\varphi$ $(\omega=(\theta,\varphi))$ on S^2 is invariant under rotation. More generally, if we define the spherical coordinates $(\theta_1, \ldots, \theta_{n-1}, \varphi)$ on the unit sphere S^n in $\mathbf{R}^{n+1}$, the measure

$$d\mu(\omega)=\sin^{n-1}\theta_1 \sin^{n-2}\theta_2 \ldots \sin\theta_{n-1}\, d\theta_1\, d\theta_2 \ldots d\theta_{n-1}\, d\varphi$$

on S^n is invariant under rotation, where the spherical coordinates are related to the rectangular coordinates $(x_1, x_2, \ldots, x_n, x_{n+1})$ in $\mathbf{R}^{n+1}$ as follows:

$$x_1=\sin\theta_1\sin\theta_2\ldots\sin\theta_{n-1}\cos\varphi,$$

$$x_2=\sin\theta_1\sin\theta_2\ldots\sin\theta_{n-1}\sin\varphi,$$

$$x_3=\sin\theta_1\ldots\sin\theta_{n-2}\cos\theta_{n-1},$$

$$\ldots$$

$$x_{n-1}=\sin\theta_1\sin\theta_2\cos\theta_3,$$

$$x_n=\sin\theta_1\cos\theta_2,$$

$$x_{n+1}=\cos\theta_1$$

$$(0\leqslant\theta_\nu\leqslant\pi,\ \nu=1,\ldots,n-1;\ 0\leqslant\varphi\leqslant 2\pi).$$

The notion of measures invariant under certain transformations is generalized to that of invariant measures, defined in the following section.

B. Definitions

Let μ be a †measure defined on a †σ-algebra $\mathfrak{B}$ of subsets of a space X, and G be a transformation group acting on X in such a way that $sA \in \mathfrak{B}$ for any $A \in \mathfrak{B}$ and $s \in G$. For every $s \in G$, define a measure $\gamma(s)\mu$ on $\mathfrak{B}$ by $(\gamma(s)\mu)(sA) = \mu(A)$. If $\gamma(s)\mu = \mu$ for all $s \in G$, then μ is called an **invariant measure** with respect to G (or G-**invariant measure**).

We consider the case where X is a †locally compact Hausdorff space and G is a †topological transformation group of X. We further suppose that $\mathfrak{B}$ is the smallest σ-algebra containing the family $\mathfrak{C}$ of all compact subsets of X and that $\mu(K) < \infty$ for all $K \in \mathfrak{C}$ (→ 270 Measure Theory). Let $C_0(X)$ be the space of all real-valued continuous functions with compact †support defined on X. For example, if X is an †oriented †Riemannian manifold and ω is the †volume element associated with the Riemannian metric on X, there exists a unique measure μ on X such that

$$\int_X f(x)\, d\mu(x) = \int_X f\omega \qquad (1)$$

for every $f(x) \in C_0(X)$. This measure μ is invariant under the group G of †isometries of X. In the case of a nonorientable X, a G-invariant measure can also be defined from the †Riemannian metric.

In the following, we consider G-invariant measures on †homogeneous spaces X of a locally compact Hausdorff topological group (abbreviated to locally compact group) G.

C. Haar Measures

Most fundamental is the case in which G is locally compact and $X = G$, with sx (resp. xs^{-1}) defined by the group multiplication law. In this case, a nonzero G-invariant measure on G is called a **left-** (resp. **right-**) **invariant Haar measure** on G. On every locally compact group, there exists a left- (right-) invariant Haar measure, which is unique up to a positive multiplicative constant **(Haar's theorem)**. For example, Haar measures on the additive group **R** of real numbers and the additive group $\mathbf{R}^n$ are the usual †Lebesgue measures. A Haar measure μ on the multiplicative group $\mathbf{R}_+^*$ of positive real numbers is given by

$$\int_0^\infty f(x)\, d\mu(x) = \int_0^\infty f(x)\, dx/x.$$

For an n-dimensional †Lie group G, a left-invariant Haar measure μ is defined by formula (1) with a left-invariant †differential form ω of degree n.

A Haar measure μ on a locally compact group G is †regular in the following sense. If $\mathfrak{O}$ is the set of all open subsets of G, then for every $A \in \mathfrak{B}$, we have

$$\mu(A) = \sup\{\mu(K) \mid K \in \mathfrak{C}, K \subset A\}$$
$$= \inf\{\mu(U) \mid U \in \mathfrak{B} \cap \mathfrak{O}, A \subset U\},$$

For $U \in \mathfrak{B} \cap \mathfrak{O}$ ($U \neq \varnothing$), we have $\mu(U) > 0$, and $\mu(A) < \infty$ for compact A. The measure $\mu(s)$ of one point s is >0 if and only if G is †discrete. The total measure $\mu(G)$ of G is finite if and only if G is compact.

D. Modular Functions

Let μ be a left-invariant Haar measure on a locally compact group G, and define the measure $\delta(s)\mu$ by $(\delta(s)\mu)(A) = \mu(As)$ for every $s \in G$. Since $\delta(s)\mu$ is also a left-invariant Haar measure, there exists a positive real number $\Delta(s)$ such that $\delta(s)\mu = \Delta(s)\mu$, by virtue of the uniqueness of the left-invariant Haar measure. The function $\Delta = \Delta_G$ on G is called the **modular function** of G. For an †integrable function on G with respect to μ, we have

$$\int_G f(xs)\, d\mu(x) = \Delta(s)^{-1} \int_G f(x)\, d\mu(x),$$

$$\int_G f(x^{-1}) \Delta(x)^{-1}\, d\mu(x) = \int_G f(x)\, d\mu(x).$$

If ν is a right-invariant Haar measure on G, we have the formulas

$$\int_G f(sx)\, d\nu(x) = \Delta(s) \int_G f(x)\, d\nu(x),$$

$$\int_G f(x^{-1}) \Delta(x)\, d\nu(x) = \int_G f(x)\, d\nu(x).$$

Moreover, $\Delta^{-1}\mu$ is a right-invariant Haar measure, while $\Delta\nu$ is a left-invariant Haar measure.

The modular function Δ of G is a continuous homomorphism of G into the multiplicative group $\mathbf{R}_+^*$ of positive real numbers. If the modular function Δ of G is equal to the constant 1, i.e., if a left-invariant Haar measure is also right-invariant, G is said to be **unimodular**. G is unimodular if G is compact, commutative, or discrete. If G is a †Lie group, we have $\Delta(s) = |\det \mathrm{Ad}(s)^{-1}|$, where $s \to \mathrm{Ad}(s)$ is the †adjoint representation. In particular, G is unimodular if G is a †semisimple Lie group, a connected †nilpotent Lie group, or a Lie group for which $\mathrm{Ad}\, G$ is compact. However,

the group $T(n;\mathbf{R})$ of right triangular matrices $(n>1)$ is not unimodular.

E. Product Measures

Let $\{G_\alpha\}_{\alpha\in A}$ be a family of locally compact groups, and let μ_α be a left-invariant Haar measure on G_α for every $\alpha\in A$. Suppose that there exists a finite subset B of A such that G_α is compact and $\mu_\alpha(G_\alpha)=1$ for $\alpha\in A-B$. The product measure $\mu=\prod_{\alpha\in A}\mu_\alpha$ is then a left-invariant Haar measure on the †Cartesian product $G=\prod_{\alpha\in A}G_\alpha$, which is also a locally compact group. Moreover, if Δ_α is the modular function of G_α, then $\Delta_G(x)=\prod_{\alpha\in A}\Delta_\alpha(x_\alpha)$ for $x=(x_\alpha)_{\alpha\in A}$.

F. Product Formula

Let H, L be two closed subgroups of a locally compact group G, and suppose that $\Omega=HL$ contains a neighborhood V of e in G. This means that Ω is an open subset of G. If we put $D=\{(s,s)\,|\,s\in H\cap L\}$, then the mapping $(s,t)\rightarrow st^{-1}$ of $H\times L$ into Ω induces a one-to-one continuous mapping φ of the quotient space $H\times L/D$ onto Ω. Suppose that φ is a homeomorphism. This is the case, for example, if G is †paracompact. Furthermore, if $H\cap L$ is compact, we have the **product formula**

$$\int_\Omega f(\omega)\,d\mu(\omega)$$

$$=a\int\int_{H\times L} f(hl)\Delta_G(l)\Delta_L(l)^{-1}\,d\mu_H(h)\,d\mu_L(l),$$

where μ, μ_H, μ_L denote left-invariant Haar measures on G, H, L, respectively, and $a>0$ is a constant independent of f.

G. Weil Measures

If A is a measurable subset with respect to a left-invariant Haar measure μ and $\mu(A)>0$, then $A^{-1}A=\{s^{-1}t\,|\,s,t\in A\}$ is a neighborhood of the identity element of G, and such subsets form a base for the neighborhood system of the identity. This shows that the topology of a locally compact group is determined by its Haar measure. Conversely, we shall consider the definition of a topology in an abstract group G with a measure μ.

Let μ be a †σ-finite measure defined on a †σ-additive family $\mathfrak{B}$ in G, such that $sA\in\mathfrak{B}$ for $A\in\mathfrak{B}$ and $s\in G$. μ is called a **Weil measure** if it satisfies the following two conditions: (W1) $\mu(sA)=\mu(A)$; (W2) if $f(x)$ is $\mathfrak{B}$-measurable, then $f(y^{-1}x)$ is $\mathfrak{B}\times\mathfrak{B}$-measurable.

If a Weil measure $\mu\neq 0$ exists in a group G, then $\{A^{-1}A\,|\,\mu(A)>0\}$ forms a base for the neighborhood system of the identity element of a topology, which makes G a locally †totally bounded topological group. If for every $s\in G$ there exists an $A\in\mathfrak{B}$ such that $\mu(A\cap sA)<\mu(A)<\infty$, $\mu(A)>0$, then G is a Hausdorff space. In this case the †completion $\bar{G}$ of G is a locally compact group, and for a suitable left-invariant Haar measure $\bar{\mu}$ on $\bar{G}$, we have $A=\bar{A}\cap G\in\mathfrak{B}$ and $\mu(A)=\bar{\mu}(\bar{A})$ for every $\bar{A}\in\bar{\mathfrak{B}}$ (the smallest σ-additive family containing the family of all compact subsets of $\bar{G}$).

H. Relatively Invariant Measures

Let G be a transformation group acting on a set X. A measure μ on X is said to be a **relatively invariant measure** with respect to G if for every $s\in G$, $\gamma(s)\mu$ is proportional to μ, i.e., $\gamma(s)\mu=\chi(s)^{-1}\cdot\mu$ $(\chi(s)\in\mathbf{R}_+^*)$. If $\mu\neq 0$, $\chi(s)$ is uniquely determined by s, and $s\rightarrow\chi(s)$ is a continuous homomorphism from G into the multiplicative group $\mathbf{R}_+^*$ of positive real numbers. We call χ the **multiplicator** of the relative invariant measure.

We now consider relatively invariant measures with respect to a locally compact group G on the †quotient space G/H of G by a closed subgroup H. Let μ, β be left-invariant Haar measures on G, H, respectively, and let $x\rightarrow x^*$ be the canonical mapping of G onto G/H. For any measure λ on G/H, there exists a unique measure $\lambda^\#$ on G satisfying the condition

$$\int_{G/H}\left(\int_H f(xh)\,d\beta(h)\right)d\lambda(x^*)=\int_G f(x)\,d\lambda^\#(x)$$

for every continuous function f with compact support on G. For every $h\in H$, we have $\delta(h)\lambda^\#=\Delta_H(h)\lambda^\#$. Conversely, for a measure ν on G such that $\delta(h)\nu=\Delta_H(h)\nu$ for every $h\in H$, there exists a unique measure λ on G/H such that $\lambda^\#=\nu$. This measure λ is called the **quotient measure** of ν by β and is denoted by $\lambda=\nu/\beta$. For a continuous homomorphism χ of G into the multiplicative group $\mathbf{R}_+^*$ of positive real numbers, a necessary and sufficient condition for the existence of a not identically zero, relatively invariant measure on G/H with the multiplicator χ is that $\chi(h)=\Delta_H(h)/\Delta_G(h)$ for every $h\in H$. If this condition is satisfied, the relatively invariant measure on G/H with multiplicator χ is unique up to a multiplicative constant and is given by the quotient measure $\nu=(\chi\mu)/\beta$ of $\chi\mu$ by β. In particular, for the existence of a G-invariant measure on G/H, it is necessary and sufficient that the modular functions Δ_G and Δ_H coincide on H. Hence, if H is compact or if G and H are unimodular, there exists a G-invariant measure on G/H.

I. Weyl's Integral Formula

Let G be a compact connected †semisimple Lie group and H a †Cartan subgroup (maximal torus) of G. Then a Haar measure μ on G can be expressed by means of a Haar measure β on H and a G-invariant measure λ on G/H. If μ, β, λ are all normalized to be of total measure 1, we have the following formula for every continuous function f on G:

$$\int_G f(g)\,d\mu(g)$$
$$=\frac{1}{w}\int_H\int_{G/H} f(ghg^{-1})J(h)\,d\lambda(g^*)\,d\beta(h)$$

(**Weyl's integral formula**), where w is the order of the †Weyl group of G and $J(h)$ is given by

$$J(\exp X)=\left|\prod_{\alpha\in P}(e^{\alpha(X)/2}-e^{-\alpha(X)/2})\right|^2,$$

with P the set of all †positive roots α of G with respect to H and X an arbitrary element of the †Lie algebra of H. For an element h of H, the element X with $h=\exp X$ is not unique, but the function J is a single-valued function. A similar formula is valid on a †symmetric Riemannian manifold. Weyl's integral formula can also be generalized to the case of noncompact semisimple Lie groups. However, it is then necessary to replace the right-hand side by a sum extended over a system of representatives of mutually nonconjugate Cartan subgroups.

J. Quasi-Invariant Measure

Suppose that a group G acts on a space X in such a way that $sA\in\mathfrak{B}$ for any $A\in\mathfrak{B}$ and $s\in G$, where $\mathfrak{B}$ is a σ-additive family of subsets of X. A measure μ defined on $\mathfrak{B}$ is called a **quasi-invariant measure** with respect to G if the measures μ and $\gamma(s)\mu$ (Section A) are equivalent for every $s\in G$. Here, two measures λ and μ defined on $\mathfrak{B}$ are equivalent if $\lambda=\varphi\mu$ (this formula means $\lambda(E)=\int_E\varphi(x)\,d\mu(x)$ for any $E\in\mathfrak{B}$) for some measurable function $\varphi(x)$ which is >0 almost everywhere with respect to μ and μ-integrable on every $A\in\mathfrak{B}$ such that $\mu(A)<\infty$.

We now consider quasi-invariant measures with respect to a locally compact group G on a quotient space G/H of G by a closed subgroup H. There are always quasi-invariant measures on G/H with respect to G, and they are all mutually equivalent. They can be constructed as follows. There exists a positive continuous function ρ on G such that $\rho(gh)=\Delta_H(h)\Delta_G(h)^{-1}\rho(g)$ for $g\in G$, $h\in H$. Then the quotient measure $\lambda=(\rho\mu)/\beta$ is a nonzero quasi-invariant measure on G/H with respect to G if μ, β are Haar measures on G, H, respectively, and it holds that $d\lambda(sx)=\rho(sg)\rho(g)^{-1}\,d\lambda(x)$ for $x\in gH\in G/H$ and $s\in G$. If G is a Lie group, we can take a function ρ of †class C^∞. If X is an infinite-dimensional †locally convex topological vector space over $\mathbf{R}$, there exists no †σ-finite Borel measure on X that is quasi-invariant with respect to translations by the elements of X [6].

References

[1] N. Bourbaki, Eléments de mathématique, Intégration, ch. 7, 8, Actualités Sci. Ind., 1306, Hermann, 1963.
[2] A. Weil, L'intégration dans les groupes topologiques et ses applications, Actualités Sci. Ind., Hermann, 1940.
[3] P. R. Halmos, Measure theory, Van Nostrand, 1950.
[4] S. Helgason, Differential geometry and symmetric spaces, Academic Press, 1962.
[5] L. H. Loomis, An introduction to abstract harmonic analysis, Van Nostrand, 1953.
[6] Y. Umemura, Measures on infinite dimensional vector spaces, Publ. Res. Inst. Math. Sci., (A) 1 (1965), 1–47.
[7] A. Haar, Der Massbegriff in der Theorie der kontinuierlichen Gruppen, Ann. Math., (2) 34 (1933), 147–169.
[8] J. von Neumann, The uniqueness of Haar's measure, Mat. Sb., 1 (1936), 721–734.
[9] H. Cartan, Sur la mesure de Haar, C. R. Acad. Sci. Paris, 211 (1940), 759–762.

226 (IV.18) Invariants and Covariants

A. The General Case

Let R be a †commutative ring. We say that a group G **acts** on R if (i) each element σ of G defines an †automorphism $f\to\sigma f$ of R, and (ii) $\sigma(\tau f)=(\sigma\tau)f$ for any $\sigma,\tau\in G, f\in R$. In this case, an element f of R is said to be G-**invariant** (or simply **invariant**) if $\sigma f=f$ for any $\sigma\in G$. An element f is called (G-)**semi-invariant** if for each σ in G there is an invariant $a(\sigma)$ such that $\sigma f=a(\sigma)f$, that is, if f is invariant up to an invariant multiplier depending on σ. A semi-invariant may also be called a **relative invariant**, and an invariant may be called an **absolute invariant**. The correspondence $\sigma\to a(\sigma)$ mod$(0:f)$ ($(0:f)=\{x\in R\mid xf=0\}$) is a †repre-

sentation of degree 1 of G, and $a(\sigma)$ is called the **multiplier** of the semi-invariant f, or the **character** defined by f.

B. Invariants of Matrix Groups

Let K be a commutative ring with a unity element. Then we can consider a **matrix group** (or **matric group**) G over K, i.e., a subgroup of the group of $n \times n$ invertible matrices over K; this latter group is called the **general linear group** of degree n over K and is denoted by $GL(n, K)$ (→ 60 Classical Groups). Assume that R is a commutative ring generated by $x_1, \ldots, x_n$ over K and an action of the group G on R is defined such that ${}^t(\sigma x_1, \ldots, \sigma x_n) = \sigma {}^t(x_1, \ldots, x_n)$ (^{t} means the †transpose of a matrix). In this case, we say that G acts on R as a matrix group. If K is a field, then the smallest †algebraic group $\bar{G}$ (in $GL(n, K)$) containing G acts on R as a matrix group, and an element f of R is $\bar{G}$-invariant if and only if f is G-invariant, and similarly for semi-invariants. These results can be generalized to the case where K is not a field.

A †homomorphism ρ of a matrix group G ($\subset GL(n, K)$) into $GL(m, K)$ is called a **rational representation** of G if there exist rational functions φ_{kl} ($1 \leqslant k, l \leqslant m$) in n^2 variables x_{ij} ($1 \leqslant i, j \leqslant n$) with coefficients in K such that $\rho((\sigma_{ij})) = (\varphi_{kl}(\sigma_{ij}))$ for all $(\sigma_{ij}) \in G$. Assume that ρ is a rational representation of a matrix group G and $\rho(G)$ acts on a ring R as a matrix group. Then we have an action of G on R defined by $\sigma f = (\rho\sigma) f$ ($\sigma \in G, f \in R$), called the **rational action** defined by ρ. If the following condition is satisfied, then the action is called **semi-reductive** (or **geometrically reductive**): If $N = f_1 K + \ldots + f_r K$ is a G-†admissible module ($f_1, \ldots, f_r \in R$) and if $f_0 \bmod N$ ($f_0 \in R$) is G-invariant, then there is a †homogeneous form h in $f_0, \ldots, f_r$ of positive degree with coefficients in K such that h is †monic in f_0 and is G-invariant. This action is called **reductive** (or **linearly reductive**) if h can always be chosen to be a linear form.

(1) Rational actions of a matrix group G in each of the following three cases are all reductive. (i) K is either the real number field or the complex number field, and G is a dense subset of a †Lie group ($\subset GL(n, K)$) which is either †semisimple or †compact. (ii) K is a field of †characteristic 0, and letting $\bar{G}$ be the smallest algebraic group containing G, the †radical of $\bar{G}$ is a †torus group (→ 13 Algebraic Groups). (iii) K is a field of characteristic $p \neq 0$, and $\bar{G}$ (as in (ii)) contains a torus group T of finite †index that is relatively prime to p.

(2) Any action of a finite group is a semireductive action. If we omit the condition that the characteristic of K is 0 in (ii), then rational actions of G are semireductive. This is known as Mumford's conjecture and was proved by W. J. Haboush.

(3) Assume that φ is a G-†admissible homomorphism of the ring R onto another ring R'. We denote the sets of G-invariants in R and R' by $I_G(R)$ and $I_G(R')$, respectively.

If the actions of G are reductive, then (i) $\varphi(I_G(R)) = I_G(R')$; (ii) $h_i \in I_G(R)$ implies $(\sum_i h_i R) \cap I_G(R) = \sum_i h_i I_G(R)$; and (iii) if K is †Noetherian, then $I_G(R)$ is finitely generated over K.

If rational actions of G are semireductive, then (i) for each element a of $I_G(R')$, there is a natural number t such that $a^t \in \varphi(I_G(R))$, and hence $I_G(R')$ is †integral over $\varphi(I_G(R))$; (ii) if $h_i \in I_G(R)$ and $f \in (\sum_i h_i R) \cap I_G(R)$, then a suitable power f^t of f is in $\sum_i h_i I_G(R)$; and (iii) if K is a †pseudogeometric ring (in particular, if K is a field), then $I_G(R)$ is finitely generated over K [3].

When $I_G(R)$ is generated by $f_1, \ldots, f_s$ over K, then $f_1, \ldots, f_s$ are called **basic invariants.**

C. Polynomial Rings

Let $\rho_1, \ldots, \rho_u$ be matrix representations of a group G over a commutative ring K of respective degrees $n_1, \ldots, n_u$. Let $x_j^{(i)}$ ($1 \leqslant i \leqslant u$, $1 \leqslant j \leqslant n_i$) be $\sum n_i$ †algebraically independent elements over K. Then we define an action of G on the †polynomial ring $K[x_1^{(1)}, \ldots, x_{n_u}^{(u)}]$ by ${}^t(\sigma x_1^{(i)}, \ldots, \sigma x_{n_i}^{(i)}) = \rho_i(\sigma) {}^t(x_1^{(i)}, \ldots, x_{n_i}^{(i)})$. In this case, a (relative) invariant is the sum of (relative) invariants that are homogeneous in each $(x_1^{(i)}, \ldots, x_{n_i}^{(i)})$. (Because of this fact, in some of the literature, a (relative) invariant means a (relative) invariant that is homogeneous in each of $(x_1^{(i)}, \ldots, x_{n_i}^{(i)})$.) On the existence of basic invariants, the following theorem is known (besides the one on (semi)reductive actions): Assume that K is a field of characteristic 0, G is dense under the †Zariski topology in an algebraic linear group $\bar{G}$ such that the †unipotent part $(\bar{G})_u$ of the radical of $\bar{G}$ is at most 1-dimensional (these conditions hold if G is a 1-dimensional Lie group), and that all the ρ_i are rational representations; then basic invariants exist (R. Weitzenböck).

Furthermore, if G is a matrix group and each ρ_i is either the †identity map or the contragredient map $A \to {}^t A^{-1}$, then the invariants are called **vector invariants**. If K is a field of characteristic zero, the basic invariants and a basis for the ideal of algebraic relations of the basic invariants are explicitly given in several cases [1, 2].

D. Classical Terminology

The classical theory of invariants considers the following objects. Let K be a field of characteristic zero (e.g., the real number field or the complex number field), and let G be $GL(n,K)$. Consider a homogeneous form F of degree d in n variables $\xi_1,\ldots,\xi_n$ with coefficients in $K: F=\sum c_{i_1\ldots i_n} m_{i_1\ldots i_n}$ $(\sum i_\alpha=d,\ m_{i_1\ldots i_n}=(d!/\prod(i_\alpha!))\xi_1^{i_1}\ldots\xi_n^{i_n})$. For each $\sigma\in G$, we define $\sigma\xi_i$ by $(\sigma\xi_1,\ldots,\sigma\xi_n)=(\xi_1,\ldots,\xi_n)\sigma^{-1}$ and then $(\sigma c)_{i_1\ldots i_n}$ by $F=\sum(\sigma c)_{i_1\ldots i_n}(\sigma m_{i_1\ldots i_n})$. Then the transformation

$$[\sigma]_d:(c_{d0\ldots0},\ldots,c_{i_1\ldots i_n},\ldots,c_{0\ldots0d})$$

$$\rightarrow((\sigma c)_{d0\ldots0},\ldots,(\sigma c)_{i_1\ldots i_n},\ldots,(\sigma c)_{0\ldots0d})$$

is a †linear transformation of a ${}_{d+n-1}C_{n-1}$-dimensional †affine space. Let us denote the matrix of the linear transformation by the same symbol $[\sigma]_d$:

$${}^t(\ldots,(\sigma c)_{i_1\ldots i_n},\ldots)=[\sigma]_d{}^t(\ldots,c_{i_1\ldots i_n},\ldots).$$

Then $\varphi_d:\sigma\rightarrow[\sigma]_d$ is a rational representation of G.

Now fix an F such that the coefficients $c_{i_1\ldots i_n}$ are independent variables, and consider the action of G on the polynomial ring $R=K[\ldots,c_{i_1\ldots i_n},\ldots]$ defined by the rational representation φ_d. If g is a relative invariant, then $\sigma g=a(\sigma)g$ with a rational representation $\sigma\rightarrow a(\sigma)$. Hence there is an integer w such that $a(\sigma)=(\det\sigma)^w$. Then g is called an **invariant of weight** w (note that g is an $SL(n,K)$-invariant); g is an absolute invariant if and only if $w=0$. The group G acts naturally on the ring $R[\xi_1,\ldots,\xi_n]$ also. Then relative invariants in this case are called **covariants**. The **weight** of a covariant and the **absolute covariant** are defined as in the case of invariants. When we want to refer to n and d, we call the invariants (covariants) **invariants (covariants) of** n**-ary forms of degree** d: binary for $n=2$; ternary for $n=3$; linear forms for $d=1$; quadratic forms for $d=2$, etc.

For example, (1) if $n=2$, $d=2$, then the †discriminant $D=c_{02}c_{20}-c_{11}^2$ is an invariant of weight 2.

(2) Assume that $d=2$ (n arbitrary), and let u_{ij} be the coefficient of $\xi_i\xi_j$ in F, or, more precisely, $u_{ij}=u_{ji}=c_{\alpha_1\ldots\alpha_n}$, where (i) if $i=j$, then $\alpha_i=2$ and the other α_k are zero and (ii) if $i\neq j$, then $\alpha_i=\alpha_j=1$ and the other α_k are zero. In this case, $D=\det(u_{ij})$ is an invariant of weight 2.

(3) If $n=2$ and $d=4$, then $g_2=c_{40}c_{04}-4c_{13}c_{31}+3c_{22}^2$, $g_3=c_{04}c_{22}c_{40}-c_{04}c_{31}^2-c_{40}c_{13}^2+2c_{13}c_{22}c_{31}-c_{22}^3$ are invariants of respective weights 4, 6. The discriminant is expressed as $2^3(g_2^3-27g_3^2)$ and is an invariant of weight 12.

(4) For $d=1$, if we denote the coefficient of ξ_i in F by c_i, then $\sum_i c_i\xi_i$ is an absolute covariant and is also a vector invariant.

(5) For arbitrary n and d, $\det(\partial^2F/\partial\xi_i\partial\xi_j)$ is a convariant of weight 2 and is termed the **Hessian**.

Instead of one form F, we may take a finite number of homogeneous forms $F_1,\ldots,F_r$ of degree $d_1,\ldots,d_r$ such that the coefficients $c^{(i)}_{\alpha_1\ldots\alpha_n}$ are algebraically independent. Then we consider an action of $GL(n,K)$ on the polynomial ring $K[\ldots,c^{(i)}_{\alpha_1\ldots\alpha_n},\ldots][\xi_1,\ldots,\xi_n]$ given by $\varphi_{\alpha_i}:\sigma\rightarrow[\sigma]_{d_i}$ on the coefficients of F_i and by $\sigma\rightarrow{}^t\sigma^{-1}$ on ξ_j, as in the case of one form. Invariants in this case are called **covariants**, and covariants containing no ξ_i are called **invariants**. Weights, absolute covariants, and absolute invariants are defined similarly. The forms $F_1,\ldots,F_r$ are called **ground forms**, and the covariant is called a **covariant with ground forms** $F_1,\ldots,F_r$.

(6) If $r=n$, then the †Jacobian $\det(\partial F_i/\partial\xi_j)$ is a covariant of weight 1.

E. Multiple Covariants

Consider the situation described in Section C for the case of polynomials, and assume that K is a field of characteristic zero, $G=GL(n,K)$, and each ρ_i is either a φ_d (d arbitrary) or the contragredient map κ. Then these invariants are called **multiple covariants**, and **weights** and **absolute multiple covariants** are defined as before. Let s be the number of ρ_i equal to κ. Then the invariants and covariants of the preceding section correspond to the cases $s=0$ and $s=1$, respectively. Now assume that $\rho_i=\kappa$ if and only if $i=1,\ldots,s$. **Gram's theorem** states: for each $\alpha=1,\ldots,s$, let H_α be a polynomial in $x_j^{(i)}$ $(i>s)$ homogeneous in $x_1^{(i)},\ldots,x_{ni}^{(i)}$ for each i, and assume that the set V of common zeros of $H_1,\ldots,H_s$ (in the affine space of dimension $\sum_{i>s}n_i$) is G-stable . Then there exists a finite number of absolute multiple covariants $c_1,\ldots,c_t$ such that V is the set of $(\ldots,a_j^{(i)},\ldots)$ $(i>s)$ satisfying the condition that $(\ldots,a_m^{(\alpha)},\ldots,a_j^{(i)},\ldots,)$ is a zero point of $c_1,\ldots,c_t$ for any $a_m^{(\alpha)}$ with $\alpha\leqslant s$.

Since every rational action of $GL(n,K)$ is reductive, the set I of absolute multiple covariants (in a fixed polynomial ring over K) is a finitely generated ring over K. Furthermore, the set of multiple covariants of a given weight is a finitely generated I-module.

If we omit the assumption that K is of characteristic zero, it is difficult to define φ_d; this difficulty can be avoided by considering transformations of coefficients $a_{i_1\ldots i_n}$ of $F=\sum a_{i_1\ldots i_n}\xi_1^{i_1}\ldots\xi_n^{i_n}$. Although we can give similar definitions in that case, the theory does not

proceed similarly because, for example, rational actions of $GL(n, K)$ are not reductive.

F. Invariants of Lie Groups

Let G be a Lie group (hence K must be either the real number field or the complex number field) and ρ be a †differentiable representation of G such that $\rho(G) \subset GL(n, K)$, and assume that an action of G on $K[x_1, \ldots, x_n]$ is defined by ρ. Then, by means of an infinitesimal transformation X_a corresponding to G, an invariant (or semi-invariant) is characterized as an element f satisfying the condition $X_a f = 0$ $(\forall X_a)$ (or $X_a f = \alpha_a f$, $\alpha_a \in K$ $(\forall X_a)$). For instance, if $G = GL(2, K)$, then f is an invariant if $X_a f = 0$ for only two infinitesimal transformations that correspond to $\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}$, respectively. Similarly, for each Lie group G, there exists a finite number of X_a such that $X_a f = 0$ for these X_a characterizes f as an invariant.

G. History

In connection with geometry, the theory of invariants, especially that of binary forms, was first studied by A. Cayley (*J. Reine Angew. Math.*, 30 (1846)). The theory was further developed by J. J. Sylvester, R. F. A. Clebsch, P. Gordan, and others. Since the theory was originated with applications to projective spaces in mind, homogeneous semi-invariants were important; this is why semi-invariants were called invariants in the classical theory. On the other hand, in the theory of binary quadratic forms, invariants of discontinuous groups were studied from the viewpoint of the theory of numbers. It was D. Hilbert who introduced clearly the notion of invariants for general groups. He proved the existence of basic invariants in the classical case, making use of the †Hilbert basis theorem. Hilbert's 14th problem (→ 196 Hilbert) is related to this result, but its answer is negative; that is, even in the case of a polynomial ring over the real number field or the complex number field, there are groups acting on the ring without basic invariants (M. Nagata). Though the theory of invariants has not been studied effectively for a long time, it is again under active study because of its importance in algebraic geometry.

References

[1] R. Weitzenböck, Invariantentheorie, Noordhoff, 1923.
[2] I. Schur and H. Grunsky, Vorlesungen über Invariantentheorie, Springer, 1968.
[3] H. Weyl, Classical groups, Princeton Univ. Press, revised edition, 1946.
[4] M. Nagata and T. Miyata, Remarks on matric groups, J. Math. Kyoto Univ., 4 (1965), 381–384.
[5] D. Mumford, Geometric invariant theory, Springer, 1965.
[6] J. Fogarty, Invariant theory, Benjamin, 1969.
[7] J. A. Dieudonné and J. B. Carrell, Invariant theory, old and new, Advances in Math., 4 (1970), 1–80.
[8] D. Hilbert, Gesammelte Abhandlungen, Band II, Springer, 1933, 1970.

227 (XIX.14) Inventory Control

Although inventory control is discussed mainly in connection with production control and operations research, inventory models have applications in many other fields. For example, an inventory (or queuing) model can be used to plan the optimal operation of a dam, and may thereby influence the engineering design of the dam.

A mathematical theory of **inventory control** must be based upon an inventory structure in which the following items are considered: (i) costs and revenue, (ii) demand, and (iii) deliveries. Item (i) involves (1) order and production costs, (2) storage costs, (3) discount rates, (4) penalties, (5) revenues (under the assumption that price and demand are not controlled by the firm), (6) costs of changing the production rate, and so on. For item (ii), there are several different situations due to the combination of predictability and stability conditions of demand for the commodity. Current theories are concerned with two particular situations. In the first, the †probability distribution of demand is known to the firm. In the second, the date of occurrence of orders is a †random variable with a known probability distribution, while the amount of demand is a known constant. A †minimax principle is applied when some of these probability distribution functions are unknown to the firm.

Problems in (iii) are due to delays (lead time periods) after an order for inventory is placed or a decision is made to produce for an uncertain demand.

The following inventory system has been studied mathematically in much of the literature. Let x_n be the initial storage level for the nth period, and ξ_n and z_n be the demand and

the amount of order, respectively, for the nth period. Then $y_n = x_n + z_n$ is the storage level after the arrival of ordered items, and $x_{n+1} = \max(0, y_n - \xi_n)$ is the initial storage at the $(n+1)$th period. Furthermore, in this period, the stock shortage $\eta_n = \xi_n - y_n$ is incurred if $y_n < \xi_n$. When the demand ξ_n is known, the inventory model is a simple deterministic one. When ξ_n is a random variable, let its density function be $\varphi(\xi)$. For constructing the model, the following three cost functions are introduced: The ordering cost $c(z)$ of order z, the holding cost $h(y)$ of storage y for one period, and the penalty cost $p(\eta)$ for a shortage η. Then the total cost for the period of storage level y after reorder is

$$I(y)=\int_0^y h(y-\xi)\varphi(\xi)d\xi+\int_y^\infty p(\xi-y)\varphi(\xi)d\xi.$$

Let x be the initial storage for the first period, and let $f_n(x)$ represent the minimal total cost of an n-period inventory model; then $f_n(x)$ satisfies the following fundamental functional equation with a fixed discount rate α $(0<\alpha<1)$:

$$f_n(x)=\min\left[\min_{y>x}\left\{c(y-x)+I(y)+\alpha\int_0^\infty f_{n-1}(y-\xi)\varphi(\xi)d\xi\right\},\ I(x)+\alpha\int_0^\infty f_{n-1}(x-\xi)\varphi(\xi)d\xi\right].$$

In this equation, the first term of the right-hand side gives the amount of order at the first period as $y_n(x)-x$, where $y_n(x)$ is a minimizing value of y, and the second term gives the cost if no order is put in. Policies of the (s, S) type are defined as follows: Order $S-x$ if $x<s$, and do not order if $x \geqslant s$. For the special case when the ordering function is given by $c(z)=C+cz$ $(z>0)$; $=0$ $(z=0)$, some sufficient conditions for the optimal policy to be of (s, S) type have been studied.

For various inventory models where the demand and lead time are random variables, the optimal policy can be derived by using queuing theory. We consider an inventory model with maximum inventory M. The queuing situation corresponding to this model is an M-channel queuing system. "Empty shelf space due to demand" corresponds to the arrival of a customer, the ordering instance corresponds to the commencement of service, and the arrival of an ordered item corresponds to the completion of service. Thus the stationary probability of storage level n, P_n, can be obtained by solving the equilibrium equation of the queuing system; then mean inventory, mean sales, and mean ordering number can be derived. This above ordering system is called "reorder for each item sold"; more general systems have been treated in a similar manner.

Another example of the queuing system approach may be illustrated by way of the planning problem for dams mentioned at the beginning of this article: The probability distribution of the amount of water stored in the dam is derived by analyzing the functional equation of a queuing system. Let the capacity of the dam be M, the storage of the dam at the nth period be x_n, and the outflow after inflow z_n be ξ_{n+1}; then the storage at each period is given by

$$x_1=x, \qquad 0<x<M,$$

$$x_{n+1}=[M-\xi_{n+1}+[x_n+z_n-M]^-]^+,$$

where $a^+=\max(0,a)$ and $a^-=\min(0,a)$.

The equations for queuing models will be the same as those given above for the inventory model if we let x_n be the waiting time of the nth customer, z_n the service time, ξ_{n+1} the interarrival time of customers n and the $n+1$, and M the service time.

References

[1] K. J. Arrow, S. Karlin, and H. E. Scarf, Studies in the mathematical theory of inventory and production, Stanford Univ. Press, 1958.
[2] H. E. Scarf, D. M. Gilford, and M. W. Shelly (eds.), Multistage inventory models and techniques, Stanford Univ. Press, 1963.
[3] T. M. Whitin, The theory of inventory management, Princeton Univ. Press, 1953.
[4] G. Hadley and T. M. Whitin, Analysis of inventory systems, Prentice-Hall, 1963.
[5] P. M. Morse, Queues, inventories and maintenance, Wiley, 1958.
[6] J. W. Cohen, Single server queue with uniformly bounded virtual waiting time, J. Appl. Probability, 5 (1968), 93–122.

228 (X.34) Isoperimetric Problems

A. The Classical Isoperimetric Problem

Two curves are called **isoperimetric** if their perimeters are equal. The term *curve* is used here to mean a †Jordan curve. The classical **isoperimetric problem** is to find, among all curves J with a given perimeter L, the curve enclosing the maximum area. This problem is also called the **special isoperimetric problem** or **Dido's problem**. Its solution is a circle. The analogous problem in 3-dimensional space has

a sphere as its solution; that is, among all closed surfaces with a given surface area, the sphere encloses the maximum volume. The following variational problem can be regarded as a generalization of the classical isoperimetric problem: To find the curve $C: y=f(x)$ that gives the maximum value of the functional $\int_C F(x,y,y')dx$ under the subsidiary condition $\int_C G(x,y,y')dx=$constant. This is sometimes called the **generalized isoperimetric problem**. The classical isoperimetric problem can be solved by variational methods. It can also be solved by using inequalities involving quantities associated with the figure in question. For example, the inequality

$$L^2-4\pi F\geqslant 0 \tag{1}$$

between the area F and the perimeter L of a Jordan curve J solves the problem, since the equality holds only for a circle. For refinements of (1) there are further inequalities due to T. Bonnesen (1921):

$$L^2-4\pi F\geqslant (L-2\pi r)^2,$$

$$L^2-4\pi F\geqslant (2\pi R-L)^2,$$

$$L^2-4\pi F\geqslant \pi^2(R-r)^2,$$

where r is the radius of the largest circle inscribed in the curve J and R the radius of the smallest circumscribed circle. These inequalities can also be used to solve the isoperimetric problem. Moreover, we have the following inequality for curves on the sphere of radius a:

$$L^2-4\pi F+F^2a^{-2}\geqslant 8\pi a^2\sin\frac{R-r}{4a(1+2\pi)}$$

(F. Bernstein, 1905). For curves on the surface of negative constant curvature $-1/a^2$, we have

$$L^2-4\pi F-F^2a^{-2}$$

$$\geqslant\frac{1}{4}a^2(4\pi+Fa^{-2})\left(\tanh\frac{R}{2a}-\tanh\frac{r}{2a}\right)^2$$

(L. A. Santaló [2]). From these inequalities, we see that the circle remains the solution to the isoperimetric problem in each of the non-Euclidean planes.

The corresponding problem in the 3-dimensional case is more difficult. Without going into detail, we cite the following example: For an †ovaloid with surface area S and volume V,

$$S^3-36\pi V^2\geqslant 0,$$

where the equality holds only for the sphere.

B. Isoperimetric Inequalities on Eigenvalues

In recent years the concept of **isoperimetric inequality** has been extended to include all inequalities connecting two or more geometric or physical quantities depending on the shape and size of a figure. For example, it includes inequalities on †eigenvalues of partial differential equations under given boundary conditions.

Lord Rayleigh conjectured in 1877 that, for the equation $\Delta u+\lambda u=0$ for a vibrating membrane on a region D of fixed area F (with $u=0$ on the boundary), the first eigenvalue λ_1 is least when D is a circle. This conjecture is true. In fact, in 1923 G. Faber and E. Krahn proved independently that

$$\lambda_1\geqslant(\pi/F)j^2 \tag{2}$$

and that the equality in (2) holds if and only if the domain D is a circle, where $j=2.4048\ldots$ is the first positive zero of the †Bessel function $J_0(x)$. For the second eigenvalue λ_2 of the same problem, the circle does not give the minimum value. I. Hong (1954) gave the inequality

$$\lambda_2\geqslant(2\pi/F)j^2$$

and showed that λ_2 approaches its greatest lower bound, $(2\pi/F)j^2$, as the shape of the domain approaches a figure consisting of two equal mutually tangential circles, each having area $F/2$ [4].

Many other results were found with regard to isoperimetric inequalities on eigenvalues of partial differential equations, for example, results related to eigenvalues for a membrane under other boundary conditions, such as $\partial u/\partial n=0$, and for other types of equations, such as $\Delta\Delta u-\lambda u=0$.

A method devised by J. Steiner and called **symmetrization** is a powerful means of discovering isoperimetric inequalities. **Steiner's symmetrization** with respect to the line l changes the plane domain P into another plane domain Q characterized as follows: Q is symmetric with respect to l, any straight line g perpendicular to l that intersects one of the domains P or Q also intersects the other, both intersections have the same length, and the intersection with Q is a segment bisected by l. This operation can be extended to a space of higher dimension by replacing l with a hyperplane. Steiner's symmetrization preserves area (or volume) and diminishes perimeter (or surface area). Steiner first used these properties to solve the classical isoperimetric problem in 1838. In 1945, G. Pólya and G. Szegö found that the electrostatic capacity of a solid is diminished by Steiner's symmetrization. The concepts developed in their papers made possible a systematic treatment of many isoperimetric inequalities and estimations of mathematical and physical quantities.

References

For the classical isoperimetric problem,
[1] W. Blaschke, Kreis und Kugel, Veit, 1916 (Chelsea, 1949).
[2] L. A. Santaló, La desigualdad isoperimetrica sobre superificies de curvatura constante negative, Rev. Univ. Tucumán, (A) 3 (1942), 243–259.
For isoperimetric inequalities on eigenvalues,
[3] G. Pólya and G. Szegö, Isoperimetric inequalities in mathematical physics, Princeton Univ. Press, 1951.
[4] I. Hong, On an inequality concerning the eigenvalue problem of a membrane, Kôdai Math. Sem. Rep. (1954), 113–114.

J

229 (XXI.29) Jacobi, Carl Gustav Jacob

Carl Gustav Jacob Jacobi (December 10, 1804–February 18, 1851) was born into a wealthy banking family in Potsdam, Germany. He was well educated at home and was highly cultured in many areas. He entered the University of Berlin, studied mathematics largely on his own through Euler's texts, and obtained the doctorate in 1825. The following year he became a private lecturer at the University of Königsberg, and in 1831 a professor. For the next seventeen years he worked vigorously in Königsberg, where his influence was considerable. Toward the end of his life, his health failed; moreover, he lost his property and met with general misfortune because of the political situation of the time. He made no further contributions after 1843; he died of smallpox at 47.

Because he had an intense personality, there were times when he invited the animosity of people; however, he did have early contact with †Abel, and in his later years he enjoyed the friendship of †Dirichlet. Jacobi's mathematical works lacked formal completeness but were very original and contributed to many fields. The †Hamilton-Jacobi equation in dynamics and the †Jacobian determinant of a differentiable mapping are well-known products of his ideas, but even more noteworthy are his contributions to the theory of elliptic functions and algebraic functions, particularly the introduction of theta functions and the treatment of the inverse problem of hyperelliptic integrals [2].

References

[1] C. W. Borchardt and K. Weierstrass (eds.), K. G. J. Jacobi's gesammelte Werke I–VII, Reimer, 1881–1891 (Chelsea, second edition, 1969).
[2] G. J. Jacobi, Fundamenta nova theoriae functionum ellipticarum, Bornträger, 1829.
[3] F. Klein, Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert I, II, Springer, 1926, 1927 (Chelsea, 1956).

230 (XXI.7) Japanese Mathematics (Wasan)

Before the introduction of Western mathematics, indigenous Japanese mathematics developed along its own characteristic lines. In Japanese, this form of mathematics is called **wasan**. The first development took place in the 8th century A.D. during the Nara era under the influence of the Tang dynasty in China. After a period of decline came another period of development from the 13th to the 17th century. During this second wave, Chinese mathematical books such as *Suanhsueh chimeng* and *Suanfa tangtsung* were imported, along with the abacus, or *soroban* as it is called in Japanese, and calculating rods, *sangi* in Japanese, with which Chinese mathematicians performed algebraic operations. Japanese mathematicians absorbed these methods and invented and developed their own written algebra, called *endan-jutsu* or *tenzan*.

The fundamental concepts of *wasan* are attributed to Seki Takakazu, or Seki Kowa (Seki is the surname; likewise for other Japanese names in this article), Takebe Katahiro, and Kurushima Yoshihiro. Its main developments stem from Ajima Naonobu and Wada Yasushi (or Wada Nei), among others. *Wasan* scholars obtained interesting and significant results, but they pursued mathematics as an art in the Japanese manner rather than as a science in the Western sense. *Wasan* had no philosophical background as did the Greek tradition, nor did it have an intimate relation with the natural sciences. Thus it lacked the character of systematized science and dissolved after the introduction of Western mathematics into the school system by the Meiji government (1867–1912).

Among *wasan* works of the earlier period, *Jinkôki* by Yoshida Mitsuyoshi (1598–1672), the first edition of which appeared in 1627, contributed much to popularize the *soroban* and to arouse general interest in mathematics.

Seki Takakazu (1642?–1708) was born in Fujioka in the Gunma prefecture. Some historians say he learned mathematics from Takahara Yoshitane, while others say he was completely self-taught. His achievements include the following: (1) the invention of *endan-jutsu*, or written algebra; (2) the discovery of determinants; (3) the solution of numerical equations by a method similar to †Horner's; (4) the invention of an iteration method to solve equations, similar to Newton's; (5) the introduction of derivatives and of discriminants of polynomials; (6) the discovery of conditions for the existence of positive and of negative roots of polynomials; (7) a method of finding maxima and minima; (8) the transformation theory of algebraic equations; (9) continued fractions; (10) the solution of some Diophantine equations; (11) the introduction of †Bernoulli numbers; (12) the study of regular polygons; (13) the calculation of π and the volume of the

sphere; (14) †Newton's interpolation formula; (15) some properties of ellipses; (16) the study of the †spirals of Archimedes; (17) the discovery of the Pappus-Guldin theorem; (18) the study of magic squares; and (19) the theoretical study of some questions arising out of mathematical recreations (called *mamako-date*, *metsuke-ji*, etc.).

Takebe Takahiro (1664–1739) was a disciple of Seki. He is the author of the book *Enri tetsujutsu*. (*Enri*, or circle theory, was one of the favorite subjects of *wasan* scholars.) It contains the formula

$$\left(\frac{1}{2}\arcsin x\right)^2=\frac{x^2}{2}+\frac{2^2\cdot x^4}{4}+\frac{2^2\cdot 4^2\cdot x^6}{6}+\ldots.$$

He also obtained other formulas of trigonometry and some approximation formulas, by means of which he compiled trigonometric tables to 11 decimal places. In collaboration with Seki he wrote the 20 volumes of *Taisei sankei* and *Fukyû tetsujutsu*, the latter elucidating the methodology of the Seki school. It contains a value of π to 42 decimal places.

Kurushima Yoshihiro (?–1757) was an original scholar influenced by Nakane Genkei (1662–1733), a disciple of Takebe. He generalized an approximation formula for sines obtained by Takebe, treated problems of maxima and minima involving trigonometric functions, improved the theory of determinants and the theory of equations, obtained a formula for $S_p=1^p+\ldots+n^p$ without using Bernoulli numbers, and found a relation among a, b, c, n by eliminating x from

$$x+(x+c)+\ldots+(x+(n-1)c)=a,$$

$$x^k+(x+c)^k+\ldots+(x+(n-1)c)^k=b.$$

Furthermore, he studied †Euler's function $\varphi(n)$ before Euler and obtained the †Laplace expansion theorem for determinants before Laplace. He is said to have contributed to *Hôen sankei*, an important work on *enri* written by Matsunaga Yoshisuke (c. 1694–1744) in which a value of π is given up to the 50th decimal place. The Seki school, continued under Takebe, Nakane, Kurushima, and Matsunaga, became a center of *wasan*. Scholars of this school lived mainly in Edo (the ancient name for Tokyo). Yamaji Nushizumi (or Shujû) (1704–1772) studied *wasan* with Nakane, Kurushima, and Matsunaga. Arima Yoriyuki (1712–83), one of his disciples, for the first time made public the teachings of this school in the book *Shûki sampô*. The practice of dedicating to Shinto shrines or Buddhist temples tablets engraved with solved mathematical problems became popular during this period.

Ajima Naonobu (1739–98) was a disciple of Yamaji. He improved *enri*, simplified its theory, and amplified its applications. He treated problems of finding volumes involving double integrals, discovered the binomial theorem for exponent $1/n$, compiled a table of logarithms to 14 decimal places, and treated Diophantine problems. No trace of demonstrative geometry is found in *wasan*, but Ajima and his school did treat geometric problems, such as †Malfatti's problem, dealing with several circles tangent to each other.

Wada Yasushi (or Wada Nei, 1787–1840) studied *wasan* with Kusaka Makoto (1764–1839), a disciple of Ajima. He made tables containing more than 100 definite integrals, including, for example,

$$\int_0^1 x^p(1-x)^q\,dx,\qquad \int_0^1 x^p(1-x^2)^q\,dx.$$

However, there is no evidence that *wasan* scholars, even in this period, knew of the fundamental theorem of infinitesimal calculus.

Apart from the Seki school, there were Tanaka Yoshizane (1651–1719), a contemporary of Seki, and his disciple Iseki Tomotoki (c. 1690). Tanaka is said to have ranked with Seki in his work on determinants and magic squares, but most of his writings have been lost. Iseki wrote a text on determinants called *Sanpô hakki* (1690), the first of its kind in the history of mathematics. A little later, Takuma-ryû (the Takuma school) was formed in Osaka. Inô Tadataka (1745–1821), famous for making the first precise map of Japan, had studied *wasan* with Takahashi Yoshitoki (1764–1804), who belonged to this school. Aida Yasuaki (1747–1817), a contemporary of Ajima, founded *Saijô-ryû*, or the "superlative school," and rivaled Fujita Sadasuke (1734–1807) of the Seki school.

Toward the end of the Tokugawa era, the study of geometric problems became popular among *wasan* scholars, such as Hasegawa Kan (1782–1383), Uchida Gokan (1805–82), and his disciple Hôdôji Zen (1820–68), who used the method of †inversion. Hasegawa wrote *Sanpô sinsho*, a popular work containing an explanation of the methods of *enri*.

The influence of Western mathematics is hardly recognizable outside astronomy, calendar making, and the compilation of logarithmic tables. However, some *wasan* scholars took a more positive attitude and began studying Western mathematics toward the close of the Tokugawa period, thus helping to lay the foundations for the development of mathematics in Japan in the new era.

Following the Meiji restoration, Kikuchi Dairoku (1855–1917), Hayashi Tsuruichi (1873–1935), Fujiwara Matsusaburo (1861–1933), and recent scholars contributed much to preserve *wasan* literature and to clarify its

content, but their undertaking has not yet been completed.

References

[1] M. Fujiwara, History of mathematics in Japan before the Meiji era (in Japanese), I–V, under the auspices of the Japan Academy, Iwanami, 1954–1960.
[2] Y. Mikami, The development of mathematics in China and Japan, Teubner, 1913 (Chelsea, 1961).
[3] A. Hirayama, K. Shimodaira, and H. Hirose (eds.), Takakazu Seki's collected works, English translation by J. Sudo, Osaka Kyoiku Tosho, 1974.

231 (III.22) Jordan Algebras

A. Definitions

Let A be a †linear space over a field K. If there is given a †bilinear mapping (multiplication) $A \times A \to A$, the space A is called a **distributive algebra** (or **nonassociative algebra**) over K. In particular, if this multiplication satisfies the associative law, A is called an **associative algebra** over K (→ 29 Associative Algebras). We assume that A is a distributive algebra over K of finite †dimension over K. Denote by $E(A)$ the associative algebra of all K-†endomorphisms of the K-linear space A. With every $a \in A$ we associate $R_a \in E(A)$, $L_a \in E(A)$ by $R_a(x) = xa$, $L_a(x) = ax$, where x belongs to A. The †subalgebra of $E(A)$ generated by the R_a, L_a $(a \in A)$ and the †identity mapping of A is called the **enveloping algebra** of A. The left, right, and two-sided †ideals of A are defined as in the case of associative algebras. An element c of A is said to be in the **center** of A if (i) $ac = ca$ and (ii) $a(bc) = (ab)c$, $a(cb) = (ac)b$, and $c(ab) = (ca)b$ for every a, b in A. We denote the product of two elements a, b in A by $a \cdot b$ in order to distinguish it from multiplication in the case of associative algebras. Denote $a \cdot a$ by $a^{\cdot 2}$ and put $A^{\cdot 2} = \{a_1 \cdot a_2 \mid a_1, a_2 \in A\}$. Define $A^{(n)}$ successively by $A^{(0)} = A$, $A^{(1)} = A^{\cdot 2}, \ldots, A^{(k+1)} = (A^{(k)})^{\cdot 2}$. Then A is called a **solvable algebra** if $A^{(n)} = 0$ for some n and a **nilalgebra** if every element of A is †nilpotent. A distributive algebra A is called a **Jordan algebra** if the following two conditions are satisfied for every a, u in A: (i) $a \cdot u = u \cdot a$ and (ii) $a^{\cdot 2} \cdot (u \cdot a) = (a^{\cdot 2} \cdot u) \cdot a$. We omit discussion of noncommutative Jordan algebras in this article. A distributive algebra A is called an **alternative algebra** if the following two conditions are satisfied for every a, u in A: (i) $a \cdot u^{\cdot 2} = (a \cdot u) \cdot u$ and (ii) $u^{\cdot 2} \cdot a = u \cdot (u \cdot a)$. An alternative algebra A is called an **alternative field** if $L_a(A) = A = R_a(A)$ for every a in A such that $a \neq 0$. Generalizing these algebras we obtain the notion of a **power associative algebra**. A distributive algebra A is a power associative algebra if every element of A generates an associative subalgebra. Jordan algebras and alternative algebras are power associative.

We now consider an algebra A over a field K. We assume that the †characteristic of K is not 2 and that A is of finite dimension over K. If A and B are associative algebras, a linear mapping $\sigma: A \to B$ is called a **Jordan homomorphism** if (i) $(a^2)^\sigma = (a^\sigma)^2$ for every a in A and (ii) $(aba)^\sigma = a^\sigma b^\sigma a^\sigma$ for every a, b in A (where we denote by a^σ the image of $a \in A$ by the mapping σ). If B does not contain †zero divisors, then $(ab)^\sigma = a^\sigma b^\sigma$ or $(ab)^\sigma = b^\sigma a^\sigma$.

Let A be an associative algebra. Define a new multiplication $\cdot$ in A by $a \cdot b = (ab + ba)/2$. We then have a Jordan algebra A^+. A subalgebra of the Jordan algebra A^+ is called a **special Jordan algebra**. Let $K[x_1, \ldots, x_n]$ be the noncommutative free ring in the indeterminates $x_1, \ldots, x_n$ (that is, $K[x_1, \ldots, x_n]$ is the associative algebra over K that has as its K-bases the free semigroup with identity element 1 over the free generators $x_1, \ldots, x_n$). The subalgebra $K[x_1, \ldots, x_n]^+$ generated by 1 and the x_i is called the **free special Jordan algebra** of n generators and is denoted by $J_0^{(n)}$. A Jordan algebra A is **special** if and only if there is an isomorphism of A into B^+, where B is some associative algebra. A Jordan algebra that is not special is called **exceptional**. All homomorphic images of $J_0^{(2)}$ are special. Denote by $\mathscr{R}$ the ideal of $J_0^{(3)}$ generated by $x^{\cdot 2} - y^{\cdot 2}$ (note that $J_0^{(3)} \subsetneq K[x, y, z]$). Then $J_0^{(3)}/\mathscr{R}$ is exceptional. A Jordan algebra is special if it contains the unity element and is generated by two elements. If A is an alternative algebra, the associated Jordan algebra A^+ is special.

Condition (ii) for a Jordan algebra A is equivalent to $[R_a, R_{a^{\cdot 2}}] = 0$. In A, we have $[R_a, R_{b \cdot c}] + [R_b, R_{c \cdot a}] + [R_c, R_{a \cdot b}] = 0$; and $[[R_c, R_a], R_b] = R_{[a,b,c]}$. Here we put $[S, T] = ST - TS$, $[a, b, c] = (a \cdot b) \cdot c - a \cdot (b \cdot c)$. Such an equation in A is called an **identity** in a Jordan algebra.

B. Structure of Jordan Algebras

A Jordan algebra A has a unique largest solvable ideal N, which contains all nilpotent ideals of A and is called the **radical** of A. If $N = 0$, A is called **semisimple**. The quotient A/N is always semisimple. A semisimple Jordan

algebra A contains the unity element and can be decomposed into a direct sum $A = A_1 \oplus \dots \oplus A_r$ of minimal ideals A_i. Each A_i is a †simple algebra. In particular, if K is of characteristic 0, there is a semisimple subalgebra S of A such that $A = S \oplus N$. Let e be an idempotent element of A, and let $\lambda \in K$. Put $A_e(\lambda) = \{x \mid x \in A, e \cdot x = \lambda x\}$. Then we have

$$A = A_e(1) \oplus A_e(1/2) \oplus A_e(0).$$

This decomposition of A is called the **Peirce decomposition** of A relative to e. Suppose that the unity element 1 is expressed as a sum of the mutually orthogonal idempotents e_i. Then, putting $A_{i,i} = A_{e_i}(1)$, $A_{i,j} = A_{e_i}(1/2) \cap A_{e_j}(1/2)$, we have $A = \sum_{i \leqslant j} \oplus A_{i,j}$. The $A_{i,j}$ are called **Peirce spaces**. Furthermore, suppose that for every i, $A_{i,i}$ is of the form $A_{i,i} = K \cdot e_i + N_i$, with N_i a nilpotent ideal of $A_{i,i}$. In this case, A is called a **reduced algebra** and the number of the e_i is called the **degree** of A.

Let D be an alternative algebra with the unity element 1 and with an †involution $^-$. Furthermore, suppose that there is a †quadratic form $Q(X)$ such that $x \cdot \bar{x} = \bar{x} \cdot x = Q(x) \cdot 1$ $(x \in D)$ and that $f(X, Y) = (Q(X+Y) - Q(X) - Q(Y))/2$ is a †nondegenerate bilinear form. Then D is called a **composition algebra**.

Reduced simple Jordan algebras A are classified into three types:

(1) $A = K \cdot 1$.

(2) There exist idempotents e_1, e_2 such that $A = K \cdot 1 \oplus K \cdot (e_1 - e_2) \oplus A_{1,2}$, where $A_{1,2} \neq 0$; and furthermore, the multiplication of A is determined as follows: Let $x = \alpha(e_1 - e_2) + a_{1,2}$, $y = \beta(e_1 - e_2) + b_{1,2}$ be in A with $\alpha, \beta \in K$, $a_{1,2}, b_{1,2} \in A_{1,2}$. Then $x \cdot y = (\alpha\beta + f(a_{1,2}, b_{1,2})) \cdot 1$, where f is a nondegenerate symmetric bilinear form.

(3) Let D be a composition algebra with the involution $^-$, and let D_n be the †total matrix algebra over D of degree n. Let $r = \operatorname{diag}\{r_1, \dots, r_n\}$ $(r_i \neq 0)$ be given in D_n. Define $J: D_n \to D_n$ by $a^J = r \cdot \bar{a} \cdot r^{-1}$. Then A is of the form $A = \{x \in D_n \mid x = x^J\}$.

Let A be a special, reduced simple Jordan algebra such that $A_{i,i} = Ke_i$. If A is of degree 2, then A is a †Clifford algebra. If A is of degree $\geqslant 3$, then A is classified into five types.

Let A be a simple Jordan algebra over a field K. Then there is an extension field P of K of finite degree such that $A_P (= A \otimes_K P)$ is isomorphic to one of the following five types: (i) P_n^+, where P_n is the total matrix algebra of degree n over P; (ii) the subalgebra of P_n^+ consisting of all symmetric matrices in P_n^+; (iii) $\{x \in P_{2m}^+ \mid x^J = x\}$, where

$$x^J = q^{-1}\,{}^t x q, \qquad q = \begin{pmatrix} 0 & I_m \\ -I_m & 0 \end{pmatrix},$$

with ${}^t x$ the transpose of x and I_m the unit matrix of degree m; (iv) an algebra generated by the generators $s_0, s_1, \dots, s_n$ together with the defining relations $s_0 \cdot s_i = s_i$, $s_i^2 = s_0$, $s_i \cdot s_j = 0$ $(i \neq j)$; (v) $\mathscr{H}_3^+$, where $\mathscr{H}_3$ is the algebra of all 3×3 Hermitian matrices over a †Cayley algebra.

C. Representations of Jordan Algebras

A **representation** S of a Jordan algebra A on a K-linear space M is a K-linear mapping $a \to S_a$ from A into the associative algebra $E(M)$ of all K-endomorphisms of M such that (i) $[S_a, S_{b \cdot c}] + [S_b, S_{c \cdot a}] + [S_c, S_{a \cdot b}] = 0$ and (ii) $S_a S_b S_c + S_c S_b S_a + S_{(a \cdot c) \cdot b} = S_a S_{b \cdot c} + S_b S_{a \cdot c} + S_c S_{a \cdot b}$ (for all a, b, c in A). A K-linear space M is called a **Jordan module** of A if there are given bilinear mappings $M \times A \to M$ (denoted by $(x, a) \to x \cdot a$), $A \times M \to M$ (denoted by $(a, x) \to a \cdot x$) such that for every $x \in M$ and every $a, b, c \in A$, (i) $x \cdot a = a \cdot x$; (ii) $(x \cdot a) \cdot (b \cdot c) + (x \cdot b) \cdot (a \cdot c) + (x \cdot c) \cdot (a \cdot b) = (x \cdot (b \cdot c)) \cdot a + (x \cdot (a \cdot c)) \cdot b + (x \cdot (a \cdot b)) \cdot c$; and (iii) $x \cdot a \cdot b \cdot c + x \cdot c \cdot b \cdot a + a \cdot c \cdot b \cdot x = (x \cdot c) \cdot (a \cdot b) + (x \cdot a) \cdot (b \cdot c) + (x \cdot b) \cdot (a \cdot c)$. As usual, there is a natural bijection between the representations of A and the Jordan modules of A. A **special representation** of a Jordan algebra A is a homomorphism $A \to E^+$, where E is an associative algebra. Among the special representations of A, there exists a unique universal one in the following sense. There exists a special representation $S: A \to U^+$ with the following property: For every special representation $\sigma: A \to E^+$ there exists a unique homomorphism $\eta: U \to E$ such that $\sigma = \eta S$. The pair (U, S) is uniquely determined. Furthermore, if A is n-dimensional over K, U is of dimension $\leqslant 2^n - 1$ over K. The pair (U, S) is called the **special universal enveloping algebra** of A. A is special if and only if $S: A \to U^+$ is injective.

A Jordan algebra A has only a finite number of inequivalent †irreducible Jordan modules. Suppose that the base field K is of characteristic 0. Let S be a representation of A. If N is the radical of A, $S(N)$ is contained in the radical of the associative algebra $S(A)^*$ generated by $S(A)$. If A is semisimple, so is $S(A)^*$. In general, the radical R of $S(A)^*$ is an ideal of $S(A)^*$ generated by $S(N)$. Furthermore, for every semisimple subalgebra T of A such that $A = T \oplus N$, we have $S(A)^* = S(T)^* \oplus R$.

References

[1] A. A. Albert, Non-associative algebras I, II, Ann. Math., (2) 43 (1942), 685–707.
[2] A. A. Albert, A structure theory for Jordan algebras, Ann. Math., (2) 48 (1947), 546–567.

[3] N. Jacobson, General representation theory of Jordan algebras, Trans. Amer. Math. Soc., 70 (1951), 509–530.
[4] F. D. Jacobson and N. Jacobson, Classification and representation of semi-simple Jordan algebras, Trans. Amer. Math. Soc., 65 (1949), 141–169.
[5] N. Jacobson and C. E. Rickart, Jordan homomorphisms of rings, Trans. Amer. Math. Soc., 69 (1950), 479–502.
[6] N. Jacobson, Structure and representations of Jordan algebras, Amer. Math. Soc. Colloq. Publ., 1968.
[7] H. Braun and M. Koecher, Jordan-Algebren, Springer, 1966.

K

232 (VII.10) Kähler Manifolds

A. Definitions

Let X be a †complex manifold of complex dimension n. Let J denote the †tensor field of type (1, 1) of †almost complex structure induced by the complex structure of X (→ 72 Complex Manifolds A). Considering J as a linear transformation of vector fields on X, we have $J^2 = -1$. A †Riemannian metric g of class C^∞ on X is called a **Hermitian metric** on X if $g(x, y) = g(Jx, Jy)$ for any two vector fields x, y on X. On a †paracompact complex manifold X there always exists a Hermitian metric. If we put $\Omega(x, y) = g(Jx, y)$ for a Hermitian metric g, then Ω is an †alternating covariant tensor field of order 2, and hence a †differential form of degree 2. We call Ω the **fundamental form** associated with the Hermitian metric g. If the †exterior derivative of Ω vanishes, i.e., $d\Omega = 0$, the given Hermitian metric g on X is called a **Kähler metric** on X, and X is called a **Kähler manifold**. With a †holomorphic local coordinate system $(z^1, \ldots, z^n)$ on a coordinate neighborhood U in X, we can write $g = \sum_{\alpha,\beta=1}^{n} g_{\alpha\bar\beta}\, dz^\alpha d\bar z^\beta$ in U, where the matrix $(g_{\alpha\bar\beta})$ is an $n \times n$ †positive definite Hermitian matrix. Then the fundamental form can be expressed as $\Omega = (\sqrt{-1}/2) \sum g_{\alpha\bar\beta}\, dz^\alpha \wedge d\bar z^\beta$.

If a complex manifold X is a Kähler manifold with a Kähler metric g, then X has the following properties: (1) For every point p of X there exists a real-valued function ψ of class C^∞ on a suitable coordinate neighborhood of p such that $g_{\alpha\bar\beta} = \partial^2\psi/\partial z^\alpha \partial\bar z^\beta$. (2) For every point p there exists a holomorphic local coordinate system at p whose real and imaginary parts form a †geodesic coordinate system at p in the weak sense. (We say a coordinate system $(x_1, \ldots, x_n)$ is a **geodesic coordinate system at** p **in the weak sense** if $[\nabla_{\partial/\partial x^i}(\partial/\partial x^j)]_p = 0$, $i, j = 1, \ldots, n$ (→ 417 Tensor Calculus).) Each of these properties characterizes a Kähler metric. The Kähler metric was introduced by E. Kähler with property (1) as the definition. W. V. D. Hodge applied it to the theory of harmonic integrals [7].

B. Harmonic Forms on Compact Kähler Manifolds (→ 194 Harmonic Integrals)

We now consider complex-valued differential forms on a compact complex manifold X of complex dimension n endowed with a Hermitian metric. The operators d and $*$ on real-valued differential forms can be uniquely extended to complex linear operators, and the inner product of real differential forms can be extended to a Hermitian inner product of the form $(\varphi, \psi) = \int_X \varphi \wedge * \overline{\psi}$. Then the †adjoint operator δ of d is also complex linear, and the decomposition $d = d' + d''$ gives rise to $\delta = \delta' + \delta''$, where δ' and δ'' are the adjoint operators of d' and d'', respectively. We also have $\delta' = (-1) * d'' *$ and $\delta'' = (-1) * d' *$. Define the operator L by $L\varphi = \Omega \wedge \varphi$, and denote by Λ the adjoint operator of L. Then for p-forms, we have $\Lambda = (-1)^p * L *$. We say that φ is **primitive** when $\Lambda\varphi = 0$. Some important properties of L and Λ follow: (3) For a p-form φ, $\Lambda\varphi^p = 0$ if and only if $L^q\varphi = 0$ ($q = \max(n-p+1, 0)$). (4) If $\Lambda\varphi = 0$ and φ is of type (r, s), then for all $0 \leqslant q \leqslant n-p$,

$$* L^q \varphi = \{(-1)^{p(p+1)/2} \times (\sqrt{-1})^{r-s} q!/(n-p-q)!\} L^{n-p-q}\varphi$$

(where $p = r + s$). (5) A p-form φ can be uniquely decomposed in the form $\varphi = \varphi_0 + L\varphi_1 + \ldots + L^r\varphi_r$ with $r = [p/2]$ and φ_i primitive (Hodge [7], Weil [13]). Properties (3)–(5) are shared by all Hermitian metrics. When the metric is Kählerian, we further have

$$Ld - dL = 0, \qquad \Lambda d' - d'\Lambda = \sqrt{-1}\,\delta'',$$

$$\Lambda d'' - d''\Lambda = -\sqrt{-1}\,\delta'.$$

From these we also obtain

$$\Delta L = L\Delta, \qquad \Delta\Lambda = \Lambda\Delta,$$

$$d'\delta'' + \delta'' d' = 0, \qquad d''\delta' + \delta' d'' = 0,$$

$$\Delta = 2(d'\delta' + \delta' d') = 2(d''\delta'' + \delta'' d''),$$

where Δ is the †Laplace-Beltrami operator $\Delta = d\delta + \delta d$. †Green's operator G commutes with d', d'', δ', and δ''. Thus when the metric is Kählerian, we have: (6) Let $L_p(X) = \sum_{r+s=p} L_{r,s}(X)$ be the decomposition of the space of p-forms into the spaces $L_{r,s}(X)$ of forms of type (r, s). Then, correspondingly, the space $H_p(X)$ of harmonic p-forms is the direct sum $H_p(X) = \sum_{r+s=p} H_{r,s}(X)$ of the spaces $H_{r,s}(X)$ of harmonic forms of type (r, s). (7) If we let $A = d''\delta'' + \delta'' d''$, then $\Delta\varphi = 0$ is equivalent to $A\varphi = 0$. Denote the projection to the space of harmonic forms by H. Then using the formula $1 = H + \Delta G$ (→ 194 Harmonic Integrals A), we can infer that not only in the †cochain complex $(\sum_p L_p(X), d)$ but also in the cochain complexes $(\sum_s L_{r,s}(X), d'')$ and $(\sum_r L_{r,s}(X), d')$, H is homotopic to the identity mapping and the respective cohomology groups of degree s and degree r of the last two complexes are both isomorphic to $H_{r,s}(X)$. (8) For a harmonic p-form φ, the forms φ_i in the decomposition (5) are harmonic. (9) The exterior powers Ω^r ($r = 0, 1, \ldots, n$) of the fundamental form Ω are non-

zero and harmonic. (10) Holomorphic differential forms are harmonic.

C. Further Properties of Compact Kähler Manifolds

The p-dimensional †de Rham cohomology group with complex coefficients of a compact Kähler manifold X is canonically isomorphic to the direct sum of the †Dolbeault cohomology groups of type (r,s), where $r+s=p$ (→ 72 Complex Manifolds D). Let b_p and $h^{r,s}$, respectively, denote the pth Betti number of X and $\dim_{\mathbf{C}} H^s(X,\Omega^r)$. Then it follows that $b_p=\sum_{r+s=p}h^{r,s}$. If φ is harmonic, then $\bar\varphi$ is also harmonic, so we have $h^{r,s}=h^{s,r}$. Therefore, if p is odd, b_p is even. This is a generalization of the fact that the first Betti number of a closed Riemann surface is even. From (9) we have $b_p\geqslant 1$ if p is even. Furthermore, a linear combination of irreducible analytic subsets of X of (complex) dimension r with positive real coefficients is a cycle of degree $2r$ that is never homologous to zero on X, since the integral of Ω^r on the cycle is never zero.

For a compact Kähler manifold X, we can define a †complex torus $\mathfrak{A}$ and a holomorphic mapping $\lambda: X\rightarrow\mathfrak{A}$ such that (i) $\mathfrak{A}$ is generated by $\lambda(X)$ as a group; (ii) any holomorphic mapping $\mu: X\rightarrow T$ can be decomposed as $\mu=\alpha\circ\lambda+c$, where T is another complex torus, α is a complex analytic homomorphism from $\mathfrak{A}$ to T, and c is a point of T. $\mathfrak{A}$ is called the **Albanese variety** of X. The set $\mathfrak{P}$ of complex analytic isomorphism classes of †complex line bundles that are trivial as topological bundles has a natural complex structure with a canonically associated structure of a family of vector bundles (→ 72 Complex Manifolds G). With this structure $\mathfrak{P}$ is, in fact, a complex torus and is called the **Picard variety** of X. Then $\mathfrak{P}$ and $\mathfrak{A}$ are constructed using $H^1(X,\mathbf{R})$ and $H^{2n-1}(X,\mathbf{R})$, respectively, and are dual complex tori. If X is a Hodge manifold (→ Section D), $\mathfrak{P}$ and $\mathfrak{A}$ are Abelian varieties (→ 3 Abelian Varieties).

A small deformation of a compact Kähler manifold is also Kählerian (a Kähler metric can be taken to be of class C^∞ with respect to parameters) [9]. But a limit (in the sense of deformation) of a Kähler manifold is not always Kählerian. An example was given by Hironaka [6].

Generalizing the notion of compact Kähler manifold, A. Fujiki [2, 3] introduced the category $\mathscr{C}$ by $X\in\mathscr{C}$ if and only if (i) X is a compact complex reduced space and (ii) there exist a compact Kähler manifold Y and a surjective holomorphic mapping: $Y\rightarrow X$. Fujiki proves among other things that when $X\in\mathscr{C}$ is a manifold, (1) every holomorphic p-form is closed, (2) $H^n(X,\mathbf{C})\cong\bigoplus_p H^p(X,\Omega_X^{n-p})$ for any $n\geqslant 0$, and (3) $H^p(X,\Omega_X^q)\cong H^q(X,\Omega_X^p)$ as a vector space, where the Ω_X^p denotes the sheaves of holomorphic p-forms on X.

Let X be a compact Kähler manifold with Kähler metric $g=\sum g_{\alpha\bar\beta}dz^\alpha d\bar z^\beta$. Then $R(g)=\sum R_{\alpha\bar\beta}(g)dz^\alpha d\bar z^\beta$ (where $R_{\alpha\bar\beta}(g)=-\partial^2/\partial z^\alpha\partial\bar z^\beta(\log(\det(g_{\alpha\bar\beta}))))$ is the associated Ricci tensor. According to Chern, $(\sqrt{-1}/2\pi)\sum R_{\alpha\bar\beta}(g)dz^\alpha\wedge d\bar z^\beta$ is a closed (1, 1) form representing the first Chern class of X. Conversely, for a compact Kähler manifold X, E. Calabi conjectured that:

(a) Given a real closed (1, 1) form $(\sqrt{-1}/2\pi)\sum R_{\alpha\bar\beta}dz^\alpha\wedge d\bar z^\beta$ which represents the first Chern class of X, there exists a unique Kähler metric g on X such that its corresponding Ricci tensor $R(g)$ coincides with $\sum R_{\alpha\bar\beta}dz^\alpha d\bar z^\beta$ and that g determines the same cohomology class as the original Kähler metric.

(b) If X has either negative or zero first Chern class, then X admits an Einstein Kähler metric g (where "Einstein" means $R(g)=(\mathrm{const})\cdot g$).

As a partial answer to (b), T. Aubin proved the existence of an **Einstein Kähler metric** on compact complex manifolds with negative first Chern class. Later, S. T. Yau gave a complete affirmative answer for both (a) and (b), developing the method (solving complex Monge-Ampère equations) of Calabi and Aubin. Among several consequences of Yau's work, the following by A. Todorov is one of the most interesting: Every point of $SO(3,19)/SO(2)\times SO(1,19)$ corresponds to at least one marked Kähler surface via period mapping.

Yau also showed that Aubin's previous work solved affirmatively the following conjecture of F. Severi: Every compact complex surface that is homotopic to $\mathbf{P}^2(\mathbf{C})$ is biholomorphic to $\mathbf{P}^2(\mathbf{C})$. Yau's proof used a delicate differential geometric analysis of the inequality $c_1(X)^2[X]\leqslant 3c_2(X)[X]$, proved by Y. Miyaoka for compact complex surfaces X of general type. The key observation is that if X is a compact complex surface with negative first Chern class and $c_1(X)^2[X]=3c_2(X)[X]$, then the universal covering space of X is an open ball $\{(z_1,z_2)\in\mathbf{c}^2;\ |z_1|^2+|z_2|^2<1\}$. For recent developments concerning Calabi's conjecture → [1, 15].

D. Hodge Manifolds

One of the most important examples of Kähler manifolds is a projective nonsingular variety over the field of complex numbers. If we

let $(\zeta_0, \dots, \zeta_N)$ be a homogeneous coordinate system in the N-dimensional projective space $\mathbf{P}^N(\mathbf{C})$, then the subset $\zeta_k \neq 0$ has its holomorphic coordinate system $(z^1, \dots, z^N)$, where $z^1 = \zeta_0/\zeta_k, \dots, z^k = \zeta_{k-1}/\zeta_k$, $z^{k+1} = \zeta_{k+1}/\zeta_k, \dots, z^N = \zeta^N/\zeta^k$. If we let $\psi_k = (1/\pi)\log(1 + \sum_j |z^j|^2)$ and $g_{\alpha\bar{\beta}} = \partial^2\psi/\partial z^\alpha \partial \bar{z}^\beta$, then $ds^2 = \sum g_{\alpha\bar{\beta}} dz^\alpha d\bar{z}^\beta$ is independent of k and determines a **standard Kähler metric** of $\mathbf{P}^N(\mathbf{C})$ that coincides with the so-called Fubini-Study metric on $\mathbf{P}^N(\mathbf{C})$ up to a constant multiplier. In this case, we have the following results: (11) Let $\mathfrak{H}$ be a hyperplane of $\mathbf{P}^N$. Then the integral $\int_Z \Omega$ of the fundamental form Ω on a 2-cycle Z is equal to the †Kronecker index KI$(Z, \mathfrak{H})$ of the †intersection of Z and $\mathfrak{H}$. Therefore (12) an integral (a period) of Ω on a 2-cycle with integral coefficients is an integer. In other words, Ω corresponds to the cohomology class in $H^2(X, \mathbf{R})$ of a cocycle with integral coefficients. Property (12) holds for the induced Kähler metric on an analytic submanifold X of $\mathbf{P}^N(\mathbf{C})$ (which is algebraic by †Chow's theorem). Property (11) also holds if we replace $\mathfrak{H}$ by the intersection Y of X and $\mathfrak{H}$. Property (8) (→ Section B) can be thought of as an expression (in terms of harmonic forms) of Lefchetz's theorem on the topology of projective algebraic varieties.

Generally, if a compact complex manifold X admits a Kähler metric with property (12), we say that the metric is a **Hodge metric** and X is a **Hodge manifold**. **Kodaira's theorem** asserts that a Hodge manifold has a biholomorphic embedding into a projective space [8]. Cohomology vanishing theorems (→ Section D) and the properties of †monoidal transformations of complex manifolds are used to prove this theorem.

On a closed Riemann surface $\mathfrak{R}$, i.e., a 1-dimensional compact complex manifold, any Hermitian metric is Kählerian. Moreover, a metric with total volume 1 is a Hodge metric. This proves that $\mathfrak{R}$ is isomorphic to an algebraic curve in a projective space. This is a proof (using Kodaira's theorem) of the existence of nonconstant meromorphic functions on $\mathfrak{R}$. The condition on the †Riemann matrix for a complex torus $T = \mathbf{C}^n/D$ (D is a discrete subgroup of rank $2n$) to be a projective algebraic variety is that T admit a Hodge metric.

Let L be the sheaf associated with a positive line bundle (i.e., its first Chern class is represented by a Hodge metric) on a compact complex manifold X. Then the **Kodaira vanishing theorem** states that $H^i(X, K_X \otimes L) = 0$ for any $i > 0$, where K_X is the canonical sheaf, i.e., the sheaf of holomorphic n-forms, $n = \dim X$. Several generalizations are known. For instance, let X be a compact complex variety with $a(X) = \dim X$ and F be a sheaf associated with a positive vector bundle on X. Take a nonsingular projective variety Y and a birational holomorphic mapping $\pi: Y \to X$ and define K_X to be the direct image $\pi_*(K_Y)$. Then $H^i(X, K_X \otimes F) = 0$ for any $i > 0$ (S. Nakano [10], H. Grauert and O. Riemenschneider [5]). Ramanujam's vanishing theorem is concerned with an effective divisor D on a nonsingular projective surface S. D is said to be **numerically connected** if the intersection number $D_1 \cdot D_2 > 0$ for every decomposition $D = D_1 + D_2$ with $D_i > 0$. If $D^2 > 0$ and D is numerically connected, then $H^1(S, \mathscr{I}_D) = 0$, where $\mathscr{I}_D$ is the sheaf of ideals defining D [12]. However, Kodaira's vanishing theorem fails for algebraic manifolds over fields of positive characteristic.

E. Examples and Other Properties

Concerning differential geometry on compact Kähler manifolds, the analytic transformation group and the †isometric transformation group are also studied. For example, let X be a compact Kähler manifold of complex dimension n. Then the Lie group of isometric transformations on X is of dimension $\leqslant n^2 + 2n$, and the equality holds if and only if $X = \mathbf{P}^n(\mathbf{C})$.

A nonalgebraic complex torus is the most important example of a compact Kähler manifold that is not algebraic. A complex torus is not an algebraic variety if it is not a Hodge manifold.

An example of a noncompact Kähler manifold is a bounded domain in $\mathbf{C}^n$ with the Kähler metric

$$ds^2 = \sum_{\alpha,\beta} (\partial^2 \log K(z, \bar{z})/\partial z^\alpha \partial \bar{z}^\beta) dz^\alpha d\bar{z}^\beta,$$

where $K(z, \bar{\zeta})$ is †Bergman's kernel function. This metric is significant because of its invariance under the analytic automorphisms of the domain. More generally, a †Stein manifold admits a complete Kähler metric [4].

References

[1] J. P. Bourguignon et al., Première classe de Chern et courbure de Ricci: preuve de la conjecture de Calabi (Séminaire Palaiseau, 1978), Astérisque 58, Société Mathématique de France.

[2] A. Fujiki, Closedness of the Douady space of compact Kähler spaces, Publ. Res. Inst. Math. Sci., 14 (1978), 1–52.

[3] A. Fujiki, On automorphism groups of compact Kähler manifolds, Inventiones Math., 44 (1978), 225–258.

[4] H. Grauert, Charakterisierung der Holomorphiegebiete durch die vollständige Kählersche Metrik, Math. Ann., 131 (1956), 38–75.

[5] H. Grauert and O. Riemenschneider, Verschwindungssätze für analytische Kohomologiegruppen auf komplexen Räumen, Inventiones Math., 11 (1970), 263–292.
[6] H. Hironaka, An example of a non-Kählerian complex-analytic deformation of Kählerian complex structures, Ann. Math., (2) 75 (1962), 190–208.
[7] W. V. D. Hodge, The theory and applications of harmonic integrals, Cambridge Univ. Press, second edition, 1952.
[8] K. Kodaira, On Kähler varieties of restricted type (An intrinsic characterization of algebraic varieties), Ann. Math., (2) 60 (1954), 28–48.
[9] K. Kodaira and D. C. Spencer, On deformations of complex analytic structures III, Ann. Math., (2) 71 (1960), 43–76.
[10] S. Nakano, On complex analytic vector bundles, J. Math. Soc. Japan, 7 (1955), 1–12.
[11] C. P. Ramanujam, Remarks on the Kodaira vanishing theorem, J. Indian Math. Soc., 36 (1972), 41–51.
[12] C. P. Ramanujam, Supplement to the article "Remarks on the Kodaira vanishing theorem," J. Indian Math. Soc., 38 (1974), 121–124.
[13] A Weil, Introduction à l'étude des variétés kähleriennes, Actualités Sci. Ind., Hermann, 1958.
[14] R. O. Wells, Jr., Differential analysis on complex manifolds, Prentice-Hall, 1973.
[15] S. T. Yau, Calabi's conjecture and some new results in algebraic geometry, Proc. Nat. Acad. Sci. US, 74 (1977), 1798–1799.
[16] S. T. Yau, On the Ricci curvature of a compact Kähler manifold and the complex Monge-Ampère equation I, Comm. Pure Appl. Math., 31 (1978), 339–411.

233 (XXI.30) Klein, Felix

Felix Klein (May 25, 1849–June 22, 1925) was one of the leading mathematicians in Germany in the latter half of the 19th century. Born in Düsseldorf and graduated from the University of Bonn, Klein went to study in Paris. In 1872 he became a professor at the University of Erlangen, and in 1886 attained a chair at the University of Göttingen, where he was employed until his death. His accomplishments cover all aspects of mathematics, but his main field was geometry. In his inaugural lecture at the University of Erlangen, he presented a bird's-eye view of all the then known fields of geometry from the standpoint of group theory, which is referred to as the †Erlangen program (→ 137 Erlangen Program). In it he stated that both Euclidean and non-Euclidean geometry are included in †projective geometry. Klein said in his last lectures [4] that he spent the greatest part of his energies in the field of †automorphic functions. These last lectures are important as historical material on the mathematics of the 19th century. Klein was a leader of reforms of mathematical education in Germany [3].

References

[1] F. Klein, Gesammelte mathematische Abhandlungen I–III, Springer, 1921–1923.
[2] R. Fricke and F. Klein, Vorlesungen über die Theorie der automorphen Funktionen, Teubner, I, 1897; II, 1901.
[3] F. Klein, Elementarmathematik von höheren Standpunkte aus, Springer, I, 1924; II, 1925; III, 1928.
[4] F. Klein, Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert I, II, Springer, 1926–1927 (Chelsea, 1956).
[5] F. Klein, Vorlesungen über nicht-Euklidische Geometrie, Springer, 1928 (Chelsea, 1960).
[6] F. Klein, Vorlesungen über höhere Geometrie, Springer, third edition, 1926 (Chelsea, 1957).

234 (XI.17) Kleinian Groups

A. Definitions

Let H^3 be the upper half-space of the 3-dimensional Euclidean space. The boundary plane of H^3 is regarded as the complex plane C. The one-point compactification $H^3 \cup \hat{C}$ of H^3 is denoted by $\hat{H}^3$, where $\hat{C}$ is the extended complex plane. Let Möb(H^3) be the group of all the Möbius transformations in $\hat{H}^3$ which are represented by an even number of compositions of reflections with respect to hemispheres in $\hat{H}^3$ orthogonal to $\hat{C}$. Here, the plane in H^3 orthogonal to C is regarded as a hemisphere orthogonal to $\hat{C}$. The hyperbolic structure can be defined naturally in H^3, and this structure is preserved by any element in Möb(H^3). A hemisphere in $\hat{H}^3$ orthogonal to $\hat{C}$ is a (hyperbolic) plane in this structure. The restriction of any element in Möb(H^3) to $\hat{C}$ is a linear fractional transformation on $\hat{C}$; conversely, for any given linear fractional transformation γ on $\hat{C}$, there is a unique Möbius

transformation in Möb(H^3) whose restriction to $\hat{C}$ coincides with γ.

A **Kleinian group** Γ is defined as a discrete subgroup of Möb(H^3) and acts discontinuously on H^3. If there exist a point $p \in H^3$ and a sequence $\{\gamma_n\}$ of distinct elements of Γ such that the sequence $\{\gamma_n(p)\}$ converges to a point $\zeta \in C$ in the usual topology of $\hat{H}^3$, then the point ζ is called a limit point of Γ, and the set $\Lambda(\Gamma)$ of all the limit points of Γ is called the **limit set** of Γ, which is a closed subset of $\hat{C}$. When $\Lambda(\Gamma) = \hat{C}$, the Kleinian group Γ is of the first kind. Otherwise, Γ is of the second kind, and the restriction of Γ to $\hat{C}$ acts discontinuously on the **region of discontinuity** $\Omega(\Gamma) = \hat{C} - \Lambda(\Gamma)$ for Γ. Usually, "a Kleinian group" means one of the second kind. Thus the Kleinian group Γ acts discontinuously on $H^3 \cup \Omega(\Gamma)$, the quotient space H^3/Γ is a 3-manifold with a hyperbolic structure induced by that of H^3, and the union of some Riemann surfaces $\Omega(\Gamma)/\Gamma$ is the boundary of this 3-manifold. A study of Kleinian groups in connection with the theory of 3-manifolds has been done recently by W. P. Thurston.

If $\Lambda(\Gamma)$ consists of at most two points, then the Kleinian group Γ is called elementary; if otherwise, Γ is nonelementary and $\Lambda(\Gamma)$ is perfect and nowhere dense in $\hat{C}$, and $\Omega(\Gamma)$ allows the Poincaré metric and has a hyperbolic structure invariant under Γ. A Kleinian group whose region of discontinuity has a nonempty connected component invariant under Γ is often called a **function group**.

B. Examples

A Fuchsian group (→ 122 Discontinuous Groups) is a Kleinian group which leaves a circular disk in $\hat{C}$ invariant. Another important Kleinian group is the so-called **Schottky group**. Let $\{C_j, C_j'\}_{j=1}^n$ be n ($\geqslant 2$) pairs of Jordan curves on C, every one of which lies in the exterior of each of the other $2n-1$ curves. Assume that there are n loxodromic (or hyperbolic) linear fractional transformations γ_j, where γ_j maps the interior of C_j onto the exterior of C_j' conformally. The group generated by these $\{\gamma_j\}$ is a Schottky group whose limit set is totally disconnected and has its nonempty subset in the interior of every C_j and C_j'. A Kleinian group is a Schottky group if and only if it consists of only loxodromic or hyperbolic elements, is finitely generated, and is free. A Schottky group is not always Fuchsian, and there are many kinds of Kleinian groups which are not Fuchsian. Quasiconformal deformation of a Fuchsian group G of the first kind yields the †Teichmüller space $T(G)$ with the center G, and every point of $T(G)$ is a **quasi-Fuchsian group** whose limit set is a quasicircle, an image of a circle on $\hat{C}$ under a quasiconformal mapping of $\hat{C}$ onto itself. To every point of the boundary of $T(G)$, there corresponds a Kleinian group called a **boundary group** for G. Among them, there is a Kleinian group whose region of discontinuity is connected and simply connected. Such a Kleinian group is called **totally degenerate**. A **web group** is a Kleinian group whose region of discontinuity consists only of Jordan domains.

C. Fundamental Sets

Let Γ be a Kleinian group acting on H^3. There is a †fundamental set for Γ in H^3 which is bounded by a countable number of hyperbolic planes in H^3 and is convex in the sense of the hyperbolic structure. The closure of this fundamental set in H^3 is called a convex fundamental polyhedron for Γ. For a point $p_0 \in H^3$ not fixed by elements of Γ except for the identity, the set $\{p \in H^3 \mid d(p, p_0) \leqslant d(\gamma(p), p_0)$ for each $\gamma \in \Gamma\}$, where d denotes the hyperbolic distance, yields a convex fundamental polyhedron for Γ and is called the **Dirichlet region** with the center p_0 for Γ. If Γ has a convex fundamental polyhedron with a finite number of sides or faces, then Γ is called **geometrically finite**. Similarly, a Dirichlet region for a Fuchsian group in the upper half complex plane H can be defined, and it is often called a normal polygon for the group. For a Kleinian group Γ, instead of fundamental sets for Γ in $\Omega(\Gamma)$, the concept of fundamental domains is sometimes useful. A **fundamental domain** for a Kleinian group Γ in $\Omega(\Gamma)$ ($\neq 0$) is an open subset D of $\Omega(\Gamma)$ such that Γ-images of any point in D, except the point itself, are not in D and such that some of the Γ-images of any point in $\Omega(\Gamma)$ lie on the closure of D in $\Omega(\Gamma)$. The **Ford fundamental region** is also useful. The action of $\gamma \in \Gamma$ on $\hat{C}$ is given by a linear fractional transformation $z \rightarrow (az+b)/(cz+d)$, $ad-bc=1$. The circle $\{z \in C \mid |cz+d|=1\}$ is called the **isometric circle** of γ with $c \neq 0$. If $\infty \in \Omega(\Gamma)$, then the set, every point of which is in the exterior of isometric circles of all $\gamma \in \Gamma$ with $c \neq 0$, is a Ford fundamental region for Γ [7].

D. Finitely Generated Kleinian Groups

Much work has been done on finitely generated Kleinian groups [2, 5, 9, 10, 12, 13], while very little has been done on nonfinitely generated Kleinian groups. If a Kleinian group Γ (of the second kind) is finitely generated,

then the quotient space $\Omega(\Gamma)/\Gamma$ is a disjoint union of a finite number of compact Riemann surfaces with at most a finite number of punctures (**Ahlfors's finiteness theorem**). Furthermore, if the above Γ is nonelementary and is generated by N generators, then the (hyperbolic) area of $\Omega(\Gamma)/\Gamma$ is not greater than $4\pi(N-1)$ (**Bers's area theorem**). These two theorems played important roles in the study of Kleinian groups after 1960. If a Fuchsian group is finitely generated, then every normal polygon for the group has a finite number of sides. In contrast with this fact, the following is suggestive: If a finitely generated Kleinian group Γ is totally degenerate, then Γ is not geometrically finite (L. Greenberg).

There is a method to obtain a Kleinian group from two simple groups. If Γ_j $(j=1,2)$ is a finitely generated Kleinian group and if the interior of a connected fundamental domain D_j for Γ_j $(j=1,2)$ contains the boundary and the exterior of D_i $(i\neq j)$, then the group generated by Γ_1 and Γ_2 is again a Kleinian group with a fundamental domain $D_1\cap D_2$ (**Klein's combination theorem**). Generalizations of this theorem are discussed by Maskit [14].

Let Γ be a nonelementary Kleinian group, and let Δ $(\neq 0)$ be an invariant union of components of $\Omega(\Gamma)$ under Γ. Denote by $A_q(\Delta,\Gamma)$ the Banach space of holomorphic automorphic forms $\varphi(z)dz^q$ in Δ for Γ satisfying $\iint_{\Delta/\Gamma}\rho(z)^{2-q}|\varphi(z)|\,dx\,dy<\infty$, where Δ/Γ is a measurable fundamental set for Γ in Δ, $z=x+iy$, and $\rho(z)|dz|$ is the Poincaré metric on Δ. The Banach space $B_q(\Delta,\Gamma)$ is the totality of holomorphic automorphic forms $\psi(z)dz^q$ for Γ in Δ with $\sup_{z\in\Delta}\rho(z)^{-q}|\psi(z)|<\infty$. If Γ is finitely generated, then $A_q(\Delta,\Gamma)=B_q(\Delta,\Gamma)$, and it is the space of all the cusp forms for Γ in Δ. If Γ is not finitely generated, then even the inclusion $A_q(\Omega(\Gamma),\Gamma)\subset B_q(\Omega(\Gamma),\Gamma)$ does not hold in general.

E. Limit Sets of Kleinian Groups

Let Γ be a Kleinian group, and let $\{\Omega_j\}$ be the collection of all the components of $\Omega(\Gamma)$. Although it was stated in [8] that $\Lambda(\Gamma)$ consists of boundaries $\partial\Omega_j$ of Ω_j, it was proved by Abikoff in [3] that the statement is incorrect; that is, the **residual limit set** $\Lambda_0(\Gamma)=\Lambda(\Gamma)-\bigcup_j\partial\Omega_j$ of Γ is not always empty. In fact, there is a web group whose region of discontinuity consists of an infinite number of components, and such a group has a nonempty residual limit set [1].

It was conjectured by Ahlfors that the 2-dimensional Lebesgue measure of $\Lambda(\Gamma)$ equals zero for every finitely generated Kleinian group Γ of the second kind. This conjecture is one of the sources of recent developments of the theory of Kleinian groups, and it is still open. Concerning this conjecture, the following fact was proved by Sullivan in [13] from the standpoint of the ergodic theory: The Beltrami coefficient for Γ with support on $\Lambda(\Gamma)$ is equal to zero. It is also known that, if Γ is geometrically finite, then the conjecture is affirmative. The study of Kleinian groups from the viewpoint of ergodic theory has also been developed [11, 15].

Let $M_t(\Gamma)$ be the †t-dimensional Hausdorff measure of $\Lambda(\Gamma)$. The **Hausdorff dimension** $d(\Gamma)$ of $\Lambda(\Gamma)$ is defined by $d(\Gamma)=\inf\{t>0\mid M_t(\Gamma)=0\}$. There are some studies on the Hausdorff dimension of $\Lambda(\Gamma)$ for several kinds of Kleinian groups Γ [3, 4, 6].

References

[1] W. Abikoff, The residual limit sets of Kleinian groups, Acta Math., 130 (1973), 127–144.
[2] L. Ahlfors et al., Advances in the theory of Riemann surfaces, Ann. math. studies 66, Princeton Univ. Press, 1971.
[3] T. Akaza, Local property of the singular sets of some Kleinian groups, Tôhoku Math. J., (2) 25 (1973), 1–22.
[4] A. F. Beardon, Hausdorff dimension of singular set of properly discontinuous groups, Amer. J. Math., 88 (1966), 722–736.
[5] L. Bers, I. Kra, et al., A crash course on Kleinian groups, Lecture notes in math. 400, Springer, 1974.
[6] R. Bowen, Hausdorff dimension of quasi-circles, Publ. Math. Inst. HES, 50 (1979), 1–25.
[7] L. R. Ford, Automorphic functions, Chelsea, second edition, 1951.
[8] R. Fricke and F. Klein, Vorlesungen über die Theorie der Automorphen Funktionen I, II, Teubner, 1897.
[9] L. Greenberg et al., Discontinuous groups and Riemann surfaces, Ann. math. studies 79, Princeton Univ. Press, 1974.
[10] W. J. Harvey, et al., Discrete groups and automorphic functions, Academic Press, 1977.
[11] E. Hopf, Ergodentheorie, Springer, 1937.
[12] I. Kra, Automorphic forms and Kleinian groups, Benjamin, 1972.
[13] I. Kra, B. Maskit, et al., Riemann surfaces and related topics, Ann. math. studies 97, Princeton Univ. Press, 1980.
[14] B. Maskit, On Klein's combination theorem, Trans. Amer. Math. Soc., 120 (1965), 499–509.
[15] M. Tsuji, Potential theory in modern function theory, Maruzen, 1959.

235 (IX.16) Knot Theory

A. General Remarks

The knot problem is a special case of the **placement problem**, stated as follows: Given two homeomorphic topological spaces X_1 and X_2 and their respective homeomorphic subsets A_1 and A_2, is there a homeomorphism $f: X_1 \to X_2$ such that $f(A_1) = A_2$? A simple closed curve in a Euclidean 3-space $\mathbf{R}^3$ (or in a 3-sphere S^3 obtained from $\mathbf{R}^3$ by one-point compactification, sometimes used in preference to $\mathbf{R}^3$) is called a **knot**. If two knots K_1 and K_2 can be mapped from one to the other by a homeomorphism of $\mathbf{R}^3$ (or of S^3), they are called **equivalent**. All knots are classified into **knot types** by this equivalence.

Let (x, y, z) be Cartesian coordinates in $\mathbf{R}^3$. Knots equivalent to $x^2 + y^2 = 1$, $z = 0$ and their knot type are called **trivial** or **unknotted**; knots of other types are called **knotted**. Knots equivalent to those given by polygonal curves in $\mathbf{R}^3$ and their knot types are called **tame**; other knots are called **wild**. We discuss only tame knots in this article except in Section H. Given a knot K in $\mathbf{R}^3$, there exists a plane such that the orthogonal projection π on it has the following two properties: (1) The image $\pi(K)$ has no multiple points other than a finite number of double points. (2) The projections of the vertices of K are not double points of $\pi(K)$. Then $\pi(K)$ is called a **regular knot projection** of K. Let $z = 0$ be the plane in question. Of the two points of K corresponding to a double point of $\pi(K)$, the one with the greater z-coordinate is called the **overcrossing point**; the one with the smaller z-coordinate is called the **undercrossing point**. Suppose that overcrossing points and undercrossing points appear alternately when we move along K in a fixed direction; then K is called **alternating**.

Knots K_1 and K_2 in $\mathbf{R}^3$ are said to be of the same **isotopy type** if there is an †isotopy $\{h_t\}$ $(0 \leqslant t \leqslant 1)$ of $\mathbf{R}^3$ such that $h_t: \mathbf{R}^3 \to \mathbf{R}^3$ are homeomorphisms with the identity h_0 and such that $h_1(K_1) = K_2$. Knots K_1 and K_2 of $\mathbf{R}^3$ are of the same isotopy type if and only if K_1 is mapped onto K_2 by an orientation-preserving homeomorphism of $\mathbf{R}^3$. Thus knots of the same isotopy type are equivalent. The converse, however, is false. A knot that can be mapped onto itself by an orientation-reversing homeomorphism of $\mathbf{R}^3$ is called **amphicheiral**. A knot K is called **invertible** if K is mapped onto itself by an orientation-preserving homeomorphism h such that $h|K$ reverses the orientation of K. Knots 4_1, 6_3, 8_3, 8_9, 8_{12}, 8_{17}, 8_{18} in Appendix A, Table 7 are amphicheiral, and 8_{17} is noninvertible.

Given a knot K, there is an orientable surface F of †genus h, say, having K as its boundary, called a **Seifert surface** of K [1]. The minimum of the genera of Seifert surfaces of K is called the **genus** of K, and there is an algorithm with which one can determine the genus of K (W. Haken, *Acta Math.*, 105 (1961)). Using †Dehn's lemma and the †sphere theorem, C. D. Papakyriakopoulos showed that for a tame knot, $\pi_i(S^3 - K) = 0$ $(i \geqslant 2)$, and further, K is unknotted if and only if $\pi_1(S^3 - K) \cong \mathbf{Z}$ [2].

The first attempt to list all knots systematically was made by P. G. Tait and C. N. Little in the 19th century; an attempt was also made by J. H. Conway in 1970. At present, the classification has been completed for all knot types with at most ten double points by using the invariants of knots introduced in Sections B–E [3, 4].

Given two knots K_1 and K_2 separated by a 2-sphere S^2, we can tie them together with a narrow strip to obtain a new knot K (Fig. 1), called the **product** or **composition** of K_1 and K_2. If the strip is chosen so that the orientations of K_1, K_2, and K are mutually compatible, then the isotopy type of K is uniquely determined by those of K_1 and K_2. The set of isotopy types forms a commutative †semigroup with respect to this product, in which the trivial knot type is an identity. Any knot is decomposed uniquely (up to isotopy) into finite products of the **prime** knots, which cannot be decomposed into products of knotted knots (H. Schubert and S.-B. Heidelberger, *Akad. Wiss.*, 3 (1949)).

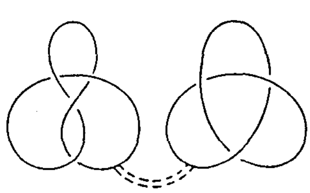


Fig. 1

Until about 1930 the theory of knots had been studied chiefly by J. W. Alexander in the United States and by K. Reidemeister, H. Seifert, and others in Germany. Little progress was made, however, until a new development of the theory was introduced by R. H. Fox [5] and his school in the United States. In Japan, significant contributions have been made by T. Homma, S. Kinoshita, K. Murasugi, and others [6].

B. Knot Groups

If K is a knot, the †fundamental group $G(K) = \pi_1(\mathbf{R}^3 - K)$ of $\mathbf{R}^3 - K$ is called the **knot group** of K. Any group G is written as a †quotient group F/N with a †free group F and its †normal subgroup N. By specifying the generating set $\{x_1, x_2, \ldots\}$ of F and a set of elements $\{r_1, r_2, \ldots\}$ in F such that N is the minimum normal subgroup of F containing $r_1, r_2, \ldots$, we obtain a system $(x_1, x_2, \ldots : r_1, r_2, \ldots)$, called a **presentation** of G, written $G = (x_1, x_2, \ldots : r_1, r_2, \ldots)$. For the knot group $G(K)$, there is a standard presentation called a Wirtinger presentation, which is obtained as follows. First, for the sake of convenience, we give an orientation to K. (The orientation is irrelevant, however, since the reverse orientation gives an isomorphic group.) Consider a regular knot projection $\pi(K)$. If $\pi(K)$ has n double points $\pi(d_i)$, then K is divided into n arcs z_i by n undercrossing points. We consider a free group F generated by n letters $x_1, \ldots, x_n$. To each double point $\pi(d_i)$ we associate a word r_i in the following manner. The two cases near $\pi(d_i)$ are illustrated in Fig. 2. For the first case, we define $r_i = x_{\lambda+1}^{-1} x_\mu x_\lambda x_\mu^{-1}$; and for the second, $r_i = x_{\lambda+1}^{-1} x_\mu^{-1} x_\lambda x_\mu$. Then the presentation $(x_1, x_2, \ldots, x_n : r_1, r_2, \ldots, r_n)$ is called a **Wirtinger presentation** of $G(K)$. Since any one of these r_i is a consequence of the rest, $G(K)$ may be written as $(x_1, x_2, \ldots, x_n : r_1, r_2, \ldots, r_{n-1})$.

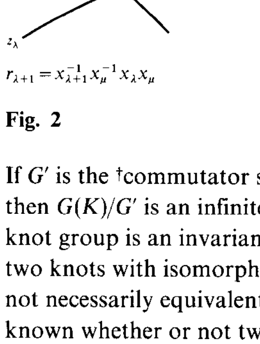


Fig. 2

If G' is the †commutator subgroup of $G(K)$, then $G(K)/G'$ is an infinite cyclic group $\mathbf{Z}$. A knot group is an invariant of a knot type, but two knots with isomorphic knot groups are not necessarily equivalent. However, it is not known whether or not two prime knots with isomorphic knot groups are equivalent. The affirmative statement is called the **general knot conjecture** and is one of the fundamental problems in knot theory. The **knot complement conjecture** is another important conjecture and states that two prime knots with isomorphic knot groups have homeomorphic complements [11]. Nevertheless, the knot group plays a prime role in knot theory. To be more precise, given a knot K in S^3, consider a †tubular neighborhood V of K. V is a solid torus. Choose two simple closed curves m and l on ∂V in such a way that (1) m and l intersect at the base point; (2) m bounds a disk in V, but m is not homologous to 0 on ∂V; (3) l bounds an orientable surface in $\overline{S^3 - V}$, but l is not homologous to 0 on ∂V. A pair $\{m, l\}$ is called a **peripheral system** for K, and m is a **meridian** and l a **longitude** of K. The meridian m and the longitude l, as elements of $G(K)$, generate a free Abelian subgroup $\rho(K)$ of $G(K)$, called the peripheral subgroup of $G(K)$. Two knots K_1 and K_2 are equivalent if and only if there is an isomorphism φ from $G(K_1)$ to $G(K_2)$ which sends $\{m_1, l_1\}$ onto a conjugate of $\{m_2^{\pm 1}, l_2^{\pm 1}\}$ in $G(K_2)$ (F. Waldhausen, *Ann. Math.*, (2) 87 (1968)). Therefore the **group system** $\{G(K), \rho(K)\}$ is an important invariant of the knot type of K.

$G(K)$ has been studied quite extensively since 1960. For example, the commutator subgroup G' of $G(K)$ is †finitely generated if and only if G' is free, and then the rank of G' is twice the genus of K [7]. $\overline{S^3 - V}$ is a fiber bundle over S^1 with the fiber an orientable surface if and only if G' is finitely generated (J. Stallings, 1962). Such a knot is called a fibered knot. $G(K)$ is residually finite for any knot K (the result requires Thurston's hyperbolization theorem), and thus the †word problem for $G(K)$ is solvable. $G(K)$ has no element of finite order [2]. The knot type whose group has a nontrivial †center is completely characterized. Also, many problems on 3-manifolds can be formulated in terms of knot groups. For example, the following conjecture, related to †Poincaré's conjecture, still remains unsolved. **Property P conjecture**: For a knotted knot K in S^3, let N_q be the smallest normal subgroup of $G(K)$ containing ml^q. If $G(K)/N_q$ is a trivial group, then $q = 0$.

C. Alexander Invariants and the Seifert Matrix

Let $(x_1, \ldots, x_n; r_1, \ldots, r_{n-1})$ be a presentation of the knot group G of a knot K, F be the free group generated by $x_1, x_2, \ldots, x_n$, and $\varphi: F \to G$, $\psi: G \to H = G/G'$ be canonical homomorphisms. Then φ and ψ can be extended to homomorphisms between †group algebras with integral coefficients $\varphi: \mathbf{Z}[F] \to \mathbf{Z}[G]$ and $\psi: \mathbf{Z}[G] \to \mathbf{Z}[H]$. For any word r in F, we have the **free derivative** $\partial r / \partial x_i$ for $i = 1, \ldots, n$ satisfying the following:

$$\partial x_i / \partial x_j = \delta_{ij}, \qquad \partial x_i^{-1} / \partial x_i = -x_i^{-1},$$

$$\partial (rs) / \partial x_i = \partial r / \partial x_i + r \cdot (\partial s / \partial x_i)$$

(Fox [6]). Now we have the **Alexander matrix** $A=(a_{ij})$ $(i=1,\dots,n-1; j=1,\dots,n)$ of the knot K defined by $a_{ij}=\psi\circ\varphi(\partial r_i/\partial x_j)$. The ideal E_k of $\mathbf{Z}[H]$ generated by all $(n-k)\times(n-k)$ minors of A is called the kth **elementary ideal** of A. $0\subseteq E_1\subseteq E_2\subseteq\dots\subseteq E_n=\mathbf{Z}[H]$, and in particular, E_1 is a principal ideal, called the **Alexander ideal** of K. Since $H\cong\mathbf{Z}$, any element of $\mathbf{Z}[H]$ is a finite sum of elements of the form mt^n $(m,n\in\mathbf{Z})$. The generator $\Delta(t)$ of the Alexander (principal) ideal is called the **Alexander polynomial** of the knot K. The elementary ideals of A, including $\Delta(t)$, are invariants of the knot type of K. $\Delta(t)$ is uniquely determined by K up to the factor $\pm t^\lambda$ and has the following properties: (i) $\Delta(1)=\pm1$; (ii) $\Delta(1/t)=t^\lambda\Delta(t)$. Conversely, any polynomial with integral coefficients satisfying (i) and (ii) is the Alexander polynomial of some knot. For example, the knot group G of the **trefoil knot** (or **cloverleaf knot**) K (Fig. 3) is presented by $G=(a,b;\ aba(bab)^{-1})$. The Alexander matrix of K is $(1-t+t^2, -1+t-t^2)$, and the Alexander polynomial of K is $\Delta(t)=1-t+t^2$.

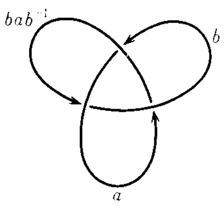


Fig. 3

For a trivial knot, the Alexander polynomial is 1, but the converse is false. The first example of a nontrivial knot with $\Delta(t)=1$ was given by Seifert (Fig. 4) [1]. The Kinoshita-Terasaka knot (Fig. 5) is another knot with this property (*Osaka J. Math.* 9 (1957)).

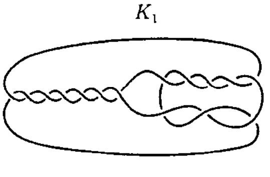


Fig. 4

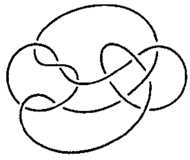


Fig. 5

The degree of the Alexander polynomial never exceeds twice the genus of the knot, and the equality holds for any alternating knots and for fibered knots. For fibered knots, $\Delta(0)=\pm1$ and this is also sufficient for alternating knots to be fibered knots.

The Seifert surface F of K is used to define an invariant of K as follows. Since $H_1(F;\mathbf{Z})$ is a free Abelian group of rank $2h$, we can choose $2h$ oriented simple closed curves $a_1,a_2,\dots,a_{2h}$ on F, whose homology classes form a free basis for $H_1(F;\mathbf{Z})$. Consider $F\times[-1,1]$ in S^3, and let v_{ij} denote the †linking number $\mathfrak{v}(a_i\times\{-1\}, a_j\times\{1\})$. Then the $2h\times 2h$ integral matrix $V=(v_{ij})$ is called the **Seifert matrix** of K. It has been shown that the determinant of the matrix ${}^tV-tV$ is the Alexander polynomial $\Delta(t)$ of K, and the †signature of the symmetric matrix ${}^tV+V$ is a useful knot invariant $\sigma(K)$, called the **signature** of K (H. F. Trotter, 1962; [8]). For a trefoil knot K (Fig. 3), $\sigma(K)=-2$. Since $\sigma(K)=0$ for an amphicheiral knot, this shows that a trefoil knot is not amphicheiral.

D. Link Theory

A link L in $\mathbf{R}^3$ is a disjoint union of knots K_i in $\mathbf{R}^3$ and is denoted by $L=K_1\cup K_2\cup\dots\cup K_r$. Two links $L=K_1\cup K_2\cup\dots\cup K_r$ and $L'=K_1'\cup K_2'\cup\dots\cup K_s'$ are said to be equivalent if $r=s$ and each K_i is mapped onto K_i', $i=1,2,\dots,r$, by a homeomorphism of $\mathbf{R}^3$. Further, we can define the isotopy type of a link, as we did in Section A. In link theory, however, we mostly consider oriented links, i.e., each component is given an orientation. Therefore, in the following, we consider only oriented links, and the **link type** means an isotopy type which preserves the orientation of each component.

Let $L=K_1\cup K_2\cup\dots\cup K_r$ be a link in $\mathbf{R}^3\subset S^3$. The **link group** $G(L)$ of L is defined as $\pi_1(S^3-L)$. Using a regular projection, a **Wirtinger presentation** of $G(L)=(x_1,x_2,\dots,x_n: r_1,r_2,\dots,r_{n-1})$ can be obtained in exactly the same manner as was described in Section B. The Alexander matrix A of L is also defined by means of the free derivative from a presentation of $G(L)$. Since G/G' is a free Abelian group of rank r, the Alexander matrix A is a matrix over an integral Laurent polynomial ring with r variables $t_1,t_2,\dots,t_r$, and hence, the first elementary ideal $E_1(A)$ may not be principal. A generator of the smallest principal ideal containing the generators of $E_1(A)$ is called the **Alexander polynomial** or the **link polynomial** of L. It is an integral polynomial $\Delta(t_1,t_2,\dots,t_r)$ which is uniquely determined up to the factor $\pm t_1^{\lambda_1}t_2^{\lambda_2}\dots t_r^{\lambda_r}$. The link polynomial is an invariant of the link type but is not an invariant of the link group. There are some necessary conditions for a polynomial to be a link polynomial, but they are not sufficient to characterize a link polynomial. By putting $t_1=t_2=\dots=t_r$, we obtain a polynomial of one variable $\tilde{\Delta}(t)$. $(1-t)\tilde{\Delta}(t)$ is called the **reduced link polynomial.** $\tilde{\Delta}(t)$ is divisible by $(1-t)^{r-2}$ and $\nabla(t)=\tilde{\Delta}(t)/(1-t)^{r-2}$ is called a **Hosokawa polynomial** of L. $\nabla(t)$ is symmetric.

$\pi_2(S^3-L)\neq0$ if and only if L splits, i.e., there is a 2-sphere S^2 in S^3 which is disjoint from L, but each component of S^3-S^2 con-

tains at least one component of L. Given two links, we can define (not uniquely) a product link in the same manner as was described in Section A, and any link type is decomposed uniquely into a finite number of prime links. The complement of a prime link cannot be determined from its group.

Given a link L of r components, we can find an orientable surface F of genus h, say, whose boundary is L, and thus we can define a Seifert matrix V for L. (→ Section C. The only difference is that $H_1(F;\mathbf{Z})$ is free Abelian of rank $2h+r-1$.) The determinant of ${}^tV-tV$ is, then, the reduced link polynomial. The signature of ${}^tV+V$ is called the signature of L [8].

For links in S^3, we can define a weaker equivalence, called the **homotopy type**. Two links are said to be homotopic (or of the same homotopy type) if one link is deformed continuously onto another under the condition that during the deformation each component of the link is allowed to cross itself, but no two components are allowed to intersect. A typical homotopy invariant of a link L is the linking number $\lambda_{ij}=\mathrm{Lk}(K_i,K_j)$. The absolute value of λ_{ij} is determined by G/G_2, where G_i, $i=1,2,\ldots$, denotes the †lower central series of $G(L)$. Using the group G/G_k, $k\geqslant 1$, J. Milnor defines a numerical link invariant $\bar{\mu}(j_1 j_2\ldots j_k)$, $1\leqslant j_1,j_2,\ldots,j_k\leqslant r$, called **Milnor's invariant** (*Ann. Math.*, (2) 59 (1954)). In particular, $\bar{\mu}(ij)=\lambda_{ij}$, and $\bar{\mu}$ completely determines the homotopy type of a link with at most three components.

E. Representations and Covering Spaces

Let K be a knot (or a link) in S^3. A †transitive homomorphism φ from the group $G(K)$ into S_n, the †symmetric group of degree n, is called a **representation** of $G(K)$ of degree n. Two representations φ_1 and φ_2 are said to be equivalent if there is an †inner automorphism ρ of S_n such that $\varphi_2=\rho\varphi_1$.

Since $G/G'\cong\mathbf{Z}$, $\varphi(G)/\varphi(G)'$ must be cyclic. Conversely, any finite group Γ with Γ/Γ' cyclic is a homomorphic image of some knot in S^3 (F. González-Acuña, 1975). If $\Gamma=S_n$ or D_{2n+1} (the †dihedral group of order $2(2n+1)$), Γ/Γ' is finite cyclic and some conditions are known for the knot group to be mapped onto Γ [5]. Given a representation φ of $G(K)$ of degree n, we obtain the n-fold †covering space C_φ of S^3-K. Equivalent representations correspond to homeomorphic covering spaces. C_φ can be completed to a covering space Σ_φ of S^3 with branch points on K. Σ_φ is an orientable closed 3-manifold and is called the n-fold branched covering space of S^3 along K. Let Λ_φ be $\Sigma_\varphi-C_\varphi$. Then Λ_φ is a link in Σ_φ. Any topological invariant of $(\Sigma_\varphi,\Lambda_\varphi)$ is an invariant of K. For example, the homology groups of Σ_φ are important invariants of K, and by using these invariants we can show that the two knots K_1 (Fig. 4) and K_2 (Fig. 5) are distinct (R. Riley, 1971). If $H_1(\Sigma_\varphi;Z)$ is a finite group, we can compute the linking number between two components of Λ_φ (R. Hartley and Murasugi, *Canad. J. Math.*, 29 (1977)). The set of these rational numbers is called the **covering linkage invariants** and is used to classify all knot types with ten double points (K. Perko, Jr. (1974)).

Since $G/G'\cong\mathbf{Z}$, every knot group has a unique representation on the cyclic group of order g. Let $\Sigma_g-\Lambda_g$ be the corresponding g-fold covering space of S^3-K, called the cyclic covering space. Then the 1-dimensional homology group $H_1(\Sigma_g-\Lambda_g)$ is isomorphic to the direct sum of $H_1(\Sigma_g)$ and $\mathbf{Z}$, and the order θ of $H_1(\Sigma_g)$ is given by $\theta=\prod_{j=0}^{g-1}\Delta(\omega^j)$, where ω denotes the primitive gth root of unity. If $H_1(\Sigma_g)$ is infinite, then the †Betti number of Σ_g is the number of roots of $\Delta(t)=0$, properly counted, that are gth roots of unity.

Since every orientable closed 3-manifold is a branched covering space of S^3 along a knot or link (Alexander (1928)), some attempt has been made to identify †simply connected 3-manifolds as branched covering spaces of S^3. However, all simply connected 3-manifolds obtained as coverings thus far are known to be S^3.

In 1979 W. Thurston proved that any g-fold ($g\geqslant 2$) branched cyclic covering space of S^3 (or of a homotopy 3-sphere) along a nontrivial knot is never simply connected. This result settles affirmatively the **Smith conjecture**: The fixed point set of any periodic self-diffeomorphism of S^3 is unknotted if it is a circle.

The proof of the conjecture is based on the study of hyperbolic structures on 3-manifolds initiated by W. Thurston (about 1977) and on the equivariant loop theorem of Meeks and Yau. A **hyperbolic manifold** is a Riemannian manifold of constant negative curvature which is complete and of finite volume. A knot (or a link) K whose complement S^3-K admits a hyperbolic-manifold structure is called a **hyperbolic knot** (or link). Two hyperbolic knots with isomorphic knot groups have homeomorphic complementary domains. A faithful representation ψ of a knot group $G(K)$ into †$PSL(2,\mathbf{C})$ is called an excellent parabolic representation if (1) $\mathrm{tr}\,\psi(m)=\pm 2$ for a meridian m and (2) $\psi(G(K))$ is a non-Abelian discrete subgroup of $PSL(2,\mathbf{C})$. According to Thurston, a knot K is hyperbolic if and only if $G(K)$ has an excellent parabolic representation.

The theory of hyperbolic knots has had a most profound influence on knot theory.

F. Braids

The theory of **braids** has been used as a tool to investigate the theory of knots. An illustration of a braid of fifth order is given in Fig. 6. In general, consider a cube D^3 written as $D^2 \times [0,1]$, where D^2 is a disk $\{(x,y) \mid 0 \leqslant x, y \leqslant 1\}$ in $\mathbf{R}^3$. Let $P_i = \left(\frac{i}{n+1}, \frac{1}{2}\right)$ be points on D^2 and define $2n$ points $A_i = P_i \times \{1\}$ and $B_i = P_i \times \{0\}$, $i = 1, 2, \ldots, n$. Then join A_i with B_{k_i}, $i = 1, 2, \ldots, n$, where $(k_1, k_2, \ldots, k_n)$ is a permutation of $(1, 2, \ldots, n)$, by means of mutually disjoint polygonal curves l_i in D^3 in such a way that (1) $l_i \cap \partial D^3 = A_i \cup B_{k_i}$ and (2) for each t, $0 \leqslant t \leqslant 1$, a disk $D_t = D^2 \times \{t\}$ intersects l_1 at exactly one point. Such a configuration is called a braid of nth order.

Two braids $\mathfrak{b}_1$ and $\mathfrak{b}_2$ are said to be **isotopic**, $\mathfrak{b}_1 \approx \mathfrak{b}_2$, if there is a homeomorphism of D^3 onto itself mapping $\mathfrak{b}_1$ onto $\mathfrak{b}_2$ such that it is an identity on ∂D^3. The isotopy relation is an equivalence relation among braids.

Suppose that we are given a braid such as Fig. 6. If we connect A_i and B_i by polygonal curves l_i' on the boundary of D^3 (Fig. 7), we obtain a **closed braid**. In general, a closed braid is a link. Conversely, any link is equivalent to a closed braid.

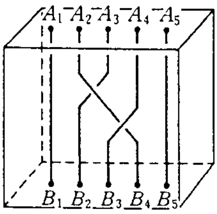


Fig. 6

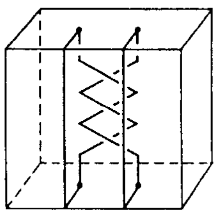


Fig. 7

The product $\mathfrak{b}_1\mathfrak{b}_2$ of two braids $\mathfrak{b}_1$ and $\mathfrak{b}_2$ is defined as the braid obtained by connecting $\mathfrak{b}_1$ and $\mathfrak{b}_2$ as shown in Fig. 8. If $\mathfrak{b}_1 \approx \mathfrak{b}_1'$ and $\mathfrak{b}_2 \approx \mathfrak{b}_2'$, then $\mathfrak{b}_1\mathfrak{b}_2 \approx \mathfrak{b}_1'\mathfrak{b}_2'$. Hence we can define the product $[\mathfrak{b}_1][\mathfrak{b}_2] = [\mathfrak{b}_1\mathfrak{b}_2]$ of equivalence classes of braids of nth order. The totality of $[\mathfrak{b}]$ forms a group $\mathfrak{B}_n$ called the **braid group** of order n. $\mathfrak{B}_n$ is generated by the equivalence classes σ_i $(i = 1, 2, \ldots, n-1)$ of braids as shown in Fig. 9(a); note σ_i^{-1} shown in Fig. 9(b). The †fundamental relations between $\{\sigma_i\}$ are $S_{ji} = 1$, where $S_{ji} = \sigma_j^{-1}\sigma_i^{-1}\sigma_j\sigma_i$ or $\sigma_j^{-1}\sigma_i^{-1}\sigma_j^{-1}\sigma_i\sigma_j\sigma_i$ according as $|j-i| \geqslant 2$ or $|i-j| = 1$.

Suppose that we are given braids $\mathfrak{b}_1, \mathfrak{b}_2$ represented as products of σ_i. Then $\mathfrak{b}_1$ and $\mathfrak{b}_2$ are equivalent if and only if the two products represent the same element in $\mathfrak{B}_n$. The problem of deciding whether $\mathfrak{b}_1$ and $\mathfrak{b}_2$ are equivalent reduces to the †word problem in $\mathfrak{B}_n$. On the other hand, the problem of deciding whether or not two closed braids are equivalent reduces

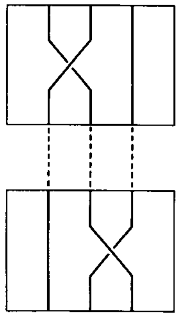


Fig. 8

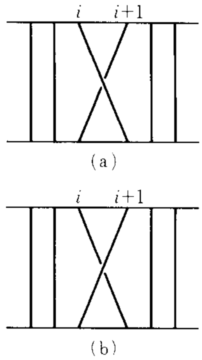


Fig. 9

to the †transformation problem in $\mathfrak{B}_n$ (→ 161 Free Groups).

The theory of braids was initiated and developed by E. Artin, who also gave a solution to the word problem in $\mathfrak{B}_n$ (*Abh. Math. Sem. Univ. Hamburg* (1926)). The transformation problem was solved by F. A. Garside (*Quart. J. Math.*, 78 (1969)).

By taking an arbitrary surface F instead of a disk D^2, we have a braid group over F, denoted by $B_n(F)$. The structure of the group $B_n(F)$ has been studied since 1962. A presentation of $B_n(F)$ for any 2-manifold is known. In particular, $B_n(F)$ has no elements of finite order except when F is a 2-sphere or a projective plane. Following the discovery of a deep connection between $B_n(F)$ and the mapping class group, the study of $B_n(F)$ has become quite important [9].

G. Higher-Dimensional Knots

The problem of knots, that is, the problem of placement of simple closed curves in $\mathbf{R}^3$, is extended to the problem of placement of q-dimensional spheres in p-space $\mathbf{R}^p$ or in a p-dimensional sphere S^p.

In this section, the explanation is restricted to the case of †combinatorial manifolds and †PL homeomorphisms between them. A similar theory can be developed for other categories (→ 114 Differential Topology).

Let D^n be the n-dimensional disk $\{(x_1, x_2, \ldots, x_n) \in \mathbf{R}^n \mid |x_i| \leqslant 1, i = 1, 2, \ldots, n\}$. We identify D^n with the subspace $D^n \times \{0\}$ of D^{n+1}. If S^q is a subcomplex of S^p, then (S^p, S^q) $(p > q)$ is called a (p,q) **sphere pair** or (p,q)**-knot**. If a q-dimensional cell B^q is a subcomplex of B^p and if $B^q \cap \dot{B}^p = \dot{B}^q$, the boundary of B^q, then (B^p, B^q) is called a (p,q) **ball pair** or (p,q)**-ball knot**. Two pairs $X = (X^p, X^q)$ and $Y = (Y^p, Y^q)$ are said to be homeomorphic if there is a homeomorphism $h: X^p \to Y^p$ such that $h(X^q) = Y^q$. We classify (p,q) ball pairs and (p,q) sphere pairs into equivalence classes via homeo-

morphisms. A (p,q) ball pair (B^p, B^q) is said to be **unknotted** (or **flat**) if it is homeomorphic to the standard pair $\Gamma^{p,q}=(D^p, D^q)$. A (p,q) sphere pair (S^p, S^q) is said to be **unknotted** if it is homeomorphic to $(\partial D^{q+1}, \partial D^{p+1})$. E. C. Zeeman showed that if $p-q\geqslant 3$, then the (p,q) ball pair and the (p,q) sphere pair are both unknotted (*Ann. Math.*, (2) 78 (1963)). Similar results have been obtained by Stallings (*Ann. Math.*, (2) 77 (1963)). Because of these results, the only interesting case is $p=q+2$. We assume further that S^q is locally flat in S^{q+2}. Let K^q be a q-knot, i.e., a q-dimensional sphere, in S^{q+2}, and let $G(K^q)=\pi_1(S^{q+2}-K^q)$, the group of K^q. Since $G(K^q)/G'\cong\mathbf{Z}$, we can define the Alexander matrix, the Alexander ideal, and the Alexander polynomial $\Delta(t)$ of K^q, as we did in Section C. An alternative description of these invariants expressed as invariants of homology groups of the infinite cyclic covering space of $S^{q+2}-K^q$ can be found in J. Levine, *Ann. Math.*, (2) 84 (1966). There are some discrepancies between 1-knot theory and q-knot theory, $q\geqslant 2$: (1) It is not known whether $\pi_1(S^4-K^2)\cong\mathbf{Z}$ implies that (S^4, K^2) is unknotted (**unknotting conjecture**). (2) $\pi_1(S^{q+2}-K^q)$ can have an element of finite order. (3) $\Delta(1)=\pm 1$, but $\Delta(t)$ may not be symmetric. (4) $\pi_2(S^4-K^2)$ may not be trivial.

Although the characterization problem of the 1-knot group $\pi_1(S^3-K)$ has not yet been solved, the same problem for the q-knot group has been completely settled by M. A. Kervaire (1963) as follows. Let G be a finitely presented group. Then G is the group of some knot K^q in S^{q+2}, $q\geqslant 3$, if and only if (i) $G/G'\cong\mathbf{Z}$, (ii) $H_2(G;\mathbf{Z})=0$, and (iii) there is an element x in G whose set of conjugates generates G. These conditions are satisfied by any q-knot group, $q\geqslant 1$, but they are not sufficient for 1-knot groups. For $q=2$, conditions (i)–(iii) are sufficient for G to be a 2-knot group in a †homotopy 4-sphere.

A q-knot K^q in S^{q+2} $(=\partial D^{q+3})$ is called **null-cobordant** if K^q is the boundary of a locally flat embedded disk in D^{q+3}. The concept of **knot cobordism** was introduced by Fox and Milnor in 1957 (*Bull. Amer. Math. Soc.*, 63 (1957)) for 1-knots and was readily extended to q-knots in S^{q+2}. Knot cobordism is a weaker equivalence than isotopy, but the set of cobordism classes forms an Abelian group C_q under the connected sum (joining the knotted spheres by a tube) or the product for 1-knots. Any 1-knot K that represents zero element in C_1 is called a **slice knot**. If K is a slice knot, then $\Delta(t)$ must be of the form $f(t)f(t^{-1})$ for some integer polynomial (Fox and Milnor (1957)) and $\sigma(K)=0$ [8]. For all $q\geqslant 1$, C_{2q-1} is not finitely generated, while $C_{2q}=0$ (Kervaire (1964)).

H. Miscellaneous Results

Results on wild knots can be found scattered throughout the mathematical literature since 1945. Many strange things can happen with wild knots or wild arcs. For example, the simple closed curve K (Fig. 10) bounds a disk, but $\pi_1(S^3-K)$ is not Abelian.

The placement problem of graphs in $\mathbf{R}^3$ has also been treated, for instance by Kinoshita (*Pacific J. Math.*, 47 (1975)).

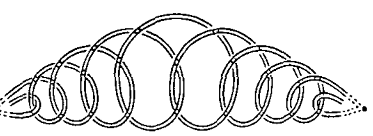


Fig. 10

References

[1] H. Seifert, Über das Geschlecht von Knoten, Math. Ann., 110 (1934), 571–592.
[2] C. D. Papakyriakopoulos, On Dehn's lemma and the asphericity of knots, Ann. Math., 66 (1957), 1–26.
[3] K. Reidemeister, Knotentheorie, Erg. Math., Springer, 1932 (Chelsea, 1948).
[4] D. Rolfsen, Knots and links, Math. lecture note 7, Publish or Perish, 1976.
[5] R. H. Fox, A quick trip through knot theory, Topology of 3-Manifolds and Related Topics, M. K. Fort, Jr. (ed.), Prentice Hall, 1962, 120–167.
[6] R. H. Crowell and R. H. Fox, Introduction to knot theory, Ginn, 1963 (Graduate Texts in Math., Springer, 1977).
[7] L. Neuwirth, Knot groups, Ann. math. studies 56, Princeton Univ. Press, 1965.
[8] K. Murasugi, On a certain numerical invariant of link types, Trans. Amer. Math. Soc., 117 (1965), 387–422.
[9] J. Birman, Braids, links, and mapping class groups, Ann. math. studies 82, Princeton Univ. Press, 1974.
[10] H. Seifert and W. Threlfall, Lehrbuch der Topologie, Teubner, 1934 (Chelsea, 1965); English translation, A textbook of topology, Academic Press, 1980.
[11] K. Murasugi, Knot theory—old and new, preprint, Univ. of Toronto, 1980.

236 (XXI.31) Kronecker, Leopold

Leopold Kronecker (December 7, 1823–December 29, 1891) was born in Liegnitz near

Breslau in Germany (now Legnica in Poland). He entered the University of Berlin in 1841 but studied at various universities throughout the country, finally studying under E. E. Kummer at Breslau. In 1845, he received his doctorate with a thesis on units of †algebraic number fields. Then he succeeded to his uncle's business in the management of banks and farms, which kept him away from publishing mathematical papers for eight years. In 1853, he published a paper on algebraic solution of equations, containing the assertions that Abelian extensions of the rational number field are contained in cyclotomic fields and that Abelian extensions of imaginary quadratic fields can be obtained using †complex multiplication. The latter is called "Kronecker's dream in his youth." It remained a conjecture until it was solved by means of †class field theory. He gave lectures at the University of Berlin, first in his capacity of academician, then as a professor, succeeding his teacher Kummer in 1883. His statement that mathematics as a whole should be based solely on the intuition of natural numbers (→ 156 Foundations of Mathematics) often brought on disputes with his colleague K. †Weierstrass. His rejection of the bold reasoning of †set theory produced anxieties for G. †Cantor. His famous statement, "Natural numbers were made by God; the rest is the work of man," can be put in contrast with the liberal statements of Cantor and R. Dedekind. Kronecker's works and lectures ranged widely over the theory of numbers, algebra, and analysis. His contribution to the theory of elliptic functions and his †limit formula for zeta functions are well known. He also did pioneering work in topology.

References

[1] K. Hensel (ed.), Leopold Kronecker's Werke I–V, Teubner, 1895–1931 (Chelsea, 1968).
[2] H. Weber, Leopold Kronecker, Jber. D. M. V., 2 (1891–1892), 5–31.

237 (IX.15) K-Theory

A. General Remarks

K-theory was introduced by M. F. Atiyah and F. Hirzebruch after the original idea was suggested by A. Grothendieck. The †Bott periodicity theorem is essential for the development of the theory. There are important applications of K-theory to †differential and algebraic topology, such as the Riemann-Roch theorems for differentiable manifolds (Atiyah and Hirzebruch [6, 7]), the solution of the vector field problem on spheres (J. F. Adams [1]), applications to immersion and embedding problems (→ 114 Differential Topology), a simple proof of the solution of the Hopf invariant problem, and the determination of J-images in stable homotopy groups of spheres (→ 202 Homotopy Theory). A review of algebraic K-theory is also included in this article.

B. Construction of $K_\Lambda(X)$

Let the basic field Λ be the real number field $\mathbf{R}$, the complex number field $\mathbf{C}$, or the quaternion field $\mathbf{H}$. Let Λ' be the †center of Λ, and X be a compact Hausdorff space. Then $\mathscr{E}_\Lambda(X)$ denotes the set of all isomorphism classes of Λ-†vector bundles over X and is a commutative semigroup under the †Whitney sum $\xi \oplus \eta$. Let $F_\Lambda(X)$ be the free Abelian group generated by the set $\mathscr{E}_\Lambda(X)$, and let $Q_\Lambda(X)$ be the subgroup of $F_\Lambda(X)$ generated by the elements of the form $\xi \oplus \eta - \xi - \eta$. Then the Abelian group $K_\Lambda(X)$ is defined as the quotient group $K_\Lambda(X) = F_\Lambda(X)/Q_\Lambda(X)$. $K_\Lambda(X)$ is called the **Grothendieck group** or **K-group**, and such a construction is called the **Grothendieck construction**. From this we obtain the canonical mapping $\theta: \mathscr{E}_\Lambda(X) \subset F_\Lambda(X) \to K_\Lambda(X)$, which is a homomorphism of the semigroups. Moreover, the pair $(K_\Lambda(X), \theta)$ is universal in the following sense: Given an Abelian group G and a semigroup homomorphism $g: \mathscr{E}_\Lambda(X) \to G$, there exists a unique group homomorphism h: $K_\Lambda(X) \to G$ such that $g = h \circ \theta$. We call h the extension of g.

Let $f: X \to Y$ be a continuous mapping from X into another compact Hausdorff space Y. Then for $\eta \in \mathscr{E}_\Lambda(Y)$, the †induced bundle $f^*(\eta) \in \mathscr{E}_\Lambda(Y)$ is defined. Since the mapping $f^*: \mathscr{E}_\Lambda(Y) \to \mathscr{E}_\Lambda(X)$ is a semigroup homomorphism, it induces a homomorphism $K_\Lambda(f): K_\Lambda(Y) \to K_\Lambda(X)$, which is the extension of $\theta \circ f^*$, so that $K_\Lambda(f) \circ \theta = \theta \circ f^*$. Usually, $K_\Lambda(f)$ is also denoted by f^*. Thus K_Λ is a †contravariant functor. According to whether $\Lambda = \mathbf{R}$, $\mathbf{C}$, or $\mathbf{H}$, the notations KO, K, or KSP are often used for K_Λ.

If $X = \{x_0\}$, the semigroup homomorphism $\mathscr{E}_\Lambda(x_0) \ni \xi \to \dim_\Lambda \xi \in \mathbf{Z}$ induces an isomorphism $K_\Lambda(x_0) \cong \mathbf{Z}$. If X is a finite CW complex with base point x_0, then the reduced group $\tilde{K}_\Lambda(X)$ is defined to be the kernel of i^*, where $i: x_0 \to X$ is the inclusion. Then we have the canonical splitting $K_\Lambda(X) \cong \mathbf{Z} \oplus \tilde{K}_\Lambda(X)$, and $\tilde{K}_\Lambda$ is a functor defined on the category of pointed finite CW complexes.

Two isomorphism classes ξ and η of $\mathscr{E}_\Lambda(X)$ are said to be **stably equivalent** if there exist trivial bundles θ_1 and θ_2 such that $\xi \oplus \theta_1 = \eta \oplus \theta_2$. An equivalence class with respect to this relation is called a **stable vector bundle**. If X is connected, then the set of all stable Λ-vector bundles can be identified with $\tilde{K}_\Lambda(X)$.

When $\Lambda = \mathbf{R}$ or $\mathbf{C}$, the †tensor product of vector bundles induces a ring structure on $K_\Lambda(X)$, and f^* becomes a ring homomorphism.

The complexification of real vector bundles $\xi \to \xi \oplus_{\mathbf{R}} \mathbf{C} = i(\xi)$ is a semigroup homomorphism $\mathscr{E}_{\mathbf{R}}(X) \to \mathscr{E}_{\mathbf{C}}(X)$ and induces a ring homomorphism $i: KO(X) \to K(X)$ such that $i \circ \theta = \theta \circ i$. If $\xi \in \mathscr{E}_{\mathbf{H}}(X)$, then ξ can be viewed as a complex vector bundle under the scalar restriction of basic field, which we shall denote by $\rho(\xi) \in \mathscr{E}_{\mathbf{C}}(X)$. The mapping $\rho: \mathscr{E}_{\mathbf{H}}(X) \to \mathscr{E}_{\mathbf{C}}(X)$ induces a homomorphism $\rho: KSP(X) \to K(X)$. Similarly, the scalar restriction from $\mathbf{C}$ to $\mathbf{R}$ induces a homomorphism $r: K(X) \to KO(X)$. All these are †natural transformations.

Let ξ be a complex vector bundle. We can formally write the †Chern class (→ 56 Characteristic Classes) $c(\xi)$ of ξ as

$$c(\xi) = \prod (1 + x_i).$$

Then the **Chern character** $ch(\xi) \in H^*(X; \mathbf{Q})$ is defined by

$$ch(\xi) = \sum \exp x_i,$$

where $\mathbf{Q}$ is the field of rational numbers. The mapping $ch: \mathscr{E}_{\mathbf{C}}(X) \to H^*(X; \mathbf{Q})$ is extended to a ring homomorphism $ch: K(X) \to H^*(X; \mathbf{Q})$. We denote by the same notation ch the ring homomorphisms $ch \circ i: KO(X) \to H^*(X; \mathbf{Q})$ and $ch \circ \rho: KSP(X) \to H^*(X; \mathbf{Q})$. These are natural transformations from the functor K_Λ to the functor $H^*(\ ; \mathbf{Q})$.

C. Cohomology Theory

$O_\Lambda(n)$ denotes $O(n)$, $U(n)$, or $Sp(n)$ according as the basic field Λ is $\mathbf{R}$, $\mathbf{C}$, or $\mathbf{H}$. Let O_Λ be the †inductive limit group with respect to the usual inclusion $O_\Lambda(n) \subset O_\Lambda(n+1)$. Provided with the weak topology, the group O_Λ becomes a †CW complex. The set of all equivalence classes of stable Λ-vector bundles corresponds bijectively to the set of †principal O_Λ-bundles. Let B_Λ be the †classifying space for the group O_Λ, X, Y be finite CW complexes with base points, $[X, Y]$ be the set of all †homotopy classes of mappings from X to Y, and $[X, Y]_0$ be the set of all †homotopy classes in the †category of pointed topological spaces. Then by the classification theorem of fiber bundles (→ 147 Fiber Bundles), we have $K_\Lambda(X) = [X, \mathbf{Z} \times B_\Lambda]$ and $\tilde{K}_\Lambda(X) = [X, \mathbf{Z} \times B_\Lambda]_0$. The space B_Λ has the structure of a weak †H-space, so that the induced group structure of the homotopy set $[X, \mathbf{Z} \times B_\Lambda]$ coincides with that of $K_\Lambda(X)$. If f is a continuous mapping, then the induced homomorphism f^* of the homotopy set $[X, \mathbf{Z} \times B_\Lambda]$ coincides with that of $K_\Lambda(X)$.

For a finite CW pair (X, A), we put $K_\Lambda^{-n}(X, A) = \tilde{K}_\Lambda(S^n(X/A))$, $n = 0, 1, 2, \ldots$, where X/A is the space obtained from X by collapsing A to a point that becomes the base point of X/A, and S^n denotes the †n-fold reduced suspension. This gives rise to a cohomology theory (indexed by nonpositive integers) (→ 201 Homology Theory, 202 Homotopy Theory).

The tensor product of vector bundles induces the following pairing, called the **cross product**:

$$K_\Lambda^{-m}(X, A) \otimes K_\Lambda^{-n}(Y, B) \to K_\Lambda^{-(m+n)}(X \times Y, X \times B \cup A \times Y).$$

When $\Lambda = \mathbf{R}$ or $\mathbf{C}$, the **cup product**

$$K_\Lambda^{-m}(X) \otimes K_\Lambda^{-n}(X) \to K_\Lambda^{-(m+n)}(X),$$

$$K_\Lambda^{-m}(X) \otimes K_\Lambda^{-n}(X, A) \to K_\Lambda^{-(m+n)}(X, A)$$

is defined as the composite of the cross product and the induced homomorphism Δ^*, where $\Delta: X \to X \times X$ is the diagonal mapping. The complexification $i: KO^{-n}(X, A) \to K^{-n}(X, A)$ preserves the cup product. The composite of

$$ch: K_\Lambda^{-n}(X, A) \to \tilde{H}^*(S^n(X/A); \mathbf{Q})$$

and the †suspension isomorphism

$$\tilde{H}^*(S^n(X/A); \mathbf{Q}) \to H^*(X, A; \mathbf{Q})$$

is denoted by

$$ch: K_\Lambda^{-n}(X, A) \to H^*(X, A; \mathbf{Q}).$$

The homomorphism ch preserves the cup product when $\Lambda = \mathbf{R}$ or $\mathbf{C}$.

D. Bott Periodicity

Let ξ_Λ be the †canonical Λ-line bundle over the Λ-projective line. The elements

$$g_{\mathbf{C}} = \theta(\xi_{\mathbf{C}}) - 1 \in \tilde{K}(S^2) = K^{-2}(x),$$

$$g_{\mathbf{H}} = \theta(\xi_{\mathbf{H}}) - 1 \in \widetilde{KSP}(S^4) = KSP^{-4}(x),$$

and

$$g_{\mathbf{R}} = g_{\mathbf{H}} \times g_{\mathbf{H}} \text{(cross product)} \in KO(S^8) = KO^{-8}(x)$$

are called **Bott generators**.

The **Bott periodicity theorem** [10] is as follows: (1) $\tilde{K}(S^2)$, $\widetilde{KSP}(S^4)$, and $\widetilde{KO}(S^8)$ are infinite cyclic groups generated by $g_{\mathbf{C}}$, $g_{\mathbf{H}}$, and $g_{\mathbf{R}}$, respectively. Moreover, $ch(g_{\mathbf{C}}) = \sigma^2$, $ch(g_{\mathbf{H}}) = \sigma^4$, and $ch(g_{\mathbf{R}}) = \sigma^8$, where $\sigma^n \in H^n(S^n; \mathbf{Z})$ is a

generator. (2) The cross products

$K^{-n}(X,A)\otimes K^{-2}(x)\to K^{-(n+2)}(X,A)$,

$KSP^{-n}(X,A)\otimes KSP^{-4}(x)\to KO^{-(n+4)}(X,A)$,

$KO^{-n}(X,A)\otimes KO^{-8}(x)\to KO^{-(n+8)}(X,A)$

are isomorphisms.

The isomorphisms

$K^{-n}(X,A)\cong K^{-(n+2)}(X,A)$,

$KSP^{-n}(X,A)\cong KO^{-(n+4)}(X,A)$,

and

$KO^{-n}(X,A)\cong KO^{-(n+8)}(X,A)$,

defined by $a\to a\times g_\Lambda$, are called **Bott isomorphisms**. They commute with f^*, the †coboundary operator δ, and ch. Identifying K^{-n} with $K^{-(n+2)}$, KSP^{-n} with $KO^{-(n+4)}$, and KO^{-n} with $KO^{-(n+8)}$, we get periodic cohomology theories $K^*=\sum_{n\in Z_2}K^n$ and $KO^*=\sum_{n\in Z_8}KO^n$, which are multiplicative cohomology theories, and ch is a multiplication-preserving homomorphism into $H^*(\ ;\mathbf{Q})$. The cohomology of a point $K_\Lambda^{-n}(x)=\tilde{K}_\Lambda(S^n)=\pi_n(B_\Lambda)=\pi_{n-1}(O_\Lambda)$ is given by Bott [11] (→ Appendix A, Table 6.IV).

E. Operations

We assume that Λ is $\mathbf{R}$ or $\mathbf{C}$. The exterior powers λ^q are basic operations in $K_\Lambda(X)$. For $\xi\in\mathscr{E}_\Lambda(X)$, the pth †exterior power of ξ, $\lambda^p(\xi)\in\mathscr{E}_\Lambda(X)$, has the following properties: $\lambda^0(\xi)=1$, $\lambda^1(\xi)=\xi$, and $\lambda^p(\xi\oplus\eta)=\sum_{q+r=p}\lambda^q(\xi)\otimes\lambda^r(\xi)$. Let G be the multiplicative group consisting of the formal power series $\in K_\Lambda(X)\{t\}$ whose constant term is 1. The assignment $\xi\to\lambda_t(\xi)=\sum\lambda^q(\xi)t^q$ gives rise to the homomorphism $\mathscr{E}_\Lambda(X)\to G$. Let $\lambda_t:K_\Lambda(X)\to G$ be its extension. The operation $\lambda^q:K_\Lambda(X)\to K_\Lambda(X)$ is defined by $\lambda_t(x)=\sum_{0\leqslant q}\lambda^q(x)t^q$.

An important series of operations ψ^k, called the **Adams operations**, is derived from the exterior powers. Put

$$\psi_{-t}(x)=-t\frac{d\lambda_t(x)}{dt}\frac{1}{\lambda_t(x)}\in K_\Lambda(X)\{t\},$$

and define $\psi^k:K_\Lambda(X)\to K_\Lambda(X)$ by $\psi_t(x)=\sum_{0\leqslant k}\psi^k(x)t^k$. When $\Lambda=\mathbf{C}$, define ψ^{-1} as the extension of $\xi\to\bar{\xi}$, where $\bar{\xi}$ is the †conjugate complex vector bundle of ξ. The operation ψ^k is a ring homomorphism preserving 1, and the relation $\psi^k\circ\psi^l=\psi^{kl}$ holds. If $\xi\in\mathscr{E}_\Lambda(X)$ is a line bundle, then $\psi^k(\theta(\xi))=(\theta(\xi))^k$. If $x\in K_\Lambda(X)$ and $ch(x)=\sum_n ch_n(x)$, where $ch_n(x)\in H^{2n}(X;\mathbf{Q})$, then

$$ch(\psi^k(x))=\sum_n k^n ch_n(x).$$

The operations λ^q and ψ^k commute with the complexification $i:KO(X)\to K(X)$. If $\beta_{\mathbf{C}}:\tilde{K}(X)\to\tilde{K}(S^2X)$ and $\beta_{\mathbf{R}}:\widetilde{KO}(X)\to\widetilde{KO}(S^8X)$ are Bott isomorphisms, then $\psi^k\circ\beta_{\mathbf{C}}=k\beta_{\mathbf{C}}\circ\psi^k$ and $\psi^k\circ\beta_{\mathbf{R}}=k^4\beta_{\mathbf{R}}\circ\psi^k$ [1].

F. Thom-Gysin Isomorphisms

Let ξ be a real oriented vector bundle of dimension n over a finite CW complex X. A reduction of the structure group $SO(n)$ of the bundle ξ to its universal covering group $Spin(n)$ is called a **spin structure** of ξ. The bundle ξ admits a spin structure if and only if $w_2(\xi)=0$, where $w_2(\xi)$ denotes the second †Stiefel-Whitney class of ξ. The bundle ξ equipped with a spin structure is called a **spin bundle**. The bundle ξ is called a **c_1-bundle** or **Spin^C bundle** if there is given a cohomology class $c_1(\xi)\in H^2(X;\mathbf{Z})$ such that $w_2(\xi)\equiv c_1(\xi)\bmod 2$. Let X^ξ be the †Thom complex of the vector bundle ξ. The group $\tilde{K}_\Lambda^*(X^\xi)$ has the structure of a $K_\Lambda^*(X)$-module. If we assume that ξ is a c_1-bundle when $\Lambda=\mathbf{C}$ and a spin bundle when $\Lambda=\mathbf{R}$, then there exists a canonical $K_\Lambda^*(X)$-module isomorphism $\varphi:K_\Lambda^n(X)\to\tilde{K}_\Lambda^{n+\dim\xi}(X^\xi)$ such that $ch\varphi(1)=\varphi'((\mathscr{A}(\xi)\exp(c_1(\xi)/2))^{-1})$ when $\Lambda=\mathbf{C}$ and $ch\varphi(1)=\varphi'(\mathscr{A}(\xi)^{-1})$ when $\Lambda=\mathbf{R}$ [6]. Here, $\varphi':H^*(X;\mathbf{Q})\to\tilde{H}^*(X^\xi;\mathbf{Q})$ is the usual †Thom-Gysin isomorphism, and $\mathscr{A}(\xi)$ is the **$\mathscr{A}$-characteristic class** of the bundle ξ defined as follows: Write the †Pontryagin class $p(\xi)$ of ξ formally as $p(\xi)=\prod(1+x_i^2)$; then the class $\mathscr{A}(\xi)$ is given by

$$\mathscr{A}(\xi)=\prod(x_i/2)/(\sinh(x_i/2)).$$

If ξ is a complex vector bundle, then its first †Chern class $c_1(\xi)$ gives a c_1-bundle structure to $r(\xi)\in\mathscr{E}_{\mathbf{R}}(X)$. In this case, the class $\mathscr{T}(\xi)=\mathscr{A}(\xi)\exp(c_1/2)$ is the **Todd class** of the complex vector bundle ξ.

G. Riemann-Roch Theorems for Differentiable Manifolds

Let M and N be connected closed differentiable manifolds. A continuous mapping f: $M\to N$ is called a **spin mapping (spin map)** if $w_1(M)=f^*w_1(N)$ and $w_2(M)=f^*w_2(N)$. If $w_1(M)=f^*w_1(N)$ and there is given a class $c_1\in H^2(M;\mathbf{Z})$ such that $w_2(M)-f^*w_2(N)\equiv c_1(\bmod 2)$, f is called a **c_1-mapping**. If we assume that f is a c_1-mapping when $\Lambda=\mathbf{C}$ and a spin mapping when $\Lambda=\mathbf{R}$, then there is a canonical homomorphism

$$f_!:K_\Lambda^n(M)\to K_\Lambda^{n+\dim N-\dim M}(N)$$

such that

$$f_!(f^*(x)\cdot y)=x\cdot f_!(y)$$

and

$ch(f_!(y))\mathscr{A}(N)=f_!(ch(y)\mathscr{A}(M)\exp(c_1/2))$

for $y\in K_\Lambda^*(M)$ and $x\in K_\Lambda^*(N)$. This is the **Riemann-Roch theorem for differentiable manifolds** [6] (→ 366 Riemann-Roch Theorems). In the second formula, the homomorphism $f_!:H^*(M;\mathbf{Q})\to H^*(N;\mathbf{Q})$ on the right-hand side is the usual †Gysin homomorphism, and if $\Lambda=\mathbf{R}$ we set $c_1=0$. The homomorphism $f_!$ depends only on the homotopy class of f and has the usual functorial properties $1_!=1$ and $(f\circ g)=f_!\circ g_!$.

H. The Atiyah-Singer Index Theorem

Let X be an n-dimensional compact differentiable manifold of class C^∞ without boundary. As we shall see later, any elliptic differential operator (or, more generally, any elliptic complex) d on X has analytic index $\operatorname{ind}_a(d)$ and topological index $\operatorname{ind}_t(d)$, the latter of which is deeply related to K-theory. The **Atiyah-Singer index theorem** asserts that $\operatorname{ind}_a(d)=\operatorname{ind}_t(d)$ [8]. We shall describe the details of the definitions and the theorem.

Let E and F be complex vector bundles of class C^∞ over X with $\dim E=s$ and $\dim F=t$ (→ 147 Fiber Bundles). Let $\Gamma(E)$ and $\Gamma(F)$ be the linear spaces over $\mathbf{C}$ consisting of C^∞-cross sections of E and F, respectively. A linear mapping d from $\Gamma(E)$ to $\Gamma(F)$ is called a **differential operator of the kth order** if d is locally expressed by some differential operator of the kth order. This means that if we choose a local coordinate neighborhood U of X and trivializations of E and F on U such that $E|U\cong U\times\mathbf{C}^s$ and $F|U\cong U\times\mathbf{C}^t$, then d is a differential operator of the kth order from $C^\infty(U,\mathbf{C}^s)$ to $C^\infty(U,\mathbf{C}^t)$. Thus d is locally expressed by the matrix form $(\sum_{|\alpha|\leqslant k}a_\alpha^{(i,j)}(x)D^\alpha)$ $(i=1,\dots,t;j=1,\dots,s)$, each component of which is a differential operator (→ 112 Differential Operators). Using this expression we define the symbol $\sigma(d)$ of d as follows: Let $T^*(X)$ be the cotangent bundle of X. Given any $\eta_x\in T_x^*(X)$, put $\sigma(d)(\eta_x)=(\sum_{|\alpha|=k}a_\alpha^{(i,j)}(x)\eta_x^\alpha)$, where η_x^α stands for $\eta_1^{\alpha_1}\dots\eta_n^{\alpha_n}$ for any multi-index $\alpha=(\alpha_1,\dots,\alpha_n)$ and for a local coordinate expression $\eta_x=(\eta_x^1,\dots,\eta_x^n)$. We call $\sigma(d)$ the **principal symbol** of d. Now a differential operator d is called **elliptic** if for each $\eta_x\neq0$ the linear mapping $\sigma(d)(\eta_x)$ gives an isomorphism from E_x onto F_x. For the elliptic differential operator d, we have $\dim\operatorname{Ker}d<\infty$ and $\dim\operatorname{Coker}d<\infty$ [5]. The **analytic index** $\operatorname{ind}_a(d)$ is defined to be the integer $\dim\operatorname{Ker}d-\dim\operatorname{Coker}d$, and it has the characteristic property that $\operatorname{ind}_a(d)$ is invariant under deformation of d.

More generally, an elliptic complex $\mathscr{E}$ on X and the analytic index $\operatorname{ind}_a(\mathscr{E})$ can be defined as follows: Given a finite number of smooth complex vector bundles $\{E_i\}_{i=1,\dots,l}$ on X and differential operators $d_i:\Gamma(E_i)\to\Gamma(E_{i+1})$, we call $\mathscr{E}=\{E_i,d_i\}_{i=1,\dots,l}$ an **elliptic complex** on X if the following two conditions are satisfied: (i) $d_{i+1}\circ d_i=0$; (ii) for any $\eta_x\in T_x^*(X)$, $\eta_x\neq0$, the (principal) symbol sequence $\to E_{i,x}\xrightarrow{\sigma(d_i)(\eta_x)}E_{i+1,x}\to$ is exact. For an elliptic complex $\mathscr{E}$, we have $\dim H^i(\mathscr{E})=\dim(\operatorname{Ker}d_i/\operatorname{Im}d_{i-1})<\infty$ [5]. The integer $\sum(-1)^i\dim H^i(\mathscr{E})$ is called the analytic index of $\mathscr{E}$. An elliptic complex with the form $0\to\Gamma(E)\xrightarrow{d}\Gamma(F)\to0$ is an elliptic operator. An important example of an elliptic complex arises from de Rham theory (→ 105 Differentiable Manifolds Q): Take $E_i=\Lambda^iT^*(X)$ and the exterior differentiations as differential operators. The elliptic complex thus obtained is the **de Rham complex**.

The topological index $\operatorname{ind}_t(\mathscr{E})$ of the elliptic complex $\mathscr{E}$ is introduced in the following way: By virtue of the exactness for $\eta_x\neq0$, the symbol sequence $\to E_{i,x}\xrightarrow{\sigma(d_i)(\eta_x)}E_{i+1,x}\to$ determines a definite element $[\sigma(d)]$ of $\tilde{K}(X^\tau)$, where X^τ is the †Thom complex associated with the cotangent bundle of X. Embed X in some Euclidean space $\mathbf{R}^N$. Then the mapping j: $T^*(X)\to T^*(\mathbf{R}^N)\cong\mathbf{R}^{2N}$ canonically induces the homomorphism $j_!:\tilde{K}(X^\tau)\to\tilde{K}((\mathbf{R}^N)^\tau)\cong\tilde{K}(S^{2N})\cong\mathbf{Z}$, and $j_!$ is obtained as in Section G from j by using a canonical complex vector bundle structure of $T^*(X)$ in $T^*(\mathbf{R}^N)$. We set $\operatorname{ind}_t(\mathscr{E})=j_![\sigma(\mathscr{E})]$ and call this the **topological index** of $\mathscr{E}$. We have

$$\operatorname{ind}_t(\mathscr{E})=ch([\sigma(d)])\mathscr{T}(X)[X^\tau],$$

where $ch([\sigma(d)])$ $(\in H^*(X^\tau;\mathbf{Q}))$ is the Chern character of $\sigma(d)$, $\mathscr{T}(X)$ $(\in H^*(X;\mathbf{Q}))$ is the Todd class of $T(X)\otimes\mathbf{C}$, and $[X^\tau]$ is the fundamental cycle of X^τ [8].

The Atiyah-Singer index theorem, in general form, asserts that $\operatorname{ind}_a(\mathscr{E})=\operatorname{ind}_t(\mathscr{E})$. For the de Rham complex E, it follows from the definition that $\operatorname{ind}_t(E)$ is equal to the Euler characteristic of X. Let X be a compact complex manifold and W be a complex analytic vector bundle on X. Applying the theorem to the Dolbeault complex with value in $W\dots\to A^{0,i}(W)\xrightarrow{\bar\partial}A^{0,i+1}(W)\to\dots$ (→ 72 Complex Manifolds), we can conclude that Hirzebruch's formulation of Riemann-Roch theorem (→ 366 Riemann-Roch Theorems B) holds not only for projective algebraic manifolds but also for compact complex manifolds. Moreover, from the index theorem we can deduce the Hirzebruch index theorem (→ 56 Characteristic Classes G) and various integrability theorems [8]. Any characteristic number which takes integral values on all oriented manifolds or weakly almost complex manifolds can be derived from the

Atiyah-Singer index theorem (R. Stong and A. Hattori).

The Hirzebruch index theorem and the Atiyah-Singer index theorem can be extended to compact manifolds with boundary in the framework of Riemannian geometry, giving a formula analogous to the Gauss-Bonnet theorem for manifolds with boundary [9].

Another generalization of the theorem is its equivariant version. Let X be a compact Hausdorff space on which a compact Lie group G acts, i.e., a †G-space. A real or complex vector bundle $\pi: E \rightarrow X$ is called a **G-vector bundle** if E is a G-space, π is an †equivariant mapping and G-actions are fiberwise linear. The set of isomorphism classes of G-vector bundles over X is an additive semigroup with respect to Whitney sums, and the Grothendieck construction on this semigroup gives an Abelian group, which is denoted by $K_G(X)$ or $KO_G(X)$, depending on the scalar field of bundles, called **equivariant K-group** of X. $K_G(X)$ and $KO_G(X)$ have commutative ring structures given by the tensor products. Now consider the case where X is a one-point space pt. Then a G-vector bundle over pt is a finite-dimensional complex or real linear representation of G. Therefore $K_G(pt)$ or $KO_G(pt)$ is the Grothendieck group of isomorphism classes of linear representations. This group also has a ring structure. It is called the representation ring of G and is denoted by $R(G)$ or $RO(G)$, respectively. When $X = G/H$, a homogeneous space of G by a closed subgroup H, the isomorphism $K_G(G/H) \approx R(H)$ (or $KO_G(G/H) \approx RO(H)$ holds true.

Let X be a compact smooth G-manifold and E and F smooth complex G-vector bundles over X, i.e., G acts smoothly on X, E, and F. A differential operator $d: \Gamma(E) \rightarrow \Gamma(F)$ is called equivariant when d commutes with induced G-actions on $\Gamma(E)$ and $\Gamma(F)$, respectively. Suppose that $d: \Gamma(E) \rightarrow \Gamma(F)$ is an equivariant elliptic differential operator. As $\Gamma(E)$ and $\Gamma(F)$ are G-modules, $\operatorname{Ker} d$ and $\operatorname{Coker} d$ are finite-dimensional G-modules, and the analytic index of d is defined by

$$\operatorname{ind}_a(d) = \operatorname{Ker} d - \operatorname{Coker} d \in R(G).$$

An argument parallel to that for the inequivariant case defines the topological index of d as

$$\operatorname{ind}_t(d) \in K_G(V) \approx K_G(pt) \approx R(G),$$

where V is a suitable finite-dimensional complex G-module. Both analytic and topological indices are also generalized for equivariant elliptic G-complexes, and the **equivariant Atiyah-Singer index theorem** asserts that

$$\operatorname{ind}_a(\mathscr{E}) = \operatorname{ind}_t(\mathscr{E}) \in R(G)$$

for any elliptic G-complex $\mathscr{E}$ [8].

The equivariant Atiyah-Singer index theorem has a close relation with the Lefschetz fixed-point theorem (of Atiyah-Bott [5]) and generalizes the Hirzebruch index theorem (also called the Hirzebruch signature theorem) to its equivariant form, the so-called G-signature theorem (→ 153 Fixed-Point Theorems).

I. J-Groups and the Adams Conjecture

Let ξ and η be real vector bundles over a finite CW complex X. ξ and η are called **fiber homotopy equivalent** if there exist fiber-preserving mappings $f: S(\xi) \rightarrow S(\eta)$, $g: S(\eta) \rightarrow S(\xi)$ between associated sphere bundles, and fiber-preserving homotopies $g \circ f \simeq 1$, $f \circ g \simeq 1$. ξ and η are called **stably fiber homotopy equivalent** if there exist trivial bundles n and m such that $\xi \oplus n$ and $\eta \oplus m$ are fiber homotopy equivalent [24]. Stable fiber homotopy equivalence is an equivalence relation, and the set of all stable fiber homotopy equivalence classes of real vector bundles over X is denoted by $J(X)$ and is called the **J-group** of X. $J(X)$ is an Abelian group with addition induced by Whitney sums of vector bundles, and we can express $J(X) = KO(X)/T(X)$ as a quotient group. The natural projection

$$J: KO(X) \rightarrow J(X) = KO(X)/T(X)$$

is called the **J-homomorphism**. When $X = S^k$, the k-sphere, the J-homomorphism can be identified with the classical stable J-homomorphism $J: \pi_{k-1}(\mathbf{O}) \rightarrow \pi_{k-1}^S$ of Hopf and Whitehead (→ 202 Homotopy Theory).

Adams [2] proposed to compute $J(X)$ by introducing two factor groups $J'(X)$ and $J''(X)$ of $KO(X)$ in such a way that these are computable and that the epimorphisms $J''(X) \rightarrow J(X) \rightarrow J'(X)$ hold whenever the following conjecture is true:

Adams conjecture. Let k be any integer and $y \in KO(X)$. There exists a nonnegative integer $e = e(k, y)$ such that $J(k^e(\psi^k - 1)y) = 0$.

Adams [2] proved this conjecture for line and plane bundles. In 1970, D. G. Quillen [25] proved this conjecture in its full generality. By intensive use of the Brauer induction theorem, Quillen reduced the problem to the case of bundles with finite structure groups and then to the case of line or plane bundles where Adams' proof [2] applies. Since then, many different proofs of the conjecture have appeared.

Adams' theory on $J(X)$ and Quillen's theorem (Adams conjecture) are utilized in determining completely the J-images in stable homotopy groups of spheres (→ 202 Homotopy Theory).

J. Algebraic K-Theory

Algebraic K-theory is a branch of algebra concerned mainly with a series of Abelian group valued functors K_n of rings (and, more generally, of certain categories), which have certain features of generalized homology theory. It originated in the K-group construction used in Grothendieck's work on the Riemann-Roch theorem. The theory was initiated in the early sixties by H. Bass, who introduced K_1 and extensively studied K_0 and K_1 in collaboration with other researchers [15–18]. Then K_2 was introduced by J. Milnor [26], and higher K-theories were constructed by Quillen and others from various viewpoints [22, I]. There is also a K-theory with respect to Hermitian structure [22, III]. Algebraic K-theory is intimately related to various other branches of mathematics, such as topology, algebraic geometry, and number theory.

The **Grothendieck group** $K_0(A)$ of a ring A is the Abelian group generated by the set of isomorphism classes $[P]$ of finitely generated †projective A-modules subject to the relation $[P \oplus P'] = [P] + [P']$ for every pair of projective modules P, P'. If A is finitely generated as a **Z**-algebra, then $K_0(A)$ is a finitely generated group. The assignment $n \to n[A]$ defines a homomorphism $\mathbf{Z} \to K_0(A)$ whose cokernel is the †projective class group of A. If A is commutative, a similar construction for the category of rank 1 projective A-modules with respect to the tensor product $\otimes$ leads to the **Picard group** $Pic(A)$. We then have an epimorphism $K_0(A) \to Pic(A)$ defined by $P \to \Lambda^r P$, where P is of rank r [16]. If X is a compact Hausdorff space, the topological $K(X)$ is isomorphic to the algebraic $K_0(A)$, where A is the ring of complex-valued continuous functions on X.

The **Whitehead group** $K_1(A)$ is defined as follows. Let $GL(A)$ be the direct limit of the sequence $\ldots \to GL_n(A) \xrightarrow{i_n} GL_{n+1}(A) \to \ldots$, where $i_n(X) = \begin{pmatrix} X & 0 \\ 0 & 1 \end{pmatrix}$. Let $E_n(A)$ be the subgroup of $GL_n(A)$ generated by all elementary matrices $1 + ae_{ij}$ $(i \neq j, a \in A)$, where the e_{ij} are matrix units. The limit $E(A)$ of $E_n(A)$ coincides with the commutator subgroup of $GL(A)$. Now define $K_1(A) = GL(A)/E(A)$ [15, 18]. For $A = \mathbf{Z}\pi$, the integral group algebra of a group π, the cokernel of the natural homomorphism $\pm\pi \to K_1(A)$ is denoted by $Wh(\pi)$. The torsion invariant of J. H. C. Whitehead is defined in $Wh(\pi)$ [20]. If A is commutative, we put $SK_1(A) = SL(A)/E(A)$, where $SL(A) = \varinjlim SL_n(A)$. By the determinant homomorphism, we have $K_1(A) \cong SK_1(A) \oplus U(A)$, where $U(A)$ is the group of units of A. $SK_1(A) = 0$ when A is a field or a local ring A. A deeper result states that $SK_1(A) = 0$ for the ring of integers of an algebraic number field (Bass, Milnor, and Serre [17]). This and related results have been applied to investigate $SK_1(\mathbf{Z}\pi)$ for a finite group π.

We define $K_2(A) = H_2(E(A); \mathbf{Z})$ (the Schur multiplier of $E(A)$) [21]. This yields a universal central extension of $E(A)$, defined by $0 \to K_2(A) \to St(A) \to E(A) \to 0$, where $St(A) = \varinjlim St_n(A)$ and $St_n(A)$ $(n \geq 3)$ is the **Steinberg group** generated by $x_{ij}(a)$ $(\in A; i, j = 1, \ldots, n, i \neq j)$ subject to the relations (i) $x_{ij}(a)x_{ij}(b) = x_{ij}(a+b)$, and (ii) the commutator $(x_{ij}(a), x_{kl}(b))$ equals $x_{il}(ab)$ for $j = k$, $i \neq l$; and equals 1 for $j \neq k$, $i \neq l$.

Let F be a field. A bimultiplicative mapping $s: F^* \times F^* \to C$ (C an Abelian group) satisfying $s(x, 1-x) = 1$ $(x \neq 0, 1)$ is called a **(Steinberg) symbol** on F. There exists a universal symbol $F^* \times F^* \to K_2(F)$ which, followed by homomorphisms $K_2(F) \to C$, yields all C-valued symbols on F. Since certain Steinberg symbols, such as the †Hilbert symbol, are important in number theory, the group $K_2(F)$ of a global or local field F is intimately related to the arithmetic of F. If F is an algebraic number field and R its ring of integers, we have an exact sequence $0 \to K_2(R) \to K_2(F) \to \coprod_{\mathfrak{p}} (R/\mathfrak{p})^* \to 0$ (the next-to-last term being the direct sum over all prime ideals $\mathfrak{p}$ of R), and $K_2(R)$ is a finite group.

Quillen [26, 27] defined **higher algebraic K-theory** based on the following topological construction. Let X be a CW complex given with a perfect normal subgroup N of $\pi_1(X)$. Attaching 2- and 3-cells suitably to X, he constructed a complex X^+ and a mapping $f: X \to X^+$ so that $\pi_1(f)$ is epimorphic, $\operatorname{Ker} \pi_1(f) = N$, and $f_*: H_*(X, f^*L) \approx H_*(X^+, L)$ for any local coefficients L over X^+. (X^+, f) is universal for pairs (Y, g), $g: X \to Y$, satisfying $\pi_1(g)(N) = 0$, i.e., there exists a mapping $g^+: X^+ \to Y$ unique up to homotopy such that $g^+ \circ f \simeq g$.

Quillen applied the above construction to $X = BGL(A)$ and $N = E(A)$, and defined $K_n(A) = \pi_n(BGL(A)^+)$ for $n \geq 1$ (the **first definition**). The universality implies that the inclusion $E(A) \subset GL(A)$ induces a mapping $BE(A)^+ \to BGL(A)^+$ which is the same as the universal covering mapping up to homotopy, which implies that $K_2(A) \approx H_2(E(A); \mathbf{Z})$, i.e., Quillen's K_2 coincides with Milnor's. It is also known that $K_3(A) \approx H_3(St(A); \mathbf{Z})$.

Quillen [22, I] defined higher algebraic K-theory for certain additive categories endowed with a class of short exact sequences, called **exact categories**. For an exact category M, he defined another category QM having the same objects as M but with morphisms changed. Making use of Segal's classifying spaces B of categories, he defined $K_i(M) = \pi_{i+1}(BQM, 0)$, where 0 is a zero object of M (the **second de-**

finition). When $M = P_A$, the exact category of finitely generated projective A-modules, $K_i(A) \approx K_i(P_A)$ [23], whereby Quillen introduced the notions of symmetric monoidal categories S and their localizations $S^{-1}S$ and showed the homotopy equivalences $BS^{-1}S_A \simeq K_0(A) \times BGL(A)^+$ and $\Omega BQP_A \simeq BS^{-1}S_A$, where $S_A = \mathrm{Iso}\, P_A$ is the subcategory of P_A whose morphisms are all isomorphisms of P_A. The second definition is used to generalize many classical results in algebraic K-theory to higher K-theory [22, I].

References

[1] J. F. Adams, Vector fields on spheres, Ann. Math., (2) 75 (1962), 603–632.
[2] J. F. Adams, On the groups $J(X)$ I–IV, Topology, 2 (1963), 181–195; 3 (1965), 137–171, 193–222; 5 (1966), 21–71.
[3] M. F. Atiyah, K-theory, Benjamin, 1967.
[4] M. F. Atiyah and R. Bott, The index problem for manifolds with boundary, Bombay Colloquium on Differential Analysis, Oxford Univ. Press (1964), 175–186.
[5] M. F. Atiyah and R. Bott, A Lefschetz fixed point formula for elliptic complexes I, Ann. Math., (2) 86 (1967), 374–407.
[6] M. F. Atiyah and F. Hirzebruch, Riemann-Roch theorems for differentiable manifolds, Bull. Amer. Math. Soc., 65 (1959), 276–281.
[7] M. F. Atiyah and F. Hirzebruch, Vector bundles and homogeneous spaces, Amer. Math. Soc., Proc. Symposia in Pure Math. (1961), 7–38.
[8] M. F. Atiyah and I. M. Singer, The index of elliptic operators I–III, Ann. Math., (2) 87 (1968), 484–604.
[9] M. F. Atiyah, V. K. Patodi, and I. M. Singer, Spectral asymmetry and Riemannian geometry I, Math. Proc. Cambridge Philos. Soc., 77 (1975), 43–69.
[10] R. Bott, Quelques remarques sur les théorèmes de périodicité, Bull. Soc. Math. France, 87 (1959), 293–310.
[11] R. Bott, The stable homotopy of the classical groups, Ann. Math., (2) 70 (1959), 313–337.
[12] R. Bott, Lectures on $K(X)$, Benjamin, 1969.
[13] P. E. Conner and E. E. Floyd, The relation of cobordism to K-theories, Lecture notes in math. 28, Springer, 1966.
[14] E. Dyer, Cohomology theories, Benjamin, 1969.
[15] H. Bass, K-theory and stable algebra, Publ. Math. Inst. HES, 22 (1964), 5–60.
[16] H. Bass and M. P. Murthy, Grothendieck groups and Picard groups of Abelian group rings, Ann. Math., (2) 86 (1967), 16–73.
[17] H. Bass, J. Milnor, and J.-P. Serre, Solution of the congruence subgroup problem for $SL_n (n \geqslant 3)$ and $Sp_{2n} (n \geqslant 2)$, Publ. Math. Inst. HES, 33 (1968), 59–137.
[18] H. Bass, Algebraic K-theory, Benjamin, 1968.
[19] R. Swan, K-theory of finite groups and orders, Lecture notes in math. 149, Springer, 1970.
[20] J. Milnor, Whitehead torsion, Bull. Amer. Math. Soc., 72 (1966), 358–426.
[21] J. Milnor, Introduction to algebraic K-theory, Ann. Math. Studies, Princeton Univ. Press, 1971.
[22] Algebraic K-theory, Battelle Inst. Conf. 1972, I–III, Lecture notes in math. 341, 342, 343, Springer, 1973.
[23] Algebraic K-theory, Evanston 1976, Lecture notes in math. 551, Springer, 1976.
[24] M. F. Atiyah, Thom complexes, Proc. London Math. Soc., 11 (1961), 291–310.
[25] D. Quillen, The Adams conjecture, Topology, 10 (1971), 67–80.
[26] D. Quillen, Cohomology of groups, Actes Congrès Intern. Math., 1970, t. 2, 47–51.
[27] J.-L. Loday, K-théorie algébrique et représentations de groupes, Ann. Sci. École Norm. Sup., 9 (1976), 309–377.

L

238 (XXI.32) Lagrange, Joseph Louis

Joseph Louis Lagrange (January 25, 1736–April 10, 1813) was born in Turin, Italy, and became an instructor at the military school there in 1753. In 1766, he was invited to Prussia by King Frederick the Great (1712–1786) and moved to Berlin, where he filled the post (formerly occupied by L. Euler) of chairman of the mathematics department in the graduate division of the University of Berlin. In 1787, he moved to Paris, where he became a professor at the recently founded Ecole Normale Supérieure; he remained in France for the rest of his life. Lagrange was chiefly responsible for the establishment of the metric system. In 1795, he became the first president of the newly established Ecole Polytechnique. During the later stages of the Napoleonic Era, he was made a count.

Mathematically, his position is between Euler and Laplace, and he is considered one of the major mathematicians of the late 18th and early 19th centuries. His notable achievements in analysis—the initiation of the †calculus of variations resulting from his research in the †isoperimetric problem, the founding of †analytical dynamics with the introduction of generalized coordinates, and the solving of the equations now known as †Lagrange's equations of motion—all have a strong algebraic flavor. Lagrange attempted to base calculus on †formal power series. He also conducted research on the solution of algebraic equations, and his work on the †permutation group of the roots of algebraic equations can be regarded as a forerunner of the achievements of Abel and Galois.

References

[1] J. L. Lagrange, Oeuvres I–XIV, M. J. A. Serret (ed.), Gauthier-Villars, 1867–1892.
[2] J. L. Lagrange, Théorie des fonctions analytiques, contenant les principes du calcul différentiel, Paris, 1797.

239 (XXI.33) Laplace, Pierre Simon

Pierre Simon Laplace (March 23, 1749–March 5, 1827) was born into a farming family in Beaumont en Auge in Normandy, France. His genius was recognized early, and in 1767 he moved to Paris, where he enjoyed the favor of J. d'Alembert. He became a professor at the Ecole Normale and the Ecole Polytechnique. During the Napoleonic Era, he was nominated for the post of Minister of the Interior; he later became a count. Following the fall of Napoleon, he became a marquis under Louis XVIII. He is often said to have lacked professional integrity, particularly in the matter of claiming priority, but on the other hand he sometimes showed independence of character and was generous to his pupils toward the close of his life. His achievements in mathematics, physics, and astronomy were so well recognized that he reached the highest social position.

Laplace's achievements reached a peak in the field of analysis, which had been initiated in the 17th century and developed in the 18th century by Euler and the mathematicians of the Bernoulli family. He applied the methods of analysis to †celestial mechanics, †potential theory, and †probability theory, obtaining remarkable results.

Without the use of formulas and in a flowing and elegant literary style [3, 5], he elucidated his various results. Concerning the origin of the solar system, in 1796 he published the nebular hypothesis—the so-called **Kant-Laplace nebular hypothesis**—which is famous as the predecessor of the theory of the evolution of the universe.

References

[1] P. S. Laplace, Oeuvres complètes I–XIV, Gauthier-Villars, 1878–1912.
[2] P. S. Laplace, Mécanique céleste, Paris, 1798–1825; English translation, Celestial mechanics I–IV, Chelsea, 1966; V (in French), Chelsea, 1969.
[3] P. S. Laplace, Exposition du système du monde, Paris, 1796.
[4] P. S. Laplace, Théorie analytique des probabilités, Courcier, 1812.
[5] P. S. Laplace, Essai philosophique sur les probabilités, Courcier, 1814.

240 (X.26) Laplace Transform

A. General Remarks

The notion of the Laplace transform can be regarded as a generalization of the notion of Dirichlet series. L. Euler applied the Laplace transform to solve certain differential equations (1737); later, independently, P. S. Laplace

applied it to solve differential and difference equations in his famous book *Théorie analytique des probabilités*, vol. 1 (1812). In this century, the Laplace transform has been used to justify Heaviside's †operational calculus, and the notion has become an important tool in applied mathematics.

Let $\alpha(t)$ be a function of †bounded variation in the interval $0 \leqslant t \leqslant R$ for every positive R. If

$$L(s)=\mathscr{L}(\alpha(t))(s)=\int_0^\infty e^{-st}d\alpha(t)$$

$$=\lim_{R\to\infty}\int_0^R e^{-st}d\alpha(t)$$

converges for some complex number s_0, then it converges for all s satisfying $\operatorname{Re}s>\operatorname{Re}s_0$. We call $L(s)$ the **Laplace-Stieltjes transform** of $\alpha(t)$. If $\alpha(t)=\int_0^t \varphi(u)du$ (where $\varphi(u)$ is †Lebesgue integrable in the interval $0 \leqslant t \leqslant R$ for every R), then we call

$$L(s)=\mathscr{L}(\varphi)(s)=\int_0^\infty e^{-st}\varphi(t)dt$$

the **Laplace transform** of $\varphi(t)$ (→ Appendix A, Table 12.I).

B. Regions of Convergence

Given a Laplace transform $L(s)$ of $\alpha(t)$, there exists a real number (or $\pm\infty$) σ_c such that the maximal region of convergence of $L(s)$ is the set of all s such that $\operatorname{Re}s>\sigma_c$. In extreme cases, when the integral never converges, we write $\sigma_c=+\infty$, and when it converges everywhere, we write $\sigma_c=-\infty$. The number σ_c is called the **abscissa of convergence** of $L(s)$, and the line $\operatorname{Re}s=\sigma_c$ the **axis of convergence** of $L(s)$. A formula to determine the abscissa of convergence in terms of $\alpha(t)$ is known. If $k=\limsup_{t\to\infty}(\log|\alpha(t)|)/t\neq 0$, then $\sigma_c=k$; if $k=0$ and $\alpha(t)$ does not converge as t tends to infinity, then $\sigma_c=0$. If $L(s)$ has a nonnegative abscissa σ_c, then $\sigma_c=\limsup_{t\to\infty}(\log|\alpha(t)|)/t$, and if $\sigma_c<0$, then $\sigma_c=\limsup_{t\to\infty}(\log|\alpha(\infty)-\alpha(t)|)/t$ (E. Landau, S. Pincherle). Generally,

$$\limsup_{t\to\infty}\frac{\log|\alpha(t)-\alpha([t])|}{t}=\sigma_c,$$

where [ ] is the †Gauss symbol (K. Kurosu, T. Kojima, T. Ugaeri, K. Knopp).

If $\int_0^\infty e^{-st}|d\alpha(t)|<\infty$, then the Laplace-Stieltjes integral $L(s)=\int_0^\infty e^{-st}d\alpha(t)$ is said to be **absolutely convergent**. There exists a real number σ_a such that $L(s)$ converges absolutely for $\operatorname{Re}s>\sigma_a$ and does not converge absolutely for $\operatorname{Re}s<\sigma_a$. We call σ_a the **abscissa of absolute convergence** of $L(s)$. There exists a real number σ_u such that $L(s)$ converges uniformly for $\operatorname{Re}s\geqslant\sigma_u+\varepsilon$ (for every $\varepsilon>0$) and fails to do so for $\operatorname{Re}s\geqslant\sigma_u-\varepsilon$. We call σ_u the **abscissa of uniform convergence** of $L(s)$. It is clear that $\sigma_c\leqslant\sigma_u\leqslant\sigma_a$. Formulas determining σ_u and σ_a are analogous to the Dirichlet series formulas given by Kojima and M. Kunieda (→ 121 Dirichlet Series B) [1].

C. Regularity

In the region of convergence $\operatorname{Re}s>\sigma_c$, $L(s)=\int_0^\infty e^{-st}d\alpha(t)$ is †holomorphic, and we have $L^{(k)}(s)=\int_0^\infty e^{-st}(-t)^k d\alpha(t)$ in $\operatorname{Re}s>\sigma_c$. If $\alpha(t)$ is monotonic, then the real point $s=\sigma_c$ on the axis of convergence is a singular point of $L(s)$. However, there may not be any singular point on the axis of convergence in general. The **abscissa of regularity** is the infimum of all σ such that $L(s)$ is holomorphic in $\operatorname{Re}s>\sigma$. Also, $L(\sigma+i\tau)=O(|\tau|)$ uniformly in $\sigma_c+\delta\leqslant\sigma<\infty$ for every positive δ as $|\tau|\to\infty$. Any analytic function that is holomorphic at ∞ can be represented by a Laplace transform. To be precise, let $f(s)=f(\infty)+\sum_{n=0}^\infty(a_n n!)/s^{n+1}$ ($|s|>c$). Then the function $\varphi(t)=\sum_{n=0}^\infty a_n t^n$ is entire, $f(s)=f(\infty)+\int_0^\infty e^{-st}\varphi(t)dt$, and $\sigma_c>c$.

D. Inversion Formulas

We say that a function $\alpha(t)$ ($t\geqslant 0$) that is of †bounded variation on any interval is normalized if $\alpha(0)=\alpha(+0)=0$ and $\alpha(t)=(\alpha(t+0)+\alpha(t-0))/2$. A normalized function $\alpha(t)$ ($t>0$) is uniquely determined by its Laplace transform. Moreover, the **inversion formula** determining $\alpha(t)$ in terms of $L(s)$ is known. That is, if $L(s)=\int_0^\infty e^{-st}d\alpha(t)$, then for $c>\max(\sigma_c,0)$,

$$\lim_{T\to\infty}\frac{1}{2\pi i}\int_{c-iT}^{c+iT}\frac{L(s)}{s}e^{st}ds=\begin{cases}\alpha(t), & t>0,\\ \alpha(+0)/2, & t=0,\\ 0, & t<0.\end{cases}$$

The integral on the left-hand side is often called the **Bromwich integral**. Suppose that $\alpha(t)=\int_0^t\varphi(u)du$, $L(s)$ converges absolutely on $\operatorname{Re}s=c$, and $\varphi(u)$ is of bounded variation in a neighborhood of $u=t$ ($t\geqslant 0$). Then we have

$$\lim_{T\to\infty}\frac{1}{2\pi i}\int_{c-iT}^{c+iT}L(s)e^{st}ds$$

$$=\begin{cases}(\varphi(t+0)+\varphi(t-0))/2, & t>0,\\ \varphi(+0)/2, & t=0,\\ 0, & t<0.\end{cases}$$

There is another form of the inversion formula by E. L. Post and D. V. Widder. Namely, set

$$L_{k,t}[f(x)]=(-1)^k f^{(k)}(k/t)(k/t)^{k+1}/k!$$

for a C^∞-function $f(x)$ ($t>0$, k is a positive

integer). Then

$$\lim_{k\to\infty}\int_0^t L_{k,u}[L(x)]\,du=\alpha(t)-\alpha(+0).$$

If $\alpha(t)=\int_0^t\varphi(u)\,du$, then for almost all t (>0),

$$\lim_{k\to\infty}L_{k,t}[L(x)]=\varphi(t)$$

(→ Appendix A, Table 12.I).

E. Representation Theorem

If $f(x)$ is of class C^∞ on (a,b) and $(-1)^k f^{(k)}(x)\geqslant 0$ for $k=1,2,\ldots$, then $f(x)$ is called **completely monotonic** in (a,b). Moreover, if $f(x)$ is continuous on $[a,b]$, then $f(x)$ is called completely monotonic on $[a,b]$. A necessary and sufficient condition for a function $f(x)$ to be completely monotonic in $0\leqslant x<\infty$ is that $f(x)=\int_0^\infty e^{-xt}\,d\alpha(t)$, where $\alpha(t)$ is bounded and increasing and the integral converges for $0\leqslant x<\infty$ (**Bernshteĭn's theorem**). A necessary and sufficient condition for $f(x)$ to be representable in the form $\int_0^\infty e^{-xt}\varphi(t)\,dt$, where $\varphi(t)\in L_p(0,\infty)$ $(p>1)$, is that (i) $f(x)$ have derivatives of all orders in $0<x<\infty$, (ii) $f(x)$ vanish at infinity, and (iii) there exist a constant M such that $\int_0^\infty|L_{k,t}[f(x)]|^p\,dt<M$ for $k=1,2,3,\ldots$. In the representation $f(x)=\int_0^\infty e^{-xt}\varphi(t)\,dt$, a necessary and sufficient condition for $\varphi(t)$ to be bounded in $0<t<\infty$ is that $f(x)$ is of class C^∞ in $0<x<\infty$ and there exists a constant M such that $|L_{k,t}[f(x)]|<M$ and $|xf(x)|<M$ for $0<x<\infty$. In order that $\varphi(t)\in L_1$ $(0\leqslant t<\infty)$, it is necessary and sufficient that (i) $f(x)$ be of class C^∞ in $0<x<\infty$, (ii) $f(x)$ vanish at infinity, and (iii) $\int_0^\infty|L_{k,t}[f(x)]|\,dt<\infty$ and

$$\lim_{\substack{j\to\infty\\k\to\infty}}\int_0^\infty|L_{k,t}[f(x)]-L_{j,t}[f(x)]|\,dt=0$$

(Widder).

F. Operations on Laplace Transforms

Let the Laplace transform of $f(x)$ be $\mathscr{L}(f(t))(s)=\int_0^\infty e^{-st}f(t)\,dt$. It is important in operational calculus to know the formula for the Laplace transform of φf, where φ is an operation. We mention here some important formulas:

$$\mathscr{L}(f(at-b))(s)=\frac{1}{a}\exp\left(-\frac{b}{a}s\right)\mathscr{L}(f(t))(s/a)$$

$$(f(at-b)=0 \text{ if } at<b,\ a>0,\ b\geqslant 0);$$

$$\mathscr{L}\left(\int_0^t f(\tau)\,d\tau\right)(s)=\frac{1}{s}\mathscr{L}(f(t))(s)$$

$$(\operatorname{Re}s>\max(0,\sigma_c));$$

and

$$\mathscr{L}(f'(t))(s)=s\mathscr{L}(f)(s)-f(0),$$

provided that $\mathscr{L}(f'(t))(s)$ converges at s (>0) and $f(t)\to f(0)$ as $t\to+0$. Generally speaking, if $f(+0),\ldots,f^{(k-1)}(+0)$ exist and the Laplace transforms $\mathscr{L}(f^{(k)}(t))(s)$ converge at s (>0), then

$$\mathscr{L}(f^{(k)}(t))(s)=s^k\mathscr{L}(f(t))(s)-f(+0)s^{k-1}-f'(+0)s^{k-2}-\ldots-f^{(k-1)}(+0).$$

Furthermore, given functions f_1 and f_2, if $\mathscr{L}(f_1)(s)$ and $\mathscr{L}(f_2)(s)$ are both convergent and if one of them converges absolutely or $\mathscr{L}(f_1*f_2)(s)$ converges, then

$$\mathscr{L}(f_1*f_2)=\mathscr{L}(f_1)\mathscr{L}(f_2).$$

G. Asymptotic Properties of the Laplace Transform

If $L(s)=\int_0^\infty e^{-st}\,d\alpha(t)$ $(s>0)$, then for any $c\geqslant 0$ and any constant A, we have

$$\limsup_{s\to+0}|s^cL(s)-A|\leqslant\limsup_{t\to\infty}|\alpha(t)t^{-c}\Gamma(c+1)-A|,$$

$$\limsup_{s\to\infty}|s^cL(s)-A|\leqslant\limsup_{t\to+0}|\alpha(t)t^{-c}\Gamma(c+1)-A|.$$

In particular, if we set $c=0$ and assume that $\alpha(t)\to A$ as $t\to\infty$, then $f(s)\to A$ as $s\to+0$; and if we assume $\alpha(t)\sim At^c/\Gamma(c+1)$ as $t\to\infty$ (or $t\to+0$), then $f(s)\sim As^{-c}$ as $s\to+0$ (or $s\to\infty$). These results are called **Abelian theorems**, because if we choose $\alpha(t)$ appropriately and change variables, we get †Abel's continuity theorem on power series: If $\sum a_n$ converges to s, then $\sum a_nx^n$ tends to s as x tends to $1-0$. More generally, we have the following theorem: If $\int_0^\infty e^{-st}\,d\alpha(t)=L(s)\to A$ as $s\to+0$, then $\lim_{t\to\infty}\alpha(t)=A$ if and only if $\beta(t)=\int_0^t u\,d\alpha(u)=o(t)$ $(t\to\infty)$.

H. Bilateral Laplace Transform

If $\alpha(t)$ is of bounded variation in every finite interval and if for some s

$$\lim_{R\to\infty}\int_0^R e^{-st}\,d\alpha(t),\qquad\lim_{R'\to\infty}\int_{-R'}^0 e^{-st}\,d\alpha(t)$$

exist, we set $L(s)=\int_{-\infty}^\infty e^{-st}\,d\alpha(t)$.

If $L(s)$ converges at $s_1=\sigma_1+i\tau_1$, $s_2=\sigma_2+i\tau_2$, then $L(s)$ converges in the vertical strip $\sigma_1<\operatorname{Re}s<\sigma_2$. If $L(s)$ converges in the strip $\sigma_c'<\operatorname{Re}s<\sigma_c''$ and diverges for $\operatorname{Re}s>\sigma_c''$ and $\operatorname{Re}s<\sigma_c'$, then each of the numbers σ_c' and σ_c'' is called

an **abscissa of convergence** of $L(s)$. If

$$\limsup_{t\to\infty}\frac{\log|\alpha(t)|}{t}=k\neq 0,$$

$$\liminf_{t\to-\infty}\frac{\log|\alpha(t)|}{t}=l\neq 0$$

with $k<l$, then k and l are the abscissas of convergence. If $\alpha(t)$ is a normalized function of bounded variation in every finite interval and $L(s)$ converges in the strip $k<\mathrm{Re}\,s<l$, then for all t

$$\lim_{T\to\infty}\frac{1}{2\pi i}\int_{c-iT}^{c+iT}\frac{L(s)}{s}e^{st}\,ds$$
$$=\begin{cases}\alpha(t)-\alpha(-\infty), & c>0, \quad k<c<l,\\ \alpha(t)-\alpha(\infty), & c<0, \quad k<c<l.\end{cases}$$

Suppose that $\alpha(t)=\int_0^t\varphi(u)\,du$, $L(s)$ converges absolutely on the line $\mathrm{Re}\,s=c$, and $\varphi(t)$ is of bounded variation in some neighborhood of $t=t_0$. Then

$$\lim_{T\to\infty}\frac{1}{2\pi i}\int_{c-iT}^{c+iT}L(s)e^{st_0}\,ds$$
$$=(\varphi(t_0+0)+\varphi(t_0-0))/2.$$

There are other formulas analogous to those for the ordinary Laplace transform.

I. Application to the Theory of Semigroups of Operators

Applying a Laplace transform to a one-parameter semigroup of bounded operators $\{U(t)\}_{t\geqslant 0}$ on a Banach space yields a natural correspondence between the infinitesimal generator A of $\{U(t)\}$ and its resolvent $(s-A)^{-1}$. Namely, given a continuous one-parameter semigroup $\{U(t)\}_{t\geqslant 0}$ on a Banach space X (→ 378 Semigroups of Operators and Evolution Equations), the infinitesimal generator A is defined by $Ax=\lim_{h\to+0}h^{-1}(U(h)-I)x$, and A is a densely defined closed operator on X for which there exist some positive numbers M and β such that

$$\|(s-A)^{-n}\|\leqslant M(\mathrm{Re}\,s-\beta)^{-n} \qquad (n=1,2,\dots),$$

is valid, provided that $\mathrm{Re}\,s>\beta$. The Laplace transform of $\{U(t)\}_{t\geqslant 0}$ is defined by

$$\mathscr{L}(U(t))(s)x=\int_0^\infty e^{-st}U(t)x\,dt$$
$$(x\in X,\ \ \mathrm{Re}\,s>\beta),$$

and the following identity is valid:

$$\mathscr{L}(U)(s)=(s-A)^{-1}.$$

Furthermore, the inversion formula

$$U(t)x=\lim_{T\to\infty}\frac{1}{2\pi i}\int_{c-iT}^{c+iT}e^{st}(s-A)^{-1}x\,ds$$
$$(c>\beta,\ t>0)$$

holds for every x in the domain of A, and the convergence is compact-uniform in $t>0$. If A is bounded, then $U(t)=\sum_{k=0}^{\infty}t^kA^k/k!$ (norm convergent), and we can write $U(t)=e^{tA}$. If A is not bounded, $U(t)$ is still regarded as an exponential function of tA, because we have $U(t)=\lim_{n\to\infty}(1+tA/n)^{-n}$ or $U(t)=\lim_{n\to\infty}e^{-tA_n}$, $A_n=A(1+tA/n)^{-1}$ being bounded (both strongly convergent). Hence the correspondence between $\{U(t)\}$ and $(s-A)^{-1}$ is considered a generalization of the formula

$$\mathscr{L}(e^{\alpha t})(s)=(s-\alpha)^{-1} \qquad (\mathrm{Re}\,s>\beta).$$

In order for us to get the inversion formula for every $x\in X$, $\{U(t)\}$ must satisfy some additional conditions. Namely, if $\{U(t)\}$ is a holomorphic semigroup, the inversion formula is valid for every $x\in X$, $t\geqslant 0$.

J. Laplace Transform of Distributions

The Laplace transform in $\mathbf{R}^n_x$ is defined by

$$\mathscr{L}(f)(\xi+i\eta)=\int_{\mathbf{R}^n_x}e^{-x(\xi+i\eta)}f(x)\,dx=\mathscr{F}(e^{-x\xi}f)(\eta)$$

when $e^{-x\xi}f(x)\in L_1$ (→ 160 Fourier Transform). For $n=1$, this definition is equivalent to the bilateral Laplace transform. For a given function $f(x)$, the set Δ of ξ for which $e^{-x\xi}f$ belongs to L_1 is convex in $\mathbf{R}^n_\xi$. $g(\zeta)=g(\xi+i\eta)=\mathscr{L}(f)(\xi+i\eta)$ is holomorphic in $\mathring{\Delta}+i\mathbf{R}^n_\eta$ provided that $\mathring{\Delta}$ is not empty. Differentiation under the integral sign gives

$$D^\alpha_\zeta g(\zeta)=\int_{\mathbf{R}^n_x}e^{-x\zeta}(-x)^\alpha f(x)\,dx,$$

and the integral converges uniformly on $K+i\mathbf{R}^n_\eta$, K being a compact subset of $\mathring{\Delta}$. Especially, if $e^{-x\xi}f(x)$ belongs to $\mathscr{S}(\mathbf{R}^n)$ for some ξ, then $\mathscr{L}(f)(\xi+i\eta)$ is defined on $\Delta_f+i\mathbf{R}^n_\eta$, where $\Delta_f=\{\xi\mid e^{-x\xi}f(x)\in\mathscr{S}(\mathbf{R}^n_x)\}$, which is also convex, $\mathscr{L}(f)(\xi+i\eta)$ belongs to $\mathscr{S}(\mathbf{R}^n)$ for every fixed $\xi\in\Delta_f$. Hence the definition of Laplace transform can be extended to some class of distributions. Namely, for every distribution $T\in\mathscr{D}'(\mathbf{R}^n_x)$, the set

$$\Delta_T=\{\xi\in\mathbf{R}^n\mid e^{-x\xi}T\in\mathscr{S}'(\mathbf{R}^n_x)\}$$

is convex. Thus the Laplace transform of T for which Δ_T is not empty is defined by

$$\mathscr{L}(T)(\xi+i\eta)=\mathscr{F}(e^{-x\xi}T)(\eta).$$

This definition can be rewritten by the use of test functions as follows. For any $\varphi\in\mathscr{S}(\mathbf{R}^n_\eta)$, $\langle L(T),\varphi\rangle_\eta=\langle\mathscr{F}(e^{-x\xi}T),\varphi\rangle_\eta=\langle e^{-x\xi}T,\mathscr{F}(\varphi)\rangle_x=\langle T,\int_{\mathbf{R}^n}e^{-x(\xi+i\eta)}\varphi(\eta)\,d\eta\rangle_x$. Moreover, if Δ_T has an interior point, $\mathscr{L}(T)$ has an explicit expression: take a ξ in $\mathring{\Delta}_T$, and let $0<\varepsilon\leqslant\mathrm{dist}(\xi,\partial\Delta_T)$; then $e^{-\varepsilon\langle x\rangle}\in\mathscr{S}(\mathbf{R}^n_x)$, $e^{\varepsilon\langle x\rangle}T\in\mathscr{S}'(\mathbf{R}^n_x)$, and $\mathscr{L}(T)(\xi+i\eta)=\langle T,e^{-x(\xi+i\eta)}\rangle_x=$

$\langle e^{\varepsilon\langle x\rangle}T, e^{-\varepsilon\langle x\rangle - x(\xi+i\eta)}\rangle_x$, where $\langle x\rangle = \sqrt{1+|x|^2}$.

In the following, assume that $T\in\mathscr{D}'(\mathbf{R}_x^n)$ and $\Omega=\mathring{\Delta}_T$ is not empty. Then $h(\zeta)=\mathscr{L}(T)(\xi+i\eta)$ is (i) holomorphic in $\Omega+i\mathbf{R}_\eta^n$ and (ii) slowly increasing in η. More precisely, for any compact subset K of Ω, there exist a positive number C_K and an integer m_K such that for every $\xi+i\eta\in K+i\mathbf{R}_\eta^n$, the following inequality holds:

$$|h(\xi+i\eta)|\leqslant C_K(1+|\eta|)^{m_K}.$$

Conversely, for a nonempty open convex set Ω and a function $h(\xi+i\eta)$ defined on $\Omega+i\mathbf{R}_\eta^n$ satisfying conditions (i) and (ii) above, the corresponding inverse Laplace transform of h is

$$\mathscr{L}^{-1}(h)(x)=e^{x\xi}\mathscr{F}_\eta^{-1}(h)(x) \qquad (\xi\in\Omega),$$

the right-hand side being independent of ξ in Ω. $T=\mathscr{L}^{-1}(h)$ is a distribution and $\mathring{\Delta}_T\supset\Omega$. Moreover, $\mathscr{L}(T)=h$ in $\Omega+i\mathbf{R}_\eta^n$. The following identities hold as in the case of functions:

$$\mathscr{L}\left(\left(\frac{\partial}{\partial x}\right)^\alpha T\right)(\zeta)=\zeta^\alpha\mathscr{L}(T)(\zeta),$$

$$\mathscr{L}(x^\alpha T)(\zeta)=\left(-\frac{\partial}{\partial\zeta}\right)^\alpha\mathscr{L}(T).$$

Roughly speaking, differentiability of T is reflected in the decreasing order of $\mathscr{L}(T)$ as $|\eta|\to\infty$.

Let Γ be a cone in $\mathbf{R}_\xi^n$ and B_r be the ball with center O and radius r in $\mathbf{R}_x^n$. The dual cone of Γ is denoted by Γ'. Assume that the support of T is contained in $B_r+\Gamma'$, and $\Omega=\mathring{\Delta}_T$ is not empty. Then (i) $\Omega+\Gamma\subset\Omega$, i.e., Ω extends in the direction of Γ, and (ii) for any $\xi^\circ\in\Omega$ and any $\varepsilon>0$, there exist some C and m such that for any $\xi+i\eta\in\xi^\circ+\Gamma+i\mathbf{R}_\eta^n$, we have

$$|\mathscr{L}(T)(\xi+i\eta)|\leqslant Ce^{(r+\varepsilon)|\xi|}(1+|\eta|)^m.$$

Conversely, if there exist a cone Γ, a domain Ω, and a holomorphic function $h(\zeta)$ on $\Omega+i\mathbf{R}_\eta^n$ satisfying (i) and (ii) with $h(\zeta)$ replacing $\mathscr{L}(T)$, then the support of $\mathscr{L}^{-1}(h)$ is contained in $B_r+\Gamma'$.

The convolution $f*T$ of a function f in $\mathscr{S}(\mathbf{R}_x^n)$ and a distribution T in $\mathscr{S}'(\mathbf{R}_x^n)$ belong to $\mathscr{S}'(\mathbf{R}_x^n)$, and if $\mathring{\Delta}_f\cap\mathring{\Delta}_T\neq\varnothing$, then $\mathring{\Delta}_{f*T}\supset\mathring{\Delta}_f\cap\mathring{\Delta}_T$ and

$$\mathscr{L}(f*T)=\mathscr{L}(f)\mathscr{L}(T) \text{ in } \mathring{\Delta}_f\cap\mathring{\Delta}_T+i\mathbf{R}_\eta^n.$$

More generally, convolution of two distributions T and S for which $\mathring{\Delta}_T\cap\mathring{\Delta}_S=\Omega\neq\varnothing$ holds is defined by

$$T*S=\mathscr{L}^{-1}(\mathscr{L}(T)\mathscr{L}(S)).$$

The formula $\mathscr{L}(T*S)=\mathscr{L}(T)\mathscr{L}(S)$ is valid by virtue of the definition.

If $\Omega+\Gamma\subset\Omega$ for some cone Γ, we have

$$\operatorname{supp}(T*S)\subset\operatorname{supp}T+\operatorname{supp}S+\Gamma'.$$

K. Moment Problem

Given a real sequence $\{\mu_n\}_0^\infty$, the problem of finding a function $\alpha(t)$ of bounded variation normalized on the interval $[0,1]$ satisfying $\mu_n=\int_0^1 t^n\,d\alpha(t)$ $(n=0,1,\dots)$, is called **Hausdorff's moment problem**. Such a function, if it exists, is unique. This problem is a discrete analog of the Laplace inversion formula, for if we put $t=e^{-u}$ and $n=s$, we have $\mu_s=\int_0^\infty e^{-su}d[-\alpha(e^{-u})]$.

Let $\lambda_{m,n}=\binom{m}{n}(-1)^{m-n}\Delta^{m-n}\mu_n$ $(m,n=0,1,\dots)$, where $\Delta^k\mu_n=\sum_{l=0}^k(-1)^l\binom{k}{l}\mu_{n+k-l}$ $(k=0,1,\dots)$. Then the problem has a solution $\alpha(t)$ if and only if there exists a positive number L such that $\sum_{n=0}^m|\lambda_{m,n}|\leqslant L$ $(m=0,1,\dots)$. Moreover, $\alpha(t)$ is nondecreasing if and only if $\lambda_{m,n}\geqslant 0$. In this case $\{\mu_n\}$ is called **completely monotonic**. When $d\alpha(t)=\varphi(t)dt$, $\varphi(t)$ belongs to $L^p(0,1)$ $(p>1)$ if and only if $(m+1)^{p-1}\sum_{n=0}^m|\lambda_{m,n}|^p\leqslant L$ $(m=0,1,\dots)$, and $\varphi(t)$ is bounded if and only if $(m+1)|\lambda_{m,n}|\leqslant L$ $(m,n=0,1,\dots)$.

Given a real sequence $\{\mu_n\}_0^\infty$, the problem of finding a nondecreasing function $\alpha(t)$ on $\mathbf{R}$ satisfying $\mu_n=\int_{-\infty}^\infty t^n\,d\alpha(t)$ $(n=0,1,\dots)$ is called **Hamburger's moment problem**, and the similar problem obtained by replacing the condition by $\mu_n=\int_0^\infty t^n\,d\alpha(t)$ is called **Stieltjes's moment problem**. In these problems uniqueness is not valid. The following is a counterexample for the Stieltjes problem given by Stieltjes himself: $\alpha_1(t)=\int_0^t\exp(-u^{1/4})du$ and $\alpha_2(t)=\int_0^t\exp(-u^{1/4})[1-\sin(u^{1/4})]\,du$ $(t>0)$ are nondecreasing in $(0,\infty)$, and both correspond to the sequence $\mu_n=4(4n+3)!$ in Stieltjes's problem. Carleman showed that the condition $\sum\mu_{2n}^{-1/2n}=\infty$ is sufficient for uniqueness in Hamburger's problem. Hamburger's problem has a solution if and only if the quadratic forms $\sum_{i,j=0}^n\mu_{i+j}\xi_i\xi_j$ $(n=0,1,\dots)$ are all non-negative definite. Stieltjes's problem has a solution if and only if the quadratic forms $\sum_{i,j=0}^n\mu_{i+j}\xi_i\xi_j$ and $\sum_{i,j=0}^n\mu_{i+j+1}\xi_i\xi_j$ $(n=0,1,\dots)$ are non-negative definite. If, in Hamburger's or Stieltjes's problem, we require only that the solution is a function of bounded variation instead of a nondecreasing one, then every sequence $\{\mu_n\}_0^\infty$ has a solution (R. P. Boas, Jr.).

References

[1] D. V. Widder, Laplace transform, Princeton Univ. Press, 1941.
[2] G. Doetsch, Theorie und Anwendung der Laplace-Transformation, Springer, 1937.
[3] B. Van der Pol and H. Bremmer, Operational calculus based on the two-sided Laplace integral, Cambridge Univ. Press, second edition, 1964.

[4] N. Dunford and J. T. Schwartz, Linear operators, pt. 1, General theory, Interscience, 1958.
[5] K. Yosida, Functional analysis, Springer, sixth edition, 1980.
[6] J. Leray, Hyperbolic differential equations, Princeton Lecture notes, Inst. for Adv. Study, 1952.
[7] L. Schwartz, Théorie des distributions, Hermann, 1966.
[8] J. A. Shohat and J. D. Tamarkin, The problem of moments, Amer. Math. Soc. Math. Surveys, no. 1, 1943.

241 (XVI.11) Latin Squares

A. Definition and Classification

A **latin square** over a set $A=\{a_1, a_2, \ldots, a_n\}$, or of order n, is an arrangement of elements of A in a square of side n such that each symbol in A occurs exactly once in each row and in each column. It is in **standard form**, or **reduced**, if the first row and the first column consist of the natural permutation. There are three different interpretations of latin squares. (1) If we identify the set of indices with A and write $z=x \circ y$ to indicate that the symbol at the row x and the column y is z, then we have a quasigroup, denoted by $(A, \circ)$ (→ Section C). (2) The set of the n^2 triplets xyz in the above relation is an †error-detecting code over the alphabet A, of word length 3 with the minimum Hamming distance 2. (3) The set of the n permutations P_i moving the natural permutation to the ith row satisfies the condition that $P_i^{-1} P_j$ is a discordant permutation if $i \neq j$.

Two latin squares L and L' are **isotopic** if under convention (2) there are three permutations p, q, r of A such that $L=\{xyz\}$, $L'=\{x'y'z'\}$, $x'=p(x)$, $y'=q(y)$, $z'=r(z)$. An isotopy class can contain more than one reduced latin square. The number of isotopy classes is denoted by L_n^*. If we admit another equivalence due to rearrangement of the three components of the words in the code (2), then we have **main classes**; the number of main classes is denoted by L_n^{**}. The total number of latin squares of order n is given by $n!(n-1)!L_n$ for the number L_n of reduced latin squares.

It is known that $L_1=L_2=L_3=1$, $L_4=4$, $L_5=56$, $L_6=9408$, $L_7=16{,}942{,}080$, $L_8=535{,}281{,}401{,}856$, $L_9=377, 597{,}570{,}964{,}258{,}816$; $L_1^*=L_2^*=L_3^*=1$, $L_4^*=L_5^*=2$, $L_6^*=22$, $L_7^*=564$, $L_8^*=1{,}676{,}257$; $L_1^{**}=L_2^{**}=L_3^{**}=1$, $L_4^{**}=L_5^{**}=2$, $L_6^{**}=12$, $L_7^{**}=147$.

B. Orthogonality

Suppose that we superimpose two latin squares $L=(A, \circ)$ and $L'=(A, \bullet)$, and if there is no coincidence for the pairs of the symbols, i.e., if $x \circ y=x' \circ y'$ and $x \bullet y=x' \bullet y'$ implies $x=x'$, $y=y'$, then L and L' are **mutually orthogonal**, and the resulting square is an **Euler square**. Actually there exists a pair of mutually orthogonal latin squares of order n, provided that $n \neq 2$ or 6; this was established by Bose, Shrikhande, and Parker only in 1959, contrary to a long-standing conjecture of Euler himself that such a set does not exist if $n \equiv 2 \pmod 4$. A set of mutually orthogonal latin squares of order n cannot contain more than $n-1$ squares, and a set of $n-1$ squares is a **complete set** of mutually orthogonal latin squares. This is equivalent to an error-correcting code of n^2 words of word length $n+1$ with minimum distance n. Again this is equivalent to a **finite projective plane** of order n, i.e., each line containing exactly $n+1$ points. If n is a power of a prime, then there exists a projective plane of order n. In general, if p^r is the minimum of the prime-power components of n, then there is a set of p^r-1 mutually orthogonal latin squares of order n.

C. Quasigroups

A **quasigroup** $(A, \circ)$ is a set A bestowed with a binary operation $\circ$, satisfying both cancellation laws; $x \circ y=x \circ z$ implies $y=z$; and $y \circ x=z \circ x$ implies $y=z$. A finite quasigroup is a group if and only if it satisfies the associative law $(x \circ y) \circ z=x \circ (y \circ z)$. A quasigroup is isotopic to a group if and only if it satisfies **Brandt's law**: $x \circ y=z \circ w$, $x \circ y'=z \circ w'$, $x' \circ y=z' \circ w$ implies $x' \circ y'=z' \circ w'$. Two groups are isotopic if and only if they are isomorphic.

A biunique mapping f of a finite group G is called a **complete mapping** if $x \to x^{-1}f(x)$ is also a biunique mapping. In this case if we denote by G' the quasigroup defined by $x \circ y=xf(y)$, then the two latin squares corresponding to G and G' are mutually orthogonal. A group of odd order has a complete mapping, while a group of even order with a cyclic 2-Sylow subgroup does not have a complete mapping. A solvable group of even order with a noncyclic 2-Sylow subgroup also has a complete mapping, and it is conjectured that the solvability condition here is redundant.

D. Room Squares

A quasigroup $(A, \circ)$ is idempotent if $x \circ x=x$ for any x and is commutative if $x \circ y=y \circ x$ for any x and y. Now a **Room square** of order $2n$

is a distribution of the $n(2n-1)$ unordered pairs of elements from $A=\{a_0, a_1, \ldots, a_{2n-1}\}$ among the $(2n-1)^2$ cells of a square of side length $2n-1$, such that each element of A appears exactly once in each row and in each column. The remaining $(2n-1)(n-1)$ cells should be empty. According to Bruck this is equivalent to a pair of two idempotent commutative quasigroups $(A; \circ)$ and $(A; \bullet)$ defined over $A'=\{a_1, \ldots, a_{2n-1}\}$, satisfying an orthogonality condition similar to that of latin squares, namely, $x \circ y = x' \circ y'$, $x \bullet y = x' \bullet y'$ implies $x=x'$, $y=y'$ or $x=y'$, $y=x'$. A Room square of order $2n$ exists if $n \geqslant 4$.

E. Number of Latin Squares

Little is known about the total number of latin squares in general. The first k rows of a latin square of order n is a $k \times n$ **latin rectangle**. The number of $k \times n$ latin rectangles is asymptotic to $(n!)^k \exp(-k(k-1)/2)$ for $k < n^{1/3}$ (Erdös, Kaplansky, and Yamamoto) and to $(n!)^k \exp(-\binom{k}{2}-\binom{k}{3}/(n-1)-\binom{k}{4}/\binom{n-1}{2})$ for $k < n^{5/12}$. They suggest an analogous asymptotic relation for the number of latin squares. The following congruences are known for the number L_n of reduced latin squares; $L_p \equiv 1 \pmod p$, $L_n \equiv 0 \pmod p$ for $n \geqslant 2p$ if p is a prime.

References

[1] J. Dénes and A. D. Keedwell, Latin squares and their applications, Akadémiai Kiadó, Budapest, and Academic Press, 1974.
[2] W. D. Wallis, A. F. Street, and J. S. Wallis, Combinatorics: Room squares, sum-free sets, Hadamard matrices, Lecture notes in math. 292, Springer, 1972.

242 (V.8) Lattice-Point Problems

A. General Remarks

Suppose that we are given in a Euclidean plane a closed †Jordan curve C of length L that bounds a †region of area F. We denote by A the number of †lattice points on the curve C or in the region bounded by C. In many cases it can be verified that $A=F+O(L)$ (O is the †Landau symbol). Specifically, if we take a circle whose center is the origin and whose radius is $\sqrt{x}$, then $A(x)=\pi x+O(\sqrt{x})$. Next, consider the closed region defined by $uv \leqslant x$, $u \geqslant 1$, and $v \geqslant 1$ on the uv-plane. Let $D(x)$ be the number of lattice points lying in this closed region. Then $D(x)=x\log x+(2C-1)x+O(\sqrt{x})$, where C is the †Euler constant. To observe another aspect of the problem of estimating the number of lattice points in a given region, consider the series

$$\sum_{\substack{m,n=-\infty \\ (m,n)\neq(0,0)}}^{+\infty} (m^2+n^2)^{-s} = f(s),$$

$$\left(\sum_{n=1}^{\infty} n^{-s}\right)^2 = g(s).$$

Then we have $f(s)=\sum_{n=1}^{\infty} r(n)n^{-s}$, $g(s)=\sum_{n=1}^{\infty} d(n)n^{-s}$, where $r(n)$ is the number of integral solutions (u,v) of the equation $u^2+v^2=n$ and $d(n)$ is the number of positive integral solutions (u,v) of the equation $uv=n$. Thus the problem of estimating $A(x)$ is identical to that of estimating $H(x)=\sum_{n\leqslant x} r(n)$; this is called **Gauss's circle problem**. We also have $D(x)=\sum_{n\leqslant x} d(n)$, and the problem of estimating the latter is called **Dirichlet's divisor problem**.

We set $P(x)=A(x)-\pi x$ and $\Delta(x)=D(x)-(x\log x+(2C-1)x)$. W. Sierpiński (1906) showed that $P(x)=O(x^{1/3})$, and G. Voronoi (1903) showed that $\Delta(x)=O(x^{1/3}\log x)$. There are further investigations concerning the estimations of $P(x)$ and $\Delta(x)$. J. G. van der Corput and E. C. Titchmarsh devised methods to estimate more general †trigonometric sums. For instance, let $f(x)$ be a real-valued function of †class C^k ($k \geqslant 3$). If $0<\lambda\leqslant f^{(k)}(x)\leqslant h\lambda$ or $0<\lambda\leqslant -f^{(k)}(x)\leqslant h\lambda$ in the interval $a\leqslant x\leqslant b$ (with $b-a\geqslant 1$), then

$$\sum_{a\leqslant n\leqslant b} \exp(2\pi i f(x)) = O(h^{2^{2-k}}(b-a)\lambda^{(2^k-2)^{-1}} + (b-\mathrm{a})^{1-2^{2-k}}\lambda^{-(2^k-2)^{-1}}).$$

As of 1968, we have an estimate slightly better than $O(x^{13/40+\varepsilon})$ for $P(x)$ and $\Delta(x)$ obtained by L. K. Hua (1942), C. J. Cheng (1963), and W. L. Yin (1959). It is conjectured that $P(x)$ and $\Delta(x)$ are $O(x^{1/4+\varepsilon})$, where ε is an arbitrary positive number. On the other hand, M. Tsuji (1953) proved that $\int_1^x P(y)/y\,dy=O(1)$. G. H. Hardy (1916) and A. E. Ingham (1941) showed that

$$\limsup_{x\to\infty}\frac{P(x)}{x^{1/4}}=\infty, \quad \liminf_{x\to\infty}\frac{P(x)}{x^{1/4}\log^{1/4}x}<0,$$

$$\limsup_{x\to\infty}\frac{\Delta(x)}{x^{1/4}\log^{1/4}x}>0, \quad \liminf_{x\to\infty}\frac{\Delta(x)}{x^{1/4}}=-\infty.$$

H. Cramér (1926) showed that

$$\frac{1}{x}\int_1^x |P(y)|\,dy=O(x^{1/4}),$$

$$\frac{1}{x}\int_1^x |\Delta(y)|\,dy=O(x^{1/4}).$$

G. Voronoi (1904) proved that if x is positive,

then

$$\sum_{n\leqslant x}' r(n)=\pi x+\sqrt{x}\sum_{n=1}^{\infty}\frac{r(n)}{\sqrt{n}}J_1(2\pi\sqrt{nx}),$$

$$\sum_{n\leqslant x}' d(n)=x\log x+(2C-1)x+1/4$$
$$+\sqrt{x}\sum_{n=1}^{\infty}\frac{d(n)}{\sqrt{n}}F(2\pi\sqrt{nx}),$$

where $\sum'$ means that when x is an integer m, the mth term $d(m)$ is replaced by $(1/2)d(m)$. Here $J_1(x)$ is the †Bessel function of the first kind, and $F(x)=(2/\pi)\int_0^\infty \cos(xu)\sin(x/u)\,du$. There are many proofs for these expansion formulas. E. Landau's proof (1920) of the estimation of $\sum_{n\leqslant x} r(n)$ is interesting from the geometric point of view, and W. Rogosinski's proof (1922) of the estimation of $\sum'_{n\leqslant x} d(n)$ uses real analytic methods in an ingenious manner. A. Oppenheim (1926) generalized these problems.

B. Other Extensions

Let $a_{\mu\nu}$ be rational, $a_{\mu\nu}=a_{\nu\mu}$, and $Q(u_1,\ldots,u_n)=\sum_{\mu,\nu=1}^{n} a_{\mu\nu}u_\mu u_\nu$ be a †positive definite quadratic form with †discriminant D. As an extension of Gauss's circle problem, it is natural to consider the number of lattice points $(m_1,\ldots,m_n)$ satisfying $Q(m_1,\ldots,m_n)\leqslant x$. In connection with the †Epstein zeta function, this problem was extended to that of estimating the sum

$$F(x)=\sum_{Q(m_1,\ldots,m_n)\leqslant x}\exp(2\pi i(\alpha_1 m_1+\ldots+\alpha_n m_n)).$$

Namely, the weight $\exp(2\pi i(\alpha_1 m_1+\ldots+\alpha_n m_n))$ is placed at each lattice point. We now define δ such that $\delta=1$ if $\alpha_1,\ldots,\alpha_n$ are all integers and $\delta=0$ otherwise. Landau (1915) obtained three exquisite proofs of

$$F(x)=\delta\frac{\pi^{n/2}}{\sqrt{D}\,\Gamma(n/2+1)}x^{n/2}+O(x^{n(n-1)/2(n+1)}).$$

I. M. Vinogradov (1960) obtained a deeper result for the special case $n=3$: $\sum_{u^2+v^2+w^2\leqslant x}1=(4/3)\pi x^{3/2}+O(x^{19/28+\varepsilon})$.

Four times the †Dedekind zeta function of the Gaussian field $\mathbf{Q}(\sqrt{-1})$ is equal to $\sum_{n=1}^{\infty} r(n)n^{-s}$. Hence Gauss's circle problem can be extended to that of estimating $H(x)$ for the Dedekind zeta function. The generalized divisor problem, including Gauss's circle problem and Dirichlet's divisor problem, was the principal theme of Landau's research after 1912. We now consider the case where the Dirichlet series $\sum_{n=1}^{\infty} F(n)n^{-s}$ is a finite product of Dedekind zeta functions. With a slight modification the following result is valid for the product of †Hecke L-functions: Let k_j $(1\leqslant j\leqslant\tau)$ be an algebraic number field of degree n_j, $\zeta_j(s)$ be the †Dedekind zeta function of k_j, and ρ_j be the residue at the pole $s=1$ of $\zeta_j(s)$. Further, we assume that

$$n_1+n_2+\ldots+n_\tau=N,$$

$$\zeta_1(s)\zeta_2(s)\ldots\zeta_\tau(s)=\sum_{n=1}^{\infty}F(n)n^{-s},$$

$$H(x)=\sum_{n\leqslant x}F(n).$$

Then

$$H(x)=x(a_1\log^{\tau-1}x+\ldots+a_{\tau-1}\log x+a_\tau)$$
$$+O(x^{(N-1)/(N+1)}\log^{\tau-1}x),$$
$$a_1=\rho_1\rho_2\ldots\rho_\tau/(\tau-1)!,$$

and the remainder O-term of the right-hand side cannot be replaced by $O(x^\theta)$ for $\theta<1/2-1/2N$. There are some algebraic results (Z. Suetuna (1929), H. Hasse and Suetuna (1931)) concerning the estimation of $H(x)$. In particular, if in the definition of $F(n)$, the $\zeta_j(s)$ are all equal to the Riemann zeta function, then we obtain $\zeta(s)^k=\sum_{n=1}^{\infty}d_k(n)n^{-s}$, where $d_k(n)$ is the number of ways of expressing n as a product of k factors. In this special case the remainder term can be replaced by $O(x^{c+\varepsilon})$, where $c=\max(1/2,(k-1)/(k+2))$ $(k\geqslant 3)$ (G. H. Hardy and J. E. Littlewood, 1922). An appropriate application of the †Artin L-function was obtained by Suetuna (*J. Fac. Sci. Univ. Tokyo*, 1925), who extended to algebraic number fields of finite degree the result obtained by Landau (1912)—which states that the number of positive integers not larger than x that can be expressed as the sum of two squares is approximately equal to $ax/\sqrt{\log x}$, where a is a positive constant.

References

[1] G. H. Hardy and J. E. Littlewood, The approximate functional equation in the theory of the zeta-function, with application to the divisor-problems of Dirichlet and Piltz, Proc. London Math. Soc., (2) 21 (1923), 39–74.
[2] L. K. Hua, Die Abschätzung von Exponentialsummen und ihre Anwendung in der Zahlentheorie, Enzykl. Math., Bd. I, 2, Heft 13, Teil I, 1959.
[3] E. G. H. Landau, Zur analytischen Zahlentheorie der definiten quadratischen Formen, S.-B. Preuss. Akad. Wiss. (1915), 458–476.
[4] E. G. H. Landau, Einführung in die elementare und analytische Theorie der algebraischen Zahlen und der Ideale, Teubner, 1918, second edition, 1927.
[5] E. G. H. Landau, Vorlesungen über Zahlentheorie II, Hirzel, 1927 (Chelsea, 1969).

[6] E. C. Titchmarsh, The theory of the Riemann zeta-function, Clarendon Press, 1951.
[7] E. G. H. Landau, Ausgewählte Abhandlungen zur Gitterpunktlehre, A. Walfisz (ed.), Deutscher Verlag der Wiss., 1962.
[8] A. Walfisz, Gitterpunkte in mehrdimensionalen Kugeln, Warsaw, 1957.

243 (II.14) Lattices

A. Definitions

When x and y are elements of an †ordered set L, the †supremum and †infimum of $\{x, y\}$, whenever they exist, are called the **join** and **meet** of x, y and denoted by $x \cup y$ and $x \cap y$, respectively. L is called a **lattice** (or **lattice-ordered set**) when every pair of its elements has a join and a meet. The following three laws hold in any lattice L: (i) $x \cup y = y \cup x$, $x \cap y = y \cap x$ (**commutative law**); (ii) $x \cup (y \cup z) = (x \cup y) \cup z$, $x \cap (y \cap z) = (x \cap y) \cap z$ (**associative law**); and (iii) $x \cup (y \cap x) = (x \cup y) \cap x = x$ (**absorption law**). Conversely, if in a set L two operations $\cup$, $\cap$ are given that satisfy (i)–(iii), then the conditions $x \cup y = y$ and $x \cap y = x$ are equivalent and define an †ordering $x \leqslant y$ in L with respect to which L becomes a lattice. The supremum and infimum of $\{x, y\}$ in this lattice coincide with the elements $x \cup y$ and $x \cap y$, respectively. Accordingly, a lattice can also be defined as an †algebraic system with operations $\cup$, $\cap$ satisfying laws (i)–(iii). The **idempotent law** $x \cup x = x \cap x = x$ holds in any lattice.

An ordered set L is called an **upper semilattice** if each pair of elements x, y always has a join (supremum) $x \cup y$, and a **lower semilattice** if each pair of elements x, y always has a meet (infimum) $x \cap y$.

B. Examples

The set $\mathfrak{P}(S)$ of all subsets of a given set S is a complete and distributive lattice with respect to the inclusion relation (→ Sections D, E). The set of all †normal subgroups of a given †group is a complete and modular lattice (→ Section F) with respect to the inclusion relation. This remains true if the normal subgroups are replaced by the †admissible subgroups with respect to a given †operator domain. This applies in particular to the case of the set of all †ideals in a given commutative †ring. The set of all subspaces of a given †projective space is a modular lattice.

C. Further Definitions

A mapping f of a lattice L into a lattice L' that satisfies the conditions $f(x \cup y) = f(x) \cup f(y)$ and $f(x \cap y) = f(x) \cap f(y)$ is called a **lattice homomorphism** (or simply **homomorphism**). A †bijective lattice homomorphism f is called a **lattice isomorphism** (or simply **isomorphism**); its inverse mapping is also a lattice isomorphism. When such an f exists, the lattices L and L' are said to be **isomorphic**. More generally, a mapping f between ordered sets is said to be order-preserving when it satisfies the condition: $x \leqslant y$ implies $f(x) \leqslant f(y)$. Any lattice homomorphism is order-preserving, but the converse is not always true; however, an order-preserving bijection is an isomorphism.

If the ordering in a lattice L is replaced by the †dual ordering, then the join and the meet are interchanged, and a new lattice L' is obtained. This new lattice is called the **dual lattice** for L.

A mapping f of a lattice L into a lattice L' satisfying the conditions $f(x \cap y) = f(x) \cup f(y)$ and $f(x \cup y) = f(x) \cap f(y)$ is called a **dual homomorphism** (or **antihomomorphism**). Moreover, when f is a bijection, f is called a **dual isomorphism** (or **anti-isomorphism**), and we say that L and L' are **dually isomorphic** (or **anti-isomorphic**) to each other.

When a lattice L' is a subset of a lattice L and the canonical injection $L' \to L$ is a lattice homomorphism, L' is called a **sublattice** of L. If a subset L' of a lattice L satisfies the condition that $x, y \in L'$ implies $x \cup y$, $x \cap y \in L'$, then two operations $\cup$, $\cap$ can be induced in L' so that L' becomes a sublattice. For example, when a, b are given elements of a lattice L, the set of elements x satisfying $a \leqslant x \leqslant b$ is a sublattice, denoted by $[a, b]$ and called an **interval** of L. When the quotient set L/R of a lattice L by an equivalence relation R in L is also a lattice and the canonical surjection $L \to L/R$ is a homomorphism, then L/R is called a **quotient lattice** of L. If an equivalence relation R in a lattice L satisfies the condition that $x \equiv x'$, $y \equiv y' \pmod{R}$ implies $x \cup y \equiv x' \cup y'$, $x \cap y \equiv x' \cap y' \pmod{R}$, then two operations $\cup$, $\cap$ can be induced in L/R so that L/R becomes a quotient lattice. The Cartesian product $L = \prod_{i \in I} L_i$ of a family $\{L_i\}_{i \in I}$ of lattices becomes a lattice if the operations $\cup$, $\cap$ are defined by $(x_i) \cup (y_i) = (x_i \cup y_i)$, $(x_i) \cap (y_i) = (x_i \cap y_i)$. This lattice is called the **direct product** of lattices $\{L_i\}_{i \in I}$.

D. Complete Lattices

An ordered set L is called a **complete lattice** if every nonempty subset of L has a supremum

and an infimum in L, and a **σ-complete lattice** if every nonempty countable subset has a supremum and an infimum. Naturally, such ordered sets are lattices. And a sort of converse holds: Any lattice is a sublattice of some complete lattice. An ordered set L is said to be **conditionally complete** when every subset †bounded from above (below) has a supremum (infimum) in L, and **conditionally σ-complete** when every countable subset bounded from above (below) has a supremum (infimum). For any ordered set L there exist a complete lattice $\bar{L}$ and an order-preserving injection $f: L \to \bar{L}$ satisfying the condition that each element $\xi \in \bar{L}$ is the supremum and infimum of the images $f(X)$ and $f(Y)$, respectively, for some sets X, $Y \subset L$. This condition is equivalent to the condition that for any complete lattice $\bar{L}'$ and order-preserving injection $f': L \to \bar{L}'$, there exists an order-preserving injection $\varphi: \bar{L} \to \bar{L}'$ for which $\varphi \circ f = f'$. Hence $(\bar{L}, f)$ is unique up to lattice isomorphisms. $\bar{L}$ is called the **completion** of the ordered set L. For example, the set of real numbers supplemented by $+\infty$ and $-\infty$ is the completion of the set of rational numbers.

E. Distributive Lattices

A lattice L is said to be **distributive** when it satisfies the following equivalent conditions (**distributive laws**) for $x, y, z \in L$: (i) $x \cup (y \cap z) = (x \cup y) \cap (x \cup z)$; (ii) $x \cap (y \cup z) = (x \cap y) \cup (x \cap z)$; and (iii) $(x \cup y) \cap (y \cup z) \cap (z \cup x) = (x \cap y) \cup (y \cap z) \cup (z \cap x)$. The dual lattices, sublattices, quotient lattices, and direct products of distributive lattices are distributive. The set $\mathfrak{P}(S)$ of subsets of a given set S is a distributive lattice, and each of its sublattices is called a **lattice of sets** in S. A distributive lattice is isomorphic to a certain lattice of sets. A homomorphism from a distributive lattice L into $\mathfrak{P}(S)$ is called a **representation** of L in S.

A lattice L is said to be **complemented** when a greatest element I and a least element 0 exist in L and for every element x, there exists an element x' satisfying $x \cup x' = I$, $x \cap x' = 0$. Such an x' is called a **complement** of x. A lattice that is distributive and complemented is called a **Boolean lattice** (or †**Boolean algebra**). In a Boolean lattice, every element has a unique complement. The lattice $\mathfrak{P}(S)$ of all subsets of a given set S is a Boolean lattice, in which S is the greatest element and $\varnothing$ the least element. A sublattice of $\mathfrak{P}(S)$ that contains the complement of each of its elements is also a Boolean lattice and is called a **Boolean lattice of sets**. Any Boolean lattice can be represented isomorphically by some Boolean lattice of sets (→ 42 Boolean Algebra).

An element a of a lattice is called a **neutral element** if for any pair of elements x and y, the sublattice generated by a, x, y is distributive. When a is neutral and has a complement, a is called a **central element**. The set of all central elements of a lattice L is called the **center** of L.

F. Modular Lattices

A lattice L is said to be **modular** if the following condition (**modular law**) is satisfied: $x \leqslant z$ implies $x \cup (y \cap z) = (x \cup y) \cap z$. A distributive lattice is always modular. The dual lattices, sublattices, quotient lattices, and direct product of modular lattices are also modular.

The †Jordan-Hölder theorem and the refinement theorem of O. Schreier on normal subgroups (→ 190 Groups) are generalized as follows to the case of any modular lattice: A pair of elements x, y in a lattice L satisfying $x \geqslant y$ is called a **quotient** and is denoted by x/y. In particular, when $x > y$ and there exists no element z such that $x > z > y$, then x/y is called a **prime quotient**, x is said to be **prime over** y, and y is said to be **prime under** x. A sequence $C: x_0, x_1, \ldots, x_k$ of elements of L satisfying the conditions $x_{i-1} \geqslant x_i$ $(1 \leqslant i \leqslant k)$ is called a **descending chain**, and k is called its **length**. Each x_{i-1}/x_i is called a quotient determined by C. When each of these quotients is prime, C is called a **composition series**. A descending chain $D: y_0, y_1, \ldots, y_l$ is called a **refinement** of a descending chain $C: x_0, x_1, \ldots, x_k$ when $x_0 = y_0$, $x_k = y_l$ and each x_i is equal to some y_j. We now define an equivalence relation between descending chains. First, a relation $x/y \approx x'/y'$ between quotients x/y and x'/y' is defined to mean that either the condition $x = x' \cup y$, $y' = x' \cap y$ or the condition $x' = x \cup y'$, $y = x \cap y'$ holds. Then x/y and x'/y' are called equivalent if there exists a finite number of quotients $q_0, q_1, \ldots, q_r$ satisfying the conditions $x/y = q_0$, $x'/y' = q_r$ and $q_{i-1} \approx q_i$ $(1 \leqslant i \leqslant r)$. Descending chains C and C' are said to be equivalent if they have the same length and the set of quotients determined by C is mapped bijectively to the set of quotients determined by C' so that the quotient and its image are equivalent. Now let L be a modular lattice. If two quotients x/y and x'/y' in L are equivalent, the intervals $[y, x]$ and $[y', x']$, considered as lattices, are isomorphic (**Dedekind's principle**). If two descending chains $C: x_0, x_1, \ldots, x_k$ and $C': y_0, y_1, \ldots, y_l$ have the same ends $x_0 = y_0$ and $x_k = y_l$, then there exist a refinement of C and a refinement of C' which are equivalent. In particular, any two composition series connecting the same elements are equivalent (→ 85 Continuous Geometry).

In a modular lattice L with a least element

0, if there exists a composition series connecting 0 and a given element a, then all such composition series have a common length k, denoted by $d(a)$ and called the **height** of the element a. If no such composition series exists, the height is defined to be $d(a)=\infty$. If two elements a, b of L satisfy $d(a\cup b)<\infty$, then $d(a\cup b)+d(a\cap b)=d(a)+d(b)$. This fact is called the **dimension theorem** of modular lattices. If L has a greatest element I, $d(I)$ is called the **height** of the lattice L.

In a complemented modular lattice L, an element which is prime over the least element 0 is called an **atomic element**. A complemented modular lattice L is said to be **irreducible** if any two atomic elements have a common complement.

G. Lattice-Ordered Groups

An ordered set G in which a group operation is defined is called an **ordered group** when $x\leqslant y$ implies $xz\leqslant yz$ and $zx\leqslant zy$ for all x, y, z in G. Moreover, if it is a †totally ordered set, the ordered group G is called a **totally ordered group**. If G is a lattice, the condition for an ordered group is equivalent to the condition that $(x\cup y)z=xz\cup yz$, $(x\cap y)z=xz\cap yz$, $z(x\cup y)=zx\cup zy$, and $z(x\cap y)=zx\cap zy$. In this case, G is called a **lattice-ordered group**. A lattice-ordered group is a distributive lattice and has neither a greatest nor a least element. If $\{x_i\}$ has a supremum in a lattice-ordered group, then we have $(\sup_i x_i)\cap y=\sup_i(x_i\cap y)$ and its dual (**complete distributive law**).

The lattice-theoretic structure of a lattice-ordered group was clarified by P. Lorentzen, A. H. Clifford, and T. Nakayama. In particular, a lattice-ordered commutative group is isomorphic (as a lattice-ordered group) to some subgroup of a direct product of totally ordered groups. A lattice-ordered group has no element of finite order other than the identity element. Conversely, a commutative group which has no element of finite order other than the identity element can be made a lattice-ordered group with respect to some total ordering. Any free group can also be made a lattice-ordered group with respect to some total ordering. K. Iwasawa (1948) and others have done further research on totally ordered groups.

An element x ($\neq e$) of a lattice-ordered group G is called **positive (negative)** when $x\geqslant e$ ($x\leqslant e$), where e is the identity. G is called an **Archimedean lattice-ordered group** when the following condition is satisfied: if, for some y, $x^n\leqslant y$ for all natural numbers n, then $x\leqslant e$. An Archimedean lattice-ordered group is isomorphic to some subgroup of a complete lattice-ordered group. Conversely, a complete lattice-ordered group is Archimedean; moreover, it is commutative and isomorphic to the direct product of some †lattice-ordered linear spaces and copies of the lattice-ordered group of rational integers (Iwasawa, 1943). In particular, any totally ordered Archimedean lattice-ordered group is isomorphic to a subgroup of the lattice-ordered group of all real numbers.

If the †minimal condition holds for the set of all positive elements in a lattice-ordered group, then the group is commutative, and each of its elements can be decomposed uniquely into a product of powers of elements that are prime over the identity element. The set of all †fractional ideals of an algebraic number field is a typical example of such a lattice-ordered group. For further reference → 310 Ordered Linear Spaces, 85 Continuous Geometry.

References

[1] G. Birkhoff, Lattice theory, Amer. Math. Soc. Colloq. Publ., 1940, revised edition, 1948.
[2] P. Dubreil and M. L. Dubreil-Jacotin, Leçons d'algèbre moderne, Dunod, second edition, 1961; English translation, Lectures on modern algebra, Oliver & Boyd, 1967.

244 (XXI.34) Lebesgue, Henri Léon

Henri Lebesgue (June 28, 1875–July 26, 1941) was born in Beauvais (Oise), studied at the Ecole Normale Supérieure, and received in 1902 the doctoral degree for his thesis concerning integration [1]. After teaching in Rennes and Poitiers, he came to Paris where he was nominated for a professorship at the Faculty of Science in 1920 and then at the Collège de France in 1921. He was one of the most influential French analysts of this century and is known as the inventor of the †Lebesgue integral. With deep insight based on intuitive geometric conceptions, he was able to initiate a new era in analysis by creating the theory of this integral. Not only was this theory the start of modern integration, it was also a turning point in the theory of †Fourier series and †potential theory. The notions of †Lebesgue dimension of topological spaces and of †Lebesgue number of compact sets are due to him. Lebesgue also made significant contributions to the †Dirichlet problem.

References

[1] H. Lebesgue, Intégrale, longueur, aire, Ann. Mat. Pura Appl., (3) 7 (1920), 231–359.
[2] H. Lebesgue, Sue les fonctions représentables analytiquement, J. Math. Pures Appl., (6) 1 (1905), 139–216.
[3] H. Lebesgue, Leçons sur l'intégration et la recherche des fonctions primitives, Gauthier-Villars, 1904, second edition, 1928.

245 (XXI.35) Leibniz, Gottfried Wilhelm

Baron Gottfried Wilhelm von Leibniz (July 1, 1646–November 4, 1716) was born the son of a professor and grew up to be a genius with encyclopedic knowledge. He took part in politics and touched all scholarly fields, contributing creatively to the development of technology as well. His posthumously published works cover theology, philosophy, mathematics, the natural sciences, history, and technology, and are classified into 41 fields. A complete edition of his works has yet to be published. *Ars combinatoria*, written upon his graduation from Altdorf in 1666, was a scheme to systematize the various fields using mathematics as a model. During his stay in Paris (1672–1676), when not involved in politics, he studied the works of †Descartes and †Pascal, as suggested by C. Huygens. He discovered the †fundamental theorem of differential and integral calculus and set up a basis for calculus with the introduction of an ingenious system of notation. After 1676, he worked on historical compilations under the Duke of Hanover.

Leibniz worked not only on the synthesis of mechanistic philosophy and medieval theological philosophy but also on the reconciliation of Protestantism and Catholicism. With his monadism he attempted to unify the old and new philosophies. In addition he worked on plans for a world academy for the development of learning and on the unification of all knowledge. This was to be accomplished using, for example, universal symbolism and universal linguistics. Under his influence, the Berlin Academy was established in 1700. After his death, his conceptions of †symbolic logic and †computers were realized.

References

[1] C. I. Gerhardt (ed.), Leibnizens mathematische Schriften I–VII, Schmidt, 1849–1863.
[2] C. I. Gerhardt (ed.), G. W. Leibniz, Philosophische Schriften I–VII, Berlin, 1875–1890.

246 (X.16) Length and Area

A. Length of a Curve

A continuous mapping C sending each point u of an interval $I=\{u \mid a\leqslant u\leqslant b\}$ to a point $p=p(u)=(x_1(u),\ldots,x_k(u))$ of the k-dimensional Euclidean space $\mathbf{R}^k$ $(k\geqslant 2)$ is called a †continuous arc or continuous curve, denoted sometimes by $C:p=p(u)$. The supremum of the length of a polygonal curve inscribed in C is called the **length** of C (in the sense of Jordan) and is denoted by $l(C)$. Namely, it is equal to $\sup_\delta \sum_{i=1}^m |p(u_i)-p(u_{i-1})|$, where δ is a partition of $I: a=u_0<u_1\ldots<u_m=b$. Let $C:p=p(u)$, and $C_n:p=p_n(u)$, $n=1,2,\ldots$ $(u\in I)$ be continuous arcs. If $p_n(u)\to p(u)$ on I, then $l(C)\leqslant\liminf_n l(C_n)$. This property is called the **lower semicontinuity** of length. Given two continuous arcs $C:p=p(u)$ $(u\in I)$ and $C_1:p=q(v)$ $(v\in I_1)$, if for any $\varepsilon>0$, there exists a homeomorphism $u=h_\varepsilon(v)$ of I_1 onto I such that $|p(h_\varepsilon(v))-q(v)|<\varepsilon$ on I_1, then C and C_1 are called **equivalent** in the sense of Fréchet. Equivalent continuous arcs have the same length. A continuous image of an open interval is called an †open arc. The length of an open arc is defined to be the supremum of the length of a continuous arc contained in the open arc, and the notion of the equivalence of open arcs is defined in the same way as for continuous arcs. An equivalence class of continuous arcs (or open arcs) is called a **Fréchet curve**, and its length is defined uniquely.

Suppose that a continuous arc C is expressed by $(x_1(u),\ldots,x_k(u))$, $u\in I$. Then the length $l(C)$ is finite if and only if every $x_i(u)$ is of †bounded variation. When $l(C)$ is finite, C is called **rectifiable**. In this case each $\partial x_i/\partial u$ exists †almost everywhere on I, and the inequality

$$l(C)\geqslant\int_a^b\left(\sum_{i=1}^k\left(\frac{\partial x_i}{\partial u}\right)^2\right)^{1/2}du \qquad (1)$$

holds. The equality holds if and only if each $x_i(u)$ is absolutely continuous. Among the continuous arcs equivalent to C, there exists a unique continuous arc C_1 such that $C_1:q=q(s)$ $(0\leqslant s\leqslant l(C))$ and the length of every subarc $q=q(s)$ $(0\leqslant s\leqslant s'(\leqslant l(C)))$ is equal to s'. C_1 is called the **representation in terms of arc length** of C. For C_1, the equality holds in (1). A similar argument is valid for any open arc. When every subarc of C is rectifiable, C is called **locally rectifiable**. If Λ_1 is the 1-dimensional †Hausdorff measure in $\mathbf{R}^k$ and $n(p)$ is the number of points on I corresponding to $p\in\mathbf{R}^k$, then $l(C)=\int n(p)\,d\Lambda_1(p)$ (M. Ohtsuka, 1951).

B. Surface Area

In this section, we deal with area of †surfaces in $\mathbf{R}^3$, using [1] as the main reference. In contrast to the situation for curves, the area of a polyhedral surface P inscribed in a given surface does not necessarily tend to a fixed value as P approximates the surface. In a letter of 1880, H. A. Schwarz gave the following example: Approximate a circular cylinder with height h and radius r by a sequence $\{P_n\}$ of inscribed polyhedral surfaces, each P_n consisting of similar triangles of height b_n and base length a_n. If b_n/a_n^2 is suitably chosen, then the surface area of P_n tends to an arbitrary value not smaller than $2\pi rh$ (Fig. 1).

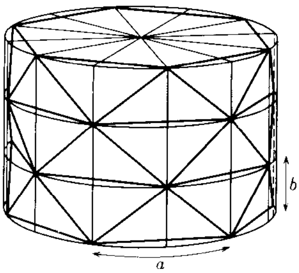


Fig. 1

C. Lebesgue Area

Suppose that we are given a plane domain A and a continuous mapping T of A into $\mathbf{R}^3$. The pair (T, A) is called a surface. Let $d(T, T', B) = \sup_{w \in B} |T(w) - T'(w)|$ for surfaces (T, A), (T', A') and a set $B \subset A \cap A'$. Let (T, A), (T_1, A_1), (T_2, A_2), ... be given. If $A_n \uparrow A$ and $d(T, T_n, A_n) \to 0$, then (T_n, A_n) (or simply T_n) is said to converge to (T, A) (or T), and the convergence is expressed by $T_n \to T$. In particular, if A consists of a finite number of triangles and T is linear on each triangle, i.e., the image of A under T consists of triangles, then the notation (P, F) is used for (T, A), and the area of $T(A)$ is denoted by $a(P, F)$. Given a surface (T, A), denote the totality of sequences $\{(P_n, F_n)\}$ converging to (T, A) by Φ, and call $\inf_\Phi \liminf_n a(P_n, F_n)$ the **Lebesgue area** of (T, A). This area is denoted by $L(T, A)$. By virtue of the definition there exists a sequence $\{(P_n, F_n)\}$ converging to (T, A) such that $a(P_n, F_n) \to L(T, A)$. Like length, Lebesgue area has the lower semicontinuity property. Namely, $T_n \to T$ implies $L(T, A) \leqslant \liminf_n L(T_n, A_n)$. When A is a Jordan domain, the same value $L(T, A)$ is obtained if Φ is replaced by the set Φ^* of all sequences $\{(P_n, F_n)\}$ such that $F_n \uparrow A$ and $P_n(w) \to T(w)$.

Let $f(x, y)$ be a continuous function defined on $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1$. Regard $f(x, y)$ as a function of y (resp. x) for a fixed x (y), and denote it by $f_x(y)$ $(f_y(x))$ and its †total variation by $V(x)$ $(V_1(y))$. When $\int_0^1 V(x)\,dx + \int_0^1 V_1(y)\,dy < \infty$, $f(x, y)$ is said to be of **bounded variation in the sense of Tonelli**. Furthermore, if $f_x(y)$ and $f_y(x)$ are †absolutely continuous for almost every x and y, respectively, then $f(x, y)$ is said to be **absolutely continuous in the sense of Tonelli**. Similar definitions are also given when f is defined in a general domain. Suppose that a surface (T, A) is expressed by a set of three functions $x = x(u, v)$, $y = y(u, v)$, $z = z(u, v)$, all of which are absolutely continuous in the sense of Tonelli, and that the partial derivatives x_u, $x_v, \ldots, z_v$ are square integrable. Then $L(T, A) = \iint_A J\,du\,dv < \infty$ (as was shown by C. B. Morrey), where $J = (J_1^2 + J_2^2 + J_3^2)^{1/2}$ with †functional determinants J_1, J_2, J_3 of the transformations $(u, v) \to (y, z), (z, x), (x, y)$.

D. The Geöcze Problem

The **Geöcze problem** is the problem of determining whether $L(T, A)$ coincides with the area obtained by using (instead of Φ) the set of all sequences $\{(P_n, F_n)\}$ such that (P_n, F_n) converges to (T, A) and each (P_n, F_n) is inscribed in (T, A). The answer is affirmative when A is a Jordan domain with $L(T, A) < \infty$. Let T be a surface expressed by a function $z = F(x, y)$ $(0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1)$ that is absolutely continuous in the sense of Tonelli, and let $\{(P_n, F_n)\}$ be as before. If the ratio of the length of the largest side and the smallest height of each triangle in (P_n, F_n) is uniformly bounded, then $a(P_n, F_n)$ tends to $L(T, A)$ [1, p. 74].

E. Geöcze Area

Consider a surface (T, A). Let E_1, E_2, E_3 be coordinate planes in $\mathbf{R}^3$, and denote by T_i $(i = 1, 2, 3)$ the composition of the mapping T and the projection of $\mathbf{R}^3$ onto E_i. Let $\partial\pi$ be the positively oriented boundary of a polygonal domain π in A and C_i be the oriented image of $\partial\pi$ by T_i. Then the †order $O(z; C_i)$ of z with respect to C_i is a measurable function of z. Set $v_i(T; \pi) = v_i = \iint_{E_i} |O(z; C_i)|\,dx\,dy$ $(z = x + iy)$ and $v(T; \pi) = v = (v_1^2 + v_2^2 + v_3^2)^{1/2}$. The quantity

$$V(T; A) = \sup_S \sum_{\pi \in S} v(T; \pi) \tag{2}$$

is called the **Geöcze area** of (T, A), where S is a finite collection of polygonal domains in A such that no two of them overlap. If $V(T, A) < \infty$, u_i is defined by $\iint_{E_i} O(z; C_i)\,dx\,dy$, and $U(T, A)$ is defined as in (2) by means of u_i, then $U(T, A) = V(T, A)$. The inequalities $V(T_i, A) \leqslant V(T, A) \leqslant V(T_1, A) + V(T_2, A) + V(T_3, A)$ hold trivially.

F. Peano Area

Consider (T, A) and π as in the previous section. Let τ be the projection of $\mathbf{R}^3$ onto a plane E, and denote by C' the image of the boundary of π under the composite mapping $\tau \circ T$. Set $v(T, \pi, E) = \iint_E |O(z; C')| d\sigma$, where $d\sigma$ is the surface element on E, and set $\psi(T, \pi) = \sup_E v(T, \pi, E)$. Define

$$P(T, A) = \sup_S \sum_{\pi \in S} \psi(T, \pi)$$

as in (2). This is called the **Peano area** of (T, A). H. Okamura defined area by integrating the †mapping degree instead of $|O(z; C')|$ [5]. L, V, P all coincide, and hence L and V are invariant under any orthogonal transformation of $\mathbf{R}^3$.

G. Other Definitions of Area

As in the definition of Peano area, consider π, τ, E, and denote the Lebesgue measure of $\tau \circ T(\pi)$ by $m(\pi, E)$. If we set $\mu(\pi) = \sup_E m(\pi, E)$, then we can define the area of (T, A) by $\sup_E \sum_{\pi \in S} \mu(\pi)$. If $v(\pi) = (m^2(\pi, E_1) + m^2(\pi, E_2) + m^2(\pi, E_3))^{1/2}$ is used instead of $\mu(\pi)$, then the **Banach area** of (T, A) is obtained, and if $\iint |O(z; C_i)| dx\, dy$ is used instead of $m(\pi, E_i)$, then the Geöcze area is obtained. Let us define various kinds of area for an arbitary †Borel set X in $\mathbf{R}^3$. Divide $\mathbf{R}^3$ into meshes $M_1, M_2, \ldots$ which are half-open cubes with diameter of equal length d, and denote by $m_i^{(j)}$ $(i = 1, 2, 3)$ the †Lebesgue measure of the projection of $M_j \cap X$ onto the ith coordinate plane. The limit of $\sum_j ((m_1^{(j)})^2 + (m_2^{(j)})^2 + (m_3^{(j)})^2)^{1/2}$ as $d \to 0$ is called the **Janzen area** of X. Denote by m_j the supremum with respect to the set of planes E in $\mathbf{R}^3$ of the Lebesgue measure of the projection of $M_j \cap X$ to E. Then $\lim \sum_j m_j$ as $d \to 0$ is called the **Gross area** of X. C. Carathéodory covered X by a countable number of convex sets $K_1, K_2, \ldots$ each of whose diameters is less than $\delta > 0$, denoted by m_j' the supremum of the Lebesgue area of the projection on K_j into E, and adopted $\lim \sum_j m_j'$ as $\delta \to 0$ as his definition of area of X. If the K_j are restricted to be spheres, then Carathéodory's area divided by $\pi/4$ is identical with the †Hausdorff measure $\Lambda_2(X)$. (For other definitions and mutual relations → [4].)

We give a measure-theoretic definition of an area for a surface (T, A) as follows. Denote by $n(p)$ the number of points in A corresponding to a point p in $\mathbf{R}^3$, and call $n(p)$ the **multiplicity function** of the mapping T. The integral $\Lambda(T, A) = (\pi/4) \int n(p) d\Lambda_2(p)$ can be taken as a definition of the area. However, a different definition of the multiplicity function is needed for the integral to be equal to $L(T, A)$ [3, 6]. In fact, if $x = \varphi(u)$, $y = \psi(u)$ $(0 \leqslant u \leqslant 1)$ represent a †Peano curve filling the square $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1$, then the Lebesgue area of the surface (T, A) defined by $A = \{0 < \mu < 1, 0 < v < 1\}$ and $T: x = \varphi(u)$, $y = \psi(u)$, $z = 0$ is zero, but $\Lambda(T, A) \geqslant 1$.

H. Mappings of Bounded Variation

Let T be a mapping of a domain A in the w-plane into the z-plane, and define π, $O = O(z; C)$, and S as in Section E. Set

$$O^+(z; C) = (|O| + O)/2,$$

$$O^-(z; C) = (|O| - O)/2,$$

$$v(T, \pi) = \iint |O(z; C)| dx\, dy,$$

$$v^\pm(T, \pi) = \iint O^\pm(z; C) dx\, dy$$

(the same signs correspond to each other),

$$V(T, A) = \sup_S \textstyle\sum_{\pi \in S} v(T, \pi),$$

$$V^\pm(T, A) = \sup_S \textstyle\sum_{\pi \in S} v^\pm(T, \pi),$$

$$N(z; T, A) = \sup_S \textstyle\sum_{\pi \in S} |O(z; C)|,$$

and

$$N^\pm(z; T, A) = \sup_S \textstyle\sum_{\pi \in S} O^\pm(z; C).$$

Then N, $N^\pm$ are lower semicontinuous in the z-plane. The integrals $W(T, A) = \iint N\, dx\, dy$, $W^+(T, A) = \iint N^+\, dx\, dy$, and $W^-(T, A) = \iint N^-\, dx\, dy$ are called the **total variation**, **positive variation**, and **negative variation** of T, respectively, and the equalities $W = W^+ + W^-$, $V = W$, and $V^\pm = W^\pm$ hold. When $W(T, A) < \infty$, T is said to be of **bounded variation**. A related notion is defined as follows: T is **absolutely continuous** if the following two conditions hold. (1) For any given $\varepsilon > 0$ there exists a $\delta > 0$ such that $\sum_{\pi \in S} v(T, \pi) \leqslant \varepsilon$ whenever the sum of the areas of $\pi \in S$ is $\leqslant \delta$. (2) For any polygonal domain π_0 such that $\pi_0 \cup \partial\pi_0 \subset A$ and any polygonal subdivision S of π_0, $V(T, \pi_0) = \sum_{\pi \in S} V(T, \pi)$. If the area of A is finite and T is absolutely continuous, then T is of bounded variation.

Let T be a continuous mapping of bounded variation of a domain A in the w-plane into the z-plane. The derivatives $V'(w)$, $V'_+(w)$, $V'_-(w)$ of the set functions $V(T, A)$, $V^+(T, A)$, $V^-(T, A)$ exist †almost everywhere (a.e.) in A and are finite. The difference $J(w) = V'_+(w) - V'_-(w)$ is called the **generalized Jacobian**, and the relation $J(w) = V'(w)$ holds a.e. If $x(u, v)$, $y(u, v)$ are differentiable a.e., then $J(w)$ coincides with the ordinary †functional determinant a.e. Next, let T be a continuous mapping of A into $\mathbf{R}^3$ with $V(T, A) < \infty$, and denote by $J_i(w)$ the generalized Jacobian of (T_i, A). Then $J(w) = (J_1^2(w)$

$+J_2^2(w)+J_3^2(w))^{1/2}$ is called the generalized Jacobian of (T, A). The relation $J(w)=\pm V'(w)$ holds a.e. in A. Therefore

$$V(T, A)\geqslant\iint_A J(w)\,du\,dv \qquad (3)$$

is valid. The equality holds if and only if each (T_i, A) is absolutely continuous.

I. Fréchet Distance

Let (T_1, A_1) and (T_2, A_2) be surfaces, and assume that the set H of homeomorphisms between A_1 and A_2 is nonempty. Define the **Fréchet distance** between two surfaces by

$$\|T_1, T_2\|=\inf_{h\in H}\sup_{w\in A_1}|T_1(w)-T_2(h(w))|.$$

It satisfies the three axioms of distance (→ 273 Metric Spaces). When $\|T_1, T_2\|=0$, T_1 and T_2 are called **equivalent** (in the sense of Fréchet). A set of all equivalent surfaces is called a **Fréchet surface**. Equivalent surfaces have equal Lebesgue areas; hence Lebesgue area is well defined for any Fréchet surface. Given a surface (T, A) with $L(T, A)<\infty$, there exists a pair (T_1, A_1), equivalent to (T, A) such that the functional determinants $J_i(w)$ exist for (T_1, A_1) and

$$V(T_1, A_1)=\iint_{A_1} J(w)\,du\,dv,$$

where $J^2(w)=\sum_{i=1}^3 J_i^2(w)$, as before. The problem of finding such a pair (T_1, A_1) is called the **representation problem**. Moreover, we can choose (T_1, A_1) to be a **generalized conformal mapping** in the following sense: Express (T_1, A_1) by $x=x(u,v)$, $y=y(u,v)$, $z=z(u,v)$. Then $x_u, x_v, \ldots, z_v$ exist a.e. in A_1 and are square integrable, $x_u^2+y_u^2+z_u^2=x_v^2+y_v^2+z_v^2$, and $x_ux_v+y_uy_v+z_uz_v=0$ a.e. in A_1.

J. Higher-Dimensional Case

In higher-dimensional spaces, many results are also known, such as area and coarea formulas [11] and a generalization of Morrey's result on Lebesgue area (C. Goffman and W. P. Ziemer, 1970).

K. Hausdorff Dimension

A. S. Besikovich [9] has shown that for every set S in a Euclidean space, there exists a real value D such that the †d-dimensional Hausdorff measure is infinite for $d<D$ and vanishes for $d>D$. This D is called the **Hausdorff dimension.**

F. Hausdorff had already considered this measure for †Cantor sets and Koch curves.

A **fractal** [10] is defined as a set for which the Hausdorff dimension strictly exceeds the †dimension of the set, which is topologically defined in a Euclidean space.

We show only one example, the coastline of a triadic Koch island constructed in the following way. Consider an equilateral triangle with sides of unit length. Remove the middle third of each side, and attach in its place a V-shaped peninsula bounded by two sides of an equilateral triangle with side length 1/3. We thus get a Star of David. Repeat the same process of formation of peninsulas for each segment of the star's sides. If we continue this process indefinitely, then we get a complicated coastline whose Hausdorff dimension is 1.2618. Of course the †length of this coastline is ∞.

B. Mandelbrot mentions many interesting examples of fractals having fractional dimensions in his book [10].

References

[1] L. Cesari, Surface area, Ann. Math. Studies, Princeton Univ. Press, 1956.
[2] L. Cesari, Recent results in surface area theory, Amer. Math. Monthly, 66 (1959), 173–192.
[3] H. Federer, Measure and area, Bull. Amer. Math. Soc., 58 (1952), 306–378.
[4] G. Nöbeling, Über die Flächenmasse im Euklidischen Raum, Math. Ann., 118 (1943), 687–701.
[5] H. Okamura, On the surface integral and Gauss-Green's theorem, Mem. Coll. Sci. Univ. Kyôto, (A., Math.) 26 (1950), 5–14.
[6] T. Radó, Length and area, Amer. Math. Soc. Colloq. Publ., 1948.
[7] T. Radó, Lebesgue area and Hausdorff measure, Fund. Math., 44 (1957), 198–237.
[8] S. Saks, Theory of the integral, Warsaw, 1937.
[9] A. S. Besikovich (Besicovitch) and H. D. Ursell, Sets of fractional dimensions V, On dimensional numbers of some continuous curves, J. London Math. Soc., 12 (1937), 18–25.
[10] B. B. Mandelbrot, Fractals: Form, chance, and dimension, Freeman, 1977.
[11] H. Federer, Geometric measure theory, Springer, 1969.

247 (XXI.36) Lie, Marius Sophus

Marius Sophus Lie (December 17, 1842–February 18, 1899), a Norwegian mathema-

tician, is famous as the founder of the theory of †Lie groups. From 1869 to 1870 he collaborated with F. Klein on †sphere geometry, which led Lie to develop the concept of continuous groups. This discovery was the stepping stone that allowed Klein to complete his ideas for the †Erlangen program. In 1872, Lie became a professor at the University of Christiania (now Oslo). In 1886, Lie succeeded Klein in the chair of mathematics at Leipzig, where he remained until 1898. Then he returned to Christiania, where a post was created for him, but he died one year later.

The continuous groups that Lie dealt with are today called the †Lie transformation group germs. With the free use of geometric concepts and analytic methods (especially the theory of †differential equations) he was able to develop his theory and apply it to the theory of differential equations. The significance of his work was not recognized until after his death. Early in the 20th century, E. †Cartan and H. †Weyl were able to complete the theory of Lie groups, and by the middle of the century, the characteristics of the Lie group as a †topological group were clarified.

References

[1] F. Engel and P. Heegaard (eds.), Sophus Lie, Gesammelte Abhandlungen I–VI, Teubner, 1922–1937.
[2] S. Lie and F. Engel, Theorie der Transformationsgruppen I–III, Teubner, 1888–1893 (Chelsea, second edition, 1970).
[3] S. Lie and G. Scheffers, Vorlesungen über continuierlicher Gruppen, Teubner, 1893 (Chelsea, second edition, 1971).

248 (IV.10) Lie Algebras

A. Basic Concepts

Let K be a †commutative ring with unity. A set $\mathfrak{g}$ is called a **Lie algebra** over K if the following four conditions are satisfied: (i) $\mathfrak{g}$ is a †left K-module, where we assume that the unity of K acts on $\mathfrak{g}$ as the identity operator. (ii) There is given a K-bilinear mapping (called the **bracket product**) $(X, Y)\rightarrow[X, Y]$ from $\mathfrak{g}\times\mathfrak{g}$ into $\mathfrak{g}$:

$$\left[\sum\alpha_i X_i, \sum\beta_j Y_j\right]=\sum\alpha_i\beta_j[X_i, Y_j]$$

for all α_i, β_j in K and X_i, Y_j in $\mathfrak{g}$. (iii) $[X, X]=0$ for every X in $\mathfrak{g}$. (Hence $[X, Y]=-[Y, X]$ for every X, Y in $\mathfrak{g}$ (**alternating law**).) (iv) $[X,[Y,Z]]+[Y,[Z,X]]+[Z,[X,Y]]=0$ for every X, Y, Z in $\mathfrak{g}$ (**Jacobi identity**). In particular, if $K=\mathbf{C}$ (the complex number field) or $K=\mathbf{R}$ (the real number field), $\mathfrak{g}$ is called a **complex Lie algebra** or a **real Lie algebra**, respectively.

For example, let $\mathfrak{A}$ be an †associative algebra over K. Putting $[X, Y]=XY-YX$, we can supply $\mathfrak{A}$ with the structure of a Lie algebra over K, which is called the Lie algebra associated with $\mathfrak{A}$. In particular, if $\mathfrak{A}$ is the †total matrix algebra K_n of degree n over K, then the Lie algebra associated with K_n is called the **general linear Lie algebra** of degree n over K and is denoted by $\mathfrak{gl}(n, K)$.

Let $\mathfrak{g}$ be a Lie algebra over K and $\mathfrak{a}$, $\mathfrak{b}$ be †K-submodules of $\mathfrak{g}$. The subset of $\mathfrak{g}$ consisting of elements of the form $\sum[A, B]$ (finite sum) with $A\in\mathfrak{a}$, $B\in\mathfrak{b}$ is denoted by $[\mathfrak{a},\mathfrak{b}]$, which is a K-submodule of $\mathfrak{g}$. A K-submodule $\mathfrak{a}$ of $\mathfrak{g}$ is called a **Lie subalgebra** of $\mathfrak{g}$ if $[\mathfrak{a},\mathfrak{a}]\subset\mathfrak{a}$. A subalgebra $\mathfrak{a}$ of $\mathfrak{g}$ is called an **ideal** of $\mathfrak{g}$ if $[\mathfrak{a},\mathfrak{g}]\subset\mathfrak{a}$ (this condition is equivalent to $[\mathfrak{g},\mathfrak{a}]\subset\mathfrak{a}$). If $\mathfrak{a}$ is a subalgebra of $\mathfrak{g}$, the restriction of the bracket product of $\mathfrak{g}$ on $\mathfrak{a}$ makes $\mathfrak{a}$ a Lie algebra over K. If $\mathfrak{a}$ is an ideal of $\mathfrak{g}$, the †quotient K-module $\mathfrak{g}/\mathfrak{a}$ is a Lie algebra over K relative to the bracket product $[X+\mathfrak{a}, Y+\mathfrak{a}]=[X, Y]+\mathfrak{a}$. This Lie algebra $\mathfrak{g}/\mathfrak{a}$ is called the **quotient Lie algebra** of $\mathfrak{g}$ modulo $\mathfrak{a}$.

Let $\mathfrak{g}_1$, $\mathfrak{g}_2$ be Lie algebras over K. A mapping $f:\mathfrak{g}_1\rightarrow\mathfrak{g}_2$ is called a **homomorphism** of $\mathfrak{g}_1$ into $\mathfrak{g}_2$ if f is K-linear and $f([X, Y])=[f(X), f(Y)]$ for every X, Y in $\mathfrak{g}_1$. A bijective homomorphism is called an **isomorphism**. Then $\mathfrak{g}_1$ is said to be **isomorphic** to $\mathfrak{g}_2$ if there exists an isomorphism from $\mathfrak{g}_1$ onto $\mathfrak{g}_2$, and we write $\mathfrak{g}_1\cong\mathfrak{g}_2$. If $f:\mathfrak{g}_1\rightarrow\mathfrak{g}_2$ is a homomorphism, then $f(\mathfrak{g}_1)$ is a subalgebra of $\mathfrak{g}_2$, and the kernel $\mathfrak{a}=f^{-1}(0)$ of f is an ideal of $\mathfrak{g}_1$. Furthermore, the homomorphism f induces an isomorphism $\tilde{f}:\mathfrak{g}_1/\mathfrak{a}\rightarrow f(\mathfrak{g}_1)$ (**homomorphism theorem**).

The **direct sum** $\mathfrak{g}_1+\mathfrak{g}_2$ of two Lie algebras $\mathfrak{g}_1$, $\mathfrak{g}_2$ over K is defined as in the case of associative algebras. Then $\mathfrak{g}_1$, $\mathfrak{g}_2$ are ideals of $\mathfrak{g}_1+\mathfrak{g}_2$.

The set $A(\mathfrak{g})$ of all automorphisms of a Lie algebra $\mathfrak{g}$ is a subgroup of the general linear group $GL(\mathfrak{g})$. $A(\mathfrak{g})$ is called the (full) **automorphism group** of $\mathfrak{g}$.

B. Representations

Let $\mathfrak{g}$ be a Lie algebra over K, and let V be a K-module. Denote by $\mathfrak{L}(V)$ the associative algebra consisting of all K-linear mappings from V into V. Denote by $\mathfrak{gl}(V)$ the Lie algebra associated with $\mathfrak{L}(V)$. (Note that if V has a basis consisting of m elements over K, then $\mathfrak{gl}(V)\cong\mathfrak{gl}(m, K)$.) A homomorphism $\rho:\mathfrak{g}\rightarrow\mathfrak{gl}(V)$

is called a **representation** (more precisely, a **linear representation**) of $\mathfrak{g}$ over V, and V is called the **representation space** of ρ. If V is a †free K-module of rank m, then m is called the **degree** of the representation ρ. We also use (ρ, V) instead of ρ to mention the representation space explicitly. The concepts concerning the representations such as †equivalence, †irreducibility, or complete reducibility are similar to the corresponding concepts found in the representation theory of associative algebras (→ 362 Representations). In particular, by taking $V=\mathfrak{g}$ and putting $\rho(X)Y=[X, Y]$ $(X, Y\in\mathfrak{g})$, we obtain a representation of $\mathfrak{g}$, called the **adjoint representation** of $\mathfrak{g}$, and $\rho(X)$ is denoted by $\mathrm{ad}(X)$. Then $\mathrm{ad}(\mathfrak{g})=\{\mathrm{ad}(X)\mid X\in\mathfrak{g}\}$ is a subalgebra of $\mathfrak{gl}(\mathfrak{g})$ and is called the **adjoint Lie algebra** of $\mathfrak{g}$.

Let (ρ, V) be a representation of $\mathfrak{g}$. Then there is associated with this representation a †symmetric bilinear form $B_\rho:\mathfrak{g}\times\mathfrak{g}\to K$ given by $B_\rho(X, Y)=\mathrm{tr}\,\rho(X)\rho(Y)$, where B_ρ satisfies the invariance property $B_\rho([X,Z], Y)=B_\rho(X,[Z, Y])$. In particular, if $\rho=\mathrm{ad}$, then we write B instead of B_ρ, and B is called the **Killing form** of $\mathfrak{g}$.

Let $\mathfrak{g}$ be a Lie algebra over $\mathbf{R}$ of dimension n, and let $\mathrm{ad}:\mathfrak{g}\to\mathfrak{gl}(\mathfrak{g})$ be the adjoint representation of $\mathfrak{g}$. Put $\det(tI-\mathrm{ad}(X))=\sum_{j=0}^{n}t^jP_j(X)$ for every element $X\in\mathfrak{g}$. Then the $P_j(X)$ are polynomial functions on $\mathfrak{g}$, and $P_n=1$. Let l be the least integer such that $P_l\neq 0$. Then $l=$ †rank $\mathfrak{g}$. An element X of $\mathfrak{g}$ is called **regular (singular)** if $P_l(X)\neq 0$ $(P_l(X)=0)$. The subset $\mathfrak{g}'$ of $\mathfrak{g}$ consisting of all regular elements of $\mathfrak{g}$ is open and dense in $\mathfrak{g}$. The subset $\mathfrak{g}-\mathfrak{g}'$ of $\mathfrak{g}$ consisting of all singular elements of $\mathfrak{g}$ is of measure zero with respect to the †Lebesgue measure of $\mathfrak{g}$, which is obtained uniquely up to positive scalar multiples by means of a linear isomorphism of $\mathfrak{g}$ onto $\mathbf{R}^n$ using any basis of $\mathfrak{g}$ over $\mathbf{R}$.

Now suppose that $\mathfrak{g}$ is reductive (→ Section G). Then an element $X\in\mathfrak{g}$ is regular if and only if the centralizer $\mathfrak{z}_X=\{Y\in\mathfrak{g}\mid\mathrm{ad}(X)Y=0\}$ of X is a Cartan subalgebra (→ Section I) of $\mathfrak{g}$. Furthermore, if $X\in\mathfrak{g}$ is regular, then $\mathrm{ad}(X)$ is a †semisimple linear endomorphism of $\mathfrak{g}$.

C. Structure of Lie Algebras

Suppose that $\mathfrak{a}$, $\mathfrak{b}$ are ideals of a Lie algebra $\mathfrak{g}$. Then $[\mathfrak{a},\mathfrak{b}]$ is also an ideal of $\mathfrak{g}$. In particular, $\mathfrak{g}$ has the following ideals: $\mathfrak{g}'=[\mathfrak{g},\mathfrak{g}]$, $\mathfrak{g}''=[\mathfrak{g}',\mathfrak{g}'],\ldots,\mathfrak{g}^{(i+1)}=[\mathfrak{g}^{(i)},\mathfrak{g}^{(i)}],\ldots$. Furthermore, we have $\mathfrak{g}\supset\mathfrak{g}'\supset\mathfrak{g}''\supset\ldots$. This series is called the **derived series** of $\mathfrak{g}$, and $\mathfrak{g}'$ is called the **derived algebra** of $\mathfrak{g}$. The Lie algebra $\mathfrak{g}$ is said to be **Abelian** if $\mathfrak{g}'=0$ and **solvable** if $\mathfrak{g}^{(k)}=0$ for some k. Now put $\mathfrak{g}^1=\mathfrak{g}$, $\mathfrak{g}^2=[\mathfrak{g},\mathfrak{g}^1]$, $\mathfrak{g}^3=[\mathfrak{g},\mathfrak{g}^2],\ldots,\mathfrak{g}^{i+1}=[\mathfrak{g},\mathfrak{g}^i],\ldots$. Then $\mathfrak{g}^1$, $\mathfrak{g}^2,\ldots$ are all ideals of $\mathfrak{g}$, and we have $\mathfrak{g}^1\supset\mathfrak{g}^2\supset\mathfrak{g}^3\supset\ldots$. This series is called the **descending central series** of $\mathfrak{g}$, and $\mathfrak{g}$ is said to be **nilpotent** if $\mathfrak{g}^k=0$ for some k. An ideal $\mathfrak{a}$ of $\mathfrak{g}$ is called **Abelian (solvable, nilpotent)** if the subalgebra $\mathfrak{a}$ is Abelian (solvable, nilpotent).

Put $\mathfrak{z}=\{A\in\mathfrak{g}\mid[X,A]=0$ for every X in $\mathfrak{g}\}$. Then $\mathfrak{z}$ is an Abelian ideal of $\mathfrak{g}$, called the **center** of $\mathfrak{g}$, and is the kernel of the adjoint representation of $\mathfrak{g}$. Define the ideals $\mathfrak{z}_1,\mathfrak{z}_2,\ldots$ of $\mathfrak{g}$ as follows: $\mathfrak{z}_1$ is the center of $\mathfrak{g}$, $\mathfrak{z}_2/\mathfrak{z}_1$ is the center of $\mathfrak{g}/\mathfrak{z}_1,\ldots,\mathfrak{z}_{i+1}/\mathfrak{z}_i$ is the center of $\mathfrak{g}/\mathfrak{z}_i,\ldots$. Then we have $0\subset\mathfrak{z}_1\subset\mathfrak{z}_2\subset\ldots$. This series is called the **ascending central series** of $\mathfrak{g}$, and $\mathfrak{g}$ is nilpotent if and only if $\mathfrak{z}_k=\mathfrak{g}$ for some k.

We assume that K is a field of characteristic 0 and Lie algebras over K are of finite dimension. Let $X_1,\ldots,X_n$ be a basis of $\mathfrak{g}$ over K. Then the n^3 elements c_{ij}^k in K defined by $[X_i, X_j]=\sum c_{ij}^kX_k$ are called the **structural constants** of $\mathfrak{g}$ relative to the basis (X_i).

D. Radicals and Largest Nilpotent Ideals

The union $\mathfrak{r}$ of all solvable ideals of $\mathfrak{g}$ is also a solvable ideal of $\mathfrak{g}$, called the **radical** of $\mathfrak{g}$. The union $\mathfrak{n}$ of all nilpotent ideals of $\mathfrak{g}$ is also a nilpotent ideal of $\mathfrak{g}$, called the **largest nilpotent ideal** of $\mathfrak{g}$. The ideal $\mathfrak{s}=[\mathfrak{r},\mathfrak{g}]$ is called the **nilpotent radical** of $\mathfrak{g}$. We have $\mathfrak{g}\supset\mathfrak{r}\supset\mathfrak{n}\supset\mathfrak{s}$.

E. Semisimplicity

A Lie algebra is called **semisimple** if its radical is 0. A semisimple Lie algebra $\mathfrak{g}$ over K is called **simple** if $\mathfrak{g}$ has no ideals other than $\mathfrak{g}$ and 0. If $\mathfrak{r}$ is the radical of a Lie algebra $\mathfrak{g}$, then $\mathfrak{g}/\mathfrak{r}$ is semisimple. Every semisimple Lie algebra is a direct sum of simple Lie algebras.

For example, put $\mathfrak{t}(n,K)=\{A=(a_{ij})\in\mathfrak{gl}(n,K)\mid a_{ij}=0$ for every $i<j\}$ and $\mathfrak{n}(n,K)=\{A=(a_{ij})\in\mathfrak{t}(n,K)\mid a_{11}=a_{22}=\ldots=a_{nn}=0\}$. Note that $\mathfrak{t}(n,K)$ is the set of all lower triangular matrices and $\mathfrak{n}(n,K)$ is the set of nilpotent lower triangular matrices. Then $\mathfrak{t}(n,K)$ is a solvable subalgebra of $\mathfrak{gl}(n,K)$, and $\mathfrak{n}(n,K)$ is a nilpotent subalgebra of $\mathfrak{gl}(n,K)$. Put $\mathfrak{sl}(n,K)=\{A\in\mathfrak{gl}(n,K)\mid\mathrm{tr}\,A=0\}$. Then $\mathfrak{sl}(n,K)$ is an ideal of $\mathfrak{gl}(n,K)$. For $n\geqslant 2$, $\mathfrak{sl}(n,K)$ is a simple Lie algebra.

F. Theorems

The following theorems are fundamental in the theory of Lie algebras:

(1) **Engel's theorem** (valid even if K is of positive characteristic): Let V be a finite-dimensional vector space over a field K such that $V\neq\{0\}$. Let $\mathfrak{g}$ be a subalgebra of $\mathfrak{gl}(V)$

consisting of nilpotent elements. Then there is a nonzero element v in V such that $Xv=0$ for every X in $\mathfrak{g}$. (Thus by choosing a suitable basis of V and identifying $\mathfrak{gl}(V)$ with $\mathfrak{gl}(n, K)$, we have $\mathfrak{g}\subset\mathfrak{n}(n, K)$, where $n=\dim V$.)

(2) **Lie's theorem**: Let (ρ, V) be an irreducible representation of a solvable Lie algebra $\mathfrak{g}$. Then $\rho(\mathfrak{g})$ is Abelian. In particular, if K is algebraically closed, then $\dim V=1$. (Thus for every representation (ρ, V) of a solvable Lie algebra $\mathfrak{g}$ over an †algebraically closed field K, we have $\rho(\mathfrak{g})\subset\mathfrak{t}(n, K)$ by choosing a suitable basis of V.)

(3) **Cartan's criterion of solvability**: Let $\mathfrak{g}$ be a subalgebra of $\mathfrak{gl}(n, K)$. Then $\mathfrak{g}$ is solvable if and only if $\operatorname{tr} XY=0$ for every $X\in\mathfrak{g}$ and $Y\in[\mathfrak{g}, \mathfrak{g}]$.

(4) **Cartan's criterion of semisimplicity**: A Lie algebra $\mathfrak{g}$ is semisimple if and only if the Killing form B of $\mathfrak{g}$ is †nondegenerate (i.e., $B(X, \mathfrak{g})=0$, $X\in\mathfrak{g}$ implies $X=0$).

(5) **Weyl's theorem**: Every representation (of finite degree) of a semisimple Lie algebra is completely reducible.

(6) **Levi decomposition**: Let $\mathfrak{r}$ be the radical of a Lie algebra $\mathfrak{g}$. Then there is a semisimple subalgebra $\mathfrak{s}$ of $\mathfrak{g}$ such that $\mathfrak{g}=\mathfrak{r}+\mathfrak{s}$, $\mathfrak{r}\cap\mathfrak{s}=0$. Furthermore, such a subalgebra $\mathfrak{s}$ is unique up to automorphisms of $\mathfrak{g}$ (A. I. Mal'tsev).

(7) **Ado's theorem** (orginally proved only for the case of characteristic 0 for K; the case of positive characteristic was proved by K. Iwasawa): Let $\mathfrak{g}$ be a finite-dimensional Lie algebra over a field K. Then there exists a representation (ρ, V) of $\mathfrak{g}$ of finite degree such that $\mathfrak{g}\cong\rho(\mathfrak{g})$.

G. Reductive Lie Algebras

A Lie algebra $\mathfrak{g}$ is called **reductive** if the radical $\mathfrak{r}$ of $\mathfrak{g}$ coincides with the center $\mathfrak{z}$ of $\mathfrak{g}$. The following four conditions for a Lie algebra $\mathfrak{g}$ are mutually equivalent: (i) $\mathfrak{g}$ is reductive; (ii) the nilpotent radical $\mathfrak{s}$ of $\mathfrak{g}$ is 0; (iii) the adjoint representation of $\mathfrak{g}$ is completely reducible; and (iv) the derived algebra $[\mathfrak{g}, \mathfrak{g}]$ of $\mathfrak{g}$ is semisimple and $\mathfrak{g}=\mathfrak{z}+[\mathfrak{g}, \mathfrak{g}]$ (direct sum), where $\mathfrak{z}$ is the center of $\mathfrak{g}$.

A representation (ρ, V) of a reductive Lie algebra $\mathfrak{g}$ is completely reducible if and only if $\rho(X)$ is diagonalizable for every X in $\mathfrak{z}$. For example, the Lie algebra $\mathfrak{gl}(n, K)$ is reductive.

H. Derivations

A linear mapping $\delta:\mathfrak{g}\to\mathfrak{g}$ is called a **derivation** of the Lie algebra $\mathfrak{g}$ if $\delta([X, Y])=[\delta(X), Y]+[X, \delta(Y)]$ for every X, Y in $\mathfrak{g}$. The set $\mathfrak{D}(\mathfrak{g})$ of all derivations of $\mathfrak{g}$ is a subalgebra of $\mathfrak{gl}(\mathfrak{g})$, and $\mathfrak{D}(\mathfrak{g})$ is called the **Lie algebra of derivations** of $\mathfrak{g}$. The adjoint Lie algebra $\operatorname{ad}(\mathfrak{g})$ is an ideal of $\mathfrak{D}(\mathfrak{g})$, and elements of $\operatorname{ad}(\mathfrak{g})$ are called **inner derivations** of $\mathfrak{g}$. If $\mathfrak{g}$ is semisimple, then $\mathfrak{D}(\mathfrak{g})=\operatorname{ad}(\mathfrak{g})\cong\mathfrak{g}$.

Now suppose that $K=\mathbf{R}$ ($K=\mathbf{C}$). Then the group $A(\mathfrak{g})$ of automorphisms of $\mathfrak{g}$ is a Lie group (complex Lie group), and the Lie algebra of $A(\mathfrak{g})$ is given by $\mathfrak{D}(\mathfrak{g})$. Lie $\delta\in\mathfrak{D}(\mathfrak{g})$ ($\subset\mathfrak{gl}(\mathfrak{g})$). Then $\exp\delta$ ($\in GL(\mathfrak{g})$) is in $A(\mathfrak{g})$. The connected subgroup $I(\mathfrak{g})$ of $A(\mathfrak{g})$ generated by $\{\exp\delta\,|\,\delta\in\operatorname{ad}(\mathfrak{g})\}$ is a †Lie subgroup of $A(\mathfrak{g})$. Furthermore, $I(\mathfrak{g})$ is a normal subgroup of $A(\mathfrak{g})$, called the **group of inner automorphisms** of $\mathfrak{g}$ or the **adjoint group** of $\mathfrak{g}$. Thus $\operatorname{ad}(\mathfrak{g})$ is the Lie algebra associated with $I(\mathfrak{g})$. The quotient group $A(\mathfrak{g})/I(\mathfrak{g})$ is called the **group of outer automorphisms** of $\mathfrak{g}$. If $\mathfrak{g}$ is semisimple, then $I(\mathfrak{g})$ coincides with the identity component of $A(\mathfrak{g})$.

I. Cartan Subalgebras

A subalgebra $\mathfrak{h}$ of a Lie algebra $\mathfrak{g}$ over K is called a **Cartan subalgebra** of $\mathfrak{g}$ if (i) $\mathfrak{h}$ is nilpotent and (ii) the normalizer $\mathfrak{n}$ of $\mathfrak{h}$ in $\mathfrak{g}$ (i.e., $\mathfrak{n}=\{X\in\mathfrak{g}\,|\,[X, \mathfrak{h}]\subset\mathfrak{h}\}$) coincides with $\mathfrak{h}$. If K is algebraically closed, then for every two Cartan subalgebras $\mathfrak{h}_1$, $\mathfrak{h}_2$ of $\mathfrak{g}$ there exists an automorphism σ of $\mathfrak{g}$ such that $\sigma(\mathfrak{h}_1)=\mathfrak{h}_2$. Furthermore, for such a σ we can take an automorphism of the form $\sigma=\exp(\operatorname{ad}(A_1))\ldots\exp(\operatorname{ad}(A_r))$ ($A_1, \ldots, A_r\in\mathfrak{g}$), where all the $\operatorname{ad}(A_i)$ are nilpotent.

J. Universal Enveloping Algebras

Let $\mathfrak{g}$ be a Lie algebra over a field K. Regarding $\mathfrak{g}$ as a vector space over K, let $T(\mathfrak{g})$ be the †tensor algebra over $\mathfrak{g}$. Let J be the †two-sided ideal of $T(\mathfrak{g})$ generated by all elements of the form $X\otimes Y-Y\otimes X-[X, Y]$ ($X, Y\in\mathfrak{g}$). The quotient associative algebra $U(\mathfrak{g})=T(\mathfrak{g})/J$ is called the **universal enveloping algebra** of $\mathfrak{g}$. The composite of the natural mappings $\mathfrak{g}\to T(\mathfrak{g})\to U(\mathfrak{g})$ is an injection $\mathfrak{g}\to U(\mathfrak{g})$, and we identify $\mathfrak{g}$ with a linear subspace of $U(\mathfrak{g})$ by this mapping. Then we have $[X, Y]=XY-YX$ ($X, Y\in\mathfrak{g}$) in $U(\mathfrak{g})$. The algebra $U(\mathfrak{g})$ has no †zero divisors. In particular, if $\mathfrak{g}$ is the Lie algebra of a connected †Lie group G, then $U(\mathfrak{g})$ is isomorphic to the associative algebra of all †left-invariant differential operators on G. For every subalgebra $\mathfrak{h}$ of $\mathfrak{g}$, the universal enveloping algebra $U(\mathfrak{h})$ of $\mathfrak{h}$ is isomorphic to the subalgebra of the associative algebra $U(\mathfrak{g})$ generated by 1 and $\mathfrak{h}$. If $\mathfrak{g}$ is the direct sum of two Lie algebras $\mathfrak{g}_1$, $\mathfrak{g}_2$, then $U(\mathfrak{g})$ is isomorphic to the †tensor product $U(\mathfrak{g}_1)\otimes_K U(\mathfrak{g}_2)$. Let $\mathfrak{a}$ be an ideal of $\mathfrak{g}$, and let $\mathfrak{A}$ be the two-sided ideal of $U(\mathfrak{g})$ generated by $\mathfrak{a}$. Then we have $U(\mathfrak{g})/\mathfrak{A}\cong$

$U(\mathfrak{g}/\mathfrak{a})$. Now put $U_0 = K \cdot 1$. Define a linear subspace U_i of $U(\mathfrak{g})$ by

$$U_i = K \cdot 1 + \mathfrak{g} + \mathfrak{g} \cdot \mathfrak{g} + \ldots + \underbrace{\mathfrak{g} \ldots \mathfrak{g}}_{i}.$$

Then we have $U_0 \subset U_1 \subset \ldots, U_i U_j \subset U_{i+j}, \bigcup_i U_i = U(\mathfrak{g})$. Thus $\{U_i\}$ defines a †filtration of $U(\mathfrak{g})$. Denote by $G = G^0 + G^1 + G^2 + \ldots$ ($G^0 = U_0$, $G^i = U_i/U_{i-1}$) the †graded ring associated with this filtration. Then we have $\mathfrak{g} = G^1 \subset G$. Let $X_1, \ldots, X_n$ be a basis of $\mathfrak{g}$, and let $S = K[Y_1, \ldots, Y_n]$ be a †polynomial ring on K in n indeterminates $Y_1, \ldots, Y_n$. Then there exists a unique algebra homomorphism $\omega: S \rightarrow G$ such that $\omega(1) = 1$, $\omega(Y_i) = X_i$ ($i = 1, \ldots, n$). Furthermore, ω is bijective, and the ith homogeneous component S^i is mapped by ω onto G^i. Thus the set of monomials $\{X_1^{i_1} X_2^{i_2} \ldots X_n^{i_n}\}$ ($i_1 \geqslant 0, \ldots, i_n \geqslant 0$) forms a basis of $U(\mathfrak{g})$ over K (the **Poincaré-Birkhoff-Witt theorem**).

Every representation (ρ, V) of $\mathfrak{g}$ over K can be extended to a unique representation (ρ', V) of $U(\mathfrak{g})$. Furthermore, ρ is irreducible (completely reducible) if and only if ρ' is irreducible (completely reducible). Given two representations ρ_1, ρ_2 of $\mathfrak{g}$, ρ_1 is equivalent to ρ_2 if and only if ρ'_1 is equivalent to ρ'_2.

Now suppose that $\mathfrak{g}$ is semisimple, and let $X_1, \ldots, X_n$ be a basis of $\mathfrak{g}$. Using the Killing form B of $\mathfrak{g}$, put $g_{ij} = B(X_i, X_j)$. Denote the inverse matrix of (g_{ij}) by (g^{ij}). Define $c \in U(\mathfrak{g})$ by $c = \sum g^{ij} X_i X_j$. The element c, called the **Casimir element** of the Lie algebra, is independent of the choice of the basis (X_i), is a well-defined element of $U(\mathfrak{g})$, and belongs to the center of $U(\mathfrak{g})$. For every absolutely irreducible representation ρ of $U(\mathfrak{g})$, $\rho(c)$ is a scalar operator, and $\operatorname{tr}\rho(c)$ is a positive rational number.

K. Complex Semisimple Lie Algebras

We assume that $K = \mathbf{C}$, although there is no essential change if we assume that K is an algebraically closed field of characteristic 0.

A subalgebra $\mathfrak{h}$ of a complex semisimple Lie algebra $\mathfrak{g}$ is a Cartan subalgebra of $\mathfrak{g}$ if and only if $\mathfrak{h}$ is a maximal Abelian subalgebra of $\mathfrak{g}$ such that $\operatorname{ad}(H)$ is diagonalizable for every H in $\mathfrak{h}$. We fix a Cartan subalgebra $\mathfrak{h}$; $\dim \mathfrak{h}$ is called the **rank** of $\mathfrak{g}$, and we denote the linear space consisting of all $\mathbf{C}$-valued forms on $\mathfrak{h}$ by $\mathfrak{h}^*$. For every α in $\mathfrak{h}^*$, let

$$\mathfrak{g}_\alpha = \{X \in \mathfrak{g} \mid \operatorname{ad}(H)X = \alpha(H)X \text{ for all } H \text{ in } \mathfrak{h}\}.$$

Then $\mathfrak{g}_\alpha$ is a linear subspace of $\mathfrak{g}$, and $\mathfrak{g}_0 = \mathfrak{h}$. Define a subset Δ of $\mathfrak{h}^*$ by

$$\Delta = \{\alpha \in \mathfrak{h}^* \mid \alpha \neq 0, \mathfrak{g}_\alpha \neq \{0\}\}.$$

Then Δ is a finite set. Elements of Δ are called **roots** of $\mathfrak{g}$ relative to $\mathfrak{h}$, and Δ is called the **root system** of $\mathfrak{g}$ relative to $\mathfrak{h}$. For every root α, $\mathfrak{g}_\alpha$ is of dimension one, and $\mathfrak{g}$ is decomposed into a direct sum of linear subspaces:

$$\mathfrak{g} = \mathfrak{h} + \sum_{\alpha \in \Delta} \mathfrak{g}_\alpha.$$

For each root α, $\mathfrak{g}_\alpha$ is called the **root subspace** corresponding to α.

The restriction $B_{\mathfrak{h}}$ of the Killing form B of $\mathfrak{g}$ on $\mathfrak{h}$ is nondegenerate. Hence for every λ in $\mathfrak{h}^*$ there exists a unique element H_λ in $\mathfrak{h}$ such that $\lambda(H) = B(H_\lambda, H)$ for all H in $\mathfrak{h}$. Thus we get a linear bijection $\mathfrak{h}^* \rightarrow \mathfrak{h}$ defined by $\lambda \rightarrow H_\lambda$. Via this bijection, $B_{\mathfrak{h}}$ gives rise to a symmetric bilinear form $(\lambda, \mu) = (H_\lambda, H_\mu)$ ($\lambda, \mu \in \mathfrak{h}^*$) on $\mathfrak{h}^*$. Denote by $\mathfrak{h}^*_{\mathbf{R}}$ the real linear subspace of $\mathfrak{h}^*$ spanned by Δ. Then the inner product (λ, μ) defined on $\mathfrak{h}^*$ is positive definite on $\mathfrak{h}^*_{\mathbf{R}}$. Hence with respect to this inner product, $\mathfrak{h}^*_{\mathbf{R}}$ is an l-dimensional Euclidean space, where $l = \dim \mathfrak{h}$. The root system Δ is a finite subset of the Euclidean space $\mathfrak{h}^*_{\mathbf{R}}$.

L. Properties of Root Systems

(i) $\alpha \in \Delta$ implies $-\alpha \in \Delta$. Furthermore, among the scalar multiples of α, only $\pm\alpha$ belong to Δ. (ii) Let $\alpha, \beta \in \Delta$. Then $2(\alpha, \beta)/(\alpha, \alpha)$ is a rational integer. (iii) Let $\alpha, \beta \in \Delta$ and $\beta \neq \pm\alpha$. Then there exist unique nonnegative integers j, i such that $\{\beta + \nu\alpha \mid \nu \in \mathbf{Z}\} \cap \Delta = \{\beta - j\alpha, \beta - (j-1)\alpha, \ldots, \beta - \alpha, \beta, \beta + \alpha, \ldots, \beta + i\alpha\}$. Furthermore, $j - i = 2(\alpha, \beta)/(\alpha, \alpha)$, $i + j \leqslant 3$. The set $\{\beta + \mathbf{Z}\alpha\} \cap \Delta$ is called the **α-string** of β.

Now let $\alpha \in \Delta$. Denote by w_α the †reflection mapping of $\mathfrak{h}^*_{\mathbf{R}}$ with respect to the hyperplane $P_\alpha = \{x \in \mathfrak{h}^*_{\mathbf{R}} \mid (\alpha, x) = 0\}$, which is orthogonal to α. Then we have $w_\alpha(\beta) = \beta - (\alpha^*, \beta)\alpha \in \Delta$ for every β in Δ, where $\alpha^* = 2\alpha/(\alpha, \alpha)$. Thus we have $w_\alpha(\Delta) = \Delta$ for every α in Δ. (iv) Let $\alpha, \beta \in \Delta$ and $\beta \neq \pm\alpha$. Then the angle θ between α and β is one of the following: 30°, 45°, 60°, 90°, 120°, 135°, 150°. Suppose, moreover, that $0 \leqslant \theta \leqslant 90°$ and $(\alpha, \alpha) \leqslant (\beta, \beta)$. Then we have the following criteria: $\theta = 30° \Leftrightarrow 3(\alpha, \alpha) = (\beta, \beta)$; $\theta = 45° \Leftrightarrow 2(\alpha, \alpha) = (\beta, \beta)$; $\theta = 60° \Leftrightarrow (\alpha, \alpha) = (\beta, \beta)$. (v) Let α, β, $\alpha + \beta \in \Delta$. Then $[\mathfrak{g}_\alpha, \mathfrak{g}_\beta] = \mathfrak{g}_{\alpha+\beta}$.

Conversely, suppose that a finite subset Δ of a finite-dimensional Euclidean space E satisfies conditions (i) and (ii) together with a part of (iii): $w_\alpha(\Delta) = \Delta$ for every α in Δ. Then Δ is a root system of some complex semisimple Lie algebra.

M. Lexicographic Linear Ordering in $\mathfrak{h}^*_{\mathbf{R}}$

Let $\lambda_1, \ldots, \lambda_l$ be a basis of $\mathfrak{h}^*_{\mathbf{R}}$ over $\mathbf{R}$. Define a linear ordering $\lambda > \mu$ on $\mathfrak{h}^*_{\mathbf{R}}$ as follows: If $\lambda = \sum \xi_i \lambda_i$, $\mu = \sum \eta_i \lambda_i$ (ξ_i, η_i all in $\mathbf{R}$), then $\lambda > \mu$ if

and only if there exists an index s $(1 \leqslant s \leqslant l)$ such that $\xi_i = \eta_i$ for $i = 1, \ldots, s-1$ and $\xi_s > \eta_s$. This linear ordering is called the **lexicographic linear ordering** of $\mathfrak{h}_{\mathbf{R}}^*$ associated with the basis (λ_i). Relative to this linear ordering, a root α is called a **positive (negative) root** if $\alpha > 0$ $(\alpha < 0)$. We denote the subset of Δ consisting of all positive (negative) roots by Δ^+ (Δ^-). A subset S of Δ coincides with Δ^+ for some lexicographic linear ordering of $\mathfrak{h}_{\mathbf{R}}^*$ if and only if the following conditions are satisfied: $\Delta = S \cup (-S)$; $S \cap (-S) = \varnothing$; $\alpha, \beta \in S$, $\alpha + \beta \in \Delta$ imply $\alpha + \beta \in S$. A positive root $\alpha \in \Delta^+$ is called a **simple root** if α cannot be expressed as the sum of two positive roots.

N. Fundamental Root Systems

Let Π be a subset of Δ consisting of l roots $\alpha_1, \ldots, \alpha_l$. Then Π is called a **fundamental root system** of Δ if (i) every element α of Δ is expressed uniquely as an integral linear combination of the α_i $(\alpha = \sum m_i \alpha_i)$, and (ii) in this expression, $m_1, \ldots, m_l$ are either all $\geqslant 0$ or all $\leqslant 0$. For any lexicographic linear ordering of $\mathfrak{h}_{\mathbf{R}}^*$, the set of all simple roots forms a fundamental root system of Δ. Moreover, every fundamental root system of Δ is obtained in this manner. Let $\Pi = \{\alpha_1, \ldots, \alpha_l\}$ be a fundamental root system of Δ. Then $\sum \mathfrak{g}_{\alpha_i} + \sum \mathfrak{g}_{-\alpha_i}$ generates $\mathfrak{g}$. The Lie algebra $\mathfrak{g}$ is not simple if and only if Π admits an orthogonal partition, i.e., $\Pi = \Pi_1 \cup \Pi_2$, $\Pi_1 \neq \varnothing$, $\Pi_2 \neq \varnothing$, $\Pi_1 \cap \Pi_2 = \varnothing$, and $(\alpha, \beta) = 0$ for every $\alpha \in \Pi_1$ and $\beta \in \Pi_2$. The l^2 integers $a_{ij} = -2(\alpha_i, \alpha_j)/(\alpha_j, \alpha_j)$ $(1 \leqslant i, j \leqslant l)$ are called the **Cartan integers** of $\mathfrak{g}$ relative to the fundamental root system Π. Then we have $a_{ii} = -2$, $a_{ij} \geqslant 0$ for $i \neq j$.

O. Borel Subalgebras and Parabolic Subalgebras

Let $\mathfrak{b} = \mathfrak{h} + \sum_{\alpha > 0} \mathfrak{g}_\alpha$. Then $\mathfrak{b}$ is a maximal solvable subalgebra of $\mathfrak{g}$. The group $I(\mathfrak{g})$ acts transitively on the set of all maximal solvable subalgebras of $\mathfrak{g}$. A maximal solvable subalgebra of $\mathfrak{g}$ is called a **Borel subalgebra** of $\mathfrak{g}$. A subalgebra of $\mathfrak{g}$ is called a **parabolic subalgebra** if it contains a Borel subalgebra of $\mathfrak{g}$. Now let Φ be any subset of a given fundamental root system $\Pi = \{\alpha_1, \ldots, \alpha_l\}$. Denote by $\Delta^-(\Phi)$ the set of all negative roots $\alpha = \sum n_i \alpha_i$ such that $n_j = 0$ for all α_j in Φ. Then $\mathfrak{p}_\Phi = \mathfrak{b} + \sum_{\alpha \in \Delta^-(\Phi)} \mathfrak{g}_\alpha$ is a parabolic subalgebra. Thus we get 2^l parabolic subalgebras $\{\mathfrak{p}_\Phi \mid \Phi \subset \Pi\}$. Every parabolic subalgebra is conjugate under $I(\mathfrak{g})$ to one and only one of the parabolic subalgebras $\{\mathfrak{p}_\Phi \mid \Phi \subset \Pi\}$.

P. Weyl's Canonical Basis

Let $H_1, \ldots, H_l$ be a basis of $\mathfrak{h}$, and let E_α be a basis of $\mathfrak{g}_\alpha$ for each root α. Then we have a basis $\{H_i, E_\alpha\}$ of $\mathfrak{g}$. Such a basis is called **Weyl's canonical basis** if the following three conditions are satisfied: (i) $\alpha(H_j) \in \mathbf{R}$ $(j = 1, \ldots, l)$ for every $\alpha \in \Delta$; (ii) the Killing form B of $\mathfrak{g}$ satisfies $B(E_\alpha, E_{-\alpha}) = -1$ for every $\alpha \in \Delta$; and (iii) if $\alpha, \beta, \alpha + \beta \in \Delta$ and $[E_\alpha, E_\beta] = N_{\alpha,\beta} E_{\alpha+\beta}$ $(N_{\alpha,\beta} \in \mathbf{C})$, then $N_{\alpha,\beta}$ is in $\mathbf{R}$ and $N_{\alpha,\beta} = N_{-\alpha,-\beta}$. The Lie algebra $\mathfrak{g}$ always has Weyl's canonical basis. For such a basis $\{H_i, E_\alpha\}$, the linear space

$$\mathfrak{g}_u = \sum \mathbf{R}\sqrt{-1}\, H_j + \sum \mathbf{R}(E_\alpha + E_{-\alpha}) + \sum \mathbf{R}(\sqrt{-1}\,(E_\alpha - E_{-\alpha}))$$

is a semisimple Lie algebra over $\mathbf{R}$. The Killing form of $\mathfrak{g}_u$ is negative definite. Every connected Lie group whose Lie algebra is $\mathfrak{g}_u$ is always compact. Furthermore, $\mathfrak{g} = \mathfrak{g}_u + \sqrt{-1}\,\mathfrak{g}_u$, $\mathfrak{g}_u \cap \sqrt{-1}\,\mathfrak{g}_u = 0$. Thus $\mathfrak{g}$ is isomorphic to the Lie algebra $\mathfrak{g}_u^{\mathbf{C}} = \mathbf{C} \otimes_{\mathbf{R}} \mathfrak{g}_u$ over $\mathbf{C}$ obtained from $\mathfrak{g}_u$ by extending the basic field $\mathbf{R}$ to $\mathbf{C}$, and $\mathfrak{g}_u$ is called the **unitary restriction** of $\mathfrak{g}$ relative to Weyl's canonical basis $\{H_i, E_\alpha\}$.

A Lie algebra $\mathfrak{a}$ over $\mathbf{R}$ is called a **real form** of $\mathfrak{g}$ if $\mathfrak{a}^{\mathbf{C}} = \mathbf{C} \otimes_{\mathbf{R}} \mathfrak{a}$ is isomorphic to $\mathfrak{g}$. When this is the case, $\mathfrak{g}$ is called a **complex form** (or the **complexification**) of $\mathfrak{a}$. Note that a real form $\mathfrak{a}$ of $\mathfrak{g}$ can be regarded as a real subalgebra of $\mathfrak{g}$ such that $\mathfrak{g} = \mathfrak{a} + \sqrt{-1}\,\mathfrak{a}$, $\mathfrak{a} \cap \sqrt{-1}\,\mathfrak{a} = 0$. A real Lie algebra $\mathfrak{a}$ is called a **compact real Lie algebra** if its Killing form is negative definite. A real Lie algebra is compact if and only if it is semisimple and is the Lie algebra of some compact Lie group. The Lie algebra $\mathfrak{a}$ of a compact Lie group A is the direct sum of its center $\mathfrak{z}$ and some compact Lie algebra; hence $\mathfrak{a}$ is reductive.

A compact real form $\mathfrak{g}_u$ of a complex semisimple Lie algebra $\mathfrak{g}$ is called a **compact form** of $\mathfrak{g}$. The group $I(\mathfrak{g})$ of all inner automorphisms of $\mathfrak{g}$ acts transitively on the set of all compact real forms of $\mathfrak{g}$ (regarding the real forms of $\mathfrak{g}$ as real subalgebras of $\mathfrak{g}$).

Q. Chevalley's Canonical Basis

A complex semisimple Lie algebra $\mathfrak{g}$ always has a basis $\{H_i, E_\alpha\}$ (consisting of a basis $H_1, \ldots, H_l$ of $\mathfrak{h}$ together with a basis E_α of $\mathfrak{g}_\alpha$ for each root α) such that (i) $\alpha(H_i) \in \mathbf{Z}$ for every $\alpha \in \Delta$ and $i = 1, \ldots, l$; (ii) $B(E_\alpha, E_{-\alpha}) = 2/(\alpha, \alpha)$ for every $\alpha \in \Delta$; and (iii) if $\alpha, \beta, \alpha + \beta \in \Delta$ and $[E_\alpha, E_\beta] = N_{\alpha,\beta} E_{\alpha+\beta}$ $(N_{\alpha,\beta} \in \mathbf{C})$, then $N_{\alpha,\beta} \in \mathbf{Z}$ and $N_{\alpha,\beta} = -N_{-\alpha,-\beta}$. Such a basis $\{H_i, E_\alpha\}$ is called **Chevalley's canonical basis**. When we take this basis, the structural constants of $\mathfrak{g}$ relative to $\{H_i, E_\alpha\}$ are all integers. Thus $\mathfrak{g}_{\mathbf{Z}} = \sum \mathbf{Z} H_i +$

$\sum \mathbf{Z}E_\alpha$ is a Lie algebra over $\mathbf{Z}$. Furthermore, $\mathfrak{g}_{\mathbf{R}}=\sum \mathbf{R}H_i+\sum \mathbf{R}E_\alpha$ is a real form of $\mathfrak{g}$ with the property that there exists a Cartan subalgebra $\mathfrak{h}_{\mathbf{R}}$ of $\mathfrak{g}_{\mathbf{R}}$ such that for each element H in $\mathfrak{h}_{\mathbf{R}}$, all the eigenvalues of $\mathrm{ad}(H)$ on $\mathfrak{g}_{\mathbf{R}}$ are contained in $\mathbf{R}$. (In fact, we may take $\sum \mathbf{R}H_i$ as $\mathfrak{h}_{\mathbf{R}}$.) Such a real form of $\mathfrak{g}$ is called a **normal real form**. The group $I(\mathfrak{g})$ acts transitively on the set of all normal real forms of $\mathfrak{g}$.

R. Weyl Groups

The reflections w_α $(\alpha\in\Delta)$ of the Euclidean space $\mathfrak{h}_{\mathbf{R}}^*$ generate a subgroup W of the group of all †congruent transformations of $\mathfrak{h}_{\mathbf{R}}^*$. W is called the **Weyl group** of $\mathfrak{g}$ relative to $\mathfrak{h}$ and is represented faithfully as a †permutation group over the finite set Δ. Hence W is a finite group. Let $\Pi=\{\alpha_1,\ldots,\alpha_l\}$ be a fundamental root system of Δ. Then W is generated by $w_{\alpha_1},\ldots,w_{\alpha_l}$. The root system Δ coincides with the set $\{w(\alpha)\mid w\in W, \alpha\in\Pi\}$. Let $\mathfrak{F}$ be the set of all fundamental root systems of Δ. Then W acts on $\mathfrak{F}$ and is †simply transitive on $\mathfrak{F}$. If $\mathfrak{g}$ is simple, two roots α, β are conjugate under W if and only if $(\alpha,\alpha)=(\beta,\beta)$. Now let Γ be the complement in $\mathfrak{h}_{\mathbf{R}}^*$ of the union of all the hyperplanes P_α $(\alpha\in\Delta)$ orthogonal to α. Then Γ is a W-stable open subset of $\mathfrak{h}_{\mathbf{R}}^*$. A connected component of Γ is called a **Weyl chamber**. W acts on the set $\mathfrak{F}_0$ of all Weyl chambers and is simply transitive on $\mathfrak{F}_0$. Let $\Pi=\{\alpha_1,\ldots,\alpha_l\}$ be a fundamental root system. Then the set $\{x\in\mathfrak{h}_{\mathbf{R}}^*\mid (x,\alpha_i)>0$ for $i=1,\ldots,l\}$ is a Weyl chamber, called the **positive Weyl chamber** associated with Π. Now fix any lexicographic linear ordering of $\mathfrak{h}_{\mathbf{R}}^*$ that has Π as the set of simple roots. For $w\in W$, put $\Delta_w^+=\{\alpha\in\Delta^+\mid w(\alpha)\in\Delta^-\}$. Denote the cardinality of Δ_w^+ by $n(w)$. Then $n(w)=0\Leftrightarrow w=1$. Furthermore, w can be expressed as a product of $n(w)$ factors $w=w_{\alpha_i}\ldots w_{\alpha_j}$, where each factor is taken from $\{w_{\alpha_1},\ldots,w_{\alpha_l}\}$ admitting repetitions. In fact, $n(w)$ is the minimum length of the expression of w as a product $w=w_{\alpha_i}\ldots w_{\alpha_j}$ $(i,j=1,\ldots,l)$. With respect to the generators $w_{\alpha_1},\ldots,w_{\alpha_l}$, W has the following system of †defining relations:

$$w_{\alpha_i}^2=1, \quad 1\leqslant i\leqslant l,$$

$$(w_{\alpha_i}w_{\alpha_j})^{m_{ij}}=1, \quad 1\leqslant i<j\leqslant l,$$

where m_{ij} is the order of $w_{\alpha_i}w_{\alpha_j}$. Thus if θ_{ij} is the angle between α_i and α_j, we have $m_{ij}=\pi/(\pi-\theta_{ij})$.

Denote by T the set of all linear transformations σ of $\mathfrak{h}_{\mathbf{R}}^*$ such that $\sigma(\Delta)=\Delta$. Then T is also the subgroup of all congruent transformations of $\mathfrak{h}_{\mathbf{R}}^*$; furthermore, T is a finite group, and W is a normal subgroup of T. Let Π be a fundamental root system of Δ and put $P=\{\alpha\in T\mid\sigma(\Pi)=\Pi\}$. Then P is a subgroup of T, and we have a semidirect product $T=P\cdot W$. Elements of P are called **particular transformations** relative to Π. The group $P\cong T/W$ is isomorphic to the group $A(\mathfrak{g})/I(\mathfrak{g})$ of outer automorphisms of $\mathfrak{g}$.

S. Classification of Complex Simple Lie Algebras

Let $\mathfrak{h}$ be a Cartan subalgebra of a complex semisimple Lie algebra $\mathfrak{g}$, and let $\Pi=\{\alpha_1,\ldots,\alpha_l\}$ be a fundamental root system relative to $\mathfrak{h}$. We associate with Π the diagram (†1-dimensional complex) indicated in Fig. 1. This diagram is called the **Dynkin diagram** of $\mathfrak{g}$ (also called the **Schläfli diagram** or **Coxeter diagram**). It is constructed as follows: with each α_i there is associated a vertex (denoted by a small open circle). These l vertices are connected by several segments as follows. Let θ_{ij} be the angle between α_i and α_j. (i) If $\theta_{ij}=150°$, α_i and α_j are connected by three oriented segments as in (1) of Fig. 1, where the orientation means $(\alpha_i,\alpha_i)>(\alpha_j,\alpha_j)$. (ii) If $\theta_{ij}=135°$, α_i and α_j are connected by two oriented segments as in (2) of Fig. 1. (iii) If $\theta_{ij}=120°$, α_i and α_j are connected by a nonoriented single segment as in (3) of Fig. 1. (iv) If $\theta_{ij}=90°$, α_i and α_j are not connected.

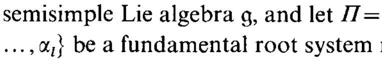


Fig. 1
A Dynkin diagram.

The Dynkin diagram of $\mathfrak{g}$ is independent of the choice of $\mathfrak{h}$, Π. Furthermore, two complex semisimple Lie algebras are isomorphic if and only if they have the same Dynkin diagram. A complex semisimple Lie algebra is simple if and only if its Dynkin diagram is connected.

Fig. 2 gives all possible Dynkin diagrams

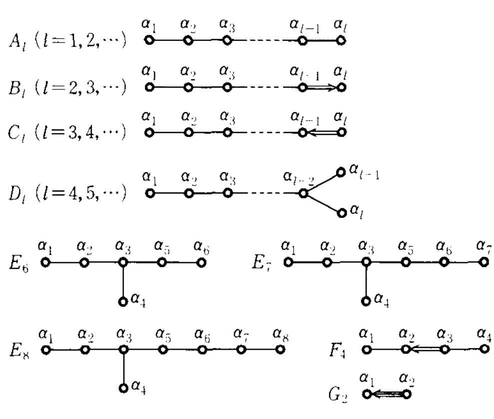


Fig. 2
Dynkin diagrams of simple Lie algebras. Note that $\dim\mathfrak{g}$ is A_l: l^2+2l; B_l: $2l^2+l$; C_l: $2l^2+l$; D_l: $2l^2-l$; E_6: 78; E_7: 133; E_8: 248; F_4: 52; G_2: 14.

associated with complex semisimple simple Lie algebras. There are seven categories. (The index l in A_l means the rank of $\mathfrak{g}$.) Among these A_l ($l \geqslant 1$), B_l ($l \geqslant 2$), C_l ($l \geqslant 3$), D_l ($l \geqslant 4$) are called **classical complex simple Lie algebras**. E_l ($l=6,7,8$), F_4, and G_2 are called **exceptional complex simple algebras**. Note that $A_1 \cong B_1 \cong C_1$, $B_2 \cong C_2$, $A_3 \cong D_3$, $D_2 = A_1 + A_1$. A_l (resp. B_l, C_l, D_l) is the Lie algebra of the complex Lie group $SL(l+1, \mathbf{C})$ ($SO(2l+1, \mathbf{C})$, $Sp(l, \mathbf{C})$, $SO(2l, \mathbf{C})$).

T. Classification of Real Simple Lie Algebras

We refer the reader to [3] and the references at the end of [3] for the classification of simple Lie algebras over a general field k; in particular, for $k = \mathbf{R}$ (→ Appendix A, Tables 5.I, 5.II), the algebras are closely related to the classification of irreducible symmetric Riemannian manifolds (→ 412 Symmetric Riemannian Spaces and Real Forms).

In particular, since compact semisimple real Lie algebras $\mathfrak{g}$ are in one-to-one correspondence (up to isomorphism) with complex semisimple Lie algebras $\mathfrak{g}^{\mathbf{C}}$ obtained as the complexification of $\mathfrak{g}$, the classification of compact real simple Lie algebras reduces to the classification of complex simple Lie algebras. Hence they are also represented by the same Dynkin diagrams. Compact real simple Lie algebras of the types A_l, B_l, C_l, and D_l are called **classical compact real simple Lie algebras**. They are the Lie algebras of the compact Lie groups $SU(l+1)$, $SO(2l+1)$, $Sp(l)$, and $SO(2l)$, respectively. Compact real simple Lie algebras of the type E_l ($l=6,7,8$), F_4, and G_2 are called **exceptional compact real simple Lie algebras**.

U. Satake Diagrams of Real Semisimple Lie Algebras

Let $\mathfrak{g}$ be a real semisimple Lie algebra, $\mathfrak{k}$ be the subalgebra associated with a †maximal compact subgroup of the †adjoint group of $\mathfrak{g}$, and $\mathfrak{p}$ be the orthogonal complement of $\mathfrak{k}$ in $\mathfrak{g}$ relative to the Killing form of $\mathfrak{g}$. Let $\mathfrak{a}$ be a maximal Abelian subalgebra contained in $\mathfrak{p}$, and let $\mathfrak{h}$ be a Cartan subalgebra of $\mathfrak{g}$ containing $\mathfrak{a}$. Denote by $\mathfrak{g}^{\mathbf{C}}$, $\mathfrak{h}^{\mathbf{C}}$ the complexifications of $\mathfrak{g}$, $\mathfrak{h}$, respectively. Let σ be the †semilinear †automorphism of $\mathfrak{g}^{\mathbf{C}}$ defined by $\sigma(X + \sqrt{-1}\, Y) = X - \sqrt{-1}\, Y$ ($X, Y \in \mathfrak{g}$). Then we have $\sigma(\mathfrak{h}^{\mathbf{C}}) = \mathfrak{h}^{\mathbf{C}}$. Thus σ acts on the root system Δ of $\mathfrak{g}^{\mathbf{C}}$ relative to $\mathfrak{h}^{\mathbf{C}}$ as follows: $(\sigma\alpha)(H) = \overline{\alpha(\sigma H)}$ ($\alpha \in \Delta$). Thus σ acts on $(\mathfrak{h}^{\mathbf{C}})^*_{\mathbf{R}}$. There is a lexicographic linear ordering of $(\mathfrak{h}^{\mathbf{C}})^*_{\mathbf{R}}$ such that $\alpha \in \Delta^+$ and $\sigma\alpha \neq -\alpha$ imply $\sigma\alpha \in \Delta^+$. Fix such an ordering, and let Π be the set of simple roots relative to the ordering. Put $\sigma = pw$, where p is a particular transformation relative to Π and w is an element in the Weyl group W. Then p induces a permutation of order 2 on the set $\{\alpha \in \Pi \mid \sigma\alpha \neq -\alpha\}$. Suppose that a vertex belonging to the Dynkin diagram of Π corresponds to a simple root α such that $\sigma\alpha = -\alpha$. Then replace the vertex by a small filled circle. Also, if two vertices are mapped to each other by p, then connect the two vertices by an arc with two arrows on the end. The diagram thus obtained is called the **Satake diagram** of $\mathfrak{g}$. The Satake diagram of $\mathfrak{g}$ is independent of the choice of $\mathfrak{k}$, $\mathfrak{a}$, $\mathfrak{h}$ and of the ordering of $(\mathfrak{h}^{\mathbf{C}})^*_{\mathbf{R}}$. Two real semisimple Lie algebras are isomorphic if and only if they have the same Satake diagram, and $\mathfrak{g}$ is simple if and only if its Satake diagram is connected. Thus real simple Lie algebras are classified by their Satake diagrams [16].

V. Iwasawa Decomposition of Real Semisimple Lie Algebras

Let $\mathfrak{g}$ be a real semisimple Lie algebra. Take $\mathfrak{k}$, $\mathfrak{a}$, $\mathfrak{h}$ and the ordering of $(\mathfrak{h}^{\mathbf{C}})^*_{\mathbf{R}}$ as in the construction of Satake diagrams (→ Section U). Let $\mathfrak{n}$ be the intersection of $\mathfrak{g}$ with the subspace $\sum(\mathfrak{g}^{\mathbf{C}})_\alpha$, where the sum is taken over $\alpha \in \Delta^+$ such that $\sigma\alpha \neq -\alpha$. Then $\mathfrak{n}$ is a nilpotent subalgebra of $\mathfrak{g}$, and we have a decomposition of $\mathfrak{g}$ into the direct sum of linear spaces: $\mathfrak{g} = \mathfrak{k} + \mathfrak{a} + \mathfrak{n}$. This decomposition is called an **Iwasawa decomposition** of $\mathfrak{g}$. Iwasawa decompositions are unique in the following sense: Let $\mathfrak{g} = \mathfrak{k}' + \mathfrak{a}' + \mathfrak{n}'$ be another Iwasawa decomposition; then there exists an inner automorphism A of $\mathfrak{g}$ such that $A\mathfrak{k} = \mathfrak{k}'$, $A\mathfrak{a} = \mathfrak{a}'$, $A\mathfrak{n} = \mathfrak{n}'$.

For the cohomology theory of Lie algebras and Lie algebras over fields of characteristic $p > 0$, in particular the theory of **restricted Lie algebras** → [3]. For the relationship between Lie algebras and the theory of finite groups (e.g., Chevalley's simple groups, †Burnside problems) → [7] and the references therein.

W. Representations

Let $\mathfrak{g}$ be a complex semisimple Lie algebra and $\mathfrak{h}$ be a Cartan subalgebra of $\mathfrak{g}$. We fix $\mathfrak{h}$ and a lexicographic linear ordering on $\mathfrak{h}^*_{\mathbf{R}}$. Let $\Pi = \{\alpha_1, \ldots, \alpha_l\}$ be the set of simple roots. Since every representation of $\mathfrak{g}$ is completely reducible, we restrict ourselves to the explanation of irreducible representations. Let (ρ, V) be a representation of $\mathfrak{g}$. For each $\lambda \in \mathfrak{h}^*$, put $V_\lambda = \{v \in V \mid \rho(H)v = \lambda(H)v \text{ for all } H \in \mathfrak{h}\}$. Then V_λ is a linear subspace of V, and λ is called a **weight** of the representation ρ (relative to $\mathfrak{h}$) if

$V_\lambda \neq \{0\}$; then $\dim V_\lambda$ is called the **multiplicity** of the weight λ. The set of all weights of ρ is a finite W-stable subset of $\mathfrak{h}_{\mathbf{R}}^*$. Denote this set by $\{\lambda_1, \ldots, \lambda_r\}$. Then V is decomposed into a direct sum: $V = V_{\lambda_1} + \ldots + V_{\lambda_r}$. The maximum element among $\{\lambda_1, \ldots, \lambda_r\}$ with respect to the given ordering of $\mathfrak{h}_{\mathbf{R}}^*$ is called the **highest weight** of ρ.

The following two theorems are basic in determining irreducible representations.

(1) **Cartan's theorem**: Let Λ_1, Λ_2 be the highest weights of irreducible representations ρ_1, ρ_2 of $\mathfrak{g}$, respectively. Then ρ_1 is equivalent to ρ_2 if and only if $\Lambda_1 = \Lambda_2$.

(2) **Cartan-Weyl theorem**: Let $\lambda \in \mathfrak{h}_{\mathbf{R}}^*$. Then there exists an irreducible representation ρ of $\mathfrak{g}$ which has λ as its highest weight if and only if (i) $2(\lambda, \alpha)/(\alpha, \alpha) \in \mathbf{Z}$ for every root $\alpha \in \Delta$ (such an element $\lambda \in \mathfrak{h}_{\mathbf{R}}^*$ is called an **integral form** on $\mathfrak{h}$); and (ii) $w(\lambda) \leqslant \lambda$ for every $w \in W$ (such an element $\lambda \in \mathfrak{h}_{\mathbf{R}}^*$ is called **dominant**).

These theorems lead to the concept of the system of fundamental representations. Put $\alpha_i^* = 2\alpha_i/(\alpha_i, \alpha_i)$, and let $\Lambda_1, \ldots, \Lambda_l$ be the basis of $\mathfrak{h}_{\mathbf{R}}^*$ dual to $\alpha_1^*, \ldots, \alpha_l^*$ ($(\Lambda_i, \alpha_j^*) = \delta_{ij}$). Then the †free Abelian group $\sum \mathbf{Z}\Lambda_i$ coincides with the module P of all integral forms. An element $\sum m_i \Lambda_i$ ($m_i \in \mathbf{Z}$) is dominant if and only if $m_1 \geqslant 0, \ldots, m_l \geqslant 0$. Denote by P^+ the semigroup in P consisting of all dominant elements in P. For $j = 1, \ldots, l$, let (ρ_j, V_j) be the irreducible representation of $\mathfrak{g}$ which has Λ_j as its highest weight. The system $\{\rho_1, \ldots, \rho_l\}$ is called the **fundamental system of irreducible representations** associated with Π. The irreducible representation that has $\Lambda = \sum m_i \Lambda_i \in P^+$ as its highest weight is constructed as follows: Put $V_j^m = V_j \otimes \ldots \otimes V_j$ (mth tensor power of V_j) and $\tilde{V} = V_1^{m_1} \otimes \ldots \otimes V_l^{m_l}$. Then $\tilde{V}$ can be regarded as a representation space of $\mathfrak{g}$ in a natural manner. Let V be the smallest $\mathfrak{g}$-stable subspace of $\tilde{V}$ containing $\tilde{V}_\Lambda = (V_1)_{\Lambda_1}^{m_1} \otimes \ldots \otimes (V_l)_{\Lambda_l}^{m_l}$. Then V gives an irreducible representation of $\mathfrak{g}$ with highest weight Λ. Thus by decomposing $V_1^{m_1} \otimes \ldots \otimes V_l^{m_l}$, we get all irreducible representations of $\mathfrak{g}$. This is why $\{\rho_1, \ldots, \rho_l\}$ is called the fundamental system of irreducible representations.

X. Relation with Representations of Compact Lie Groups

Let G be a compact, connected, semisimple Lie group. Then every Cartan subalgebra $\mathfrak{h}$ of the Lie algebra $\mathfrak{g}$ of G is Abelian. We call $\dim \mathfrak{h}$ the **rank** of $\mathfrak{g}$ or of G. Let H be the connected Lie subgroup of G associated with $\mathfrak{h}$. Then H is a **maximal torus (toroidal subgroup)** of G. Furthermore, every maximal torus of G is conjugate to H in G. Also, every element of G is conjugate to an element of H in G. Let $N = N(H)$ be the †normalizer of H in G. Then every $\sigma \in N$ induces an automorphism $\mathrm{Ad}(\sigma)$ of $\mathfrak{g}$, which induces an automorphism (denoted by the same symbol $\mathrm{Ad}(\sigma)$) of the complexification $\mathfrak{g}^{\mathbf{C}}$ of $\mathfrak{g}$. We have $\mathrm{Ad}(\sigma)(\mathfrak{h}^{\mathbf{C}}) = \mathfrak{h}^{\mathbf{C}}$, and $\mathrm{Ad}(\sigma)$ preserves the root system of $\mathfrak{h}^{\mathbf{C}}$. Furthermore, the restriction of $\mathrm{Ad}(\sigma)$ on $(\mathfrak{h}^{\mathbf{C}})_{\mathbf{R}}^*$ is an element w_σ of the Weyl group W of $\mathfrak{g}^{\mathbf{C}}$ relative to $\mathfrak{h}^{\mathbf{C}}$. Via this mapping $\sigma \rightarrow w_\sigma$, we have $N/H \cong W$. Thus we identify N/H with W.

Since $(\mathfrak{h}^{\mathbf{C}})_{\mathbf{R}} = \sqrt{-1}\,\mathfrak{h}$, we can define a lexicographic linear ordering using a basis of $\mathfrak{h}$. Fix such an ordering. Now every representation ρ of G over a complex vector space V induces a representation $d\rho$ of $\mathfrak{g}$ over V, where $d\rho$ is the differential of ρ (→ 249 Lie Groups). Then we have a representation $d\rho$ of $\mathfrak{g}^{\mathbf{C}}$ over V. We also call a weight λ relative to $\mathfrak{h}^{\mathbf{C}}$ of the representation $d\rho$ of $\mathfrak{g}^{\mathbf{C}}$ a weight of the representation ρ of G relative to H. Representations ρ_1, ρ_2 of G are equivalent if and only if representations $d\rho_1, d\rho_2$ of $\mathfrak{g}^{\mathbf{C}}$ are equivalent. Denote by Λ_1, Λ_2 the highest weights of $d\rho_1, d\rho_2$, respectively. Then ρ_1 is equivalent to ρ_2 if and only if $\Lambda_1 = \Lambda_2$.

The module P of all integral forms in $(\mathfrak{h}^{\mathbf{C}})_{\mathbf{R}}^*$ coincides with the set of all elements in $(\mathfrak{h}^{\mathbf{C}})_{\mathbf{R}}^*$ that are weights (relative to $\mathfrak{h}^{\mathbf{C}}$) of some representation of $\mathfrak{g}^{\mathbf{C}}$. Denote by P_G the subset of P consisting of all elements in P that are weights (relative to H) of some representation of G. Then P_G is a submodule of P such that $[P : P_G] < \infty$. Put $P_G^+ = P_G \cap P^+$. Then the mapping (representation ρ) → (the highest weight Λ of ρ) induces a bijective mapping from the set of all classes of irreducible representations of G onto P_G^+.

The exponential mapping $\exp: \mathfrak{h} \rightarrow H$ is a surjective homomorphism from $\mathfrak{h}$ to H. The kernel Γ_G of this homomorphism is a †lattice group in $\mathfrak{h}$ of rank l ($= \dim \mathfrak{h}$); a basis $H_1, \ldots, H_l$ of Γ_G over $\mathbf{Z}$ is also a basis of $\mathfrak{h}$ over $\mathbf{R}$. Now define linear forms $\lambda_1, \ldots, \lambda_l$ on $\mathfrak{h}$ by $(\lambda_i, H_j) = \delta_{ij}$. Then we have

$$P_G = 2\pi\sqrt{-1}\sum \mathbf{Z}\lambda_j.$$

In other words, an element $\lambda \in (\mathfrak{h}^{\mathbf{C}})_{\mathbf{R}}^*$ is in P_G if and only if $\lambda(H) \in 2\pi\sqrt{-1}\,\mathbf{Z}$ for every element H in Γ_G. This characterization of P_G also characterizes $P_G^+ = P_G \cap P^+$. In particular, G is simply connected $\Leftrightarrow P = P_G \Leftrightarrow \Gamma_G = 2\pi\sqrt{-1}\cdot \sum \mathbf{Z}\alpha_i^*$, where $\alpha_1, \ldots, \alpha_l$ are simple roots and $\alpha_i^* = 2\alpha_i/(\alpha_i, \alpha_i)$ for $i = 1, \ldots, l$. We also have $G = I(\mathfrak{g})$ = the group of inner automorphisms of $\mathfrak{g} \Leftrightarrow P_G = \sum \mathbf{Z}\alpha_i \Leftrightarrow \Gamma_G = 2\pi\sqrt{-1}\sum \mathbf{Z}\varepsilon_i$, where $\varepsilon_1, \ldots, \varepsilon_l$ are elements of $(\mathfrak{g}^{\mathbf{C}})_{\mathbf{R}}^*$ defined by $(\alpha_i, \varepsilon_j) = \delta_{ij}$. In general, we have $P \supset P_G \supset \sum \mathbf{Z}\alpha_i$, $2\pi\sqrt{-1}\sum \mathbf{Z}\alpha_i^* \subset \Gamma_G \subset 2\pi\sqrt{-1}\,\mathbf{Z}\varepsilon_i$. Furthermore, the †fundamental group $\pi_1(G)$ of G is given by $P/P_G \cong \Gamma_G/(2\pi\sqrt{-1}\sum \mathbf{Z}\alpha_i^*)$. Also,

the kernel of the adjoint representation $G \to I(G) = I(\mathfrak{g})$ is given by $P_G / \sum \mathbf{Z}\alpha_i \cong (2\pi\sqrt{-1} \cdot \sum \mathbf{Z}\varepsilon_i)/\Gamma_G$.

Y. Invariant Measure on G

Keeping the definitions in this section the same as in the previous section, we see that every root α defines a representation $h \to \chi_\alpha(h)$ of the group H of dimension 1 over $(\mathfrak{g}^{\mathbf{C}})_\alpha$ as follows: Let E_α be a basis of $(\mathfrak{g}^{\mathbf{C}})_\alpha$. Then $\mathrm{Ad}(h)E_\alpha = \chi_\alpha(h)E_\alpha$. We have $\chi_\alpha(h) = e^{\alpha(X)}$ if $h = \exp X$ $(X \in \mathfrak{h})$, i.e., $\chi_\alpha \circ \exp = e^\alpha$. We let e^α stand for χ_α: $e^\alpha(h) = \chi_\alpha(h)$, $h \in H$. Now let dg, dh, dm be the †invariant measures on G, H, $M = G/H$, respectively, normalized by

$$\int_G dg = 1, \quad \int_H dh = 1, \quad \int_M dm = 1.$$

Then for every continuous function f on G, we have

$$\int_G f(g)\,dg = \frac{1}{w}\int_H \left(\int_M f(m,h)\,dm\right)\Omega(h)\,dh, \qquad (1)$$

where w is the order of the Weyl group W and $f(m,h)$ is a function on $M \times H$ defined by $f(m,h) = f(ghg^{-1})$, $m = gH$. Note that $f(m,h)$ is well defined. Finally, $\Omega(h)$ is a function on H defined by

$$\Omega(h) = \prod_{\alpha \in \Delta} (e^{\alpha(X)/2} - e^{-\alpha(X)/2})$$

for $h = \exp X$ $(X \in \mathfrak{h})$, and $\Omega(h)$ is called the **density** on H. Denote by $D(h)$ the same product as $\Omega(h)$, letting α range over Δ^+. Then we have $\Omega(h) = D(h)\overline{D(h)} = |D(h)|^2 \geqslant 0$. In particular, if f is a †class function on G (i.e., $f(xyx^{-1}) = f(y)$ for every $x, y \in G$), $f(m,h) = f(h)$. Hence (1) is simplified into the following integral formula for class functions:

$$\int_G f(g)\,dg = \frac{1}{w}\int_H f(h)\Omega(h)\,dh. \qquad (2)$$

Z. Weyl's Character Formula

Keeping the definitions in this section the same as those in the previous section, we let (ρ, V) be an irreducible representation of G and χ_ρ be the character of $\rho (\chi_\rho(g) = \mathrm{tr}\,\rho(g))$. Let Λ be the highest weight of ρ. Then since Λ determines ρ up to equivalence, χ_ρ must be determined by Λ. In fact, χ_ρ is given via Λ by **Weyl's character formula** (3). (Note that $G = \bigcup gHg^{-1}$. Hence χ_ρ is determined by its restriction on H.) Now for $h \in H$, we have

$$\chi_\rho(h) = \frac{\xi_{\Lambda+\delta}(h)}{\xi_\delta(h)}, \qquad (3)$$

where ξ_λ $(\lambda \in P)$ is the alternating sum

$$\xi_\lambda(h) = \sum_{w \in W} \det(w) e^{(w(\lambda))(X)}, \quad h = \exp X. \qquad (4)$$

Finally, δ in (3) is given by

$$\delta = \frac{1}{2}\sum_{\alpha \in \Delta^+} \alpha,$$

that is, δ is the half-sum of the positive roots. In particular, we have $\xi_\delta(h) = D(h)$. Denote by $m_\Lambda(\lambda)$ the multiplicity $\dim V_\lambda$ of a weight λ of ρ. Then we have $\chi_\rho(h) = \sum_\lambda m_\Lambda(\lambda)e^\lambda(h)$. Furthermore, we have $m_\Lambda(\Lambda) = 1$, $m_\Lambda(\lambda) = m_\Lambda(w(\lambda))$ $(w \in W)$, and $m_\Lambda(\lambda)$ is given by **Kostant's formula**:

$$m_\Lambda(\lambda) = \sum_{w \in W} \det(w) P(w(\Lambda + \delta) - (\lambda + \delta)), \qquad (5)$$

where for each $\mu \in P$, $P(\mu)$ is the number of ways μ can be expressed as a sum of positive roots. Thus $P(\mu)$ is the number of nonnegative integral solutions $\{k_\alpha\}$ of $\mu = \sum_{\alpha \in \Delta^+} k_\alpha \alpha$.

Now suppose that (ρ_1, V_1), (ρ_2, V_2) are irreducible representations of G. Their tensor product $(\rho_1 \otimes \rho_2, V_1 \otimes V_2)$ is decomposed into a direct sum of irreducible constituents: $\rho_1 \otimes \rho_2 = \sum m(\mu)\rho_\mu$, where ρ_μ is the irreducible representation of G which has μ as its highest weight and $m(\mu)$ is the multiplicity of ρ_μ in $\rho_1 \otimes \rho_2$. Then $m(\mu)$ is given by **Steinberg's formula**:

$$m(\mu) = \sum_{w \in W}\sum_{w' \in W} \det(ww') P(w(\Lambda_1 + \delta) + w'(\Lambda_2 + \delta) - (\mu + 2\delta)), \qquad (6)$$

where $\rho_1 = \rho_{\Lambda_1}$, $\rho_2 = \rho_{\Lambda_2}$. The partition function $P(\mu)$ that appears in (5) and (6) satisfies the following recursive formula (B. Kostant):

$$P(\mu) = -\sum_{w \in W, w \neq 1} \det(w) P(\mu - (\delta - w(\delta))). \qquad (7)$$

References

[1] Sém. "Sophus Lie," Ecole Norm. Sup., 1954–1955.
[2] N. Bourbaki, Eléments de mathématique, Groupe et algèbres de Lie., Actualités Sci. Ind., Hermann, ch. 1–3, 1960; ch. 4–6, 1968.
[3] N. Jacobson, Lie algebras, Interscience, 1962.
[4] S. Helgason, Differential geometry, Lie groups, and symmetric spaces, Academic Press, 1978.
[5] E. Cartan, Oeuvres complètes pt. I, Gauthier-Villars, 1952.
[6] C. Chevalley, Theory of Lie groups I, Princeton Univ. Press, 1946; Théorie des groupes de Lie, Actualités Sci. Ind., Hermann, II, 1951; III, 1954.

[7] I. Kaplansky, Lie algebras, Lectures on modern mathematics 1, T. L. Saaty (ed.), Wiley, 1963, 115–132.
[8] H. Weyl, Theorie der Darstellung kontinuierlicher halb-einfacher Gruppen durch linearen Transformationen I–III, Math. Z., 23 (1925), 271–309; 24 (1926), 328–376, 377–395 (Gesammelte Abhandlungen II, Springer, 1968, 543–647).
[9] H. Weyl, The classical groups, Princeton Univ. Press, revised edition, 1946.
[10] I. D. Ado, Über die Darstellung von Lieschen Gruppen durch linearen Substitutionen, Bull. Soc. Physico-Math. Kazan, (3) 7 (1934–1935), 3–43.
[11] F. R. Gantmacher, On the classification of real simple Lie groups, Mat. Sb., 5 (47) (1939), 217–250.
[12] E. B. Dynkin, The structure of semisimple Lie algebras, Amer. Math. Soc. Transl., 17 (1950). (Original in Russian, 1947.)
[13] R. Bott, Homogeneous vector bundles, Ann. Math., (2) 66 (1957), 203–248.
[14] I. Satake, On representations and compactifications of symmetric Riemannian spaces, Ann. Math., (2) 71 (1960), 77–110.
[15] H. Freudenthal, Oktaven, Ausnahmegruppen und Oktavengeometrie, Mathematisch Instituut der Rijksuniversiteitte, Utrecht, mimeographed note, 1951.
[16] S. Araki, On root systems and an infinitesimal classification of irreducible symmetric spaces, J. Math. Osaka City Univ., 13 (1962), 1–34.
[17] J.-P. Serre, Lie algebras and Lie groups, Benjamin, 1965.
[18] J.-P. Serre, Algèbres de Lie semi-simples complexes, Benjamin, 1966.
[19] J. E. Humphreys, Introduction to Lie algebras and representation theory, Springer, 1972.
Also → references in [3, 4].

249 (IV.9) Lie Groups

A. Definitions

A set G is called a **Lie group** if there is given on G a structure satisfying the following three axioms: (i) G is a group; (ii) G is a †paracompact, †real analytic manifold (G need not be connected); and (iii) the mapping $G \times G \to G$ defined by $(x, y) \to xy^{-1}$ is †real analytic (→ 105 Differentiable Manifolds).

For simplicity, real analyticity is denoted by C^ω: for example, C^ω-functions, C^ω-mappings. If we replace real analyticity by †complex analyticity in axioms (ii) and (iii), then we have the axioms (i), (ii′), and (iii′) of a **complex Lie group**. We consider the real analytic case, since the complex analytic case can be dealt with similarly.

Every element σ of G defines a mapping $G \to G$ given by $x \to \sigma x$ ($x \to x\sigma$), denoted by L_σ (R_σ) and called the **left (right) translation** of G by σ. L_σ, R_σ are automorphisms of G as a C^ω-manifold. Therefore, given a †vector field X on G, a †differential form ω on G, or, in general, a †tensor field T on G, we can apply L_σ and R_σ in a natural manner to the given tensor field and obtain a tensor field $L_\sigma T$, $R_\sigma T$ on G. A tensor field T on G is called **left (right) invariant** if $L_\sigma T = T$ ($R_\sigma T = T$) for every $\sigma \in G$. A tensor field on G is of class C^ω if it is left or right invariant.

B. Lie Algebras of Lie Groups

Let G be a Lie group. Then the set $\mathfrak{X}(G)$ of all C^ω-vector fields on G has the structure of a vector space over the real number field $\mathbf{R}$. Furthermore, with respect to the bracket operation $[X, Y] = XY - YX$ ($X, Y \in \mathfrak{X}(G)$), $\mathfrak{X}(G)$ forms a †Lie algebra over $\mathbf{R}$. Denote by $\mathfrak{g}$ the subset of $\mathfrak{X}(G)$ consisting of all left invariant vector fields on G. Then $\mathfrak{g}$ is a †subalgebra of the Lie algebra $\mathfrak{X}(G)$. Thus $\mathfrak{g}$ is also a Lie algebra over $\mathbf{R}$. This Lie algebra $\mathfrak{g}$ is called the **Lie algebra of the Lie group** G. The linear mapping from $\mathfrak{g}$ into the †tangent space $T_e(G)$ of G at the identity element e given by $X \to X_e$ is bijective. Hence $\dim \mathfrak{g} = \dim G$. The Lie algebra $\mathfrak{g}$ is often identified with $T_e(G)$ via this bijection.

C. Simply Connected Covering Lie Groups

For any finite-dimensional Lie algebra $\mathfrak{g}$ over $\mathbf{R}$, there exists a †connected Lie group G that has $\mathfrak{g}$ as its Lie algebra. Such Lie groups are all †locally isomorphic. Among these groups, there exists a †simply connected one that is unique up to isomorphism. This group is called the **simply connected covering Lie group** of the Lie algebra $\mathfrak{g}$.

D. Lie Subgroups

A subgroup H of a Lie group G is called a **Lie subgroup** if (i) H has the structure of a Lie group and (ii) the inclusion mapping $\varphi: H \to G$ is a C^ω-mapping, and the †differential $d\varphi$ is injective at every point of H (i.e., as a manifold, H is a †submanifold of G). Moreover, if H is a connected manifold, it is called a **connected Lie subgroup**. For a subgroup H of a Lie group

G, there exists at most one structure of a Lie group on H that makes H a connected Lie subgroup. (If H is not assumed to be connected, then this uniqueness statement is not generally true.) A subgroup H of a Lie group G is a connected Lie subgroup if and only if H is arc-wise connected (M. Kuranishi and H. Yamabe).

Let H be a Lie subgroup of a Lie group G. Then the tangent space $T_e(H)$ is a subspace of $T_e(G)$. Under the isomorphism $T_e(G) \cong \mathfrak{g}$ mentioned in Section B, the subspace $T_e(H)$ corresponds to $\mathfrak{h}$, which is a †subalgebra of the Lie algebra $\mathfrak{g}$. We call $\mathfrak{h}$ the **subalgebra of $\mathfrak{g}$ associated with the Lie subgroup** H, and $\mathfrak{h}$ can be identified with the Lie algebra of H in a natural manner. By this mapping $H \to \mathfrak{h}$, we have a bijection from the set of all connected Lie subgroups of G onto the set of all subalgebras of $\mathfrak{g}$. For example, the connected Lie subgroup G' of G corresponding to the †derived algebra $\mathfrak{g}'$ of $\mathfrak{g}$ is the †commutator subgroup of G. The Lie algebra of a normal subgroup of G is an †ideal of $\mathfrak{g}$. Conversely, if the Lie algebra $\mathfrak{h}$ of a connected Lie subgroup H of a connected Lie group G is an ideal of $\mathfrak{g}$, then H is a normal subgroup of G.

A connected Lie group G is called **semisimple**, **simple**, **solvable**, **nilpotent**, or **Abelian** (or **commutative**), respectively, whenever the Lie algebra $\mathfrak{g}$ is (→ 248 Lie Algebras). This definition of solvability, nilpotency, and commutativity agrees with the corresponding definition in which G is regarded as an abstract group.

E. Closed Subgroups

The topology of a Lie subgroup H (which we call the **inner topology** of H) as a submanifold of a Lie group G need not coincide with the relative topology of H regarded as a subspace of a topological space G. The inner topology of H coincides with the relative topology of H if and only if H is closed in G. Conversely, for every closed subgroup H of G, H has the structure of a Lie subgroup such that the inner topology and relative topology coincide (E. Cartan). Moreover, such a structure of a Lie subgroup on H is unique. We always regard closed subgroups of a Lie group as Lie subgroups in this sense. Also, we denote the Lie algebras of Lie groups $G, H, \ldots$ by the corresponding lower-case German letters $\mathfrak{g}, \mathfrak{h}, \ldots$.

F. Homogeneous Spaces

Let G be a Lie group, and let H be a closed subgroup of G. Then the †quotient topological space (coset space) $M = G/H$ of the topological group G modulo H has the structure of a C^ω-manifold such that the canonical mapping $\pi: G \to M$ and the action of G on $M: G \times M \to M$ (defined by $(g, xH) \to gxH$) are both C^ω-mappings. Furthermore, such a C^ω-manifold structure on M is unique. The C^ω-manifold M thus obtained is called the **homogeneous space** of G over H (→ 199 Homogeneous Spaces). Put $\pi(e) = p$. Then $\mathfrak{h}$ is the kernel of the differential mapping $T_e(G) = \mathfrak{g} \to T_p(M)$. Thus we can identify the tangent space $T_p(M)$ of $M = G/H$ at p with the quotient linear space $\mathfrak{g}/\mathfrak{h}$ via this differential mapping.

G. Quotient Lie Groups

Suppose that H is a closed normal subgroup of a Lie group G. Then G/H has the structure of a quotient group together with the structure of a manifold as a homogeneous space. Now G/H is a Lie group with respect to these two structures. The Lie group G/H is called the **quotient Lie group** of G over H. In this case, $\mathfrak{h}$ is an †ideal of $\mathfrak{g}$, and the quotient Lie algebra $\mathfrak{g}/\mathfrak{h}$ is isomorphic to the Lie algebra of the Lie group G/H.

H. Direct Product of Lie Groups

Let G_1, G_2 be Lie groups. Then the direct product $G_1 \times G_2$ satisfies axioms (i), (ii), and (iii) of a Lie group (→ Section A) in a natural manner. The Lie group $G_1 \times G_2$ thus obtained is called the **direct product** of the Lie groups G_1, G_2. The Lie algebra of $G_1 \times G_2$ can be identified with the direct sum of $\mathfrak{g}_1$, $\mathfrak{g}_2$.

I. Cartan Subgroups

A subgroup H of a group G is called a **Cartan subgroup** of G if H is a maximal nilpotent subgroup of G and moreover, for every subgroup H_1 of H of finite index in H, $[N(H_1): H_1] < \infty$, where $N(H_1) = \{g \in G \mid gH_1g^{-1} = H_1\}$ is the †normalizer of H_1 in G. A connected Lie group G always has a Cartan subgroup. Furthermore, every Cartan subgroup H of G is closed; hence H is a Lie subgroup. The Lie algebra $\mathfrak{h}$ of H is a †Cartan subalgebra of the Lie algebra $\mathfrak{g}$. This mapping $H \to \mathfrak{h}$ is a bijection from the set of all Cartan subgroups of G onto the set of all Cartan subalgebras of $\mathfrak{g}$ [7].

J. Borel Subgroups

Let G be a complex connected semisimple Lie group. A maximal connected solvable Lie subgroup of G is called a **Borel subgroup** of G.

Any two Borel subgroups are conjugate in G. A subgroup P of G is called a **parabolic subgroup** if P contains a Borel subgroup of G. Every parabolic subgroup is a connected closed Lie subgroup of G.

K. Simple Examples

Let V be a finite-dimensional vector space over $\mathbf{R}$. Let $\mathfrak{L}(V)$ be the †associative algebra of linear endomorphisms of V. The †general linear group $GL(V)=\{x\in\mathfrak{L}(V)\mid\det x\neq 0\}$ over V is a Lie group. The Lie algebra of $GL(V)$ is denoted by $\mathfrak{gl}(V)$. We can identify $\mathfrak{gl}(V)$ with $\mathfrak{L}(V)$, and we have $[X,Y]=XY-YX$ for all $X, Y\in\mathfrak{gl}(V)$. If $\dim V=n$, $GL(V)$ may be identified with the group $GL(n,\mathbf{R})$ of all real nonsingular $n\times n$ matrices. Then $\mathfrak{gl}(V)$ may be identified with the Lie algebra $\mathfrak{gl}(n,\mathbf{R})$ of all $n\times n$ real matrices.

The following are examples of (closed) Lie subgroups of the general linear group: the †special linear group $SL(n,\mathbf{R})$, whose Lie algebra is $\{X\in\mathfrak{gl}(n,\mathbf{R})\mid\operatorname{tr}X=0\}$; the †unitary group $U(n)$, whose Lie algebra is $\{X\in\mathfrak{gl}(n,\mathbf{C})\mid {}^tX+\bar{X}=0\}$; the †orthogonal group $O(n)$, whose Lie algebra is $\{X\in\mathfrak{gl}(n,\mathbf{R})\mid {}^tX+X=0\}$; and the †symplectic group $Sp(n)$, whose Lie algebra is $\{X\in\mathfrak{gl}(2n,\mathbf{C})\mid JX+{}^tXJ=0, {}^tX+\bar{X}=0\}$, where

$$J=\begin{pmatrix}0 & I\\ -I & 0\end{pmatrix}$$

and I is the $n\times n$ unit matrix.

Examples of complex Lie groups are: the complex †general linear group $GL(n,\mathbf{C})$, whose Lie algebra is $\mathfrak{gl}(n,\mathbf{C})$; the complex †orthogonal group $O(n,\mathbf{C})$, whose Lie algebra is $\{X\in\mathfrak{gl}(n,\mathbf{C})\mid {}^tX+X=0\}$; and the complex †sympletic group $Sp(n,\mathbf{C})$, whose Lie algebra is $\{X\in\mathfrak{gl}(2n,\mathbf{C})\mid JX+{}^tXJ=0\}$, where J is the matrix given above.

L. Compact Simple Lie Groups

If a connected Lie group G has a †compact real Lie algebra, then G is compact. Thus for each compact real simple Lie algebra $\mathfrak{g}$, the simply connected covering Lie group G is compact, and the center of G is a finite group. The connected compact Lie groups $SU(l+1)$, $SO(2l+1)$, $Sp(l)$, $SO(2l)$ have, respectively, the compact real simple Lie algebras A_l $(l\geqslant 1)$, B_l $(l\geqslant 2)$, C_l $(l\geqslant 3)$, D_l $(l\geqslant 4)$ as their Lie algebras; these groups are called **classical compact simple Lie groups**. $Sp(n)$ $(n\geqslant 2)$ and $SU(n)$ $(n\geqslant 2)$ are simply connected, but $SO(n)$ $(n=3$ or $n\geqslant 5)$ is not, and has the fundamental group of order 2. The universal covering group of $SO(n)$ (i.e., the simply connected, connected Lie group that has the same Lie algebra as $SO(n)$) is $Spin(n)$. The connected compact Lie groups that have E_l $(l=6,7,8)$, F_4, or G_2 as their Lie algebras are called **exceptional compact simple Lie groups**. The simply connected covering Lie groups of E_8, F_4, G_2 have the identity group $\{e\}$ as their center. Hence they coincide with their †adjoint groups. The simply connected covering Lie groups E_6, E_7 have as their centers the groups of orders 3 and 2, respectively. Hence their adjoint groups are not simply connected.

M. Complex Simple Lie Groups

For a complex Lie group G, the definition of the Lie algebra $\mathfrak{g}$ is similar to that for real Lie groups. Then $\mathfrak{g}$ is a Lie algebra over the complex number field $\mathbf{C}$ and is called the **complex Lie algebra of the complex Lie group** G. The complex connected Lie groups $SL(l+1,\mathbf{C})$, $SO(2l+1,\mathbf{C})$, $Sp(l,\mathbf{C})$, $SO(2l,\mathbf{C})$, which have the †classical complex simple Lie algebras A_l, B_l, C_l, D_l as their Lie algebras, respectively, are called **classical complex simple Lie groups**. Complex connected Lie groups that have †exceptional complex simple Lie algebras E_l $(l=6,7,8)$, F_4, G_2 as their Lie algebras are called **exceptional complex simple Lie groups**.

N. Homomorphisms

A mapping $\varphi: G_1\to G_2$ from a Lie group G_1 into a Lie group G_2 is called an **analytic homomorphism** (**C^ω-homomorphism**) if (i) φ is a homomorphism between groups and (ii) φ is a C^ω-mapping between manifolds. Moreover, if φ is bijective and φ^{-1} is also a C^ω-mapping, then φ is called an **analytic isomorphism** (**C^ω-isomorphism**). Two Lie groups G_1, G_2 are said to be **isomorphic** as Lie groups if there exists a C^ω-isomorphism from G_1 onto G_2; we denote this situation by $G_1\cong G_2$. Now let $\varphi: G_1\to G_2$ be a C^ω-homomorphism. Then the †differential $(d\varphi)_{e_1}: T_{e_1}(G_1)\to T_{e_2}(G_2)$ induces a Lie algebra homomorphism $d\varphi:\mathfrak{g}_1\to\mathfrak{g}_2$. The kernel of $d\varphi$ is the Lie algebra of the kernel of φ. Suppose that G_1 is connected and simply connected. Then for every Lie group G_2 and every Lie algebra homomorphism $\psi:\mathfrak{g}_1\to\mathfrak{g}_2$, there exists a unique C^ω-homomorphism $\varphi: G_1\to G_2$ such that $\psi=d\varphi$. We can replace condition (ii) in the definition of C^ω-homomorphism by the following weaker one: (ii′) φ is a continuous mapping. In particular, for a given topological group G, there exists at most one structure of a Lie group on G preserving the group structure and the topology. A topological group G admits a structure of a Lie group (preserving the group structure and the topology) if and

only if G is a †locally Euclidean topological group (→ 423 Topological Groups N).

O. Representations

Let G be a Lie group, and let V be a finite-dimensional vector space over **C**(**R**). Then a continuous (hence C^ω-) homomorphism $\rho: G \to GL(V)$ is called a **complex (real) representation** of G. The linear space V is called the **representation space** of ρ, and dim V is called the **degree** of ρ. To be more precise, a representation is denoted by (ρ, V) instead of ρ. The function $\chi_\rho: G \to \mathbf{C}$ defined by $\chi_\rho(g) = \operatorname{tr}\rho(g)$ $(g \in G)$ is called the **character** of ρ (→ 362 Representations). The representation (ρ, V) of G gives rise to a representation $(d\rho, V)$ of $\mathfrak{g}$. Suppose G is connected. Then two representations (ρ_i, V_i) $(i=1,2)$ of G are †equivalent if and only if the representations $d\rho_1$ and $d\rho_2$ of $\mathfrak{g}$ are equivalent. For the †direct sum of representations and the contragredient representation, we have $d(\rho_1+\rho_2)=d\rho_1+d\rho_2$, $d({}^t\rho^{-1}) = -{}^t(d\rho)$. For the †tensor product of representations, we have $(d(\rho_1 \otimes \rho_2))(X) = (d\rho_1)(X) \otimes 1_{V_2} + 1_{V_1} \otimes (d\rho_2)(X)$. (For representations of compact Lie groups → 248 Lie Algebras W.)

P. Adjoint Representations

An element σ of a Lie group G defines an analytic automorphism $\varphi_\sigma: x \to \sigma x \sigma^{-1}$ of G. The †differential $\mathrm{Ad}(\sigma)$ of φ_σ is an automorphism of the Lie algebra $\mathfrak{g}$. Since $\mathrm{Ad}(\sigma) \in GL(\mathfrak{g})$, we have a representation $\sigma \to \mathrm{Ad}(\sigma)$ of G whose representation space is $\mathfrak{g}$. This is called the **adjoint representation** of G, and the image $\mathrm{Ad}(G)$ of G is called the **adjoint group** of G. $\mathrm{Ad}(G)$ is a Lie subgroup of $GL(\mathfrak{g})$. The †center Z of G is a closed subgroup of G, and $G/Z \cong \mathrm{Ad}(G)$. The analytic homomorphism $\sigma \to \mathrm{Ad}(\sigma)$ from G into $GL(\mathfrak{g})$ gives rise to a Lie algebra homomorphism $\mathfrak{g} \to \mathfrak{gl}(\mathfrak{g})$ by taking the differential of $\sigma \to \mathrm{Ad}(\sigma)$. Denote this Lie algebra homomorphism by $X \to \mathrm{ad}(X)$ $(X \in \mathfrak{g})$. Then $\mathrm{ad}(X)Y = [X, Y]$ for every X, Y in $\mathfrak{g}$. Thus $\mathrm{ad}: \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g})$ coincides with the †adjoint representation of the Lie algebra $\mathfrak{g}$. If G is connected, then the Lie subalgebra $\mathrm{ad}(\mathfrak{g})$ (the †adjoint Lie algebra of $\mathfrak{g}$) of $\mathfrak{gl}(\mathfrak{g})$ is associated with the connected Lie subgroup $\mathrm{Ad}(G)$ of $GL(\mathfrak{g})$. Thus the adjoint group $\mathrm{Ad}(G)$ coincides with the group $I(\mathfrak{g})$ of †inner automorphisms of $\mathfrak{g}$ (the †adjoint group of $\mathfrak{g}$).

If G is connected and semisimple, $\mathfrak{g}$ is also semisimple. Furthermore, we have $\mathfrak{D}(\mathfrak{g}) = \mathrm{ad}(\mathfrak{g})$. Hence the adjoint group $\mathrm{Ad}(G) = I(\mathfrak{g})$ of G coincides with the connected component of the †automorphism group $A(\mathfrak{g})$ of $\mathfrak{g}$ containing the identity element. If G is compact or semisimple, then every representation of G is completely reducible.

Let G be a connected Lie group of dimension n, $\mathfrak{g}$ be the †Lie algebra of G, and $\mathrm{Ad}: G \to GL(\mathfrak{g})$ be the adjoint representation of G. Put $\det((t+1)I - \mathrm{Ad}(x)) = \sum_{j=0}^n t^j D_j(x)$ for every $x \in G$. Then each D_j is an analytic function on G, and $D_n = 1$. Let l be the least integer such that $D_l \neq 0$. Then $l = \operatorname{rank} G = \operatorname{rank} \mathfrak{g}$. An element x of G is called **regular (singular)** if $D_l(x) \neq 0$ $(D_l(x) = 0)$. The subset G^* of G consisting of all regular elements of G is open and dense in G. The subset $G - G^*$ of G consisting of all singular elements of G is of measure zero with respect to the left-invariant †Haar measure of G.

Now suppose that $\mathfrak{g}$ is †reductive. Then an element $x \in G$ is regular if and only if the centralizer $\mathfrak{z}_x = \{X \in \mathfrak{g} \mid \mathrm{Ad}(x)X = 0\}$ is a †Cartan subalgebra of $\mathfrak{g}$. Furthermore, if $x \in G$ is regular, then $\mathrm{Ad}(x)$ is a semisimple linear endomorphism of $\mathfrak{g}$.

Q. Canonical Coordinates

With each element X of the Lie algebra $\mathfrak{g}$ of the Lie group G there is associated a **one-parameter subgroup** of G, i.e., a continuous homomorphism $t \to \varphi(t)$ from the additive group **R** of real numbers into G, such that $d\varphi(X_0) = X$, where $X_0 = d/dt$ is the basis of the Lie algebra of **R**. Furthermore, the continuous homomorphism φ is unique. Putting $\varphi(1) = \exp X$, we define the **exponential mapping** $\exp: \mathfrak{g} \to G$. Then we have $\varphi(t) = \exp tX$ for every $t \in \mathbf{R}$. In particular, for the case $G = GL(n, \mathbf{C})$, $\mathfrak{g} = \mathfrak{gl}(n, \mathbf{C})$, we have $\exp X = \sum X^m/m!$ for every $X \in \mathfrak{g}$. Thus $\exp X$ coincides with the usual exponential mapping of matrices (→ 269 Matrices). The mapping $\exp: \mathfrak{g} \to G$ is a C^ω-mapping whose differential at $X = 0$ is a bijection from $T_0(\mathfrak{g}) = \mathfrak{g}$ onto $T_e(G)$. Thus for each basis $X_1, \ldots X_n$ of $\mathfrak{g}$, there exists a positive real number ε with the following property: $\{\exp(\sum x_i X_i) \mid |x_i| < \varepsilon \ (i=1,\ldots,n)\}$ is an open neighborhood of the identity element e in G on which $\sigma = \exp(\sum x_i X_i) \to (x_1, \ldots, x_n)$ $(|x_i| < \varepsilon$, $i = 1, \ldots, n)$ is a †local coordinate system. These local coordinates are called the **canonical coordinates of the first kind** associated with the basis (X_i) of $\mathfrak{g}$. Similarly, we have a local coordinate system $\tau = (\exp x_1 X_1) \ldots (\exp x_n X_n) \to (x_1, \ldots, x_n)$ in a neighborhood of e; these $x_1, \ldots, x_n$ are called the **canonical coordinates of the second kind** associated with the basis (X_i) of $\mathfrak{g}$.

Let $\varphi: G_1 \to G_2$ be a continuous homomorphism from a Lie group G_1 into a Lie group G_2. Then $\varphi(\exp X) = \exp(d\varphi(X))$ for every $X \in \mathfrak{g}_1$. The Lie subalgebra $\mathfrak{h}$ associated with a

connected Lie subgroup H of G is characterized by using the exponential mapping as follows: $\mathfrak{h}=\{X\in\mathfrak{g}\mid \exp tX\in H \text{ for all } t\in\mathbf{R}\}$.

R. Multiplication Functions

Fix a basis $X_1,\ldots,X_n$ of the Lie algebra $\mathfrak{g}$ of a Lie group G. Let $(u_1,\ldots,u_n)$ be the canonical coordinates of the first kind associated with the basis (X_i). With respect to this coordinate system, let (x_i), (y_i), (z_i) be the coordinates of the elements σ, τ, $\sigma\tau$, respectively. Then each z_i is a real analytic function in $(x_1,\ldots,x_n;$ $y_1,\ldots,y_n)$. The †Taylor expansion of z_i at $x_1=x_2=\ldots=y_1=y_2=\ldots=0$ is given as follows: Put $S=\sum x_iX_i$, $T=\sum y_iX_i$. Then

$$z_i=\sum_{k,l=0}^{\infty}\frac{1}{k!\,l!}(S^kT^lu_i)_0$$

$$=x_i+y_i+(STu_i)_0+(\text{terms of degree }\geqslant 3).$$

Furthermore, let (w_i) be the coordinates of $\tau^{-1}\sigma^{-1}\tau\sigma$. Then we have

$$w_i=([T,S]u_i)_0+(\text{terms of degree }\geqslant 3).$$

Let $\mathfrak{g}^*$ be the set of all $\mathbf{R}$-valued linear forms on $\mathfrak{g}$, that is, $\mathfrak{g}^*$ is the dual vector space of $\mathfrak{g}$. We can identify the elements of $\mathfrak{g}^*$ with the left-invariant †differential forms of degree 1 on G. These forms are called **Maurer-Cartan differential forms**. Let $\omega_1,\ldots,\omega_n$ be a basis of $\mathfrak{g}^*$ which is dual to the basis $X_1,\ldots,X_n$ of $\mathfrak{g}$: $\omega_i(X_j)=\delta_{ij}$ $(1\leqslant i,j\leqslant n)$. The †exterior derivative $d\omega_k$ of ω_k is a left-invariant differential form of degree 2 on G and is expressed as a linear combination of the †exterior products $\omega_i\wedge\omega_j$ as follows:

$$d\omega_k=-\frac{1}{2}\sum_{i,j}c_{ij}^k\omega_i\wedge\omega_j, \tag{1}$$

where (c_{ij}^k) are the †structural constants of $\mathfrak{g}$ with respect to the basis (X_i), i.e., $[X_i,X_j]=\sum_k c_{ij}^kX_k$.

Using the canonical coordinates of the first kind (u_i) associated with the basis (X_i), put $\omega_k=\omega_k(u,du)=\sum_j A_{kj}(u)du_j$. Then the matrix $\mathfrak{A}=(A_{kj}(u))$ is given by

$$\mathfrak{A}=I+\frac{\mathfrak{X}}{2!}+\frac{\mathfrak{X}^2}{3!}+\ldots, \tag{2}$$

where $\mathfrak{X}=(c_{ij}(u))$, $c_{ij}(u)=\sum_k c_{jk}^iu_k$, i.e., $-\mathfrak{X}$ is the matrix of $\operatorname{ad}(\sum u_kX_k)$ relative to the basis (X_i). Thus, in particular, if G is a nilpotent group, $\mathfrak{X}$ is a †nilpotent matrix.

Once the functions $A_{kj}(u)$ are known, then the functions $z_i=f_i(x_1,\ldots,x_n;y_1,\ldots,y_n)$ describing the multiplication in G (note that $\sigma\leftrightarrow(x_i)$, $\tau\leftrightarrow(y_i)$, $\sigma\tau\leftrightarrow(z_i)$) are obtained as follows. By the left-invariance of the ω_i, we have the following **Maurer-Cartan system of differential equations**:

$$\omega_i(z,dz)=\omega_i(y,dy),\qquad 1\leqslant i\leqslant n. \tag{3}$$

Regarding $x_1,\ldots,x_n$ as parameters, put $z_i=\varphi_i(y_1,\ldots,y_n)$. Then (3) is equivalent to

$$\sum_j A_{ij}(z)\frac{\partial\varphi_j}{\partial y_k}=A_{ik}(y),\qquad 1\leqslant i,k\leqslant n. \tag{3'}$$

Now (3′) is †completely integrable, by (1). By solving the system (3′) of differential equations under the initial conditions

$$\varphi_j(0,\ldots,0)=x_j,\qquad 1\leqslant j\leqslant n, \tag{4}$$

we obtain the multiplication functions $z_i=f_i(x_1,\ldots,x_n;y_1,\ldots,y_n)$ [1].

By the exponential mapping $\exp:\mathfrak{g}\to G$ from the Lie algebra $\mathfrak{g}$ of a Lie group G, a sufficiently small neighborhood U of the zero element of $\mathfrak{g}$ is mapped bijectively and bi-real-analytically onto a neighborhood V of the identity element of G (→ Section Q above). Now by taking U sufficiently small, the product $(\exp X)(\exp Y)$ for X, Y in U can be explicitly given in the form $\exp Z$ by the **Campbell-Hausdorff formula**. Namely, $Z\in\mathfrak{g}$ is given by

$$Z=c_1(X,Y)+c_2(X,Y)+\ldots,$$

where

$$c_1(X,Y)=X+Y$$

and $c_2,c_3,\ldots$ are defined recursively by

$$(n+1)c_{n+1}(X,Y)=\frac{1}{2}[X-Y,c_n]$$

$$+\sum_{p\geqslant 1,\,2p\leqslant n}K_{2p}\sum{}'[c_{k_1},[\ldots[c_{k_{2p}},X+Y]\ldots],$$

where the second summation is taken over all $2p$-tuples $(k_1,\ldots,k_{2p})$ of positive integers $k_1,\ldots,k_{2p}$ satisfying $k_1+\ldots+k_{2p}=n$. Here the $K_2,K_4,K_6,\ldots$ are rational numbers defined by the Taylor expansion

$$\frac{z}{1-e^{-z}}-\frac{1}{2}z=1+\sum_{p=1}^{\infty}K_{2p}z^{2p}.$$

For example,

$$c_2=\frac{1}{2}[X,Y],$$

$$c_3=\frac{1}{12}[[X,Y],Y]$$
$$-\frac{1}{12}[[X,Y],X],$$

$$c_4=-\frac{1}{48}[Y,[X,[X,Y]]]$$
$$-\frac{1}{48}[X,[Y,[X,Y]]],$$

etc. [23].

S. Maximal Compact Subgroups

Any compact subgroup of a connected Lie group G is contained in a maximal compact subgroup of G. Any two maximal compact subgroups of G are conjugate in G. Let K be a maximal compact subgroup of G. Then K is connected and G is homeomorphic to the direct product of K with a Euclidean space $\mathbf{R}^m$ (**Cartan-Mal'tsev-Iwasawa theorem**).

T. Iwasawa Decomposition

Let G be a connected semisimple Lie group. Suppose that the center of G is a finite group. Let $\mathfrak{g}=\mathfrak{k}+\mathfrak{a}+\mathfrak{n}$ be an †Iwasawa decomposition of the Lie algebra $\mathfrak{g}$ of G (→ 248 Lie Algebras). Then the connected Lie subgroups K, A, N associated with $\mathfrak{k}$, $\mathfrak{a}$, $\mathfrak{n}$, respectively, are all closed subgroups of G. Furthermore, K is a maximal compact subgroup of G, A is isomorphic to the additive group $\mathbf{R}^s$ for a suitable s, and N is a simply connected nilpotent Lie group. The mapping $K\times A\times N\rightarrow G$ given by $(k,a,n)\rightarrow kan$ is bijective and is an isomorphism of analytic manifolds. The decomposition $G=KAN$ is called an **Iwasawa decomposition** of G. An Iwasawa decomposition is unique in the following sense: Let $G=K'A'N'$ be another Iwasawa decomposition. Then there exists an element g in G such that $gKg^{-1}=K'$, $gAg^{-1}=A'$, $gNg^{-1}=N'$ [9, 3].

U. Complexification of Compact Lie Groups

Let G be a compact Lie group. Denote by $C(G)$ the commutative associative algebra of all $\mathbf{C}$-valued continuous functions defined on G relative to the usual multiplication in $C(G)$. Then for each $\sigma\in G$, $L_\sigma(R_\sigma)$ acts on $C(G)$ by $L_\sigma f=f\circ L_\sigma$ $(R_\sigma f=f\circ R_\sigma)$. Now put $\mathfrak{o}(G)=\{f\in C(G)\,|\,\dim\sum_{\sigma\in G}\mathbf{C}L_\sigma f<\infty\}$. Then $\mathfrak{o}(G)$ is a subalgebra of $C(G)$, called the **representative ring** of G. Elements of $\mathfrak{o}(G)$ are called **representative functions** of G because for an element $f\in C(G)$, $f\in\mathfrak{o}(G)$ if and only if there exists a continuous matrix representation $\sigma\rightarrow(d_{ij}(\sigma))$ of G such that f is a $\mathbf{C}$-linear combination of the d_{ij}. For each $\sigma\in G$, L_σ and R_σ preserve $\mathfrak{o}(G)$. With respect to the uniform norm $\|f\|_\infty=\max_{\sigma\in G}|f(\sigma)|$ of $C(G)$, $\mathfrak{o}(G)$ is everywhere dense in $C(G)$ (**Peter-Weyl theory**) (→ 69 Compact Groups B). This implies the existence of a faithful (=injective) representation $\rho:G\rightarrow GL(n,\mathbf{C})$ of G. Furthermore, $\mathfrak{o}(G)$ is a finitely generated algebra over $\mathbf{C}$. Thus there exists a representation $\sigma\rightarrow(d_{ij}(\sigma))$ such that $\mathfrak{o}(G)$ is generated over $\mathbf{C}$ by the d_{ij}. Such a representation is called a **generating representation**.

Denote the group of all automorphisms of the algebra $\mathfrak{o}(G)$ by A. Let $\tilde{G}$ be the centralizer in A of the subgroup $\{L_\sigma\,|\,\sigma\in G\}$ of A. Then we have a bijective mapping $\alpha\rightarrow\alpha'$ defined by $\alpha'(f)=(\alpha(f))(e)$ $(f\in\mathfrak{o}(G))$ from the group $\tilde{G}$ onto the set $\mathrm{Hom}_{\mathbf{C}}(\mathfrak{o}(G),\mathbf{C})$, which is an affine variety. Thus every generating representation $\rho:\sigma\rightarrow(d_{ij}(\sigma))_{1\leqslant i,j\leqslant n}$ defines a faithful matrix representation $\tilde{\rho}$ of $\tilde{G}$ by $\tilde{\rho}(\alpha)=(\alpha'(d_{ij}))_{1\leqslant i,j\leqslant n}$ by means of the bijection $\alpha\rightarrow\alpha'$ from $\tilde{G}$ onto $\mathrm{Hom}_{\mathbf{C}}(\mathfrak{o}(G),\mathbf{C})$. Furthermore, the image $\tilde{\rho}(\tilde{G})$ is an †algebraic subgroup of $GL(n,\mathbf{C})$ given by $\{\beta(d_{ij})_{1\leqslant i,j\leqslant n}\,|\,\beta\in\mathrm{Hom}_{\mathbf{C}}(\mathfrak{o}(G),\mathbf{C})\}$. Thus $\tilde{G}$ has the structure of a †linear algebraic group, which is actually independent of the choice of a generating representation. (Hence $\tilde{G}$ also has the structure of a complex Lie group.) Now $\sigma\rightarrow R_\sigma$ defines an injective continuous homomorphism from G into $\tilde{G}$, and we can regard G as a subgroup of $\tilde{G}$. Then for $\alpha\in\tilde{G}$, α is in G if and only if $\alpha(\bar{f})=\overline{\alpha(f)}$ for every f in $\mathfrak{o}(G)$ (**Tannaka duality theorem**) (→ 69 Compact Groups D). G is a maximal compact subgroup of $\tilde{G}$, and every maximal compact subgroup of $\tilde{G}$ is conjugate to G in $\tilde{G}$. Then complex Lie algebra $\tilde{\mathfrak{g}}$ of $\tilde{G}$ is isomorphic to the complexification $\mathfrak{g}^{\mathbf{C}}=\mathbf{C}\otimes_{\mathbf{R}}\mathfrak{g}$ of the Lie algebra $\mathfrak{g}$ of G. $\tilde{G}$ is homeomorphic to the direct product of G with a Euclidean space $\mathbf{R}^N$, where $N=\dim G$. For a complex analytic function φ defined on $\tilde{G}$, $\varphi=0\Leftrightarrow\varphi|_G=0$. In particular, $\tilde{G}$ is the closure of G relative to the †Zariski topology of $\tilde{G}$. The group $\tilde{G}$ is called the **Chevalley complexification** of G, which we denote by $G^{\mathbf{C}}$. Let G_1, G_2 be compact Lie groups, and let $\varphi:G_1\rightarrow G_2$ be a continuous homomorphism. Then φ can be extended uniquely to a rational homomorphism $\varphi^{\mathbf{C}}:G_1^{\mathbf{C}}\rightarrow G_2^{\mathbf{C}}$. In particular, every complex representation (ρ, V) can be extended uniquely to a †rational representation $(\rho^{\mathbf{C}}, V)$ of $G^{\mathbf{C}}$. Every complex analytic representation $\tilde{\rho}$ of $G^{\mathbf{C}}$ is a rational representation, and $\tilde{\rho}=\rho^{\mathbf{C}}$, where $\rho=\tilde{\rho}|_G$. Thus $\tilde{\rho}$ is completely reducible. Also, we have a bijection from the classes of irreducible continuous representations of G onto the classes of irreducible complex analytic representations of $G^{\mathbf{C}}$. For a closed subgroup H of G, the complexification $H^{\mathbf{C}}$ of H coincides with the closure of H in $G^{\mathbf{C}}$ relative to the Zariski topology of $G^{\mathbf{C}}$.

If (ρ, V) is a generating representation of G, then $G^{\mathbf{C}}\cong\rho^{\mathbf{C}}(G^{\mathbf{C}})\subset GL(V)$. Furthermore, $\rho^{\mathbf{C}}(G^{\mathbf{C}})$ is an algebraic subgroup of $GL(V)$ that is completely reducible on V. Conversely, let F be an algebraic subgroup of $GL(V)$ that is completely reducible on V. Then there exists a compact Lie group G such that $G^{\mathbf{C}}\cong F$ (as algebraic groups). If an algebraic subgroup F of $GL(n,\mathbf{C})$ satisfies ${}^t\bar{F}=F$, then $F\cong G^{\mathbf{C}}$, where $G=F\cap U(n)$. Now let F be a connected semisimple complex Lie group. Then $F\cong G^{\mathbf{C}}$ for

every maximal compact subgroup G of F. In particular, F has a faithful representation.

For example, let $G \cong U(n)$, $O(n)$, $SO(n)$, $Sp(n)$, respectively. Then $G^{\mathbf{C}} \cong GL(n, \mathbf{C})$, $O(n, \mathbf{C})$, $SO(n, \mathbf{C})$, $Sp(n, \mathbf{C})$, respectively [1].

V. History

S. Lie was the first to consider Lie groups, in the late 19th century; at that time they were called continuous groups. His motivation was to treat the various geometries from a group-theoretic point of view (→ 137 Erlangen Program) and to investigate the relationship between differential equations and the group of transformations preserving their solutions. Lie groups were studied locally, and the notion of Lie algebras was introduced. It was observed that the properties of Lie groups reflect the properties of Lie algebras to a large degree. Even in this early stage, the notions of solvable and semisimple Lie algebras were introduced together with proofs of basic properties. The Galois theory of linear differential equations (by E. Vessiot and others) is contained in Lie's theory. Early in the 20th century, the theory of infinite-dimensional Lie groups was studied by E. Cartan. After him, however, there was a long lull until it was taken up again by M. Kuranishi and others in the 1950s. We restrict ourselves to the consideration of finite-dimensional Lie groups.

From 1900 to 1930, E. Cartan and H. Weyl obtained a complete classification of the semisimple Lie algebras and determined their representations and characters. They also devised useful methods for investigating the structure of these algebras, and did pioneering work on global structures of the underlying manifolds of Lie groups. After them (1930–1950), these results were systematized and refined by C. Chevalley, Harish-Chandra, and others. In the same period, K. Iwasawa [9] clarified Cartan's idea, showing that the only Lie groups that are topologically important are compact Lie groups. He also obtained the Iwasawa decomposition, which has become a basic tool in the study of semisimple Lie groups. At the same time, Iwasawa contributed to Hilbert's fifth problem, which seeks to characterize Lie groups among topological groups. This problem was solved by A. M. Gleason, D. Montgomery, L. Zippin, and H. Yamabe in 1952. Since 1950, the topological properties of Lie groups have attracted considerable attention. The methods of Cartan, who postulated de Rham's theory, and H. Hopf, who used the properties of groups extensively, were succeeded by systematic applications of the general theory of algebraic topology. In particular, the topological theory of †fiber bundles was applied to homogeneous spaces G/H by A. Borel and others. Thus homology groups of Lie groups were completely determined, and homotopy groups of Lie groups were determined to a considerable extent. At the present time, the following fields, rather than Lie groups themselves, are objects of extensive study: structure and analysis of homogeneous spaces, algebraic groups, infinite-dimensional unitary representations, and finite or discrete subgroups (→ 13 Algebraic Groups, 109 Differential Geometry, 427 Topology of Lie Groups and Homogeneous Spaces, 437 Unitary Representations).

References

[1] C. Chevalley, Theory of Lie groups I, Princeton Univ. Press, 1946.
[2] L. S. Pontryagin, Topological groups, first edition, Princeton Univ. Press, 1939, second edition, Gordon & Breach, 1966. (Original in Russian, 1954.)
[3] S. Helgason, Differential geometry and symmetric spaces, Academic Press, 1962.
[4] E. Cartan, Sur la structure des groupes de transformations finis et continus, Thesis, Paris, Nony, 1894. (Oeuvres complètes, Gauthier-Villars, 1952, pt. I, vol. 1, p. 137–287.)
[5] H. Weyl, The structure and representation of continuous groups, Lecture notes, Institute for Advanced Study, Princeton, 1935.
[6] H. Weyl, The classical groups, Princeton Univ. Press, revised edition, 1946.
[7] S. Lie, Theorie der Transformationsgruppen, Teubner, I, 1888; II, 1890; III, 1893 (Chelsea, 1970).
[8] F. Peter and H. Weyl, Die Vollständigkeit der primitiven Darstellungen einer geschlossenen kontinuierlichen Gruppe, Math. Ann., 97 (1927), 737–755. (Weyl. Gesammelte Abhandlungen, Springer, 1968, vol. 3, p. 58–75.)
[9] K. Iwasawa, On some types of topological groups, Ann. Math., (2) 50 (1949), 507–558.
[10] G. D. Mostow, A new proof of E. Cartan's theorem on the topology of semisimple groups, Bull. Amer. Math. Soc., 55 (1949), 969–980.
[11] C. Chevalley, On the topological structure of solvable groups, Ann. Math., (2) 42 (1941), 668–675.
[12] D. Montgomery and L. Zippin, Topological transformation groups, Interscience, 1955.
[13] A. Borel, Topology of Lie groups and characteristic classes, Bull. Amer. Math. Soc., 61 (1955), 397–432.

[14] Y. Matsushima, Differentiable manifolds, Dekker, 1972. (Original in Japanese, 1965.)
[15] G. P. Hochschild, The structure of Lie groups, Holden-Day, 1965.
[16] P. M. Cohn, Lie groups, Cambridge Univ. Press, 1957.
[17] J.-P. Serre, Lie algebras and Lie groups, Benjamin, 1965.
[18] H. Freudenthal, Linear Lie groups, Academic Press, 1969.
[19] N. Bourbaki, Eléments de mathématique, groupes et algèbras de Lie, Actualités Sci. Ind., Hermann, 1960–1975, chs. 1–8.
[20] J. F. Adams, Lectures on Lie groups, Benjamin, 1969.
[21] G. Warner, Harmonic analysis on semi-simple Lie Groups I, II, Springer, 1972.
[22] N. R. Wallach, Harmonic analysis on homogeneous spaces, Dekker, 1973.
[23] V. S. Varadarajan, Lie groups, Lie algebras, and their representations, Prentice-Hall, 1974.
[24] S. Helgason, Differential geometry, Lie groups, and symmetric spaces, Academic Press, 1978.

250 (XVII.3) Limit Theorems in Probability Theory

A. General Remarks

Let S_n be the number of successes in n †Bernoulli trials with probability p for success. Then the classical de Moivre-Laplace limit theorem implies that

$$P\left(\frac{S_n-np}{\sqrt{np(1-p)}}<x\right)\to\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-u^2/2}du \qquad (n\to\infty),$$

for all x. It is also known that for any $\varepsilon>0$

$$P(|n^{-1}S_n-p|>\varepsilon)\to 0 \qquad (n\to\infty),$$

which is a special case of the law of large numbers. General forms of such convergences of sequences of †random variables are explained in this article.

In †probability theory, theorems concerning †convergence in distribution, †convergence in probability, and †almost sure convergence of sequences of †random variables are subsumed under the generic term of **limit theorems**. When the †probability distributions of a sequence of random variables converge to the distribution F (→ 341 Probability Measures F), then F is called the **limit distribution**.

B. Convergence in Distribution of Sums of Independent Random Variables

A sequence of random variables $\{X_{nk}\}$ $(1\leqslant k\leqslant k_n, n\geqslant 1)$ $(k_n\to\infty)$ is said to be **infinitesimal** if $\max_{1\leqslant k\leqslant k_n}P(|X_{nk}|\geqslant\varepsilon)\to 0$ $(n\to\infty)$ for every $\varepsilon>0$. When $X_{n1},\ldots,X_{nk_n}$ are †independent for every n and $\{X_{nk}\}$ $(1\leqslant k\leqslant k_n, n\geqslant 1)$ is infinitesimal, the set of all limit distributions for the sums $S_n=X_{n1}+\ldots+X_{nk_n}-A_n$ (the A_n are suitable constants) coincides with the class of infinitely divisible distributions (→ 341 Probability Measures G). Suppose that the †characteristic function f of an infinitely divisible distribution F is given by †Lévy's canonical form

$$\log f(t)=i\gamma t-\frac{\sigma^2t^2}{2}+\int_{-\infty}^{0-}\left(e^{itx}-1-\frac{itx}{1+x^2}\right)dM(x)+\int_{0+}^{\infty}\left(e^{itx}-1-\frac{itx}{1+x^2}\right)dN(x),$$

and the distribution function of X_{nk} is $F_{nk}(x)$. Then a necessary and sufficient condition for the distribution of $S_n-\gamma_n$ to converge to F for some sequence $\{\gamma_n\}$ is that (i) for every †continuity point x of $M(x)$ and $N(x)$,

$$\sum_{k=1}^{k_n}F_{nk}(x)\to M(x) \qquad (x<0),$$

$$\sum_{k=1}^{k_n}(F_{nk}(x)-1)\to N(x) \qquad (x>0),$$

$$\lim_{\varepsilon\to 0}\limsup_{n\to\infty}\sum_{k=1}^{k_n}\left(\int_{|x|<\varepsilon}x^2dF_{nk}(x)-\left(\int_{|x|<\varepsilon}xdF_{nk}(x)\right)^2\right)$$
$$=\lim_{\varepsilon\to 0}\liminf_{n\to\infty}\sum_{k=1}^{k_n}\left(\int_{|x|<\varepsilon}x^2dF_{nk}(x)-\left(\int_{|x|<\varepsilon}xdF_{nk}(x)\right)^2\right)$$
$$=\sigma^2.$$

Applying these results to special distribution functions, we obtain various kinds of limit distributions.

(1) The Central Limit Theorem. In the central limit theorem, the limit distribution is a †normal distribution. A necessary and sufficient condition for the distributions of the sums $S_n=B_n^{-1}(X_1+\ldots+X_n)-A_n$ $(B_n\to\infty)$ of a sequence of independent random variables $\{X_n\}$ to converge to the †standard normal distribution $N(0,1)$ and for $\{B_n^{-1}X_k\}$ $(1\leqslant k\leqslant n,$

$n \geq 1$) to become infinitesimal is that

$$\lim_{n\to\infty} \sum_{k=1}^{n} \int_{|x|>\varepsilon B_n} dF_n(x)=0 \quad \text{and}$$

$$\lim_{n\to\infty} B_n^{-2} \sum_{k=1}^{n} \left(\int_{|x|<\varepsilon B_n} x^2\,dF_k(x) - \left(\int_{|x|<\varepsilon B_n} x\,dF_k(x) \right)^2 \right) = 1,$$

where $F_k(x)$ is the distribution function of X_k. When the †variance of each X_k is finite, we put

$$B_n^2 = \sum_{k=1}^{n} V(X_k), \quad A_n = B_n^{-1} \sum_{k=1}^{n} E(X_k),$$

where $V(X)$, $E(X)$ stand for the variance and mean of X, respectively. Then the necessary and sufficient condition is replaced by the **Lindeberg condition**:

$$\lim_{n\to\infty} B_n^{-2} \sum_{k=1}^{n} \int_{|x|>\varepsilon B_n} x^2\,dF_k(x+E(X_k))=0 \quad \text{for every } \varepsilon>0.$$

In particular, if the X_n have the same distribution with finite variance, the corresponding Lindeberg condition is satisfied, and the central limit theorem holds. When the variables give outcomes of independent †Bernoulli trials, the proposition is reduced to the classical theorem of de Moivre and Laplace. Moreover, if X_n has the finite †absolute moment $m_{2+\delta}^{(n)} = E(|X_n|^{2+\delta})$ of order $2+\delta$ and $E(X_n)=0$, then the Lindeberg condition is implied by the **Lyapunov condition**

$$B_n^{-(2+\delta)} \sum_{k=1}^{n} m_{2+\delta}^{(k)} \to 0 \qquad (n\to\infty).$$

(2) The Law of Small Numbers. In order that the distributions of the sums $S_n = X_{n1} + \ldots + X_{nk_n}$ of infinitesimal independent random variables converge to the †Poisson distribution $P(\lambda)$ with mean λ, it is necessary and sufficient that for every $\varepsilon \in (0,1)$,

$$\lim_{n\to\infty} \sum_{k=1}^{k_n} \int_{R_\varepsilon} dF_{nk}(x)=0, \qquad R_\varepsilon = \{x \mid |x-1| \geq \varepsilon, \quad |x| \geq \varepsilon\};$$

$$\lim_{n\to\infty} \sum_{k=1}^{k_n} \int_{|x-1|<\varepsilon} dF_{nk}(x)=\lambda;$$

$$\lim_{n\to\infty} \sum_{k=1}^{k_n} \int_{|x|<\varepsilon} x\,dF_{nk}(x)=0; \quad \text{and}$$

$$\lim_{n\to\infty} \sum_{k=1}^{k_n} \left(\int_{|x|<\varepsilon} x^2\,dF_{nk}(x) - \left(\int_{|x|<\varepsilon} x\,dF_{nk}(x) \right)^2 \right) = 0.$$

From this proposition follows the classical **law of small numbers**: If the probability of success p_k for the kth outcome of independent trials satisfies $kp_k = \lambda$ with constant λ independent of k, then the total number S_n of success up to the nth trial converges in distribution to the Poisson distribution with mean λ.

(3) The Law of Large Numbers. In the law of large numbers the limit distribution is the †unit distribution. Given a sequence of independent random variables X_n with distribution function $F_n(x)$ with mean a_n, a necessary and sufficient condition for $n^{-1}\sum_{k=1}^{n}(X_k - a_k)$ to converge to 0 in probability is that

$$\lim_{n\to\infty} \sum_{k=1}^{n} \int_{|x|>n} dF_k(x+a_k)=0,$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \int_{|x|<n} x\,dF_k(x+a_k)=0, \quad \text{and}$$

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{k=1}^{n} \left(\int_{|x|<n} x^2\,dF_k(x+a_k) - \left(\int_{|x|<n} x\,dF_k(x+a_k) \right)^2 \right) = 0.$$

In particular, this is the case when (i) X_n has the finite variance $V(X_n)$ and $n^{-2}\sum_{k=1}^{n} V(X_k) \to 0$; or (ii) the X_n, $n \geq 1$, obey the same distribution with finite mean.

(4) Convergence to Quasistable Distributions. The set of limit distributions for the sums $S_n = B_n^{-1}(X_1 + \ldots + X_n) - A_n$ of identically distributed independent random variables $\{X_n\}$ (with suitable constants A_n, B_n) coincides with the set of quasistable distributions (→ 341 Probability Measures G). If $G(x)$ is the distribution function of the X_i, in order for the limit distribution to be normal it is necessary and sufficient that

$$\lim_{K\to\infty} K^2 \int_{|x|>K} dG(x) \Big/ \int_{|x|<K} x^2\,dG(x)=0.$$

In order for the limit distribution to be quasistable with index α $(0<\alpha<2)$ it is necessary and sufficient that

$$\lim_{x\to\infty} \frac{G(-x)}{1-G(x)} = \frac{c_1}{c_2}, \quad c_1+c_2>0; \text{ and}$$

$$\lim_{x\to\infty} \frac{1-G(x)+G(-x)}{1-G(ax)+G(-ax)} = a^\alpha \quad \text{for every } a>0.$$

In this case the characteristic function of the limit distribution is given by Lévy's canonical form with $M(x) = c_1|x|^{-\alpha}$, $N(x) = -c_2|x|^{-\alpha}$, $\sigma=0$.

(5) Refinement of Central Limit Theorem. Let $\{X_n\}$ be a sequence of identically distributed independent random variables with $E(X_i)=0$, $\sigma^2 = V(X_i)$, $E(|X_i|^3) < \infty$. Let $\Phi_n(x)$ be the distribution function of $S_n = (X_1 + \ldots + X_n)/$

$\sigma\sqrt{n}$ and $\Phi(x)$ the distribution function of the normal distribution $N(0,1)$. Then we have

$$\Phi_n(x)-\Phi(x)$$
$$=\frac{\exp(-x^2/2)}{\sqrt{2\pi n}}(Q(x)+R(x))+o\left(\frac{1}{\sqrt{n}}\right)$$

uniformly in x, where $Q(x)=(1-x^2)E(X_1^3)/6\sigma^3$; $R(x)$ is identically zero when X_i does not have a †lattice distribution; but $R(x)=d\sigma^{-1}R_1((x+a_n)\sigma\sqrt{n}\,d^{-1})$, $R_1(x)=[x]-x+1/2$, and $a_n=-\sqrt{n}\,a\sigma^{-1}$ when X_i has the lattice distribution with maximal span d: $P(X_i\in\{a+kd\,|\,k=0,\pm1,\ldots\})=1$. More accurate asymptotic expansions for $\Phi_n(x)$ are derived when the X_i have absolute moments of higher orders [1]. Asymptotic expressions are also obtained when the X_i are not identically distributed [2] or the limit distribution is †stable [3].

(6) The Local Limit Theorem. The local limit theorem is concerned with the density function of the limit distribution. Let $\{X_n\}$ be a sequence of identically distributed lattice variables with finite variances as in (5). Let

$$S_n=(\sigma\sqrt{n})^{-1}\sum_{k=1}^{n}(X_k-E(X_k)),$$

$$s_{n,j}=(\sigma\sqrt{n})^{-1}d(j-nE(X_1)).$$

Then uniformly in j,

$$\sigma\sqrt{n}\,P(S_n=s_{nj})-(2\pi)^{-1/2}\exp(-s_{nj}^2/2)\to0 \qquad (n\to\infty).$$

The classical **de Moivre-Laplace theorem** for Bernoulli trials is a special case of this theorem. When X_i has a density function, a corresponding limit theorem holds. Local limit theorems are derived also when (i) higher moments exist, (ii) the limit distribution is stable, or (iii) the component variables are not identically distributed [2, 3].

(7) Large Deviations. Let $\{X_n\}$ be a sequence of identically distributed independent random variables with $E(X_i)=0$, $0<\sigma^2=V(X_i)<\infty$. Let $\Phi_n(x)$ be the distribution function of $S_n=(X_1+\ldots+X_n)/\sigma\sqrt{n}$ and $\Phi(x)$ the distribution function of the standard normal distribution $N(0,1)$. In many problems in the field of applications of probability theory, the asymptotic behavior of $1-\Phi_n(x_n)$ as $x_n\to\infty$ with n is of interest. This is called the problem of **large deviation**. Assume that $E(\exp(a|X_i|))<\infty$ for some $a>0$. Then for $x_n\geqslant1$, $x_n=o(\sqrt{n})$ and $n\to\infty$ the following assertions hold:

$$\frac{1-\Phi_n(x_n)}{1-\Phi(x_n)}=\exp\left(\frac{x_n^3}{\sqrt{n}}\lambda\left(\frac{x_n}{\sqrt{n}}\right)\right)\left(1+O\left(\frac{x_n}{\sqrt{n}}\right)\right),$$

$$\frac{\Phi_n(-x_n)}{\Phi(-x_n)}=\exp\left(\frac{-x_n^3}{\sqrt{n}}\lambda\left(\frac{-x_n}{\sqrt{n}}\right)\right)\left(1+O\left(\frac{x_n}{\sqrt{n}}\right)\right),$$

where $\lambda(z)$ is a power series involving the †semi-invariants of X_i and is convergent for sufficiently small $|z|$ [4, 5]. This result has been generalized for the case of random variables not identically distributed [5, 6]. Bounds for $1-\Phi_n(x_n)$ have also been derived by several authors (→ [7]).

C. The Strong Law of Large Numbers and Its Refinements

A sequence of random variables $\{X_n\}$ is said to obey the **strong law of large numbers** if $n^{-1}\sum_{k=1}^{n}(X_k-a_k)$ tends to zero with probability 1 as $n\to\infty$, when constants a_k are properly chosen. When the component variables of the sequence are independent, a useful sufficient condition for validity of the strong law of large numbers is

$$\sum_{n=1}^{\infty}\frac{1}{n^2}E(X_n-E(X_n))^2<\infty.$$

According to Birkhoff's individual ergodic theorem (→ 136 Ergodic Theory B), it is necessary and sufficient that the mean of a component variable exist for a sequence of identically distributed independent variables in order for it to obey the strong law of large numbers.

Suppose that $\{X_n\}$ is a sequence of independent random variables with $E(X_n)=0$, $\sigma_n^2=E(X_n^2)<\infty$, $b_n^2=\sigma_1^2+\ldots+\sigma_n^2\to\infty$, and $P(|X_n|<\lambda_nb_n,\ n=1,2,\ldots)=1$ for a decreasing sequence λ_n tending to 0 as $n\to\infty$. If $\lambda_n=O(1/\varphi^3(b_n^2))$ for a certain increasing continuous function φ, the probability that $X_1+\ldots+X_n>b_n\varphi(b_n^2)$ for infinitely many n equals 0 or 1 according as the integral $\int_1^{\infty}\frac{1}{t}\varphi(t)e^{-\varphi^2(t)/2}dt$ converges or diverges [8]. In particular, when

$$\varphi(t)=(2\log_{(2)}t+3\log_{(3)}t+2\log_{(4)}t+\ldots+2\log_{(k-1)}t+(2+\varepsilon)\log_{(k)}t)^{1/2},$$

where $\log_{(2)}t=\log\log t$, etc., the relevant probability is equal to 0 or 1 according as ε is positive or negative. If we take $\varphi(t)=2\log_{(2)}t$, we are led to **Khinchin's law of the iterated logarithm**: If

$$|X_n|=o(b_n/\sqrt{\log_{(2)}b_n^2}),$$

then

$$P\left(\limsup_{n\to\infty}\frac{X_1+\ldots+X_n}{\sqrt{2b_n^2\log_{(2)}b_n^2}}=1\right)=1.$$

D. Functionals of Sums of Independent Variables

(1) Consider a **recurrent event** which occurs at successive time periods $\tau_1,\tau_1+\tau_2,\ldots$, where

$\{\tau_n\}$ is a sequence of nonnegative independent random variables with the same distribution. Let F be the distribution function of τ_1, $m = E(\tau_1)$, $\sigma^2 = V(\tau_1) < \infty$, and N_n be the number of occurrences to time n. Then

$$\lim_{n\to\infty} P\left(N_n \geqslant \frac{n}{m} - \frac{\sigma\sqrt{n}}{m^{3/2}}x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-u^2/2}du.$$

When $\sigma^2 = \infty$ and $P(\tau_1 < \infty) = 1$, a necessary and sufficient condition for suitably normalized N_n to converge in distribution is that for every $a > 0$ the relation $(1 - F(x))/(1 - F(ax)) \to a^\alpha (x \to \infty)$, $0 \leqslant \alpha < 2$, hold. When $0 \leqslant \alpha < 1$, $P(N_n(1 - F(n)) < x) \to \Psi_\alpha(x)$. When $1 < \alpha < 2$, $P((N_n - nm^{-1})m^{1+1/\alpha}b_n^{-1} < x) \to \Psi_\alpha(x)$ for b_n such that $1 - F(b_n) \sim n^{-1}$, where $\Psi_0(x) = 1 - e^{-x} (x > 0)$, $\Psi_\alpha(x) = 1 - \Phi_\alpha(x^{-1/\alpha})$ $(x > 0)$ for $0 < \alpha < 1$, $\Psi_\alpha(x) = 1 - \Phi_\alpha(-x)$ for $1 < \alpha < 2$, and Φ_α is the quasistable distribution function whose characteristic function is

$$\exp(-|t|^\alpha(\cos 2^{-1}\pi\alpha - i\,\mathrm{sgn}\,t \sin 2^{-1}\pi\alpha)\Gamma(1-\alpha)).$$

In particular, we have

$$\Psi_{1/2}(\sqrt{2/\pi}\,x) = \sqrt{2/\pi}\int_0^x \exp(-u^2/2)\,du.$$

Moreover, the strong law of large numbers and the law of the iterated logarithm connected with N_n have been obtained [9].

(2) Let $\{X_n\}$ be a sequence of independent identically distributed random variables taking integer values with the same distribution, and put $S_n = X_1 + \ldots + X_n$. If the maximum span of the X_i is d and $0 < E(X_i) \leqslant +\infty$, then

$$E\left(\sum_{k=1}^{\infty} \chi_{\{nd\}}(S_k)\right) \to d/E(X_i) \qquad (n \to \infty),$$

$$\to 0 \qquad (n \to -\infty);$$

$$t^{-1}E\left(\sum_{k=1}^{\infty} \chi_{(0,t)}(S_k)\right) \to 1/E(X_i) \qquad (t \to \infty),$$

where χ_E is the †indicator function of the set E, and $1/E(X_i)$ is defined as zero when $E(X_i) = \infty$. If the limit distribution of the normalized sums S_n of $\{X_n\}$ is stable with index β $(1 \leqslant \beta \leqslant 2)$, X_i is symmetric for $\beta = 1$, and $E(X_i) = 0$ for $\beta > 1$, then the event that S_n returns to 0 is a recurrent event. For these cases, setting $\alpha = 1 - \beta^{-1}$, the conditions in (1) imposed on the distribution of the recurrence time τ_1 are satisfied, and the limit distribution for $N_n = \chi_{\{0\}}(S_1) + \ldots + \chi_{\{0\}}(S_n)$ is determined. In this connection, we can show that a similar result is valid when $\chi_{\{0\}}$ is replaced by a member of a considerably wider class of functions, and that the limit distribution is determined as well for M_n which is the number of occurrences of the event $(S_k \geqslant 0, S_{k+1} \leqslant 0)$, $1 \leqslant k \leqslant n$. Similar theorems are established for the case when X_i has nonlattice distribution [10–12].

(3) Let $\{X_n\}$ be a sequence of independent random variables with the same distribution, and put $S_n = X_1 + \ldots + X_n$, $L_n = \chi_{(0,\infty)}(S_1) + \ldots + \chi_{(0,\infty)}(S_n)$. If $n^{-1}E(L_n) \to \alpha$ $(0 \leqslant \alpha \leqslant 1)$, then $P(L_n \leqslant nx) \to G_\alpha(x)$, where G_α is a distribution function on the unit interval $[0, 1]$ such that

$$G_0(0+) - G_0(0-) = 1,$$

$$G_\alpha(x) = \pi^{-1}\sin\pi\alpha\int_0^x u^{\alpha-1}(1-u)^{-\alpha}du$$

for $0 < \alpha < 1$, and $G_1(1+0) - G_1(1-0) = 1$.

When $\alpha = 1/2$, the corresponding limit distribution is given by $G_{1/2}(x) = 2\pi^{-1}\arcsin\sqrt{x}$ (the **arcsine law**). These L_n and N_n, M_n in (2) are considered as †additive functionals of the †Markov process S_n. When $E(X_i) = 0$ and $E(X_i^2) = 1$, we have the limiting relation

$$\lim_{n\to\infty} P\left(\max_{1\leqslant k\leqslant n} S_k < x\sqrt{n}\right) = \begin{cases} 0 & (x \leqslant 0), \\ \sqrt{\dfrac{2}{\pi}}\displaystyle\int_0^x \exp\left(-\frac{u^2}{2}\right)du & (x > 0), \end{cases}$$

$$\lim_{n\to\infty} P\left(\max_{1\leqslant k\leqslant n} |S_k| < x\sqrt{n}\right) = \frac{4}{\pi}\sum_{m=0}^{\infty}\frac{(-1)^m}{2m+1}\exp\left(-\frac{(2m+1)^2\pi^2}{8x^2}\right) \qquad (x \geqslant 0).$$

Furthermore, the limit distributions of $n^{-1}(S_1^2 + \ldots + S_n^2)$ and $n^{-3/2}(|S_1| + \ldots + |S_n|)$ are also determined explicitly [14]. When X_1 is nonnegative, let ν_x be the number of S_n $(n = 1, 2, \ldots)$ not exceeding x, and write $Y_x = S_{\nu_x+1} - x$, $Y'_x = x - S_{\nu_x}$. Then a fundamental theorem of †renewal theory guarantees the existence of the limit distributions of Y_x, Y'_x as $x \to \infty$ [15].

Extensive studies have also been devoted to limit theorems for sums of dependent random variables, especially for †Markov chains [16–20].

E. Convergence in Distribution for Stochastic Processes

(1) General Theory. Let T be a finite or infinite time interval, and let $(\Omega, \mathfrak{B}, P)$ and S be a probability space and topological state space, respectively. Given an S-valued stochastic process $X = (X(t, \omega), t \in T, \omega \in \Omega)$, we denote by $\mathfrak{B}(S^T)$ the σ-algebra generated by the †cylinder sets and denote by P_X and $P_X^{t_1 \ldots t_n}$, respectively, the probability measures over $\mathfrak{B}(S^T)$ induced by the processes X and $(X(t_1, \omega), \ldots, X(t_k, \omega))$. Suppose that there is a sequence of stochastic processes X_n $(n = 1, 2, \ldots)$ whose †sample paths are contained in a subset E of S^T. We can introduce a topology τ on E so that if $\mathfrak{B}_n$ $(n =$

$1, 2, \ldots$) is the P_{X_n}-completion of $\mathfrak{B}(S^T)$, the topological σ-algebra $\mathfrak{B}_E^\tau$ on E becomes a subfamily of $\mathfrak{B}_n \cap E$ ($n = 1, 2, \ldots$). If there exists a probability measure P_0 over $(E, \mathfrak{B}_E^\tau)$ such that

$$\lim_{n\to\infty} \int_E f(y)\,dP_{X_n} = \int_E f(y)\,dP_0$$

for every bounded τ-continuous real function f on E, then the probability distributions of $\{X_n\}$ are said to converge to P_0.

Let (E, ρ) be a †complete metric separable space and P' the Lebesgue measure on $\Omega' = [0, 1]$. If a sequence of E-valued random variables $\{X_n(\omega), 1 \leqslant n < \infty\}$ has the probability distribution P_0 over the σ-algebra $\mathfrak{B}_E^0$ of Borel sets of (E, ρ) as its limit distribution, then it can be shown that there is a sequence of E-valued random variables $\{\xi_n(\omega')\}$ ($\omega' \in \Omega'$; $0 \leqslant n < \infty$) such that the probability distribution of $\xi_n(\omega')$ coincides with that of $X_n(\omega)$ ($1 \leqslant n < \infty$), $\xi_0(\omega')$ has the probability distribution P_0, and $P'(\xi_n(\omega') \to \xi_0(\omega')) = 1$ [21].

Consider a sequence of real-valued stochastic processes $\{X_n(t, \omega)\}$. If (i) there is a probability measure P_0 on $\mathfrak{B}(\mathbf{R}^T)$ such that $P_{X_n}^{t_1 \ldots t_k}$ converges to $P_0^{t_1 \ldots t_k}$ for every choice of $(t_1, \ldots, t_k)$ and (ii)

$$\lim_{h\downarrow 0} \limsup_{n\to\infty} \sup_{|s-t| \leqslant h} P(|X_n(s) - X_n(t)| > \varepsilon) = 0$$

for every $\varepsilon > 0$, then there exists a sequence of real-valued processes $\{\xi_n(t, \omega')\}$ ($\omega' \in \Omega'$; $0 \leqslant n < \infty$) such that for every $t \in T$, $\xi_n(t)$ converges in probability to $\xi_0(t)$, ξ_n and X_n ($1 \leqslant n < \infty$) induce the same probability measure over $\mathfrak{B}(\mathbf{R}^T)$, and ξ_0 has the probability distribution P_0 and is continuous in probability.

(2) Convergence in Distribution of Stochastic Processes Whose Sample Functions Have †Discontinuities of Only the First Kind. Yu. V. Prokhorov and A. V. Skorokhod carried out a systematic study of convergence in distribution for those processes whose sample functions are (S, ρ)-valued and have discontinuities of only the first kind, where (S, ρ) is a complete metric separable space. Let D be the set of functions $x(t)$ from $T = [0, 1]$ to S having discontinuities of only the first kind and right continuous for $0 \leqslant t < 1$ with $x(1-0) = x(1)$, and introduce a metric ρ_D by $\rho_D(x, y) = \inf_\lambda \{\sup_{t\in T} |t - \lambda(t)| + \sup_{t \in T} \rho(x(t), y(\lambda(t)))\}$, where the infimum is taken over all homeomorphisms λ on T. Then (i) in the notation of (1), $\mathfrak{B}_D^{\rho_D} = \mathfrak{B}(S^T) \cap D$, (ii) (D, ρ_D) is †separable but not necessarily †complete, and (iii) there is a complete separable metric ρ_D' in D which is equivalent to ρ_D (i.e., ρ_D and ρ_D' induce the same topology in D). When almost all sample functions of X are elements of D, we write $X \in D$. Given $X_n \in D$ ($n \geqslant 0$), a necessary and sufficient condition for P_{X_n} to converge to P_{X_0} over $\mathfrak{B}_D^{\rho_D}$ is that (i) $P_{X_n}^{t_1 \ldots t_k}$ converge to $P_{X_0}^{t_1 \ldots t_k}$ for any $(t_1, \ldots, t_n) \subset N$, where N is a dense subset of $[0, 1]$ containing 0, 1; and that (ii) $\lim_{h\downarrow 0} \limsup_{n\to\infty} P_{X_n}(\Delta_D(h, X_n) > \delta) = 0$ for every $\delta > 0$, where $\Delta_D(h, x) = \sup\{\min(\rho(x(t_1), x(t)), \rho(x(t_2), x(t))) \mid t - h < t_1 < t < t_2 < t + h\}$.

If C is a subspace of D consisting of all continuous functions, ρ_D is equivalent to $\rho_C(x, y) = \max\{\rho(x(t), y(t)) \mid 0 \leqslant t \leqslant 1\}$, and $\Delta_C(h, x) = \max\{\rho(x(s), x(t)) \mid |s - t| \leqslant h\}$ corresponds to Δ_D. For a sequence of processes $X_n \in C$, a necessary and sufficient condition for P_{X_n} to converge to P_{X_0} over $\mathfrak{B}_C^{\rho_C}$ is that (i) $P_{X_n}^{t_1 \ldots t_k}$ converge to $P_{X_0}^{t_1 \ldots t_k}$ for any $(t_1, \ldots, t_k) \subset N$ (N is a dense set as in the case of D); and that (ii) $\lim_{h\downarrow 0} \limsup_{n\to\infty} P_{X_n}(\Delta_C(h, X_n) > \delta) = 0$ for every $\delta > 0$.

If P_{X_n} converges to P_{X_0} over $\mathfrak{B}_D^{\rho_D}$ (over $\mathfrak{B}_C^{\rho_C}$), then there exists a sequence $\xi_n(t, \omega') \in D$ ($\in C$) ($t \in T$, $\omega' \in \Omega'$) for $0 \leqslant n < \infty$ such that $P_{\xi_n} = P_{X_n}$ for $0 \leqslant n < \infty$ and $P'(\rho_D(\xi_n, \xi_0) \to 0) = 1$ ($P'(\rho_C(\xi_n, \xi_0) \to 0) = 1$).

The following theorem, due to Prokhorov, is useful for practical applications. Let $\{X_\alpha(t) \mid 0 \leqslant t < 1, \alpha \in A\}$ be a family of real-valued processes that satisfy

$$E(|X_\alpha(t) - X_\alpha(s)|^a) \leqslant c|t - s|^{1+b}, \qquad \alpha \in A,$$

where a, b, c are positive constants independent of α. If the family of distributions of $\{X_\alpha(0) \mid \alpha \in A\}$ is tight (→ 341 Probability Measures F), then $X_\alpha(t) \in C$ ($\alpha \in A$), and $\{P_{X_\alpha}\}$ is a tight family of probability distributions over C.

(3) Convergence in Distribution of Markov Processes. Consider a Markov process $\{X(t), 0 \leqslant t \leqslant 1\}$ with complete metric separable space (S, ρ) as its state space and with topological σ-algebra $\mathfrak{B}_\rho$. Suppose that its transition probabilities $P(s, x; t, A)$, $0 \leqslant s < t \leqslant 1$, $A \in \mathfrak{B}_\rho$ are measurable in $x \in S$. Define $V^\varepsilon(x) = \{y \mid \rho(x, y) \geqslant \varepsilon\}$, and introduce two kinds of conditions: (D) for every $\varepsilon > 0$ and $0 \leqslant t \leqslant 1$, $\lim_{h\downarrow 0} \sup_{x \in S, t_1, t_2} P(t_1, x; t_2, V^\varepsilon(x)) = 0$, where t_1, t_2 are subject to the conditions $t \leqslant t_1 < t_2 \leqslant t + h$, $t - h \leqslant t_1 < t_2 < t$, $1 - h \leqslant t_1 < t_2 \leqslant 1$; (C) for every $\varepsilon > 0$, $\sup_{x \in S, t-s \leqslant h} P(s, x; t, V^\varepsilon(x)) \leqslant h \Psi_\varepsilon(h)$, where $\Psi_\varepsilon(h)$ is such that $\Psi_\varepsilon(h) \downarrow 0$ as $h \downarrow 0$ for every $\varepsilon > 0$. According as $\{X(t), 0 \leqslant t \leqslant 1\}$ satisfies conditions (D) or (C), there is a process $X'(t)$ which is equivalent (→ 407 Stochastic Processes) to $X(t)$ belonging respectively to (D) or (C). Convergence in distribution of Markov processes is based on this fact. Suppose that a sequence of Markov processes $\{X_n(t) \mid 0 \leqslant t \leqslant 1\}$, $0 \leqslant n < \infty$, satisfies (i) condition (D) uniformly in n (condition (C) with Ψ_ε independent of n), (ii) with N as in (2), $P_{X_n}^{t_1 \ldots t_k}$ con-

verges to $P_{X_0}^{t_1 \dots t_k}$ for every $(t_1, \dots, t_k) \subset N$. Then P_{X_n} converges to P_{X_0} over $\mathfrak{B}_D^{\rho p}$ (over $\mathfrak{B}_C^{\rho c}$). Convergence in distribution of Markov processes is also implied by the convergence of the †generators of the Markov processes [24, 25].

(4) Donsker's Invariance Principle. Donsker's invariance principle is a kind of central limit theorem for stochastic processes. Let $\{\xi_n\}$ be a sequence of identically distributed independent (real-valued) random variables with $E(\xi_i)=0$ and $0<\sigma^2=V(\xi_i)<\infty$, and let $S_n=\xi_1+\ldots+\xi_n$, $n \geqslant 1$, $S_0=0$. Define random elements $X_n \in C = C[0,1]$, $n \geqslant 1$, by

$$X_n(t,\omega) = \frac{1}{\sigma\sqrt{n}} S_{[nt]}(\omega) + (nt-[nt]) \frac{1}{\sigma\sqrt{n}} \xi_{[nt]+1}(\omega) \qquad (0 \leqslant t \leqslant 1),$$

when $[nt]$ denotes the integer part of the real number nt. Thus $X_n(i/n, \omega) = S_i/\sigma\sqrt{n}$, and X_n is linear on the interval $[(i-1)/n, i/n]$ for $i = 1, 2, \dots, n$. The C-valued random variables X_n induce their probability distributions P_{X_n} over $\mathfrak{B}_C^{\rho c}$, $n \geqslant 1$. Let $X(t)$, $0 \leqslant t \leqslant 1$, be a (real-valued) †Wiener process (→ 45 Brownian Motion A) with $X(0)=0$ and P_X its probability distribution over $\mathfrak{B}_C^{\rho c}$. P_X is called a **Wiener measure**. Then **Donsker's invariance principle** [26] asserts that P_{X_n} converges to P_X.

If we define $X_n' \in D$, $n \geqslant 1$, by

$$X_n'(t,\omega) = \frac{1}{\sigma\sqrt{n}} S_{[nt]}(\omega) \qquad (0 \leqslant t \leqslant 1),$$

and if we consider P_X as a probability measure on the space $(D, \mathfrak{B}_D^{\rho p})$ which is concentrated on C (i.e., $P_X(C)=1$), then the above convergence theorem holds also for the distributions $P_{X_n'}$ of X_n', $n \geqslant 1$, over $\mathfrak{B}_D^{\rho p}$ [20].

The arcsine law (→ Section D (3)) is a nice example of an application of the invariance principle. Define a function f on C by $f(x) = \int_0^1 \chi_{(0,\infty)}(x(t))\,dt$, $x \in C$. Then it is shown that f is measurable and continuous except on a set of Wiener measure 0. Hence the invariance principle implies that the probability distributions of $\{f(X_n)\}$ converge to the distribution of $f(X)$. It is known that $f(x)$ obeys the arcsine law $P(f(X)<a) = 2\pi^{-1} \arcsin\sqrt{a}$, $0 \leqslant a \leqslant 1$ (→ 45 Brownian Motion E). In the random walk case, where $P(\xi_i=1)=P(\xi_i=-1)=1/2$, $nf(X_n)$ is the number of i, $1 \leqslant i \leqslant n$, for which S_{i-1} and S_i are both nonnegative.

(5) Strassen's Invariance Principle. An invariance principle for the law of the iterated logarithm is explained here. Let $\{\xi_n\}$, $\{S_n\}$, and $\{X_n(t)\}$ be the same as in subsection (4) above. Define

$$T_n(t,\omega) = (2\log\log n)^{-1/2} X_n(t,\omega) \qquad (0 \leqslant t \leqslant 1,\ n \geqslant 3).$$

Strassen's invariance principle [28] asserts that with probability 1 the set $\{T_n, n \geqslant 3\}$ is †relatively compact in $C = C[0,1]$, and the set of its limit points coincides with the set K of absolutely continuous functions x on the interval $[0,1]$ such that $x(0)=0$ and $\int_0^1 x'(t)^2\,dt \leqslant 1$. This principle is based on the following fact. Let $X(t)$, $0 \leqslant t < \infty$, be a †Wiener process with $X(0)=0$, and

$$Z_n(t) = (2n\log\log n)^{-1/2} X(nt), \qquad (0 \leqslant t \leqslant 1,\ n \geqslant 3).$$

Then with probability 1 the set $\{Z_n, n \geqslant 3\}$ is relatively compact in C with its limit set equal to K. As applications of the invariance principle we obtain, e.g., (i) the ordinary law of the iterated logarithm

$$P\left(\limsup_{n\to\infty} S_n/(2n\log\log n)^{1/2} = \sigma\right) = 1,$$

and (ii) if $\sigma=1$, for any $a \geqslant 1$,

$$P\left(\limsup_{n\to\infty} (2n\log\log n)^{-a/2} n^{-1} \sum_{i=1}^n |S_i|^a = 2(a+2)^{(a/2)-1} a^{-a/2} \left(\int_0^1 \frac{dt}{\sqrt{1-t^a}}\right)^{-a}\right) = 1,$$

in particular,

$$P\left(\limsup_{n\to\infty} (2n\log\log n)^{-1/2} n^{-1} \sum_{i=1}^n |S_i| = 3^{-1/2}\right) = 1.$$

F. Convergence of Empirical Distributions

Let $\{X_k(\omega) | 1 \leqslant k \leqslant n\}$, $\{Y_j(\omega) | 1 \leqslant j \leqslant m\}$ be independent random variables with the same distribution function F. Let $N_n(x,\omega)$ be the number of k $(k=1,\dots,n)$ such that $X_k \leqslant x$. Then $F_n(x,\omega) = n^{-1}N_n(x,\omega)$ is a distribution function in x called an **empirical distribution function**. According to the Glivenko-Cantelli theorem, $H_n^+ = \sup_x |F_n(x,\omega) - F(x)| \to 0$ $(n\to\infty)$ with probability 1. Let $G_m(x,\omega)$ be the empirical distribution function constructed from the Y_j, and put

$$H_n = \sup_x (F_n(x,\omega) - F(x)),$$

$$H^+(n,m) = \sup_x |F_n(x,\omega) - G_m(x,\omega)|,$$

$$H(n,m) = \sup_x (F_n(x,\omega) - G_m(x,\omega)).$$

If F is continuous, then the following results

hold. According to **Kolmogorov's theorem**, as $n\to\infty$ we have

$$P(\sqrt{n}\,H_n^+ < x)\to K(x),$$

where

$$K(x)=\sum_{k=-\infty}^{\infty}(-1)^k\exp(-2k^2x^2)\qquad(x>0);$$

$$=0\qquad(x\leqslant 0).$$

On the other hand, **Smirnov's theorem** asserts that (i)

$$P(\sqrt{n}\,H_n<x)=0\qquad(x\leqslant 0);$$

$$=1-(1-n^{-1/2}x)^n-x\sqrt{n}\sum_{k=r+1}^{n-1}n^{-n}\binom{n}{k}$$

$$\times(k-x\sqrt{n})^k(n-k+x\sqrt{n})^{n-k-1}$$

$$(0<x\leqslant\sqrt{n},\ r=[x\sqrt{n}]);$$

$$=1\qquad(x>\sqrt{n}),$$

and, as $n\to\infty$,

$$P(\sqrt{n}\,H_n<x)\to L(x),$$

$$L(x)=1-\exp(-2x^2)\qquad(x>0),$$

$$=0\qquad(x\leqslant 0);$$

(ii) when $n\to\infty$, $mn^{-1}\to r$, r a constant,

$$P((nm)^{1/2}(n+m)^{-1/2}H^+(n,m)<x)\to K(x),$$

$$P((nm)^{1/2}(n+m)^{-1/2}H(n,m)<x)\to L(x).$$

In terms of convergence in distribution for Markov processes, these results are interpreted as follows: $F(X_k)$ is uniformly distributed over $[0,1]$. Let $\nu_n(t)$ be the number of k ($k=1,2,\ldots,n$) such that $F(X_k)\leqslant t$, and put $X_n(t)=\sqrt{n}(n^{-1}\nu_n(t)-t)$ $(0<t\leqslant 1)$, $X_n(0)=0$. Then $E(X_n(t))=0$, $E(X_n(t)X_m(t))=\min(s,t)-st$, $H_n^+=\sup_t|X_n(t)|$, and $H_n(t)=\sup_t X_n(t)$. Let $\{X(t)\,|\,0\leqslant t\leqslant 1\}$ be a †Gaussian process with $E(X(t))=0$, $E(X(s)X(t))=\min(s,t)-st$, and put $H^+=\sup_t|X(t)|$, $H=\sup_t X(t)$. Then P_{X_n} converges to P_X over $\mathfrak{B}_D^{\rho_D}$. Therefore

$$P(\sqrt{n}\,H_n^+<x)\to P(H^+<x)\qquad(n\to\infty),$$

$$P(\sqrt{n}\,H_n<x)\to P(H<x)\qquad(n\to\infty).$$

Similarly, if we denote the number of j such that $F(Y_j)\leqslant t$ by $\mu_m(t)$, then $H^+(n,m)=\sup_t|n^{-1}\nu_n(t)-m^{-1}\mu_m(t)|$, and therefore

$$P((nm)^{1/2}(n+m)^{-1/2}H^+(n,m)<x)$$

$$\to P((1+r)^{-1/2}\sup_t|X(t)-\sqrt{r}\,Y(t)|<x)$$

when $n\to\infty$, $nm^{-1}\to r$, where $\{Y(t)\,|\,0\leqslant t\leqslant 1\}$ is independent of $\{X(t)\,|\,0\leqslant t\leqslant 1\}$ and distributed according to the same law as for the latter. The exact as well as asymptotic expressions for the distributions H_n^+, $H^+(n,m)$ are also obtained [33, 34]. Thus the limit distributions can be calculated by using the distributions of the processes $\{X(t)\}$ and $\{Y(t)\}$. The results above are obtained by applying Donsker's invariance principle. The Gaussian process $\{X(t)\,|\,0\leqslant t\leqslant 1\}$ introduced above is called a **Brownian bridge** because it is obtained from a Brownian motion $B(t)$ by conditioning $B(0)=B(1)=0$.

References

[1] B. V. Gnedenko and A. N. Kolmogorov, Limit distributions for sums of independent random variables, Addison-Wesley, 1954. (Original in Russian, 1949.)
[2] V. V. Petrov, Asymptotic expansions for distributions of sums of independent random variables, Theory of Prob. Appl., 4 (1959), 208–211. (Original in Russian, 1959.)
[3] H. Cramér, On asymptotic expansions for sums of independent random variables with a limiting stable distribution, Sankhyā, (A) 25 (1963), 13–24.
[4] H. Cramér, Sur un nouveau théorème-limite de la théorie des probabilités, Actualités Sci. Ind., 736, Hermann, 1938, 5–23.
[5] V. V. Petrov, A generalization of Cramér's limit theorem, Selected Transl. Math. Stat. and Prob., Amer. Math. Soc., 6 (1966), 1–8. (Original in Russian, 1954.)
[6] Yu. V. Linnik, On the probability of large deviations for the sums of independent variables, Proc. 4th Berkeley Symp. on Math. Stat. and Prob., II, Univ. of California Press, 1961, 289–306.
[7] S. V. Nagaev, Large deviations of sums of independent random variables, Ann. Probability, 7 (1979), 745–789.
[8] W. Feller, The general form of the so-called law of the iterated logarithm, Trans. Amer. Math. Soc., 54 (1943), 373–402.
[9] W. Feller, Fluctuation theory of recurrent events, Trans. Amer. Math. Soc., 67 (1949), 98–119.
[10] G. Maruyama, Fourier analytic treatment of some problems on the sums of random variables, Natural Sci. Report, Ochanomizu Univ., 6 (1955), 7–24.
[11] N. Ikeda, Fluctuation of sums of independent random variables, Mem. Fac. Sci. Kyushu Univ., (A) 10 (1956), 15–28.
[12] G. Kallianpur and H. Robbins, The sequence of sums of independent random variables, Duke Math. J., 21 (1954), 285–307.
[13] F. Spitzer, A combinatorial lemma and its application to probability theory, Trans. Amer. Math. Soc., 82 (1956), 323–339.
[14] P. Erdös and M. Kac, On certain limit theorems of the theory of probability, Bull. Amer. Math. Soc., 52 (1946), 292–302.
[15] E. B. Dynkin, Some limit theorems for

sums of independent random variables with infinite mathematical expectations, Selected Transl. Math. Stat. and Prob., Amer. Math. Soc., 1 (1961), 171–189. (Original in Russian, 1955.)
[16] R. L. Dobrushin, Central limit theorem for nonstationary Markov chains I, II, Theory of Prob. Appl., 1 (1956), 65–80, 329–383. (Original in Russian, 1956.)
[17] S. V. Nagaev, Some limit theorems for stationary Markov chains, Theory of Prob. Appl., 2 (1957), 378–406. (Original in Russian, 1957.)
[18] R. L. Dobrushin, Limit theorems for a Markov chain of two states, Selected Transl. Math. Stat. and Prob., Amer. Math. Soc., 1 (1961), 97–134. (Original in Russian, 1953.)
[19] I. A. Ibragimov and Yu. V. Linnik, Independent and stationary sequences of random variables, Wolters-Noordhoff, 1971. (Original in Russian, 1965.)
[20] P. Billingsley, Convergence of probability measures, Wiley, 1968.
[21] A. V. Skorokhod, Limit theorems for stochastic processes, Theory of Prob. Appl., 1 (1956), 261–290. (Original in Russian, 1956.)
[22] Yu. V. Prokhorov, Convergence of random processes and limit theorems in probability theory, Theory of Prob. Appl., 1 (1956), 155–214. (Original in Russian, 1956.)
[23] A. V. Skorokhod, Limit theorems for Markov processes, Theory of Prob. Appl., 3 (1958), 202–246. (Original in Russian, 1958.)
[24] A. V. Skorokhod, Studies in the theory of random processes, Addison-Wesley, 1965. (Original in Russian, 1961.)
[25] D. W. Stroock and S. R. S. Varadhan, Multidimensional diffusion processes, Springer, 1979.
[26] M. D. Donsker, An invariance principle for certain probability limit theorems, Mem. Amer. Math. Soc., 6 (1951), 1–12.
[27] D. Freedman, Brownian motion and diffusion, Holden-Day, 1971.
[28] V. Strassen, An invariance principle for the law of the interated logarithm, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 3 (1964), 211–226.
[29] A. N. Kolmogorov, Sulla determinazione empirica di una legge di distribuzione, Giorn. Ist. Ital. Attuari, 4 (1933), 83–91.
[30] N. V. Smirnov, Approximate laws of random variables from empirical data (in Russian), Uspekhi Mat. Nauk, (old ser.) 10 (1944), 179–206.
[31] W. Feller, On the Kolmogorov-Smirnov limit theorems for empirical distributions, Ann. Math. Statist., 19 (1948), 177–189.
[32] J. L. Doob, Heuristic approach to the Kolmogorov-Smirnov theorems, Ann. Math. Statist., 20 (1949), 393–403.
[33] V. S. Korolyuk, On the discrepancy of empiric distributions for the case of two independent samples, Selected Transl. Math. Stat. and Prob., Amer. Math. Soc., 4 (1963), 105–121. (Original in Russian, 1955.)
[34] V. S. Korolyuk, Asymptotic expansions for the criteria of fit of A. N. Kolmogorov and N. V. Smirnov (in Russian), Izv. Akad. Nauk SSSR, Ser. Mat., 19 (1955), 103–124.

251 (XII.9) Linear Operators

A. General Remarks

In †functional analysis, when we talk about an operator or a mapping T from one space to another, it is important to specify not only the domain $D(T)$ on which T is defined and the range $R(T)$ onto which T maps $D(T)$ but also the spaces X and Y of which $D(T)$ and $R(T)$, respectively, are subsets. Thus a mapping T from a subset $D(T)$ of a (real or complex) linear space X to a linear space Y is called an **operator** from X to Y. The image of $x \in X$ under T is customarily denoted by Tx. An operator T from X to Y is said to be a **linear operator (linear transformation** or **additive operator)** if (i) $D(T)$ is a †linear subspace of X and (ii) $T(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 Tx_1 + \alpha_2 Tx_2$ for all $x_1, x_2 \in D(T)$ and all scalars α_1, α_2. The set of all continuous linear operators T from X to Y with $D(T) = X$ is denoted by $L(X, Y)$. We denote $L(X, X)$ by $L(X)$.

For simplicity we suppose throughout this article that X and Y are †Banach spaces. In this case $L(X, Y)$ consists of all †bounded linear operators from X to Y. Hence $L(X, Y)$ and $L(X)$ are denoted by $\mathbf{B}(X, Y)$ and $\mathbf{B}(X)$, respectively. Some of the statements given remain true in more general situations (→ 424 Topological Linear Spaces). More information about operators in $\mathbf{B}(X, Y)$ can be found elsewhere (→ 37 Banach Spaces; 68 Compact and Nuclear Operators). Examples are grouped together in Section O.

B. Operations on Operators

When T_1 and T_2 are operators from X to Y, the **sum** $T_1 + T_2$ is the operator defined by $D(T_1 + T_2) = D(T_1) \cap D(T_2)$ and $(T_1 + T_2)x = T_1 x + T_2 x$, $x \in D(T_1 + T_2)$. When T_1 is from X to Y and T_2 is from Y to Z, the **product** $T_2 T_1$ is the operator defined by $D(T_2 T_1) = \{x \in D(T_1) \mid T_1 x \in D(T_2)\}$ and $T_2 T_1 x = T_2(T_1 x)$, $x \in D(T_2 T_1)$. The **product** αT of a scalar α and

an operator T is defined in a similar and obvious way. These operators are linear whenever T_1 and T_2 are linear. For an operator T from X to Y the subset $\Gamma(T)=\{(x, Tx)|x\in D(T)\}$ of the product space $X\times Y$ is said to be the **graph** of T. If $\Gamma(T_1)\subset\Gamma(T_2)$, the operator T_2 is said to be an **extension** of T_1 (we write $T_1\subset T_2$). If a linear operator T from X to Y satisfies the condition that $x\neq 0$ implies $Tx\neq 0$, then T has the **inverse operator** T^{-1}, which is a linear operator satisfying $D(T^{-1})=R(T)$ and $R(T^{-1})=D(T)$.

C. Convergence of Operators

The following three topologies are most frequently used in the linear space $\mathbf{B}(X, Y)$, which becomes a †locally convex topological linear space under each of them: (1) the **uniform operator topology**, (2) the **strong operator topology**, and (3) the **weak operator topology**. These topologies are determined by the respective fundamental systems of neighborhoods of 0 consisting of all sets of type (1) $\{T\mid \|T\|<\varepsilon\}$, (2) $\{T\mid \|Tx\|<\varepsilon, x\in\mathbf{X}\}$, and (3) $\{T\mid |f(Tx)|<\varepsilon, x\in\mathbf{X}, f\in\mathbf{Y}'\}$, where ε varies over all positive numbers, $\mathbf{X}$ over all finite subsets of X, and $\mathbf{Y}'$ over all finite subsets of Y', the †dual space of Y. The uniform operator topology is thus the metric topology determined by the †norm in $\mathbf{B}(X, Y)$ (→ 37 Banach Spaces C). Convergence of T_n to T with respect to one of these topologies is referred to as (1) **uniform convergence**, (2) **strong convergence**, or (3) **weak convergence**. We have such convergence if and only if (1) $\|T_n-T\|\to 0$, (2) $\|(T_n-T)x\|\to 0$ for every $x\in X$, or (3) $|f(T_nx-Tx)|\to 0$ for every $x\in X$ and $f\in Y'$.

D. Closed Operators

Closed operators play an important role when we deal, as is frequent in applications, with operators that are not necessarily continuous. An operator T from X to Y is said to be a **closed operator** if the graph of T is closed in $X\times Y$, or equivalently, if $x_n\in D(T)$, $x_n\to x$, and $Tx_n\to y$ imply $x\in D(T)$ and $Tx=y$. If T is continuous and $D(T)$ is closed, then T is closed. Conversely, if a linear operator T is closed and $D(T)$ is closed, then T is necessarily continuous (the **closed graph theorem**). An operator T from X to Y is said to be a **closable operator** if T has a closed extension. A linear operator is closable if and only if $x_n\in D(T)$, $x_n\to 0$, and $Tx_n\to y$ imply $y=0$. The closure of the graph of a closable operator T is the graph of the smallest closed extension $\bar{T}$ of T. Thus $\bar{T}$ is also called the **closure** of T. When T is a linear operator from X to Y with $D(T)$ dense in X, the **dual operator** (or **conjugate operator** or **adjoint operator**) T' of T is defined to be the operator from Y' to X' determined by the relations $D(T')=\{f\in Y'\mid$ there exists $g\in X'$ such that $g(x)=f(Tx)$ for every $x\in D(T)\}$ and $T'f=g$, $f\in D(T')$ (→ 37 Banach Spaces). T' is always closed. If T is a closable linear operator with dense domain, then its smallest closed extension is obtained as the bidual T'' restricted to $D(\bar{T})=\{x\in X\cap D(T'')\mid T''x\in Y\}$.

For a closed linear operator T we define its **nullity** nul T to be equal to $\dim N(T)$, where $N(T)=\{x\in D(T)\mid T_x=0\}$ is the **null space** (i.e., the †kernel) of T, and its **deficiency** def T to be equal to $\dim Y/R(T)$. When at least one of nul T and def T is finite, we define the **index** of T by ind $T=$ nul $T-$ def T. A closed linear operator T is said to be a **Fredholm operator** if $R(T)$ is closed, nul$(T)<\infty$, and def$(T)<\infty$ (→ 68 Compact and Nuclear Operators).

The domain $D(T)$ of a closed linear operator T from a Banach space X to another Y turns out to be a Banach space under the **graph norm** $\|x\|_X+\|Tx\|_Y$, and T is looked upon as a bounded linear operator from this Banach space $D(T)$ to Y.

E. Operators between Hilbert Spaces

Throughout this section we suppose that X and Y are complex †Hilbert spaces. Let T be a densely defined linear operator from X to Y. Instead of the dual T', it is sometimes convenient to use the operator T^* from Y to X determined by the relation $(x, T^*y)=(Tx, y)$, $x\in D(T)$. The operator T^* is called the **adjoint operator** (or **Hilbert space adjoint**) of T. By means of the antilinear isomorphism π_X from X onto X' given by $(\pi_X x)(u)=(u, x)$ (†Riesz's theorem), T^* is related to T' by $T^*=\pi_X^{-1}T'\pi_Y$. The correspondence $T\to T'$ is linear, while the correspondence $T\to T^*$ is antilinear. $D(T^*)$ is dense in Y if and only if T is closable, and in this case the smallest closed extension $\bar{T}$ of T coincides with $T^{**}=(T^*)^*$.

If a densely defined linear operator T in X (i.e., T is from X to X) satisfies $T\subset T^*$, then T is said to be a **symmetric operator** (or **Hermitian operator**). If $T=T^*$, then T is said to be a **self-adjoint operator**. A symmetric operator is always closable. A symmetric operator T is said to be **essentially self-adjoint** if the closure of T is self-adjoint. A self-adjoint operator T is said to be **nonnegative** or **positive** or **positive semidefinite** if $(Tx, x)\geqslant 0$ for every $x\in D(T)$. An operator $T\in\mathbf{B}(X, Y)$ is said to be **partially isometric** if there exists a closed linear subspace M of X such that T is isometric on M

(i.e., $\|Tx\|=\|x\|$, $x\in M$) and zero on the †orthogonal complement of M. The closed linear subspace M (TM) is called the **initial (final) set** of T. An operator $T\in\mathbf{B}(X,Y)$ is partially isometric if and only if T^*T and TT^* are (orthogonal) †projections. The ranges of these projections are the initial and final sets of T, respectively. In particular, if $M=X$, then T is said to be an **isometric operator** or an **isometry**. If $X=Y=M=TM$ or, more generally, if $X=M$ and $Y=TM$, then T is said to be a **unitary operator**. $T\in\mathbf{B}(X)$ is unitary if and only if $T^*=T^{-1}$. A closed linear operator T with dense domain is said to be **normal** if $T^*T=TT^*$. The set of self-adjoint (or unitary) operators in X forms a subclass of the class of all normal operators. The structure of normal operators, especially that of self-adjoint operators, has been studied in detail by means of spectral analysis (→ 390 Spectral Analysis of Operators). Let T be a densely defined closed linear operator from X to Y. Then there exist a nonnegative self-adjoint operator P in X and a partially isometric operator W with initial set $\overline{R(P)}=\overline{R(T^*)}$ such that $T=WP$. The operators P and W are determined uniquely by these requirements. This is called the **canonical decomposition** or the **polar decomposition** of T. For linear operators T in a complex Hilbert space X, the notion of **numerical range** $W(T)=\{(Tx,x)\mid x\in D(T),\ \|x\|=1\}$ is also important. $W(T)$ is a convex set such that $\|T\|/2\leqslant\sup|W(T)|=\sup\{|\lambda|\mid\lambda\in W(T)\}\leqslant\|T\|$. If T is normal, then the closure of $W(T)$ coincides with the †closed convex hull of the †spectrum $\sigma(T)$, and we have $\sup|W(T)|=\|T\|$. A normal operator T is a nonnegative self-adjoint operator if and only if its numerical range is in the positive ray $\lambda\geqslant 0$.

F. Resolvents and Spectra

Let T be a closed linear operator in a complex Banach space X, and I be the identity in X. The set $\rho(T)$ of all complex numbers λ such that $\lambda I-T$ has an inverse in $\mathbf{B}(X)$ is called the **resolvent set** of T, and the complement $\sigma(T)$ of $\rho(T)$ is called the **spectrum** of T. For $\lambda\in\rho(T)$ the operator $R(\lambda;T)=(\lambda I-T)^{-1}\in\mathbf{B}(X)$ (or sometimes $-R(\lambda;T)$) is called the **resolvent** of T. If $\lambda_0\in\rho(T)$, then $\{\lambda\mid|\lambda-\lambda_0|<\|R(\lambda_0;T)\|^{-1}\}\subset\rho(T)$. Hence $\rho(T)$ is an open, and $\sigma(T)$ a closed, set. If T is bounded, then $\sigma(T)$ is a nonempty compact set. However, $\rho(T)$ may be empty (example (2) in Section O) or the whole plane $\mathbf{C}$ in general. The operator $R(\lambda;T)$ is an †analytic operator function of λ in $\rho(T)$ and satisfies the (first) **resolvent equation**

$$R(\lambda_1;T)-R(\lambda_2;T)=(\lambda_2-\lambda_1)R(\lambda_1;T)R(\lambda_2;T)$$

for every $\lambda_1,\lambda_2\in\rho(T)$. For every $T\in\mathbf{B}(X)$ the limit $r(T)=\lim\|T^n\|^{1/n}\leqslant\|T\|$, $n\to\infty$, exists and coincides with the †**spectral radius** $\sup|\sigma(T)|$ of T. An operator $T\in\mathbf{B}(X)$ with $r(T)=0$ is called **quasinilpotent** or **generalized nilpotent**. Concerning the dual operator, we have $\rho(T')=\rho(T)$ and $R(\lambda;T')=R(\lambda;T)'$. If X is a Hilbert space, then $\rho(T^*)=\overline{\rho(T)}$ and $R(\lambda;T^*)=R(\bar{\lambda};T)^*$, where $\bar{\ }$ stands for the complex conjugate.

G. Operational Calculus

In operator theory, the term **operational calculus** generally indicates a way of defining "functions" $f(T)$ of an operator T so that a kind of algebraic homomorphism is established between a set of complex-valued functions f and the corresponding set of operators $f(T)$. The functions and operators that must be taken into consideration depend on the nature of the problems to be solved, and accordingly there are several versions of operational calculus. We describe two typical ones.

(1) Let T be a self-adjoint operator in a complex Hilbert space X with the †spectral resolution $T=\int_{-\infty}^{\infty}\lambda\,dE(\lambda)$, and let f be a complex-valued Borel measurable function on $\mathbf{R}$. Then the operator $f(T)$ is uniquely determined by the following relations (→ 390 Spectral Analysis of Operators):

$$D(f(T))=\left\{x\,\middle|\,\int_{-\infty}^{\infty}|f(\lambda)|^2\,d(E(\lambda)x,x)<\infty\right\};$$

$$(f(T)x,y)=\int_{-\infty}^{\infty}f(\lambda)d(E(\lambda)x,y),\qquad x\in D(f(T)),\quad y\in X.$$

Then $f(T)$ is normal and the correspondence $f\mapsto f(T)$ satisfies the following relations: (i) $(\alpha f+\beta g)(T)\supset\alpha f(T)+\beta g(T)$; (ii) $(fg)(T)\supset f(T)g(T)$; and (iii) $f(T)^*=\bar{f}(T)$. If g is a bounded function, the extensions in (i) and (ii) can be replaced by equalities.

(2) Let X be a complex Banach space, $T\in\mathbf{B}(X)$, and $\mathscr{F}(T)$ the set of all functions holomorphic in a neighborhood of $\sigma(T)$. We define an operator $f(T)\in\mathbf{B}(X)$, $f\in\mathscr{F}(T)$, by

$$f(T)=\frac{1}{2\pi i}\int_C f(t)R(t;T)\,dt,\qquad (*)$$

where C is a closed curve consisting of a finite number of rectifiable Jordan arcs encircling a domain that contains $\sigma(T)$ in its interior and lies with its boundary completely in the domain in which f is holomorphic. The integral does not depend on the curve. In this case relations (i) and (ii) hold with equality in place of extension. Instead of (iii) we have (iv) $\mathscr{F}(T)=\mathscr{F}(T')$ and $f(T')=f(T)'$. The integral

appearing in (*) is sometimes called the **Dunford integral**. In both situations described in (1) and (2) the **spectral mapping theorem** $\sigma(f(T))=f(\sigma(T))$ holds. When these two ways of defining $f(T)$ are possible, the resulting operators coincide.

Another kind of operational calculus can be constructed, for example, when T is the generator of a certain semigroup of operators [1, 2, 4].

H. Isolated Singularities of the Resolvent

Let T be a densely defined closed linear operator in a complex Banach space X and λ_0 an isolated point of $\sigma(T)$. Take a sufficiently small circle C around λ_0 and put

$$E=\frac{1}{2\pi i}\int_C R(\lambda;T)\,d\lambda,$$

which is a projection in X. Then the Laurent expansion around λ_0 of the resolvent is given by

$$R(\lambda;T)=\sum_{n=-\infty}^{\infty}A_n(\lambda-\lambda_0)^n,$$

with $A_n=(-1)^{n+1}(\lambda_0I-T)^{-(n+1)}E$ for $n<0$. When the dimension ν of the range of E is finite, λ_0 is a †pole of $R(\lambda;T)$ with order not exceeding ν, and λ_0 is an †eigenvalue of T with †multiplicity not exceeding ν. Furthermore, E is then a projection onto the †root subspace belonging to the eigenvalue λ_0.

I. Extension of Symmetric Operators

In applications we frequently encounter the problem of finding self-adjoint extensions of a given symmetric operator. Let T be a closed symmetric operator in a complex Hilbert space X. Then $T\pm iI$ is one-to-one, and its range $R_\pm$ is a closed linear subspace of X. The operator $V_T=(T-iI)(T+iI)^{-1}$ from R_+ onto R_- is isometric, and $(I-V_T)R_+$ is dense in X. We call V_T the **Cayley transform** of T. Conversely, let V be an isometric operator from a closed linear subspace M of X onto another one N such that $(I-V)M$ is dense in X. Then the operator $T=i(I+V)(I-V)^{-1}$ is a closed symmetric operator satisfying $V_T=V$. Thus the correspondence $T\rightarrow V_T$ is one-to-one onto; $T\subset S$ if and only if $V_T\subset V_S$, and T is self-adjoint if and only if V_T is unitary. The dimension $n_\pm$ of the subspaces $X\ominus N_\pm=\{x\mid T^*x=\pm ix\}$ are called the **deficiency indices** of T. Denoting the †residual spectrum of T by $\sigma_r(T)$ and putting $\Pi_\pm=\{\lambda\mid \operatorname{Im}\lambda\gtrless 0\}$, we have the following propositions: (i) According as $n_+>0$ or $n_+=0$, $\Pi_+\subset\sigma_r(T)$ or $\Pi_+\subset\rho(T)$ (similarly for n_- in place of n_+); (ii) T has a self-adjoint extension if and only if $n_+=n_-$; and (iii) T is self-adjoint if and only if $n_+=n_-=0$, or, equivalently, if and only if $\Pi_\pm\subset\rho(T)$. When T is symmetric but not closed, the previous arguments can be applied to the closure $\bar{T}$ of T. The deficiency indices of $\bar{T}$ are also called the deficiency indices of T. These arguments can be performed similarly with λ and $\bar{\lambda}$ ($\operatorname{Im}\lambda\neq 0$) in place of i and $-i$.

We now describe two more concrete criteria for T to have self-adjoint extensions. (1) Semibounded operators: If there exists a real number γ such that $(Tx,x)\geqslant\gamma\|x\|^2$ for every $x\in D(T)$, then T has a self-adjoint extension satisfying a similar inequality with the same constant γ. The structure of all self-adjoint extensions of T with the same lower bound γ was studied in detail by M. G. Kreĭn [8]. Among such extensions the **Friedrichs extension** is distinguished as the one having the smallest form domain. (2) Real operators: If T commutes with a conjugation in X, namely, if there exists an antilinear mapping J from X onto X such that $(Jx,Jy)=(y,x)$, $J^2=I$, and $JT=TJ$, then T has a self-adjoint extension (→ Section O (2)).

J. Dissipative Operators

A linear operator T in a Hilbert space X is said to be **dissipative** (resp. **accretive**) if $\operatorname{Re}(Tx,x)\leqslant 0$ (resp. $\geqslant 0$) for every $x\in D(T)$. To extend the definition to operators in a Banach space X, we define, for each $x\in X$, Fx to be the set of all x' in the dual space X' such that $\langle x,x'\rangle=\|x\|^2=\|x'\|^2$. Fx is not empty by virtue of the Hahn-Banach extension theorem (→ 37 Banach Spaces). The multivalued mapping $F:x\rightarrow x'$ is called the **duality mapping**. A linear operator T in X is called **dissipative** (resp. **accretive**) if for each $x\in D(T)$ there exists an $x'\in Fx$ such that $\operatorname{Re}\langle Tx,x'\rangle\leqslant 0$ (resp. $\geqslant 0$), or equivalently if $\|x-\lambda Tx\|\geqslant\|x\|$ for every $x\in D(T)$ and $\lambda>0$ (resp. $\lambda<0$). A dissipative operator T is called **m-dissipative** if $R(I-\lambda T)=X$ for a (and all) $\lambda>0$. Then the half-plane $\operatorname{Re}\lambda>0$ is in the resolvent set $\rho(T)$, and we have $\|(\lambda I-T)^{-1}\|\leqslant 1/\operatorname{Re}\lambda$, $\operatorname{Re}\lambda>0$. In particular, T is closed, and if X is †reflexive, then the domain $D(T)$ is dense. A dissipative operator is called a **maximal dissipative operator** if it has no strict extension that is dissipative. An m-dissipative operator is a maximal dissipative operator. If X is a Hilbert space, then conversely a maximal dissipative operator with dense domain is m-dissipative. Every dissipative operator with dense domain in a Hilbert space can be extended to an m-dissipative operator. A linear operator T in a Banach

space is the †infinitesimal generator of a contraction semigroup of class (C^0), i.e., a †semigroup $\{T_t \mid t \geqslant 0\}$ of class (C^0) such that $\|T_t\| \leqslant 1$ for all $t \geqslant 0$, if and only if T is an m-dissipative operator with dense domain.

The notion of dissipative and accretive operators is also defined for nonlinear operators (→ 286 Nonlinear Functional Analysis).

K. Subnormal Operators and Hyponormal Operators

A bounded linear operator T in a complex Hilbert space X is said to be **subnormal** if there exists a bounded normal operator N in a complex Hilbert space Y, containing X as a closed linear subspace such that $Tx = Nx$ for all $x \in X$. The operator N is called a **normal extension** of T. T is subnormal if and only if $\sum_{m,n}(T^{*m}T^n x_m, x_n) \geqslant 0$ for every finite sequence $\{x_n\}$ of X. Then its normal extension N is determined uniquely up to †unitary equivalence under the minimality requirement that there is no †reducing subspace (→ Section L) for N between X and Y. Every normal or isometric operator is subnormal. The spectrum of a subnormal operator T is obtained from the spectrum of its minimal normal extension N by adding some bounded components of its complement. An operator $T \in \mathbf{B}(X)$ is said to be **hyponormal** if the **self-commutator** $[T^*, T] = T^*T - TT^*$ is nonnegative. Every subnormal operator is hyponormal. If a hyponormal operator T has a **cyclic element** x, i.e., the †linear span of $\{T^n x \mid n = 0, 1, \ldots\}$ is dense in X, then the self-commutator is a †trace class operator (C. Berger and B. Shaw). The planar Lebesgue measure of the spectrum of any hyponormal operator T is not less than $\pi\|[T^*, T]\|$ **(Putnam's theorem)** [10].

L. Invariant Subspaces

A closed linear subspace M of a Banach space X is said to be **invariant** under an operator $T \in \mathbf{B}(X)$ if T maps M into M. If M is invariant under all operators in $\mathbf{B}(X)$ that commute with T, then M is said to be **hyperinvariant** under T. When X is a Hilbert space and M is invariant under both T and T^*, then M is said to **reduce** T. That M reduces T is characterized by the commutativity $TP = PT$, where P is the orthogonal projection to M. The question of whether an arbitrary nonzero bounded linear operator in a separable infinite-dimensional Hilbert space has a nontrivial invariant subspace still remains open but work has progressed significantly in recent years. (1) Every nonzero †compact operator has a nontrivial hyperinvariant subspace (V. I. Lomonosov). (2) An operator T has a nontrivial invariant subspace if it is subnormal or if $\|T\| \leqslant 1$ and the spectrum $\sigma(T)$ covers the whole unit disk of the complex plane (S. Brown, 1979). (3) If T is hyponormal, a certain power T^n has a nontrivial invariant subspace (Berger, 1979).

The complete description of all invariant subspaces is usually difficult even for an operator of quite simple type. For the †shift operator (→ Section O (6)) on the †Hardy space H_2 on the unit circle this was done by A. Beurling [11, 12]. For the †integral operator on the space $L_2(0, 1)$ with †kernel $k(t, s) = \max(t - s, 0)$ the set of all invariant subspaces is †totally ordered with respect to inclusion [13].

M. Dilations

Let X be a Hilbert space and $T \in \mathbf{B}(X)$. A bounded linear operator U in a Hilbert space Y, containing X as a closed linear subspace, is said to be a **dilation** (also **strong** or **power dilation**) of T if $T^n x = PU^n x$ for all $x \in X$, $n = 1, 2, \ldots$, where P is the orthogonal projection from Y to X. If U is unitary, it is called a **unitary dilation** of T. Every contraction $T \in \mathbf{B}(X)$, i.e., an operator T with $\|T\| \leqslant 1$, has a unitary dilation U in a suitably constructed Hilbert space Y (the **dilation theorem**). Such a unitary dilation U is uniquely determined up to †unitary equivalence under the minimality requirement that there is no reducing subspace for U between X and Y. A corollary is the **von Neumann inequality**: For any contraction T and any complex polynomial $p(\lambda)$, the inequality $\|p(T)\| \leqslant \sup_{|\lambda|<1} |p(\lambda)|$ holds. Furthermore, if an operator $S \in \mathbf{B}(X)$ commutes with a contraction T, then there exists a dilation V of S in $\mathbf{B}(Y)$ such that $\|V\| = \|S\|$ and V commutes with the unitary dilation U of T (the **lifting theorem**). If a contraction T is **completely nonunitary** in the sense that the restriction of T to any nontrivial reducing subspace is not unitary, then the minimal unitary dilation has †absolutely continuous spectrum, and the operational calculus $p \mapsto p(T)$ is extended to all functions in the †Hardy space H_∞ [15].

N. Functional Models for Contractions

A **canonical model** for an operator is a "natural" representation of the operator in terms of simpler operators and in a context in which more structure is present. There are several canonical models for bounded linear operators [13–16]. Here, we follow the approach of B. Sz.-Nagy and C. Foiaş that was developed

in connection with unitary dilations of contractions [15, 16].

An †analytic operator function $\Theta(\lambda)$ defined on the open unit disk with values in $\mathbf{B}(X, Y)$, where X and Y are separable complex Hilbert spaces, is denoted by $\{X, Y, \Theta\}$. An analytic operator function $\{X, Y, \Theta\}$ is said to be **contractive** if $\|\Theta(\lambda)\| \leqslant 1$ for all λ. If, in addition, $\|\Theta(0)x\| < \|x\|$ for all nonzero $x \in X$, then it is said to be **purely contractive**. Each contractive $\{X, Y, \Theta\}$ is uniquely decomposed into the direct sum $\Theta(\lambda) = \Theta_0(\lambda) \oplus V$, where $\{X_0, Y_0, \Theta_0\}$ is a purely contractive analytic operator function with closed linear subspaces $X_0 \subseteq X$ and $Y_0 \subseteq Y$, and V is an isometric operator from $X \ominus X_0$ into $Y \ominus Y_0$. The operator function Θ_0 is called the **purely contractive part** of Θ.

Let $\{X, Y, \Theta\}$ be purely contractive. Then the boundary value $\Theta(\zeta) = \text{s-lim}_{r \to 1} \Theta(r\zeta)$ exists for almost all ζ with respect to the Lebesgue measure on the unit circle. Let $\Delta(\zeta) = (I - \Theta^*(\zeta)\Theta(\zeta))^{1/2}$ for $|\zeta| = 1$. Then the operator functions $\Theta(\zeta)$ and $\Delta(\zeta)$ defined on the unit circle induce, respectively, the operators $\Theta \in B(L_2^X, L_2^Y)$ and $\Delta \in B(L_2^X)$ defined by $(\Theta f)(\zeta) = \Theta(\zeta)f(\zeta)$ and $(\Delta f)(\zeta) = \Delta(\zeta)f(\zeta)$ for $|\zeta| = 1$ and $f \in L_2^X$, where L_2^X is the †Hilbert space of X-valued square integrable functions on the unit circle. Consider the Hilbert space $\mathbf{K} = L_2^Y \oplus \overline{R(\Delta)}$ and the unitary operator $\mathbf{U}$ defined by $\mathbf{U}(f \oplus g)(\zeta) = \zeta f(\zeta) \oplus \zeta g(\zeta)$. Consider the linear subspace $\mathbf{H} = [H_2^Y \oplus \overline{R(\Delta)}] \ominus \{\Theta f \oplus \Delta f \mid f \in H_2^X\}$ of $\mathbf{K}$, where H_2^X is the X-valued †Hardy space that is a subspace of L_2^X. Then the operator $\mathbf{T} = \mathbf{P_H}\mathbf{U}$ restricted to $\mathbf{H}$, where $\mathbf{P_H}$ is the orthogonal projection from $\mathbf{K}$ to $\mathbf{H}$, is a completely nonunitary contraction with $\mathbf{U}$ as its minimal unitary dilation. $\mathbf{T}$ is called the contraction generated by $\{X, Y, \Theta\}$. Two purely contractive analytic operator functions $\{X_1, Y_1, \Theta_1\}$ and $\{X_2, Y_2, \Theta_2\}$ can generate contractions which are unitarily equivalent if and only if there are isometries V_1 from Y_1 onto Y_2 and V_2 from X_1 onto X_2 such that $V_1\Theta_1(\lambda) = \Theta_2(\lambda)V_2$ for all λ in the unit disk. In this case, Θ_1 and Θ_2 are said to coincide.

Now let a completely nonunitary contraction T be given. Consider the operator $D_T = (I - T^*T)^{1/2}$ and the closed linear subspace $\mathfrak{D}_T = \overline{R(D_T)}$, and define similarly D_{T^*} and $\mathfrak{D}_{T^*}$ by using T^* instead of T. Then the function $\Theta_T(\lambda) = -T + \lambda D_{T^*}(I - \lambda T^*)^{-1} D_T$ restricted to $\mathfrak{D}_T$ becomes a purely contractive analytic operator function with values in $\mathbf{B}(\mathfrak{D}_T, \mathfrak{D}_{T^*})$ and is called the **characteristic operator function** of T. The contraction generated by the characteristic operator function is unitarily equivalent to T and is called the **functional model** (or the **Sz.-Nagy-Foiaş model**) for T. The spectrum of T coincides with the set consisting of the λ_0 in the open unit disk at which $\Theta_T(\lambda_0)$ fails to have bounded inverse, together with the ζ_0 on the unit circle at which $\Theta_T(\cdot)$ fails to have an analytic extension to a neighborhood Ω of ζ_0 which is unitary on the intersection of Ω and the unit circle. There is a one-to-one correspondence between the invariant subspaces M for T and the **regular factorizations** of Θ_T as a product $\Theta_T(\lambda) = \Theta_2(\lambda) \cdot \Theta_1(\lambda)$ of two contractive analytic operator functions $\{Y, \mathfrak{D}_{T^*}, \Theta_2\}$ and $\{\mathfrak{D}_T, Y, \Theta_1\}$ with a suitable Hilbert space Y. Here the regularity of the factorization means that for almost all ζ in the unit circle the range of $(I - \Theta_2^*(\zeta)\Theta_2(\zeta))^{1/2}$ and the range of $(I - \Theta_1(\zeta)\Theta_1^*(\zeta))^{1/2}$ are linearly independent. Moreover, the characteristic operator function of the restriction of T to M coincides with the purely contractive part of Θ_1, while the characteristic operator function of $P_{M^\perp}T$ restricted to $M^\perp = X \ominus M$ coincides with the purely contractive part of Θ_2, where $P_{M^\perp}$ is orthogonal projection to $M^\perp$.

O. Examples of Linear Operators

(1) Integral Operators. Let E_j, $j = 1, 2$, be linear spaces consisting of measurable functions defined on measure spaces Ω_j with measures μ_j. Let $k(t,s)$ be a measurable function on $\Omega_2 \times \Omega_1$, and define $D(R)$ to be the set of all $x \in E_1$ such that $(Kx)(t) = \int_{\Omega_1} k(t,s)x(s)\,d\mu_1(s)$ belongs to E_2, where the integral is assumed to be absolutely convergent almost everywhere. The mapping that assigns Kx to each $x \in D(R)$ determines a linear operator K from E_1 to E_2 with domain $D(K)$. K is called an **integral operator**, and $k(t,s)$ the **kernel** (or **integral kernel**) (of K). As an example, let $E_j = L_p(\Omega_j)$. $1 \leqslant p \leqslant \infty$, and suppose there exists an $M > 0$ such that

$$\int_{\Omega_1} |k(t,s)|\,d\mu_1(s) \leqslant M,$$

$$\int_{\Omega_2} |k(t,s)|\,d\mu_2(t) \leqslant M.$$

Then $K \in \mathbf{B}(E_1, E_2)$ with $\|K\| \leqslant M$ (→ 68 Compact and Nuclear Operators C; [17]). An integral operator is said to be **Hermitian** if the kernel satisfies $k(t,s) = \overline{k(s,t)}$. A bounded Hermitian integral operator is self-adjoint in $L_2(\Omega)$.

(2) Differential Operators. For $X = L^2(0,1)$, let $\mathfrak{D}_0$ be the set of all $x \in C^2(0,1)$ with compact support in $(0,1)$ and $\mathfrak{D}_1$ the set of all $x \in C^1(0,1)$ such that $x'(t)$ is absolutely continuous in $(0,1)$ with $x'' \in X$. Then the operators T_j, $j = 0, 1$, determined by $(T_j x)(t) = -x''(t)$, $x \in \mathfrak{D}_j$, are linear in X. $T_0^* = T_1$, so that T_0 is a

symmetric operator. Furthermore, T_0 is a real operator with respect to the conjugation $x \to \bar{x}$. Since two linearly independent solutions of $(T_1 - \lambda I)x = 0$ both belong to X, the deficiency indices $n_\pm$ of T_0 are 2. (Note that $\rho(T_1) = \varnothing$.) Self-adjoint extensions of T_0 are obtained by restricting the domain of T_1 by boundary conditions (→ 112 Differential Operators; [1,4]).

(3) Fourier Transforms. For every $x \in L_p(\mathbf{R}^n)$, $1 \leqslant p \leqslant 2$,

$$(Ux)(t) = \lim_{m\to\infty} (2\pi)^{-n/2} \int_{|t| \leqslant m} \exp(-its)x(s)\,ds$$

converges in the norm of $L_q(\mathbf{R}^n)$, $p^{-1} + q^{-1} = 1$. The operator U thus defined belongs to $\mathbf{B}(L_p, L_q)$. When $p = q = 2$, U is a unitary operator in L^2 (→ 160 Fourier Transform).

(4) Singular Integral Operators. Let $\Omega(t)$ be a bounded measurable function on $\mathbf{R}^n$ homogeneous of degree 0. Suppose that the integral of Ω over the unit sphere vanishes and that Ω satisfies the condition

$$\int_0^1 \delta^{-1}\Bigg(\sup_{\substack{|t|=|s|=1\\|t-s|<\delta}} |\Omega(t) - \Omega(s)|\Bigg) d\delta < \infty.$$

Then for every $x(t) \in L_p(\mathbf{R}^n)$, $1 < p < \infty$,

$$(Tx)(t) = \lim_{\varepsilon \downarrow 0} \int_{|t-s|>\varepsilon} \frac{\Omega(t-s)}{|t-s|^n} x(s)\,ds$$

converges in the norm of $L_p(\mathbf{R}^n)$, and the operator T so defined is a bounded linear operator in $L_p(\mathbf{R}^n)$. If $n = 1$ and $\Omega(t) = \pi^{-1} t/|t|$, then T is the †Hilbert transform. If $n > 1$ and $\Omega(t) = \Gamma((n+1)/2)\pi^{-(n+1)/2} t_j/|t|$, $j = 1, \ldots, n$, then T is called the **Riesz transform**.

In general, integral operators on $L_p(\mathbf{R}^n)$ defined by kernels $k(t,s)$ satisfying the estimate $|k(t,s)| \leqslant C|t-s|^{-n}$ are called **Calderón-Zygmund singular integral operators**. These have been widely investigated [18, 19].

(5) Pseudodifferential Operators. Let $a(t,\tau)$ be a function defined on $\mathbf{R}^n \times \mathbf{R}^n$. Then the operator T defined by

$$(Tx)(t) = (2\pi)^{-n/2} \int e^{it\tau} a(t,\tau)\hat{x}(\tau)\,d\tau$$

is called a **pseudodifferential operator**, and $a(t,\tau)$ is called the **symbol** of T, where $\hat{x}$ denotes the Fourier transform of x. This is a generalization of the Calderón-Zygmund singular integral operator. The following is a sufficient condition for T to be a bounded linear operator in $L_p(\mathbf{R}^n)$ for every $1 < p < \infty$: There exists a constant C such that the inequality

$$|\partial_\tau^\alpha a(t,\tau)| \leqslant C(1+|\tau|)^{-|\alpha|}$$

holds for every †multi-index α with $|\alpha| \leqslant n+2$, and there exists a function $\omega(\delta)$ such that $\int_0^1 \omega(\delta)^2 \delta^{-1}\,d\delta < \infty$ and that

$$\sup_{|t-s|\leqslant\delta} |\partial_\tau^\alpha a(t,\tau) - \partial_\tau^\alpha a(s,\tau)| \leqslant \omega(\delta)(1+|\tau|)^{-|\alpha|}$$

for every multi-index α with $|\alpha| \leqslant n+2$ (T. Muramatu and M. Nagase, *Proc. Japan Acad.* (1979); [19]). Pseudodifferential operators play a decisive role in the modern theory of partial differential equations (→ 345 Pseudodifferential Operators).

(6) Toeplitz Operators. Let L_2 be the L_2-space on the unit circle in the complex plane with respect to the linear Lebesgue measure, and let H_2 be the †Hardy space. Each bounded measurable function φ on the unit circle gives rise to the **Toeplitz operator** T_φ in H_2 defined by $T_\varphi f = P(\varphi f)$ for $f \in H_2$, where P is the orthogonal projection from L_2 to H_2. The matrix (α_{nm}) of the Toeplitz operator T_φ with respect to the complete orthonormal basis $\{\chi_n \mid 0 \leqslant n < \infty\}$, where $\chi_n(\zeta) = \zeta^n$ for $|\zeta| = 1$, is given by $\alpha_{nm} = \hat{\varphi}_{n-m}$ where $\hat{\varphi}_k$ is the kth †Fourier coefficient of φ. If φ is in the †Hardy space H_∞, then the Toeplitz operator T_φ is subnormal. When $\varphi(\zeta) = \zeta$, the Toeplitz operator T_φ is called the **shift operator** (or **shift**). A Toeplitz operator T_φ cannot be compact except when $\varphi = 0$. When φ is a continuous function, the Toeplitz operator T_φ becomes a Fredholm operator if and only if φ does not vanish on the unit circle, and in this case the index of T_φ is equal to minus the winding number of the curve traced out by φ with respect to the origin. Toeplitz operators play an important role in approximation theory and prediction theory [20].

References

[1] N. Dunford and J. T. Schwartz, Linear operators, Interscience, I, 1958; II, 1963; III, 1971.
[2] E. Hille and R. S. Phillips, Functional analysis and semigroups, Amer. Math. Soc. Colloq. Publ., second edition, 1957.
[3] F. Riesz and B. Sz.-Nagy, Functional analysis, Ungar, 1955. (Original in French, 1952.)
[4] K. Yosida, Functional analysis, Springer, sixth edition, 1978.
[5] M. H. Stone, Linear transformations in Hilbert space and their applications, Amer. Math. Soc. Colloq. Publ., 1932.
[6] N. I. Akhiezer and U. M. Glazman, Theory of linear operators in Hilbert space, Pitman, 1980. (Original in Russian, 1966.)
[7] I. C. Gokhberg (Gohberg) and M. G. Kreĭn, Introduction to the theory of linear non-self-adjoint operators in Hilbert space,

Amer. Math. Soc. Transl. of Math. Monographs, 1969. (Original in Russian, 1965.)
[8] M. G. Kreĭn, Theory of self-adjoint extensions of semibounded Hermitian operators and its application I, II (in Russian), Math. Sb., 20 (62) (1947), 431–495; 21 (63) (1947), 365–404.
[9] P. R. Halmos, A Hilbert space problem book, Springer, second edition, 1982.
[10] K. Clancey, Seminormal operators, Springer, 1979.
[11] H. Helson, Lectures on invariant subspaces, Academic Press, 1964.
[12] H. Radjavi and P. Rosenthal, Invariant subspaces, Springer, 1973.
[13] M. S. Brodskiĭ, Triangular and Jordan representations of linear operators, Amer. Math. Soc. Transl. of Math. Monographs, 1971. (Original in Russian, 1969.)
[14] C. Pearcy (ed.), Topics in operator theory, Amer. Math. Soc. Surveys, 1974.
[15] B. Sz.-Nagy and C. Foiaş, Harmonic analysis of operators on Hilbert space, North-Holland and Akademia Kiado, 1970.
[16] N. K. Nikolskiĭ, Lectures on shift operators (in Russian), Nauka, 1980.
[17] M. A. Krasnoselskiĭ, P. P. Zabreiko, E. I. Pustylnik, and P. E. Sobolevskiĭ, Integral operators in spaces of summable functions, Noordhoff, 1976. (Original in Russian, 1966.)
[18] E. M. Stein, Singular integrals and differentiability properties of functions, Princeton Univ. Press, 1970.
[19] R. R. Coifman and Y. Meyer, Au delà des opérateurs pseudodifféretiels, Astérisque, 57 (1982).
[20] R. G. Douglas, Banach algebra techniques in operator theory, Academic Press, 1972.

252 (XIII.6) Linear Ordinary Differential Equations

A. General Remarks

Let $p_1(x), \ldots, p_n(x)$, $q(x)$ be known functions of a real (or complex) variable x. An ordinary differential equation

$$y^{(n)}+p_1(x)y^{(n-1)}+\ldots+p_n(x)y=q(x) \qquad (1)$$

containing an unknown function y and its derivatives $y', y'', \ldots, y^{(n)}$ of order up to n is called a **linear ordinary differential equation of the nth order**. In particular, a linear differential equation

$$y^{(n)}+p_1(x)y^{(n-1)}+\ldots+p_n(x)y=0 \qquad (1')$$

with $q(x)\equiv 0$ is said to be **homogeneous**. If $q(x)\neq 0$, (1) is said to be **inhomogeneous**. The existence and uniqueness theorems for solutions to the initial value problem are valid for (1) (→ 316 Ordinary Differential Equations (Initial Value Problems)). Moreover, the following theorem holds: Let D denote an interval in the real line or a domain in the complex plane. If the coefficients $p_k(x)$, $q(x)$ are continuous in an interval D, then every solution of (1) has D as its interval of definition. If $p_k(x)$, $q(x)$ are holomorphic in a domain D, then every solution of (1) is continued analytically along any path in D. Combining this theorem and the existence and uniqueness theorems we have the following theorem: Let $p_k(x)$, $q(x)$ be continuous in an interval D. Then for every point x_0 in D and every n-tuple of numbers $\eta, \eta', \ldots, \eta^{(n-1)}$, there exists one and only one solution $y(x)$ of (1) satisfying the initial conditions

$$y(x_0)=\eta,\ y'(x_0)=\eta', \ldots, y^{(n-1)}(x_0)=\eta^{(n-1)} \qquad (2)$$

and such that $y(x), y'(x), \ldots, y^{(n)}(x)$ are all continuous in D. If $p_k(x)$, $q(x)$ are holomorphic in D, then there exists one and only one solution $y(x)$ of (1) satisfying (2) and such that $y(x)$ is †complex analytic (not necessarily single-valued) in D.

It follows that every point of discontinuity (or singular point) of a solution of (1) is a point of discontinuity (or singular point) for at least one of the functions $p_k(x)$, $q(x)$ (→ 254 Linear Ordinary Differential Equations (Local Theory)).

B. Fundamental Systems of Solutions

The totality of solutions of a homogeneous linear ordinary differential equation forms a linear space (over the real or complex field). That is, any linear combination $y(x)=\sum_{i=1}^m C_i y_i(x)$ of the solutions $y_1, y_2, \ldots, y_m$ of (1′), where the C_i are arbitrary constants, is also a solution of (1′). This is called the **principle of superposition**. More than $n+1$ solutions of (1′) are always linearly dependent, that is, if $m\geqslant n+1$, we can find m constants C_1, $C_2, \ldots, C_m$, not all equal to zero, such that $\sum_{i=1}^m C_i y_i(x)=0$. Equation (1′) has n linearly independent solutions. For instance, the n solutions $y_1, y_2, \ldots, y_n$ defined by the initial conditions

$$y_1(x_0)=1, \quad y_1'(x_0)=0, \ldots, y_1^{(n-1)}(x_0)=0,$$

$$y_2(x_0)=0, \quad y_2'(x_0)=1, \ldots, y_2^{(n-1)}(x_0)=0,$$

$$\ldots$$

$$y_n(x_0)=0, \quad y_n'(x_0)=0, \ldots, y_n^{(n-1)}(x_0)=1 \qquad (3)$$

are linearly independent. Such a system of n

linearly independent solutions $y_1, \ldots, y_n$ of (1′) is called a **fundamental system of solutions** of (1′). In terms of a fundamental system of solutions $y_1, \ldots, y_n$, each solution y of (1′) is represented uniquely in the form $y(x)=\sum_{i=1}^{n} C_i y_i(x)$.

C. Liouville's Formula

In order for n solutions $y_1, y_2, \ldots, y_n$ of (1′) to be linearly independent, it is necessary and sufficient that the †Wronskian determinant $W(y_1, y_2, \ldots, y_n) \neq 0$ in D. Furthermore, the coefficients $p_k(x)$ can be represented in terms of an arbitrary fundamental system of solutions $y_1, y_2, \ldots, y_n$ since the coefficient of $y^{(n-k)}$ in the expansion of

$$\frac{(-1)^n W(y, y_1(x), y_2(x), \ldots, y_n(x))}{W(y_1(x), y_2(x), \ldots, y_n(x))} \tag{4}$$

is identically equal to $p_k(x)$ in D. Using this equality for $p_1(x)$, we obtain **Liouville's formula**:

$$W(y_1(x), \ldots, y_n(x)) = W(y_1(x_0), \ldots, y_n(x_0)) \exp\left(-\int_{x_0}^{x} p_1(t)\,dt\right). \tag{5}$$

D. Lagrange's Method of Variation of Constants

The difference of two solutions of the inhomogeneous equation (1) is a solution of the homogeneous equation (1′). Consequently, the †general solution of (1) can be represented as the sum of a †particular solution of (1) and the †general solution of (1′). Since a particular solution of (1) can be obtained from an arbitrary fundamental system of solutions $y_1, y_2, \ldots, y_n$ of (1′), (1) can be solved if a fundamental system of solutions for (1′) is known. In fact, if we consider $C_1, C_2, \ldots, C_n$ in the representation $y=\sum_{i=1}^{n} C_i y_i(x)$, not as constants, but as functions of x, and determine them by the conditions

$$y_1(x)C_1'(x)+y_2(x)C_2'(x)+\ldots+y_n(x)C_n'(x)=0,$$

$$y_1'(x)C_1'(x)+y_2'(x)C_2'(x)+\ldots+y_n'(x)C_n'(x)=0,$$

$$\ldots$$

$$y_1^{(n-1)}(x)C_1'(x)+y_2^{(n-1)}(x)C_2'(x)+\ldots + y_n^{(n-1)}(x)C_n'(x)=q(x), \tag{6}$$

then $y(x)=\sum_{i=1}^{n} C_i(x) y_i(x)$ is a solution of (1). This is always possible because from (6) we obtain

$$\frac{dC_i}{dx}=\frac{q(x)W_i(x)}{W(y_1(x), \ldots, y_n(x))}, \tag{7}$$

where $W_i(x)$ is the †cofactor of $y_i^{(n-1)}(x)$ in the determinant $W(y_1(x), \ldots, y_n(x))$. This method is called **Lagrange's method of variation of constants** (or **variation of parameters**).

E. Linear Ordinary Differential Equations with Constant Coefficients

A linear ordinary differential equation

$$y^{(n)}+a_1 y^{(n-1)}+\ldots+q_{n-1}y'+a_n y=0 \tag{8}$$

with constant coefficients a_i has $y=\exp rx$ as a solution if r is a root of the algebraic equation

$$f(r)=r^n+a_1 r^{n-1}+\ldots+a_{n-1}r+a_n=0, \tag{9}$$

called the **characteristic equation** of (8). Let $r_1, r_2, \ldots, r_m$ be the distinct roots of (9), and suppose that the root r_i has multiplicity μ_i ($i=1,2,\ldots,m$). Then the set of functions

$$e^{r_1 x}, xe^{r_1 x}, \ldots, x^{\mu_1-1}e^{r_1 x}; \ldots;$$

$$e^{r_m x}, xe^{r_m x}, \ldots, x^{\mu_m-1}e^{r_m x} \tag{10}$$

is a fundamental system of solutions of (8).

F. D'Alembert's Method of Reduction of Order

Let $y_1(x)$ be a solution, not identically equal to zero, of the homogeneous equation (1′). By substituting $y=y_1 z$ into (1′), we see that z' satisfies a linear differential equation of order $n-1$. This method is called **d'Alembert's method of reduction of order**. Since linear ordinary differential equations of the first order can be integrated by quadrature (→ Appendix A, Table 14.I), a homogeneous linear ordinary differential equation of the second order can be integrated completely if one solution of the equation that does not identically vanish is known.

So far we have outlined a general theory of solutions of (1) in the domain where solutions are continuous or complex analytic, but in order to have thorough knowledge of all the solutions, we have to examine their behavior also in the neighborhood of a singular point (which is a discontinuity point or a singular point for at least one of the coefficients) (→ 253 Linear Ordinary Differential Equations (Global Theory), 254 Linear Ordinary Differential Equations (Local Theory)). Also, boundary value problems are important as well as initial value problems described before, especially for second-order equations in connection with mathematical physics [7]. For these → 315 Ordinary Differential Equations (Boundary Value Problems), 390 Spectral Analysis of Operators.

G. Systems of Linear Ordinary Differential Equations of the First Order

Let the $f_{ij}(x)$ be known functions. A **system of linear differential equations of the first order**

$$dy_1/dx=f_{11}(x)y_1+\ldots+f_{1n}(x)y_n+g_1(x),$$

$$dy_2/dx=f_{21}(x)y_1+\ldots+f_{2n}(x)y_n+g_2(x),$$

$$\ldots$$

$$dy_n/dx=f_{n1}(x)y_1+\ldots+f_{nn}(x)y_n+g_n(x) \qquad (11)$$

with n unknowns $y_1, y_2, \ldots, y_n$ contains (1) as a special case, since (1) is transformed into (11) by setting $y=y_1, y'=y_2, \ldots, y^{(n-1)}=y_n$. A system (11) with $g_i(x)\equiv 0$, i.e., a system

$$dy_1/dx=f_{11}(x)y_1+\ldots+f_{1n}(x)y_n,$$

$$\ldots$$

$$dy_n/dx=f_{n1}(x)y_1+\ldots+f_{nn}(x)y_n \qquad (11')$$

is said to be **homogeneous**, while (11) is said to be **inhomogeneous**.

Suppose that the $f_{ij}(x)$, $g_i(x)$ are continuous in an interval D. Then the following theorem holds: To every point x_0 in D and every initial condition

$$y_1(x_0)=b_1,\ y_2(x_0)=b_2, \ldots, y_n(x_0)=b_n,$$

there corresponds one and only one solution that is continuous in D. If the $f_{ij}(x)$, $g_i(x)$ are holomorphic in a domain D, there corresponds one and only one solution that is complex analytic (not necessarily single-valued) in D.

It follows that every point of discontinuity (singular point) of a solution of (11) is a discontinuity (singular) point for at least one of the coefficients $f_{ij}(x)$, $g_i(x)$.

H. Fundamental Systems of Solutions

If $m\geqslant n+1$, m solutions $(y_{1i}, y_{2i}, \ldots, y_{ni})$ $(i=1,2,\ldots,m)$ of (11′) are linearly dependent, i.e., we can find constants $C_1, C_2, \ldots, C_m$, not all equal to zero, such that $\sum_{i=1}^{m} C_i y_{ki}(x)=0$ $(k=1,2,\ldots,n)$. System (11′) has n linearly independent solutions. To see this, we have only to choose the initial conditions so that

$$y_{ik}(x_0)=1, \quad i=k,$$
$$\quad\;\; =0, \quad i\neq k.$$

Such a system of n linearly independent solutions $(y_{1i}, y_{2i}, \ldots, y_{ni})$ $(i=1,2,\ldots,n)$ is called a **fundamental system of solutions** of (11′). In terms of this fundamental system, any solution $(y_1, \ldots, y_n)$ of (11′) is represented uniquely in the form $y_k(x)=\sum_{j=1}^{n} C_j y_{kj}(x)$ $(k=1,2,\ldots,n)$.

The linear independence of n solutions $(y_{11}, \ldots, y_{1n}), \ldots, (y_{n1}, \ldots, y_{nn})$ is equivalent to the condition that the determinant

$$\Delta(x)=\begin{vmatrix} y_{11}(x) & y_{12}(x) & \ldots & y_{1n}(x) \\ y_{21}(x) & y_{22}(x) & \ldots & y_{2n}(x) \\ & & \ldots & \\ y_{n1}(x) & y_{n2}(x) & \ldots & y_{nn}(x) \end{vmatrix}$$

does not vanish in D. Corresponding to Liouville's formula (5), we have

$$\Delta(x)=\Delta(x_0)\exp\left(\sum_{i=1}^{n}\int_{x_0}^{x} f_{ii}(t)\,dt\right).$$

I. Method of Variation of Constants

The general solution $(Y_1, Y_2, \ldots, Y_n)$ of the inhomogeneous equation (11) is given as the sum of the general solution $(y_1, y_2, \ldots, y_n)$ of (11′) and one particular solution $(Y_{10}, Y_{20}, \ldots, Y_{n0})$ of (11), i.e., in the form

$$(y_1+Y_{10}, y_2+Y_{20}, \ldots, y_n+Y_{n0}).$$

To obtain the particular solution $(Y_{10}, Y_{20}, \ldots, Y_{n0})$, we take any fundamental system of solutions

$$y_1=\varphi_{1k}(x),\ y_2=\varphi_{2k}(x), \ldots, y_n=\varphi_{nk}(x), \qquad k=1,2,\ldots,n,$$

of (11′) and consider the constants u_k in the linear combination

$$y_i=\sum_{k=1}^{n}\varphi_{ik}(x)u_k, \qquad i=1,2,\ldots,n, \qquad (12)$$

as functions of x. Substituting (12) into (11), we obtain a system of differential equations

$$\sum_{k=1}^{n}\varphi_{ik}(x)u_k'(x)=g_i(x), \qquad i=1,2,\ldots,n, \qquad (13)$$

with unknowns u_k. Since the y_i form a fundamental system, the determinant of the matrix with elements $\varphi_{ik}(x)$ does not vanish. Hence (13) can be solved in the form

$$u_k'(x)=G_k(x), \qquad k=1,2,\ldots,n,$$

and the $u_k(x)$ can be obtained by quadrature. Consequently, it follows that one particular solution can be given, in terms of the $u_k(x)$, in the form (12). This method is also called the **method of variation of constants**.

J. Systems of Linear Ordinary Differential Equations with Constant Coefficients

Suppose that the coefficients f_{ij} in (11′) are all constants. Then n columns of the matrix $\exp(Fx)$ form a system of fundamental solutions of (11′), where F denotes the matrix $[f_{ij}]$. Therefore the general solution has the form

$$y_j=\sum_{k=1}^{m}P_{jk}(x)e^{\lambda_k x}, \qquad j=1,2,\ldots,n.$$

Here $\lambda_1, \lambda_2, \ldots, \lambda_m$ are the distinct roots of the **characteristic equation**

$$\begin{vmatrix} f_{11}-\lambda & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22}-\lambda & \cdots & f_{2n} \\ & & \cdots & \\ f_{n1} & f_{n2} & \cdots & f_{nn}-\lambda \end{vmatrix} = 0,$$

and if λ_k has multiplicity e_k $(\sum_{j=1}^m e_j = n)$, $P_{jk}(x)$ is a polynomial of degree at most $e_k - 1$ which contains e_k arbitrary constants.

Suppose that the coefficients $f_{jk}(x)$ in (11') are all periodic functions having the same period ω. Then there exists a linear transformation $y_j = \sum q_{jk}(x) z_k$ in which the q_{jk} are periodic with period ω, such that the original equation is reduced to $dz_j/dx = \sum c_{jk} z_k$, where the c_{jk} are constants (**Floquet's theorem**). Hence if we can find such a linear transformation, we can integrate the original equation.

K. Adjoint Differential Equations

Consider a linear homogeneous ordinary differential equation

$$F(y) = p_0 y^{(n)} + p_1 y^{(n-1)} + \ldots + p_n y = 0.$$

Integration by parts of $\int \bar{z} F(y) dx$ gives

$$\bar{z} F(y) - y\overline{G(z)} = d[R(y,z)]/dx, \qquad (14)$$

where

$$G(z) = (-1)^n((\bar{p}_0 z)^{(n)} - (\bar{p}_1 z)^{(n-1)} + \ldots + (-1)^n \bar{p}_n z),$$

$$R(y,z) = \sum_{k=1}^{n} \sum_{h=0}^{k-1} (-1)^h y^{(k-h-1)} (p_{n-k} \bar{z})^{(h)}.$$

The equation $G(z) = 0$ is called the **adjoint differential equation** of $F(y) = 0$, and the identity (14) is called the **Lagrange identity**. By integrating (14), we have **Green's formula**

$$\int_{x_0}^{x_1} (\bar{z} F(y) - y\overline{G(z)}) dx = R[y,z](x_1) - R[y,z](x_0).$$

The adjoint differential equation of the adjoint equation of $F(y) = 0$ coincides with $F(y) = 0$. If y is a solution of $F(y) = 0$, then the solution z of the adjoint differential equation satisfies the $(n-1)$st-order differential equation $R(y,z) = \text{constant}$. When $G(y) = F(y)$, $F(y) = 0$ is called a **self-adjoint differential equation**. In the case of second-order equations with real coefficients, its general form is

$$F(y) = d(p\,dy/dx)/dx + qy = 0. \qquad (15)$$

For systems of differential equations, the **adjoint system** of (11') is defined by

$$dz_j/dx = -\bar{f}_{1j} z_1 - \bar{f}_{2j} z_2 - \ldots - \bar{f}_{nj} z_n, \qquad j = 1, 2, \ldots, n. \qquad (16)$$

Conversely, (11') is the adjoint system to (16). If $(y_1, y_2, \ldots, y_n)$, $(z_1, z_2, \ldots, z_n)$ are solutions of (11'), (16), respectively, then we have $\sum_{j=1}^n y_j \bar{z}_j = \text{constant}$. The system (11') is called a **self-adjoint system of differential equations** if (11') coincides with (16), i.e., if $f_{jk}(x) = -\bar{f}_{kj}(x)$.

L. Laplace and Euler Transforms

When the coefficients $p_i(x)$ $(i = 1, 2, \ldots, n)$ in (1') are rational functions, it often happens that we can find a solution in the form

$$y(x) = \int_a^b v(t) e^{xt} dt. \qquad (17)$$

Namely, we can often find a suitable function $v(t)$ so that the †Laplace transform (17) of $v(t)$ is a solution of (1'). Similarly, it is often possible to find a solution of (1') as the Euler transform

$$y(x) = \int_a^b v(t)(1-x)^{\rho-1} dt \qquad (18)$$

of some suitable $v(t)$. These transforms are used for the integral representation of special functions.

M. Linear Ordinary Differential Equations and Special Functions

A number of transcendental functions, such as †hypergeometric functions, †Bessel functions, †Legendre functions, etc., and Hermite polynomials, Laguerre polynomials, Jacobi polynomials, etc., are defined by linear ordinary differential equations of the second order (→ 389 Special Functions).

References

[1] E. A. Coddington and N. Levinson, Theory of ordinary differential equations, McGraw-Hill, 1955.
[2] K. O. Friedrichs, Lectures on advanced ordinary differential equations, Gordon & Breach, 1965.
[3] E. L. Ince, Ordinary differential equations, Dover, 1956.
[4] P. Hartman, Ordinary differential equations, Wiley, 1964.
[5] E. Hille, Lectures on ordinary differential equations, Addison-Wesley, 1968.
[6] M. A. Naĭmark (Neumark), Linear differential operators, I, II, Ungar, 1967, 1968. (Original in Russian, 1954.)
[7] K. Yosida, Lectures on differential and integral equations, Wiley, 1960. (Original in Japanese, 1950.)

253 (XIII.8) Linear Ordinary Differential Equations (Global Theory)

A. General Remarks

Let there be given a linear differential equation of the nth order

$$y^{(n)}+a_1(x)y^{(n-1)}+\ldots+a_n(x)y=0, \tag{1}$$

or a system of linear differential equations

$$\mathbf{y}'=A(x)\mathbf{y}, \tag{2}$$

which is the vector-matrix expression of

$$y_j'=\sum_{k=1}^{n} a_{jk}(x)y_k, \quad j=1,\ldots,n, \tag{2'}$$

where the coefficients $a_k(x)$, $a_{jk}(x)$ are complex analytic functions of x in a certain complex domain D. The solutions of (1) or (2) are known to be holomorphic when the coefficients are all holomorphic. However, at a singular point of at least one of the coefficients, a †branch point of the solution usually appears. Thus a solution of (1) or (2) is, in general, a †multiple-valued analytic function of x. The object of global theory is the function-theoretic study of this function—that is, determination of its †Riemann surface and investigation of its behavior on the Riemann surface.

At a †regular singular point of (1) or (2), any solution can be expressed explicitly by the combination of elementary functions and power series convergent within a circle around the singular point. In the presence of an †irregular singular point, instead of a convergent expression, we can construct an asymptotic expansion valid within a certain sector whose vertex is situated at the singular point (→ 254 Linear Ordinary Differential Equations (Local Theory)). Once such expressions have been obtained, the remaining task is to find the relations connecting those locally valid expressions. Therein lies the main and most difficult part of global theory. The problem of determining the relations, called the **connection formulas**, is the **connection problem**.

Equation (1) or the system (2) is said to be of **Fuchsian type** if all singular points of (1) or (2) are regular singular points. If (1) is an equation of Fuchsian type defined on the Riemann sphere and having singularities at $a_1, a_2, \ldots, a_m, a_{m+1}=\infty$, then we have the **Fuchs relation**

$$\sum_{j=1}^{m+1}\sum_{k=1}^{n}\rho_{jk}=(m-1)n(n-1)/2,$$

where $\rho_{jl},\ldots,\rho_{jn}$ are the †exponents at a_j of (1).

B. Monodromy Groups

Suppose that the coefficients of equation (2) are defined on a certain Riemann surface $\mathfrak{F}$ with singular points $a_1, a_2, \ldots$. By deleting a_1, $a_2, \ldots$ from $\mathfrak{F}$, another Riemann surface $\mathfrak{F}'$ is obtained. Choose a point x on $\mathfrak{F}'$, and let $Y(x)$ be any fixed branch of a †fundamental system of solutions of (2). Also, let Γ be a circuit on $\mathfrak{F}'$ starting from x. By an analytic continuation along Γ, another branch of $Y(x)$ is obtained, which we denote by $Y(x\Gamma)$. It is known that in this case these two branches are connected by the relation $Y(x\Gamma)=Y(x)C_\Gamma$, where C_Γ is an $n\times n$ constant matrix, and also that if Γ_1 and Γ_2 are †homotopic circuits, $C_{\Gamma_1}=C_{\Gamma_2}$. So the branch $Y(x\Gamma)$ and the matrix C_Γ are determined by the †homotopy class γ to which Γ belongs. Thus we can write $Y(x\gamma)$ or C_γ instead of $Y(x\Gamma)$ or C_Γ.

Now let G be the †fundamental group of $\mathfrak{F}'$. Since $Y(x\gamma_2\gamma_1)=Y(x\gamma_2)C_{\gamma_1}=Y(x)C_{\gamma_2}C_{\gamma_1}$ for any $\gamma_1, \gamma_2\in G$, the correspondence $\gamma\rightarrow C_\gamma$ defines a †representation of G. The group $g=\{C_\gamma \mid \gamma\in G\}$, which is naturally homomorphic to G, is called the **monodromy group** of equation (2).

For equation (1), we can also define the monodromy group of the equation by $g=\{C_\gamma \mid \gamma\in G\}$, where C_γ is a matrix such that $(y_1(x\gamma),\ldots,y_n(x\gamma))=(y_1(x),\ldots,y_n(x))C_\gamma$, where $y_1(x),\ldots,y_n(x)$ are linearly independent solutions of (1).

If the equation is of Fuchsian type, the global problem can be regarded as solved when the monodromy group of the equation has been completely determined.

If $\mathfrak{F}$ is a complex sphere and the equation is of Fuchsian type, the number of singular points of the coefficients is of course finite. Let $a_1,\ldots,a_m$ be those singular points, and γ_k be the homotopy class of $\mathfrak{F}'$ determined by a closed curve Γ_k that encloses only one singular point a_k. Then the monodromy group g is generated by the matrices $C_{\gamma_1},\ldots,C_{\gamma_m}$. Obviously $C_{\gamma_1},\ldots,C_{\gamma_m}$ are not necessarily independent. At least one relation $C_{\gamma_1},\ldots,C_{\gamma_m}=I$ (a unit matrix) always holds. In this case, †Jordan canonical forms of C_{γ_k} are all determined from the convergent expression for the fundamental system of solutions valid around a_k constructed by the famous †Frobenius method. However, the calculation of C_{γ_k} itself is generally impossible.

If $n=2$, $m=3$, and the coefficients are all rational functions of x, equation (1) or (2) is completely determined if we fix the roots of †indicial equations at every a_k, as long as the equation is of Fuchsian type. Therefore the monodromy group is determined by the values of the roots of indicial equations. Since these roots are calculated purely algebraically, the

monodromy group is determined by algebraic procedure in this case [5].

Let $n=2$, $m=3$ in equation (1). Denote by a, b, c the three singular points and by λ, λ'; μ, μ'; ν, ν' the roots of indicial equations at a, b, c, respectively. As was mentioned in the previous paragraph, the equation is determined uniquely by these nine quantities. Hence they also determine a family of functions consisting of all the solutions of the equation. This family is usually written as

$$P\begin{Bmatrix} a & b & c & \\ \lambda & \mu & \nu & x \\ \lambda' & \mu' & \nu' & \end{Bmatrix}$$

and is called the **P-function of Riemann** [2]. A simple transformation of variables reduces it to the totality of solutions of Gauss's †hypergeometric differential equation

$$x(1-x)y''+(\gamma-(\alpha+\beta+1)x)y'-\alpha\beta y=0. \tag{3}$$

Some solutions of (3) are expressed by a **hypergeometric integral**

$$\int_C t^{\alpha-\gamma}(t-1)^{\gamma-\beta-1}(t-x)^{-\alpha}dt,$$

where C is a suitably chosen path of integration [2] (→ 206 Hypergeometric Functions).

Calculation of the monodromy group of the equation (1) or (2) is still an unsolved problem except for the case $n=2$, $m=3$ and a few other particular cases.

C. Equations with an Irregular Singular Point

In the presence of an irregular singular point, complete knowledge of the monodromy group is still insufficient for the solution of the global problem. It is only the structure of the Riemann surface that is known from the monodromy group, and the behavior of the solution on the Riemann surface still remains to be studied. At an irregular singular point, the solution can be expressed only by an †asymptotic series valid within a certain sector, and the same solution possesses completely different expressions in different sectors. This is called †Stokes's phenomenon (→ 254 Linear Ordinary Differential Equations (Local Theory)). Thus, to complete the global theory, connection formulas between different asymptotic expressions must be established.

For a second-order linear equation with two singular points, one of which is regular and the other irregular of the first rank, the problem is completely solved. In this case, the equation can be reduced to a †confluent hypergeometric differential equation [2] (→ 167 Functions of Confluent Type). The problem is also partly solved for a linear equation of higher order with two singular points, one of which is regular and the other irregular (K. Okubo, *J. Math. Soc. Japan*, 1963).

If two singular points are both irregular, even the monodromy group cannot be calculated in general. For such a case, G. D. Birkhoff proposed a method of reducing one of the singular points to a regular one. He showed that this procedure is possible under certain assumptions on the monodromy matrix (*Math. Ann.*, 1913).

D. Riemann's Problem

As a noteworthy result for equation (1) of Fuchsian type with algebraic coefficients, Poincaré's theory deserves special mention. According to his theory, a solution of (1) can be uniformized in the form $y=f(z)$, $x=g(z)$, where f and g are single-valued analytic functions of z. Although it is known generally that any analytic function admits such uniformization (→ 367 Riemann Surfaces), Poincaré's theory gives a more explicit and efficient uniformizing construction. As uniformizing parameter z, we may take a ratio of two independent solutions of a certain linear differential equation of the second order that is determined from (1), and f and g are, in general, †Fuchsian functions, i.e., †automorphic functions for a certain †Fuchsian group, save for a few exceptional cases in which they are rational or †elliptic functions.

Brief mention should be made of **Riemann's problem** as a problem closely related to the global theory of linear differential equations. This problem was taken up by †Hilbert in his famous Paris lecture as the 21st problem, and hence is often called the Riemann-Hilbert problem. The problem can be stated as follows: Suppose that we are given a Riemann surface $\mathfrak{F}$, points $a_1, a_2, \ldots$ on $\mathfrak{F}$, and a group g of $n\times n$ matrices homomorphic to the fundamental group of $\mathfrak{F}-\{a_1,a_2,\ldots\}$. Then find an equation of the form (2) such that (i) the coefficient $A(x)$ is single-valued and meromorphic on $\mathfrak{F}$; (ii) the singular points are all regular and situated at $a_1, a_2, \ldots$; and (iii) the monodromy group of the equation coincides with g if a fundamental system of solutions is suitably chosen. Extensive research was done by many mathematicians, and finally H. Röhrl succeeded in solving the problem (*Math. Ann.*, 1957).

E. Isomonodromic Deformations

M. E. R. Fuchs considered the equation

$$d^2y/dx^2=p(x)y, \tag{4}$$

where $p(x)$ is given by

$$p(x)=\frac{a}{x^2}+\frac{b}{(x-1)^2}+\frac{c}{(x-t)^2}+\frac{d}{x(x-1)}+\frac{3}{4(x-\lambda)^2}+\frac{\alpha}{x(x-1)(x-t)}+\frac{\beta}{x(x-1)(x-\lambda)},$$

and he proposed the following problem: Obtain conditions among the parameters $a, b, c, d, t, \lambda, \alpha, \beta$ such that the monodromy group of (4) is kept invariant when these parameters vary, under the hypothesis that a fundamental system of solutions around $x=\lambda$ of (4) does not contain logarithmic terms. It is clear that a, b, c, d remain constant. Fuchs obtained a necessary and sufficient condition which is stated as follows: λ, α, β are considered as functions of t and λ satisfies the †Painlevé equation (VI) and α, β are rational functions of λ and $d\lambda/dt$ (*Math. Ann.*, 63 (1907)).

The result of Fuchs was extended by R. Garnier into two directions: Considering equations of the form (4) with irregular singular points, Garnier derived the Painlevé equations (I)–(V). Then from equation (4) with

$$p(x)=\frac{a}{x^2}+\frac{b}{(x-1)^2}+\sum_{j=1}^{N}\frac{c_j}{(x-t_j)^2}+\frac{d}{x(x-1)}+\sum_{j=1}^{N}\frac{3}{4(x-\lambda_j)^2}+\sum_{j=1}^{N}\frac{\alpha_j}{x(x-1)(x-t_j)}+\sum_{j=1}^{N}\frac{\beta_j}{x(x-1)(x-\lambda_j)},$$

he obtained a system of partial differential equations called the **Garnier system** (*Ann. Sci. Ecole Norm. Sup.*, 1912).

For equations with irregular singular points, the problem must be modified as follows: Deform equations so that not only the monodromy group but also the system of †Stokes multipliers are kept invariant. This modified problem is called the **isomonodromic deformation**.

L. Schlesinger studied the isomonodromic deformation of the system

$$\frac{dy}{dx}=\left(\sum_{j=1}^{N}\frac{A_j}{x-t_j}\right)y,$$

and he derived a †Pfaffian system, called the **Schlesinger equations** (*Crelles J.*, 1912).

The research of the theory of isomonodromic deformation has recently become active. K. Okamoto has investigated the isomonodromic deformation of Painlevé equations, Garnier systems, and linear ordinary differential equations in detail, and T. Miwa and M. Jimbo have extended the Schlesinger equations. Also → 288 Nonlinear Ordinary Differential Equations (Global Theory).

References

[1] E. A. Coddington and N. Levinson, Theory of ordinary differential equations, McGraw-Hill, 1955.
[2] E. Hille, Ordinary differential equations in the complex domain, Wiley, 1976.
[3] L. Schlesinger, Handbuch der Theorie der linearen Differentialgleichungen I, II_1, II_2, Teubner, 1895–1898.
[4] E. Hilb, Lineare Differentialgleichungen in Komplexen Gebiet, Enzykl. Math., Bd. II, 2, Heft 4 (1913), 471–562.
[5] C. E. Picard, Traité d'analyse III, Gauthier-Villars, 1896.
[6] A. Lappo-Danilevskii, Mémoires sur la théorie des systèmes des équations differentielles linéaires, Chelsea, 1953.
[7] N. I. Mushelishvili, Singular integral equations, Noordhoff, 1953. (Original in Russian, 1946.)
[8] J. Plemelj, Problems in the sense of Riemann and Klein, Wiley, 1964.

254 (XIII.7) Linear Ordinary Differential Equations (Local Theory)

A. Singular Points

Consider a system of n linear ordinary differential equations

$$\frac{d\mathbf{y}}{dx}=A(x)\mathbf{y}, \tag{1}$$

where the independent variable x belongs to a domain in the †Riemann sphere, $\mathbf{y}$ is a complex n-dimensional column vector $(y_1(x), y_2(x), \ldots, y_n(x))$, and the $n\times n$ matrix $A(x)$ has †complex analytic functions as elements. A singular point $x=a$ $(\neq\infty)$ of $A(x)$ is called a **singular point** of the system (1). Let

$$d\mathbf{y}/dt=B(t)\mathbf{y} \tag{2}$$

be the system transformed from (1) by the change of variable $t=1/x$. The point $x=\infty$ is a singular point of the system (1) if $t=0$ is a singular point of (2). By definition a point $x=a$ or $x=\infty$ is a singular point of (1) if $t=0$ is a singular point of the transformed equation (2) by the change of variable $t=x-a$ or $t=1/x$. It follows that by the use of †local coordinates the notion of singular points of the system can

be extended to the case when $A(x)$ is complex analytic on a †Riemann surface.

Consider a single nth-order differential equation defined on a Riemann surface

$$y^{(n)}+a_1(x)y^{(n-1)}+\ldots+a_n(x)y=0. \qquad (3)$$

The above definition of singular points may easily be adapted for use in equation (3), because (3) can be converted into a system (1) by a simple transformation.

B. Classification of Singular Points

Suppose that $A(x)$ is holomorphic in a deleted neighborhood $0<|x|<R$ of $x=0$. Then any solution of (1) can be continued analytically in $0<|x|<R$ and is not necessarily single-valued (→ 252 Linear Ordinary Differential Equations). For simplicity, we denote by $xe^{2\pi i}$ the terminal point of a closed path starting and ending at x and surrounding $x=0$ once in the positive sense. Let $\mathbf{y}_1(x), \ldots, \mathbf{y}_n(x)$ be a fundamental system of solutions of (1). Then the fundamental matrix solution $Y(x)=(\mathbf{y}_1(x), \ldots, \mathbf{y}_n(x))$ undergoes a linear transformation $Y(xe^{2\pi i})=Y(x)M$ when x is transformed to $xe^{2\pi i}$, where M is a nonsingular constant matrix. The matrix M is the **monodromy matrix** (or **circuit matrix**) of $Y(x)$ at $x=0$. If we take a matrix S such that $M=e^{2\pi i S}$, then there exists a matrix $P(x)$ whose elements are holomorphic functions in $0<|x|<R$ and such that $Y(x)$ has the form $Y(x)=P(x)x^S$, where x^S is defined by $x^S=\exp(S\log x)$.

If for a solution $\mathbf{y}(x)$ of (1) and an arbitrary sector Σ there is a positive number r such that $\lim_{x\to 0}|x|^r\mathbf{y}(x)=0$, $x=0$ is a **regular singular point** of the solution $\mathbf{y}(x)$; and if there is no such number, $x=0$ is an **irregular singular point of** $\mathbf{y}(x)$. If $x=0$ is a regular singular point of all the solutions of (1), it is a **regular singular point** of the system (1). If some of the solutions have $x=0$ as an irregular singular point, then it is an **irregular singular point** of the system (1). A necessary and sufficient condition for $x=0$ to be a regular singular point of the system is that an arbitrary fundamental matrix solution $Y(x)$ of (1) in the form $Y(x)=P(x)x^S$ as described above have $x=0$ as a pole of $P(x)$. In the same way, we can give the definition of regular singular points and irregular singular points for equation (3).

C. Regular Singular Points

Consider equation (3). A necessary and sufficient condition for $x=0$ to be a regular singular point of (3) is that every $a_k(x)$ have a pole of order at most k at $x=0$. Consequently, we can write (3) in the form

$$x^n y^{(n)}+A_1(x)x^{n-1}y^{(n-1)}+\ldots+A_n(x)y=0, \qquad (4)$$

where each $A_k(x)$ is holomorphic at $x=0$. Equation (4) has a fundamental system of solutions of the form $y=x^{\rho_k}P_k(x,\log x)$, where ρ_k are n roots of the **indicial equation** at $x=0$ of (4):

$$\rho(\rho-1)\ldots(\rho-n+1)+A_1(0)\rho(\rho-1)\ldots(\rho-n+2)+\ldots+A_{n-1}(0)\rho+A_n(0)=0,$$

and $P_k(x,L)$ is a polynomial in L of degree at most equal to the number of roots of the indicial equation that are congruent to ρ_k (modulo integers) with coefficients holomorphic functions of x. The quantities ρ_k are called the **exponents** at $x=0$ of (3). If the real part of ρ_k is largest among the real parts of the roots ρ_j that are congruent to ρ_k modulo integers, there exists a solution with the exponent ρ_k and not containing the logarithmic term. (A number a is congruent to b modulo integers when $a-b$ is an integer.) In particular, no solution contains the logarithmic terms when no pair of roots of the indicial equation is congruent modulo integers. Frobenius's method is convenient for finding these solutions (→ Appendix A, Table 14).

We return to system (1). The following theorem gives a simple sufficient condition for $x=0$ to be a regular singular point, but there is no simple necessary condition: $x=0$ is a regular singular point of (1) if $x=0$ is a simple pole of $A(x)$. Then (1) is written as

$$x(d\mathbf{y}/dx)=C(x)\mathbf{y}, \qquad (5)$$

where $C(x)$ is holomorphic at $x=0$. Equation (5) has a fundamental system of solutions of the form $\mathbf{y}_k=x^{\rho_k}\mathbf{p}_k(x,\log x)$, $k=1,\ldots,n$, where the exponents ρ_k are n roots of the indicial equation at $x=0$ of (5): $\det(C(0)-\rho I)=0$, and the $\mathbf{p}_k(x,L)$ are vector functions whose components are polynomials in L with coefficients holomorphic at $x=0$. If none of the exponent-differences $\rho_j-\rho_k$ $(j\neq k)$ is equal to an integer, then there is a fundamental system of solutions of the form $\mathbf{y}_k=x^{\rho_k}\mathbf{p}_k(x)$. The above result is derived from the following theorem: There exists a transformation $\mathbf{y}=P(x)\mathbf{z}$ that takes (5) into a system

$$x(d\mathbf{z}/dx)=D\mathbf{z},$$

where $P(x)$ is a matrix holomorphic at $x=0$ and D is a constant matrix.

If all solutions of (1) or (3) are meromorphic at $x=0$, then $x=0$ is called an **apparent singular point**.

D. Irregular Singular Points

If at least one of the coefficients $a_j(x)$ of equation (3) has a pole of order more than j, then $x=0$ is an irregular singular point of (3). Let m_j be the order of the pole of $a_j(x)$ at $x=0$, with the convention that $m_j=\infty$ when $a_j(x)=0$; and assume that $m_0=0$ for simplicity. Let A_ν ($\nu=0,1,\ldots,n$) be the points with coordinates (ν,m_ν) in the Euclidean plane and Π be the upper half of the boundary polygon of the †convex hull of the set $\{A_0,A_1,\ldots,A_n\}$. This polygon is called the **Newton diagram** of equation (3). Let (ν,r_ν) be the intersection of the straight line $x=\nu$ with Π, and consider the nonincreasing sequence of numbers $\{\sigma_j\}$, $\sigma_\nu=r_\nu-r_{\nu-1}$ ($r_0=0$). We assume μ to be an integer such that $\sigma_1\geqslant\sigma_2\geqslant\ldots\geqslant\sigma_\mu>1\geqslant\sigma_{\mu+1}\geqslant\ldots\geqslant\sigma_n$.

To the first μ sections of the diagram Π with slopes $\sigma_j>1$ there correspond μ formal solutions of the following form that are formally linearly independent:

$$y=\exp(\lambda_k(x))x^{\rho_k}P_k(x,\log x),$$

where the $\lambda_k(x)$ are polynomials in fractional powers of x^{-1} and the $P_k(x,L)$ are polynomials in L with coefficients given by formal series in fractional powers of x. Associated with the sections of Π corresponding to those $n-\mu$ numbers $\sigma_j\leqslant 1$, there are $n-\mu$ linearly independent formal solutions of similar form but without the exponential term.

Consider a system

$$x^{r+1}(d\mathbf{y}/dx)=A(x)\mathbf{y}, \qquad (6)$$

where r is a positive integer and $A(x)$ is holomorphic at $x=0$. The system (6) has n formally linearly independent solutions:

$$\mathbf{y}_k=\exp(\lambda_k(x))x^{\rho_k}\mathbf{p}_k(x,\log x),$$

where the $\lambda_k(x)$ are polynomials in fractional powers of x^{-1} and the $\mathbf{p}_k(x,L)$ are n-vectors whose components are polynomials in L with coefficients given by formal power series in fractional powers of x. This result is obtained from the following theorem: There exists a transformation $\mathbf{y}=P(x)\mathbf{z}$ that changes (6) formally into the system

$$x^{r+1}(d\mathbf{z}/dx)=(\Lambda(x)+Jx^r)\mathbf{z},$$

where $P(x)$ is a matrix with elements given by formal power series in fractional powers of x, $\Lambda(x)$ is a diagonal matrix with diagonal elements $x^{r+1}\lambda_k'(x)$, and J is a constant matrix in Jordan canonical form. A necessary and sufficient condition for $x=0$ to be a regular singular point of (6) is that $\Lambda(x)$ vanish identically.

Unfortunately, formal power series appearing in the formal solutions of (3) and (6) are in general divergent series. H. Poincaré, by introducing the notion of †asymptotic expansions, first proved under a very restrictive hypothesis that these formal solutions represent asymptotically actual solutions in any small sector. Contributions in this direction had been made, notably by W. J. Trjitzinski and J. Malmquist, but a decisive result was obtained by M. Hukuhara, whose method is also applicable to the study of regular singular points.

Let $y_1^1(x),\ldots,y_n^1(x)$ ($\Phi^1(x)$) be a fundamental system of solutions of (3) (fundamental matrix solution of (6)) that are expressed asymptotically by n formal solutions in a sector D_1 and $y_1^2(x),\ldots,y_n^2(x)$ ($\Phi^2(x)$) be another fundamental system of solutions of (3) (fundamental matrix solution of (6)) of the same nature expressed by the same formal solutions but in a different sector D_2. Then the two systems (two matrix solutions) are, in general, not the same system (matrix solution), but they are connected by a constant matrix C:

$$(y_1^1(x),\ldots,y_n^1(x))=(y_1^2(x),\ldots,y_n^2(x))C$$

$$(\Phi^1(x)=\Phi^2(x)C).$$

This is the so-called **Stokes phenomenon**, and the elements of the matrix C are called **Stokes multipliers**. The problem of determining these multipliers is a kind of †connection problem, and the above equality is a †connection formula (→ 253 Linear Ordinary Differential Equations (Global Theory)).

For an equation of the form (3) (system of the form (6)) with an irregular singular point at $x=0$, the polynomials $\lambda_1(x),\ldots,\lambda_n(x)$, the quantities $\rho_1,\ldots,\rho_n$, and a set of Stokes multipliers form a complete system of invariants under linear transformations of the form

$$z=q_1(x)y+q_2(x)y'+\ldots+q_n(x)y^{(n-1)}$$

$$(z=Q(x)y),$$

where the $q_j(x)$ (the elements of $Q(x)$) are meromorphic at $x=0$.

Even if equation (3) has $x=0$ as an irregular singular point, it may happen that (3) admits solutions holomorphic at $x=0$. This was first studied by O. Perron and then by F. Lettenmeyer, M. Hukuhara, and Iwano, and by H. Komatsu. G. D. Birkhoff proved that when the monodromy matrix at zero of the system (6) can be diagonalized, there is a nonsingular matrix $P(x)$ ($\det P(0)\neq 0$) such that the linear transformation $\mathbf{y}=P(x)\mathbf{z}$ transforms (6) into

$$x^r\frac{d\mathbf{z}}{dx}=\left(\sum_{k=0}^{r-1}B_kx^k\right)\mathbf{z}.$$

E. Singularities with Respect to a Parameter

An analogous theory has been obtained for a system of first-order linear differential equa-

tions with a small complex parameter ε:

$$\varepsilon^h \frac{dy}{dx} = A(x, \varepsilon)y,$$

$$A(x, \varepsilon) \cong \sum_{k=0}^{\infty} A_k(x)\varepsilon^k, \qquad (7)$$

where h is a positive integer, the $n \times n$ matrices $A_k(x)$ $(k = 0, 1, \ldots)$ are single-valued holomorphic functions of x in a neighborhood D of the origin, and the †asymptotic expansion is valid when $\varepsilon \to 0$ in a sector Σ. When all eigenvalues of the matrix $A_0(0)$ are distinct, there is a matrix of formal solutions of the form $P(x, \varepsilon)e^{Q(x, \varepsilon)}$, where $P(x, \varepsilon)$ is a formal series of the form $A(x, \varepsilon)$, $Q(x, \varepsilon)$ is a diagonal matrix with polynomials in $1/\varepsilon$ of degree h as diagonal elements, and all the coefficients are single-valued holomorphic functions of x on D^* $(\subset D)$. In particular, the coefficient of ε^{-h} in the jth diagonal element of $Q(x, \varepsilon)$ is $\int_0^x \mu_j(t)\,dt$, where $\mu_j(x)$ is an eigenvalue of the matrix $A_0(x)$. If we take a certain subsector Σ^* of Σ, the matrix $P(x, \varepsilon)e^{Q(x, \varepsilon)}$ represents an actual matrix solution in Σ^*. When there is no †turning point (→ Section F), it is always possible, even if there are multiple eigenvalues in $A_0(0)$, to construct formal solutions that are asymptotic representations of some solutions in some sector. Similar theories were developed for cases where more than two parameters appear.

F. Turning Points

Consider the point $x = 0$ in the system (7), and set $n = 2$, $h = 1$, and $A_0(x) = \begin{pmatrix} 0 & 1 \\ x & 0 \end{pmatrix}$. Then $A_0(x)$ has a multiple root when $x = 0$ and distinct roots when $x \neq 0$. As in this example, when the Jordan canonical form of the leading matrix $A_0(x)$ has different structure for $x = 0$ and for $x \neq 0$, the coefficients of the formal power series in ε are not single-valued holomorphic functions of x, and have worse singularities as the order becomes higher. Consequently, in the neighborhood of $x = 0$, we cannot construct actual solutions that can be represented asymptotically by these formal solutions. Such a point is called a **turning point** (or **transition point**) of the system (7).

If there is a nonsingular formal transformation $y = T(x, \varepsilon)z$; $T(x, \varepsilon) = \sum T_k(x)\varepsilon^k$ $(\det T_0(0) \neq 0)$ having similar analytic properties to those of $A(x, \varepsilon)$, and if the transformed system has a well-known form, it is possible to give analytic meaning to the formal transformation $T(x, \varepsilon)$. Then the transformed system is called a **related differential equation** of (7); in the above example, $\varepsilon(dz/dx) = A_0(x)z$ is a related differential equation that has well-known solutions expressed by †Bessel functions. What is meant by well-known here is that the behavior of all solutions is known in the entire complex plane for a fixed ε. However, it is not easy to find a suitable related differential equation for an arbitrarily given system.

References

[1] E. A. Coddington and N. Levinson, Theory of ordinary differential equations, McGraw-Hill, 1955.

[2] E. L. Ince, Ordinary differential equations, Dover, 1956.

[3] E. Hille, Ordinary differential equations in the complex domain, Wiley, 1976.

[4] L. Bieberbach, Theorie der gewöhnlichen Differentialgleichungen auf funktionentheoretischen Grundlage dargestellt, Springer, 1953.

[5] W. Wasow, Asymptotic expansions for ordinary differential equations, Wiley, 1965.

[6] A. Erdélyi, Asymptotic expansions, Dover, 1956.

[7] P. Hartman, Ordinary differential equations, Wiley, 1964.

255 (XIX.2) Linear Programming

A. Problems

A **linear programming problem** is a special type of mathematical programming (→ 264 Mathematical Programming) in which all the functions involved, i.e., the objective function and the constraints, are linear in the variables.

The simplest typical case is formulated as follows.

Problem I: Maximize $z = \mathbf{c}'\mathbf{x}$, under the condition

(C) $\quad A\mathbf{x} = b$ and $\mathbf{x} \geqslant \mathbf{0}$,

where $\mathbf{x} \in \mathbf{R}^n$ is the vector to be determined, $\mathbf{c}$ is a constant n-real vector, $\mathbf{b}$ is a constant m-vector, and A is an $m \times n$ real matrix.

Without any loss of generality we can assume that the rank of A is equal to m, because otherwise the condition is either redundant or inconsistent. This also implies that $m \leqslant n$.

Any vector which satisfies condition (C) is called a **feasible solution**, and that which maximizes z among all feasible solutions is called an **optimal solution**. Let A_1 be an $m \times m$ nonsingular submatrix of A, i.e., a matrix consisting of m linearly independent columns of the matrix A. Denote by A_2 the $m \times (n - m)$ matrix

formed by the remaining columns of A. Let $\mathbf{x}_1$ be the m-vector consisting of the components of $\mathbf{x}$ corresponding to A_1 and $\mathbf{x}_2$ be the vector of the remaining components. Then the equation $A\mathbf{x}=\mathbf{b}$ can be written as $A_1\mathbf{x}_1+A_2\mathbf{x}_2=\mathbf{b}$, and one of its solutions is given by $\mathbf{x}_1=A_1^{-1}\mathbf{b}$ and $\mathbf{x}_2=\mathbf{0}$, which is called a **basic solution**. Furthermore, if $\mathbf{x}_1\geqslant\mathbf{0}$, the basic solution is called a **basic feasible solution**, and if it is also optimal, it is a **basic optimal solution**. Then the following theorem can be proved.

Theorem: If there exists a feasible solution, there is also a basic feasible solution, and if there exists an optimal solution, there is also a basic optimal solution.

Corresponding to a basic solution, we can derive the expression $\mathbf{x}_1+A_1^{-1}A_2\mathbf{x}_2=A_1^{-1}\mathbf{b}$; $z-(\mathbf{c}_2'-\mathbf{c}_1'A_1^{-1}A_2)\mathbf{x}_2=0$. Such an expression is called a **basic form** of the problem, and the components of $\mathbf{x}_1$ are called the **basic variables**. The basic solution $\mathbf{x}_1=A_1^{-1}\mathbf{b}$, $\mathbf{x}_2=\mathbf{0}$, is an optimal solution if and only if

(Q) $A_1^{-1}\mathbf{b}\geqslant\mathbf{0}$ and $\mathbf{q}=\mathbf{c}_2-A_2'A_1^{-1\prime}\mathbf{c}_1\leqslant\mathbf{0}$,

which is called the optimality criterion.

Since there are at most ${}_nC_m$ basic solutions, we can always find an optimal solution, if there is one, in a finite number of steps.

Another form of the linear programming problem is as follows.

Problem II: Maximize $z=\mathbf{c}'\mathbf{x}$, under the condition

(C′) $A\mathbf{x}\leqslant\mathbf{b}$ and $\mathbf{x}\geqslant\mathbf{0}$.

The two formulations are equivalent because Problem II can be transformed into Problem I by introducing a new nonnegative m-vector $\mathbf{s}$ and writing the equation as $A\mathbf{x}+\mathbf{s}=\mathbf{b}$. Such a vector $\mathbf{s}$ is called the vector of **slack variables**. Conversely, Problem I can be formulated as Problem II by imposing the conditions $A\mathbf{x}\leqslant\mathbf{b}$ and $A\mathbf{x}\geqslant\mathbf{b}$ instead of the equality $A\mathbf{x}=\mathbf{b}$.

Sometimes it happens that for some of the variables nonnegativity conditions are not assumed. Then for the vector $\mathbf{x}^0$ of such variables we can define $\mathbf{x}^+-\mathbf{x}^-=\mathbf{x}^0$, $\mathbf{x}^+\geqslant0$, and $\mathbf{x}^-\geqslant0$, and we can assume that all the variables are nonnegative.

B. Duality

Problem II has the following **dual problem**.

Problem III: Minimize $w=\mathbf{b}'\mathbf{y}$, under the condition

(C″) $A'\mathbf{y}\geqslant\mathbf{c}$ and $\mathbf{y}\geqslant\mathbf{0}$.

For Problem I, if we restate the equality condition as $A\mathbf{x}\leqslant\mathbf{b}$ and $-A\mathbf{x}\leqslant-\mathbf{b}$, then the dual problem can be written as: Minimize $w=\mathbf{b}'\mathbf{y}^+-\mathbf{b}'\mathbf{y}^-$ under the condition $A'\mathbf{y}^+-A'\mathbf{y}^-\geqslant\mathbf{c}$; $\mathbf{y}^+\geqslant\mathbf{0}$, $\mathbf{y}^-\geqslant\mathbf{0}$. Then by putting $\mathbf{y}=\mathbf{y}^+-\mathbf{y}^-$, we can formulate the dual problem as follows.

Problem IV: Minimize $w=\mathbf{b}'\mathbf{y}$, under the condition

(C‴) $A'\mathbf{y}\geqslant\mathbf{c}$, with no restriction on the sign of $\mathbf{y}$.

In contrast to the dual problem, the original problem is called the **primary problem**. The following theorem is basic to the theory of linear programming.

Duality theorem: If either one of the primary or dual problems has an optimal solution, then the other also has an optimal solution, and it holds that $\max z=\min w$. Moreover, if x^* and y^* are optimal solutions of the two problems, we have $\mathbf{c}'\mathbf{x}^*=\mathbf{b}'\mathbf{y}^*=\mathbf{y}^{*\prime}A\mathbf{x}^*$.

Conversely, if $\mathbf{x}^*$ and $\mathbf{y}^*$ are feasible solutions of the primary and dual problems and if we have $\mathbf{c}'\mathbf{x}^*=\mathbf{b}'\mathbf{y}^*=\mathbf{y}^{*\prime}A\mathbf{x}^*$, then $\mathbf{x}^*$ and $\mathbf{y}^*$ are optimal solutions of the respective problems.

Let $\mathbf{x}_1^*=A_1^{-1}\mathbf{b}$, $\mathbf{x}_2^*=\mathbf{0}$ be a basic optimal solution of Problem I. Then the condition in its dual problem is written as $A_1'\mathbf{y}\geqslant\mathbf{c}_1$ and $A_2'\mathbf{y}\geqslant\mathbf{c}_2$. Put $\mathbf{y}^*=A_1^{-1\prime}\mathbf{c}_1$; then from the optimality condition it is shown that $A_2'\mathbf{y}^*\geqslant\mathbf{c}_2$, and also $A_1'\mathbf{y}^*=\mathbf{c}_1$; hence $\mathbf{y}=\mathbf{y}^*$ is a feasible solution of the dual problem. It is obvious that $\mathbf{c}'\mathbf{x}^*=\mathbf{c}_1'\mathbf{x}_1^*=\mathbf{c}_1A_1^{-1}\mathbf{b}=\mathbf{b}'\mathbf{y}^*=\mathbf{y}^{*\prime}A_1\mathbf{x}_1^*=\mathbf{y}^{*\prime}A\mathbf{x}^*$, and for any feasible solution y of the dual problem, we have $\mathbf{b}'\mathbf{y}=\mathbf{x}^{*\prime}A_1'\mathbf{y}\geqslant\mathbf{x}^{*\prime}\mathbf{c}=\mathbf{b}'\mathbf{y}^*$, which establishes that $\mathbf{y}^*$ is an optimal solution.

The duality theorem can be formulated in a more general way: Let V and W be closed convex cones in $\mathbf{R}^m$ and $\mathbf{R}^n$, respectively. Then we state the following problems.

Primary problem: Maximize $z=\mathbf{c}'\mathbf{x}$, $\mathbf{x}\in\mathbf{R}^n$, under the condition $\mathbf{b}-A\mathbf{x}\in V$ and $\mathbf{x}\in W$.

Dual problem: Minimize $w=\mathbf{b}'\mathbf{y}$, $\mathbf{y}\in\mathbf{R}^m$, under the condition $A'\mathbf{y}-\mathbf{c}\in W^*$ and $\mathbf{y}\in V^*$, where W^* and V^* are the dual cones of W and V, respectively.

We can also write $\max z=\infty$ if the value of z is not bounded, and $\max z=-\infty$ if there is no $\mathbf{x}$ which satisfies the condition; similarly, $\min w=-\infty$ if w is not bounded from below, and $\min w=\infty$ if there is no $\mathbf{y}$ satisfying the condition.

Theorem: If either $\max z\neq-\infty$ or $\min w\neq\infty$, we have $\max z=\min w$. Moreover, if $-\infty<\max z=\min w=\mathbf{c}'\mathbf{x}^*=\mathbf{b}'\mathbf{y}^*<\infty$, we have $\mathbf{y}^{*\prime}A\mathbf{x}^*=\mathbf{c}'\mathbf{x}^*=\mathbf{b}'\mathbf{y}^*$.

Define the expression $\varphi(\mathbf{x},\mathbf{y})=\mathbf{c}'\mathbf{x}+\mathbf{b}'\mathbf{y}-\mathbf{y}'A\mathbf{x}=\mathbf{c}'\mathbf{x}+\mathbf{y}'(\mathbf{b}-A\mathbf{x})=\mathbf{b}'\mathbf{y}-\mathbf{x}'(A'\mathbf{y}-\mathbf{c})$ for $\mathbf{x}\in W$ and $\mathbf{y}\in V^*$. φ can be regarded as a Lagrangian form for both the primary and dual problems. And if $(\mathbf{x}^*,\mathbf{y}^*)$ is a pair of op-

timal solutions, we have $\varphi(\mathbf{x}^*, \mathbf{y}) \geqslant \varphi(\mathbf{x}^*, \mathbf{y}^*) \geqslant \varphi(\mathbf{x}, \mathbf{y}^*)$ for all $\mathbf{x} \in W$ and $\mathbf{y} \in V^*$, which means that $(\mathbf{x}^*, \mathbf{y}^*)$ is a **saddle point** of the function φ. Now suppose that in the primary problem the constraint vector $\mathbf{b}$ can change, and consider $\max z$ to be a function of $\mathbf{b}$, denoted by $z(\mathbf{b})$. Then under a small change of $\mathbf{b}$ at least some of the optimal solutions of the dual problem will remain optimal; hence for any vector $\mathbf{a}$ we have

$$\inf_{\mathbf{y}^* \in Y^*} \mathbf{a}'\mathbf{y}^* \leqslant \lim \frac{1}{t}(y(\mathbf{b} + t\mathbf{a}) - z(\mathbf{b})) \leqslant \sup_{\mathbf{y}^* \in Y^*} \mathbf{a}'\mathbf{y}^*,$$

where Y^* is the set of the optimal solutions of the dual problem. Because of this property the solution of the dual problem is called the vector of **shadow prices** or of the **imputed costs** of the constraints. Symmetrically, the solution of the primary problem describes the rate of change in the value of the objective function under a small change in the constraint vector in the dual problem.

The duality theorem is closely related to some theorems on systems of linear inequalities and convex cones in a finite-dimensional Euclidean space; especially, the following are equivalent to or easily derivable from the duality theorem. **Minkowski-Farkas theorem**: Given an equation $A\mathbf{x} = \mathbf{b}$, where $\mathbf{b}$ is an element of $\mathbf{R}^n$, a necessary and sufficient condition for a solution $\mathbf{x} \geqslant 0$ to exist is that $\mathbf{u}'\mathbf{b} \geqslant 0$ hold for any vector $\mathbf{u}$ such that $\mathbf{u}'A \geqslant 0$. **Stiemke theorem**: For a matrix A one of the following two alternatives holds: (i) $A\mathbf{x} = 0$, $\mathbf{x} > 0$ have a solution; (ii) $\mathbf{u}'A \geqslant 0$ has a solution. **Tucker's theorem on complementary slackness**: For any matrix A, the two systems of linear inequalities (i) $A\mathbf{x} = 0$, $\mathbf{x} \geqslant 0$, and (ii) $\mathbf{u}'A \geqslant 0$ have solutions $\mathbf{x}$, $\mathbf{u}$ satisfying $A'\mathbf{u} + \mathbf{x} > 0$. The minimax theorem for †zero-sum two-person games (→ 173 Games Theory C) with finite number of strategies for both players is also shown to be equivalent to the duality theorem. It is formulated as

$$\min_{\mathbf{y} \in S_2} \max_{\mathbf{x} \in S_1} \mathbf{y}'M\mathbf{x} = \max_{\mathbf{x} \in S_1} \min_{\mathbf{y} \in S_2} \mathbf{y}'M\mathbf{x},$$

where M is the payoff matrix and S_1 and S_2 are †simplexes of mixed strategies.

C. Algorithms

The most commonly used method for solving linear programming problems numerically is the **simplex method** introduced by van Dantzig [1] and its variations. It gives a procedure starting from one basic feasible solution to reach an optimal basic solution in a finite number of steps by improving the value of the objective function at each step. Let

$$x_i + \sum_{j \in J} d_{ij} x_j = g_i, \quad i \in I,$$

$$z + \sum_{j \in J} f_j x_j = v$$

be a basic form for Problem I, where I denotes the set of the basic variables and J the set of nonbasic variables. Then the feasibility implies that $d_i \geqslant 0$ for all $i \in I$. Furthermore, if $f_j \geqslant 0$ for all $j \in J$, the basic solution $x_i = g_i$ for i in I and $x_j = 0$ for j in J is optimal. If $f_{j^*} < 0$ for some j^* in J, define $r_i = g_i / d_{ij^*}$ for i in I and for which $d_{ij} > 0$. Let $r_{i^*} = \min r_i$; then we can delete i^* from the set I and add j^* to it and get a new basis, and by simple algebraic calculation we get a new basic form corresponding to the new basic feasible solution, for which the value of z is increased by $-f_{j^*} r_{i^*}$. If r_{i^*} is always positive, we can get an optimal basic solution, since the above procedure cannot continue indefinitely; and even when r_i becomes zero (the degenerate case), we can avoid infinite circular repetition by using a method proposed by R. G. Bland [16]. And $f_{j^*} < 0$ and $d_{ij} \leqslant 0$ for all i implies that the value of z is unbounded. The array of the coefficients of the basic form is called the **simplex tableau**, and the method is called the simplex method. In order to obtain a basic feasible solution, the **two-phase simplex method** is used. Assume that in Problem I the vector $\mathbf{b}$ is nonnegative. Then we introduce a new m-vector $\mathbf{u}$ called the vector of **artificial variables**, and formulate an additional problem as follows.

Problem Ia: Maximize $t = -\mathbf{1}'\mathbf{u}$ under the condition $A\mathbf{x} + \mathbf{u} = \mathbf{b}$ and $\mathbf{x} \geqslant \mathbf{0}$, $\mathbf{u} \geqslant \mathbf{0}$.

This problem can be solved by the simplex method starting from the basic solution $\mathbf{u} = \mathbf{b}$, $\mathbf{x} = \mathbf{0}$, and if we get to an optimal solution with $t = 0$, we have a feasible solution for the original problem, from whence we can proceed with the original problem.

If there some inequalities in the condition we transform them to equalities by introducing slack variables, and then apply the simplex procedure. Once an optimal basic solution is obtained, a solution of the dual problem is easily obtained from the relation $\mathbf{y}^* = A_1^{-1}\mathbf{b}$. The dual simplex method utilizes the primary-dual relationship.

Recently, L. G. Khachiyan [7] proposed a new linear-programming method that gives an optimal solution within predetermined accuracy of approximation in a number of steps bounded by a polynomial in the numbers of the variables and constraints, a property the simplex algorithm fails to have. His method is basically an iterative procedure to find a solution $\mathbf{x}$ satisfying a system of strict inequalities

$\mathbf{a}_i'\mathbf{x} < b_i$, $i=1,\ldots,m$, where the b_i and the components of the $\mathbf{a}_i$ are all integers. Let $\mathbf{x}^{(0)}$ be any vector and $B^{(0)}=KI$, where K is a positive constant defined in terms of the $\mathbf{a}_i$ and the b_i, and the sequence of vectors $\mathbf{x}^{(k)}$ and matrices $B^{(k)}$ are defined by

$$\mathbf{x}^{(k+1)}=\mathbf{x}^{(k)}-\frac{1}{n+1}\frac{B^{(k)}\mathbf{a}_i}{\sqrt{\mathbf{a}_i'B^{(k)}\mathbf{a}_i}},$$

$$B^{(k+1)}=\frac{n^2}{n^2-1}\left(B^{(k)}-\frac{2}{n+1}\frac{B^{(k)}\mathbf{a}_i\mathbf{a}_i'B^{(k)}}{\mathbf{a}_i'B^{(k)}\mathbf{a}_i}\right),$$

when the vector $\mathbf{x}^{(k)}$ violates the inequality $\mathbf{a}_i'\mathbf{x} < b_i$, and continue until a solution is obtained or the number of steps reaches some constant. In the latter case it can be proved that the system of inequalities does not have any feasible solution. This algorithm can be applied to solve Problem II in the following way: Due to the duality theorem, the optimal solution of Problem II can be characterized as vectors satisfying $\mathbf{b}'\mathbf{y} \leqslant \mathbf{c}'\mathbf{x}$, $A\mathbf{x} \leqslant \mathbf{b}$, $-A'\mathbf{y} \leqslant -\mathbf{c}$, $\mathbf{x} \geqslant 0$, $\mathbf{y} \geqslant 0$. Then the system is approximated by another system with integer coefficients, and then by a system of strict inequalities, and the above algorithm can be applied to this approximated system. This method is based on a principle entirely different from the simplex method of computing successively the centers of ellipsoids of indefinitely decreasing size and containing a subset of feasible solutions.

Although Khachiyan's method has a mathematically appealing property, in practical applications the simplex method and its variations still seem to be the most efficient general method for the numerical solution of linear programming problems.

Some special types of linear programming problems allow specific algorithm for solution, e.g., the **transportation problem** has the structure: Minimize $\sum_i\sum_j c_{ij}x_{ij}$ under the condition $\sum_j x_{ij} \geqslant a_i$, $\sum_i x_{ij} \leqslant b_j$, $x_{ij} \geqslant 0$. This can be solved by any one of several simple intuitive methods. Various types of linear programming problems are discussed as problems of maximizing flows on networks (→ 281 Network Flow Problems). Also, problems with the further condition that some or all of the variables are integers can profitably be discussed separately (→ 215 Integer Programming).

D. Generalizations and Applications

Linear programming in the sequence space $(l)=\{X=\{x_j\} \mid \sum_{j=1}^{\infty}|x_j|<\infty\}$ was treated by P. C. Rosenbloom [12]. In this case, the requirements for variables X are given in terms of †linear functionals $\lambda_i \in (l)^*=(m)$ $(i=1,2,\ldots,k)$, for example, as (i) equalities $\lambda_i(X)=c_i$ or (i′) inequalities $\lambda_i(X) \leqslant c_i$, and (ii) $x_j \geqslant 0$ $(\forall j)$. The space (l) is supplied with the †weak topology as the conjugate space of $(c_0)=\{X=\{x_j\} \mid \lim_{j\to\infty}x_j=0\}$. Let $\mathfrak{F}$ be the set of all X satisfying (i) (or (i′)) and (ii). Let $\lambda(X)$ be a †weakly upper semicontinuous functional, and suppose that the $\lambda_i(x)$ are †weakly lower semicontinuous, $\mathfrak{F} \neq \varnothing$, and $\lambda(X)$ is upper bounded. Then the maximal value of $\lambda(X)$ is attained at an extreme point of $\mathfrak{F}$, and furthermore, if the space is completely regular, then the solution is unique. If $\mathfrak{F}$ is bounded, it is the **convex hull** of its extreme points, i.e., the smallest closed convex set containing its extreme points. By applying the theory in this section to the family of functions that can be expressed as $f(x)=\sum_{j=1}^{\infty}a_j\varphi_j(x)$ in terms of a given system of functions $\varphi_j(x)$ $(j=1,2,\ldots)$ on $\mathbf{R}$, S. N. Bernshteĭn's approximation theory of function systems, the theory of absolutely monotonic functions, and several inequalities in the theory of functions of a complex variable can be treated in a unified fashion. If the theory is further extended to the case of †finitely additive measures defined by means of linear functionals on the Banach space of bounded functions, we may treat the extremal problems of linear functionals on the function space of all $f(x)$ that can be expressed as $f(x)=\int_S K(x,s)\,d\mu(s)$, and apply it to obtain the interpolation formula of nonnegative †harmonic functions, results of Carathéodory and Fejér on its Fourier coefficients, an analog of Harnack's theorem for the †heat conduction equation, and so on.

The extension of linear programming theory to linear topological spaces is due to L. Hurwicz [5]. Let $\mathscr{X}$ be a linear space, $\mathscr{Y}$, $\mathscr{Z}$ be linear topological spaces, P_Y, P_Z the nonnegativity cones of $\mathscr{Y}$, $\mathscr{Z}$, respectively, which are closed convex cones containing inner points, and D a convex set in $\mathscr{X}$. We assume that F and A are linear mappings from D into $\mathscr{Y}$ and $\mathscr{Z}$, $\mathscr{Y}=\mathbf{R}$, $\mathscr{Z}$ is †locally convex, and P_X, P_Z are closed convex cones in $\mathscr{X}$ and $\mathscr{Z}$, respectively, and we consider the following.

Problem L: Maximize $F(X)=X^*(X)$ under the condition $G(X)=A(X)-B \geqslant 0$, $X \geqslant 0$, $B \in \mathscr{Z}$.

Put $\Phi(X,Z^*)=X^*(X)+Z^*(A(X)-B)$. If we define a linear mapping T from $\mathscr{W}=\mathscr{X}\times\mathbf{R}$ into $\mathscr{V}=\mathscr{Z}\times\mathscr{W}$ by $T((X,\rho))=(A(X)-\rho B,(X,\rho))$, where $\rho \in \mathbf{R}$, then under the condition that the image under T^* of the nonnegativity cone of $\mathscr{V}^*$ is a †regularly convex set in $\mathscr{W}^*$, X_0 is a solution of problem L if and only if there exists a $Z_0^* \in \mathscr{Z}^*$ such that (X_0,Z_0^*) is a nonnegative saddle point of Φ.

Under conditions expressed in simpler terms, Isii [6] proved the coincidence of the

supremum of the objective function and $\inf_{Z^*}\sup_{X\in D}$ of the †Lagrangian form, gave conditions for the supremum and the infimum to be attained, and developed a theory that generalizes the †Chebyshev inequality and the †Neyman-Pearson fundamental lemma for testing statistical hypotheses.

E. History

The theory of linear programming is closely related to the theory of convex cones, convex polyhedra, and systems of linear inequalities. Concerning †polyhedral convex cones, †convex polyhedra in $\mathbf{R}^n$, and the algebraic theory of systems of linear inequalities, we have classical results due to P. Gordon (1873), J. Farkas [2], E. Stiemke [13], H. Weyl [15], etc., and a later refinement due to A. W. Tucker [10, paper 1].

The application of linear systems to economics was made possible through the works of J. von Neumann, especially his †game theory and balanced linear growth model [14]. These and the interindustrial input-output analysis of W. Leontief [11] led to the works assembled in 1951 by T. C. Koopmans and others [9]. Concerning practical computation and applications to industry, there are isolated and long-neglected works by L. V. Kantorovich [8]; however, the main part of the method was developed in the United States, especially after the discovery of the simplex method by G. B. Dantzig [1] and his followers.

Linear programming has become one of the most important techniques in operations research, and, following upon the development of computers, has found wide application to practical problems. Most contemporary large-scale computers are equipped with programs to solve linear systems problems.

References

[1] G. B. Dantzig, Linear programming and extension, Princeton Univ. Press, 1963.
[2] J. Farkas, Über die Theorie der einfachen Ungleichungen, J. Reine Angew. Math., 124 (1902), 1–27.
[3] S. I. Gass, Linear programming: Methods and applications, McGraw-Hill, second edition, 1964.
[4] A. Haar, Über linear Ungleichungen, Acta Sci. Math. Szeged., 2 (1924), 1–14.
[5] L. Hurwicz, Programming in linear spaces, Studies in Linear and Non-Linear Programming, K. J. Arrow, L. Hurwicz, and H. Uzawa (eds.), Stanford Univ. Press, 1958, 38–102.
[6] K. Isii, Inequalities of the types of Chebyshev and Cramér-Rao and mathematical programming, Ann. Inst. Stat. Math., 16 (1964), 277–293.
[7] L. G. Khachiyan, A polynomial algorithm in linear programming, Soviet Math. Dokl., 20 (1979), 191–194. (Original in Russian, 1979.)
[8] L. V. Kantorovich, Mathematical methods of organization and planning production, Management Sci., 6 (1960), 366–422. (Original in Russian, 1904.)
[9] T. C. Koopmans (ed.), Activity analysis of production and allocation, Wiley, 1951.
[10] H. W. Kuhn and A. W. Tucker (eds.), Linear inequalities and related systems, Ann. Math. Studies, Princeton Univ. Press, 1956.
[11] W. Leontief, The structure of American economy 1919–1939, Oxford Univ. Press, second edition, 1951.
[12] P. C. Rosenbloom, Quelques classes de problèmes extrémaux, Bull. Soc. Math. France, 79 (1951), 1–58; 80 (1952), 183–215.
[13] E. Stiemke, Über positive Lösungen homogener linearer Gleichungen, Math. Ann., 76 (1914–1915), 340–342.
[14] J. von Neumann, Über ein ökonomisches Gleichungssystem und eine Verallgemeinerung des Brouwerschen Fixpunktsatzes, Ergebnisse eines Math. Kolloquiums, 8 (1937), 73–83; English translation, A model of general economic equilibrium, Rev. Economic Studies, 13 (1945–1946), 1–9 (Collected works, Pergamon, 1963, vol. 6, 29–37).
[15] H. Weyl, Elementare Theorie der konvexen Polyeder, Comment. Math. Helv., 7 (1934–1935), 290–306.
[16] R. G. Bland, New finite pivoting rules for the simplex method, Core discussion papers 7612, Center for Operations Research and Economics, Univ. Catholique de Louvain, 1976.

256 (III.8) Linear Spaces

A. Definition

Suppose that we are given a set L and a †field K satisfying the following two requirements: (i) Given an arbitrary pair (a, b) of elements in L, there exists a unique element $a+b$ (called the **sum** of a, b) in L; (ii) given an arbitrary element α in K and an arbitrary element a in L, there exists a unique element αa (called the **scalar multiple** of a by α) in L. The set L is called a **linear space over** K (or **vector space over** K) if the following eight conditions are satisfied: (i) $(a+b)+c=a+(b+c)$; (ii) there exists an ele-

ment $0 \in L$, called the **zero element** of L, such that $a+0=0+a=a$ for all $a \in L$; (iii) For any $a \in L$, there exists an element $x=-a \in L$ satisfying $a+x=x+a=0$; (iv) $a+b=b+a$; (v) $\alpha(a+b)=\alpha a+\alpha b$; (vi) $(\alpha\beta)a=\alpha(\beta a)$; (vii) $(\alpha+\beta)a=\alpha a+\beta a$; (viii) $1a=a$ (where 1 is the †unity element of K). An element of K is called a **scalar**, and an element of L is called a **vector**. K is called the **field of scalars (basic field** or **ground field)** of the linear space L.

In the definition of linear spaces, K can be noncommutative. L is also called a **left linear space** over K since the scalars act on L from the left ($a \rightarrow \alpha a, a \in L, \alpha \in K$). A **right linear space** is similarly defined. Actually, a left (right) linear space is a unitary left (right) K-module. If K is commutative, it is not necessary to specify left or right, since we can identify αa and $a\alpha$. In this article we consider only linear spaces over commutative fields. A similar theory can be established for linear spaces over noncommutative fields (→ 277 Modules).

If K is the field of real numbers $\mathbf{R}$ or the field of complex numbers $\mathbf{C}$, a linear space over K is called a **real linear space** or **complex linear space**, respectively. In the following discussion, we fix a field K, and by a linear space we mean a linear space over K.

Examples. (1) Geometric vectors: In a †Euclidean space or, more generally, an †affine space, the set of vectors $\overrightarrow{PQ}$ associated with points P, Q in the space forms a linear space.

(2) n-tuples in K: K^n denotes the set of all sequences $(\alpha_1, \ldots, \alpha_n)$ of n elements in a field K. Defining two operations by $(\alpha_1, \ldots, \alpha_n)+(\beta_1, \ldots, \beta_n)=(\alpha_1+\beta_1, \ldots, \alpha_n+\beta_n)$, $\lambda(\alpha_1, \ldots, \alpha_n)=(\lambda\alpha_1, \ldots, \lambda\alpha_n)(\lambda \in K)$, the set K^n forms a linear space over K. An element of K^n is called an **n-tuple** in K, and α_i is called the ith **component** of $(\alpha_1, \ldots, \alpha_n)$. In general, the **inner product** of $a=(\alpha_1, \ldots, \alpha_n)$, $b=(\beta_1, \ldots, \beta_n)$ is defined by $(a, b)=\sum_{i=1}^n \alpha_i\beta_i$. However, when $K=\mathbf{C}$ (complex number field), we usually define (a, b) to be $\sum_{i=1}^n \alpha_i\bar{\beta}_i$.

(3) Sequences in K: All infinite sequences in a field K form a linear space over K under the operations defined in example (2).

(4) K-valued functions: Given a nonempty set I and a field K, let K^I be the set of all K-valued functions defined on I. Defining two operations by $(f+g)(x)=f(x)+g(x)$, $(\lambda f)(x)=\lambda f(x)$ $(x \in I, \lambda \in K)$, the set K^I forms a linear space over K. In particular, if we put $I=\{1, \ldots, n\}$, then the space K^I can be identified with the space of n-tuples given in example (2), and if we put $I=\mathbf{N}$ (natural numbers), then we obtain the space given in example (3). Let K be the field $\mathbf{R}$ of real numbers and I an interval in $\mathbf{R}$. The set $C(I)$ of all continuous functions on I, the set $D(I)$ of all differentiable functions on I, and the set $A(I)$ of all †real analytic functions on I are all linear spaces contained in the space $\mathbf{R}^I$.

(5) Polynomials in K: $K[X_1, \ldots, X_n]$ denotes the set of all polynomials of n variables with coefficients in a field K. This forms a linear space under the usual operations.

B. Linear Mappings

Let L, M be linear spaces over a field K. A mapping φ from L to M is called a **linear mapping** or **linear operator** if φ satisfies the following two conditions; (i) $\varphi(a+b)=\varphi(a)+\varphi(b)$; and (ii) $\varphi(\lambda a)=\lambda\varphi(a)$ $(a, b \in L, \lambda \in K)$. Namely, a linear mapping is a K-homomorphism between K-modules (→ 277 Modules). Regarding K as a linear space, a linear mapping $L \rightarrow K$ is called a **linear form**. A linear mapping from L to L is called a **linear transformation** of L. The identity mapping of L is a linear transformation. Given linear spaces L, M, N, and linear mappings $\varphi: L \rightarrow M$ and $\psi: M \rightarrow N$, the composite $\psi \circ \varphi: L \rightarrow N$ is also a linear mapping. If a linear mapping $\varphi: L \rightarrow M$ is †bijective, then the inverse mapping $\varphi^{-1}: M \rightarrow L$ is also a linear mapping. Such a mapping φ is called an **isomorphism**, and we write $L \cong M$ if there exists an isomorphism $L \rightarrow M$. A linear transformation $L \rightarrow L$ is called **regular** (or **nonsingular**) if it is an isomorphism.

Examples. (1) Let L be the linear space formed by all geometric vectors in a Euclidean space (affine space) E. Then a †motion (†affine transformation) of E induces a linear transformation $L \rightarrow L$.

(2) Let (α_{ij}) be an $m \times n$ †matrix in K. Assigning $(\eta_1, \ldots, \eta_m)$ to $(\xi_1, \ldots, \xi_n)$, where $\eta_i=\sum_{j=1}^n \alpha_{ij}\xi_j$ $(1 \leqslant i \leqslant m)$, we have a linear mapping $K^n \rightarrow K^m$.

(3) Assigning the †derivative f' to a real-valued differentiable function f on an interval I, we have a linear mapping $D(I) \rightarrow \mathbf{R}^I$.

C. Linear Combinations

Let L be a linear space over a field K. An element of L of the form $\alpha_1 a_1+\ldots+\alpha_n a_n$ $(\alpha_i \in K, a_i \in L)$ is called a **linear combination** of $a_1, \ldots, a_n$. A sequence $a_1, \ldots, a_n$ of elements in L is called **linearly dependent** if there exists a sequence $\alpha_1, \ldots, \alpha_n$ of elements in K such that not all the α_i are equal to 0 and $\alpha_1 a_1+\ldots+\alpha_n a_n=0$. A sequence of elements in L is called **linearly independent** if it is not linearly dependent. Suppose that there exists a linearly independent sequence of n elements in a linear

space L, and no sequence of $n+1$ elements in L is linearly independent. Then n is called the **dimension** of L and is denoted by $\dim L$. If there exists such a number n, L is said to be **finite-dimensional**. Otherwise, L is said to be **infinite-dimensional**. In an infinite-dimensional linear space, there exist linearly independent sequences of elements having arbitrary length. The linear space K^n of n-tuples in K is of dimension n.

A sequence $(a_1, \ldots, a_n)$ of elements in a linear space L is called a basis if every element a of L is uniquely written in the form $a = \alpha_1 a_1 + \ldots + \alpha_n a_n$ ($\alpha_i \in K$, $i = 1, \ldots, n$). This means that the linear mapping $K^n \rightarrow L$ assigning $\alpha_1 a_1 + \ldots + \alpha_n a_n \in L$ to $(\alpha_1, \ldots, \alpha_n) \in K^n$ is bijective and hence an isomorphism. The condition that $(a_1, \ldots, a_n)$ is a basis of L is equivalent to any two of the following three conditions: (i) $(a_1, \ldots, a_n)$ is linearly independent; (ii) every element of L is a linear combination of $a_1, \ldots, a_n$; (iii) L is of dimension n. It follows that the length n of a basis $(a_1, \ldots, a_n)$ is equal to the dimension and hence is independent of the choice of basis. In the expression $a = \sum \alpha_i a_i$, α_i is called the **ith component** (or **ith coordinate**) of the element a relative to the basis $(a_1, \ldots, a_n)$.

D. Spaces of Linear Mappings (Finite-Dimensional Case)

Let L, M, and N be finite-dimensional linear spaces over a field K. The set $\mathrm{Hom}_K(L, M)$ of all linear mappings $L \rightarrow M$ is a linear space under the operations defined by $(\varphi + \varphi')(a) = \varphi(a) + \varphi'(a)$, $(\lambda\varphi)(a) = \lambda\varphi(a)$ ($a \in L$, $\lambda \in K$).

Let $(a_1, \ldots, a_l)$, $(b_1, \ldots, b_m)$ be bases of L, M, respectively. Then any linear mapping $\varphi: L \rightarrow M$ can be represented by an $m \times l$ matrix (α_{ij}) determined by $\varphi(a_j) = \sum_{i=1}^{m} b_i \alpha_{ij}$ $(1 \leqslant j \leqslant l)$. This assignment $\varphi \rightarrow (\alpha_{ij})$ gives an isomorphism from the linear space $\mathrm{Hom}_K(L, M)$ to the linear space of all $m \times l$ matrices (→ 269 Matrices). In addition, let $(c_1, \ldots, c_n)$ be a basis of N, and let the $n \times m$ matrix (β_{ki}) represent a linear mapping $\psi: M \rightarrow N$. Then the composite mapping $\psi \circ \varphi: L \rightarrow N$ is represented by the product $(\beta_{ki})(\alpha_{ij})$ of the matrices (β_{ki}) and (α_{ij}). The set of all linear transformations of a linear space N of dimension n forms an †associative algebra over K which is isomorphic to the †total matrix algebra $M_n(K)$ of degree n under the correspondence $\varphi \rightarrow (\alpha_{ij})$. Its †invertible elements are regular linear transformations, and they form a group which is denoted by $GL(N)$ and called the †**general linear group** on N. This corresponds to the group $GL(n, K)$ formed by all $n \times n$ †invertible matrices under the isomorphism $\varphi \rightarrow (\alpha_{ij})$.

E. Infinite-Dimensional Linear Spaces

In this section, we consider only the algebraic aspects of infinite-dimensional linear spaces (for the topological aspects → 422 Topological Abelian Groups L; 424 Topological Linear Spaces). Let $\{a_\lambda\}_{\lambda \in \Lambda}$ be a family of elements in a linear space L. A linear combination of the family is an element of L in the form $\sum_{\lambda \in \Lambda} \alpha_\lambda a_\lambda$ ($\alpha_\lambda \in K$, where $\alpha_\lambda = 0$ except for a finite number of λ). The family $\{a_\lambda\}_{\lambda \in \Lambda}$ is called **linearly independent** if no linear combination $\sum_{\lambda \in \Lambda} \alpha_\lambda a_\lambda$ is equal to 0 unless all the coefficients α_λ are equal to 0. The family $\{a_\lambda\}_{\lambda \in \Lambda}$ is called a **basis** of L if every element of L is uniquely written in the form $\sum_{\lambda \in \Lambda} \alpha_\lambda a_\lambda$. These notions are generalizations of those defined for finite Λ. In general, Λ is an infinite set. Any linear space L has a basis (→ 34 Axiom of Choice and Equivalents C). The cardinality of a basis is determined by L. Two linear spaces are isomorphic if and only if their bases have the same cardinality.

F. Subspaces and Quotient Spaces

Let L be a linear space over a field K. A nonempty subset N of L forms a linear space over K under the induced operations if the following two conditions hold: (i) $a, b \in N$ implies $a + b \in N$; and (ii) $\lambda \in K$, $a \in N$ imply $\lambda a \in N$. In this case, the subset N is called a **linear subspace** of L (or simply **subspace** of L). The canonical mapping $\varphi: N \rightarrow L$ defined by $\varphi(a) = a$ ($a \in N$) is an injective linear mapping.

Let S be a nonempty subset of L. The set of all linear combinations of elements in S forms the smallest subspace of L containing S; this space is called the subspace **generated** (or **spanned**) by S. For subspaces N, N' of L, the intersection $N \cap N'$ and the **sum** $N + N' = \{a + a' \mid a \in N, a' \in N'\}$ are both subspaces. Similar propositions hold for an arbitrary number of subspaces. If N, N' are of finite dimension, the equality $\dim N + \dim N' = \dim(N \cap N') + \dim(N + N')$ holds. We say that L is decomposed into the **direct sum** of N, N' if every element of L can be uniquely written in the form $a + a'$ ($a \in N$, $a' \in N'$). This is the case if and only if L is generated by N and N' and $N \cap N' = \{0\}$. In this case, N' is called a **complementary subspace** of N. Any subspace has a complementary subspace. For direct products and sums of linear spaces → 277 Modules F.

An equivalence relation R in a linear space L is said to be **compatible** with the operations in L if the following two conditions hold: (i) $R(a, a')$ and $R(b, b')$ imply $R(a + b, a' + b')$; (ii) $R(a, a')$ implies $R(\lambda a, \lambda a')$ ($\lambda \in K$). Then the †quotient set L/R, namely, the set of all

equivalence classes, forms a linear space over K under the induced operations; this is called the **quotient linear space** (or simply **quotient space**) of L with respect to R. The canonical mapping $\varphi: L \rightarrow L/R$, given by $a \in \varphi(a)$ $(a \in L)$, is a surjective linear mapping. The equivalence class N containing 0 forms a subspace of L, and the equivalence class containing $a \in L$ is the coset $a+N=\{a+b \mid b \in N\}$. We have $R(a, a')$ if and only if $a-a' \in N$. Conversely, for any subspace N, there exists an equivalence relation R compatible with the operations determined by $R(a, a')$ if and only if $a-a' \in N$. The quotient linear space L/R thus obtained is denoted by L/N and called the **quotient (linear) space** of L by N. If L/N is finite-dimensional, its dimension is called the **codimension** of N relative to L and is denoted by $\operatorname{codim} N$.

Let $\varphi: L \rightarrow M$ be a linear mapping of linear spaces. Its **image** $\varphi(L)$ is a subspace of M, and the **kernel** $N=\{a \in L \mid \varphi(a)=0\}$ is a subspace of L. The mapping φ induces an isomorphism $\varphi: L/N \rightarrow \varphi(L)$ (→ 277 Modules E). If L is finite-dimensional, $\dim L - \dim N = \dim \varphi(L)$. The dimension of the image of φ is called the **rank** of φ, and the dimension of the kernel of φ is called the **nullity** of φ. The rank and nullity of an $m \times n$ matrix (α_{ij}) (→ 269 Matrices) are the respective rank and nullity of the linear mapping $K^n \rightarrow K^m$ represented by the matrix (α_{ij}) (→ Section B, example (2)).

G. Dual Spaces

Let L be a linear space over a field K. The set $\mathrm{Hom}_K(L, K)$ of all linear forms on L is a linear space, denoted by L^* and called the **dual (linear) space** of L. The space L^* is the †dual module of L as a K-module (→ 277 Modules). For elements a of L and a^* of L^*, we denote the element $a^*(a)$ by $\langle a, a^* \rangle$ and call it the **inner product** of a and a^*. For a linear mapping $\varphi: L \rightarrow M$, we define a linear mapping ${}^t\varphi: M^* \rightarrow L^*$ by $({}^t\varphi)(b^*) = b^* \circ \varphi$ $(b^* \in M^*)$. The mapping ${}^t\varphi$ is called the **dual mapping (transposed mapping** or **transpose)** of φ, and is determined by the relation $\langle a, {}^t\varphi(b^*) \rangle = \langle \varphi(a), b^* \rangle$ $(a \in L,\ b^* \in M^*)$. We have ${}^t(\varphi_1+\varphi_2) = {}^t\varphi_1 + {}^t\varphi_2$, ${}^t(\psi \circ \varphi) = {}^t\varphi \circ {}^t\psi$, ${}^t1_L = 1_{L^*}$. If φ is †surjective, then ${}^t\varphi$ is †injective, and if φ is injective, then ${}^t\varphi$ is surjective. If φ is bijective, then ${}^t\varphi$ is also bijective. The rank of ${}^t\varphi$ coincides with the rank of φ if the rank of φ is finite. For an isomorphism $\varphi: L \rightarrow M$, the inverse mapping ${}^t\varphi^{-1} = \check{\varphi}: L^* \rightarrow M^*$ of ${}^t\varphi$ is called the **contragredient** of φ. We have $(\psi \circ \varphi)\check{} = \check{\psi} \circ \check{\varphi}$.

Given a subspace N of a linear space L, the subspace $\{a^* \in L^* \mid \langle a, a^* \rangle = 0\ (a \in N)\}$ of L^* is denoted by $N^\perp$ and is called the subspace (of L^*) **orthogonal** to N. Then we have the canonical isomorphisms $(L/N)^* \cong N^\perp$, $N^* \cong L^*/N^\perp$. Similarly, given a subspace N' of L^*, we obtain the subspace $N'^\perp$ of L orthogonal to N': $N'^\perp = \{a \in L \mid \langle a, a^* \rangle = 0\ (a^* \in N')\}$. Thus we have a one-to-one correspondence between the finite-codimensional subspaces of L and the finite-dimensional subspaces of L^* by assigning $N^\perp$ to N and $N'^\perp$ to N'. The codimension of N is equal to the dimension of $N^\perp$. If L is finite-dimensional, we have a canonical isomorphism $L \cong (L^*)^*$ and a one-to-one correspondence $N \rightarrow N' = N^\perp$ between the set $\{N\}$ of all subspaces of L and the set $\{N'\}$ of all subspaces of L^*. These properties of correspondence between the subspaces of L and L^* are called **duality properties** of the linear space.

Let $(e_1, \ldots, e_n)$ be a basis of a linear space L. Then the system of elements $(e_1^*, \ldots, e_n^*)$ in L^* defined by the relation $\langle e_j, e_i^* \rangle = 0$ $(i \neq j)$, $\langle e_i, e_i^* \rangle = 1$ forms a basis of L^*, called the **dual basis** of $(e_1, \ldots, e_n)$. Thus a finite-dimensional linear space L can be identified with its dual space utilizing the isomorphism given by assigning each element of the dual basis to the corresponding element of the basis in a natural manner.

H. Multilinear Mappings

Let L, M, N be linear spaces over a field K and f be a mapping from the Cartesian product $M \times N$ to L. Suppose that for any fixed $b \in N$, the mapping $M \rightarrow L$ assigning $f(x, b) \in L$ to $x \in M$ is linear, and for any fixed $a \in M$, the mapping $N \rightarrow L$ assigning $f(a, y) \in L$ to $y \in N$ is also linear. Then f is called a **bilinear mapping** from $M \times N$ to L. The set of all bilinear mappings from $M \times N$ to L forms a linear space under the operations $(f+g)(x, y) = f(x, y) + g(x, y)$, $(\lambda f)(x, y) = \lambda f(x, y)$ $(\lambda \in K)$; this space is denoted by $\mathscr{L}(M, N; L)$. In general, for linear spaces $M_1, \ldots, M_n$, a mapping $f: M_1 \times \ldots \times M_n \rightarrow L$ is called a **multilinear mapping** if it is linear in each variable. The set of all multilinear mappings from $M_1 \times \ldots \times M_n$ to L forms a linear space, denoted by $\mathscr{L}(M_1, \ldots, M_n; L)$. If $L = K$, a bilinear mapping and a multilinear mapping are called a **bilinear form** and **multilinear form**, respectively.

Suppose, in particular, that $M_1 = \ldots = M_n = M$. A multilinear mapping $f: M_1 \times \ldots \times M_n \rightarrow L$ is called **symmetric** if $f(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) = f(x_1, \ldots, x_n)$ $(x_i \in M)$ for any permutation σ of $\{1, \ldots, n\}$. Also, f is called **alternating** if $f(x_1, \ldots, x_i, \ldots, x_j, \ldots, x_n) = 0$ for $x_i = x_j$, $i \neq j$. In this case, $f(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) = \operatorname{sgn} \sigma \cdot f(x_1, \ldots, x_n)$ for any permutation σ ($\operatorname{sgn} \sigma$ is $+1$ if σ is †even and -1 if σ is †odd). On the other hand, f is called **skew-symmetric** (or **antisymmetric**) if it satisfies this equality. Therefore, if the charac-

teristic of the field K is different from 2, a skew-symmetric mapping is alternating.

Let M, N be linear spaces over a field K and Φ be a bilinear form on $M \times N$. The mappings $d_\Phi : N \to M^*$, $s_\Phi : M \to N^*$ defined by $\Phi(x, y) = (d_\Phi(y))(x) = (s_\Phi(x))(y)$ $(x \in M, y \in N)$ are linear mappings. If M, N are finite-dimensional, then d_Φ and s_Φ have the same rank, called the **rank** of Φ. Let $(x_1, \ldots, x_m)$, $(y_1, \ldots, y_n)$ be bases of M, N and $(x_1^*, \ldots, x_m^*)$, $(y_1^*, \ldots, y_n^*)$ be their dual bases. Then we have $d_\Phi(y_j) = \sum_{i=1}^m x_i^* \Phi(x_i, y_j)$, $s_\Phi(x_i) = \sum_{j=1}^n \Phi(x_i, y_j) y_j^*$. The matrix $(\Phi(x_i, y_j))$ is called the **matrix of a bilinear form** Φ relative to the given bases, and its rank coincides with the rank of Φ. If d_Φ, s_Φ are both injective, they are also bijective, and in this case Φ is said to be **nondegenerate**. Then each of d_Φ and s_Φ can be regarded as the transpose of the other, and M, N can be identified with N^*, M^*, respectively. In particular, if Φ is a nondegenerate bilinear form on $M \times M$, we have an isomorphism from M to its dual space M^*; identifying M with M^* by this isomorphism, M is said to be **self-dual**.

Let M be a linear space over a field K. A mapping $Q : M \to K$ is called a **quadratic form** on M if the following two conditions hold: (i) $Q(\alpha x) = \alpha^2 Q(x)$ $(\alpha \in K, x \in M)$; and (ii) the mapping $\Phi : M \times M \to K$ defined by $\Phi(x, y) = Q(x+y) - Q(x) - Q(y)$ $(x, y \in M)$ is a bilinear form on $M \times M$. In this case, Φ is called the **bilinear form associated with the quadratic form** Q, and it can be shown to be symmetric. We have $\Phi(x, x) = 2Q(x)$ $(x \in M)$ and $Q(x) = (1/2)\Phi(x, x)$ if the characteristic of $K \neq 2$. In general, for any bilinear form $f : M \times M \to K$, the mapping $Q : M \to K$ defined by $Q(x) = f(x, x)$ is a quadratic form. If $(x_1, \ldots, x_n)$ is a basis of M, a quadratic form Q is expressed as follows: $Q(\sum_{i=1}^n \xi_i x_i) = \sum_{\{i,j\}} \alpha_{ij} \xi_i \xi_j$ (the sum over all unordered pairs $\{i, j\}$), where $\alpha_{ii} = Q(x_i)$, $\alpha_{ij} = \Phi(x_i, x_j)$ $(i \neq j)$ (→ 348 Quadratic Forms). A **metric vector space** is a linear space M supplied with a nondegenerate quadratic form Q on M, and is denoted by (M, Q). The bilinear form Φ associated with Q gives an **inner product** $\Phi(x, y)$ $(x, y \in M)$.

I. Tensor Products

Let M, N be linear spaces over a field K. The **tensor product** $M \otimes N$ of M, N is defined as follows and can be used to "linearize" bilinear mappings from $M \times N$ to any linear space. Let F be the linear space generated by $M \times N$ and R be the subspace of F generated by all elements of the forms $(x+x', y) - (x, y) - (x', y)$, $(x, y+y') - (x, y) - (x, y')$, $(\alpha x, y) - \alpha(x, y)$, $(x, \alpha y) - \alpha(x, y)$ $(x, x' \in M, y, y' \in N, \alpha \in K)$. Then the quotient space F/R is denoted by $M \otimes N$, and the canonical projection $F \to M \otimes N$ is denoted by ψ. Given an element $(x, y) \in M \times N$, we denote the image $\psi((x, y))$ by $x \otimes y$. The bilinear mapping $M \times N \to M \otimes N$ assigning $x \otimes y$ to (x, y) is called the **canonical bilinear mapping**. We have, by definition, $(x+x') \otimes y = x \otimes y + x' \otimes y$, $x \otimes (y + y') = x \otimes y + x \otimes y'$, $(\alpha x) \otimes y = \alpha(x \otimes y) = x \otimes (\alpha y)$ $(\alpha \in K)$. To emphasize the basic field K, we sometimes write $M \otimes_K N$ instead of $M \otimes N$.

The tensor product can be characterized by the property that for any linear space L and any bilinear mapping $f : M \times N \to L$, there exists a unique linear mapping $\varphi : M \otimes N \to L$ satisfying $f(x, y) = \varphi(x \otimes y)$. Thus assigning the bilinear mapping $f : M \times N \to L$ defined by $f(x, y) = \varphi(x \otimes y)$ to a linear mapping $\varphi : M \otimes N \to L$, we obtain an isomorphism $\mathrm{Hom}(M \otimes N, L) \cong \mathscr{L}(M, N; L)$. Every element of $M \otimes N$ can be expressed as a finite sum of elements of the form $x \otimes y$ $(x \in M, y \in N)$. If $\{x_i\}_{i \in I}$, $\{y_j\}_{j \in J}$ are bases of M, N, respectively, then the family $\{x_i \otimes y_j\}_{i \in I, j \in J}$ forms a basis of $M \otimes N$. Hence if M and N are of finite dimension, $\dim(M \otimes N) = \dim M \dim N$.

Let $M_1, M_2, \ldots$ be linear spaces over a field K. We have a unique isomorphism $M_1 \otimes M_2 \to M_2 \otimes M_1$ that assigns $x_2 \otimes x_1$ to $x_1 \otimes x_2$ $(x_i \in M_i)$. We also have a unique isomorphism $(M_1 \otimes M_2) \otimes M_3 \to M_1 \otimes (M_2 \otimes M_3)$ that assigns $x_1 \otimes (x_2 \otimes x_3)$ to $(x_1 \otimes x_2) \otimes x_3$ $(x_i \in M_i)$; hence we can identify $(M_1 \otimes M_2) \otimes M_3$ and $M_1 \otimes (M_2 \otimes M_3)$, and we denote them simply by $M_1 \otimes M_2 \otimes M_3$. In general, assigning $x_1 \otimes \ldots \otimes x_n$ to $(x_1, \ldots, x_n)$, we obtain the canonical multilinear mapping $M_1 \times \ldots \times M_n \to M_1 \otimes \ldots \otimes M_n$. As before, given any linear space L, we have the natural isomorphism $\mathrm{Hom}(M_1 \otimes \ldots \otimes M_n, L) \cong \mathscr{L}(M_1, \ldots, M_n; L)$. Conversely, given linear spaces $M_1, \ldots, M_n$, the space $M_1 \otimes \ldots \otimes M_n$ can be characterized as a linear space N with a given multilinear mapping $\psi : M_1 \times \ldots \times M_n \to N$ such that (i) N is generated by the image $\psi(M_1 \times \ldots \times M_n)$; and (ii) for any multilinear mapping $f : M_1 \times \ldots \times M_n \to L$, there exists a unique linear mapping $f' : N \to L$ satisfying $f = f' \circ \psi$. The tensor product $M_1 \otimes \ldots \otimes M_n$ is sometimes written as $\bigotimes_{i=1}^n M_i$, and an element $x_1 \otimes \ldots \otimes x_n$ is written as $\bigotimes_{i=1}^n x_i$.

Given linear mappings $f_i : M_i \to M_i'$ $(1 \leqslant i \leqslant n)$, there exists a unique linear mapping $f : M_1 \otimes \ldots \otimes M_n \to M_1' \otimes \ldots \otimes M_n'$ satisfying $f(x_1 \otimes \ldots \otimes x_n) = f_1(x_1) \otimes \ldots \otimes f_n(x_n)$; we denote the mapping f by $f_1 \otimes \ldots \otimes f_n$ or $\bigotimes_{i=1}^n f_i$ and call it the **tensor product** of the f_i $(1 \leqslant i \leqslant n)$. The assignment $(f_1, \ldots, f_n) \to f_1 \otimes \ldots \otimes f_n$ gives an isomorphism $\bigotimes_{i=1}^n \mathrm{Hom}(M_i, M_i') \to \mathrm{Hom}(\bigotimes_{i=1}^n M_i, \bigotimes_{i=1}^n M_i')$ if the M_i are finite-dimensional. If in particular $M_1' = \ldots = M_n' =$

K, we have an isomorphism $\bigotimes_{i=1}^n M_i^* \to (\bigotimes_{i=1}^n M_i)^*$ under the identification $\bigotimes_{i=1}^n M_i' = K$ given by the assignment $x_1' \otimes \ldots \otimes x_n' \to x_1' \ldots x_n'$. Explicitly, the isomorphism f: $\bigotimes_{i=1}^n M_i^* \to (\bigotimes_{i=1}^n M_i)^*$ is determined by $f(\bigotimes_{i=1}^n x_i^*)(\bigotimes_{i=1}^n x_i) = \prod_{i=1}^n \langle x_i, x_i^* \rangle$ $(x_i^* \in M_i^*, x_i \in M_i)$.

J. Tensors

Let $E^{(\lambda)}$ $(1 \leqslant \lambda \leqslant k)$ be linear spaces over a field K. If $E^{(1)} = \ldots = E^{(k)} = E$, then $\bigotimes_{\lambda=1}^k E^{(\lambda)}$ is written $\bigotimes^k E$ and called the **tensor space of degree** k of E ($\bigotimes^0 E$ denotes K). Also, $(\bigotimes^p E) \otimes (\bigotimes^q E^*)$ is written $T_q^p(E)$, where E^* is the dual space of E. We have $T_0^p(E) = \bigotimes^p E$, $T_q^0(E) = \bigotimes^q E^*$, and $T_0^0(E) = K$. $T_q^p(E)$ is called the **tensor space of type** (p, q) of E, and each of its elements is called a **tensor of type** (p, q). In particular, a tensor of type $(p, 0)$ is called a **contravariant tensor of degree** p, and a tensor of type $(0, q)$ is called a **covariant tensor of degree** q. A tensor of type $(0, 0)$ is a **scalar**. An element of $T_0^1(E) = E$ is called a **contravariant vector**, and an element of $T_1^0(E) = E^*$ is called a **covariant vector**. If $p \neq 0$, $q \neq 0$, a tensor of type (p, q) is called a **mixed tensor**.

Let $(e_1, \ldots, e_n)$ be a basis of E and $(f^1, \ldots, f^n)$ be the basis of E^* dual to $(e_1, \ldots, e_n)$. Then the tensors $e_{i_1} \otimes \ldots \otimes e_{i_p} \otimes f^{j_1} \otimes \ldots \otimes f^{j_q}$ $(i_\lambda, j_\mu = 1, \ldots, n;\ \lambda = 1, \ldots, p;\ \mu = 1, \ldots, q)$ form a basis of $T_q^p(E)$. Therefore any tensor of type (p, q) can be written uniquely in the form

$$t = \sum \xi_{j_1 \ldots j_q}^{i_1 \ldots i_p} e_{i_1} \otimes \ldots \otimes e_{i_p} \otimes f^{j_1} \otimes \ldots \otimes f^{j_q}.$$

Also, $\xi_{j_1 \ldots j_q}^{i_1 \ldots i_p}$ is called the **component** of t relative to the basis $(e_1, \ldots, e_n)$, the index i_λ is called a **contravariant index**, and the index j_μ is called a **covariant index**.

Let $(\bar{e}_1, \ldots, \bar{e}_n)$ be another basis of E and $(\bar{f}^1, \ldots, \bar{f}^n)$ be its dual basis. Suppose that we have

$$\bar{e}_i = \sum_{j=1}^n \alpha_i^j e_j, \quad \bar{f}^i = \sum_{j=1}^n \beta_j^i f^j.$$

Then we have

$$\sum_{k=1}^n \beta_k^i \alpha_j^k = \delta_j^i$$

and the transformation formula

$$\bar{\xi}_{j_1 \ldots j_q}^{i_1 \ldots i_p} = \sum_{k_1, \ldots, k_p} \sum_{l_1, \ldots, l_q} \beta_{k_1}^{i_1} \ldots \beta_{k_p}^{i_p} \alpha_{j_1}^{l_1} \ldots \alpha_{j_q}^{l_q} \xi_{l_1 \ldots l_q}^{k_1 \ldots k_p},$$

where the $\xi_{l_1 \ldots l_q}^{k_1 \ldots k_p}$ are the components of t relative to $(e_1, \ldots, e_n)$ and the $\bar{\xi}_{j_1 \ldots j_q}^{i_1 \ldots i_p}$ are the components of t relative to $(\bar{e}_1, \ldots, \bar{e}_n)$.

In the tensor calculus, an index appearing after the symbol $\sum$ is called a **dummy index** if it appears in both the upper and the lower positions. For example, in the expression $\sum_{i=1}^n \xi_i y^i$, the index i is a dummy. As a convention, we sometimes omit the symbol $\sum_{i=1}^n$ for a dummy index i; for example, by $\xi_i y^i$ we mean the sum $\sum_{i=1}^n \xi_i y^i$. This convention is called **Einstein's convention**. Using it, we write the previous transformation formula as

$$\bar{\xi}_{j_1 \ldots j_q}^{i_1 \ldots i_p} = \beta_{k_1}^{i_1} \ldots \beta_{k_p}^{i_p} \alpha_{j_1}^{l_1} \ldots \alpha_{j_q}^{l_q} \xi_{l_1 \ldots l_q}^{k_1 \ldots k_p}.$$

We have a nondegenerate bilinear form Φ on $T_q^p(E) \times T_p^q(E)$ determined by

$$\Phi\left(\bigotimes_{i=1}^p x_i \otimes \bigotimes_{j=1}^q y_j^*, \bigotimes_{j=1}^q y_j \otimes \bigotimes_{i=1}^p x_i^* \right) = \prod_{i=1}^p \langle x_i, x_i^* \rangle \prod_{j=1}^q \langle y_j, y_j^* \rangle.$$

Thus the space $T_q^p(E)$ can be identified with the dual space of $T_p^q(E)$, and vice versa (→ Section H). In this identification, the basis $(e_{i_1} \otimes \ldots \otimes e_{i_p} \otimes f^{j_1} \otimes \ldots \otimes f^{j_q})$ of $T_q^p(E)$ and the basis $(e_{j_1} \otimes \ldots \otimes e_{j_q} \otimes f^{i_1} \otimes \ldots \otimes f^{i_p})$ of $T_p^q(E)$ are dual to each other. In addition, combining the natural isomorphism $T_p^q(E)^* \cong \mathscr{L}(\prod^q E, \prod^p E^*; K)$ with the duality $T_p^q(E)^* = T_q^p(E)$, we have a natural isomorphism $T_q^p(E) \to \mathscr{L}(\prod^q E, \prod^p E^*; K)$. Explicitly, identifying an element $t \in T_q^p(E)$ with the multilinear form $\prod^q E \times \prod^p E^* \to K$ corresponding to it under the natural isomorphism, we have

$$t(x_1, \ldots, x_q, y_1^*, \ldots, y_p^*) = \zeta_{j_1 \ldots j_q}^{i_1 \ldots i_p} \xi_1^{j_1} \ldots \xi_q^{j_q} \eta_{i_1}^1 \ldots \eta_{i_p}^p,$$

where $\{\xi_\lambda^j\}$ is the component of $x_\lambda \in E = T_0^1(E)$, $\{\eta_i^\mu\}$ is the component of $y_\mu^* \in E^* = T_1^0(E)$, and $\{\zeta_{j_1 \ldots j_q}^{i_1 \ldots i_p}\}$ is the component of $t \in T_q^p(E)$. By the natural isomorphisms $T_0^p(E) \cong \mathscr{L}(\prod^p E^*, K)$, $T_p^0(E) = \mathscr{L}(\prod^p E, K)$, a contravariant tensor of degree p and a covariant tensor of degree p can be identified with a multilinear p-form on E^* and on E, respectively.

K. Tensor Algebras

There exists a unique bilinear mapping $T_q^p(E) \times T_s^r(E) \to T_{q+s}^{p+r}(E)$ which assigns the element $x_1 \otimes \ldots \otimes x_p \otimes y_1 \otimes \ldots \otimes y_r \otimes x_1^* \otimes \ldots \otimes x_q^* \otimes y_1^* \otimes \ldots \otimes y_s^*$ to the pair $(x_1 \otimes \ldots \otimes x_p \otimes x_1^* \otimes \ldots \otimes x_q^*, y_1 \otimes \ldots \otimes y_r \otimes y_1^* \otimes \ldots \otimes y_s^*)$; we denote the element assigned to the latter by $t \otimes u$, where $t = x_1 \otimes \ldots \otimes x_p \otimes x_1^* \otimes \ldots \otimes x_q^*$, $u = y_1 \otimes \ldots \otimes y_r \otimes y_1^* \otimes \ldots \otimes y_s^*$, and call it the **product** of t and u. If the components of t, u, $t \otimes u$ are

$$\{\xi_{j_1 \ldots j_q}^{i_1 \ldots i_p}\}, \quad \{\eta_{l_1 \ldots l_s}^{k_1 \ldots k_r}\}, \quad \{\zeta_{b_1 \ldots b_{q+s}}^{a_1 \ldots a_{p+r}}\},$$

then we have

$$\zeta_{j_1 \ldots j_q l_1 \ldots l_s}^{i_1 \ldots i_p k_1 \ldots k_r} = \xi_{j_1 \ldots j_q}^{i_1 \ldots i_p} \eta_{l_1 \ldots l_s}^{k_1 \ldots k_r}.$$

Let $T(E)$ be the direct sum of $T_q^p(E)$ $(p, q = 0, 1, 2, \ldots)$. Then $T(E)$ is an associative algebra over K whose product is a natural

extension of the product $\otimes$. We call $T(E)$ the **tensor algebra** on E. The direct sum of $T_0^p(E)$ $(p=0,1,2,\ldots)$ forms a subalgebra of $T(E)$, also called the **(contravariant) tensor algebra**.

L. Contractions

The **contraction** of $T_q^p(E)$ relative to the kth contravariant index and the lth covariant index is by definition the linear mapping $C_l^k: T_q^p(E)\to T_{q-1}^{p-1}(E)$ determined by assigning $\langle x_k, x_l^*\rangle x_1\otimes\ldots\otimes x_{k-1}\otimes x_{k+1}\otimes\ldots\otimes x_p\otimes x_1^*\otimes\ldots\otimes x_{l-1}^*\otimes x_{l+1}^*\ldots\otimes x_q^*$ to $x_1\otimes\ldots\otimes x_p\otimes x_1^*\otimes\ldots\otimes x_q^*$, where $\langle x_k, x_l^*\rangle$ is the inner product of x_k, x_l^*. For a tensor t of type (p,q), the tensor $C_l^k(t)$ of type $(p-1,q-1)$ is called the **contracted tensor** of t. If the components of t are $\xi_{j_1\ldots j_q}^{i_1\ldots i_p}$, the components of $C_l^k(t)$ are given by

$$\eta_{j_1\ldots j_{q-1}}^{i_1\ldots i_{p-1}}=\sum_{s=1}^{n}\xi_{j_1\ldots j_{l-1}sj_l\ldots j_{q-1}}^{i_1\ldots i_{k-1}si_k\ldots i_{p-1}}.$$

M. Tensor Representations

For a linear mapping $f: E\to F$, the tensor product $f\otimes\ldots\otimes f: T_0^p(E)=\bigotimes^p E\to\bigotimes^p F=T_0^p(F)$ is denoted by f^p. The f^p $(p=0,1,2,\ldots)$ give an algebra homomorphism $\sum_{p=0}^{\infty}T_0^p(E)\to\sum_{p=0}^{\infty}T_0^p(F)$. Next, let f be an isomorphism and $\check{f}={}^t f^{-1}$ be its contragredient. Then f_q denotes the tensor product $\check{f}\otimes\ldots\otimes\check{f}: T_q^0(E)=\bigotimes^q E^*\to\bigotimes^q F^*=T_q^0(F)$, and f_q^p the tensor product $f^p\otimes f_q: T_q^p(E)\to T_q^p(F)$. The mapping f_q^p is an isomorphism, and the system $\{f_q^p\}$ $(p,q=0,1,2,\ldots)$ gives rise to an algebra isomorphism $T(E)\to T(F)$. In particular, if f is a nonsingular linear transformation of E, then f_q^p is a nonsingular linear transformation of the linear space $T_q^p(E)$, and the assignment $f\to f_q^p$ gives a group homomorphism $GL(E)\to GL(T_q^p(E))$; this homomorphism is called a **tensor representation** of the group $GL(E)$.

N. Symmetric and Alternating Tensors

A contravariant tensor of degree p is called **symmetric (alternating)** if the corresponding multilinear p-form, under the natural isomorphism $T_0^p(E)\cong\mathscr{L}(\prod^p E^*, K)$, is symmetric (alternating). A covariant tensor is also called **symmetric (alternating)** if the corresponding multilinear form is symmetric (alternating). A **skew-symmetric** (or **antisymmetric**) tensor is defined similarly. We reformulate these definitions under the assumption that the field K is not of characteristic 2. Let $\mathfrak{S}_p$ be the group of all permutations of $1,\ldots,p$ (the †symmetric group of degree p). For any $\sigma\in\mathfrak{S}_p$, we have a unique linear transformation $T_0^p(E)\to T_0^p(E)$ assigning $x_{\sigma^{-1}(1)}\otimes\ldots\otimes x_{\sigma^{-1}(p)}$ to $x_1\otimes\ldots\otimes x_p$. This transformation is nonsingular and is also denoted by σ. Similarly, we have a unique nonsingular linear transformation of $T_p^0(E)$, also denoted by σ. An element $t\in T_0^p(E)$ (or $\in T_p^0(E)$) is symmetric if and only if $\sigma t=t$ for all $\sigma\in\mathfrak{S}_p$, while t is alternating if and only if $\sigma t=(\operatorname{sgn}\sigma)t$ for all $\sigma\in\mathfrak{S}_p$. Let the $\xi^{i_1\ldots i_p}$ (or the $\xi_{i_1\ldots i_p}$) be the components of t. Then t is symmetric (alternating) if and only if the components are symmetric (alternating) relative to permutations of the indices $i_1,\ldots,i_p$. The linear transformation $S_p=\sum_{\sigma\in\mathfrak{S}_p}\sigma$ of $T_0^p(E)$ or $T_p^0(E)$ is called the **symmetrizer**, and $A_p=\sum_{\sigma\in\mathfrak{S}_p}(\operatorname{sgn}\sigma)\sigma$ is called the **alternizer**. For any t, $S_p t$ is a symmetric tensor, and $A_p t$ is an alternating tensor.

The subspace of $T_0^p(E)$ consisting of all symmetric (or alternating) tensors is invariant under the transformation $R: GL(E)\to GL(T_0^p(E))$, the tensor representation where $\varphi(t)=R(\varphi)(t)$ for $\varphi\in GL(E)$ and $t\in T_0^p(E)$.

O. Exterior Product

For simplicity, we assume that the basic field K is of characteristic 0. We denote by N_p the kernel of the alternizer $A_p: T_0^p(E)\to T_0^p(E)$, namely, the subspace consisting of all t satisfying $A_p t=0$, and by $\bigwedge^p E$ the quotient space $T_0^p(E)/N_p$. The image of $x_1\otimes\ldots\otimes x_p$ $(x_i\in E)$ under the natural mapping $T_0^p(E)\to\bigwedge^p E$ is denoted by $x_1\wedge\ldots\wedge x_p$ and is called the **exterior product** of $x_1,\ldots,x_p$. The linear space $\bigwedge^p E$ is called the p**-fold exterior power** of E. We have

$$x_1\wedge\ldots\wedge(x_i+x_i')\wedge\ldots\wedge x_p$$
$$=x_1\wedge\ldots\wedge x_i\wedge\ldots\wedge x_p+x_1\wedge\ldots\wedge x_i'\wedge\ldots\wedge x_p,$$
$$x_1\wedge\ldots\wedge(\alpha x_i)\wedge\ldots\wedge x_p$$
$$=\alpha(x_1\wedge\ldots\wedge x_i\wedge\ldots\wedge x_p),\qquad \alpha\in K,$$

and for every $\sigma\in\mathfrak{S}_p$,

$$x_{\sigma(1)}\wedge\ldots\wedge x_{\sigma(p)}=(\operatorname{sgn}\sigma)x_1\wedge\ldots\wedge x_p.$$

A_p induces a natural isomorphism $\bigwedge^p E\cong\mathfrak{A}^p$, where $\mathfrak{A}^p$ consists of all contravariant alternating tensors of degree p. Thus an element of $\bigwedge^p E$ can be identified with a contravariant alternating tensor of degree p. Then we have $A_p(x_1\otimes\ldots\otimes x_p)=x_1\wedge\ldots\wedge x_p$. Similarly, $\bigwedge^p E^*$ is identified with the linear space consisting of all covariant alternating tensors of degree p. An element of $\bigwedge^p E$ is sometimes called a p**-vector**, and an element of $\bigwedge^p E^*$ is called a p**-covector** (→ 90 Coordinates B). If $(e_1,\ldots,e_n)$ is a basis of E, then the $e_{i_1}\wedge\ldots\wedge e_{i_p}$ $(i_1<i_2<\ldots<i_p)$ form a basis of $\bigwedge^p E$, and any element of $\bigwedge^p E$ is written in

the form $t=\sum_{i_1<\ldots<i_p}\alpha^{i_1\ldots i_p}e_{i_1}\wedge\ldots\wedge e_{i_p}$ or $t=(1/p!)\sum_{i_1,\ldots,i_p}\alpha^{i_1\ldots i_p}e_{i_1}\wedge\ldots\wedge e_{i_p}$. In the latter form, $\alpha^{i_1\ldots i_p}$ is alternating relative to permutations of $i_1,\ldots,i_p$, and it is the component of t. The dimension of $\bigwedge^pE$ is equal to $\binom{n}{p}$, and $\bigwedge^pE=0$ if $p>n$. If $(f^1,\ldots,f^n)$ is the dual basis of $(e_1,\ldots,e_n)$, the inner product of an element $t=\sum_{i_1<\ldots<i_p}\alpha^{i_1\ldots i_p}e_{i_1}\wedge\ldots\wedge e_{i_p}\in\bigwedge^pE$ and an element $s=\sum_{i_1<\ldots<i_p}\beta_{i_1\ldots i_p}f^{i_1}\wedge\ldots\wedge f^{i_p}\in\bigwedge^pE^*$ is defined by

$$\langle s,t\rangle=\sum_{i_1<\ldots<i_p}\alpha^{i_1\ldots i_p}\beta_{i_1\ldots i_p}.$$

Then we have $\langle x_1\wedge\ldots\wedge x_p, y_1\wedge\ldots\wedge y_p\rangle=\det(\langle x_i,y_j\rangle)$, where $\langle x_i,y_j\rangle$ is the inner product of $x_i\in E$ and $y_j\in E^*$. By this inner product, we can identify $\bigwedge^pE^*$ with the dual space of $\bigwedge^pE$.

The tensor product $\otimes^pE\times\otimes^qE\to\otimes^{p+q}E$ induces naturally a mapping $\Phi:\otimes^pE/N_p\times\otimes^qE/N_q\to\otimes^{p+q}E/N_{p+q}$. Using this bilinear mapping $\Phi:\bigwedge^pE\times\bigwedge^qE\to\bigwedge^{p+q}E$, we define the **exterior product** $t\wedge s$ of an element $t\in\bigwedge^pE$ and an element $s\in\bigwedge^qE$ by $t\wedge s=\Phi(t,s)$. Then $t\wedge s$ is an element of $\bigwedge^{p+q}E$, and we have $t\wedge s=(-1)^{pq}s\wedge t$, $(x_1\wedge\ldots\wedge x_p)\wedge(x_{p+1}\wedge\ldots\wedge x_{p+q})=x_1\wedge\ldots\wedge x_{p+q}$.

We denote by $\bigwedge E$ the direct sum of $\bigwedge^pE$ ($p=0,1,2,\ldots,n$), and define the product of two elements $x=\sum_{p=0}^n x^p$, $y=\sum_{p=0}^n y^p$ ($x^p,y^p\in\bigwedge^pE$) by $x\wedge y=\sum_{p+q=0}^n x^p\wedge y^q$. Then the product $\wedge$ satisfies the associative law. We call $\bigwedge E$ the **exterior algebra** (or **Grassmann algebra**) of the linear space E. If E is of dimension n, $\bigwedge E$ is of dimension 2^n. If $(e_1,\ldots,e_n)$ is a basis of E relative to K, $\bigwedge E$ is sometimes written as $\bigwedge_K(e_1,\ldots,e_n)$. The exterior algebra $\bigwedge E^*$ of the dual space E^* is similarly defined and can be considered as the dual space of $\bigwedge E$.

P. Semilinear Mappings

Let L be a linear space over a field K and L' be a linear space over a field K'. A pair (φ,ρ) consisting of a mapping $\varphi:L\to L'$ and a mapping $\rho:K\to K'$ is called a **semilinear mapping** if the following four conditions hold (for convenience here we write the scalars to the right of the vectors): (i) $\varphi(a+b)=\varphi(a)+\varphi(b)$; (ii) $\varphi(a\lambda)=\varphi(a)\rho(\lambda)$; (iii) $\rho(\alpha+\beta)=\rho(\alpha)+\rho(\beta)$; and (iv) $\rho(\alpha\beta)=\rho(\alpha)\rho(\beta)$ ($a,b\in L$; $\lambda,\alpha,\beta\in K$). In this case, φ is sometimes said to be semilinear relative to ρ (→ 277 Modules L). Conditions (iii) and (iv) mean that ρ is a field homomorphism. If $K=K'$ and ρ is the identity, then the semilinear mapping $\varphi:L\to L'$ is a linear mapping. If $L=L'$, $K=K'$ (and ρ is an automorphism), φ is called a **semilinear transformation** relative to ρ.

For semilinear mappings $(\varphi,\rho):L\to L'$, $K\to K'$ and $(\varphi',\rho'):L'\to L''$, $K'\to K''$, where L'' is a linear space over K'', the composite $(\varphi'\circ\varphi,\rho'\circ\rho)$ is also a semilinear mapping. If a basis $(e_1,\ldots,e_n)$ for L over K and a basis $(e'_1,\ldots,e'_{n'})$ for L' over K' are given, a semilinear mapping $(\varphi,\rho):L\to L'$, $K\to K'$ determines a matrix (α_{ij}) by the relation $\varphi(e_j)=\sum_{i=1}^{n'}e'_i\alpha_{ij}$ $(1\leqslant j\leqslant n)$. Conversely, a homomorphism ρ and an $n'\times n$ matrix $A=(\alpha_{ij})$ determine a semilinear mapping φ by this relation. Hence for fixed bases a semilinear mapping is represented by a pair (A,ρ), where A is an $n'\times n$ matrix. If a semilinear mapping φ' relative to ρ' is represented by (A',ρ'), the composite $(\varphi'\circ\varphi,\rho'\circ\rho)$ is represented by $(A'A^{\rho'},\rho'\circ\rho)$, where $A^{\rho'}$ is the matrix $(\rho'(\alpha_{ij}))$.

Let $\varphi:L\to L$ be a semilinear transformation relative to an automorphism $\rho:K\to K$. Suppose that φ is represented by (A,ρ) relative to a basis $(e_1,\ldots,e_n)$ for L and by (B,ρ) relative to another basis $(f_1,\ldots,f_n)$. If we define a matrix $P=(p_{ij})$ by the relation $f_j=\sum_{i=1}^n e_ip_{ij}$ $(1\leqslant j\leqslant n)$, we have $B=P^{-1}AP^\rho$. Two pairs (A,ρ), (B,ρ) having the relation $B=P^{-1}AP^\rho$ are said to be **similar**.

Q. Sesquilinear Forms

Let K be a field (not necessarily commutative) and J be its †antiautomorphism. For left linear spaces M, N over K, a mapping $\Phi:M\times N\to K$ is called a (right) **sesquilinear form** relative to J if the following four conditions are satisfied: (i) $\Phi(x+x',y)=\Phi(x,y)+\Phi(x',y)$; (ii) $\Phi(x,y+y')=\Phi(x,y)+\Phi(x,y')$; (iii) $\Phi(\alpha x,y)=\alpha\Phi(x,y)$; and (iv) $\Phi(x,\alpha y)=\Phi(x,y)\alpha^J$ ($x,x'\in M$; $y,y'\in N$; $\alpha\in K$). If J is the identity automorphism, then K is necessarily commutative and Φ is a bilinear form (→ Section H). As an example of K and J, we may take K as the field of complex numbers and J as complex conjugation. In general, for a left linear space E over K, we denote by E^J the right linear space with the scalar multiplication $x\lambda=\lambda^{J^{-1}}x$ ($x\in E$, $\lambda\in K$). Then condition (iv) becomes (iv′) $\Phi(x,y\alpha)=\Phi(x,y)\alpha$; and if K is commutative, Φ is a bilinear form on $M\times N^J$. For a sesquilinear form Φ on $M\times N$, we have the linear mappings $d_\Phi:N^J\to M^*$, $s_\Phi:M^{J^{-1}}\to N^*$ defined by the relation $\Phi(x,y)=\langle x,d_\Phi(y)\rangle=\langle y,s_\Phi(x)\rangle^J$ ($x\in M$, $y\in N$). If M, N are finite-dimensional, d_Φ, s_Φ have the same rank, which is called the **rank** of Φ. We assume that all linear spaces are finite-dimensional.

Let $(x_1,\ldots,x_m)$, $(y_1,\ldots,y_n)$ be bases of M, N and $(x_1^*,\ldots,x_m^*)$, $(y_1^*,\ldots,y_n^*)$ be their dual bases. Then we have $d_\Phi(y_j)=\sum_{i=1}^m x_i^*\Phi(x_i,y_j)$, $s_\Phi(x_i)=\sum_{j=1}^n y_j^*\Phi(x_i,y_j)^{J^{-1}}$. The matrix $(\Phi(x_i,y_j))$ is called the **matrix of the sesquilinear form**

Φ relative to the given bases; its rank is equal to the rank of Φ. If d_Φ, s_Φ are both injective (therefore bijective), Φ is said to be **nondegenerate**. Let $\Phi': M' \times N' \to K$ be another sesquilinear form relative to J. Then for any linear mapping $u: M \to M'$, there exists a unique linear mapping $u^*: N' \to N$ such that $\Phi'(u(x), y') = \Phi(x, u^*(y'))$ $(x \in M, y' \in N')$; this is called the **left-adjoint** linear mapping of u. Similarly, for any linear mapping $v: N \to N'$, there exists a unique linear mapping $v^*: M' \to M$ such that $\Phi'(x', v(y)) = \Phi(v^*(x'), y)$ $(x' \in M', y \in N)$; this is called the **right-adjoint** linear mapping of v. We have $u^* = d_\Phi^{-1} \circ {}^t u \circ d_{\Phi'}$, $v^* = s_\Phi^{-1} \circ {}^t v \circ s_{\Phi'}$. In particular, let u, v be isomorphisms. Then we have $\Phi(x, y) = \Phi'(u(x), v(y))$ $(x \in M, y \in N)$ if and only if $u^{-1} = v^*$, $v^{-1} = u^*$.

A sesquilinear form on $M \times M$ is called simply a **sesquilinear form** on M. Let J be an †involution (namely $J = J^{-1}$), and write $\lambda^J = \bar{\lambda}$ $(\lambda \in K)$. If condition (v) $\Phi(x, y) = \overline{\Phi(y, x)}$ $(x, y \in M)$ holds, Φ is called a **Hermitian form** on M. On the other hand, if the condition (v′) $\Phi(x, y) = -\overline{\Phi(y, x)}$ holds, Φ is called an **anti-Hermitian form** (or **skew-Hermitian form**) on M. In particular, if $J = 1_K$, then a Hermitian form (anti-Hermitian form) is a symmetric bilinear form (antisymmetric bilinear form). A linear space M supplied with a nondegenerate Hermitian form Φ is called a **Hermitian linear space**, and $\Phi(x, y)$ is called the **Hermitian inner product** (or simply **inner product**) of $x, y \in M$.

References

[1] O. Schreier and E. Sperner, Einführung in die analytische Geometrie und Algebra I, II, Vandenhoeck & Ruprecht, fourth edition, 1959; English translation, Introduction to modern algebra and matrix theory, Chelsea, 1959.
[2] N. Bourbaki, Eléments de mathématique, Algèbre, ch. 2, 3, 7, 8, 9, Actualités Sci. Ind. 1236b, 1044, 1179a, 1261a, 1272a, Hermann, 1958–1964.
[3] I. M. Gel'fand, Lectures on linear algebra, Interscience, 1961. (Original in Russian, 1951.)
[4] N. Jacobson, Lectures in abstract algebra II, Van Nostrand, 1953.
[5] C. Chevalley, Fundamental concepts of algebra, Academic Press, 1956.
[6] W. H. Graeub, Lineare Algebra, Springer, 1958.
[7] P. R. Halmos, Finite-dimensional vector spaces, Van Nostrand, second edition, 1958.
[8] R. Godement, Cours d'algèbre, Hermann, 1963.
[9] D. A. Raĭkov, Vector spaces, Noordhoff, 1965. (Original in Russian, 1962.)
[10] S. Lang, Algebra, Addison-Wesley, 1965.

257 (V.17) Local Fields

A. General Remarks

A field k that is †complete with respect to a †discrete valuation is called a **local field** if its field of residue classes is finite. (Real and complex number fields are sometimes also called local fields; these, however, are not considered in this article.) A local field k is isomorphic either to the †completion with respect to a †$\mathfrak{p}$-adic valuation determined by a prime ideal $\mathfrak{p}$ of a number field of finite degree or to the field of †formal power series of one variable over a finite field. In the former case, k is called a **$\mathfrak{p}$-adic number field**. We let $\mathfrak{o}$ stand for the †valuation ring of k, $\mathfrak{p}$ stand for the †valuation ideal of k, p stand for the †characteristic of the field $\mathfrak{o}/\mathfrak{p}$ of residue classes, and $N(\mathfrak{p})$ stand for the number of the elements of $\mathfrak{o}/\mathfrak{p}$. An †additive valuation of k whose set of values coincides with the set of all rational integers, is denoted by $\operatorname{ord}\alpha$ $(\alpha \in k)$; here we understand $\operatorname{ord} 0 = \infty$. The †normal (multiplicative) valuation of k is defined by $|\alpha| = (N(\mathfrak{p}))^{-\operatorname{ord}\alpha}$ (→ 439 Valuations).

B. Construction of Local Fields

A $\mathfrak{p}$-adic number field k is an extension of finite degree of the p-adic field $\mathbf{Q}_p$. If $n = [k:\mathbf{Q}_p] = ef$, where e is the †ramification index of $k/\mathbf{Q}_p$ and f is the †relative degree of $k/\mathbf{Q}_p$ (→ Section D), then there exists one and only one field F such that $k \supset F \supset \mathbf{Q}_p$, $[k:F] = e$, $[F:\mathbf{Q}_p] = f$, and $F/\mathbf{Q}_p$ is †unramified. The field κ of residue classes of F is isomorphic to $GF(p^f)$, and F is uniquely determined by κ by means of Witt vectors (→ 449 Witt Vectors). Every residue class $(\neq \bar{0})$ of $\mathfrak{o}_F/\mathfrak{p}_F \cong \kappa$ contains one and only one mth root of unity (m is a divisor of $p^f - 1$), and F is obtained by adjoining to $\mathbf{Q}_p$ a primitive $(p^f - 1)$th root. Then k is a totally ramified extension of F and is obtained by adjoining to F a root of an †Eisenstein polynomial (→ 337 Polynomials F).

C. The Topology of k

Taking $\mathfrak{p}^m$ $(m = 0, 1, 2, \ldots)$ as a †base for a neighborhood system of 0, k becomes a †locally compact †totally disconnected †topological field, and $\mathfrak{p}^m$ $(m = 0, 1, 2, \ldots)$ are compact subgroups of the additive group k. The multiplicative group $k^\times$ of nonzero elements of k is a locally compact †Abelian group, and the $\mathfrak{u}^{(m)} = \{\alpha \in \mathfrak{o} \mid \alpha \equiv 1 \pmod{\mathfrak{p}^m}\}$ $(m = 0, 1, 2, \ldots)$ form a base for the neighborhood system of 1. The

†character group in the sense of Pontryagin of the additive group k is isomorphic to k. This isomorphism is obtained by the following natural correspondence: For a p-adic field k, denote by φ the composition of the natural mapping of $\mathbf{Q}_p$ onto $\mathbf{Q}_p/\mathbf{Z}_p$ ($\cong \mathbf{Z}[1/p]/\mathbf{Z} \subset \mathbf{Q}/\mathbf{Z}$) and the †trace Tr from k to $\mathbf{Q}_p$, and put $\chi_x(y)=\exp(2\pi\sqrt{-1}\ \varphi(xy))$ $(y\in k)$; for the field k of the formal power series over a finite field κ, put $\chi_x(y)=\psi(xy)$ $(y\in k)$ with $\psi(\alpha)=\exp(2\pi\sqrt{-1}\cdot Tr(\operatorname{Res}\alpha)/p)$ $(\alpha\in k)$, where $\operatorname{Res}\alpha$ is the residue of $\alpha\in k$ and Tr is the trace from κ to $\mathbf{Z}/p\mathbf{Z}$. Then in both cases χ_x is a character of k, and $x\rightarrow\chi_x$ gives an isomorphism between k and the character group.

D. Ramification Theory

A valuation v of k has a unique †prolongation to an extension K of finite degree over k (we denote the prolongation also by v). K is complete under the valuation and is therefore a local field. Denoting by $\mathfrak{k}$ and κ the field of residue classes of K and k, respectively, we call $[\mathfrak{k}:\kappa]=f$ the **relative degree** of K/k, and $e=[v(K^\times):v(k^\times)]$ the **ramification index** of K/k. Then we have the equality $[K:k]=ef$. If $e=1$, we call K/k an **unramified extension**.

An unramified extension K/k is †normal, and its †Galois group is a cyclic group generated by the Frobenius automorphism, i.e., the element σ of the Galois group of K/k such that $\alpha^\sigma\equiv\alpha^{N(\mathfrak{p})} \pmod{\mathfrak{p}}$ for any element α in the valuation ring of K. For a given natural number f, there exists one and only one unramified extension of degree f over k in an algebraic closure of k.

Let K/k be a normal extension of finite degree with Galois group G, let Π be a †prime element of K, that is, a generator of the †valuation ideal $\mathfrak{P}$ of K, and put $V^{(i)}=\{\sigma\in G \mid \Pi^\sigma\equiv \Pi \pmod{\mathfrak{P}^{i+1}}\}$. Then $V^{(i)}$ is independent of the choice of Π. We call $V^{(0)}$ the **inertia group** and $V^{(i)}$ the ith **ramification group**. Then $V^{(i)}$ is normal in G, $[G:V^{(0)}]=f$, $[V^{(0)}:1]=e$, and $V^{(1)}$ is the p-Sylow subgroup of $V^{(0)}$. Furthermore, $G/V^{(0)}$ and $V^{(0)}/V^{(1)}$ are cyclic, and $V^{(i)}/V^{(i+1)}$ $(i=1,2,\dots)$ is an Abelian group of type $(p,p,\dots,p)$. Ramification theory for Abelian extensions is described in Section F.

For $x\in k$, the series $\exp(x)=\sum_{n=0}^{\infty}x^n/n!$ (resp. $\log(1+x)=\sum_{n=1}^{\infty}(-1)^{n-1}x^n/n$) converges for $\operatorname{ord}x>e/(p-1)$, where e is the ramification index of $k/\mathbf{Q}_p$ (resp. $\operatorname{ord}x>0$). The additive group $\mathfrak{p}^m$ and the multiplicative group $\mathfrak{u}^{(m)}$ $(m>e/(p-1))$ are isomorphic as topological groups under the mappings $x\rightarrow y=\exp(x)$, $y\rightarrow x=\log(y)$. If we fix an element $\pi\in k$ with $\operatorname{ord}\pi=1$, then an arbitrary element $x\in k$ with $\operatorname{ord}x=r$ is uniquely expressed in the form $x=\pi^r\zeta\alpha$, $\zeta^{p^f-1}=1$, $\alpha\in\mathfrak{u}^{(1)}$. The group $\mathfrak{u}^{(1)}$ is a multiplicative group on which $\mathbf{Z}_p$ operates, and the structure of $\mathfrak{u}^{(1)}$ as a $\mathbf{Z}_p$-group can be determined explicitly (→ [2, ch. II]).

E. Cohomology

For a normal extension K/k with Galois group G, we may consider the †cohomology groups $H^r(G,K^\times)$ $(r=1,2,\dots)$ of G operating on the multiplicative group $K^\times$. In particular, the 2-cohomology group $H^2(G,K^\times)$ is important in local class field theory and theory of algebras over k.

If C is the separable algebraic closure of k, i.e., the †maximal separable field over k in the †algebraic closure of k, and Γ is the Galois group of C/k, we can consider the 2-cohomology group $H^2(\Gamma,C^\times)$. Here we take as cocycles only those mappings $f(\sigma,\tau)$ of $\Gamma\times\Gamma$ into $C^\times$ that are continuous with respect to the †Krull topology of Γ and the discrete topology of $C^\times$. A fundamental theorem about the structure of $H^2(\Gamma,C^\times)$ states that the cocycles of $H^2(\Gamma,C^\times)$ that split in an extension K/k of degree n are exactly those cocycles that split in the unramified extension of degree n over k. Here a cocycle is said to **split** in K if it belongs to the †kernel of the homomorphism $H^2(\Gamma,C^\times)\overset{\text{res}}{\rightarrow}H^2(H,C^\times)$, where H is the subgroup of Γ corresponding to K, and res is the mapping obtained by restricting σ, τ of $f(\sigma,\tau)$ to the elements of H.

Let K/k be a normal extension of degree n with the Galois group G, and let H be the subgroup of Γ corresponding to K. Then the fundamental theorem combined with the exact sequence

$$0\rightarrow H^2(G,K^\times)\rightarrow H^2(\Gamma,C^\times)\rightarrow H^2(H,C^\times) \qquad (1)$$

known in the theory of Galois cohomology (→ 59 Class Field Theory H, 200 Homological Algebra I), yields $H^2(G,K^\times)\cong\mathbf{Z}/n\mathbf{Z}$ (cyclic group of order n), and $H^2(\Gamma,C^\times)\cong\mathbf{Q}/\mathbf{Z}$, where $\mathbf{Q}$ is the additive group of rational numbers and $\mathbf{Z}$ is the additive group of rational integers. These isomorphisms can be given in canonical form as follows: Denote the unramified extension of degree n by K_n/k, and the Frobenius automorphism of K_n/k by σ. Then an element c of $H^2(G_n,K_n^\times)$ (where G_n is the Galois group of K_n/k) is represented by the cocycle

$$f(\sigma^i,\sigma^j)=a^{[(i+j)/n]-[i/n]-[j/n]}, \quad i,j\in\mathbf{Z},$$

with $a\in k^\times$, and conversely, every $a\in k^\times$ determines an element c of $H^2(G_n,K_n^\times)$ in this manner. Under these conditions, the correspondence between c and a gives rise to an isomorphism $H^2(G_n,K_n^\times)\cong k^\times/N_{K_n/k}(K_n^\times)$.

Next define an element $\operatorname{inv} c$ of $\mathbf{Q}/\mathbf{Z}$ by $\operatorname{inv} c = (\operatorname{ord} a)/n \pmod{\mathbf{Z}}$. Then since $c \in H^2(\Gamma, C^\times)$ splits in an unramified extension of degree n by the fundamental theorem, the exact sequence (1) determines in a natural way an element c' of $H^2(G_n, K_n^\times)$ corresponding to c. Putting $\operatorname{inv} c = \operatorname{inv} c'$, we can show that $H^2(\Gamma, C^\times) \ni c \rightarrow \operatorname{inv} c$ gives an isomorphism $H^2(\Gamma, C^\times) \cong \mathbf{Q}/\mathbf{Z}$. We call $\operatorname{inv} c$ the **invariant** of $c \in H^2(\Gamma, C^\times)$. The invariant of an element of the cohomology group $H^2(G, K^\times)$ is defined to be the invariant of the corresponding element of $H^2(\Gamma, C^\times)$, which is determined by the exact sequence (1). Mapping an element of $H^2(G, K^\times)$ to its invariant, we obtain an isomorphism $H^2(G, K^\times) \cong \mathbf{Z}/n\mathbf{Z}$.

F. Local Class Field Theory

Let K/k be a normal extension of degree n with the Galois group G, let $f(\sigma, \tau)$ be a cocycle representing the element of $H^2(G, K^\times)$ with the invariant $1/n$, and put

$$\left(\frac{K/k}{\sigma}\right) = \prod_{\tau \in G} f(\tau, \sigma)^{-1}.$$

Then $\sigma \rightarrow \left(\frac{K/k}{\sigma}\right)$ gives an isomorphism between G/G' (where G' is the commutator subgroup of G) and $k^\times / N_{K/k}(K^\times)$. It follows from this that $[k^\times : N_{E/k}(E^\times)] \leqslant [E:k]$ for any extension E/k of finite degree, and the equality holds if and only if E/k is Abelian. The inverse mapping of $G \ni \sigma \rightarrow \left(\frac{K/k}{\sigma}\right) \in k^\times / N_{K/k}(K^\times)$ for an Abelian extension K/k is written as $k^\times \ni \alpha \rightarrow (\alpha, K/k) \in G$, and $(\alpha, K/k)$ is called the **norm-residue symbol**. If L/k is Abelian and K/k is a subfield of L, then the restriction of $(\alpha, L/k)$ to K coincides with $(\alpha, K/k)$. Let k_a be the maximal Abelian extension of k, i.e., the union of all Abelian extensions of finite degree over k. Then for any $\alpha \in k^\times$, an element (α, k) of the Galois group $G(k_a/k)$ of k_a/k is uniquely determined by $(\alpha, k)(\gamma) = (\alpha, k(\gamma)/k)(\gamma)$, $\gamma \in k_a$. The mapping $k^\times \ni \alpha \rightarrow (\alpha, k) \in G(k_a/k)$ is a one-to-one continuous homomorphism, and the image is †dense in $G(k_a/k)$. It has also been proved that there exists one and only one Abelian extension K/k with $N_{K/k}(K^\times) = A$ for any given closed subgroup A of finite index of $k^\times$. Therefore closed subgroups A of finite index of $k^\times$ are in one-to-one correspondence with finite Abelian extensions K of k through the relation $A = N_{K/k}(K^\times)$, and in this case A is called the subgroup of $k^\times$ corresponding to K/k.

Let K/k be an Abelian extension of finite degree, A be its corresponding subgroup of $k^\times$, and $V^{(i)}$ $(i = 0, 1, 2, \ldots)$ be ramification groups of K/k. Furthermore, define constants $v_1, v_2, \ldots, v_r$ by

$$V^{(0)} = V^{(1)} = \ldots = V^{(v_1)} \supsetneqq V^{(v_1+1)} = \ldots$$
$$= V^{(v_2)} \supsetneqq \ldots \supsetneqq V^{(v_{r-1}+1)} = \ldots = V^{(v_r)}$$
$$\supsetneqq V^{(v_r+1)} = (1),$$

denote the order of $V^{(v_i+1)}$ by n_i $(i = 1, 2, \ldots, r)$, and put $u_\rho = v_0 + (n_0/n_0)(v_1 - v_0) + \ldots + (n_{\rho-1}/n_0)(v_\rho - v_{\rho-1})$, $\rho = 1, 2, \ldots, r$ (here we understand $v_0 = -1$, $n_0 = [V^{(0)} : 1]$). Then $u_1, \ldots, u_r$ are rational integers, and we have

$$A\mathfrak{u}^{(0)} = \ldots = A\mathfrak{u}^{(u_1)} \supsetneqq A\mathfrak{u}^{(u_1+1)} = \ldots$$
$$= A\mathfrak{u}^{(u_2)} \supsetneqq \ldots \supsetneqq A\mathfrak{u}^{(u_r+1)} = A$$

(H. Hasse). If m is the smallest integer with $A \supseteq \mathfrak{u}^{(m)}$, then $\mathfrak{p}^m$ is called the **conductor** of K/k. The above results of Hasse show that $m = u_r + 1$. On the other hand, it is known that the correspondence between $A\mathfrak{u}^{(u_\rho)}$ and $V^{(v_\rho)}$ $(\rho = 1, 2, \ldots, r)$ is given by the norm-residue symbol $(\alpha, K/k)$ (→ 59 Class Field Theory).

G. Theory of Algebras

By the general theory of †crossed products of algebras, the structure of the †Brauer group formed by the classes of †normal simple algebras over a local field k is obtained directly from results concerning cohomology (→ Section F). Namely, a †normal simple algebra over k splits over a separable extension of degree n if and only if it splits over the unramified extension of degree n, and the Brauer group of k is isomorphic to $\mathbf{Q}/\mathbf{Z}$. Furthermore, the †exponent (the order as an element of the Brauer group) of a normal simple algebra $\mathfrak{A}$ over k coincides with the †Schur index, and if $[\mathfrak{A} : k] = n^2$, then $\mathfrak{A}$ is expressed as a crossed product with respect to any normal extension of degree n over k. The invariant of the factor set (2-cocycle) that appears in this crossed product expression is called the **invariant** of $\mathfrak{A}$ (→ 29 Associative Algebras; for the properties of $\mathfrak{A}$ as a topological ring and as a topological group of the group of invertible elements of $\mathfrak{A}$ → 6 Adeles and Ideles).

H. Explicit Formulas

Let K/k be an Abelian extension of finite degree. When we have an explicit formula for the norm-residue symbol $(\alpha, K/k)$, we say that we have an explicit reciprocity law.

Let k be a $\mathfrak{p}$-adic number field containing a primitive mth root ζ_m of unity, and let β be an element in $k^\times$. Let p be a prime number contained in $\mathfrak{p}$. Since the Kummer extension $K = k(\sqrt[m]{\beta})$ over k is Abelian, the Hilbert norm-residue symbol $(\alpha, \beta)_m$ is defined by

$(\alpha, K/k)(\sqrt[m]{\beta}) = (\alpha, \beta)_m \sqrt[m]{\beta}$, where $(\alpha, \beta)_m$ is an mth root of unity. In this case, the problem of obtaining an explicit reciprocity law is solved if we can express the symbol $(\alpha, \beta)_m$ in terms of α, β and suitable parameters depending on the ground field k.

In particular, if $p=2$, $m=2$, we have the simple formula given by Hasse, $(\alpha, \beta)_2 = (-1)^{Tr((\alpha-1)(\beta-1)/4)}$, where α, β are two units in k satisfying $\alpha \equiv \beta \equiv 1 \pmod 2$ and Tr is the trace from k to $\mathbf{Q}_p$. Similar formulas for the complementary laws are also known [10].

On the other hand, if $m=p$ is an odd prime and $k=\mathbf{Q}_p(\zeta_p)$, we have the following formulas for a prime element $\lambda_p = 1-\zeta_p$ and two units α, β satisfying $\alpha \equiv 1 \pmod{p^2}$, $\beta \equiv 1 \pmod p$:

$$(\alpha, \beta)_p = \zeta_p^{(1/p)Tr(\zeta_p \log\alpha \cdot D\log\beta)}, \qquad (2)$$

$$(\zeta_p, \beta)_p = \zeta_p^{(1/p)Tr(\log\beta)}, \qquad (3)$$

$$(\lambda_p, \beta)_p = \zeta_p^{-(1/p)Tr((\zeta_p/\lambda_p)\log\beta)}, \qquad (4)$$

where $D\log\beta = (1/\beta)\sum_{i=1}^{\infty} i b_i \lambda_p^{i-1}$, while the b_i are determined by the λ_p-expansion $\beta = \sum_{i=0}^{\infty} b_i \lambda_p^i$, $b_i \in \mathbf{Z}_p$ [5]. Furthermore, we have an explicit Kummer-Hilbert formula deduced from (2) in terms of Kummer's logarithmic differential quotients [8, 10]. Concerning the complementary laws (3), (4), the following Artin-Hasse formulas are known for $k = \mathbf{Q}_p(\zeta_{p^n})$ and $\beta \equiv 1 \pmod p$:

$$(\zeta_{p^n}, \beta)_{p^n} = \zeta_{p^n}^{(1/p^n)Tr(\log\beta)}, \qquad (3')$$

$$(\lambda_{p^n}, \beta)_{p^n} = \zeta_{p^n}^{-(1/p^n)Tr((\zeta_{p^n}/\lambda_{p^n})\log\beta)}, \qquad (4')$$

where $\lambda_{p^n} = 1-\zeta_{p^n}$ [9]. Utilizing these formulas, K. Iwasawa obtained a formula for $(\alpha, \beta)_{p^n}$ that is a natural generalization of (2) [13]. A. Wiles has shown that a generalization of the Artin-Hasse-Iwasawa formulas can be obtained in terms of the Lubin-Tate formal groups (*Ann. Math.*, (2) 107 (1978)).

Concerning $(\alpha, \beta)_{p^n}$, we have **Shafarevich's reciprocity law** [11]. To explain it, let k_0 be the inertia field in k, i.e., the maximal subfield in k that is unramified over $\mathbf{Q}_p$. For an arbitrary integer a in k_0, we consider the **Artin-Hasse function** $E(a,x) = \exp(-L(a,x))$, where $L(a,x) = \sum_{i=0}^{\infty}(a^{\sigma_i}/p^i)x^{p^i}$ with the †Frobenius automorphism σ of $k_0/\mathbf{Q}_p$. We choose a prime element $\tilde{\pi}$ in k_0 such that $\zeta_{p^n} = E(1, \tilde{\pi})$ and an integer $\bar{a}$ in a suitable unramified extension field of k_0 such that $\bar{a}^\sigma - \bar{a} = a$ for a given integer a in k_0. Given any p^n-primary element x in k (i.e., $k(x^{1/p^n})$ is unramified over k), there exists an integer a in k_0 such that $x \approx E(a) = E(p^n\bar{a}, \tilde{\pi})$, where $x \approx y$ ($x, y \in k^\times$) means $x = y \cdot u$ for an element u in $k^{\times p^n}$. Furthermore, if δ is an element of $k^\times$, we have the canonical decomposition formula

$$\delta \approx \pi^{d^*} E(d) \prod_{\substack{1 \leqslant j < e_0 p \\ (j,p)=1}} E(d_j, \pi^j),$$

where π is a prime element in k, $d^* \in \mathbf{Z}$, d, d_i are integers in k_0, and $e_0 = e/(p-1)$ with the ramification index e of k. The explicit formula due to Shafarevich and Hasse is expressed as

$$(\alpha, \beta)_{p^n} = \zeta_{p^n}^{Tr_0(a^* b - a b^* + c)},$$

where Tr_0 is the trace from k_0 to $\mathbf{Q}_p$ and a, a^* (resp. b, b^*) are determined by the canonical decompositions of α (resp. β) as stated above. The element c in k_0 is determined by

$$\prod_{\substack{1 \leqslant i,j < e_0 p \\ (i,p)=(j,p)=1}} E(i a_i b_j, \pi^{i+j}) = \gamma \approx E(c) \prod_{\substack{1 \leqslant j < e_0 p \\ (j,p)=1}} E(c_j, \pi^j)$$

for odd p, and by

$$(-1)^{a^* b^*} \prod_{\substack{1 \leqslant i,j < 2e_0 \\ (i,2)=(j,2)=1}} E(i a_i b_j, \pi^{i+j})$$
$$\times \prod_{\mu, \nu=1}^{\infty} E((i2^{\mu-1} + j2^{\nu-1}) a_i^{\sigma^\mu} b_j^{\sigma^\nu}, \pi^{i2^\mu + j2^\nu})$$
$$= \gamma \approx E(c) \prod_{\substack{1 \leqslant j < 2e_0 \\ (j,2)=1}} E(c_j, \pi^j)$$

for $p=2$. Several formulas for the case where k is a function field are also known [1, 5].

Let k be a general local field. When K is an extension field of k obtained by adjoining π^n-division points $\{\lambda\}$ of the Lubin-Tate formal group F defined over the integer ring of k, the extension K/k is totally ramified and Abelian, and we have an explicit formula: $(u^{-1}, K/k)\lambda = [u]_F(\lambda)$, where u is a unit in k and $[u]_F$ is an endomorphism of F corresponding to the unit u. In particular, this formula implies the cyclotomic reciprocity law [12].

The problem of obtaining an explicit reciprocity law arises in the problem of obtaining the reciprocity law for power residues from the law of reciprocity in global class field theory. The problem is closely connected with T. Kubota's recent results (e.g., [14]), which clarify the analytic meaning of the reciprocity law in algebraic number fields.

References

[1] E. Artin, Algebraic numbers and algebraic functions, Lecture notes, Princeton Univ., 1950–1951 (Gordon & Breach, 1967).
[2] H. Hasse, Zahlentheorie, Akademie-Verlag, third edition, 1969; English translation, Number theory, Springer, 1980.
[3] H. Hasse, Normenresttheorie galoisscher Zahlkörper mit Anwendungen auf Führer und Diskriminante abelscher Zahlkörper, J. Fac. Sci. Univ. Tokyo, 2 (1934), 477–498.
[4] M. Deuring, Algebren, Erg. Math., Springer, second edition, 1968.
[5] E. Artin and J. Tate, Class field theory, Lecture notes, Princeton Univ., 1951 (Benjamin, 1967).

[6] J.-P. Serre, Corps locaux, Actualités Sci. Ind., Hermann, 1962; English translation, Local fields, Springer, 1979.
[7] O. F. G. Schilling, The theory of valuations, Amer. Math. Soc. Math. Surveys, 1950.
[8] T. Takagi, On the law of reciprocity in the cyclotomic corpus, Proc. Phys.-Math. Soc. Japan, 4 (1922), 173–182.
[9] E. Artin and H. Hasse, Die beiden Ergänzungssätze zum Reziprozitätsgesetz der l^n-ten Potenzreste im Körper der l^n-ten Einheitswurzeln, Abh. Math. Sem. Univ. Hamburg, 6 (1928), 146–162.
[10] H. Hasse, Bericht über neuere Untersuchungen und Probleme aus der Theorie der algebraischen Zahlkörper II, Jber. Deutsch. Math. Verein. (1930) (Physica Verlag, 1965).
[11] I. R. Shafarevich, A general reciprocity law, Amer. Math. Soc. Transl., 4 (1950), 73–106. (Original in Russian, 1950.)
[12] J. Lubin and J. Tate, Formal complex multiplication in local fields, Ann. Math., (2) 81 (1965), 380–387.
[13] K. Iwasawa, On explicit formulas for the norm residue symbol, J. Math. Soc. Japan, 20 (1968), 151–165.
[14] T. Kubota, Ein arithmetischer Satz über eine Matrizengruppe, J. Reine Angew. Math., 222 (1966), 55–57.
[15] J. W. S. Cassels and A. Frölich (eds.), Algebraic number theory, Academic Press, 1967.

258 (XX.24) Lorentz Group

A. Lorentz Group

(1) Minkowski space. A 4-dimensional real vector space M with an indefinite inner product between two vectors x and y defined by
$x\cdot y=xG{}^ty=x^0y^0-x^1y^1-x^2y^2-x^3y^3$,

$$G=\begin{pmatrix}1&0&0&0\\0&-1&0&0\\0&0&-1&0\\0&0&0&-1\end{pmatrix}=(g_{\mu\nu}),$$

is called a **Minkowski space**. A vector $x\in M$ is classified according to the signs of $x\cdot x$ and x^0 as follows.

$x\cdot x>0,\ x^0>0$: future timelike } **timelike**,
$x\cdot x>0,\ x^0<0$: past timelike }

$x\cdot x=0,\ x^0>0$: future lightlike }
$x\cdot x=0,\ x^0=0$: origin } **lightlike**,
$x\cdot x=0,\ x^0<0$: past lightlike }

$x\cdot x<0$: **spacelike**.

The set of all future (or past) timelike vectors is called the **future** (or **past**) **cone** and is denoted by V_+ (or V_-). The set of lightlike vectors is called the **light cone**. The set of spacelike vectors is called the **side cone**.

Minkowski space is used in the †special theory of relativity (with units such that the velocity of light is 1), where $\{(x-y)\cdot(x-y)\}^{1/2}$ is called the **proper time** between two mutually timelike events x and y.

(2) Inhomogeneous Lorentz group. The group of all one-to-one mappings of the Minkowski space onto itself preserving the proper time between any timelike pair of points is called the **full inhomogeneous Lorentz group** or **Poincaré group** and is denoted by $\mathscr{P}$. An element g of $\mathscr{P}$ is a linear transformation

$$(gx)^\mu=\sum_{\nu=0}^{3}\Lambda^\mu_\nu x^\nu+a^\mu\qquad(\mu=0,1,2,3),$$

where $a\in M$ and $\Lambda=(\Lambda^\mu_\nu)$ is any real matrix satisfying $\Lambda G({}^t\Lambda)=G$. We write $g=(a,\Lambda)$ and $gx=\Lambda x+a$. Then

$$(a_1,\Lambda_1)(a_2,\Lambda_2)=(a_1+\Lambda_1a_2,\Lambda_1\Lambda_2).$$

The normal subgroup of $\mathscr{P}$ consisting of $(a,1)$, $a\in M$, is called the **translation group** on M. The subgroup of $\mathscr{P}$ consisting of $(0,\Lambda)$ is called the **full homogeneous Lorentz group** and is denoted by $\mathscr{L}$ or $O(1,3)$. It consists of the following four connected components:

$$\mathscr{L}^\uparrow_+=\{\Lambda\in\mathscr{L}\mid\det\Lambda=1,\Lambda V_+=V_+\},$$
$$\mathscr{L}^\downarrow_+=\{\Lambda\in\mathscr{L}\mid\det\Lambda=1,\Lambda V_+=V_-\},$$
$$\mathscr{L}^\uparrow_-=\{\Lambda\in\mathscr{L}\mid\det\Lambda=-1,\Lambda V_+=V_+\},$$
$$\mathscr{L}^\downarrow_-=\{\Lambda\in\mathscr{L}\mid\det\Lambda=-1,\Lambda V_+=V_-\}.$$

The identity mapping, the **space-time inversion** $x\to-x$, the **space inversion** $x^0\to x^0$, $x^i\to-x^i$ $(i=1,2,3)$, and the **time reversal** $x^0\to-x^0$, $x^i\to x^i$ are their typical elements, respectively. If $\det\Lambda=1$, $\Lambda\in\mathscr{L}$ is called **proper**. If $\Lambda V_+=V_+$, $\Lambda\in L$ is called **orthochronous**. The identity component $\mathscr{L}^\uparrow_+$ is called the **restricted** homogeneous Lorentz group. (This usage in mathematical physics may differ from the general terminology for $O(m,n)$, in which the identity component is called proper.)

The same terminology is used for $\mathscr{P}$, which also consists of four components $\mathscr{P}^\uparrow_+$, $\mathscr{P}^\downarrow_+$, $\mathscr{P}^\uparrow_-$, and $\mathscr{P}^\downarrow_-$.

(3) Universal covering group. For $x\in M$, let

$$\tilde{x}=\sum_{\mu=0}^{3}x^\mu\sigma_\mu,$$

$$\sigma_0=\begin{pmatrix}1&0\\0&1\end{pmatrix},\quad\sigma_1=\begin{pmatrix}0&1\\1&0\end{pmatrix},\quad\sigma_2=\begin{pmatrix}0&-i\\i&0\end{pmatrix},$$

$$\sigma_3=\begin{pmatrix}1&0\\0&-1\end{pmatrix},$$

where $\sigma=(\sigma_1,\sigma_2,\sigma_3)$ is called the **Pauli spin matrix**. For any 2×2 matrix A with determinant 1 (i.e., $A\in SL(2,\mathbf{C})$),

$$A\tilde{x}A^*=\tilde{x}_A, \qquad x_A=\Lambda(A)x,$$

defines $\Lambda(A)\in\mathscr{L}_+^\uparrow$, and $A\to\Lambda(A)$ is a two-to-one mapping from $SL(2,\mathbf{C})$ onto $\mathscr{L}_+^\uparrow$. By this covering mapping, $SL(2,\mathbf{C})$ becomes the universal covering group for $\mathscr{L}_+^\uparrow$.

If A is unitary (i.e., $A\in SU(2)$), $\Lambda(A)$ leaves x^0 invariant, and $SU(2)$ is the universal covering group of $\mathscr{O}(3)_+$ (the 3-dimensional **proper orthogonal group** or **proper rotation group**).

For a complex vector $z\in M+iM$,

$$A\tilde{z}^+B=\tilde{z}_{A,B}, \qquad z_{A,B}=\Lambda(A,B)z$$

defines a complex Lorentz transformation $\Lambda(A,B)$, and $(A,B)\to\Lambda(A,B)$ gives a covering mapping from $SL(2,\mathbf{C})\times SL(2,\mathbf{C})$ onto the **proper complex Lorentz group** $\mathscr{L}(\mathbf{C})_+$, consisting of complex 4×4 matrices Λ satisfying $\Lambda G({}^t\Lambda)=G$ and $\det\Lambda=1$.

B. Finite-Dimensional Representations

(1) $\mathscr{L}_+^\uparrow$. Any continuous representation of $SL(2,\mathbf{C})$ on a finite-dimensional complex vector space is a direct sum of irreducible representations.

Let $E=\mathbf{C}^2$ be the 2-dimensional complex vector space on which $SL(2,\mathbf{C})$ is acting. Let a representation $\pi_{k,n}$ of $SL(2,\mathbf{C})$ on the $(k+n)$-fold tensor product $E^{\otimes(k+n)}$ be

$$\pi_{k,n}(A)=A^{\otimes k}\otimes(\bar{A})^{\otimes n},$$

where $\bar{A}$ is the complex conjugate of A. A vector ξ in this representation space is called a **mixed spinor** of rank (k,n) and its components are written as $\xi^{\lambda_1\ldots\lambda_k\dot{\mu}_1\ldots\dot{\mu}_n}$, with **undotted indices** $\lambda_1,\ldots,\lambda_k$ and **dotted indices** $\dot{\mu}_1\ldots\dot{\mu}_n$ all taking values 1 or 2. If $n=0$ (or $k=0$), it is called an **undotted** (or **dotted**) spinor of rank k (or n). A spinor with upper indices, called a **contravariant spinor**, can be converted to a spinor with lower indices (partially or totally), called a **covariant spinor**, by

$$\varepsilon=(\varepsilon_{\lambda\mu})=\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}.$$

For example,

$$\xi^{\lambda_1}{}_{\lambda_2}{}^{\dot{\mu}_1}{}_{\dot{\mu}_2}=\sum_{\kappa,\dot{\nu}}\xi^{\lambda_1\kappa\dot{\mu}_1\dot{\nu}}\varepsilon_{\kappa\lambda_2}\varepsilon_{\dot{\nu}\dot{\mu}_2}.$$

The restriction of $\pi_{k,n}$ to the subspace of spinors which are invariant under permutations of undotted indices and under those of dotted indices is an irreducible representation of dimension $(k+1)(n+1)$, which is denoted by $[k,n]$. The set of $[k,n]$, $k=0,1,\ldots,n=0,1,\ldots$, exhausts all finite-dimensional irreducible representations of $SL(2,\mathbf{C})$.

The tensor product of two irreducible representations is decomposed as follows:

$$[k_1,n_1]\otimes[k_2,n_2]=\sum\nolimits^{\oplus}[k,n],$$

where $[k,n]$ appears with multiplicity 1 for k, n satisfying $k_1+k_2\geqslant k\geqslant|k_1-k_2|$, $n_1+n_2\geqslant n\geqslant|n_1-n_2|$. The complex conjugate representation of $[k,n]$ is $[n,k]$.

For $x\in M$, $\tilde{x}$ is a mixed spinor of rank $(1,1)$.

(2) $SU(2)$. The restriction of $[k,0]$ $(k=0,1,\ldots,)$ to the subgroup $SU(2)$ is an irreducible representation (called the **spinor representation of rank** k) and exhausts all irreducible representations of $SU(2)$. $[0,n]$ is equivalent to $[n,0]$ for $SU(2)$ due to $\bar{A}=\varepsilon A\varepsilon^{-1}$ for $A\in SU(2)$. $[k,n]$ for $SU(2)$ is equivalent to

$$[k,0]\otimes[n,0]=\sum_{r=|k-n|}^{k+n}{}^{\oplus}[r,0].$$

An explicit coefficient in this decomposition is called a †**Clebsch-Gordan coefficient**.

Any representation of $\mathscr{O}(3)_+$ is a representation of $SU(2)$. The irreducible representation of $SU(2)$ with rank r is an irreducible representation of $\mathscr{O}(3)_+$ if r is even. It is an irreducible **double-valued representation** of $\mathscr{O}(3)_+$ if r is odd. Likewise, the irreducible representation of $SL(2,\mathbf{C})$ with rank (k,n) is an irreducible representation of $\mathscr{L}_+^\uparrow$ if $k+n$ is even and double-valued if $k+n$ is odd.

(3) $\mathscr{L}^\uparrow=\mathscr{L}_+^\uparrow\cup\mathscr{L}_-^\uparrow$. The space inversion P induces an automorphism of $\mathscr{L}_+^\uparrow$ by $\Lambda\to P\Lambda P$, that of $SL(2,\mathbf{C})$ by $A\to\varepsilon\bar{A}\varepsilon^{-1}$ and the associated mapping $U(A)\to V(A)=U(\varepsilon\bar{A}\varepsilon^{-1})$ of irreducible representations $[k,n]\to[n,k]$. Thus an irreducible representation of $\mathscr{L}^\uparrow$ is obtained on $[k,n]\oplus[n,k]$ if $n\neq k$, and two inequivalent irreducible representations on $[n,n]$ corresponding to two choices (± 1) for the operator representing P. A vector in $[k,n]\oplus[n,k]$ is called a **bispinor of rank** (k,n). The wave function of the Dirac equation is a bispinor of rank $(1,0)$.

C. Unitary Representations

(1) Invariance in quantum mechanics. In the special theory of relativity, each $g\in\mathscr{P}_+^\uparrow$ represents a symmetry of the system, namely, a transformation of states ψ of the system to states $g\psi$ such that "physical relations" between two states ψ_1 and ψ_2 should be the same as those between $g\psi_1$ and $g\psi_2$. In quantum mechanics, a (pure) state is represented by a unit ray (i.e., a set of vectors $e^{i\theta}\Psi$, $0\leqslant\theta<2\pi$ for a unit vector Ψ) in a Hilbert space, and $|(\Psi_1,\Psi_2)|^2$ is supposed to be an observable quantity, called the transition probability. Thus the special theory of relativity in quantum-mechanical situation implies a continuous representation of $\mathscr{P}_+^\uparrow$ by bijective map-

pings of unit rays in a Hilbert space preserving the transition probability. By the **Wigner theorem**, such a mapping is implemented by either a unitary or antiunitary operator $U(g)$ as a mapping $\{e^{i\theta}\Psi\}\to\{e^{i\theta}U(g)\Psi\}$ (→ e.g., [12]). In order for this to be a group of transformations of unit rays, U should be a projective representation: $U(g_1)U(g_2)=u(g_1,g_2)U(g_1g_2)$, where $u(g_1,g_2)$ is a 2-cocycle with a complex value of modulus 1. In the case of $\mathscr{P}_+^\uparrow$, each $U(g)$ has to be unitary (true for any connected Lie group), and there is a choice of $U(a,A)$ from the unitary ray $\{e^{i\theta}U(a,\Lambda(A))|0\leqslant\theta<2\pi\}$ so as to make $U(a,A)$ $(a\in M, A\in SL(2,\mathbf{C}))$ a continuous unitary representation of the universal covering group $\tilde{\mathscr{P}}_+^\uparrow=\{(a,A)|a\in M,\ A\in SL(2,\mathbf{C})\}$ of $\mathscr{P}_+^\uparrow$ [2].

(2) $\tilde{\mathscr{P}}_+^\uparrow$. Any continuous unitary representation of $\tilde{\mathscr{P}}$ on a separable Hilbert space is a direct integral of continuous irreducible unitary representations of $\tilde{\mathscr{P}}_+^\uparrow$ (called irreducible representations in the following). Any continuous unitary representation of the translation group $\{(a,1)|a\in M\}$, the group being commutative, is of the form

$$(U(a,1)\Psi)(p)=e^{ia\cdot p}\Psi(p),\qquad \Psi\in\int H(p)\,d\mu(p).$$

The parameter $p\in M$ is called the **energy-momentum 4-vector** or **4-momentum**. In an irreducible representation of $\tilde{\mathscr{P}}_+^\uparrow$, the measure $d\mu$ is equivalent to an invariant measure on an orbit of the conjugacy action $g(a,1)g^{-1}=(\Lambda(A)a,1)$, $g=(b,A)\in\tilde{\mathscr{P}}_+^\uparrow$, i.e., on one of the following orbits.

$m_\pm$: $p^2\equiv p\cdot p=m^2\quad(m>0),\quad \pm p^0>0,$

$0_\pm$: $p^2=0,\quad \pm p^0>0,$

0: $p=0,$

im: $p^2=-m^2\quad(m>0).$

The parameter $(p^2)^{1/2}$ is called the **mass**.

For each orbit the subgroup of all $A\in SL(2,\mathbf{C})$ that do not move a predetermined point q_λ on the orbit λ, is called the **little group** (of λ or of q_λ). Up to isomorphism, the little group contains the following.

$m_\pm$: $SU(2)$, consisting of all unitary $A\in SL(2,\mathbf{C})$, which leaves $q_{m_\pm}=(\pm m,0,0,0)$ invariant.

$0_\pm$: The 2-fold covering group of the 2-dimensional Euclidean group, consisting of all

$$A(\theta,z)=\begin{pmatrix}e^{i\theta} & e^{-i\theta}z\\ 0 & e^{-i\theta}\end{pmatrix},\quad z\in\mathbf{C},\quad \theta\in\mathbf{R}$$

$(q_{0_\pm}=(\pm1,0,0,\pm1))$, with the product

$$A(\theta_1,z_1)A(\theta_2,z_2)=A(\theta_1+\theta_2,z_1+e^{2i\theta_1}z_2).$$

0: $SL(2,\mathbf{C})$.

im: $SL(2,\mathbf{C})$ consisting of all real $A\in SL(2,\mathbf{C})$, which leaves $q_{im}=(0,0,m,0)$ invariant.

Any irreducible representation of $\tilde{\mathscr{P}}_+^\uparrow$ is uniquely determined (up to unitary equivalence) by an orbit λ and an irreducible representation D of the little group $G_\lambda\subset SL(2,\mathbf{C})$ on a Hilbert space K as follows (as an †induced representation):

$$\Psi=\int\Psi(p)\,d\mu(p)\in H=\int H(p)\,d\mu(p),$$

$$[U(a,A)\Psi](p)=e^{ia\cdot p}D(R(A,p))\Psi(\Lambda(A)^{-1}p),$$

$$R(A,p)=L(p)^{-1}AL(\Lambda(A^{-1})p),$$

where $H(p)=K$, $d\mu(p)$ is a measure (unique up to a multiplicative constant) on λ, invariant under the action of $\mathscr{L}_+^\uparrow$ on λ $(p\to\Lambda p,\Lambda\in\mathscr{L}_+^\uparrow)$, $L(p)$ is a fixed element of $SL(2,\mathbf{C})$ for each $p\in\lambda$ such that $\Lambda(L(p))q_\lambda=p$, and $R(A,p)$ is called the **Wigner rotation**.

(3) The orbit $m_\pm$. The irreducible representation with the orbit $m_\pm$ and the irreducible representation of $SU(2)$ with rank $2j$ $(j=0,1/2,1,\dots)$ is denoted by $[m_\pm,j]$, where j is called the **spin**. It is given by

$$[U(a,A)\Psi](p)=e^{ip\cdot a}A^{\otimes 2j}\Psi(\Lambda(A)^{-1}p),$$

$$(\Psi,\Phi)=\int(\Psi(p),(m/\pm\tilde{p})^{\otimes 2j}\Phi(p))\times(\pm2p^0)^{-1}\,dp^1\,dp^2\,dp^3,$$

where $\Psi(p)\in\mathbf{C}^{2j}$ for each p and p is restricted to the orbit $m_\pm$.

(4) The orbit $0_\pm$. Irreducible representations of the little group can be classified by the orbit in the spectrum of the normal Abelian subgroup consisting of $A(0,z)$, $z\in\mathbf{C}$. Since the spectrum is a plane and $A(\theta,z)$ acts as a rotation by an angle 2θ, an orbit is a circle of radius ρ. The case $\rho=0$ is further classified completely by the spin $j=0,\ \pm1/2,\ \pm1,\dots$ (used for the description of a massless particle of spin $|j|$ and **helicity** $\sigma=\operatorname{sign}j$). It is denoted by $[0_\pm,j]$ and is given by

$$[U(a,A)\Psi](p)=e^{ip\cdot a}(A^\sigma)^{\otimes 2|j|}\Psi(\Lambda(A)^{-1}p),$$

$$(\Psi,\Phi)=\int(\Psi(p),(\pm\tilde{p}^\sigma)^{\otimes 2|j|}\Phi(p))\times(\pm2p^0)^{-1}\,dp^1\,dp^2\,dp^3,$$

where $A^\sigma=A$ or $\bar{A}$ and $\tilde{p}^\sigma=\varepsilon^t\tilde{p}\varepsilon^{-1}$ or $\tilde{p}$ depending on $\sigma=1$ or -1, and $\Psi(p)$ belongs to the quotient of $\mathbf{C}^n$ and the subspace consisting of all χ with $(\chi,(\tilde{p}^\sigma)^{\otimes 2|j|}\chi)=0$ (the quotient is of one dimension).

The case $\rho\neq0$ (called **continuous spin**) is completely classified by ρ and $U(0,-1)=\pm1$.

The representation D of the little group for $\rho \neq 0$ is given by

$$[D(A(z,\theta))\Psi](\varphi)$$
$$=\exp i\{\rho \operatorname{Re} e^{-i\varphi} z + k\theta\}\Psi(\varphi - 2\theta),$$

where $k=0$ or 1 (accordingly $U(0,-1)=1$ or -1), $\varphi \in \mathbf{R} \bmod 2\pi$ and $\Psi \in L_2([0,2\pi], d\varphi)$.

(5) The orbit 0. Irreducible representations of the little group $SL(2,\mathbf{C})$ are as follows: The **principal series** with two parameters $-\infty < \rho < \infty$ and $m = 0, \pm 1, \ldots$ are given by

$$[D(A)f](z) = (bz+d)^m |bz+d|^{2i\rho - m - 2} f({}^tAz),$$
$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$
$${}^tAz = (az+c)/(bz+d),$$

where f is an L_2-function of $z = x+iy$ with respect to $dx\,dy$. The **supplementary series** with a parameter $0 < \rho < 1$ is given by

$$[D(A)f](z) = |bz+d|^{2\rho-2} f({}^tAz),$$
$$(f_1, f_2) = \int \overline{f_1(z_1)} f_2(z_2) \times |z_1 - z_2|^{-2-2\rho} dx_1\,dx_2\,dy_1\,dy_2.$$

Together with the identity representation $D(A) = 1$, they exhaust all possibilities up to unitary equivalence.

(6) The orbit im. Irreducible representations of the little group $SL(2,\mathbf{R})$ are as follows: The **principal continuous series** with parameters $\sigma = 0, 1$, $\rho \geqslant 0$ for $\sigma = 0$ and $\rho > 0$ for $\sigma = 1$ are given by

$$[D(A)f](x) = f({}^tAx)|bx+d|^{i\rho-1}(\operatorname{sign}(bx+d))^\sigma,$$

where $f \in L_2((-\infty,\infty), dx)$. The **supplementary series** with a parameter $0 \leqslant \rho < 1$ are given by

$$[D(A)f](x) = |bx+d|^{-1-\rho} f({}^tAx),$$
$$(f_1, f_2) = \int\int \overline{f(x_1)} f(x_2) |x_1 - x_2|^{-1-\rho} dx_1\,dx_2.$$

The **principal discrete series** with a parameter $n = 1, 2, 3, \ldots$ are given by

$$[D(A)f](z) = (bz+d)^{-n} f({}^tAz),$$
$$(f_1, f_2) = \int_{-\infty}^{\infty} dx \int y^{n-2} dy \overline{f_1(z)} f_2(z),$$

where f is holomorphic in either the upper or lower half-plane, (so that the y-integration is over $(0,\infty)$ or $(-\infty, 0)$ accordingly), $z = x + iy$, and the two choices give rise to two inequivalent representations for each n. Together with the identity representation $D(A) = 1$, they exhaust all possibilities up to unitary equivalence.

D. Infinitesimal Generators

In any continuous unitary representation U of $\mathscr{P}_+^\uparrow$,

$$i^{-1} \lim_{t\to\infty} t^{-1}(U(\gamma(t)) - 1)$$

(defined on the dense set of all those vectors on which the limit exist) for any one-parameter subgroup $\gamma(t)$ of $\mathscr{P}_+^\uparrow$ is a self-adjoint operator called the infinitesimal generator for $\gamma(t)$. If $\gamma(t) = (ta, 1)$ $(a \in M)$, the generator is $P \cdot a$, where the components P^0, P^1, P^2, P^3 of P mutually commute and are called **energy-momentum operators** (or **4-momentum operators**). If $\gamma(t) = (0, \exp(it\,\sigma_k/2))$, $k = 1, 2, 3$ (the rotation by an angle $-t$ around the kth axis), the generator is written as $J^k = M^{jl} = -M^{lj}$ ((jlk) is a cyclic permutation of (123)) and is called the **angular momentum operator**. If $\gamma(t) = (0, \exp(t\sigma_k/2))$, $k = 1, 2, 3$ (the pure Lorentz transformation along the kth axis with a velocity $\tanh t$), it is written as $N^k = M^{0k} = -M^{k0}$, and $M^{\mu\nu}$ $(\mu, \nu = 0, 1, 2, 3)$ is called the **angular momentum tensor**. The value of the scalar $P \cdot P$ is the square of the mass and, when $P \cdot P \leqslant 0$, the alternatives of P^0 being positive or negative definite or zero give the invariants of irreducible representations, distinguishing different orbits. The angular momentum in the center of mass coordinates,

$$w^\kappa = 2^{-1} \sum \varepsilon^{\kappa\lambda\mu\nu} P_\lambda M_{\mu\nu},$$

gives another invariant $w \cdot w$, called the **Pauli-Lubanski vector**, where the quantity with lower indices are obtained from those with upper indices by contraction with the Minkowski metric g, e.g., $P_\nu = \sum_\mu P^\mu g_{\mu\nu}$; ε is totally antisymmetric in its indices and $\varepsilon^{0123} = 1$. If $P \cdot P > 0$, then $w \cdot w = -j(j+1) P \cdot P$ defines the spin j. In the representation with orbit $0_\pm$ and $\rho = 0$, $w^\mu = \sigma j P^\mu$ defines the spin $j \geqslant 0$ and the helicity $\sigma = \pm 1$.

References

[1] V. Bargmann, Irreducible unitary representations of the Lorentz group, Ann. Math., 48 (1947), 568–640.
[2] V. Bargmann, On unitary ray representations of continuous groups, Ann. Math., 59 (1954), 1–46.
[3] C. Chevalley, The algebraic theory of spinors, Columbia Univ. Press, 1954.
[4] I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, Representations of the rotation and Lorentz groups and their applications, Pergamon Press, 1963. (Original in Russian, 1958.)
[5] I. M. Gel'fand and M. A. Naĭmark, Uni-

tary representations of the Lorentz group, J. Phys., 10 (1946), 93–94; Izv. Akad. Nauk SSSR, 11 (1947), 411–504.
[6] M. A. Naĭmark, Linear representations of the Lorentz group, Pergamon Press, 1964. (Original in Russian, 1958.)
[7] M. Schaaf, The reduction of the product of two irreducible unitary representations of the proper orthochronous quantum-mechanical Poincaré group, Lecture notes in physics 5, Springer, 1970.
[8] A. S. Wightman, L'invariance dans la mecanique quantique relativiste, Dispersion Relations and Elementary Particles, C. De Witt and R. Omnes (eds.), Wiley, 1960, 161–226.
[9] E. P. Wigner, Unitary representations of the inhomogeneous Lorentz group, Ann. Math., 40 (1939), 149–204.
[10] E. P. Wigner, Group theory and its applications to quantum mechanics of atomic spectra, Academic Press, 1959. (Original in German, 1931.)
[11] M. Sugiura, Unitary representations and harmonic analysis, Kodansha, Halsted Press, 1975.
[12] U. Uhlhorn, Representation of symmetry transformations in quantum mechanics, Arkiv för Physik, 23 (1963), 307–340.

M

259 (XX.11) Magnetohydrodynamics

Magnetohydrodynamics (also called **hydromagnetics** or **magnetofluid dynamics**) is concerned with the motion of an electrically conductive fluid in the presence of a magnetic field. An electromotive force (e.m.f.), induced by the motion of the fluid in the magnetic field, in turn induces an electric current that perturbs the original magnetic field. On the other hand, an electromagnetic force due to the magnetic field deforms the original motion. Many important and interesting phenomena result from such interactions of the magnetic field and the motion of the fluid.

In ordinary magnetohydrodynamics we assume that (i) the fluid is continuous, (ii) electric conductivity σ is not negligible, and (iii) fluid velocity is small compared with the velocity of light, i.e., $\max(L^2/(c^2T^2), U^2/c^2) \ll \min(1, R_m)$, where L is a representative length, T a representative time, U a representative velocity, and $R_m = \sigma\mu UL$ (μ is the magnetic permeability) is a nondimensional number called the **magnetic Reynolds number**. In this case we can neglect in the †Maxwell equations the displacement and convective currents in comparison with the conductive current $\mathbf{J}$, and write

$$\operatorname{div}\mathbf{B}=0, \quad \operatorname{rot}\mathbf{H}=\mathbf{J}, \quad \operatorname{rot}\mathbf{E}=-\partial\mathbf{B}/\partial t, \tag{1}$$

$$\rho_e = \operatorname{div}\varepsilon\mathbf{E}, \tag{2}$$

where ε is the dielectric constant, and

$$\mathbf{B}=\mu\mathbf{H}, \quad \mathbf{J}=\sigma(\mathbf{E}+\mathbf{v}\times\mathbf{B}). \tag{3}$$

Here the latter equation, **Ohm's law** for a moving medium, is valid when the effects of temperature gradient, the Hall effect, etc., are small. The motion of fluids (→ 205 Hydrodynamics) is governed by the †equation of continuity

$$\partial\rho/\partial t + \operatorname{div}\rho\mathbf{v}=0, \tag{4}$$

the †equation of motion

$$\partial(\rho\mathbf{v})/\partial t = -\operatorname{div}(\rho\mathbf{v}\otimes\mathbf{v}-\mathbf{P}-\mathbf{T}) \tag{5}$$

($\mathbf{P}$ is the mechanical stress tensor, $\mathbf{T}$ is the Maxwell stress tensor $T_{ij}=\mu(H_iH_j-\frac{1}{2}H^2\delta_{ij})$) or

$$\rho D\mathbf{v}/Dt = \operatorname{div}\mathbf{P}+\mathbf{K}, \quad \mathbf{K}=\mathbf{J}\times\mathbf{B}, \tag{5'}$$

the †equation of state, and the †energy equation. (From assumption (iii), the force $\rho_e\mathbf{E}$ on the electric charge ρ_e can be neglected compared with the force $\mathbf{J}\times\mathbf{B}$ on the electric current, and (2) can be separated from the other equations in order to determine ρ_e.)

When μ and σ are uniform, we can eliminate $\mathbf{E}$ and $\mathbf{J}$ from (1) and Ohm's law to obtain the **induction equation**

$$\partial\mathbf{B}/\partial t = \operatorname{rot}(\mathbf{v}\times\mathbf{B})+\lambda\Delta\mathbf{B}, \tag{6}$$

where $\Delta = \operatorname{grad}\operatorname{div} - \operatorname{rot}\operatorname{rot}$, $\lambda = 1/(\mu\sigma)$. This is of the same form as the equation $\partial\boldsymbol{\omega}/\partial t = \operatorname{rot}(\mathbf{v}\times\boldsymbol{\omega})+\nu\Delta\boldsymbol{\omega}$ (ν is the kinematic viscosity) for the †vorticity $\boldsymbol{\omega}$ of an ordinary †incompressible viscous fluid. We call $\lambda=1/(\mu\sigma)$ the **magnetic viscosity**, and the ratio of the first term (the convection term) to the second one (the diffusion term) in the right-hand side of (6) is the magnetic Reynolds number $R_m = UL/\lambda = \mu\sigma UL$. $R_m=\infty$ corresponds to the †perfect-fluid case as $\sigma\to\infty$ or $L\to\infty$. In this case, the magnetic flux moves with the fluid as if both were frozen together, as in the †Helmholtz theorem about vorticity. The existence of a transverse wave of velocity $\alpha=\sqrt{\mu H^2/\rho}$ along magnetic lines of force in the fluid, owing to the tension μH^2 (T_{ij} except for the magnetic pressure $-\frac{1}{2}\mu H^2\delta_{ij}$) of the magnetic flux frozen to the fluid, was noted for the first time by H. Alfvén (1943), and this wave is called the **Alfvén wave**. In a compressible perfect fluid ($R_m=\infty$, $P_{ij}=-p\delta_{ij}$, where p is the pressure), (1)–(5) reduce to a system of †hyperbolic partial differential equations, yielding as †characteristic surfaces in addition to the pure Alfvén wave two magnetosonic waves of phase velocities

$$a_{\pm} = \left(\frac{1}{\sqrt{2}}\right) \times \left(a^2+\alpha^2 \pm \sqrt{(a^2+\alpha^2)-4a^2\alpha^2\cos^2\theta}\right)^{1/2}$$

(θ is the angle between the magnetic field and the wave normal, a is the velocity of sound interfering with the Alfvén wave). We call a_+ and a_- the **fast wave** and the **slow wave**, respectively. **Hydromagnetic dynamo theory** explains the generation and maintenance of the magnetic field inside the earth on the basis of (6). Applications are also made to cosmic problems and MHD generation of electricity. Mercury, liquid sodium, etc., can be used to verify the theoretical results. Extrapolation may be made to the plasma used in thermonuclear fusion to the extent that the hydrodynamic treatment is valid as a first approximation.

References

[1] T. G. Cowling, Magnetohydrodynamics, Interscience, 1957.
[2] H. Alfvén and C. G. Falthammar, Cosmical electrodynamics, Clarendon Press, second edition, 1963.
[3] L. D. Landau and E. M. Lifshitz, Electro-

dynamics of continuous media, Course of theoretical physics 8, Pergamon Press, 1960.
[4] H. Cabannes, Theoretical magnetofluid-dynamics, Academic Press, 1970.
[5] Progress of theoretical physics, suppl. no. 24, 1962.

260 (XVII.6) Markov Chains

A. General Remarks

We consider a random process $\{X_t\}$ ($t \geqslant 0$ or $t=0,1,2,\ldots$) that is governed by some probability law. One of the most important is the process whose probability law of X_t under the condition $X_{s_1}=a_1,\ldots,X_{s_n}=a_n$ ($s_1<s_2<\ldots<s_n<t$) coincides with that under the condition $X_{s_n}=a_n$. This is called the Markov property, and a process with this property is called a Markov process (→ 261 Markov Processes). If the process takes place in a finite or countable set S, it is called a **Markov chain**.

A Markov process is specified by its †transition probability (i.e., the system of the probability laws of X_t given $X_s=x$ for $s<t$) and initial distribution (→ 261 Markov Processes). The transition probability of a Markov chain is denoted by $p_{s,t}(x,y)$, while that of a general Markov process is denoted by $P(s,x,t,A)$. Before proceeding to a sophisticated definition of Markov chains, we give some examples. (a) Suppose that $\xi_1,\xi_2,\ldots$ are mutually †independent real †random variables on a probability space $(\Omega,\mathfrak{B},P)$ and $f_1,f_2,\ldots$ are Borel measurable functions from $\mathbf{R}^2$ to $\mathbf{R}$. The process $\{X_n\}$ which is defined by the recurrence relation $X_0=\xi_0$, $X_n=f_n(\xi_n,X_{n-1})$ is a Markov process with its transition probability given by $P(n-1,x,n,A)=P\{f_n(\xi_n,x)\in A\}$. In particular, if the possible values of each ξ_n are at most countable, $\{X_n\}$ becomes a Markov chain. It is †temporally homogeneous, i.e., $p_{n,n+1}(x,y)$ is independent of n, if the $\{\xi_n\}$ are †identically distributed and $f_1=f_2=\ldots$. Moreover, if each ξ_n is integer-valued and $f_n(x,y)=x+y$, $\{X_n\}$ is a 1-dimensional random walk. (b) **Ehrenfest model of diffusion**. Suppose that N molecules are distributed in two containers A and B. At each trial a molecule is chosen at random and moved from its container to the other. Let X_n be the number of molecules in A after the nth trial. Then $\{X_n\}$ is a temporally homogeneous Markov chain with $p_{n,n+1}(i,i-1)=i/N$, $p_{n,n+1}(i,i+1)=(N-i)/N$ for $0\leqslant i\leqslant N$ ($p_{n,n+1}(i,j)=0$ otherwise). (c) Population model. Consider a population in which there is no interaction among individuals. Suppose that during a time interval $(t,t+\Delta t]$ each individual gives birth to a new one with probability $\lambda\Delta t+o(\Delta t)$ and dies with probability $\mu\Delta t+o(\Delta t)$. Let X_t be the population size at time t. Then the process $\{X_t\}$ becomes a birth and death process, the most typical of Markov chains with continuous time parameter (→ Section G). Numerous Markov chains with special structure are extensively studied in various fields of applications of probability theory such as queuing theory (→ Section H), the theory of †branching processes, the stochastic theory of †population genetics, and others [1, 3]. In this article, however, we are mainly concerned with the theoretical aspects of Markov chains.

Hereafter, we consider only †temporally homogeneous Markov chains, which are reformulated as a system of †stochastic processes in the following way. Let S be a finite or countable set (called the †state space of the motion) and T be $\{0,1,2,\ldots\}$ or $[0,\infty)$ (called the **time parameter space**). A family $(X_t,P_x)_{t\in T,x\in S}$ is called a **Markov chain** if P_x, the †probability law of the process X_t starting at x, is subjected to the condition

$$P_x(X_{s_1+t_n}=y_1,\ldots,X_{s_m+t_n}=y_m\mid X_{t_1}=x_1,\ldots,X_{t_n}=x_n)$$
$$=P_{x_n}(X_{s_1}=y_1,\ldots,X_{s_m}=y_m), \tag{1}$$

for every $t_j, s_k\in T$ ($j=1,\ldots,n$; $k=1,\ldots,m$) such that $t_1<t_2<\ldots<t_n$, $0<s_1<s_2\ldots<s_m$. Then the function defined by

$$p_t(x,y)=P_x(X_t=y),\qquad t\in T,\qquad x,y\in S$$

satisfies

$$0\leqslant p_t(x,y)\leqslant 1, \tag{2}$$

$$\sum_{y\in S}p_t(x,y)=1, \tag{3}$$

$$p_{t+s}(x,y)=\sum_{z\in S}p_t(x,z)p_s(z,y) \tag{4}$$

because of (1) and the general properties of probability laws. We call $p_t(x,y)$ ($t\in T$, $x,y\in S$) the **transition probability** (or **transition function**) of the Markov chain, and relation (4) is called the **Chapman-Kolmogorov equation**. Furthermore, the matrix $p_t=(p_t(x,y))$ is called the **transition matrix**.

Conversely, for a given $p_t(x,y)$ with properties (2), (3), and (4), we can construct a Markov chain that satisfies

$$P_x(X_{t_1}=x_1,\ldots,X_{t_n}=x_n)$$
$$=p_{t_1}(x,x_1)p_{t_2-t_1}(x_1,x_2)\ldots p_{t_n-t_{n-1}}(x_{n-1},x_n)$$

by using Kolmogorov's extension theorem. Such a Markov chain is essentially unique. If we are given a **stochastic matrix** $p=(p_{x,y})$, i.e., a matrix satisfying $0\leqslant p_{x,y}\leqslant 1$ and $\sum_{y\in S}p_{x,y}=1$, the components $p_n(x,y)$ of the iterated matrix p^n satisfy (2)–(4), and hence there exists a

Markov chain with discrete parameter having $p_n(x, y)$ as its transition probability.

Let $\mu_x = P(X_0 = x)$ be the initial distribution over S at time 0. Then the distribution $\mu_y^{(t)} = P(X_t = y)$ at time t is obtained from $\mu_y^{(t)} = \sum_{x \in S} \mu_x p_t(x, y)$. In particular, if $\mu_y^{(t)}$ is independent of t, i.e.,

$$\mu_y = \sum_{x \in S} \mu_x p_t(x, y), \tag{5}$$

then μ is called the **invariant distribution** of the Markov chain. An arbitrary real nonnegative solution of (5) is called an **invariant measure.**

When S is the set of all d-dimensional lattice points and $p_{x,y} = \pi_{x-y}$, the associated Markov chain is called a **(general) random walk**. In particular, if $\pi_x = 2^{-n}$ for every neighboring point x of the origin and $=0$ otherwise, it is called a **standard random walk** (or simply **random walk**).

Now, when the equal sign in (3) is replaced by $<$, we consider the space S^* obtained by adjoining ∂ (†death point) to S, and define

$$p_t^*(x, y) = p_t(x, y), \qquad x, y \in S;$$

$$p_t^*(x, \partial) = 1 - \sum_{y \in S} p_t(x, y), \qquad x \in S;$$

$$p_t^*(\partial, \partial) = 1; \qquad p_t^*(\partial, x) = 0.$$

Then we can associate a Markov chain (X_t, P_x^*) on S^* with $\{p_t^*\}$. If X_t equals ∂, the particle of the process is regarded as extinct at the random time $\zeta = \inf\{t \mid X_t = \partial\}$, called the **lifetime** (or **killing time**). In this case, the process X_t restricted to t smaller than ζ is also called a Markov chain on S with the transition probabilities $\{p_t(x, y)\}$. Then the conditions $\sum_{y \in S} p_t(x, y) = 1$ and $P_x(\zeta = \infty) = 1$ are equivalent, and the chain is called **conservative** if $P_x(\zeta = \infty) = 1$ for every x.

In this section we have restricted ourselves to the temporally homogeneous case. In the temporally inhomogeneous case, we have to consider the probability laws $P_{x,t}$ of the path starting from $x \in S$ at time t, instead of P_x. Equation (1) becomes

$$P_{x,t}(X_{s_1} = y_1, \ldots, X_{s_m} = y_m \mid X_{t_1} = x_1, \ldots, X_{t_n} = x_n)$$
$$= P_{x_n, t_n}(X_{s_1} = y_1, \ldots, X_{s_m} = y_m),$$
$$t_1 < \ldots < t_n < s_1 < \ldots < s_n.$$

For the rest of this article, we consider only the homogeneous case.

B. Markov Chains with Discrete Parameter

Hereafter, the one-step and the n-step transition probability of a Markov chain with discrete parameter will be denoted by $P(x, y)$ and $P_n(x, y)$ instead of $p_1(x, y)$ and $p_n(x, y)$, respectively.

Let $\mathfrak{X} = (X_n, P_x)$ be a conservative Markov chain. For a subset A of the state space S, $\sigma_A = \min\{n \geqslant 1 \mid X_n \in A\}$ ($\min \varnothing = +\infty$) is called the **hitting time** for the set A. If $P_x(\sigma_y < \infty) > 0$ ($=0$), we write $x \to y$ ($x \not\to y$). Then $x \to y$ is equivalent to the existence of $n \geqslant 1$ with $P_n(x, y) > 0$. When $x \to y$ and $y \to x$, we write $x \leftrightarrow y$. The set of all x for which there exists $y \neq x$ with $x \to y$ and $y \not\to x$ is denoted by F and is called the **dissipative part** of S. For elements of $S - F$, the relation $\leftrightarrow$ is an equivalence relation. Each equivalence class E_α is called an **ergodic class.** When $F = \varnothing$ and S consists of a single class, the chain $\mathfrak{X}$ is called **irreducible** (or **ergodic**). For an ergodic class E, the greatest common divisor d of $\{n > 1 \mid P_n(x, x) > 0\}$ for $x \in E$ does not depend on the choice of $x \in E$ and is called the **period** of the class E. A class with period 1 is called **aperiodic**. Set $G_n = \{y \in E \mid P_{kd+n}(x, y) > 0$ for some $k \geqslant 0\}$ ($n = 1, 2, \ldots, d$) for a fixed $x \in E$. Then we have a decomposition of $E: E = \bigcup G_n$, $G_n \cap G_m = \varnothing$ ($n \neq m$), and $\sum_{z \in G_{n+1}} P(y, z) = 1$ ($y \in G_n$). Each G_n is called a **cyclic part**. The decomposition may depend on the choice of x but is unique up to ordering.

The point x is called **recurrent** or **nonrecurrent (transient)** according as $P_x(\sigma_x < \infty) = 1$ or <1. A necessary and sufficient condition for x to be recurrent is $\sum_{n=0}^{\infty} P_n(x, x) = \infty$. The probability P_x ($X_n = x$ occurs infinitely often) is 1 or 0 according as x is recurrent or nonrecurrent. A chain is called **recurrent** or **nonrecurrent** according as every point in S is recurrent or nonrecurrent. A recurrent point x is called **positive recurrent** or **null recurrent** according as $m_x \equiv E_x(\sigma_x)$ is finite or infinite. Let E be an ergodic class. If there exists a positive recurrent element in E, then all elements in E are also positive recurrent, and the class E is called **positive recurrent**. We can define null recurrence or nonrecurrence of an ergodic class similarly. For a finite state space, every Markov chain has at least one ergodic class, and all ergodic classes are positive recurrent.

C. Limit Theorems for Markov Chains, and Recurrent Events

The properties of a recurrent chain are reduced to those of an irreducible recurrent chain, since a recurrent chain is decomposed into ergodic classes. We assume that $\mathfrak{X} = (X_n, P_x)$ is an irreducible recurrent chain. Concerning the transition function we have the following limit theorems: Let d be the period of $\mathfrak{X}$ and $S = \bigcup_{r=1}^{d} G_r$ be the decomposition into cyclic parts. If $x \in G_d$, then

$\lim_{n\to\infty} P_{nd+r}(x, y) = dm_y^{-1}$ $(y\in G_r)$, $=0$ $(y \notin G_r)$. These results are sometimes referred to as the **basic limit theorem of** $P_n(x, y)$. For every x, $y\in S$, $\lim_{n\to\infty} N^{-1}\sum_{n=1}^{N} P_n(x, y) = m_y^{-1}$. Furthermore, $r_{x,y} = \lim_{N\to\infty}(\sum_{n=1}^{N} P_n(y, y)/\sum_{n=1}^{N} P_n(x, x))$ exists and is finite. Then $r_{x,y}$ is an invariant measure as a function of y, and every invariant measure is a constant multiple of it. The chain is positive recurrent if and only if $\sum_y r_{x,y} < \infty$, and then we have $r_{x,y} = m_x m_y^{-1}$. Consequently there exists a real invariant measure μ_x (real solution of (5)) such that $0 < \sum_{x\in S}|\mu_x| < \infty$ if and only if the chain is positive recurrent, and in this case μ_x is a constant multiple of m_x^{-1}.

Let μ be an invariant measure, and set $I_\mu(f) = \sum_{x\in S}\mu_x f(x)$. For f, g such that $I_\mu(|f|) < \infty$, $0 < I_\mu(|g|) < \infty$, and $I_\mu(g) \neq 0$, we have

$$P_x\left(\lim_{N\to\infty}\left(\sum_{n=1}^{N} f(X_n)\Big/\sum_{n=1}^{N} g(X_n)\right) = I_\mu(f)/I_\mu(g)\right) = 1.$$

The †law of large numbers, the †central limit theorem, and the †law of the iterated logarithm hold for the asymptotic behavior of the sum $\sum_{n=1}^{N} f(x_n)$ as $N\to\infty$.

Let $\mathscr{E} = \{E_1, E_2, \ldots\}$ be a sequence of events on a probability space $(\Omega, \mathscr{B}, P)$, and let $u_n = P(E_n)$. The time intervals $\{\tau_k\}$ between the successive occurrences of $\mathscr{E}$ are called the **recurrence time of** $\mathscr{E}$. More precisely, the $\{\tau_k\}$ are successively defined by $\tau_1 = \inf\{n\geq 1 \mid E_n$ occurs$\}$, $\tau_k = \inf\{n \geq \tau_1 + \ldots + \tau_{k-1} + 1 \mid E_n$ occurs$\} - (\tau_1 + \ldots + \tau_{n-1})$ $(k\geq 2)$. To avoid complications we assume here that all the τ_k are finite †almost surely. The sequence $\mathscr{E}$ is called a (persistent) **recurrent event** if the τ_k $(k\geq 1)$ are mutually †independent †random variables with a common †distribution $\{f_n\}$. More generally, if the distribution of τ_1 is allowed to be different from $\{f_n\}$, $\mathscr{E}$ is called a **delayed** recurrent event. The basic relation on a recurrent event is the **renewal equation**

$$u_n = b_n + u_1 f_{n-1} + u_2 f_{n-2} + \ldots + u_{n-1} f_1,$$

where $\{b_n\}$ is the distribution of τ_1. Suppose that the greatest common divisor of $\{n\geq 1 \mid f_n > 0\}$ is d. Then we obtain

$$\lim_{n\to\infty} u_{nd+r} = d\beta_r/\mu, \qquad r = 1, 2, \ldots d, \qquad (6)$$

where $\mu = \sum n f_n$ is the **mean recurrence time** and $\beta_r = \sum_{n\geq 0} b_{nd+r}$. This fact is known as the **renewal theorem**. Let $\mathfrak{X} = (X_n, P_x)$ be an irreducible recurrent Markov chain. For fixed x and y, consider the events $E_n = \{X_n = y\}$ under the measure P_x. Then, $\mathscr{E} = \{E_1, E_2, \ldots\}$ is a delayed recurrent event with $u_n = P_n(x, y)$, $f_n = P_y(\sigma_y = n)$, and $b_n = P_x(\sigma_y = n)$. Applying (6), we obtain the basic limit theorem of $P_n(x, y)$ stated at the beginning of this section. We next consider the number N_n of occurrences of a recurrent event $\mathscr{E}$ up to time n. The limiting behavior of N_n has been extensively studied (→ 250 Limit Theorems in Probability Theory D).

D. Potential Theory for Markov Chains

For a given Markov chain,

$$G(x, y) = \sum_{n=0}^{\infty} P_n(x, y)$$

is well defined, admitting possibly the value ∞. If $G(x, y)$ is not identically ∞, we can define a (generalized) †potential with kernel $G(x, y)$ (→ 45 Brownian Motion; 261 Markov Processes; 338 Potential Theory). For a real function φ over S, the function $G\varphi(x) = \sum_{y\in S} G(x, y)\varphi(y)$ is called the **potential** with charge φ if the sum exists. Even if $G\varphi$ does not exist, the infinite sum $\sum_{n=0}^{\infty} P_n\varphi$ may exist. In this case this sum is also called the potential with charge φ.

A real function f $(-\infty < f \leq +\infty)$ over S is defined to be **superharmonic** (or **superregular**) if $Pf \leq f$. Here P is the operator associated with the kernel $P(x, y)$, that is, $Pf(x) = \sum_{y\in S} P(x, y) f(y)$. Furthermore, if $f\geq 0$ and $f\geq Pf$, f is called **excessive**, and if $-\infty < f < +\infty$ and $f = Pf$, f is called **harmonic**. If $Pf\leq f$ at a point x, f is called **superharmonic at** x, etc. A potential with charge $\varphi\geq 0$ is always excessive. For an irreducible recurrent chain, every nonnegative superharmonic function is constant.

(1) For a nonrecurrent chain we can consider the potential $f = G\varphi$ for every function φ with finite support. G satisfies $(P - I)G = -I$ and $\lim_{n\to\infty} P_n G = 0$, where I is the unit matrix. Consequently, the operator $P - I$ corresponds to the †Laplacian Δ of †Newtonian potential theory, and the equation $(P - I)f = 0$ corresponds to the Laplace equation $\Delta f = 0$. If the limit $w = \lim_{n\to\infty} P_n f$ exists for a function f on S, f can be expressed uniquely as the sum of the potential $G\varphi$ $(\varphi = f - Pf)$ and a harmonic function w. This decomposition is called the **Riesz decomposition**, following the terminology used in Newtonian potential theory (→ 338 Potential Theory). Let E be a finite subset of S and $\sigma_E^* = \min\{n\geq 0 \mid X_n\in E\}$. Then $f(x) = P_x(\sigma_E^* < \infty)$ is a (unique) potential which is harmonic at $x\in E^c$ and takes the value 1 on E. This is called the **equilibrium potential** of the set E, and its total charge $C(E) = \sum_{x\in S}(f(x) - Pf(x))$ is called the **capacity** of E. The †maximum principle and the †balayage principle (→ 338 Potential Theory) are valid in this potential theory.

(2) For a recurrent chain, we cannot define the potential kernel as in (1), since $G(x, x) = \infty$. However, we can define a kernel analogous

to the case of the †logarithmic potential. We assume that the chain is irreducible. When the limit

$$A(x, y) = \lim_{n\to\infty} (G_n(x, x)\mu_y\mu_x^{-1} - G_n(x, y)),$$

$$G_n(x, y) = \sum_{k=0}^{n} P_k(x, y),$$

exists for an invariant measure μ, the chain is called (right) **normal**. A chain is normal if and only if

$$\lambda_E(\cdot) = \lim_{n\to\infty} \sum_{z\in S} P_n(x, z) P_z(\sigma_E \in \cdot)$$

exists. If $\lambda_E(\cdot)$ exists, it is independent of x. Every positive recurrent chain is normal. For a normal chain we obtain the following results. Let φ be a function with finite support. The potential f of φ exists if and only if $\sum_{x\in S}\mu_x\varphi(x) = 0$, and $f = -A\varphi$ in this case. $A\varphi$ satisfies the conditions $(P-1)A\varphi = \varphi$ and $\lim_{n\to\infty} P_n A\varphi = 0$. A function f is a bounded potential of a function with finite support E if and only if f is bounded and harmonic in E^c and satisfies $\sum_x \lambda_E(x) f(x) = 0$. An irreducible recurrent chain is not necessarily normal. For every such chain, however, there exists a kernel $W(x, y)$ such that $(P-I)W\varphi = -\varphi$ for any function φ with finite support and null charge (i.e., $\sum_{x\in S}\mu_x\varphi(x) = 0$). Such a kernel W is called a **weak potential kernel**.

E. Random Walks

Consider a random walk defined on the set S of all lattice points in a d-dimensional Euclidean space. Let $S^+ = \{x \mid 0 \to x\}$, $\bar{S} = \{z \mid z = x - y, x, y \in S^+\}$. F. Spitzer obtained the following results for the random walk with $S = \bar{S}$. The random walk is recurrent if the following conditions are satisfied: (i) $d = 1$, $\sum |x| P(0, x) < \infty$ ($|x|$ is the distance between 0 and x), and $m \equiv \sum x P(0, x) = 0$; (ii) $d = 2$, $m = 0$, and $\sigma^2 \equiv \sum |x|^2 P(0, x) < \infty$. When $d \geqslant 3$, the random walk is always nonrecurrent. The measure $\mu_x \equiv 1$ is invariant whether the random walk is recurrent or not.

Every recurrent random walk is right normal, and the potential kernel A satisfies $(P-I)A = I$. Several interesting results are known on the uniqueness of the kernel A satisfying $(P-I)A = I$ [7].

For the case $m = 0$, there are a number of results similar to those for †Brownian motion, including: (i) When $d = 1$ and $0 < \sigma < \infty$,

$$\lim_{n\to\infty} P_0(\max_{0\leqslant k\leqslant n} |X_k| \leqslant \sigma\sqrt{n}\, x) = 1 - F(x^{-2}), \qquad x > 0,$$

where

$$F(x) = 1 - \frac{4}{\pi}\sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \exp\left(-\frac{\pi^2}{8}(2k+1)^2 x\right);$$

(ii) The **arc sine law**: Let T_n be the number of $k \leqslant n$ for which X_k becomes >0. When $d = 1$ and $0 < \sigma < \infty$,

$$\lim_{n\to\infty} P_0(T_n \leqslant nx) = \frac{2}{\pi}\arcsin\sqrt{x}, \qquad 0 \leqslant x \leqslant 1;$$

(iii) The **Wiener test**: The set E is called **recurrent** if $P_x(\sigma_E < \infty) = 1$ holds for every x. When $d = 3$ and $\sigma < \infty$, a set E is recurrent if and only if $\sum_{n=1}^{\infty} C(E_n) 2^{-n} = \infty$, where $E_n = E \cap \{x \mid 2^n \leqslant |x| < 2^{n+1}\}$.

F. Markov Chains with Continuous Time Parameter

Suppose that the transition probability $p_t(x, y)$ is measurable in t. Then $p_t(x, y)$ is uniformly continuous in a complement of any neighborhood of $t = 0$, and there exists $m_{x,y} = \lim_{t\uparrow\infty} p_t(x, y)$ for which $m_{x,y} = \sum_{z\in S} m_{x,z} p_t(z, y) = \sum_{z\in S} p_t(x, z) m_{z,y}$. If $p_t(x, y) = 0$ for all $x \in S$ and $t > 0$, y is called a **fictitious state**. Let F be the set of all fictitious states of S. Then the restriction of $p_t(x, y)$ to $S - F$ gives a transition probability on $S - F$. If $F = \varnothing$ and the family $\mathbf{p} = \{p_t(\cdot, y) \mid t > 0, y \in S\}$ separates points of S, the transition probability $p_t(x, y)$ is called **standard**. Then $p_t(x, y)$ is standard if and only if $\lim_{t\downarrow 0} p_t(x, y) = \delta_{x,y}$. When $F = \varnothing$ and $\mathbf{p}$ does not separate, points of S can be reduced to the standard case by a suitable identification of states. We assume that $p_t(x, y)$ is standard. Then $\lim_{t\downarrow 0} t^{-1}(p_t(x, y) - \delta_{x,y}) = q_{x,y}$ exists and satisfies $0 \leqslant q_{x,y} < \infty$ if $x \neq y$. The matrix $Q = (q_{x,y})$ over S is called the **Q-matrix** of the transition matrix $p_t = (p_t(x, y))$, and we write $Q = p_0'$. Set $q_x = -q_{x,x}$ $(\geqslant 0)$. We call x a **stable state** if $q_x < \infty$ and an **instantaneous state** if $q_x = \infty$. If $q_x < \infty$, $p_t'(x, y)$ (the derivative with respect to t) exists and is continuous in $t > 0$.

When every point of S is stable, $\pi(x, y)$ defined by

$$\pi(x, y) = \begin{cases} q_{x,y} q_x^{-1} & \text{if } x \neq y \\ 0 & \text{if } x = y \end{cases}$$

satisfies

$$0 \leqslant \pi(x, y) \leqslant 1, \qquad \pi(x, x) = 0,$$

$$\sum_{y\in S} \pi(x, y) \leqslant 1. \tag{7}$$

From the †Kolmogorov-Chapman equation we can formally derive **Kolmogorov's backward equation**

$$p_t'(x, y) = -q_x p_t(x, y) + q_x \sum_{z\in S} \pi(x, z) p_t(z, y) \tag{8}$$

and its dual, **Kolomogorov's forward equation**

$$p_t'(x,y) = -q_y p(x,y) + \sum_{z\in S} p_t(x,z)q_z\pi(z,y). \qquad (9)$$

Strictly speaking, these equations hold only under suitable conditions: For instance, a conservative transition probability $p_t(x,y)$ satisfies (8) if and only if $\sum_{y\in S}\pi(x,y) = 1$. If there exists $\xi_x > 0$ $(x\in S)$ such that $\sum_{x\in S}\xi_x p_t(x,y) \leqslant \xi_y$ $(\forall t>0)$, then $p_t(x,y)$ satisfies (9) if and only if $\sum_{y\in S}\xi_y q_y \pi(y,x) = \xi_x$. Conversely, given $0\leqslant q_x<\infty$ and π satisfying (7), there exist in general many solutions of (8) and (9) with the initial condition $\lim_{t\downarrow 0} p_t(x,y) = \delta_{x,y}$. Among them, the minimal solution $p_t^0(x,y)$ exists. The chain with $p_t^0(x,y)$ as its transition probability is called the **minimal chain** associated with $\{q,\pi\}$.

For the path of a Markov chain with standard transition probability, there exists a separable measurable †modification X_t, but in general there exists neither a right continuous modification nor a modification having the †strong Markov property for all †stopping times. (For detailed properties of paths → [8, 9].) Set $\tau_0 = 0$, $\tau_n = \inf\{t>\pi_{n-1} \mid X_t \neq X_{\tau_{n-1}}\}$ $(n\geqslant 1)$, and $\tau_\infty = \lim_{n\to\infty}\tau_n$. If each point x of S is stable, τ_1 is subject to the exponential distribution $P_x(\tau_1>t) = e^{-q_x t}$, so that $E_x(\tau_1) = q_x^{-1}$, $P_x(X_{\tau_1} = y) = \pi(x,y)$, and τ_1 and X_{τ_1} are independent with respect to the P_x-measure. Furthermore $p_t^0(x,y) = P_x(X_t = y, \tau_\infty > t)$ is the minimal solution. For the minimal chain there exists a right continuous modification with left-hand limits, which has the strong Markov property.

For a finite Markov chain with a standard transition probability, all states are stable and both (8) and (9) are fulfilled. Furthermore, the transition probability satisying (8) and (9) is the unique minimal solution.

D. Williams [10] obtained the following remarkable result concerning Markov chains having only instantaneous states. Let Q be a matrix on S with $-q_{x,x} = \infty$ $(\forall x)$, $0\leqslant q_{x,y}<\infty$ $(\forall x,y;\, x\neq y)$. Then Q is the Q-matrix of a transition matrix if and only if the following two conditions are satisfied: (i) $\sum_{z\neq x,y}\min(q_{x,z}, q_{y,z}) < \infty$ $(\forall x,y;\, x\neq y)$, (ii) for some infinite subset A of S, $\sum_{y\in A\setminus\{x\}} q_{x,y} < \infty$ $(\forall x)$. Williams [11] also studied a counterpart of Kolmogorov's backward equation for the instantaneous case.

G. Birth and Death Processes

A Markov chain $\mathfrak{X}$ with state space $S = \{0, 1, 2, \dots\}$ is called a **birth and death process** if its Q-matrix $Q = (q_{n,m})$ satisfies the following conditions: $0\leqslant q_0 = q_{0,1} < \infty$, $0 < q_n = q_{n,n-1} + q_{n,n+1} < \infty$ $(n\geqslant 1)$ and $q_{n,m} = 0$ $(m\neq n+1, n-1)$, where $q_n = -q_{n,n}$ as before. Usually we write λ_n for $q_{n,n+1}$ and μ_n for $q_{n,n-1}$. The parameters λ_n and μ_n are called the **infinitesimal birth** and **death rates**, respectively. In particular, the chain $\mathfrak{X}$ is called a **birth process** if $\mu_n = 0$ and a **death process** if $\lambda_n = 0$. The birth and death process satisfying $\lambda_n > 0$ and $\mu_n > 0$ for $n\geqslant 1$ has a number of properties similar to those of a 1-dimensional †diffusion process.

For the rest of this section we assume that $\lambda_0 = 0$ (i.e., 0 is a †trap) and $\lambda_n > 0$, $\mu_n > 0$ $(n\geqslant 1)$. The sequence x_n, called the **natural scale**, is defined as follows: $x_1 = \mu_1^{-1}$, $x_2 = \mu_1^{-1} + \lambda_1^{-1}$, $x_n = \mu_1^{-1} + \lambda_1^{-1} + \dots + (\mu_2\dots\mu_{n-1})(\lambda_1\dots\lambda_{n-1})^{-1}$ $(n\geqslant 3)$, $x_\infty = \lim_{n\to\infty} x_n$. The measure m with the mass $m_n = \lambda_1\dots\lambda_{n-1}(\mu_2\dots\mu_n)^{-1}$ $(n\geqslant 2)$ and $= 1$ $(n=1)$ at the point x_n is called the **canonical measure**. Then $p_t(x_i, x_k) = p_t(i,k)m_k^{-1}$ is a transition probability on $E = \{x_1, x_2, \dots\}$ and $f(x_i, t) = p_t(x_i, x_k)$ satisfies a differential-difference equation

$$\partial f/\partial t = D_m f^+, \qquad (10)$$

which is equivalent to (8). Here, $f^+(x_n) = (f(x_{n+1}) - f(x_n))(x_{n+1} - x_n)^{-1}$ and $D_m g(x_n) = (g(x_n) - g(x_{n-1}))m_n^{-1}$. The operator $f\to D_m\cdot f^+$ is similar to Feller's expression for the infinitesimal generators of 1-dimensional diffusion processes.

Every birth and death process is obtained from †Brownian motion by †time change. Furthermore, x_∞ is regarded as a boundary point and is classified as a natural, exit, entrance, or regular boundary point. Every birth and death process is determined by (10) and the boundary condition at x_∞ (→ 115 Diffusion Processes).

H. Markov Chains in Queuing Theory

A queue is formed when customers arrive at random times at some facility and request service of some kind. Such phenomenon often arises in service systems. The **queuing model** generally embodies the following mathematical structure. Suppose that the nth customer arrives at time τ_n, waits for a time interval w_n until the beginning of his service, and departs after a service time v_n. Usually the models we consider are specified by the following assumptions. The interarrival times $u_n = \tau_{n+1} - \tau_n$ $(n\geqslant 1)$ are mutually †independent †random variables distributed according to a common †distribution function $F(x)$ with †mean λ^{-1}. Similarly, $\{v_n\}$ is a sequence of independent random variables with a common distribution function $G(x)$ with mean μ^{-1}. The sequences $\{u_n\}$ and $\{v_n\}$ are mutually independent. The queuing discipline is "first come, first served."

In case of many servers, this means that the first customer waiting in the queue is served as soon as any one of the servers becomes free. Besides this standard queuing model, various other queuing models have been studied in different fields of application. There is a huge literature on **queuing theory**, i.e., the mathematical analysis of queuing models. A compendium of results obtained by around the 1960s, together with an extensive bibliography on this theory, is contained in [15]. Some recent aspects of queuing theory are developed in [17]. Here, we are concerned only with some simple queuing models and their associated Markov chains.

There is a standardized notation, introduced by D. G. Kendall [14], for identifying standard queuing models. In **Kendall's notation** $A/B/s$, s represents the number of servers, while A and B indicate the types of distributions $F(x)$ and $G(x)$, respectively. Distributions frequently used in the first two places of Kendall's notation are the following: $M=$ an †exponential distribution, $D=$ a †unit distribution, $E_k=$ a †gamma distribution of order k (called a k-**Erlang distribution** in queuing theory), G(or GI) $=$ a general distribution, and so on. The model $M/\cdot/\cdot$ is often said to have **Poisson input**, for the number A_t of arrivals during a time interval $(0,t]$ forms a †Poisson process.

One of the important problems in queuing theory is to analyze the number Q_t of customers waiting or being served at time t. The process Q_t is called the **queue length**. If $s=\infty$, then Q_t indicates the number of busy servers at time t. The simplest and most extensively studied queuing model is $M/M/s$. In this case, the process Q_t becomes a birth and death process with $\lambda_j=\lambda$, $\mu_j=j\mu$ $(1\leqslant j\leqslant s)$, and $=s\mu$ $(j>s)$. Kolmogorov's forward equation is given by

$$p_t'(i,0)=-\lambda p_t(i,0)+\mu p_t(i,1),$$

$$p_t'(i,j)=\lambda p_t(i,j-1)-(\lambda+j\mu)p_t(i,j)+(j+1)\mu p_t(i,j+1),\qquad 1\leqslant j\leqslant s-1,$$

$$p_t'(i,j)=\lambda p_t(i,j-1)-(\lambda+s\mu)p_t(i,j)+s\mu p_t(i,j+1),\qquad j\geqslant s.$$

This equation can be solved explicitly by the method of †generating functions. The limit distribution $\lim_{t\to\infty}p_t(i,j)=p_j$, independent of i, exists if and only if $s\mu>\lambda$. In particular,

$$p_j=\begin{cases}(1-\rho)\rho^j & \text{if } s=1,\\ e^{-\rho}\rho^j/j! & \text{if } s=\infty,\end{cases}$$

where $\rho=\lambda/\mu$. (In the single-server case, ρ is called the **traffic intensity**.) Unless the queuing model is $M/M/s$, Q_t is no longer a Markov chain. In some special cases, such as $E_k/M/1$, the properties of Q_t can be reduced to those of some birth and death process.

The analysis of Q_t itself is difficult in general. However, if either $F(x)$ or $G(x)$ is of exponential type, the method of **embedded** ($=$ **imbedded**) **Markov chains** is useful. For example, in the system $M/G/1$ we examine the queue length only at times $t_0=0, t_1, t_2, \ldots$, where t_n is the departure time of the nth customer. This embedded process $X_n=Q_{t_n}$ becomes a Markov chain on $S=\{0,1,2,\ldots\}$. For practical purposes we could content ourselves with results concerning $\{X_n\}$ instead of the original process Q_t. The transition probability $P(i,j)$ of $\{X_n\}$ is given by

$$P(0,j)=k_j,\quad P(i,j)=k_{j-i+1},\qquad i\geqslant 1,\tag{11}$$

where

$$k_j=\begin{cases}\displaystyle\int_0^\infty e^{-\lambda x}\frac{(\lambda x)^j}{j!}dG(x) & \text{if } j\geqslant 0,\\ k_j=0 & \text{if } j<0.\end{cases}$$

This chain is irreducible and aperiodic. Moreover, it is transient, null recurrent, or positive recurrent according as the traffic intensity $\rho=\lambda/\mu>1$, $=1$, or <1. In the last case, the generating function $P(s)$ of the limit distribution $\{P_j\}$ is given by

$$P(s)=\frac{(1-\rho)(1-s)K(s)}{K(s)-s},$$

where $K(s)$ is the generating function of $\{k_j\}$. Similarly, if the system is $GI/M/1$, the embedded process $X_n=Q_{\tau_n-}$ is the Markov chain with transition probability

$$P(i,0)=\alpha_i,\quad P(i,j)=l_{i-j+1},\quad j\geqslant 1,$$

where

$$l_j=\begin{cases}\displaystyle\int_0^\infty e^{-\mu x}\frac{(\mu x)^j}{j!}dF(x) & \text{if } j\geqslant 0,\\ l_j=0 & \text{if } j<0,\end{cases}$$

and $\alpha_i=\sum_{j>i}l_j$. The limiting behavior of this chain is analogous to that of the preceeding chain (11). In the positive recurrent case $(\rho<1)$, the limit distribution is given by $p_j=(1-\zeta)\zeta^j$, where ζ is the solution of $L(s)=s$ in $(0,1)$ for the generating function $L(s)$ of $\{l_j\}$. These results are extended to the cases with many servers.

For the general single-server queuing model $GI/G/1$, when we consider the waiting times $\{w_n\}$ instead of Q_t, we obtain the recurrence relation $w_{n+1}=\max\{0,w_n+v_n-u_n\}$. This implies that $\{w_n\}$ is a Markov process with discrete parameter, over $S=[0,\infty)$ an uncountable state space, having the structure of a Markov chain with transition probability (11). Several methods are exploited for the analysis of $\{w_n\}$ [3, 16, 17].

I. Boundary Value Problems for Markov Chains

To discuss the behavior of the path of a Markov chain beyond the time τ_∞, we have to introduce a suitable boundary of the state space S. Among several conceivable boundaries, the Martin boundary, which is most frequently utilized, is explained later in this section. The name comes from its similarity to the †Martin boundary in the theory of harmonic functions. Most results stated in this section are also valid for the discrete time parameter case.

Let $\mathfrak{X}=(X_t, P_x)$ be the minimal, nonrecurrent chain associated with $\{q,\pi\}$ for which τ_∞ equals the lifetime ζ. Harmonic functions, etc., are defined as in the discrete time parameter case (with p replaced by π). Let $\gamma\equiv\gamma(x)$ be a measure such that $0<\gamma G(y)\equiv\sum_x\gamma(x)G(x,y)<\infty$, where $G(x,y)=\int_0^\infty p_t(x,y)\,dt$. Let ρ_1 be the metric in the one-point compactification of the state space S equipped with the discrete topology. Set $K(x,y)=G(x,y)/\gamma G(y)$, and define

$$\rho_2(y,y')=\sum_{n=1}^{\infty}\frac{1}{2^n}\frac{|K(x_n,y)-K(x_n,y')|}{1+|K(x_n,y)-K(x_n,y')|},$$

$$\{x_n\}=S.$$

The set ∂S of all points adjoined to S by the completion of S relative to $\rho=\rho_1+\rho_2$ is called the **Martin boundary** of S. $M=S\cup\partial S$ is a compact separable space. By the definition of ρ, $K(x,\xi)=\lim_{y\to\xi}K(x,y)$ exists, is continuous in ξ, and is superharmonic in x. A nonnegative superharmonic function u is called **minimal** if every superharmonic function v such that $0\leqslant v\leqslant u$ is a constant multiple of u. The set $\partial S_1=\{\xi\,|\,K(\cdot,\xi)$ is minimal harmonic$\}$, called the **essential part** of ∂S, is an †F_σ-set. Then every γ-integrable nonnegative superharmonic function u is represented by $u(x)=\int K(x,\xi)\mu(d\xi)$ by means of a unique measure μ on $M_1=S\cup\partial S_1$, called the **canonical measure** of u. In particular, if u is harmonic, μ is concentrated on ∂S. A number of representation problems in analysis, such as the Hausdorff †moment problem, can be considered to be representation problems of suitable Markov chains. Let u be a γ-integrable nonnegative superharmonic function and (X_t, P_x^u) the Markov chain (called the **u-chain)** having $p_t^u(x,y)=u(x)^{-1}u(y)p_t(x,y)$ $(0/0=0)$ as its transition probability. Then $X_{\zeta-}=\lim_{t\uparrow\zeta}X_t$ exists and

$$P_x^u(X_{\zeta-}\in B)=u(x)^{-1}\int_B K(x,\xi)\mu(d\xi),$$

where μ is the canonical measure of u.

A measure v on S is called a **superharmonic measure (harmonic measure)** if $vq(\pi-I)\leqslant 0$ $(=0)$, that is, $\sum_{y\in S}v_yq_y(\pi_{yx}-\delta_{yx})\leqslant 0$ $(=0)$. Fix a function $g\geqslant 0$ such that $0<Gg<\infty$ (→ Section D), and set $K^*(x,y)=G(x,y)Gg(x)^{-1}$. Define the metric ρ_2^* similar to ρ_2, using the function family $\{K^*(\cdot,y)\}$ $(y\in S)$. The set adjoined to S by the completion relative to $\rho^*=\rho_1+\rho_2^*$ is called the **dual Martin boundary**. Extend K^* to $S\cup\partial S^*$ and denote by ∂S_1^* the set of all $\eta\in\partial S^*$ such that $K^*(\eta,\cdot)$ is a minimal superharmonic measure. Then every superharmonic measure v with $\int v(dx)g(x)<\infty$ is represented uniquely as $v=\int\mu(d\eta)K^*(\eta,\cdot)$ in terms of a measure μ on $S\cup\partial S_1^*$.

Let $\xi\in M_1$, and denote the $K(\cdot,\xi)$-chain by $(X_t, P_x^{*,\xi})$. Then $P_x^\xi(\zeta<\infty)=0$ or $=1$. We call ξ a **passive boundary point** in the first case and an **exit boundary point** in the second. On the other hand, denote by $(X_t, P_x^{*,\eta})$ the chain having $P^{*,\eta}(x,y)=K^*(\eta,x)^{-1}K^*(\eta,y)p_t(y,x)$ as its transition probability. Then $P_x^{*,\eta}(\zeta<\infty)=0$ or $=1$. We call η a **dual passive boundary point** in the first case and an **entrance boundary point** in the second.

Fix $\{q,\pi\}$, and let $\mathfrak{X}=(X_t, P_x)$ be the minimal chain associated with $\{q,\pi\}$. Let $(\partial S)_{\mathrm{ex}}$ and $(\partial S^*)_{\mathrm{en}}$ be the sets of exit and entrance boundary points, respectively. If $P_x(X_{\zeta-}\in(\partial S)_{\mathrm{ex}})=0$, (8) has no solution other than the minimal solution. But if $P_x(X_{\zeta-}\in(\partial S)_{\mathrm{ex}})>0$ for some $x\in S$, (8) has infinitely many solutions. Furthermore, if $(\partial S^*)_{\mathrm{en}}\neq\varnothing$, we can obtain infinitely many solutions satisfying (8) and (9). There are many open problems in this connection.

J. Concluding Remarks

(1) **Literature and some historical remarks.** Besides the monographs on countable Markov chains, such as [6, 8], many of the standard textbooks on stochastic processes contain chapters on Markov chains (e.g., [1–4]). Among others, W. Feller [1] gave an elementary and elegant treatment of the basic theory of Markov chains with discrete parameter, which the bulk of the recent textbooks follow. The terminology for the classification of states we use here follows the first edition of [1]. Since the second edition, Feller has used the terms "**persistent**" and "**ergodic**" instead of "recurrent" and "positive recurrent," respectively. S. Karlin's book [3] contains some excellent chapters on applications of Markov chains.

The potential theory of nonrecurrent Markov chains is essentially contained in that of general nonrecurrent Markov processes due to G. A. Hunt [5]. The potential theory of recurrent chains was developed by J. G. Kemeny and J. L. Snell (*J. Math. Anal. Appl.*, 3 (1961)). A greater part of Kemeny, Snell, and

A. W. Knapp [6] is devoted to the potential-theoretic aspects of Markov chains, including boundary theory. F. Spitzer's book [7] is a definitive work on the theory and application of random walks. The proof of normality of recurrent random walk is the highlight of the book. P. Lévy's article [9] is a pioneering work on Markov chains with continuous parameter involving instantaneous states. Feller and H. P. McKean first gave an example of Markov chains having only instantaneous states (*Proc. Nat. Acad. Sci. US*, 42 (1956)). K. L. Chung's monograph [8] is the only book devoted to the foundation of the theory of Markov chains with continuous parameter. The exposition of birth and death processes in this article follows Feller [12]. (For details of birth and death processes → [3].) The first systematic treatment of queuing theory from the point of view of stochastic processes is due to D. G. Kendall [13, 14], and this has greatly influenced subsequent work in this field. As the standard textbooks of queuing theory we mention N. U. Prabhu [16] and ch. 14 of Karlin [3]. The Martin boundary theory of nonrecurrent Markov chains was developed by J. L. Doob [18], Hunt [19], and T. Watanabe [20]. There are many papers on the extension of the boundary theory to more general Markov processes (e.g., H. Kunita and T. Watanabe, *Illinois J. Math.*, 9 (1965)). The exposition in this article was taken from Kunita and Watanabe (*Sûgaku*, 13 (1961); 14 (1962)). Feller (*Trans. Amer. Math. Soc.*, 83 (1956)) was the first to introduce the notion of an †ideal boundary of Markov chains. He also studied the †boundary conditions for Kolmogorov's equations (*Ann. Math.*, (2) 65 (1957)).

(2) **General Markov chains**. A Markov process with discrete parameter is often called a Markov chain whether the state space is countable or not. We here call such a process a **general Markov chain**. A considerable amount of the results on Markov chains—on potential theory, limit theorems, random walks, and so on—are extended to general Markov chains. Classical results are found in Doob [2]. Later, Sparre E. Andersen (*Math. Scand.*, 1 (1953); 2 (1954)) initiated the fluctuation theory of 1-dimensional nonlattice random walks. T. E. Harris carried out an important study on the existence of invariant measures of recurrent general Markov chains (*Proc. 3rd Berkeley Symp. on Math. Stat. Prob.*, vol. 2, 1956). S. Orey (*Pacific J. Math.*, 9 (1959)) extended the limit theorems of Markov chains to a class of recurrent general Markov chains (called Harris chains). A systematic treatment of general Markov chains is given in D. Revuz [21].

References

[1] W. Feller, An introduction to probability theory and its applications, I, Wiley, third edition, 1968.
[2] J. L. Doob, Stochastic processes, Wiley, 1953.
[3] S. Karlin, A first course in stochastic processes, Academic Press, 1966; second edition, with H. M. Taylor, Academic Press, 1975; Karlin and Taylor, A second course in stochastic processes, Academic Press, 1981.
[4] L. Breiman, Probability, Addison-Wesley, 1968.
[5] G. A. Hunt, Markov processes and potentials, Illinois J. Math., 1 (1957), 44–93, 316–369; 2 (1958), 151–213.
[6] J. G. Kemeny, J. L. Snell, and A. W. Knapp, Denumerable Markov chains, Van Nostrand, 1966.
[7] F. Spitzer, Principles of random walk, Springer, second edition, 1976.
[8] K. L. Chung, Markov chains with stationary transition probabilities, Springer, 1960.
[9] P. Lévy, Systèmes markoviens et stationaires, Cas dénombrable, Ann. Sci. Ecole Norm. Sup., (3) 68 (1951), 327–381.
[10] D. Williams, The Q-matrix problem for Markov chains, Bull. Amer. Math. Soc., 81 (1975), 1115–1118.
[11] D. Williams, The Q-matrix problem 2: Kolmogorov backward equations, Lecture notes in math. 511, Springer, 1976, 505–519.
[12] W. Feller, The birth and death processes as diffusion processes, J. Math. Pures Appl., (9) 38 (1959), 301–345.
[13] D. G. Kendall, Some problems in the theory of queues, J. Roy. Stat. Soc., (B) 13 (1951), 151–185.
[14] D. G. Kendall, Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain, Ann. Math. Statist., 24 (1954), 338–354.
[15] T. L. Saaty, Elements of queuing theory with applications, McGraw-Hill, 1961.
[16] N. U. Prabhu, Queues and inventories, Wiley, 1965.
[17] A. A. Borovkov, Stochastic processes in queuing theory, Springer, 1976. (Original in Russian, 1972.)
[18] J. L. Doob, Discrete potential theory and boundaries, J. Math. Mech., 8 (1959).
[19] G. A. Hunt, Markov chains and Martin boundaries, Illinois J. Math., 4 (1960), 313–340.
[20] T. Watanabe, On the theory of Martin boundaries induced by countable Markov processes, Mem. Coll. Sci. Univ. Kyoto, 33 (1960–1961), 39–108.

[21] D. Revuz, Markov chains, North-Holland and Elsevier, 1975.

261 (XVII.5) Markov Processes

A. General Remarks

Let $\{X_t\}_{t\in T}$ be a †stochastic process defined on the †probability space $(\Omega, \mathfrak{B}, P)$. The †state space S of X_t is the set of real numbers $\mathbf{R}$ or N-dimensional Euclidian space $\mathbf{R}^N$. In general, S may be a †locally compact Hausdorff space satisfying the second countability axiom. T is an interval $[0,\infty)$ or a set $\{0,1,2,\ldots\}$. (T may also be any interval in the real line or a set $\{\ldots,-2,1,0,1,2,\ldots\}$.) We call this $\{X_t\}_{t\in T}$ a **Markov process** if, for any choice of points $s_1<s_2<\ldots<s_n<t$ in T, the †conditional probability distribution of X_t relative to $X_{s_1}, X_{s_2}, \ldots, X_{s_n}$ is equal to the conditional probability distribution of X_t relative to X_{s_n}. Namely, for $A\in\mathfrak{B}(S)$ and $x_1,\ldots,x_n\in S$,

(1) $$P(X_t\in A\mid X_{s_1}=x_1,\ldots,X_{s_n}=x_n) = P(X_t\in A\mid X_{s_n}=x_n),$$

where $\mathfrak{B}(S)$ is the least †σ-algebra that contains all open sets of S.

For a Markov process $\{X_t\}_{t\in T}$, the distribution of X_0 is called the **initial distribution** of $\{X_t\}_{t\in T}$. The conditional probability distribution $P(X_t\in A\mid X_s=x)$ is denoted by $p(s,x,t,A)$ and is called the **transition probability** of $\{X_t\}_{t\in T}$. This is a function of $s,t\in T$ $(s\leqslant t)$, $x\in S$, and $A\in\mathfrak{B}(S)$ that has the following properties:

(2) For fixed s and t, $P(s,x,t,A)$ is a †probability measure in A and is $\mathfrak{B}(S)$-measurable in x.

(3) $$P(s,x,s,A)=1 \quad \text{if} \quad x\in A, \qquad =0 \quad \text{if} \quad x\notin A.$$

(4) The **Chapman-Kolmogorov equality**,

$$P(s,x,u,A)=\int_S P(s,x,t,dy)P(t,y,u,A) \qquad (s<t<u).$$

In view of (1), the †finite-dimensional distribution of a Markov process is completely determined by its initial distribution and its transition probability.

Moreover, for a given function $P(s,x,t,A)$ satisfying (2), (3), and (4), and for a given probability distribution μ on $\mathfrak{B}(S)$, there exists a Markov process $\{X_t\}_{t\in T}$ with transition probability $P(s,x,t,A)$ and initial distribution μ. In this sense, the transition probability is a characteristic quantity of the Markov process.

If the transition probability depends only on the difference between s and t, that is, if there exists a function $P(t,x,A)$ of t, x, and A such that $P(s,x,t,A)=P(t-s,x,A)$, then the Markov process is called **temporally homogeneous**. If S is an additive group and there exists a function $P(s,t,A)$ of s, t, and A such that $P(s,x,t,A)=P(s,t,A-x)$, where $A-x=\{y-x\mid y\in A\}$, then the Markov process is called **spatially homogeneous**. When $S=\mathbf{R}^N$, a spatially homogeneous Markov process is an †additive process. A Markov process whose state space S is countable is called a †Markov chain (→ 260 Markov Chains). A Markov process whose sample path (→ Section B) is continuous with probability 1 (→ 407 Stochastic Processes) is called a †diffusion process (→ 115 Diffusion Processes). Consider the case $T=[0,\infty)$ and $S=\mathbf{R}^N$, and assume that the transition probability has a †density $p(s,x,t,y)$ with respect to Lebesgue measure and satisfies certain analytic conditions that ensure the continuity of the sample path, etc. Then A. N. Kolmogorov proved that $p(s,x,t,y)$ satisfies the †Fokker-Planck partial differential equation. Conversely, he raised the problem of finding conditions for the existence and uniqueness of the transition probability satisfying the given Fokker-Planck equation [1] (→ 115 Diffusion Processes). W. Feller extended this equation to an †integrodifferential equation of a certain type and solved the problem partially by classical analytic methods [2]. On the other hand, S. N. Bernstein [3] and P. Lévy [4] made probabilistic approaches to the problem, and K. Itô constructed Markov processes directly by solving the corresponding stochastic differential equations [5] (→ 406 Stochastic Differential Equations).

A profound study of the structure of Brownian motion and the additive process by P. Levy [4, 6], the theory of †martingales initiated by J. L. Doob [7], and the †stochastic calculus due to K. Itô [5] have served as useful apparatus in analyzing the structure of Markov processes probabilistically. The theory of Markov processes has also retained a close relationship to various aspects of mathematical analysis—especially to functional analysis and †potential theory. The link lies in an obvious fact that, by virtue of the formula

(5) $$T_tf(x)=\int_S p(t,x,dy)f(y),$$

the transition probability $p(t,x,A)$ of a (tem-

porarily homogeneous) Markov process induces a semigroup $\{T_t, t \geqslant 0\}$ of linear positive operators on a function space, e.g., on the space $B(S)$ of all bounded measurable functions on S.

The study of diffusion processes has played important roles in the theory of Markov processes. When $S=\mathbf{R}^1$, the structure of diffusion processes was completely clarified by W. Feller [8], E. B. Dynkin [9], K. Itô and H. P. McKean [10], and others (→ 115 Diffusion Processes). In succession, G. A. Hunt [11] and E. B. Dynkin [9] isolated fairly general and practically useful concepts in Markov processes, including the right continuity of sample paths and the strong Markov property. Based on these properties, probabilistic potential theory and the theory of additive functionals were developed.

We now present a formulation for the temporally homogeneous Markov process with continuous time parameter [9, 11–14].

B. Fundamental Notions

Adjoin a point ∂ to S as a point at infinity when S is noncompact and as an isolated point when S is compact, set $\bar{S}=S\cup\{\partial\}$, and let $\mathfrak{B}(\bar{S})$ be the σ-algebra that consists of all the Borel sets in $\bar{S}$. Let $\tilde{W}$ be the set of all right continuous functions $w(t)$ whose †discontinuities are at most of the first kind and such that $w(t)=\partial$ for $t\geqslant s$ if $w(s)=\partial$. By convention, we set $w(\infty)=\partial$. Let $\zeta(w)$ be the minimum of the t-values such that $w(t)=\partial$. For $w\in\tilde{W}$, w_t^- and w_t^+ in $\tilde{W}$ are defined as $w_t^-(s)=w(\min(t,s))$ and $w_t^+(s)=w(s+t)$, respectively. Let W be a subset of $\tilde{W}$ that is closed under the operation $w\rightarrow w_t^+$ and $w\rightarrow w_t^-$, and let $\mathfrak{B}=\mathfrak{B}(W)$ be the σ-algebra generated by the sets $\{w\in W\,|\,w(t)\in A\}$ $(A\in\mathfrak{B}(\bar{S}))$. We often write $X_t(w)$ for $w(t)$. The subclass of $\mathfrak{B}$ that consists of sets represented by $\{w\in W\,|\,w_t^-\in B\}$ $(B\in\mathfrak{B})$ is denoted by $\mathfrak{B}_t$. Suppose that the family of probability measures $\{P_x\}$ $(x\in\bar{S})$ on (W,B) satisfies the following conditions:

(6) For a fixed B in $\mathfrak{B}$, $P_x(B)$ is $\mathfrak{B}(\bar{S})$-measurable in x.

(7) $P_x(X_0(w)=x)=1$ for $x\in\bar{S}$.

(8) The **Markov property** $P_x(w_t^+\in B\,|\,\mathfrak{B}_t)=P_{X_t(w)}(B)$ holds with P_x-measure 1 for $B\in\mathfrak{B}$.

Then the triple $\mathfrak{M}=(X_t, W, P_x\,|\,x\in\bar{S})$ is called a **Markov process**. This is a mathematical model for the random motion of a particle moving in S whose †probability law is independent of its past history once the present position of the particle is known (→ 260 Markov Chains). We call S the **state space** of $\mathfrak{M}$, W the **path space**, and an element w in W a **path**, respectively. P_x represents the probability law of a particle that starts from x. In view of the interpretation that the particle vanishes if it reaches ∂, we call $\zeta(w)$ the **lifetime** (or **terminal time**) and ∂ the **terminal point**.

Now, for any $s_1<s_2<\ldots<s_n$ and t,

$$P_x(X_{s_n+t}\in A\,|\,X_{s_1},\ldots,X_{s_n})$$
$$=P_x(X_{s_n+t}\in A\,|\,X_{s_n})=P_{X_{s_n}}(X_t\in A)$$

holds with P_x-measure 1 for any x. This equality means that $\{X_t\}_{t\geqslant 0}$ is a Markov process on S in the sense mentioned before with initial distribution δ_x (a unit measure at x) and the transition probability given by $P(t,x,A)=P_x(X_t\in A)$. The restriction of this transition probability to $(t,x,A)\in[0,\infty)\times S\times\mathfrak{B}(S)$ satisfies the following properties:

(2′) For a fixed $t\geqslant 0$, $P(t,x,A)$ is $\mathfrak{B}(S)$-measurable in x and is a measure in $A\in\mathfrak{B}(S)$ with $P(t,x,S)\leqslant 1$.

(3′) $P(0,x,A)=\delta_x(A)$, $x\in S$, $A\in\mathfrak{B}(S)$.

(4′) $P(s+t,x,A)=\int_S P(s,x,dy)P(t,y,A)$.

(9) $\lim_{t\downarrow 0}T_tf(x)=f(x)$, $x\in S$, holds for any bounded continuous function f on S, where T_tf is defined by formula (5).

$\{T_t\}$ is called the **semigroup** of the Markov process $\mathfrak{M}$ because it enjoys the property $T_{s+t}=T_sT_t$ on $B(S)$. By convention, any function $f\in B(S)$ is extended to a function on $\bar{S}$ by setting $f(\partial)=0$. (9) then follows from the expression $T_tf(x)=E_x(f(X_t))$, where $E_x(\ )$ represents the integral with respect to P_x.

A Markov process $\mathfrak{M}$ is called **conservative** if $P(t,x,S)=1$ holds for every x in S and every t. A point x in S is called a **recurrent point** if for every neighborhood U of x and every $t>0$, the path starting from x returns to U after time t with probability 1. Under some conditions, x is recurrent if and only if $\int_0^\infty p(t,x,U)dt=\infty$ for any neighborhood U of x. $\mathfrak{M}$ is called **recurrent** if all points in S are recurrent. Otherwise, it is called **transient**. $\mathfrak{M}$ is of course conservative if it is recurrent. There are other instances in which $\{X_t\}$ is called recurrent, if a particle starting from any point in S reaches any neighborhood of any other point in S in finite time with probability 1. In this case, under suitable conditions for regularity, the †mixing property holds, whence follows the †ergodic property (→ 136 Ergodic Theory). If the transition probability of $\{X_t\}$ has an †invariant measure with total mass 1, †Birkhoff's individual ergodic theorem holds [7].

Let $\{\mathfrak{F}_t\}$ be an increasing family of σ-subalgebras of $\mathfrak{B}(W)$ such that X_t is $\mathfrak{F}_t$-

measurable for each $t \geqslant 0$. A $[0, \infty)$-valued function σ on W is said to be an ($\{\mathfrak{F}_t\}$-)**Markov time** (or **stopping time**) if $\{\omega \mid \sigma(\omega) \leqslant t\} \in \mathfrak{F}_t$, $t \geqslant 0$. The σ-algebra $\mathfrak{F}_\sigma$ is then defined by

$$\mathfrak{F}_\sigma = \{B \in \mathfrak{B} \mid B \cap \{\sigma \leqslant t\} \in \mathfrak{F}_t \text{ for any } t \geqslant 0\}.$$

If there exists a family $\{\mathfrak{F}_t\}$ as above and if for every $\{\mathfrak{F}_t\}$-Markov time σ and for any $x \in S$, $s > 0$, and $A \in \mathfrak{B}(\bar{S})$,

$$(10) \quad P_x(X_{\sigma+s} \in A \mid \mathfrak{F}_\sigma) = P_{X_\sigma}(X_s \in A)$$

holds with P_x-measure 1, then it is said that $\mathfrak{M}$ has the **strong Markov property** with respect to $\{\mathfrak{F}_t\}$, and such an $\mathfrak{M}$ is called a **strong Markov process**. Since a constant time is also a Markov time, the strong Markov property imposes stronger conditions than the Markov property. A sufficient condition for a Markov process $\mathfrak{M}$ to have the strong Markov property is that the space $C(S)$ of all bounded continuous functions on S be invariant under the semigroup of $\mathfrak{M}$. In this case, we can take $\mathfrak{F}_t = \bigcap_{s>t} \mathfrak{B}_s$.

It is convenient to enlarge the algebra $\mathfrak{B}_t$ as follows. Let μ be a measure on $(\bar{S}, \mathfrak{B}(\bar{S}))$, and write $\mathfrak{B}(\bar{S})_\mu$ for the completion of $\mathfrak{B}(\bar{S})$ by μ. Write $\bar{\mathfrak{B}}(\bar{S})$ for the intersection of all $(\mathfrak{B}(\bar{S}))_\mu$, where μ runs over all probability measures on $\mathfrak{B}(\bar{S})$. Let P_μ be the measure on $(W, \mathfrak{B})$ defined by $P_\mu(\Lambda) = \int_{\bar{S}} \mu(dx) P_x(\Lambda)$; and let $\bar{\mathfrak{B}}_t = \bigcap_\mu (\mathfrak{B}_t)_\mu$ and $\bar{\mathfrak{B}} = \bigcap_\mu (\mathfrak{B})_\mu$, where $(\mathfrak{B})_\mu$ is the completion of $\mathfrak{B}$ by P_μ and $(\mathfrak{B}_t)_\mu$ is the σ-algebra obtained by adjoining to $\mathfrak{B}_t$ all P_μ-null sets in $(\mathfrak{B})_\mu$. A Markov process $\mathfrak{M}$ is called a **Hunt process** if $\bar{\mathfrak{B}}_t = \bigcap_{s>t} \bar{\mathfrak{B}}_s$ for any $t \geqslant 0$, $\mathfrak{M}$ has the strong Markov property with respect to $\{\bar{\mathfrak{B}}_t\}$, and $\mathfrak{M}$ has left quasicontinuity in the following sense: If $\{\sigma_n\}$ is any increasing sequence of $\{\bar{\mathfrak{B}}_t\}$-Markov times and $\sigma = \lim \sigma_n$, then for $\sigma < \infty$, $\lim X_{\sigma_n} = X_\sigma$ holds except for a set of P_x-measure 0 for any $x \in S$. For a set A of $\bar{S}$, let

$$\sigma_A = \inf\{t \mid t > 0, \quad x_t \in A\} \text{ if such a } t \text{ exists,}$$

$$= \infty \text{ if } x_t \notin A \text{ for all } t > 0.$$

Then σ_A is called the **hitting time** for A. (Sometimes the condition $t > 0$ in the definition of σ_A is replaced by $t \geqslant 0$.) If $\mathfrak{M}$ is a Hunt process, then σ_A is a $\{\bar{\mathfrak{B}}_t\}$-Markov time for any Borel set A [14, 15]. For a subset A, $\tau_A = \sigma_{A^c}$ is called an exit time from A. If $\mathfrak{M}$ is a strong Markov process, the exit time τ_a from a point a is subject to the exponential distribution

$$P_a(\tau_a > t) = e^{-\lambda(a)t}, \quad 0 \leqslant \lambda(a) \leqslant \infty.$$

In particular, a is called an **instantaneous state** if $\lambda(a) = \infty$ and a **trap** if $\lambda(a) = 0$. For a Hunt process $\mathfrak{M}$, $P_x(B) = 0$ or 1 if B is in $\bar{\mathfrak{B}}_0$. This is called **Blumenthal's zero-one law**.

A function $P(t, x, A)$ of $t \geqslant 0$, $x \in S$, and $A \in B(S)$ is said to be a **transition function** on S if it satisfies (2′), (3′), and (4′). Denote by $C_\infty(S)$ the set of those functions in $C(S)$ vanishing at infinity. A transition function defines by (5) a linear operator on $B(S)$ which satisfies the semigroup property $T_s T_t = T_{s+t}$. If T_t leaves the space $C_\infty(S)$ invariant for every $t \geqslant 0$ and if (9) holds for any $f \in C_\infty(S)$, then the transition function (resp. semigroup T_t) is called a **Feller transition function** (resp. Feller semigroup). Any Feller transition function on S admits a Hunt process; namely, for a given Feller transition function $P(t, x, A)$, there exists a Hunt process $\mathfrak{M}$ (on $\bar{S}$) whose transition probability coincides with $P(t, x, A)$ for $t \geqslant 0$, $x \in S$, and $A \in \mathfrak{B}(S)$. Such a process $\mathfrak{M}$ is often called a Feller process. For instance, any spatially homogeneous Markov process on $\mathbf{R}^N$ is a Feller process. If a Feller transition function satisfies an additional condition that, for any compact set K and neighborhood G of K,

$$t^{-1} P(t, x, S - G) \to 0, \quad t \to 0,$$

uniformly on K, then the associated Hunt process $\mathfrak{M}$ is a diffusion, namely, P_x (X_t is continuous in $t < \zeta$) $= 1$, $x \in S$ [2, 7].

C. Generators of Markov Processes

For a given Markov process $\mathfrak{M}$, the generator of $\mathfrak{M}$ is defined by

$$(11) \quad \mathfrak{G}f(x) = \lim_{t \downarrow 0} t^{-1}(T_t f(x) - f(x)), \quad x \in S,$$

where $\{T_t\}$ is the semigroup of $\mathfrak{M}$. The choice of the domain $\mathfrak{D}(\mathfrak{G})$ of $\mathfrak{G}$ depends on the situation. When $\{T_t\}$ is a Feller semigroup, it is a strongly continuous semigroup on $C_\infty(S)$. Let $\mathfrak{G}$ be its infinitesimal generator in the sense of Hille and Yosida (→ 378 Semigroups of Operators and Evolution Equations). Then $\mathfrak{D}(\mathfrak{G})$ consists of those functions $f \in C_\infty(S)$ for which the convergence in the right-hand side of (11) takes place uniformly on S. A necessary and sufficient condition for a linear operator on $C_\infty(S)$ to be the infinitesimal generator of a Feller semigroup is known [8].

For a general strong Markov process $\mathfrak{M}$, one way of defining $\mathfrak{D}(\mathfrak{G})$ is as follows [16, 17]. Denote by $\tilde{C}(S)$ the set of all functions $f \in B(S)$ which are finely continuous in the sense that

$$(12) \quad P_x(f(X_t) \text{ is right continuous in } t \geqslant 0) = 1, \quad x \in S.$$

$\mathfrak{D}(\mathfrak{G})$ is then defined as the set of those functions $f \in \tilde{C}(S)$ for which the convergence in the right-hand side of (11) takes place boundedly in x with limit functions belonging to $\tilde{C}(S)$. $\mathfrak{G}$ is called the **generator** of the strong Markov process $\mathfrak{M}$. The following three conditions are

equivalent:

(13) $f\in\mathfrak{D}(\mathfrak{G}),\quad \mathfrak{G}f=g.$

(14) $f, g\in C(S)$ and $T_tf(x)-f(x)$

$$=\int_0^t T_sg(x)\,ds,\quad t\geqslant 0.$$

(15) $f, g\in\tilde{C}(S)$ and $f(X_t)-f(X_0)-\int_0^t g(X_s)\,ds$ is a †martingale on $(W,\mathfrak{B}_t,P_x)$ for each $x\in S$.

Furthermore, let σ be a Markov time such that $E_x(\sigma)<\infty$. Then we have **Dynkin's formula**: for $f\in\mathfrak{D}(\mathfrak{G})$,

$$f(x)=-E_x\left(\int_0^\sigma \mathfrak{G}f(X_t)\,dt\right)+E_x(f(X_\sigma)).$$

If $\mathfrak{G}f$ is continuous at x, then

$$\mathfrak{G}f(x)=\lim_{U\downarrow\{x\}}\frac{E_x(f(X_{\tau_U}))-f(x)}{E_x(\tau_U)}\quad \text{if } x \text{ is not a trap,}$$
$$=0\qquad\qquad \text{if } x \text{ is a trap,}$$

where U is an open neighborhood of x.

This representation of $\mathfrak{G}$ suggests that, when S is a smooth manifold, $\mathfrak{G}f$ for smooth functions f often becomes an elliptic partial differential operator or a mixture of it and an integral operator satisfying a certain maximum principle. In many practical cases, the data given to us are the coefficients of such a linear operator L carrying a certain linear space $\mathfrak{D}(L)\subset B(S)$ into $B(S)$. Thus a problem arises as to the existence and uniqueness of a strong Markov process $\mathfrak{M}$ whose generator $\mathfrak{G}$ is an extension of L. In connection with (15), Stroock and Varadhan [18] formulated this problem as the **martingale problem** concerning the existence and uniqueness of a probability measure P_x on W such that $P_x(X_0=x)=1$ and $f(X_t)-f(X_0)-\int_0^t Lf(X_s)\,ds$ is a martingale on $(W,\mathfrak{B}_t,P_x)$ for every $f\in\mathfrak{D}(L)$. Since this formulation refers directly to probability measures on W, it is useful in the study of stochastic differential equations and of the convergence of Markov processes as well.

Let m be a positive Radon measure on S which is strictly positive on each nonempty open set. A Markov process $\mathfrak{M}$ is said to be ***m*-symmetric** if

$$\int_S T_tf(x)g(x)m(dx)$$
$$=\int_S f(x)T_tg(x)m(dx)\quad(\leqslant+\infty)$$

holds for any $t>0$ and nonnegative measurable functions f, g. Then $\{T_t\}$ is realized uniquely as a strongly continuous semigroup of symmetric operators on the real L^2-space $L^2(S;m)$. Let $\mathfrak{G}$ be its infinitesimal generator. $\mathfrak{G}$ is then a nonpositive definite self-adjoint operator on $L^2(S;m)$, and the following definition of the symmetric form $\mathscr{E}$ on $L^2(S;m)$ makes sense: $\mathfrak{D}[\mathscr{E}]=\mathfrak{D}(\sqrt{-\mathfrak{G}})$, $\mathscr{E}(f,g)=(\sqrt{-\mathfrak{G}}\,f,\sqrt{-\mathfrak{G}}\,g)$, $f,g\in\mathfrak{D}[\mathscr{E}]$, where $(\ ,\ )$ denotes the inner product in $L^2(S;m)$. $\mathscr{E}$ is called the **Dirichlet form** of an m-symmetric Markov process $\mathfrak{M}$. A Dirichlet form (on $L^2(S;m)$) is by definition a closed symmetric form $\mathscr{E}$ on $L^2(S;m)$ such that $u\in\mathfrak{D}[\mathscr{E}]$ implies $v=\min(\max(u,0),1)\in\mathfrak{D}[\mathscr{E}]$ and $\mathscr{E}(v,v)\leqslant\mathscr{E}(u,u)$. A Dirichlet form is called **regular** if the space $\mathfrak{D}[\mathscr{E}]\cap C_0(S)$ is dense in $\mathfrak{D}[\mathscr{E}]$ and in $C_0(S)$, where $C_0(S)$ is the space of continuous functions with compact support. For a given regular Dirichlet form $\mathscr{E}$, there exists uniquely in a certain sense an m-symmetric Hunt process whose Dirichlet form is the given $\mathscr{E}$ [19]. In many cases of symmetric Markov processes, the form $\mathscr{E}(f,f)$ rather than the generator $\mathfrak{G}f$ admits an explicit expression for smooth functions f.

When $\mathfrak{M}$ is the N-dimensional Brownian motion, the generator of its semigroup on $L^2(\mathbf{R}^N)$ is given by $\mathfrak{G}u=\frac{1}{2}\Delta u$, $\mathfrak{D}(\mathfrak{G})=\{u\in L^2(\mathbf{R}^N)\mid\Delta u\in L^2(\mathbf{R}^N)\}$, and its Dirichlet form is given by

$$\mathscr{E}(u,v)=\frac{1}{2}\int_{\mathbf{R}^N}\operatorname{grad}u\cdot\operatorname{grad}v\,dx,$$

$$\mathfrak{D}[\mathscr{E}]=\left\{u\in L^2(\mathbf{R}^N)\,\middle|\,\frac{\partial u}{\partial x_i}\in L^2(\mathbf{R}^N),\quad 1\leqslant i\leqslant N\right\}.$$

D. Markov Processes and Potential Theory

Analytic notions and relations in classical potential theory can be interpreted in terms of Brownian motion (→ 45 Brownian Motion). The relevant probabilistic notions and relations have been formulated not only for Brownian motion but also for a general Hunt process under the name of probabilistic potential theory [11, 14, 20].

Let $\mathfrak{M}$ be a Hunt process. A set $A\subset\bar{S}$ is called **nearly Borel measurable** if for each probability measure μ on $\bar{S}$ there exist A_1, $A_2\in\mathfrak{B}(\bar{S})$ such that $A_1\subset A\subset A_2$ and $P_\mu(X_t\in A_2-A_1$ for some $t\geqslant 0)=0$. $\mathfrak{B}^n(S)$ denotes the family of all nearly Borel subsets of S. Then $\mathfrak{B}(S)\subset\mathfrak{B}^n(S)\subset\overline{\mathfrak{B}(S)}$. The hitting time σ_A for a nearly Borel set A is still a $\{\mathfrak{B}_t\}$-Markov time, and the probability $P_x(\sigma_A>0)$ is either zero or one according to Blumenthal's zero-one law. x is said to be a **regular point** of A in the former case and an **irregular point** of A in the latter case. Let A^r denote the totality of regular points of A.

A set $A\subset S$ is called **finely open** if for any $x\in A$ there exists a set $C=C(x)\in\mathfrak{B}^n$ containing

$S-A$ such that $x \notin C^r$. Any open set is finely open by virtue of the right continuity of the paths. The **fine topology** defined by the fine open subsets of S is therefore finer than the original topology. A nearly Borel measurable function f on S is finely continuous if and only if it satisfies (12). A nonnegative $\overline{\mathfrak{B}(S)}$-measurable function f is called **α-excessive** ($\alpha \geqslant 0$) if $e^{-\alpha t}T_t f \leqslant f$ for any t and $e^{-\alpha t}T_t f(x) \rightarrow f(x)$ $(t \rightarrow 0)$. A 0-excessive function is called simply **excessive**. Any α-excessive function is nearly Borel measurable and finely continuous. The function $f(x)=G_\alpha g(x)=\int_0^\infty e^{-\alpha t}T_t g(x)\,dt$ (G_α is called the **resolvent** of $\mathfrak{M}$) is α-excessive for any nonnegative $\overline{\mathfrak{B}(S)}$-measurable function g. If f is α-excessive and $A \subset S$ is nearly Borel measurable, then the function $p_A^\alpha f(x)=E_x(e^{-\alpha\sigma_A}f(X_{\sigma_A}); \sigma_A<\infty)$ is α-excessive again. When $\mathfrak{M}$ is the Brownian motion on $\mathbf{R}^N$ ($N \geqslant 3$), the class of excessive functions coincides with the class of nonnegative †superharmonic functions. Also, the fine topology is identical with the †Cartan fine topology.

A set $A \subset S$ is called **polar** if A is contained in a set $D \in \mathfrak{B}^n$ such that $P_x(\sigma_D<\infty)=0$ for any $x \in S$. A is called **thin** if A is contained in a set $D \in \mathfrak{B}^n$ such that D^r is empty. A set contained in a countable union of thin sets is called **semipolar**. If A is semipolar, then $X_t \in A$ occurs for at most countably many values of t with P_x-measure 1, $x \in S$. The set $A-A^r$ is semipolar for any $A \in \mathfrak{B}^n$. Any polar set is semipolar. The converse is also true when $\mathfrak{M}$ is Brownian motion but not true when $\mathfrak{M}$ is the uniform motion to the right on $\mathbf{R}^1$.

Under certain conditions of duality or symmetry for the Hunt process, the notion of capacity can be introduced by generalizing the classical †Newtonian capacity. Then a set of zero capacity is identified with a polar set or its weaker version [11, 19]. The stochastic solution of the classical †Dirichlet problem can also be formulated for a wide class of diffusions. Thus, for a domain $D \subset S$, the solution of the equation $\mathfrak{G}u(x)=0$, $x \in D$, with boundary function f is expressed as $u(x)=p_{\partial D}^0 f(x)$, $x \in D$, and the boundary behavior of u is studied probabilistically [9].

We have so far discussed mainly potential theory for transient Markov processes. However, we can also establish potential theory for recurrent Markov processes following the model of the †logarithmic potential in the case of 2-dimensional Brownian motion.

E. Additive Functionals

The notion of additive functionals was first introduced in relation to the study of time change of Markov processes, and in particular, to the study of the †local time of Brownian motion. Later it was studied in relation to potential theory and †martingale theory. Additive functionals play an important role in the study of Markov processes. For a Markov process $\mathfrak{M}$, a function $\varphi=\varphi_t(w)$ of t and w is called a (right continuous and homogeneous) **additive functional** if the following conditions are satisfied: (1) $-\infty<\varphi_t(w)\leqslant\infty$; (2) $\varphi_t(w)$ is right continuous in t ($t<\zeta$), $\varphi_{\zeta-}$ exists, and $\varphi_t=\varphi_{\zeta-}$ for $t \geqslant \zeta$; (3) $\varphi_t(w)$ is $\mathfrak{B}_t$-measurable in w for fixed t; and (4) for any t and s,

(16) $\varphi_{s+t}(w)=\varphi_s(w)+\varphi_t(w_s^+)$ holds with P_x-measure 1 for any $x \in S$.

We call φ a **perfect additive functional**, if it satisfies (1), (2), (3) and satisfies (16) for all w. Two additive functionals are **equivalent** if they are equal except on a set of P_x-measure 0 for any t and x, and φ is called nonnegative if there exists a nonnegative additive functional which is equivalent to φ. The concept of continuous additive functionals can be defined similarly. The function $\alpha=\alpha_t(w)$ of t and w is called a (right continuous and homogeneous) **multiplicative functional** if it can be expressed by an additive functional φ as

(17) $\alpha_t(w)=\exp(-\varphi_t(w))$.

An additive functional φ is said to be **natural** if φ_t and the path X_t of M have no jumps in common with P_x-measure 1 for any $x \in S$.

For a nonnegative additive functional φ,

(18) $$u_\alpha(x)=E_x\left(\int_0^\infty e^{-\alpha t}\,d\varphi_t\right)$$

is α-excessive. Let $\mathfrak{M}$ be a Hunt process. A finite α-excessive function u_α can be expressed in the form (18) by a unique nonnegative natural additive functional φ if for every increasing sequence $\{\sigma_n\}$ of Markov times such that $\sigma_n \uparrow \zeta$, $E_x(e^{-\alpha\sigma_n}u_\alpha(X_{\sigma_n}))\rightarrow 0$ holds. It is possible to choose a continuous φ corresponding to u_α if and only if for every increasing sequence $\{\sigma_n\}$ of Markov times, $E_x(e^{-\alpha\sigma_n}u_\alpha(X_{\sigma_n}))\rightarrow E_x(e^{-\alpha\sigma}u_\alpha(X_\sigma))$ holds, where $\sigma=\lim\sigma_n$. In particular, u_α may correspond to a continuous φ if the following two conditions are satisfied: (1) u_α is bounded and $e^{-\alpha t}T_t u_\alpha(x)\rightarrow 0$ $(t\rightarrow\infty)$, (2) $e^{-\alpha t}T_t u_\alpha(x)\rightarrow u_\alpha(x)$ $(t \rightarrow 0)$ uniformly in x [4]. In the case of Brownian motion, any nonnegative continuous additive functional φ can be characterized by a certain σ-finite measure μ on S charging no polar set, which is called a smooth measure [9, 21]. Such a characterization can be extended to more general Hunt processes satisfying certain conditions of duality or symmetry with respect to a basic measure m, and the correspondence between φ and μ is specified by the following formula of Revuz [22]:

For any function $f \geqslant 0$,

$$(19) \quad \lim_{t \downarrow 0} \frac{1}{t} E_m\left(\int_0^t f(X_s)\,d\varphi_s\right) = \int_S f(x)\mu(dx).$$

An additive functional φ of a Hunt process $\mathfrak{M}$ is called a **martingale additive functional** if $E_x(\varphi_t^2) < \infty$ and $E_x(\varphi_t) = 0$ for any $t > 0$ and $x \in S$. In fact, φ_t is then a square integrable †martingale on $(W, \{\mathfrak{B}_t\}_{t \geqslant 0}, P_x)$ for each $x \in S$. Let $\mathscr{M}$ be the set of all martingale additive functionals. The study of the class $\mathscr{M}$ constitutes a special aspect in the general theory of †semimartingales [15, 16, 23–26]. In particular, the quadratic variation $\langle \varphi \rangle$ of $\varphi \in \mathscr{M}$ can be realized as a nonnegative continuous additive functional that satisfies $E_x(\langle \varphi \rangle_t) = E_x(\varphi_t^2)$ for all $x \in S$ and $t > 0$. The additive functional $\langle \varphi, \psi \rangle$ is similarly defined for $\varphi, \psi \in \mathscr{M}$. Consider for $\varphi \in \mathscr{M}$ a measurable function f on S such that $E_x(\int_0^t f(X_s)^2\,d\langle \varphi \rangle_s)$ is finite. Then the **stochastic integral** $f \cdot \varphi$ can be defined as an element of $\mathscr{M}$ satisfying

$$(20) \quad E_x((f \cdot \varphi)_t \psi_t) = E_x\left(\int_0^t f(X_s)\,d\langle \varphi, \psi \rangle_t\right), \quad \psi \in \mathscr{M}.$$

$(f \cdot \varphi)_t$ is often denoted by $\int_0^t f(X_s)\,d\varphi_s$. The space of all continuous functionals in $\mathscr{M}$ is denoted by $\mathscr{M}_c$. Its orthogonal complement in the sense of $\langle\ ,\ \rangle$ is denoted by $\mathscr{M}_d$. The structure of $\mathscr{M}_d$ is known, and each functional in $\mathscr{M}_d$ can be represented by means of the so-called Lévy system, which is a pair consisting of a certain kernel on S and a certain nonnegative continuous additive functional [24, 25]. If $\mathscr{M}$ is an N-dimensional Brownian motion, then $\mathscr{M} = \mathscr{M}_c$, and any $\varphi \in \mathscr{M}$ takes the form

$$(21) \quad \varphi_t = \sum_{i=1}^N \int_0^t b_i(X_s)\,dX_t^i,$$

where the integral appearing on the right-hand side is defined by (20) for the ith coordinate process $X_t^i - X_0^i \in \mathscr{M}$ and for a measurable function b_i on $\mathbf{R}^N$ with $E_x(\int_0^t b_i(X_s)^2\,ds) < \infty$, $x \in \mathbf{R}^N$ [23, 28].

Replacing (16) by the "associative law"

$$\varphi_t^s(w) + \varphi_u^t(w) = \varphi_u^s(w),$$

we can also define temporally inhomogeneous additive functionals.

F. Transformation of Markov Processes

There are several methods by which a given Markov process can be transformed to a new one. Here we mention some important transformations.

Transformation by a Multiplicative Functional. For a Markov process $\mathfrak{M}$, let α be a multiplicative functional such that $E_x(\alpha_t) \leqslant 1$ and $P_x(\alpha_0 = 1) = 1$. Set

$$P^\alpha(t, x, \Gamma) = E_x(\alpha_t \chi_\Gamma(X_t)) \quad (\Gamma \in \mathfrak{B}(S)),$$

$$P^\alpha(t, x, \{\partial\}) = 1 - P^\alpha(t, x, S).$$

Then $P^\alpha(t, x, E)$ is a transition probability on $\bar{S}$ and corresponds to a Markov process $\mathfrak{M}^\alpha = (S, W^\alpha, P_x, x \in \bar{S})$. We call $\mathfrak{M}^\alpha$ a **transformation of $\mathfrak{M}$ by a multiplicative functional** α. It is possible to choose $W^\alpha = \tilde{W}$. If $\mathfrak{M}$ is a strong Markov process, so is $\mathfrak{M}^\alpha$. Conversely, let $\mathfrak{M}$ and $\mathfrak{M}'$ be two Markov processes with the same path space $\tilde{W}$ on the same state space S. If the probability law P_x' of $\mathfrak{M}'$ is absolutely continuous relative to P_x of $\mathfrak{M}$, then $\mathfrak{M}'$ is a transformation of $\mathfrak{M}$ by a certain multiplicative functional of $\mathfrak{M}$. This transformation includes killing, transformation by drift, and superharmonic transformation as special cases.

(1) Killing. Transformation by a multiplicative functional α is called **killing** if $0 \leqslant \alpha_t \leqslant 1$ holds. In fact, $\mathfrak{M}^\alpha$ can be constructed as follows: A particle going along a path w of $\mathfrak{M}$ is "killed" (jumped to ∂) in such a way that its surviving probability up to time t is $\alpha_t(w)$. For a nonnegative bounded continuous function $c(x)$ on S, let

$$\alpha_t(w) = \exp\left(-\int_0^t c(X_s(w))\,ds\right).$$

Then α satisfies the condition given previously. If the semigroup $\{T_t\}$ of $\mathfrak{M}$ is a Feller semigroup and has the generator $\mathfrak{G}$, then the semigroup $\{T_t\}$ of $\mathfrak{M}^\alpha$ is also a Feller semigroup, its generator $\mathfrak{G}^\alpha$ has the same domain as $\mathfrak{G}$, and $\mathfrak{G}^\alpha = \mathfrak{G} - c$ [29].

(2) Transformation by drift. For $\varphi \in \mathscr{M}$, let $\alpha_t = \exp\{\varphi_t - \langle \varphi \rangle_t / 2\}$. Then α_t is a multiplicative functional. The transformation determined by α is called a **transformation by drift**. Let $\mathfrak{M}$ be the N-dimensional Brownian motion and $\varphi \in \mathscr{M}$ be the functional expressed as in (21) with bounded $b_1, \ldots, b_N$; then

$$\langle \varphi \rangle_t = \int_0^t \sum_{i=1}^N b_i^2(X_s)\,ds,$$

and the above formula for α gives a transformation by drift. Moreover, if $b_1, \ldots, b_N$ are in $C_\infty(S)$, then the semigroup of $\mathfrak{M}^\alpha$ is a Feller semigroup, and for a bounded function f with bounded continuous derivatives up to the second order, f is in the domain of $\mathfrak{G}^\alpha$ and $\mathfrak{G}^\alpha f = (1/2)\Delta f + \sum_{i=1}^N b_i(\partial f / \partial x_i)$ [9, 16].

(3) Superharmonic transformation. Let u be an excessive function of $\mathfrak{M}$ and $A = \{x \mid 0 < u(x) < \infty\}$. Set

$$\alpha_t(w) = u(X_t(w))/u(X_0(w)) \quad \text{if} \quad X_0(w) \in A$$
$$= 0 \quad \text{if} \quad X_0(w) \notin A.$$

Then α_t is a multiplicative functional. The

transformation defined by α, first introduced by Doob, is called a **superharmonic transformation**. The transition probability $P(t,x,\Gamma)$ of $\mathfrak{M}^\alpha$ is equal to $u(x)^{-1}\int_\Gamma P(t,x,dy)u(y)$ if $x\in A$, 0 if $x\notin A$ and $t>0$, and $\delta_x(\Gamma)$ if $x\notin A$ and $t=0$, for Γ in $\mathfrak{B}(S)$. In particular, if $\mathfrak{M}$ is a Feller process and u is a continuous function such that $0<c\leqslant u\leqslant k<\infty$, then $\mathfrak{M}^\alpha$ is also a Feller process and $\mathfrak{G}^\alpha f=u^{-1}\mathfrak{G}(uf)$, where the domain of $\mathfrak{G}^\alpha$ is the set of f for which uf is in the domain of $\mathfrak{G}$.

Time Change. We understand the term **time change** in a broad sense, including the following two important special cases.

(4) Time change by an additive functional. Let $\mathfrak{M}$ be a Hunt process and φ be a nonnegative continuous additive functional such that $P_x(\varphi_0=0)=1$. Set $S^*=\{x\mid P_x(\varphi_t(w)>0$ for every $t>0)=1\}$, and assume that S^* is locally compact. Let $\tilde{W}^*$ be the set of all right continuous functions on S^*. They have discontinuities of at most the first kind, and $\mathfrak{B}^*$ is the σ-algebra on $\tilde{W}^*$ generated by all †Borel cylinder sets. Set $P_x^*(B)=P_x(X_{\varphi_t^{-1}(w)}(w)\in B)$ $(B\in\mathfrak{B}^*)$, where φ_t^{-1} is a right continuous inverse function of φ_t. Then $\mathfrak{M}^*=\{S^*,\tilde{W}^*,P_x^*, x\in\bar{S}^*\}$ is a Hunt process on S^*, and we say that $\mathfrak{M}^*$ is obtained by time change from $\mathfrak{M}$ by φ. Roughly speaking, $\mathfrak{M}^*$ can be considered to be a Markov process with paths $X_t^*(w)=X_{\varphi_t^{-1}(w)}(w)$. The resolvent of $\mathfrak{M}^*$ is given by $E_x(\int_0^\infty e^{-\lambda\varphi_t}f(X_t)d\varphi_t)$. Suppose that $a(x)$ is a continuous function on S such that $0<c\leqslant a(x)\leqslant k<\infty$, and set $\varphi_t(w)=\int_0^t a(X_s(w))ds$. Then $S^*=S$, $\tilde{W}^*=\tilde{W}$, and $\mathfrak{B}^*=\mathfrak{B}$, and $\mathfrak{M}^*$ has the same fine topology as $\mathfrak{M}$. The domain of the generator $\mathfrak{G}^*$ of $\mathfrak{M}^*$ coincides with that of $\mathfrak{G}$ of $\mathfrak{M}$ and $\mathfrak{G}^*f=a^{-1}\mathfrak{G}f$. Let $\mathfrak{M}$ and $\mathfrak{M}'$ be Markov processes with the same state space and the same path space $\tilde{W}$. If they have the same hitting probabilities $\{H_K(x,\cdot)=P_x(X_{\sigma_K}\in\cdot)\}$, then each one can be obtained from the other by time change by a strictly increasing additive functional. The converse is also true [14].

(5) Subordination. This concept, introduced by Bochner, was then extended as follows: Let $e^{-\psi(\lambda)}$ be the †Laplace transform of an †infinitely divisible distribution with support $[0,\infty)$, and let $F_t(\cdot)$ be the distribution with Laplace transform $e^{-t\psi(\lambda)}$. Let $\{T_t\}$ be a Hille-Yosida semigroup on a certain †Banach space, and set $T_t^\psi=\int_0^\infty T_sF_t(ds)$ (†Bochner integral). Then $\{T_t^\psi\}$ is also a semigroup and is called the **subordination** of $\{T_t\}$ by ψ. If $\mathfrak{G}$ is a generator of $\{T_t\}$, then $-\psi(-\mathfrak{G})$ is a generator of $\{T_t^\psi\}$.

In particular, we can assume $\{T_t\}$ to be a nonnegative semigroup on $C(S)(C_\infty(S))$ such that $T_t1=1$, and $\{X_t\}$ to be a Markov process corresponding to the semigroup T_t. Let $\{\Psi(t)\}$ be an additive process that is independent of $\{X_t\}$ and satisfies $E(e^{-\lambda\Psi(t)})=e^{-t\Psi(\lambda)}$. Set $Y_t(\omega)=X_{\Psi(t,\omega)}(\omega)$; then $\{Y_t\}$ is a Markov process corresponding to the semigroup $\{T_t^\psi\}$. The operation by which we obtain $\{Y_t\}$ from $\{X_t\}$ by using $\{\Psi(t)\}$ is also called **subordination**. In particular, if $\{\Psi(t)\}$ is a one-sided stable process of the αth order (→ 5 Additive Processes), this operation is called the **subordination of the αth order**. If $\{X_t\}$ is an additive process, then the process obtained from it by subordination is also an additive process. The subordination of the αth order of Brownian motion gives a †symmetric stable process of the 2αth order. Let $\{\psi_1(t)\}$ and $\{\psi_2(t)\}$ be independent of $\{X_t\}$. Then the superposition of two subordinations of $\{X_t\}$ by $\psi_1(t)$ and $\psi_2(t)$ coincides with the subordination by $\{\Psi_1(\Psi_2(t,\omega),\omega)\}$ [30].

(6) Reversed processes. Let $\{X_t\}_{t\in T}$ be a Markov process on $(\Omega,\mathfrak{B},P)$ and $X_t^*=X_{-t}$ for $t\in T^*=\{t\mid -t\in T\}$. Then $\{X_t^*\}_{t\in T^*}$ is a Markov process and is called a **reversed process** of $\{X_t\}$. If the state space of S is countable, then the transition probability $P(s,x,t,y)$ of $\{X_t\}$ and $P^*(s,x,t,y)$ of $\{X_t^*\}$ satisfy the following condition: $P^*(s,x,t,y)=Q(-t,y)P(-t,y,-s,x)Q(-s,x)^{-1}$, where $Q(t,x)=P(X_t(\omega)=x)$ and we assume $Q(t,x)\neq 0$.

For a given temporally homogeneous Markov process $\{X_t\}$ with state space S, let $\{T_t\}$ be the semigroup corresponding to $\{X_t\}$. Then a †σ-finite measure m on $(S,\mathfrak{B}(S))$ is called a **subinvariant measure** (or **excessive measure**) if the inequality $\int_S T_tf(x)m(dx)\leqslant\int_S f(x)m(dx)$ holds for every nonnegative function f. The measure m is called an **invariant measure** if the equality holds instead. Let $\{X_t^*\}$ be another Markov process whose state space is S and whose semigroup is $\{T_t^*\}$. If for some σ-finite measure m and for every nonnegative f and g the equality $\int_S T_tf(x)g(x)m(dx)=\int_S f(x)T_t^*g(x)m(dx)$ holds, then $\{X_t^*\}$ is called the **dual process** of $\{X_t\}$. m is then a subinvariant measure of $\{X_t\}$ and $\{X_t^*\}$.

The concept of reversed process is related to that of the dual process. For example, if m is an invariant probability measure of $\{X_t\}$ and if the distribution of X_0 is m, then the distribution of X_t is also m for every t, and the reversed process of $\{X_t\}$ coincides with the dual process [23].

References

[1] A. N. Kolmogorov, Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung, Math. Ann., 104 (1931), 415–458.
[2] W. Feller, Zur Theorie der stochastichen Prozesse, Math. Ann., 113 (1936), 113–160.

[3] S. Bernstein, Equations differentielles stochastiques, Actualités Sci. Ind., Hermann, 1938, 5–31.
[4] P. Lévy, Processus stochastiques et mouvement brownien, Gauthier-Villars, 1948.
[5] K. Itô, On stochastic differential equations, Mem. Amer. Math. Soc., 1951.
[6] P. Lévy, Théorie de l'addition des variables aléatoires, Gauthier-Villars, 1937.
[7] L. Doob, Stochastic processes, Wiley, 1953.
[8] W. Feller, On second order differential operators, Ann. Math., (2) 61 (1955), 90–105.
[9] E. B. Dynkin, Markov Processes I, II, Springer, 1965. (Original in Russian, 1963.)
[10] K. Itô and H. P. McKean, Jr., Diffusion processes and their sample paths, Springer, 1965.
[11] G. A. Hunt, Markov processes and potentials, Illinois J. Math., 1 (1957), 44–93, 316–369; 2 (1958), 151–213.
[12] K. Itô, Lectures on stochastic processes, Tata Inst. of Fundamental Research, Bombay, 1960.
[13] P. A. Meyer, Processus de Markov, Lecture notes in math. 26, Springer, 1967.
[14] R. M. Blumenthal and R. K. Getoor, Markov processes and potential theory, Academic Press, 1968.
[15] C. Dellacherie and P. A. Meyer, Probabilités et potentiel, two volumes, Hermann, 1975, 1980.
[16] N. Ikeda and S. Watanabe, Stochastic differential equations and diffusion processes, Kodansha and North-Holland, 1981.
[17] K. Itô, Stochastic processes, Lecture Notes Series, no. 16 Aarhus Univ., Matematisk Institute, 1969.
[18] D. W. Stroock and S. R. S. Varadhan, Multidimensional diffusion processes, Springer, 1979.
[19] M. Fukushima, Dirichlet forms and Markov processes, Kodansha and North-Holland, 1980.
[20] J. L. Doob, Semimartingales and subharmonic functions, Trans. Amer. Math. Soc., 77 (1954), 86–121.
[21] H. P. McKean and H. Tanaka, Additive functionals of the Brownian path, Mem. Coll. Sci. Univ. Kyoto, (A. Math.) 33 (1961), 479–506.
[22] D. Revuz, Mesures associées aux fonctionnelles additives de Markov I, Trans. Amer. Math. Soc., 148 (1970), 501–531.
[23] M. Motoo and S. Watanabe, On a class of additive functionals of Markov processes, J. Math. Kyoto Univ., 4 (1965), 429–469.
[24] H. Kunita and S. Watanabe, On square integrable martingales, Nagoya Math. J., 30 (1967), 209–245.
[25] S. Watanabe, On discontinuous additive functionals and Lévy measures of a Markov process, Japan. J. Math., 34 (1964), 53–70.
[26] P. A. Meyer, Integral stochastiques I–III, Lecture notes in math. 39, Springer, 1967.
[27] P. A. Meyer, Martingales local fonctionneles additives I, Lecture notes in math. 649, Springer, 1978.
[28] A. D. Wenzell, Additive functionals of the multidimensional Wiener process, Dokl. Akad. Nauk SSSR, 139 (1961), 13–16; 142 (1962), 1223–1226.
[29] M. Kac, On some connections between probability theory and differential and integral equations, Proc. 2nd Berkeley Symp. Math. Stat. Prob., Univ. of California Press (1951), 189–215.
[30] S. Bochner, Harmonic analysis and the theory of probability, Univ. of California Press, 1955.

262 (XVII.11) Martingales

A. General Remarks

Let $(\Omega, \mathfrak{B}, P)$ be a †probability space and T a time parameter set. For each $t \in T$, let $\mathfrak{F}_t$ be a †σ-algebra such that $\mathfrak{F}_s \subset \mathfrak{F}_t \subset \mathfrak{B}$ $(s<t)$. Without loss of generality we can assume that the probability space $(\Omega, \mathfrak{B}, P)$ is †complete and each $\mathfrak{F}_t$ contains all measurable subsets of Ω with P-measure zero. A real-valued †stochastic process $\{X_t\}_{t \in T}$ on $(\Omega, \mathfrak{B}, P)$ (which is also denoted by $(X_t, t \in T)$) is called a **martingale** with respect to $\mathfrak{F}_t$ provided that (i) X_t is $\mathfrak{F}_t$-measurable and $E(|X_t|)<\infty$; and (ii) if $s<t$, then

$$E(X_t \mid \mathfrak{F}_s) = X_s \quad \text{(a.s.)}, \tag{1}$$

where (a.s.) means †almost surely, which will be omitted when there is no room for confusion. In this case, we also say that $(X_t, \mathfrak{F}_t, t \in T)$ is a martingale. If the equality in (1) is replaced by the inequality $\leqslant$ $(\geqslant)$, $\{X_t\}_{t \in T}$ is called a **supermartingale (submartingale)**. For the case of martingales the values of X_t may be complex numbers. We write martingale, submartingale, and supermartingale as (M), (SbM), and (SpM), respectively, for short. If $(X_t, \mathfrak{F}_t, t \in T)$ is an (M) then $(X_t, \mathfrak{B}_t, t \in T)$ is also an (M), where $\mathfrak{B}_t = \mathfrak{B}(X_u, u \in T, u \leqslant t)$. When the family of σ-algebras involved in the definition of an (SbM) $\{X_t\}_{t \in T}$ is not explicitly mentioned, we understand that $\{X_t\}_{t \in T}$ is an (M) with respect to $\mathfrak{B}_t$. This convention is used for (SbM) and (SpM) also. The term martingale is due to J. Ville. P. Lévy had already made use of the concept in his work, but it was J. L. Doob who estab-

lished a systematic theory of martingales [2]. These concepts now provide us with not only basic tools but also fundamental principles in the theory of stochastic processes.

The terminology "super-" and "submartingales" comes from the analogy to super- and subharmonic functions. As a mathematical model of games, however, they correspond on the contrary to subfair and superfair games, respectively. One of the fundamental mathematical principles in games was formulated by Doob as the following **optional sampling theorem** for martingales. To state it, we suppose for the moment that $T=\mathbf{Z}^+=\{0,1,2,\ldots\}$; a similar result also holds if $T=\mathbf{R}^+$ and X_t is right-continuous. Let σ and τ be bounded $(\mathfrak{F}_n)$-†stopping times (σ is bounded if $m\in\mathbf{Z}^+$ exists such that $\sigma\leqslant m$ a.s.). If $(X_n, n\in\mathbf{Z}^+)$ is an (M) ((SbM), (SpM)) with respect to $(\mathfrak{F}_n)$, then

$$X_{\sigma\wedge\tau}=E(X_\sigma\,|\,\mathfrak{F}_\tau)\qquad(\leqslant,\geqslant).$$

In particular, $E(X_{\sigma\wedge\tau})=E(X_\sigma)$ $(\leqslant,\geqslant)$. Conversely, these properties characterize martingales: If X_n is $\mathfrak{F}_n$-measurable, $E(|X_n|)<\infty$ and $E(X_{\sigma\wedge\tau})=E(X_\sigma)$ $(\leqslant,\geqslant)$ for any bounded stopping times σ and τ, then $(X_n,\mathfrak{F}_n,n\in\mathbf{Z}^+)$ is an (M) ((SbM), (SpM)).

The following are some basic properties of martingales: (1) For any given (SbM) X_t, $m(t)=E(X_t)$ is an increasing function of t, and X_t is an (M) if and only if $m(t)=$ constant. (2) Let X_t and Y_t be (SbM). Then aX_t+bY_t $(a,b\geqslant 0)$ and $\sup(X_t,Y_t)$ are (SbM). (3) Let X_t be (SbM) and $f(x)$ an increasing †convex function defined in $(-\infty,\infty)$. For any given $t_0\in T$, if $E(|f(X_{t_0})|)<\infty$, then $(f(X_t), t\in(-\infty,t_0]\cap T)$ is an (SbM) and furthermore, when X_t is an (M), $(f(X_t), t\in(-\infty,t_0]\cap T)$ is an (SbM) even if $f(x)$ is not increasing. In particular, $X_t^+=\sup(X_t,0)$ is an (SbM) if X_t is an (SbM), and $|X_t|$ is an (SbM) if X_t is an (M). (4) Let X_t be an (SbM). If a, b, $t\in T$ and $a<t<b$, then $E(|X_t|)\leqslant 2E(|X_b|)-E(X_a)$. (5) If X_t is an (SbM) and $X_t\geqslant 0$ $(t\in T)$ and $t_1\in T$, then the family of †random variables $\{X_t\,|\,t\in(-\infty,t_1]\cap T\}$ is uniformly integrable (see (6)). (6) If X_t is an (SbM), $t_n\in T$ and $t_n\downarrow$, then $\{X_{t_n}\}$ is uniformly integrable if and only if $\lim_{n\to\infty}E(X_{t_n})>-\infty$. Here a family of random variables $\{X_t\}_{t\in T}$ is said to be **uniformly integrable** if we have

$$\limsup_{n\to\infty}\sup_t\int_{A_{n,t}}|X_t(\omega)|\,dP(\omega)=0,\tag{2}$$

where $A_{n,t}=\{\omega\,|\,|X_t(\omega)|>n\}$.

Example 1. For any sequence of random variables $Y_1, Y_2, \ldots$, if the relations

$$E(Y_{n+1}\,|\,Y_1,\ldots,Y_n)\geqslant 0,\qquad n=1,2,\ldots,\tag{3}$$

hold, then

$$X_n=\sum_{\nu=1}^{n}Y_\nu$$

is an (SbM). If the inequality sign in (3) is replaced by the equality sign, then X_n is an (M). In particular, if $\{Y_n\}$ is a sequence of independent random variables such that $E(Y_n)=0$, then

$$X_n=\sum_{\nu=1}^{n}Y_\nu$$

is an (M).

B. Martingale Inequalities and Convergence Theorems

The following inequalities, which are consequences of the above optional sampling theorem, are due to Doob. Let $(X_j, 1\leqslant j\leqslant n)$ be a nonnegative (SbM) and $X_n^*=\max_{1\leqslant j\leqslant n}X_j$. Then

$$\lambda P(X_n^*>\lambda)\leqslant E(X_n;X_n^*>\lambda)\leqslant E(X_n)\qquad(\lambda>0).$$

From this we have that if $(X_j, 1\leqslant j\leqslant n)$ is an (M) such that $E(|X_n|^p)<\infty$, then

$$P(|X|_n^*>\lambda)\leqslant\lambda^{-p}E(|X_n|^p)\qquad(p\geqslant 1)$$

and

$$E(|X|_n^{*p})\leqslant\left(\frac{p}{p-1}\right)^p E(|X_n|^p)\qquad(p>1),$$

where $|X|_n^*=\max_{1\leqslant j\leqslant n}|X_j|$. If $U(I)$ is the upcrossing number of an interval $I=[a,b]$ by a sample sequence of an (SbM) $(X_j, 1\leqslant j\leqslant n)$ (i.e., the number of pairs (i,j), $1\leqslant i\leqslant j\leqslant n$, such that $X_i\leqslant a$, $X_j\geqslant b$ and $a<X_k<b$ for $i<k<j$), then

$$E(U(I))\leqslant\frac{1}{b-a}(E[(X_n-a)^+]-E[(X_1-a)^+]).$$

Using these inequalities, we have the following **convergence theorems**: (i) Let $(X_n, 1\leqslant n<\infty)$ be an (SbM). (a) If $\sup_n E(X_n^+)<\infty$, then $\lim_{n\to\infty}X_n=X_\infty$ exists with probability 1 and $E(|X_\infty|)<\infty$. In particular, every nonpositive (SbM) and nonnegative (SpM) converge to integrable random variables a.s. (b) Furthermore, if $\{X_n\,|\,1\leqslant n<\infty\}$ is uniformly integrable, then $\lim_{n\to\infty}X_n=X_\infty$ exists with probability 1 by (a), and $(X_n\,|\,1\leqslant n\leqslant\infty)$ is also an (SbM). (c) If X_n is an (SbM) such that $\{E(|X_n|)\}$ is bounded, then $\lim_{n\to\infty}X_n$ exists, and if $(X_n, 1\leqslant n\leqslant\infty)$ is an (SbM), then $\lim_{n\to\infty}E(X_n)\leqslant E(X_\infty)$, where the equality holds if and only if $\{X_n\,|\,1\leqslant n<\infty\}$ is uniformly integrable. (ii) If $(X_n, -\infty<n\leqslant -1)$ is an (SbM), then $\lim_{n\to-\infty}X_n=X_{-\infty}$ exists and $-\infty\leqslant X_{-\infty}<\infty$. Furthermore, if $E(X_{-\infty})>-\infty$, then $-\infty<X_{-\infty}<\infty$ and $(X_n, -\infty\leqslant n\leqslant -1)$ is a uniformly integrable (SbM). (iii) Let $(X_1,X_2,\ldots,Z)$ be an (SbM). (a) $\lim_{n\to\infty}X_n=X_\infty$ exists and $\lim_{n\to\infty}E(X_n)\leqslant E(X_\infty)\leqslant E(Z)$. (b) $\lim_{n\to\infty}E(X_n)=E(X_\infty)$ if and only if $\{X_n\,|\,1\leqslant n<\infty\}$ is uniformly integrable, and in

this case $(X_1, X_2, \ldots, X_\infty, Z)$ is an (SbM).

(iv) Let $\{\mathfrak{F}_n\}$ $(-\infty < n < \infty$ and $\mathfrak{F}_n \subset \mathfrak{F}_{n+1} \subset \mathfrak{B})$ be a sequence of σ-algebras on $(\Omega, \mathfrak{B}, P)$. Put $\mathfrak{F}_{-\infty} = \bigcap_n \mathfrak{F}_n$ and $\mathfrak{F}_{+\infty} = \bigvee_n \mathfrak{F}_n$ (the smallest σ-algebra containing all $\mathfrak{F}_n$). If Z is a random variable with $E(|Z|) < \infty$, then $\lim_{n\to\pm\infty} E(Z \mid \mathfrak{F}_n) = E(Z \mid \mathfrak{F}_{\pm\infty})$ (a.s.). We mention some applications of these convergence theorems.

Example 2. Let $(\Omega, \mathfrak{B}, P)$ be a probability space and $\{\pi_n\}$ $(n = 1, 2, \ldots)$ a sequence of partitions of Ω into $\mathfrak{B}$-measurable sets with positive P-measure such that for each n, π_{n+1} is finer than π_n. Let π_n be $\{M_1^{(n)}, M_2^{(n)}, \ldots\}$, denote the smallest σ-algebra containing $\{M_j^{(n)}\}_{j=1,2,\ldots}$ by $\mathfrak{F}_n$ for each n, and set $\mathfrak{F}_\infty = \bigvee_n \mathfrak{F}_n$. For a given †completely additive set function φ on $(\Omega, \mathfrak{F}_\infty)$, if we define $X_n(\omega)$ by $X_n(\omega) = \varphi(M_j^{(n)})/P(M_j^{(n)})$ for $\omega \in M_j^{(n)}$, $j = 1, 2, \ldots$, then $(X_n, \mathfrak{F}_n, 1 \leqslant n < \infty)$ is an (M), and $\lim_{n\to\infty} X_n = X_\infty$ exists. If $\hat{P}$ is the restriction of P to $\mathfrak{F}_\infty$, then φ is †absolutely continuous with respect to $\hat{P}$ if and only if $\{X_n \mid 1 \leqslant n < \infty\}$ is uniformly integrable, and in this case, $X_\infty = d\varphi/d\hat{P}$ with $\hat{P}$-measure 1. If φ is †singular with respect to $\hat{P}$, then $X_\infty = 0$ with $\hat{P}$-measure 1.

Example 3. Let $X_1, X_2, \ldots$ be any sequence of random variables and Z an integrable random variable that is measurable with respect to $\mathfrak{B}(X_1, X_2, \ldots)$. Then $\lim_{n\to\infty} E(Z \mid X_1, X_2, \ldots, X_n) = Z$ (a.s.). In particular, if $X_1, X_2, \ldots$ are independent and Z is $\mathfrak{B}(X_n, X_{n+1}, \ldots)$-measurable for every n, then Z is equal to a constant a.s. This is the so-called †Kolmogorov zero-one law.

Recently, much work has been done to develop an analog of the classical theory of H^p-spaces of harmonic functions in the framework of martingale theory. Let $(\Omega, \mathfrak{F}, P)$ and $\{\mathfrak{F}_n\}_{n\in\mathbf{Z}^+}$ be fixed, and let $X = (X_n)$ be a uniformly integrable (M) with respect to $\{\mathfrak{F}_n\}$. Then $X_n = E(X_\infty \mid \mathfrak{F}_n)$, where $X_\infty = \lim_{n\to\infty} X_n$, and we can identify $X = (X_n)$ with X_∞. Set $X^* = \sup_n |X_n|$ and $[X, X] = \sum_{n=0}^\infty (X_n - X_{n-1})^2$ $(X_{-1} = 0)$. By Doob's inequality above,

$$\left(\frac{p-1}{p}\right)\|X^*\|_p \leqslant \|X_\infty\|_p \leqslant \|X^*\|_p, \qquad 1 < p < \infty,$$

where $\| \quad \|_p$ is the usual $L^p(\Omega, P)$-norm, $1 \leqslant p \leqslant \infty$. Burkholder and Gundy, and Davis obtained the following inequality: There exist positive constants c_p and C_p depending only on p such that

$$c_p \|[X, X]^{1/2}\|_p \leqslant \|X^*\|_p \leqslant C_p \|[X, X]^{1/2}\|_p, \qquad 1 \leqslant p < \infty.$$

If we set

$$H^p = \{X \mid \|X\|_{H^p} \equiv \|[X, X]^{1/2}\|_p < \infty\}, \quad p \geqslant 1,$$

then H^p is a Banach space which can be identified (by the identification of X and X_∞) with L^p for $p > 1$. $H^1 \subsetneq L^1$, however, and if we set

$$BMO = \{X \mid \|X\|_{BMO}^2 \equiv \sup_n \|E(|X_\infty - X_{n-1}|^2 \mid F_n)\|_\infty < \infty\},$$

then BMO is a Banach space which can be identified with H^{1*}, the dual space of H^1. In particular, Fefferman's inequality holds:

$$|E(X_\infty Y_\infty)| \leqslant \sqrt{2}\|X\|_{H^1}\|Y\|_{BMO}, \quad X \in H^1, \quad Y \in BMO.$$

This is an analog of the classical Fefferman's theorem, and $X \in BMO$ is an analog of a function of bounded mean oscillation, the notion of which is due to John and Nirenberg. For details → [1, 3]; for an approach using conformal martingales in continuous time → [4].

C. Sample Functions

Let $(X_t, \mathfrak{F}_t, t \in T)$ be an (SbM). Here the parameter set T may be an arbitrary subset of $(-\infty, \infty)$. However, we can always find an interval $I \supset T$ and for each $t \in I$ a σ-algebra $\tilde{\mathfrak{F}}_t$ and a random variable $\tilde{X}_t$ such that $(\tilde{X}_t, \tilde{\mathfrak{F}}_t, t \in I)$ is an (SbM) and $P(\tilde{X}_t = X_t) = 1$, $\tilde{\mathfrak{F}}_t = \mathfrak{F}_t$ for every $t \in T$. Therefore we can assume without loss of generality that the parameter set T is an interval. Furthermore, we assume that the stochastic process $\{X_t\}_{t\in T}$ is †separable. Using inequalities and convergence theorems for the sample sequences of (SbM)'s with discrete parameters, we obtain the following properties of the †sample functions of (SbM)'s with continuous parameters: (i) The sample function of an (SbM) $\{X_t\}$ is bounded on every finite interval $[a, b] \subset T$ with probability 1. (ii) Let T_0 be the interior of T. Then $P(X_{t+0}$ and X_{t-0} exist for all $t \in T_0) = 1$, and for each $t \in T_0$, $\lim_{s\uparrow t} E(X_s) \leqslant E(X_{t-0}) \leqslant E(X_t) \leqslant E(X_{t+0}) \leqslant \lim_{s\downarrow t} E(X_s)$. (iii) Let D be the set of fixed discontinuity points of $\{X_t\}$ (t is called a †fixed discontinuity point of $\{X_t\}$ if $P(X_{t-0} = X_t = X_{t+0}) \neq 1$). Then D is an at most countable set.

We assume for simplicity that the parameter set is $R^+ = [0, \infty)$ and the sample functions of $\{X_t\}$ are right continuous with probability 1. In this case, if $(X_t, \mathfrak{F}_t, t \in \mathbf{R}^+)$ is an (SbM), then $(X_t, \mathfrak{F}_{t+}, t \in \mathbf{R}^+)$ is an (SbM), where $\mathfrak{F}_{t+} = \bigcap_{s>t} \mathfrak{F}_s$. Therefore we can assume that $\mathfrak{F}_t = \mathfrak{F}_{t+}$ for all $t \in \mathbf{R}^+$. Let A be an interval and $\{\tau_\alpha\}_{\alpha\in A}$ a family of stopping times such that $\tau_\alpha \leqslant \tau_\beta < \infty$ whenever $\alpha < \beta$. Put $X_\alpha^* = X_{\tau_\alpha}$ and $\mathfrak{F}_\alpha^* = \mathfrak{F}_{\tau_\alpha}$. Then $(X_\alpha^*, \mathfrak{F}_\alpha^*, \alpha \in A)$ is called the stochastic process obtained by an **optional sampling** (or a **time change**) from $(X_t, \mathfrak{F}_t, t \in T)$. By the optional sampling theorem of Section A, we can conclude that $(X_\alpha^*, \mathfrak{F}_\alpha^*, \alpha \in A)$ is also an (SbM) if at least one of the following con-

ditions is satisfied: (1) $X_t \leq 0$ (a.s.) for all $t \in \mathbf{R}^+$; (2) $\{X_t, t \in \mathbf{R}^+\}$ is uniformly integrable; (3) for each $\alpha \in A$, τ_α is bounded with probability 1.

D. Decompositions of Submartingales [1, 6]

If an (SbM) $(X_t, \mathfrak{F}_t, t \in \mathbf{R}^+)$ is uniformly integrable, $\lim_{t \to \infty} X_t = X_\infty$ exists and $(X_t, \mathfrak{F}_t, t \in [0, \infty])$ is an (SbM) for which $\mathfrak{F}_\infty = \bigvee_t \mathfrak{F}_t$. In addition, when the family $\{X_\tau \mid \tau \in \mathfrak{T}\}$ is uniformly integrable, it is said to belong to class (D). Here $\mathfrak{T}$ denotes the collection of all stopping times with respect to $(\mathfrak{F}_t)$. If for each a $(0 < a < \infty)$ the family $\{X_\tau \mid \tau \in \mathfrak{T}$ and $\tau \leq a\}$ is uniformly integrable, the family is said to belong locally to class (D).

If an (SbM) X_t is uniformly integrable, X_t is decomposed as

$$X_t = E(X_\infty \mid \mathfrak{F}_t) - (E(X_\infty \mid \mathfrak{F}_t) - X_t), \tag{4}$$

and if we take appropriate †versions of conditional expectations, $E(X_\infty \mid \mathfrak{F}_t)$ becomes a right continuous (M). In the decomposition (4), $M_t = E(X_\infty \mid \mathfrak{F}_t)$ is an (M) and $Z_t = E(X_\infty \mid \mathfrak{F}_t) - X_t$ is a potential, i.e., a nonnegative right continuous (SpM) with $\lim_{t \to \infty} Z_t = 0$ (a.s.). The decomposition (4) is called the **Riesz decomposition** of the (SbM) X_t; the names "potential" and "Riesz decomposition" come from †potential theory in view of the obvious similarity.

A stochastic process $(A_t, t \in \mathbf{R}^+)$ on $(\Omega, \mathfrak{B}, P)$ is called a (right continuous) **increasing process** provided that (i) A_t is $\mathfrak{F}_t$-measurable, $E(A_t) < \infty$ for each t, and (ii) with probability 1, the sample function is a right continuous and increasing function with $A_0 = 0$. If $E(A_\infty) < \infty$, where $A_\infty = lim_{t \to \infty} A_t$, the stochastic process is said to be **integrable**. We have the following **Doob-Meyer decomposition theorem**: (i) A potential X_t is decomposed as

$$X_t = E(A_\infty \mid \mathfrak{F}_t) - A_t \tag{5}$$

by a suitably chosen integrable increasing process if and only if X_t belongs to class (D). (ii) An (SbM) X_t is decomposed into $X_t = X_t' + A_t$, where X_t' is an (M) and A_t is an increasing process if and only if X_t belongs locally to class (D). (iii) In (i) and (ii), A_t can be chosen to be †predictable; under this condition, these decompositions are unique. Furthermore, A_t can be chosen to be continuous if and only if X_t is **regular** in the sense that for any sequence $\tau_n \in \mathfrak{T}$ such that $\tau_n \uparrow \tau$, $\lim_{n \to \infty} E(X_{\tau_n \wedge a}) = E(X_{\tau \wedge a})$ for every $a > 0$.

E. Semimartingales [1, 6, 7]

†Lévy processes and †Itô processes are naturally generalized to a class of stochastic processes called semimartingales. This class of processes is important since it appears to be the most general class for which a systematic theory of stochastic calculus can be developed (→ 406 Stochastic Differential Equations). Furthermore, we can define some quantities called local characteristics of semimartingales which are, in a sense, a generalization of Lévy-Khinchin characteristics for Lévy processes. In many interesting cases, the law of a semimartingale can be determined if we specify its local characteristics to be given functionals of the sample paths. This way of characterizing stochastic processes is known as a martingale problem, a concept introduced by Stroock and Varadhan.

A stochastic process X_t, $t \in \mathbf{R}^+$, on $(\Omega, \mathfrak{B}, P)$ and $(\mathfrak{F}_t)$ is called a **semimartingale** if it can be represented as

$$X_t = X_0 + M_t + V_t,$$

where X_0 is $\mathfrak{F}_0$-measurable, M_t $(M_0 = 0)$ is a right continuous **local martingale** with respect to $(\mathfrak{F}_t)$, i.e., there exists $\sigma_n \in \mathfrak{T}$ such that $\sigma_n \uparrow \infty$ and $M_t^{\sigma_n} = M_{t \wedge \sigma_n}$ is a uniformly integrable (M) with respect to $(\mathfrak{F}_t)$ for each n, and V_t $(V_0 = 0)$ is a right continuous †$(\mathfrak{F}_t)$-adapted process such that $t \in [0, T] \mapsto V_t$ is of bounded variation a.s. for every $T > 0$. The class of semimartingales is known to be invariant under C^2-transformations (Itô's formula), time changes, and absolutely continuous changes of the basic probability P (Girsanov's theorem).

Example 4. Let $X_t = (X_t^1, X_t^2, \ldots, X_t^n)$ be a system of semimartingales such that all X_t^i, $X_t^i X_t^j - \delta^{ij} t$, $i, j = 1, 2, \ldots, n$, are continuous (local) martingales with respect to $(\mathfrak{F}_t)$. Then X_t is an n-dimensional Wiener process such that $\mathfrak{B}(X_u - X_v, u, v \geq t)$ and the $\mathfrak{F}_t$ are independent for every t. For this reason such an X_t is often called a Wiener martingale.

Example 5. Let X_t be a semimartingale whose sample paths are increasing step functions with jumps of size 1 a.s. If $X_t - ct$ $(c > 0)$ is an (M) with respect to $(\mathfrak{F}_t)$ then X_t is a †Poisson process with the same independence property as Example 4. For a similar characterization of Lévy processes in the martingale framework → [5].

The notion of semimartingales can also be defined for processes taking values in a differentiable manifold [8].

References

[1] C. Dellacherie and P. A. Meyer, Théorie des martingales, Hermann, 1980, chs. 5, 8.
[2] J. L. Doob, Stochastic processes, Wiley, 1953.
[3] A. M. Garsia, Martingale inequalities, Benjamin, 1973.

[4] R. K. Getoor and M. J. Sharpe, Conformal martingales, Inventiones Math., 16 (1972), 271–308.
[5] B. Grigelionis, On the martingale characterization of stochastic processes with independent increments, Lit. Math. J., 17 (1977), 75–86.
[6] N. Ikeda and S. Watanabe, Stochastic differential equations and diffusion processes, Kodansha and North-Holland, 1980.
[7] J. Jacod, Calcul stochastique et problèmes de martingales, Lecture notes in math. 714, Springer, 1979.
[8] L. Schwartz, Semi-martingales sur variétés, et martingales conformes sur des variétés analytiques complexes, Lecture notes in math. 780, Springer, 1980.

263 (XVI.8) Mathematical Models in Biology

A. History

In certain quantitative studies in biology, particularly in epidemiology, population ecology, and developmental biology, there arise mathematically describable models. Such mathematical descriptions of biological phenomena have sometimes been proposed by mathematicians and sometimes by biologists. We call these descriptions **mathematical models in biology**.

The first mathematical model was proposed by R. Ross [1] in 1911. He undertook a theoretical investigation of the propagation of malaria. His object was to give an analysis of the propagation of malaria in a certain locality under somewhat simplified conditions. His assumption was that both emigration and immigration were negligible and that there was no increase of population. In such a locality, the propagation of malaria is considered to be determined in general by two factors which evolve continuously and simultaneously. On the one hand, the number of new infections depends upon the number and infectivity of the mosquitoes; on the other hand, at the same time the infectivity of mosquitoes is determined by the number of people in the given locality and the frequency of infection among them. Ross expressed the uninterrupted and simultaneous dependence of the first component on the second and that of the second on the first by means of a system of first-order differential equations.

As a much simpler case, following P. F. Verhulst [2], R. Pearl and L. J. Reed [3] proposed in 1920 a population model describing how a population of animals living in a fixed region initially increases, but after some period approaches a saturation point. Their model is the simple differential equation

$$\frac{du}{dt}=(A-ku)u, \tag{1}$$

where u is the population at time t and A and k are positive constants. The solution of this equation with positive initial data u_0 is expressed by a monotone-increasing curve that tends to the saturation value A/k as t tends to infinity. We call this equation a **logistic equation**. The solution fits the growth pattern of some real populations of insects; but we also have another model, which was exploited by experimental ecologists who were engaged in the study of a kind of bean weevil. That model is expressed by the difference equation [4]

$$u_{n+1}=f(u_n), \tag{2}$$

where u_n is a population of the insects in the nth generation and $f(u)$ is called the **reproduction function**; the latter is often a simple function, but is not necessarily monotonic.

B. Population Model with Two Species

V. Volterra [5] proposed the following model for the **prey-predator relation** between two species. We denote the prey population by u, and the predator population by v. These populations satisfy the system of equations

$$\frac{du}{dt}=(A-kv)u,$$

$$\frac{dv}{dt}=-(B-hu)v, \tag{3}$$

where A, B, k, and h are positive constants. The orbit of the solution of this system passing through (u_0, v_0) in the first quadrant is closed, enclosing the point $(B/h, A/K)$ and staying in the first quadrant. An integral for this system is given by

$$u^{-B}v^{-A}e^{hu+kv}=C, \tag{4}$$

where C is a constant of integration. Consequently, the solution starting at a point in the first quadrant is periodic, with the period depending on the initial data. The average populations over one period,

$$\bar{u}=\frac{1}{T}\int_t^{t+T}u(\xi)\,d\xi,\qquad \bar{v}=\frac{1}{T}\int_t^{t+T}v(\xi)\,d\xi, \tag{5}$$

do not depend on the initial data.

Many mathematical models in biology are expressed in terms of ordinary differential equations [6, 7].

C. Fundamental Equations

Some models are used to describe the spatial distributions, that is, the spatial patterns of populations. For example, we can consider equations for the pattern of a population with migration.

We denote by $p(t,x)$ the population density at time t and position x in some region V in $\mathbf{R}^2$. $v(t,x)$ is the velocity of the migration. $f(t,x)$ is a source (or supply) term for the population. Then we get

$$\frac{d}{dt}\int_V p\,dx = -\int_{\partial V} pv\mathbf{n}\,ds + \int_V f\,dx, \qquad (6)$$

where $\mathbf{n}$ is the outer unit normal to the boundary ∂V. Consequently, we get a partial differential equation

$$\frac{\partial p}{\partial t} + \nabla\cdot(pv) = f.$$

Here we assume, for example, that v is a function of x, p; then we get a †hyperbolic equation

$$\frac{\partial p}{\partial t} + \nabla\cdot(v(x,p)p) = f.$$

Also, if we assume $v = -d(x,p)\nabla p/p$ and $f = f(p)$, then we get

$$\frac{\partial p}{\partial t} = \nabla(d(x,p)\nabla p) + f(p) \quad \text{on} \quad V, \qquad (7)$$

which is an equation of †parabolic type. We impose some boundary condition at the boundary ∂V; for example, the homogeneous †Neumann condition

$$\frac{\partial p}{\partial \mathbf{n}} = 0 \quad \text{on} \quad \partial V. \qquad (8)$$

In this case, if V is a convex region and $d(x,p)$ = constant, then H. Matano [8] showed that only a constant solution of the stationary problem

$$d\Delta p + f(p) = 0 \qquad (9)$$

and (8) can be stable as a limit of the solution of the nonstationary problem (7) and (8) as t tends to ∞. (He assumed that $f(p)$ is a smooth function of p.)

D. A Diffusive Prey and Predator Population Model

R. May [4] and M. Mimura [10] independently proposed mathematical models to explain some aspects of the population patterns of plankton in water. Their work can be considered to be the completion of earlier attempts by the ecologist J. H. Steele [9]; Mimura and Nishida proved that Steele's original model does not yield a solution describing the observed population patterns, and Mimura and May introduced corrections arising from their ecological viewpoint. The model is the system

$$\frac{\partial p}{\partial t} = d_1\frac{\partial^2 p}{\partial x^2} + kp - \mu p^2 - \frac{apq}{1+bp},$$

$$\frac{\partial q}{\partial t} = d_2\frac{\partial^2 q}{\partial x^2} - \gamma q^2 + \frac{apq}{1+bp},$$

$$\text{on} \quad [0,1]\times[0,\infty), \qquad (10)$$

where $p(t,x)$ and $q(t,x)$ are the prey and predator populations, respectively, and d_1 and d_2 are diffusion coefficients. k, μ, γ, a, and b are positive constants. If b, d_1, d_2, and μ are zero, we get Volterra's prey and predator equation. Mimura considered the Cauchy problem with boundary conditions

$$\frac{\partial p}{\partial x}(t,0) = \frac{\partial p}{\partial t}(t,1) = 0,$$

$$\frac{\partial q}{\partial x}(t,0) = \frac{\partial q}{\partial t}(t,1) = 0. \qquad (11)$$

In this case, under some assumptions on d_1, d_2 and on k, μ, γ, a, Mimura [11] succeeded in proving some stationary patterns of p and q which are stable but not constant for the variable x. He used bifurcation theory to prove existence of a stationary solution near the equilibrium point. For the case in which d_1 is sufficiently small, he used the singular perturbation theory introduced by Fife [12].

The same procedure can be applied to the following system of partial differential equations, which was proposed by Gierer and Meinhardt in developmental biology [14]:

$$\frac{\partial a}{\partial t} = D_a\frac{\partial^2 a}{\partial x^2} + \rho_0\rho + \frac{c\rho a^2}{h} - \mu a,$$

$$\frac{\partial h}{\partial t} = D_h\frac{\partial^2 h}{\partial x^2} + c'\rho' a^2 - vh, \qquad (12)$$

where $a(t,x)$ is the concentration of a short-range activator and $h(t,x)$ is that of a long-range inhibitor. D_a, D_h, ρ_0, ρ, c, c', μ, and v are all positive constants. Gierer and Meinhardt showed numerically that a system such as (12) exhibits some interesting spatial patterns. Under the boundary conditions (11), Mimura obtained a rigorous proof of the numerical results [13].

These models may not be realistic from the biological viewpoint, but they do afford some insight, so that biologists may now begin to devise more realistic ones.

For example, the model given by (10) and (11) is one by means of which we can explain **pattern formation** in spatially homogenous environments. (We call the existence of such

patterns "patchiness.") The model shows that without any local environmental inhomogeneity, we can still find some patchiness in the population pattern of the plankton.

There are many models that explain biological phenomena only metaphorically. An example is the use of catastrophe theory by R. Thom [15] to model morphogenesis.

E. Population Genetics

Consider the development of a genetic structure in a population consisting of N individuals. We assume two kinds, a and A, of genes; and an individual has one of the genotypes aa, AA, aA. We assume that every generation is of a constant size N, and we consider the number of genes, not genotypes. The number X_n of a genes in the nth generation is the result of a stochastic process. In the simpest model it is a †Markov chain with the †transition probability p_{ij} from $X_n=i$ to $X_{n+1}=j$ being expressed by a †binomial distribution:

$$p_{ij}=\binom{2N}{j}p_i^j(1-p_i)^{2N-j}, \quad p_i=\frac{i}{2N}.$$

Models that take mutation, natural selection, and migration into account can be given by appropriate changes of p_i. Since the works of R. A. Fisher and S. Wright, natural selection and migration models have been the object of much research. A powerful method in the analysis of these Markov chains is the diffusion approximation [16–18]. Consider the time t and the ratio $x=i/(2N)$ of a genes as continuous variables. Then the partial differential equation

$$\frac{\partial p}{\partial t}=\frac{1}{4N}\frac{\partial^2}{\partial y^2}\{y(1-y)p\}-\frac{\partial}{\partial y}\{a(y)p\} \tag{13}$$

is obtained for the transition probability density $p(t,x,y)$ of the ratio of a genes. Here, $a(y)$ is a polynomial with degree $\leqslant 2$. M. Kimura found a solution of (13) in terms of special functions and solved many problems in absorption probability, limit distribution, speed of convergence, etc. Equation (13) is †Kolmogorov's forward equation in the theory of †diffusion processes, but the coefficients are degenerate at the boundaries (0 and 1). S. Karlin and J. McGregor [19] found that the foregoing discrete models can be taken as †branching processes under the condition that the number of individuals in each generation is $2N$. The relation between discrete and continuous models has been studied in general dimensions as a problem in the convergence of stochastic processes [20, 21].

For related topics → 40 Biometrics.

References

[1] R. Ross, The prevention of malaria, London, 1911.
[2] P. F. Verhulst, Notice sur la loi que la population suit dans son accroisement, Corr. Math. Phys. Publ. A, 10 (1838), 113.
[3] R. Pearl and L. J. Reed, On the rate of growth of the population of the United States since 1790 and its mathematical representation, Proc. Nat. Acad. Sci. US, 6 (1920), 275.
[4] R. M. May, Theoretical ecology: Principles and applications, Blackwell, 1976.
[5] V. Volterra, Leçon sur la théorie mathématique de la lutte pour la vie, Gauthier-Villars, 1931.
[6] E. C. Pielou, An introduction to mathematical ecology, Wiley, 1969.
[7] J. M. Smith, Models in ecology, Cambridge Univ. Press, 1974.
[8] H. Matano, Asymptotic behavior and stability of solutions of semi-linear diffusion equations, Publ. Res. Inst. Math. Sci., 15 (2) (1979), 401–454.
[9] J. H. Steele, Spatial heterogeneity and population stability, Nature, 248 (1974), 83.
[10] M. Mimura, Y. Nishiura, and M. Yamaguti, Some diffusive prey and predator systems and their bifurcation problems, Bifurcation Theory and Applications in Scientific Disciplines, O. Gurel and O. E. Rössler (eds.), Ann. New York Acad. Sci., 316 (1979), 490–510.
[11] M. Mimura, Asymptotic behaviors of a parabolic system related to a planktonic prey and predator model, SIAM J. Appl. Math., 37, no. 3 (1979), 499–512.
[12] P. C. Fife, Boundary and interior transition layer phenomena for pairs of second-order differential equations, J. Math. Anal. Appl., 54 (1976), 497–521.
[13] M. Mimura and M. Yamaguti, Pattern formation in interacting and diffusing system in population biology, Advances in Biophysics, M. Kotani (ed.), vol. 15, Tokyo Univ. Press, 1982.
[14] G. Meinhardt, A theory of biological pattern formation, Kybernetik, 12 (1972), 30–39.
[15] R. Thom, La stabilite structurellé et la morphogénèse, Benjamim, 1972.
[16] J. F. Crow and M. Kimura, An introduction to population genetics theory, Harper & Row, 1970.
[17] W. Feller, Diffusion processes in genetics, Proc. 2nd Berkeley Symp. Math. Statist. Probab., Univ. of California Press, 1951, 227–246.
[18] W. J. Ewens, Mathematical population genetics, Springer, 1979.

[19] S. Karlin and J. McGregor, Direct product branching processes and related induced Markov chains I, Springer, 1965, 111–145.
[20] K. Sato, Diffusion processes and a class of Markov chains related to population genetics, Osaka J. Math., 13 (1976), 631–659.
[21] S. N. Ethier, A class of degenerate diffusion processes occurring in population genetics, Comm. Pure Appl. Math., 29 (1976), 483–493.

264 (XIX.1) Mathematical Programming

A. Introduction

Mathematical programming is a method useful in operations research, industrial engineering, systems engineering, etc., and it is concerned with optimization in general. If we want to minimize cost or loss or to maximize some effect or profit under various circumstances, the corresponding mathematical problem is usually formulated in the form of a mathematical-programming problem. Computational methods in mathematical programming have seen great development hand in hand with remarkable progress in computer technology, and we are now able to deal with large-scale problems practically.

B. General Definitions

A **mathematical-programming problem** in the broadest sense of the word is one of finding a maximum or minimum of a given function $f: X \rightarrow R$ (where X is a set and R is an ordered set). However, in a narrower sense, it usually means a problem where X is a closed subset of the n-demensional Euclidean space $\mathbf{R}^n$ (or, more generally, a Banach space) and f is a real-valued continuous function. Often, X is defined as the set of points in $\mathbf{R}^n$ that satisfy a system of equalities and inequalities of the form $g_i(\mathbf{x}) \leqslant 0$, $\geqslant 0$, or $=0$ for some given real-valued functions g_i $(i=1, \ldots, m)$ defined on $\mathbf{R}^n$. In mathematical programming, special terminology is sometimes used; e.g., X is conventionally called the **feasible region**, a point of X a **feasible solution**, the solution of the problem itself the **optimal solution**, f the **objective function**, and the equalities and inequalities defined in terms of the g_i are the **constraints**.

C. Types of Mathematical-Programming Problems

Mathematical-programming problems can be classified from various viewpoints, from which the problems get their various names. (i) **Linear programming** (→ 255 Linear Programming): The objective function f and all the functions g_i of constraints are linear. **Nonlinear programming**: Some of f and the g_i are nonlinear. **Convex programming**: f and all the g_i are convex with inequalities of the form $g_i(\mathbf{x}) \leqslant 0$ (and, consequently, X is a convex set); the problem is to minimize f (→ 88 Convex Analysis, 255 Linear Programming, 292 Nonlinear Programming). (ii) **Disjunctive programming**: X is not connected. **Integer programming**: X is a subset of the lattice points of integer coordinates in $\mathbf{R}^n$ (→ 215 Integer Programming). (iii) **Parametric programming**: f and/or the g_i contain parameters, and the problem is to analyze the behavior of the optimal solution and/or the feasible region when the parameters vary. (iv) **Stochastic programming**: The parameters in (iii) are random variables. (v) **Multiobjective programming**: A vector-valued function $f: \mathbf{R}^n \rightarrow \mathbf{R}^k$ $(k \geqslant 2)$ is taken as the objective function, where a certain order relation is defined in $\mathbf{R}^k$ (such as the Cartesian product of the order in $\mathbf{R}$ or the lexicographic order based on the order in $\mathbf{R}$). (vi) when the constraints and/or the objective function are endowed with special mathematical structures, special names are accordingly given. For example, †network flow problems, whose f and g_i are defined with reference to a graph (→ 186 Graph Theory), are called **network programming**; if f and the g_i have an iterative or repetitive structure, the name **multistage programming** is used; **dynamic programming** (→ 127 Dynamic Programming) can be regarded as a kind of multistage programming.

D. Mathematical-Programming Problems of Special Type

In this encyclopedia, independent articles appear for those types of mathematical-programming problems that have been mathematically well investigated and are most frequently used in practice (→ 127 Dynamic Programming, 215 Integer Programming, 255 Linear Programming, 292 Nonlinear Programming, 349 Quadratic Programming). The following are some other typical problems that have been systematically studied.

(i) **Fractional programming**: The objective function to be minimized within the set X

takes the form $f(\mathbf{x})=C(\mathbf{x})/D(\mathbf{x})$ (C, D: $\mathbf{R}^n \rightarrow R$), where $D(\mathbf{x})$ is assumed to be positive in X. If X is convex, C is convex and of positive value, and D is concave, the function of a real parameter λ defined as $h(\lambda) \equiv \min_{\mathbf{x}\in X}[C(\mathbf{x})-\lambda D(\mathbf{x})]$ is monotone decreasing and convex in λ. The problem of determining $h(\lambda)$ for each value of λ is a convex-programming problem, and the solution of the convex-programming problem for $\lambda=\lambda_0$ such that $h(\lambda_0)=0$ is the optimal solution of the original problem. In particular, if C and D are linear, i.e., if $C(\mathbf{x})=\mathbf{c}'\cdot\mathbf{x}+c_0$ and $D(\mathbf{x})=\mathbf{d}'\cdot\mathbf{x}+d_0$, and, moreover, if $X=\{\mathbf{x}\in\mathbf{R}^n \mid A\mathbf{x}\leqslant\mathbf{b}, \mathbf{x}\geqslant\mathbf{0}\}$, where $A\in\mathbf{R}^{m\times n}$, $\mathbf{b}\in\mathbf{R}^m$ and the inequalities are to be read componentwise (**linear fractional programming**), then by introducing new variables $\mathbf{y}=\mathbf{x}/D(\mathbf{x})$ and $y_0=1/D(\mathbf{x})$, we can reduce the original problem to the linear-programming problem $\min(\mathbf{c}'\mathbf{y}+c_0y_0 \mid A\mathbf{y}-\mathbf{b}y_0\leqslant\mathbf{0}, \mathbf{d}'\cdot\mathbf{y}+d_0y_0=1, \mathbf{y}\geqslant\mathbf{0}, y_0\geqslant 0\}$.

(ii) **Geometric programming**: The objective function $f(\mathbf{x})=g_0(\mathbf{x})$ and the feasible region $X=\{\mathbf{x}\in\mathbf{R}^n \mid g_i(\mathbf{x})\leqslant 0, i=1,\ldots,m; \mathbf{x}\geqslant\mathbf{0}\}$ are defined in terms of functions $g_i(\mathbf{x})$ of the so-called polynomial type: $g_i(\mathbf{x})=\sum_k c_{ik_j}x_j^{a_{ik_j}}$ ($i=0,1,\ldots,m$). The dual (→ 292 Nonlinear Programming) of a geometric-programming problem is easier to treat than the original, because it is a problem of finding the minimum of a convex function under constraints of linear equalities and inequalities.

(iii) **Nonconvex quadratic programming** and **bilinear programming**: The former is the problem of minimizing an objective function of the form $f(\mathbf{x})=\mathbf{c}'\cdot\mathbf{x}+\frac{1}{2}\mathbf{x}'Q\mathbf{x}$ (where Q is a symmetric matrix not necessarily nonnegative definite) under linear constraints (equalities and inequalities), and the latter is to minimize $f(\mathbf{x}_1,\mathbf{x}_2)=\mathbf{c}_1'\cdot\mathbf{x}_1+\mathbf{c}_2'\cdot\mathbf{x}_2+\mathbf{x}_1'Q\mathbf{x}_2$ over the region $X=\{(\mathbf{x}_1,\mathbf{x}_2)\in\mathbf{R}^{n_1}\times\mathbf{R}^{n_2} \mid A_1\mathbf{x}_1\leqslant\mathbf{b}_1, A_2\mathbf{x}_2\leqslant\mathbf{b}_2, \mathbf{x}_1\geqslant\mathbf{0}, \mathbf{x}_2\geqslant\mathbf{0}\}$, where $Q\in\mathbf{R}^{n_1\times n_2}$, $A_i\in\mathbf{R}^{m_i\times n_i}$, and $\mathbf{b}_i\in\mathbf{R}^{m_i}$ ($i=1,2$). These two types of problems are related to each other, and a number of computational techniques, such as specially elaborated versions of the cutting-plane method (→ 215 Integer Programming B), have been developed for them.

References

[1] G. B. Dantzing, Linear programming and extensions, Princeton Univ. Press, 1963.
[2] D. G. Luenberger, Optimization by vector space methods, Wiley, 1969.
[3] J. F. Shapiro, Mathematical programming —Structures and algorithms, Wiley, 1979.
[4] W. Dinkelbach, On nonlinear fractional programming, Management Sci., 13 (1967), 492–498.
[5] S. Schaible, Duality in fractional programming—A unified approach, Operations Res., 24 (1976), 452–461.
[6] S. Schaible, Fractional programming I—Duality, Management Sci., 22 (1976), 858–867.
[7] R. J. Duffin, E. L. Peterson, and C. Zener, Geometric programming—Theory and applications, Wiley, 1967.
[8] H. Konno, A cutting plane algorithm for solving bilinear programs, Math. Program., 11 (1976), 14–27.
[9] H. Konno, Maximization of a convex quadratic function under linear constraints, Math. Program., 11 (1976), 117–128.

265 (XXI.9) Mathematics in the 17th Century

The 17th century abounds in remarkable events in the history of science: the work on mechanics by Galileo (1564–1642); the discovery of analytic geometry by R. †Descartes (1596–1650); the early research in the theory of probability by P. de †Fermat (1601–1665) and B. †Pascal (1623–1662); the discovery of †mathematical induction by Pascal; and the discovery of the **infinitesimal calculus** (i.e., †differential calculus and †integral calculus) by I. †Newton (1642–1727) and G. W. †Leibniz (1646–1716). Compared with these events, the results of mathematical research from the Middle Ages to the 16th century seem minute. These results nonetheless exist, and historians of mathematics are now reevaluating them, particularly those of the 15th and 16th centuries.

Before Galileo, Tycho Brahe (1546–1601) kept precise records of astronomical observations. J. Kepler (1571–1630), motivated by a mystic faith in the "harmony of the universe," studied Brahe's records and discovered three laws on the motion of planets. He also treated a question of cubature in his paper on the form and volume of the wine barrel (1615). His contemporaries J. Napier (1550–1617) and J. Bürgi (1552–1632) discovered logarithms, which helped astronomers tremendously in their calculations. Napier used the concept of velocity in his introduction of logarithms; thus analysis began to germinate. Galileo founded the modern approach to the concepts of velocity and acceleration in his *Dialogue on two new sciences* (1638). Using a telescope he built, he discovered four of Jupiter's moons and observed sunspots. His espousal of the heliocentric theory of Copernicus (1473–1543) led to his denunciation before the Inquisition, which

ordered him to refrain from holding or defending the theory. This is the most famous episode of his life, but his most significant contribution to science lies in his foundations of theoretical mechanics, which he freed from the Aristotelian tradition, thereby opening the way for Newton. F. Cavalieri (1598–1647), a disciple of Galileo, applied the notion of *indivisibilis* (originating in scholastic philosophy) to questions of quadrature in his *Geometry of the indivisible* (1635). This idea influenced Pascal and J. Wallis (1616–1703).

Descartes established the method of analytic geometry in his *Geometry* (1637), published as an appendix to his *Discourse on method.* The use of †coordinates can be traced back to Apollonius of Perga; Fermat used them occasionally, but Descartes made the first clear formulation of the method of representing general figures by means of equations, an essential step beyond Greek geometry. He also surpassed F. †Viète (1540–1603) by abolishing the restriction that quantities represented by letters should be of one dimension.

Contemporary to Descartes, Fermat made remarkable contributions to †number theory and Pascal to †projective geometry, and through their correspondence there began the theory of probability; both also made precursory contributions to analysis. Fermat treated questions on maxima and minima of functions and tangents of curves; Pascal answered some questions on tangents, centers of gravity, quadrature, and cubature concerning †cycloids. Pascal also contributed to hydrostatics, made positive use of the idea of the point of infinity in projective geometry, and clearly formulated the principle of mathematical induction in his theory of arithmetic triangles, the so-called †Pascal's triangles. (Freudenthal [4] established that the first discovery of the principle of mathematical induction is due to Pascal; the exact date of the discovery was studied by Hara [5].)

In England, Wallis and I. Barrow (1630–1677) preceded Newton. Wallis solved questions concerning quadrature and cubature (by bold use of the methods established by Cavalieri), infinite series, and interpolation. Barrow was Newton's teacher. He came close to the fundamental theorem of calculus, and Newton certainly owed some ideas which led to his discovery to Barrow's suggestions. Newton completed his method of fluxions, corresponding to our differential calculus, toward 1669–1671, but his paper on this method was published only after his death (1736). In his main work, *Principia mathematica philosophiae naturalis* (1687), he used this method and its converse, without naming them, to solve the †two-body problem. The work begins with three laws of mechanics and covers the motion of the moon and hydromechanics. Leibniz discovered infinitesimal calculus slightly later than Newton but independently. He invented convenient new symbols that gave great impetus to the development of calculus: the symbols dx and $\int$ are due to him. Leibniz was in Paris in 1672–1677, where he made the acquaintance of Father Arnauld (1612–1694) of Port Royal (the monastery to which Pascal belonged) and the Dutch physicist C. Huygens (1629–1695). Through their suggestions, he studied the work of Descartes and Pascal. Leibniz's first papers on calculus were published in 1684 in the scientific journal *Acta Eruditorum*, which he also edited. The methods of calculus he initiated were transmitted to mathematicians of the †Bernoulli family and then to L. †Euler, who developed them into the wide field of analysis.

Thus the mathematics of the 17th century went clearly beyond Greek mathematics. The importance of numbers over diagrams was recognized, and mathematicians were no longer hesitant to use infinity. Moreover, people became aware of the importance of experimental methods in science. The position of mathematics as an important method of natural science was established; mathematics became a rational basis of scientific research.

It was also in this century that a peculiar kind of mathematics was developed in Japan by T. Seki (1642?–1708). However, it lacked the Greek tradition of viewing logical foundations as being important, and furthermore had no connection with natural sciences; consequently, it did not see subsequent development comparable to that of Western mathematics (→ 230 Japanese Mathematics (Wasan)).

References

[1] M. Cantor, Vorlesungen über Geschichte der Mathematik, Teubner, II, 1892; III, 1898.
[2] P. L. Boutroux, L'idéal scientifique des mathématiciens dans l'antiquité et les temps moderns, Presses Universitaires de France, new edition, 1955.
[3] D. T. Whiteside, Patterns of mathematical thought in the later seventeenth century, Arch. History Exact Sci., 1 (1961), 179–388.
[4] H. Freudenthal, Zur Geschichte der vollständigen Induktion, Arch. Internat. Histoire Sci., 6, no. 22 (1953), 17–37.
[5] K. Hara, Pascal et l'induction mathématique, Rev. Histoire Sci. Appl., 15 (1962), 287–302.
[6] N. Bourbaki, Eléments d'histoire des mathématiques, Hermann, 1960.

266 (XXI.10) Mathematics in the 18th Century

During the Age of Enlightenment, mathematical analysis developed steadily after its initiation in the preceding century. It found numerous applications in theoretical physics and contributed to the growth of rationalistic thought.

The central figures in mathematics during the late 17th and early 18th centuries were I. †Newton (1642–1727) and G. W. †Leibniz (1646–1716). C. Maclaurin (1698–1746) of Scotland followed Newton, but no mathematician of Newton's stature appeared in Great Britain. An unfortunate dispute over the priority of discovery of infinitesimal calculus arose between Newton and Leibniz, after which their followers came into conflict. This prevented the members of the English school from giving up their inconvenient notation system, which hindered their progress in calculus.

On the Continent, Leibniz was succeeded by the mathematicians of the †Bernoulli family and by L. †Euler (1707–1783), who brought about brilliant developments in calculus and its applications. They solved various kinds of †differential equations and invented the †calculus of variations. F. †Viète used the term *analysis* in the sense of algebra as a heuristic method; the same term meant algebraic treatment of infinite series in Newton's usage. It was during this century that analysis secured a position as a branch of mathematics independent of algebra and geometry. †Analytical dynamics, initiated by Euler, was further developed by J. L. †Lagrange (1736–1813) and P. S. de †Laplace (1749–1827). Laplace, in systematizing †celestial mechanics and the †theory of probability, showed what a powerful instrument analysis was. A. M. Legendre (1752–1833) investigated †elliptic integrals and opened the way for C. F. †Gauss and other mathematicians of the next century.

The growth of the Ecole Polytechnique, established during the time of the French Revolution, contributed to the brilliant progress of French mathematics. Lagrange, Laplace, and Legendre were all active in Paris during this period, as were S. D. Poisson (1781–1840) and J. B. J. †Fourier (1768–1830), both of whom made major contributions to analysis, and G. Monge (1746–1818), L. Carnot (1753–1823), and J. V. Poncelet (1788–1867). A problem proposed by Fourier in his theory of heat propagation gave rise to an important question of analysis, one that later formed the basis of †harmonic analysis. Fourier and Poisson aimed at clarifying the laws of nature, while Monge, Carnot, and Poncelet developed †projective geometry and †descriptive geometry for their purely geometric interest. Monge also did precursory work on †differential geometry.

The mathematics of this century left many remarkable results in geometry and analysis and their applications; however, it inherited its methods from the preceding century and lacked critical spirit. Mathematicians were more interested in obtaining new results than in reflecting upon the rigor of their methods. Reexamination and reestablishment of the foundations of mathematics were left to the next century.

References

[1] M. Cantor, Vorlesungen über Geschichte der Mathematik III, Teubner, 1898.
[2] D. J. Struik, A concise history of mathematics, Dover, 1948, third edition, 1967.
[3] N. Bourbaki, Eléments d'histoire des mathématiques, Hermann, 1960.

267 (XXI.11) Mathematics in the 19th Century

The 19th century was a critical period in the history of mathematics. When the century began, the memory of the French Revolution was still fresh, and World War I followed closely upon the turn of the 20th century. During this time mathematics made enormous progress and left a tremendous legacy to the present century. Increased personal liberty released people from traditions and allowed culture to spread to wider classes of society, producing a greater reservoir of talent. Research activities were intensified in the universities, and many specialists collaborated or competed with each other. The century can be divided into three periods: the first 20 years, during which many new fields of mathematics arose; the next 30 years, a period of further development; and the latter half of the century, when these fields attained maturity.

In 1801, *Disquisitiones arithmeticae* by the young C. F. †Gauss (1777–1855) appeared. It contained a systematized theory of numbers, ushering in a new era of mathematics. In France, many mathematicians studied at the Ecole Polytechnique, established during the French Revolution. Among them, A. L. †Cauchy (1789–1857) was one of the most prominent. He gave the exact definitions of

limit and convergence, thus giving solid foundations to calculus some 150 years after its discovery. N. H. †Abel (1802–1829) and C. G. J. †Jacobi (1804–1851) studied †elliptic functions during the same time, producing results sensational to their contemporaries. Gauss gave a rigorous proof of the existence of roots of algebraic equations in the field of complex numbers; Abel proved the algebraic nonsolvability of algebraic equations of degree $\geqslant 5$; and E. †Galois (1811–1832) created his theory of algebraic equations, which began a new phase in algebra. J. V. Poncelet (1788–1867), another graduate of the Ecole Polytechnique, developed †projective geometry along the lines pursued by G. Monge (1746–1818). His research was continued in Germany by A. F. Möbius (1790–1868), J. Steiner (1796–1863), and J. Plücker (1801–1868). Steiner investigated, in particular, †algebraic curves and surfaces by synthetic methods; Plücker introduced projective coordinates, thus enlarging the usage of analytic methods in geometry. These brilliant results in the first period of the 19th century were all achieved by young mathematicians, most of them still in their twenties.

The new geometry was developed in the period 1830–1840 by K. G. C. von Staudt (1798–1867) in Germany and M. Chasles (1793–1880) in France. In the 1840s, the theory of †invariants was taken up in connection with geometry; outstanding in this domain were the English mathematicians A. Cayley (1821–1895) and J. J. Sylvester (1814–1897). P. G. L. †Dirichlet (1805–1859) endeavored to simplify Gauss's number theory and introduced the †Dirichlet series in his computation of †class numbers of binary †quadratic forms. He also initiated the theory of †trigonometric series by giving a rigorous proof of a theorem on expansion in †Fourier series, introduced by J. B. J. †Fourier (1768–1830) in his theory of heat propagation. Another notable event was the independent and almost simultaneous discovery of †non-Euclidean geometry by J. Bolyai (1802–1860) and N. I. Lobachevskiĭ (1793–1856), which aroused philosophical interest since it changed the character of axioms. The invention of †quaternions by W. R. Hamilton (1805–1865), publication of *Ausdehungslehre* (theory of extensions) by H. G. Grassmann (1809–1877), and development of the algebra of logic by G. Boole (1815–1864) also occurred during this period, but these notions received neither full comprehension nor sympathy until later.

During the latter half of the century, G. F. B. †Riemann (1826–1866) and K. T. W. †Weierstrass (1815–1897) were prominent. Both have had great influence on the mathematics of the 20th century, the former by his brilliant and abundant production, the latter by his mature and critical spirit. Riemann lived only 40 years, and published, in rapid succession, his epoch-making ideas on the theory of functions of a complex variable, †Abelian functions, trigonometric series, †foundations of geometry, †distribution of primes, and †zeta functions. Weierstrass was already 49 when, after teaching in a country *gymnasium* (secondary or college preparatory school), he became a professor at the University of Berlin. The theory of the functions of a complex variable, initiated by Cauchy in the 1820s, had to wait for the contribution of these men to attain completion in the form of the theory of †elliptic functions. Riemann's influence is also considerable in algebraic geometry and the theory of †differential equations. Weierstrass reformed the †calculus of variations. His critical approach uncovered pathological functions, such as continuous nowhere differentiable functions (→ 106 Differential Calculus) and Peano space-filling curves (→ 93 Curves J), and to real analysis, based largely on the set theory of G. †Cantor (1848–1918).

Concerning the foundations of mathematics, Cantor, M. C. Méray, and J. W. R. Dedekind (1831–1916) established the theory of irrational numbers. Dedekind and G. Peano (1858–1932) developed the theory of natural numbers; their results brought about the "arithmetization" of mathematics and led to the research in the foundations of mathematics of the present century.

Cayley and F. †Klein (1849–1925) interpreted non-Euclidean geometry by means of metrics introduced in projective geometry. Toward the end of the century, D. †Hilbert (1862–1943) examined the roles of axioms of congruence, continuity, and parallelism in Euclidean geometry, thus initiating the study of †axiom systems in general.

The theory of †groups, in particular finite groups, was developed around 1870 by C. Jordan (1838–1922), G. Frobenius (1849–1917), and W. S. Burnside (1852–1927). M. S. †Lie (1842–1899) applied infinitesimal transformations to differential equations, and Klein applied the groups of linear transformations to geometry. The discovery of †automorphic functions by Klein and H. †Poincaré (1854–1912) was another brilliant application of the theory of groups. In the algebraic theory of numbers, originated by Gauss, E. E. †Kummer (1810–1893) developed the idea of "ideal numbers" (→ 14 Algebraic Number Fields); Dedekind then established the theory of †ideals. L. †Kronecker (1823–1891), an admirer of Abel's work, studied algebraic equations and discovered that every †Abelian extension of the

rational number field is contained in a †cyclotomic field; he believed that relations of a similar kind would hold between †modular equations of elliptic functions with †complex multiplications and Abelian extensions of †imaginary quadratic fields, and enunciated a famous conjecture known as his "dream in his youth."

Finally, we note the appearance of another important mathematician, S. V. Kovalevskaya (1850–1891). After studying with Weierstrass, in 1884 she was invited by G. M. Mittag-Leffler (1846–1927) to teach at the University of Stockholm, where she remained until her death.

Toward the end of the 19th century, the subjects of mathematical research became highly diversified. Branches were further ramified into more specialized branches, while unexpected relations were found between previously unconnected fields. The situation became so complicated that it was difficult to view mathematics as a whole. It was under these circumstances that in 1898, at the suggestion of F. Meyer and under the sponsorship of the Academies of Göttingen, Berlin, and Vienna, a project was initiated to compile an encyclopedia of the mathematical sciences. The *Enzyklopädie der mathematischen Wissenschaften* was completed in 20 years; it provides a useful overview of the mathematics of the 19th century.

Toward the end of the century, the International Congress of Mathematicians (ICM) was established to foster communication among mathematicians from all parts of the world. Before World War I broke out, the ICM met in Zürich (1896), Paris (1900), Heidelberg (1904), Rome (1908), and Cambridge, Mass. (1912). During this period, mathematical societies were formed in many countries, e.g., the London Mathematical Society (1865), the Société Mathématique de France (1872), the American Mathematical Society (1888), the Deutsche Mathematiker Vereinigung (1907), and the Mathematical Society of Tokyo (1877), which later became the Physico-Mathematical Society of Japan and then was split in two in 1946. The present Mathematical Society of Japan evolved from this division.

Five years after the 1872 reform of the Japanese educational system, the University of Tokyo was established, and D. Kikuchi (1855–1917) and R. Fujisawa (1861–1933) taught at the Department of Mathematics during its early ears. Under their influence, Japanese research in European-style mathematics (based on Greek traditions) began. The Universities of Kyoto and Tôhoku were established in 1897 and 1911, respectively. From the beginning of the 20th century, original results were obtained and published in the *Proceedings of the Physico-Mathematical Society of Japan* and in the journals of the faculties of science of these universities. In 1911, the *Tôhoku Mathematical Journal* was founded by T. Hayashi (1873–1935). In 1920, a paper on †class field theory by T. †Takagi (1875–1960) was published in the *Journal of the College of Science of the University of Tokyo*. Thus the position of Japanese mathematics gradually came to be established.

References

[1] F. Klein, Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert, Springer, I, 1926; II, 1927 (Chelsea, 1956).
[2] N. Bourbaki, Les éléments d'histoire des mathématiques, Hermann, second edition, 1969.
[3] Enzyklopädie der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen, Teubner, 1898–1934.

268 (XIV.10) Mathieu Functions

A. Mathieu's Differential Equation

The 2-dimensional †Helmholtz equation $(\Delta+k^2)\Psi=0$ $(\Delta=\partial^2/\partial x^2+\partial^2/\partial y^2)$, separated in †elliptic coordinates ξ, η given by $x=c\cosh\xi\cos\eta$, $y=c\sinh\xi\sin\eta$, has a solution of the form $\Psi=X(\xi)Y(\eta)$ whose factors $X(\xi)$, $Y(\eta)$ satisfy

$$d^2u/dz^2+(a-2q\cos 2z)u=0, \tag{1}$$

$$d^2u/dz^2-(a-2q\cosh 2z)u=0, \tag{2}$$

respectively, where a is an arbitrary constant and $q=k^2c^2/4$. By the substitution $z\to\pm iz$, (1) becomes (2). (1) and its solutions are known as **Mathieu's differential equation** and the **Mathieu functions**, and (2) and its solutions are called the **modified Mathieu differential equation** and the **modified Mathieu functions**.

B. Hill's Differential Equation

Hill's differential equation is a linear ordinary differential equation of the second order:

$$d^2u/dx^2+F(x)u=0, \tag{3}$$

with $F(x+2\pi)=F(x)$. It is named after G. W. Hill, who investigated it in his study of lunar motion. This equation includes Mathieu's differential equation and †Lamé's differential equation as particular cases, and by suitable

transformations, †Legendre's differential equation and the †confluent hypergeometric differential equations as well.

While $F(x)$ is periodic, solutions of (3) are not necessarily so. There always exists, however, a particular solution that is **quasiperiodic** in the sense that

$$u(x+2\pi)=\sigma u(x), \qquad \sigma=\text{constant} \tag{4}$$

(Floquet's theorem). That is, the differential equation (3) has a solution of the form

$$u(x)=e^{\mu x}\varphi(x), \tag{5}$$

where $\varphi(x+2\pi)=\varphi(x)$ and μ (defined by $\sigma=e^{2\pi\mu}$) is called the **characteristic exponent**.

Being a particular case of Hill's equation, (1) has a general solution of the form

$$u(z)=Ae^{\mu z}\varphi(z)+Be^{-\mu z}\varphi(-z), \tag{6}$$

where $\varphi(z+\pi)=\varphi(z)$. For those values of a (called **eigenvalues**) that make the characteristic exponent μ equal to 0 or i, $u(z)$ has period π or 2π and is called a **Mathieu function of the first kind**, also called an **elliptic cylinder function** when employed in problems of diffraction by an elliptic cylinder. Sometimes it is called simply the Mathieu function, and other solutions of (1) are referred to as **general Mathieu functions**. The formula

$$F(x)=\sum_{m=-\infty}^{\infty}a_m e^{imx} \tag{7}$$

suggests, in conjunction with Floquet's theorem, a solution of (3) in the form

$$u=e^{\mu x}\sum_{n=-\infty}^{\infty}b_n e^{inx}. \tag{8}$$

Substituting (7) and (8) into (3) and comparing coefficients of $e^{(\mu+in)x}$, we have infinitely many linear equations

$$(\mu+in)^2 b_n+\sum_{m=-\infty}^{\infty}a_m b_{n-m}=0, \qquad n=\ldots,-2,-1,0,1,2,\ldots. \tag{9}$$

By eliminating the b_n in (9), we also obtain an infinite determinantal equation called **Hill's determinantal equation**,

$$\Delta(\mu)=|B_{rs}|=0, \tag{10}$$

where the elements B_{rs} of **Hill's determinant** $\Delta(\mu)$ are such that

$$B_{rs}=\begin{cases}1 & \text{if } r=s,\\ \dfrac{a_{r-s}}{(\mu+ir)^2+a_0} & \text{if } r\neq s.\end{cases}$$

Here an **infinite determinant** $D=|B_{mn}|(m,n=-\infty,\ldots,\infty)$ is defined as the limit, if it exists, of $D_m=\det(B_{ij})$ $(i,j=-m,\ldots,m)$ as $m\to\infty$. The formula (10) can be reduced to a simpler form

$$\sin^2\pi i\mu=\Delta(0)\sin^2\pi\sqrt{a_0}. \tag{11}$$

This determines a characteristic exponent μ, which in turn determines the b_n in (9) and a solution (8). This procedure is called **Hill's method of solution**.

Applying Hill's method of solution to (1), we have an equation for the characteristic exponent μ:

$$\sin^2(\pi/2)i\mu=\Delta(0)\sin^2(\pi/2)\sqrt{a}, \tag{12}$$

where the infinite determinant $\Delta(0)=|B_{mn}|$ has elements such that

$$B_{mn}=\begin{cases}1 & \text{if } m=n\\ \dfrac{-q}{a-4m^2} & \text{if } n=m\pm1\\ 0 & \text{otherwise}\end{cases} \qquad (m,n=\ldots,-1,0,1,\ldots).$$

When $q\to0$, we have

$$\Delta(0)=1+2q^2\frac{\pi}{8\sqrt{a}(1-a)}\cot\frac{\pi\sqrt{a}}{2}+O(q^4). \tag{13}$$

Thus if $q=0$, we have $\Delta(0)=1$, and $a=4n^2$, $(2n+1)^2$ correspond to $\mu=0$, i in (12).

C. Mathieu Functions of the First Kind

Mathieu functions of the first kind are further classified into the following four types:

$$ce_{2n}(z,q)=\sum A_r^{(2n)}\cos 2rz, \qquad (a_{2n}), \tag{14.1}$$

$$se_{2n+1}(z,q)=\sum B_{2r+1}^{(2n+1)}\sin(2r+1)z, \qquad (b_{2n+1}), \tag{14.2}$$

$$ce_{2n+1}(z,q)=\sum A_{2r+1}^{(2n+1)}\cos(2r+1)z, \qquad (a_{2n+1}), \tag{14.3}$$

$$se_{2n+2}(z,q)=\sum B_{2r+2}^{(2n+2)}\sin(2r+2)z, \qquad (b_{2n+2}) \tag{14.4}$$

$(n,r=0,1,2,\ldots)$, where the appended terms in parentheses are eigenvalues, ordered by $a_{2n}<b_{2n+1}<a_{2n+1}<b_{2n+2}$ for a given q, and increasing with n. Each of these series converges absolutely and uniformly for all finite z and has n zeros in $0<z<\pi/2$. In addition, orthonormality relations

$$\int_0^{2\pi}ce_m(x)se_n(x)\,dx=0,$$

$$\int_0^{2\pi}ce_m(x)ce_n(x)\,dx=\int_0^{2\pi}se_m(x)se_n(x)\,dx=\pi\delta_{mn}$$

hold. When $q\to0$, $ce_0(z)\to1/\sqrt{2}$, $ce_m(z)\to\cos mz$, and $se_m(z)\to\sin mz$.

For small q we assume that the quantities involved have power series expansions in q,

e.g.,

$$a=m^2+\alpha q+\beta q^2+\dots,$$

$$ce_m(z)=\cos mz+qF_1(z)+q^2F_2(z)+\dots,$$

and we substitute them in (1) to determine successively $\alpha, \beta, \dots, F_1(z), F_2(z), \dots$ (**Mathieu's method**). For larger q, (14) is substituted in (1) to give recurrence formulas for the coefficients, e.g., for $ce_{2n}(z)$,

$$-aA_0^{(2n)}+qA_0^{(2n)}=0,$$

$$2qA_0^{(2n)}+(4-a)A_2^{(2n)}+qA_4^{(2n)}=0,$$

$$qA_{2r-2}^{(2n)}+(4r^2-a)A_{2r}^{(2n)}+qA_{2r+2}^{(2n)}=0, \qquad r>1. \tag{15}$$

After we eliminate the $A_{2r}^{(2n)}$, the formulas lead to an equation for the eigenvalues a_{2n}:

$$\begin{vmatrix} a & -q & 0 & 0 & 0 & \dots \\ -2q & a-4 & -q & 0 & 0 & \dots \\ 0 & -q & a-16 & -q & 0 & \dots \\ & & \dots & & & \end{vmatrix}=0, \tag{16}$$

or, equivalently, to a †continued fraction

$$a=\frac{2q^2}{a-4-}\frac{q^2}{a-16-}\frac{q^2}{a-36-}\dots. \tag{17}$$

Given q, we can find the a_{2n} from (17) and determine the $A_{2r}^{(2n)}$ for each a_{2n} from (15) (**Ince-Goldstein method**).

D. Mathieu Functions of the Second Kind and Modified Mathieu Functions

There exists only one (half-)periodic solution of (1) corresponding to each (half-)periodic eigenvalue (→ Section E). Therefore other solutions corresponding to the same a_m or b_m and independent of $ce_m(z,q)$ or $se_m(z,q)$ are nonperiodic. They are called the **Mathieu functions of the second kind** and are denoted by $fe_m(z,q)$ or $ge_m(z,q)$.

By the substitution $z\to iz$ in (14) we obtain formulas for the **modified Mathieu functions of the first kind**,

$$Ce_m(z,q)=ce_m(iz,q),$$

$$Se_m(z,q)=-ise_m(iz,q). \tag{18}$$

When $q\to0$, then $Ce_0(z)\to1/\sqrt{2}$, $Ce_m(z)\to\cosh mz$, $Se_m(z)\to\sinh mz$. Similarly, by the substitution $z\to iz$ we obtain **modified Mathieu functions of the second kind** from Mathieu functions of the second kind. In addition, we introduce **modified Mathieu functions of the third kind** as those linear combinations of modified Mathieu functions of the first and second kinds that have the asymptotic form $e^{-y}y^{-1/2}$ $(y=q^{1/2}e^z)$ as $z\to\infty$. In addition to the Fourier expansion (7), expansion of the Mathieu functions in terms of †Bessel functions is possible, e.g., after taking $q=h^2$,

$$Ce_{2n}(z,q)=\sum A_{2r}\cosh 2rz \tag{19.1}$$

$$=(A_0)^{-1}ce_{2n}(\pi/2,q)\times\sum(-1)^rA_{2r}J_{2r}(2h\cosh z) \tag{19.2}$$

$$=(A_0)^{-1}ce_{2n}(0,q)\times\sum A_{2r}J_{2r}(2h\sinh z) \tag{19.3}$$

$$=(A_0)^{-2}ce_{2n}(0,q)ce_{2n}(\pi/2,q)\times\sum(-1)^rA_{2r}J_r(he^{-z})J_r(he^z). \tag{19.4}$$

These series converge absolutely and uniformly for all finite z. Replacing the J on the right-hand sides of (19) by $N_{2r}(2h\cosh z)$, $N_{2r}(2h\sinh z)$, $J_r(he^{-z})N_r(he^z)$, respectively, in an obvious way we obtain infinite series for a function that again satisfies (2) and is denoted by $Fey_{2n}(z,q)$. In a similar manner other modified Mathieu functions of the second kind $Gey_{2n+1}(z,q)$, $Fey_{2n+1}(z,q)$, $Gey_{2n+2}(z,q)$ can be obtained. These are more convenient for practical applications than $fe_m(iz,q)$ and $ge_m(iz,q)$ since they converge more rapidly.

The equations

$$d^2u/dz^2+(a+2q\cos 2z)u=0, \tag{20}$$

$$d^2u/dz^2-(a+2q\cosh 2z)u=0 \tag{21}$$

obtained from (1) and (2) by the substitution $q\to-q$ are the results of separating $(\Delta-k^2)\varphi=0$. In general, if $f(z,q)$ is a solution of (1), then $f(\pi/2-z,q)$ is a solution of (20). Thus the formulas

$$ce_{2n}(z,-q)=(-1)^nce_{2n}(\pi/2-z,q), \tag{22.1}$$

$$ce_{2n+1}(z,-q)=(-1)^nse_{2n+1}(\pi/2-z,q), \tag{22.2}$$

$$se_{2n+1}(z,-q)=(-1)^nce_{2n+1}(\pi/2-z,q), \tag{22.3}$$

$$se_{2n+2}(z,-q)=(-1)^nse_{2n+2}(\pi/2-z,q) \tag{22.4}$$

can be adopted as definitions of $ce_m(z,q)$, $se_m(z,q)$ for $q<0$ (**Ince's definition**). Accordingly, an expansion of Ce holds in terms of modified Bessel functions I_m in place of the J_m in (19), which in turn becomes a solution of (21) if we replace I_m by $(-1)^mK_m/\pi$:

$$Fek_{2n}(z,-q)=(-1)^n(\pi A_0)^{-1}ce_{2n}(\pi/2,q)\times\sum A_{2r}K_{2r}(2h\sinh z). \tag{23}$$

In a similar manner, $Fek_{2n+1}(z,-q)$, $Gek_{2n+1}(z,-q)$, $Gek_{2n+2}(z,-q)$ can be defined. They decrease exponentially as $z\to\infty$ and hence are precisely the modified Mathieu functions of the third kind.

E. Stability

Let $u_1(x)$, $u_2(x)$ be a fundamental system of solutions of (3) such that $u_1(0)=1$, $u_1'(0)=0$;

$u_2(0)=0$, $u_2'(0)=1$; then σ, μ in (4), (5) are given by

$$\sigma=e^{2\pi\mu}, \qquad 2\pi\mu=\operatorname{arccosh} A,$$

$$2A=u_1(2\pi)+u_2'(2\pi). \tag{24}$$

There are two values of μ satisfying (24); let them be μ_1 and μ_2 ($\mu_2=-\mu_1$). If $F(x)$ is a real function, then $u_1(x)$, $u_2(x)$ are also real functions and A is a real number. It can be seen from (24) that $A<-1$, $1<A<0$, $0<A<1$, $1<A$ correspond to $\pm 2\pi\mu=\operatorname{arccosh}|A|+\pi i$, $i(\arccos|A|+\pi)$, $i\arccos A$, $\operatorname{arccosh} A$, respectively. Consequently, when $|A|<1$, the general solution of (3) neither diverges nor vanishes as x tends to infinity. Such solutions are called **stable solutions** of Hill's equation, or sometimes **Hill's functions**. When $|A|>1$, either $e^{\mu_1 x}$ or $e^{\mu_2 x}$ tends to infinity with x. Such solutions are called **unstable solutions**. If $A=1$ (-1), we have $\mu=0$ $(i/2)$ and a solution of the form of $u=\varphi(x)$ $(e^{ix/2}\varphi(x))$ with the period 2π (4π), called a **periodic (half-periodic) solution**.

When we apply the Mathieu functions to physical and engineering sciences, such as the theory of oscillation and quantum mechanics, it is convenient to modify (3) in the form

$$d^2u/dx^2+(\lambda+\gamma\Phi(x))u=0 \tag{25}$$

involving parameters λ, γ. Then with γ fixed, there exist countably many values of λ (called eigenvalues) corresponding to periodic or half-periodic solutions of (25). Let them be $\lambda_0 \leqslant \lambda_1 \leqslant \ldots$ or $\bar\lambda_1 \leqslant \bar\lambda_2 \leqslant \ldots$, respectively; then we have

$$\lambda_0<\bar\lambda_1\leqslant\bar\lambda_2<\lambda_1\leqslant\lambda_2<\ldots<\bar\lambda_{2k-1}\leqslant\bar\lambda_{2k}$$
$$<\lambda_{2k-1}\leqslant\lambda_{2k}<\ldots.$$

The values of λ in the open intervals $(\lambda_{2k-2}, \bar\lambda_{2k-1})$ and $(\bar\lambda_{2k}, \lambda_{2k-1})$ correspond to stable solutions, while the λ's in other intervals correspond to unstable solutions (**Haupt's theorems**).

When both λ and γ vary, the $\lambda\gamma$-plane is divided into regions corresponding to stable or unstable solutions according as the characteristic exponent μ is purely imaginary or not. For example, when $\Phi(x)=2\cos x$ in equation (25), we obtain the Mathieu equation. In this case the $\lambda\gamma$-plane is divided as shown in Fig. 1, where shaded (unshaded) regions correspond to stable (unstable) solutions and boundary curves give eigenvalues corresponding to periodic or half-periodic solutions.

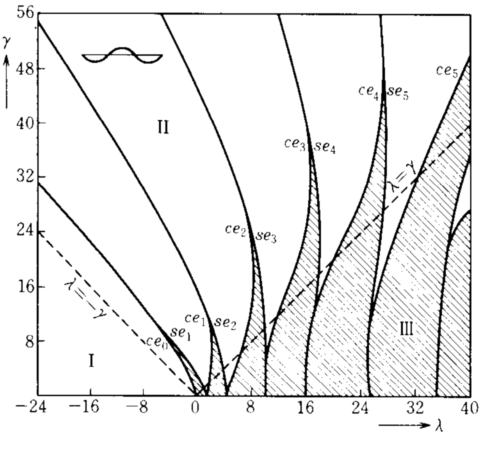


Fig. 1

References

[1] E. T. Whittaker and G. N. Watson, A course of modern analysis, Cambridge Univ. Press, fourth edition, 1958.

[2] M. J. O. Strutt, Lamésche-, Mathieusche-, und verwandte Funktionen in Physik und Technik, Erg. Math., Springer, 1932 (Chelsea, 1967).

[3] N. W. McLachlan, Theory and application of Mathieu functions, Clarendon Press, 1947.

[4] J. Meixner and F. W. Schäfke, Mathieusche Funktionen und Sphäroidfunktionen, Springer, 1954.

Also → references to 133 Ellipsoidal Harmonics.

269 (III.2) Matrices

A. General Remarks

Let K be a †ring or a †field. As examples of such K, we may take the real number field **R** and the complex number field **C**. By a **matrix** with elements in K, we mean an array of mn elements a_{ik} $(i=1,\ldots,m;\ k=1,\ldots,n)$ of K arranged in a rectangular form:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & 22 & \cdots & a_{2n} \\ & & \cdots & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

The element a_{ik} is called the (i,k)**-element** (**entry** or **component**). (Sometimes, instead of using parentheses, we use $\| \quad \|$ or $[\quad]$.) More precisely, this matrix is said to be an m **by** n **matrix** ($m\times n$ **matrix** or **matrix of** (m,n)**-type**). In particular, an $n\times n$ matrix is called a **square matrix** of **degree** (or **order**) n, while a matrix in general is sometimes called a **rectangular matrix**. Each horizontal n-tuple in an $m\times n$ matrix is called a **row** of the matrix, and each vertical m-tuple is called a **column** of the matrix. We often abbreviate the notation for the matrix given previously by writing (a_{ik}) or

simply A. A square matrix is called a **diagonal matrix** if all its components are zero except possibly for diagonal components a_{ii}, that is, if $a_{ik}=0$ for $i \neq k$. If the components a_{ii} of a diagonal matrix are all equal, it is called a **scalar matrix**. An $n \times n$ matrix whose (i,k)-element is equal to δ_{ik} is called the **unit matrix** (or **identity matrix**) of degree n, where δ_{ik} is the **Kronecker delta**, which is defined by $\delta_{ii}=1$ and $\delta_{ij}=0$ for $i \neq j$. We denote this matrix by I_n, or simply by I if there is no need to specify n.

A $1 \times n$ matrix $(a_1, a_2, \ldots, a_n)$ is called a **row vector** of dimension n, and an $m \times 1$ matrix

$$\begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$$

is called a **column vector** of dimension m. The m rows and n columns of an $m \times n$ matrix A are called row vectors and column vectors of A.

B. Operations on Matrices

Two matrices $A=(a_{ik})$ and $B=(b_{ik})$ are said to be equal if and only if they are of the same type and $a_{ik}=b_{ik}$ $(i=1,\ldots,m; k=1,\ldots,n)$. If both A and B are $m \times n$ matrices, we define the **sum** of A and B by $A+B=(a_{ik}+b_{ik})$. The product of two matrices A and B is defined by $AB=(c_{ik})$, $c_{ik}=\sum_j a_{ij}b_{jk}$, provided that the number of columns of A is equal to the number of rows of B. We further define the (left and right) multiplication of a matrix A by an element a of K by $aA=(aa_{ik})$ and $Aa=(a_{ik}a)$. The set of all matrices over K of the same type forms a †K-module. The multiplication of matrices satisfies the associative law and the left and right distributive laws with respect to addition. Thus the set of all $n \times n$ matrices over K forms a ring, which is called the **total matrix algebra** (or **full matrix algebra**) of degree n over K; it is usually denoted by $M_n(K)$ or K_n. If K has the †unity element 1, then I_n is the unity element of $M_n(K)$. The matrix whose components are all 0 is called the **zero matrix** and is denoted by the same symbol 0. Suppose that K has the unity element 1. Let E_{ik} be the matrix whose (i,k)-element is 1 and whose other elements are all 0. Then every matrix $A=(a_{ik})$ in $M_n(K)$ can be expressed uniquely as $A=\sum a_{ik}E_{ik}$, a linear combination of E_{ik}. The matrix E_{ik} is called a **matrix unit**. We have $E_{ij}E_{kl}=0$ if $j \neq k$, $E_{ik}E_{kl}=E_{il}$, and $aE_{ik}=E_{ik}a$ for all $a \in K$.

Let A be a square matrix in K. If there exists a matrix A^{-1} such that $AA^{-1}=A^{-1}A=I$, then A^{-1} is called the **inverse matrix** of A, and A is called a **regular matrix (nonsingular matrix** or **invertible matrix**). The inverse A^{-1} is unique if it exists. When K is commutative, A is regular if and only if its †determinant $|A|$ is a †regular element of K; in particular, when K is a field, A is regular if and only if $|A| \neq 0$. A square matrix that is not regular is called a singular matrix.

Let $A=(a_{ik})$ be an $m \times n$ matrix. Then the $n \times m$ matrix (b_{ik}) such that $b_{ik}=a_{ki}$ for all i and k is called the **transposed matrix** of A, and is usually denoted by ${}^tA$. Transposing a matrix amounts to changing rows into columns and vice versa. When K is commutative, $AB=C$ implies ${}^tC={}^tB\,{}^tA$. A square matrix such that ${}^tA=A$ is called a **symmetric matrix**, and one such that ${}^tA=-A$ is called an **alternating (skew-symmetric** or **antisymmetric) matrix**. A square matrix $A=(a_{ik})$ is called an **upper (lower) triangular matrix** if $a_{ik}=0$ for $i>k$ $(i<k)$.

C. The Kronecker Product of Matrices

We assume that the ring K is commutative. Let A be an $m \times n$ matrix (a_{ik}), let B be an $r \times s$ matrix (b_{jl}) in K, and write $c_{\lambda,\mu}=a_{ik}b_{jl}$ by means of indexes $\lambda=(i,j)$ and $\mu=(k,l)$. The **Kronecker product** of A and B, usually denoted by $A \otimes B$, is defined as the $mr \times ns$ matrix $C=(c_{\lambda,\mu})$. By an appropriate choice of λ and μ, it can be expressed as

$$\begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ & & \cdots & \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix}$$

or

$$\begin{pmatrix} b_{11}A & b_{12}A & \cdots & b_{1s}A \\ b_{21}A & b_{22}A & \cdots & b_{2s}A \\ & & \cdots & \\ b_{r1}A & b_{r2}A & \cdots & b_{rs}A \end{pmatrix}$$

We have the formulas

$$A \otimes (B_1+B_2)=A \otimes B_1 + A \otimes B_2,$$

$$(A_1+A_2) \otimes B = A_1 \otimes B + A_2 \otimes B,$$

$$c(A \otimes B)=(cA) \otimes B = A \otimes (cB),$$

$$(A_1 \otimes B_1)(A_2 \otimes B_2)=(A_1A_2) \otimes (B_1B_2),$$

provided that the sums and products can be defined.

Matrices correspond to †linear mappings of free K-modules (→ Section L). The Kronecker product corresponds to the †tensor product of the corresponding linear mappings.

D. The Rank of a Matrix

Let A be an $m \times n$ matrix (a_{ik}) in a field K. If there exists a nonzero †minor of A of degree r, and if all minors of degree $\geqslant r+1$ are equal to

0, then the number r is called the **rank** of A. Denoting the rank of a matrix A by $\rho(A)$, we have $\rho(PAQ) \leqslant \rho(A)$ for any matrices P, Q for which the product PAQ exists; the equality holds if P and Q are regular (square) matrices. The rank r of A is equal to the maximum number of †linearly independent row vectors of A, and it is also equal to the maximum number of linearly independent column vectors of A. The number $n-\rho(A)$ is called the **(column) nullity** of the matrix A. It is equal to the dimension of the linear space consisting of solutions of homogeneous linear equations $Ax=0$, that is, the number of fundamental solutions of these equations. (The **row nullity** $m-\rho(A)$ of A is equal to the dimension of the linear space consisting of solutions of $xA=0$.) Even if K is not commutative, we can define the rank of a matrix A as the maximum number of left (right) linearly independent row (column) vectors of A (→ 256 Linear Spaces F).

E. Elementary Divisors

Let $\mathfrak{o}$ be a †principal ideal domain (e.g., the ring $\mathbf{Z}$ of rational integers, or the †polynomial ring $K[x]$ over a field K), and let A be a matrix with elements in $\mathfrak{o}$. Then, multiplying appropriate invertible matrices from the left and right, we can transform A into a diagonal matrix of the form

$$\begin{pmatrix} e_1 & & & & & & 0 \\ & e_2 & & & & & \\ & & \ddots & & & & \\ & & & e_r & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ 0 & & & & & & 0 \end{pmatrix},$$

where r denotes the rank of A, each $e_i \neq 0$, and each e_{i+1} is divisible by e_i. These $e_1, e_2, \ldots, e_r$ are called the **elementary divisors** of A and are uniquely determined up to unit factors. This fact can be generalized to some extent for the case where $\mathfrak{o}$ is noncommutative (→ 256 Linear Spaces P). Using the concept of determinant, the elementary divisors can also be defined as follows. Let d_k denote the greatest common divisor of all minors of degree k of A. The elements $e_i = d_i/d_{i-1}$ $(i=1, \ldots, r; d_0=1)$ of $\mathfrak{o}$ are the elementary divisors of A. The numbers d_i $(1 \leqslant i \leqslant r)$ are called the **determinant factors** of A.

F. Characteristic Polynomials and Eigenvalues

Let $A=(a_{ik})$ be a square matrix of degree n in a field K. The determinant $F(x)=|xI-A|$ is a polynomial in x with coefficients in K, called the **characteristic polynomial** of A. The algebraic equation $F(x)=0$ is called the **characteristic equation** of A, and its roots $\lambda_1, \lambda_2, \ldots, \lambda_n$ are called the **eigenvalues (proper values** or **characteristic roots)** of A. The †determinant of A, $\det A$, is equal to $\prod_{i=1}^n \lambda_i$, the product of the eigenvalues. The sum of eigenvalues, $\sum_{i=1}^n \lambda_i = \sum_{i=1}^n a_{ii}$, is called the **trace** (Ger., **Spur**) or **diagonal sum** of A, and is denoted by $\operatorname{tr} A$ or $\operatorname{Sp}(A)$. In case K is †algebraically closed, there exists a nonzero vector x that satisfies the equations $Ax=\lambda x$ if and only if λ is an eigenvalue of A. This solution x is called an **eigenvector (proper vector** or **characteristic vector)** belonging to the eigenvalue λ. In particular, if all the a_{ik} are real and A is symmetric, then all eigenvalues of A are real, and the characteristic equation $F(x)=0$ of A is called a **secular equation**. Every square matrix A satisfies its characteristic equation, i.e., $F(A)=0$. This is called the **Hamilton-Cayley theorem**, and it is useful in numerical calculation of inverse matrices.

Let A be a square matrix. As is clear from above, there exist monic polynomials $f(x)$ $(\neq 0)$ in $K[x]$ such that $f(A)=0$. Let $\varphi(x)$ be such a polynomial of the least degree. Then every $f(x)$ is divisible by $\varphi(x)$. This $\varphi(x)$ is called a **minimal polynomial** of A. Let $e_1(x), \ldots, e_n(x)$ be the elementary divisors of the matrix $xI-A$. Then $e_n(x)=\varphi(x)$. Now if $F(x)=x^n$, A is said to be a **nilpotent matrix**, and if $F(x)=(x-1)^n$, A is said to be a **unipotent matrix**.

G. Jordan Normal Form

Two square matrices A and B are said to be **similar** to each other if $B=P^{-1}AP$ with a regular matrix P. The matrices A and B are similar if and only if $xI-A$ and $xI-B$ have the same elementary divisors. Now suppose that A has elements in a field K and that every eigenvalue of A is in K. Then there is a regular matrix P with elements in K such that $P^{-1}AP$ is of the form

$$P^{-1}AP = \begin{pmatrix} A_1 & & & 0 \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_s \end{pmatrix},$$

where

$$A_i = \begin{pmatrix} \lambda_i & 1 & & 0 \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_i \end{pmatrix}.$$

(A_i is a square matrix of a certain degree, say m_i; if $m_i=1$ then $A_i=(\lambda_i)$.) This matrix $P^{-1}AP$ is called the **Jordan normal form** of A. It is a diagonal matrix if and only if the minimal

polynomial of A has no multiple roots. If this is the case, A is said to be **semisimple**.

H. The Exponential Function of a Matrix

Let $A_\nu=(a_{ij}^{(\nu)})$ $(\nu=0,1,2,\ldots)$ be square matrices with elements in the complex number field **C**. The series $A_0+A_1+\ldots+A_\nu+\ldots$ is said to be convergent if the series of components $a_{ij}^{(0)}+a_{ij}^{(1)}+\ldots+a_{ij}^{(\nu)}+\ldots$ is convergent for each (i,j). For every square matrix A, the following series is convergent:

$$I+A+\frac{1}{2!}A^2+\ldots+\frac{1}{\nu!}A^\nu+\ldots.$$

We denote this series by **exp** A. Then $\exp({}^tA)={}^t(\exp A)$; $\exp(-A)=(\exp A)^{-1}$; and if $AB=BA$, $\exp(A+B)=\exp A\exp B$. Moreover, $\det(\exp A)=\exp(\operatorname{tr}A)$. Furthermore, if we set $F(t)=\exp(tA)$, then $dF(t)/dt=AF(t)$ (→ 431 Transformation Groups).

I. Normal Matrices, Unitary Matrices, and Hermitian Matrices

Let $A=(a_{ik})$ be a square matrix with elements in the complex number field **C**. Then the **adjoint matrix** A^* of A is the transposed conjugate ${}^t\bar{A}=(\bar{a}_{ki})$, where $\bar{a}$ is the complex conjugate of a. If $AA^*=A^*A$, then A is said to be a **normal matrix**. A matrix U such that $U^*U=I$, that is, $U^{-1}=U^*$, is normal. Following G. Frobenius, we call it a **unitary matrix** (in the following discussion, we shall call it simply a u-matrix). A matrix H such that $H^*=H$ is also normal. It is called a **Hermitian matrix** (which we shall refer to as an h-matrix). An h-matrix P such that $P^2=P$ is called a **projection matrix**.

The set of all u-matrices forms a †group with respect to matrix multiplication. If A is a normal matrix and U is a u-matrix, U^*AU is a normal matrix. If H is an h-matrix, Q^*HQ is also an h-matrix for any matrix Q. Every normal matrix A can be transformed into a diagonal matrix by a suitable u-matrix (i.e., there exists a u-matrix U such that $U^*AU=U^{-1}AU$ is a diagonal matrix), and conversely, every matrix with this property is a normal matrix. Thus, in particular, any h-matrix or u-matrix can be transformed into a diagonal matrix by a suitable u-matrix. Moreover, every real symmetric matrix, which is naturally an h-matrix, can be transformed into a diagonal matrix by an orthogonal matrix; this will be defined later. If $A_1,\ldots,A_m$ are mutually commutative normal matrices, we can transform them into diagonal matrices by the same u-matrix; that is, there exists a u-matrix U such that $U^{-1}A_iU$ is of diagonal form for all i. All eigenvalues of an h-matrix are real, and all eigenvalues of a u-matrix have absolute value 1.

If all eigenvalues of an h-matrix H are positive (positive or zero), then H is said to be **positive definite (positive semidefinite)**. For an h-matrix H, $\exp H$ is a positive definite h-matrix, and conversely, every positive definite h-matrix can be expressed as $\exp H$ by a unique h-matrix H. Furthermore, every regular matrix A can be uniquely expressed as $A=UH$ (or $A=H'U'$) by a u-matrix U (or U') and a positive definite h-matrix H (or H'), and A is normal if and only if $UH=HU$.

A square matrix A such that $A^*=-A$ is called a **skew-Hermitian matrix** (simply **skew h-matrix** or **anti-Hermitian matrix**). All eigenvalues of a skew h-matrix are purely imaginary numbers. If A is a skew h-matrix, $\exp A$ is a u-matrix. Conversely, if a u-matrix U lies in a sufficiently small neighborhood of the identity matrix I (i.e., if all elements of $U-I$ have sufficiently small absolute values), U can be uniquely expressed in the form $\exp A$ with a skew h-matrix A.

J. Orthogonal Matrices

An h-matrix whose components are all real is necessarily symmetric. A u-matrix whose components are all real, that is, a real matrix R such that $R^{-1}={}^tR$, is called an **orthogonal matrix**. The totality of orthogonal matrices forms a group with respect to matrix multiplication. If S is a real symmetric matrix, there exists an orthogonal matrix T for which $T^{-1}ST$ is a diagonal matrix; that is, S can be transformed into a diagonal form by T. Every orthogonal matrix R can be transformed by an orthogonal matrix T into the diagonal form

$$T^{-1}RT=\begin{pmatrix}1&&&&&&&&\\&\ddots&&&&&&&\\&&1&&&&&&\\&&&-1&&&&&\\&&&&\ddots&&&&\\&&&&&-1&&&\\&&&&&&P_1&&\\&&&&&&&\ddots&\\&&&&&&&&P_t\end{pmatrix}$$

$$P_j=\begin{pmatrix}\cos\theta_j&\sin\theta_j\\-\sin\theta_j&\cos\theta_j\end{pmatrix}\quad\text{for}\quad j=1,\ldots,t.$$

The determinant $|R|$ of an orthogonal matrix R is either 1 or -1. If $|R|=1$, then R is called a **proper orthogonal matrix**. If A is a real alternating matrix, then $\exp A$ is a proper orthogonal matrix; conversely, any pro-

per orthogonal matrix in a sufficiently small neighborhood of the identity matrix I can be expressed uniquely in this form. Moreover, there exists a one-to-one correspondence between real alternating matrices A and proper orthogonal matrices R without eigenvalues equal to -1, given by $R=(I-A)(I+A)^{-1}$ and $A=(I-R)(I+R)^{-1}$. This is called a **Cayley transformation.**

A (complex) square matrix T with the property ${}^tT=T^{-1}$ is called a **complex orthogonal matrix**. This matrix can be uniquely expressed as $T=R\exp(iA)$, where R is an orthogonal matrix, A is a real alternating matrix, and $i^2=-1$.

K. Infinite Matrices

By an **infinite matrix** with elements in a ring K, we mean an array of elements of K with infinite numbers of rows and columns as follows:

$$\begin{pmatrix} \cdots & \cdots & \cdots & \cdots \\ \cdots & a_{\sigma\tau} & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix},$$

where σ, τ are indices denoting the row and column for the element $a_{\sigma\tau}$ and each index ranges over an infinite set Γ. Equality, addition, and multiplication by an element of K of infinite matrices are defined in the same manner as for ordinary matrices. Generally, however, multiplication of infinite matrices cannot be defined. If the elements of each row (column) of an infinite matrix are zero except for a finite number of them, then it is called a **row (column) finite matrix**. For row (column) finite matrices $A=(a_{\sigma\tau})$ and $B=(b_{\sigma\tau})$, the product $AB=(c_{\sigma\tau})$ is defined by $c_{\sigma\tau}=\sum_\nu a_{\sigma\nu}b_{\nu\tau}$ for all $\sigma,\tau\in\Gamma$. By this definition the totality of such infinite matrices forms a ring.

Now let K be the complex number field and Γ the set of natural numbers, and consider an infinite matrix (a_{ik}). This is called a **bounded matrix** if the following inequality holds for arbitrary x_i and y_k:

$$\left|\sum_{i,k=1}^{m,n} a_{ik}x_iy_k\right| \leqslant M\left(\sum_{i=1}^{m}|x_i|^2\right)^{1/2}\left(\sum_{k=1}^{n}|y_k|^2\right)^{1/2},$$

where M is a constant. The set of all bounded matrices forms a ring. If $\varphi_1, \varphi_2, \ldots$ is a complete orthonormal system of a †Hilbert space $\mathfrak{H}$, then for any continuous linear operator A we have $A\varphi_k=\sum_{i=1}^{\infty}\varphi_i a_{ik}$, and A corresponds to a bounded matrix (a_{ik}). By this correspondence we have a ring isomorphism between the ring of continuous linear operators of $\mathfrak{H}$ and the ring of bounded matrices (→ 197 Hilbert Spaces).

L. Linear Mappings

Let L, L' be free right K-modules with bases $a_1,\ldots,a_m$ and $b_1,\ldots,b_n$, respectively. Then, to a †linear mapping f of L into L', there corresponds an $n\times m$ matrix A such that

$$f:\sum a_ic_i=(a_1,\ldots,a_m)\begin{pmatrix}c_1\\ \vdots\\ c_m\end{pmatrix}\mapsto(b_1,\ldots,b_n)A\begin{pmatrix}c_1\\ \vdots\\ c_m\end{pmatrix}.$$

Conversely, for any $n\times m$ matrix A with elements in K, there is a unique linear mapping f as above (with respect to these bases). In this sense, if a linear mapping g of L' (into some free K-module) corresponds to a matrix B, then the product $g\circ f$ corresponds to the product BA. (For left modules, we can observe transpositions.) If $a'_1,\ldots,a'_m$ and $b'_1,\ldots,b'_n$ are another pair of bases, then there are regular matrices P, Q such that $(a'_1,\ldots,a'_m)=(a_1,\ldots,a_m)P$ and $(b'_1,\ldots,b'_n)=(b_1,\ldots,b_n)Q$. With respect to these new bases, the corresponding matrix to f is $Q^{-1}AP$. In the particular case where $L=L'$, $a_i=b_i$ (for all i), we see that (1) under a fixed basis of L, a linear transformation of L is nothing but a matrix of degree m and (2) a change of basis corresponds to the transformation $A\mapsto P^{-1}AP$ by the base-change matrix P.

Thus, notions applicable to matrices can be applied to linear mappings. For instance, an **eigenvector (characteristic vector** or **proper vector)** of a linear transformation φ of L is an element $a(\neq 0)$ of L such that $\varphi(a)=a\alpha$ $(\alpha\in K)$. If K is a field, then the **characteristic polynomial** $\chi(X)$, and therefore also the **eigenvalues (proper values** or **characteristic roots)**, are invariant under base change. We add a little more on the case where K is a field. For an eigenvalue α of φ, the subspace $N_\alpha=\{a\in L\mid\varphi(a)=\alpha a\}$ of L is called the **eigenspace** belonging to α. Furthermore, the space $N'_\alpha=\{a\in L\mid(\varphi-\alpha)^k(a)=0$ for some $k>0\}$ is a subspace of L containing N_α and is sometimes called an **eigenspace in a weaker sense**. If all roots of $\chi(X)=0$ are in K, then L is decomposed into the direct sum of $N'_{\alpha_1},\ldots,N'_{\alpha_s}$, where $\alpha_1,\ldots,\alpha_s$ are the distinct roots of the equation $\chi(X)=0$. The dimension of N'_{α_i} is equal to the multiplicity of the root α_i in the equation $\chi(X)=0$. This fact is the basis of the Jordan normal form. **Minimal polynomials** of linear mappings are well defined.

A linear transformation φ of L is called **semisimple** if L has the structure of a †semi-

simple $K[X]$-module determined by $Xa = \varphi(a)$. Hence φ is semisimple if and only if the minimal polynomial $\mu(X)$ of φ has no square factor different from constants in $K[X]$. In particular, the condition that all roots of $\mu(X) = 0$ are in K and simple is sufficient for semisimplicity of φ. This condition is equivalent to the condition that φ is represented by a †diagonal matrix relative to some basis of L. Then L is decomposed into the direct sum of the eigenspaces N_α, and φ is said to be **diagonalizable**.

If K is a †perfect field, any linear transformation φ of L is represented as the sum of a semisimple linear transformation φ_s and a nilpotent linear transformation φ_n: $\varphi = \varphi_s + \varphi_n$ (**Jordan decomposition**). Also, φ_s and φ_n commute with each other, and they are uniquely determined by φ. We call φ_s and φ_n the **semisimple** and **nilpotent component** of φ, respectively. Furthermore, φ_s and φ_n can be represented as polynomials of φ without a constant term. The transformation φ is nonsingular if and only if φ_s is nonsingular. A nonsingular linear transformation φ is called **unipotent** if φ_s is equal to the identity transformation 1_L. Any nonsingular linear transformation φ is uniquely represented as a product of a semisimple linear transformation and a unipotent linear transformation, which are commutative: $\varphi = \varphi_s \varphi_u$ (**multiplicative Jordan decomposition**). Here φ_s is the semisimple component and $\varphi_u = 1_L + \varphi_s^{-1}\varphi_n$ is unipotent (φ_u is called the **unipotent component** of φ; → 13 Algebraic Groups).

M. Linear Equations

Let K be a †field, $f_i = a_{i1}X_1 + \ldots + a_{in}X_n$ ($i = 1, \ldots, m$) be m †linear forms: $K^n \to K$, and $b_1, \ldots, b_m$ be m given elements of K. Then a set of n elements $x_1, \ldots, x_n$ of K, or an n-tuple $\mathbf{x} = (x_1, \ldots, x_n) \in K^n$, satisfying the **system of m linear equations**

$$a_{i1}x_1 + \ldots + a_{in}x_n = b_n, \quad i = 1, \ldots, m, \tag{1}$$

is called a **solution** of equations (1). In particular, a system with $b_1 = \ldots = b_m = 0$, i.e.,

$$a_{i1}x_1 + \ldots + a_{in}x_n = 0, \quad i = 1, \ldots, m, \tag{2}$$

is called a **system of linear homogeneous equations**. In the theory of linear equations, the field K need not be commutative. In those cases where K is noncommutative, we have to distinguish between right multiplication and left multiplication. We make this distinction by adding the word "right" or "left" in parentheses whenever it is necessary to do so.

If $\mathbf{x}_1, \ldots, \mathbf{x}_t$ are solutions of the system (2), then any of their (right) linear combinations $\sum_i \mathbf{x}_i c_i$ ($c_i \in K$) is also a solution of (2), so that the solutions of (2) form a †(right) linear space L over K. L is the kernel of the (left) linear mapping $K^n \to K^m$ given by $\mathbf{x} \to (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x}))$. Let $s = \dim L$. If $s = 0$, the system (2) has only $\mathbf{x} = 0$ as a solution, called the **trivial solution**. If $s > 0$, L has a basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_s\}$. Then $\mathbf{x}_1, \ldots, \mathbf{x}_s$ are (right) linearly independent solutions of (2), and every solution of (2) is a (right) linear combination of them. We say that $\mathbf{x}_1, \ldots, \mathbf{x}_s$ form a **system of fundamental solutions** of (2). Let r denote the maximum number of (left) linearly independent forms in the set $\{f_1, f_2, \ldots, f_m\}$. Then we have $s + r = n$, so that the system (2) has a nontrivial solution if and only if $r < n$, and the number of linearly independent fundamental solutions is $n - r$. Accordingly, if the number of equations is less than the number of unknown quantities, there always exists a nontrivial solution.

Suppose now that the system (1) has a solution $\mathbf{x}_0$. Then every solution $\mathbf{x}$ of (1) can be written $\mathbf{x}_0 + \mathbf{y}$ for a solution $\mathbf{y}$ of (2); and conversely, for any solution $\mathbf{y}$ of (2), $\mathbf{x}_0 + \mathbf{y}$ is a solution of (1). Furthermore, in order for (1) to have a solution, it is necessary and sufficient that $\sum_i c_i b_i = 0$ whenever $\sum_i c_i f_i = 0$ ($c_i \in K$), i.e., that the following two matrices have the same †rank:

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ & \cdots & \\ a_{m1} & \cdots & a_{mn} \end{pmatrix},$$

$$\tilde{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} & b_1 \\ & \cdots & & \\ a_{m1} & \cdots & a_{mn} & b_m \end{pmatrix}.$$

In particular, equation (1) has a unique solution if and only if A and $\tilde{A}$ have the same rank n. If $m = n$ (i.e., A is a square matrix), (1) has a unique solution if and only if A has the inverse A^{-1}, and then the solution is given by ${}^t\mathbf{x} = A^{-1}\,{}^t\mathbf{b}$, where $\mathbf{b} = (b_1, b_2, \ldots, b_n)$.

We have also the following result. Let K' be an †extension field of K. If a system of linear equations in K has a solution in K', it has a solution contained in K. In particular, if (2) in K has a nontrivial solution in K', it has a nontrivial solution already in K, and any system of fundamental solutions in K is itself a system of fundamental solutions in K'.

Suppose now that K is commutative. Then we have an explicit formula for solving equation (1) by means of †determinants.

Consider first the case $m = n$. Let $\Delta = |A|$, the determinant of the matrix A. If $\Delta \neq 0$, then (1) has a unique solution, given by $x_k = \Delta_k / \Delta$ ($k = 1, \ldots, n$), where Δ_k is the determinant obtained from Δ by replacing its kth column,

$a_{1k}, a_{2k}, \ldots, a_{nk}$ by $b_1, b_2, \ldots, b_n$. This is called **Cramer's rule**. To consider the general case, let r be the common rank of A and $\tilde{A}$. We may assume that $|a_{ik}| \neq 0$ $(i, k=1, \ldots, r)$, appropriately changing the order of the equations and unknowns. Then by Cramer's rule we can solve the first r equations for $x_1, \ldots, x_r$, assigning arbitrary values to $x_{r+1}, \ldots, x_n$. In order for equation (2) to have a nontrivial solution, it is necessary and sufficient that the rank of A be less than n, and hence that $|A|=0$ if A is a square matrix. For geometric applications of the theory of linear equations → 7 Affine Geometry, and for numerical solution → 302 Numerical Solution of Linear Equations.

N. Positive Matrices

A **positive** (or **nonnegative**) matrix is a matrix whose elements are all positive (or nonnegative) real numbers. A nonnegative square matrix A is said to be **reducible** (or, more precisely, reducible by permutations) if by certain permutations of rows and columns A is reduced to a form

$$\begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}$$

with square matrices A_1, A_2; otherwise A is called **irreducible**.

Theorem (Perron). For each positive matrix A, there is a positive real number r such that (1) r is a simple root of the characteristic equation for A and (2) every other eigenvalue of A has absolute value less than r. Furthermore, (3) if v is an eigenvector corresponding to r, then the components of v are all positive or all negative.

Theorem (Frobenius). If A is an irreducible nonnegative matrix, then there is a positive number r satisfying (1) above and such that every eigenvalue of A has absolute value at most r. Furthermore, (3) above also holds in this case.

Theorem. If A is a nonnegative matrix, then there is a nonnegative real number r that is an eigenvalue and such that absolute values of other eigenvalues are at most r. If v is an eigenvector corresponding to this r, then the components of v are all nonnegative or all nonpositive.

References

See references to 103 Determinants.

270 (X.12) Measure Theory

A. General Remarks

Roughly speaking, a measure is a nonnegative function of subsets of a space completely additive in the sense that the measure of the union of a sequence of mutually disjoint sets is the sum of the measures of the sets. This concept was introduced as a mathematical abstraction of length, area, volume, mass distribution, probability distribution, etc. Measure theory is the basis of integration theory, and these two theories play a fundamental role in modern mathematics, particularly in analysis, functional analysis, and probability theory.

B. Important Classes of Sets

Let X be an abstract space. The power set of X, the class of all subsets of X, is denoted by 2^X. Let $\mathfrak{A}$ be a nonempty subclass of 2^X. If $\mathfrak{A}$ is closed under intersections, i.e., if $A, B \in \mathfrak{A}$ implies $A \cap B \in \mathfrak{A}$, then $\mathfrak{A}$ is called a **multiplicative class** on X. If $\mathfrak{A}$ is closed under finite unions and complements, then $\mathfrak{A}$ is called an **algebra** (or **finitely additive class** or **field**) on X. If $\mathfrak{A}$ is closed under countable unions and complements, $\mathfrak{A}$ is called a **σ-algebra** (**countably additive class, completely additive class**, or **Borel field**). If $\mathfrak{A}$ is closed under montone limits, i.e., if for every sequence $A_n \in \mathfrak{A}$, $n=1, 2, \ldots$, monotone increasing or monotone decreasing, the union or intersection of the sequence belongs to $\mathfrak{A}$, then $\mathfrak{A}$ is called a **monotone class** on X. If $X \in \mathfrak{A}$ and if $\mathfrak{A}$ is closed under countable disjoint unions and proper differences ($A \setminus B$ for $A \supset B$), then $\mathfrak{A}$ is called the **Dynkin class** on X. For any $\mathfrak{C} \subset 2^X$ the least σ-algebra that includes $\mathfrak{C}$ is called the σ-algebra generated by $\mathfrak{C}$; this is written as $\sigma[\mathfrak{C}]$. Similarly, we denote by $\mathfrak{M}[\mathfrak{C}]$ $(\mathfrak{D}[\mathfrak{C}])$ the monotone class (Dynkin class) generated by $\mathfrak{C}$. The following theorems are often useful.
The monotone class theorem: If $\mathfrak{C}$ is an algebra, then $\mathfrak{M}[\mathfrak{C}]=\sigma[\mathfrak{C}]$.
The Dynkin class theorem: If $\mathfrak{C}$ is a multiplicative class, then $\mathfrak{D}[\mathfrak{C}]=\sigma[\mathfrak{C}]$.

C. Measurable Spaces (Borel Spaces)

A space X endowed with a σ-algebra $\mathfrak{B}$ on X is called a **measurable space** (or **Borel space**) and is denoted by $(X, \mathfrak{B})$. A set belonging to $\mathfrak{B}$ is called a **$\mathfrak{B}$-measurable set** (or **Borel set**) in $(X, \mathfrak{B})$. It is obvious that the empty set $\varnothing$ and

the whole space X are $\mathfrak{B}$-measurable. $\mathfrak{B}$-measurability is preserved by complements, namely, if A is $\mathfrak{B}$-measurable, then so is A^c. Similarly $\mathfrak{B}$-measurability is preserved by countable set operations such as countable unions, countable intersections, differences, etc., but not always by uncountable unions.

For a sequence $\{A_n\}$ $(n \in \mathbf{N})$ of subsets of X, the **superior limit** (or **limit superior**) and the **inferior limit** (or **limit inferior**) are defined by $\limsup_{n\to\infty} A_n = \bigcap_{m=1}^{\infty}(\bigcup_{n=m}^{\infty} A_n)$ and $\liminf_{n\to\infty} A_n = \bigcup_{m=1}^{\infty}(\bigcap_{n=m}^{\infty} A_n)$, respectively. The symbols $\overline{\lim}$ and $\underline{\lim}$ may be used in place of lim sup and lim inf, respectively. The inferior limit is always contained in the superior limit, and when they coincide, they define the **limit** of $\{A_n\}$, which is denoted by $\lim_{n\to\infty} A_n$. If $\{A_n\}$ is monotone increasing (or decreasing), $\lim_{n\to\infty} A_n$ is $\bigcup_{n=1}^{\infty} A_n$ (or $\bigcap_{n=1}^{\infty} A_n$) (→ 87 Convergence L). $\mathfrak{B}$-measurability is preserved by these limits.

Let Y be a subset of a measurable space $(X, \mathfrak{B})$. Then the class $\mathfrak{B} \cap Y = \{B \cap Y \mid B \in \mathfrak{B}\}$ is a σ-algebra on Y. Y is regarded as a measurable space endowed with $\mathfrak{B} \cap Y$ unless stated otherwise. Suppose that we are given a family of measurable spaces $(X_\lambda, \mathfrak{B}_\lambda)$, $\lambda \in \Lambda$, where Λ is an arbitrary index set. Let X be the †Cartesian product of X_λ, $\lambda \in \Lambda$, and π_λ the projection from X to X_λ for each $\lambda \in \Lambda$. Then X is regarded as a measurable space endowed with the σ-algebra generated by the class $\{\pi_\lambda^{-1}(B_\lambda) \mid \lambda \in \Lambda, B_\lambda \in \mathfrak{B}_\lambda\}$, denoted by $\prod_{\lambda\in\Lambda} \mathfrak{B}_\lambda$, unless stated otherwise.

Let T be a topological space. The σ-algebra on T generated by the open subsets of T is called the **topological σ-algebra (Borel field)** on T, denoted by $\mathfrak{B}(T)$. Some authors define $\mathfrak{B}(T)$ to be the σ-algebra generated by compact subsets of T, which is equivalent to the definition mentioned above if T is locally compact and σ-compact. T is regarded as a measurable space endowed with $\mathfrak{B}(T)$ unless stated otherwise. Hence $\mathbf{R}^n$ is a measurable space $(\mathbf{R}^n, \mathfrak{B}^n)$, where $\mathfrak{B}^n = \mathfrak{B}(\mathbf{R}^n)$. A $\mathfrak{B}(T)$-measurable subset of T is called a **Borel subset** of T. If the †characteristic function of a set is a †Baire function, then the set is a Borel set. Such a set is called a **Borel set in the strict sense** (or a **Baire set**). The union (intersection) of at most a countable number of closed (open) sets is called an **F_σ set** (**G_δ set**). Both F_σ sets and G_δ sets are Borel sets. Denote by $\mathfrak{F}_0$ the class of all closed subsets of X and by $\mathfrak{G}_0$ the class of all open subsets, and define classes $\mathfrak{F}_\xi$ ($\mathfrak{G}_\xi$) for an ordinal number ξ to be the sets expressible as a countable union (intersection) of the sets belonging to $\bigcup_{\eta<\xi} \mathfrak{F}_\eta$ ($\bigcup_{\eta<\xi} \mathfrak{G}_\eta$). Then by the construction, every set belonging to either $\mathfrak{F}_\xi$ or $\mathfrak{G}_\xi$ is a Borel set. If $\omega_1 < \xi$, then $\mathfrak{F}_\xi = \mathfrak{F}_{\omega_1}$, $\mathfrak{G}_\xi = \mathfrak{G}_{\omega_1}$. (For the definition of ω_1 → 312 Ordinal Numbers.) In particular, in a †completely normal topological space (for example, a metric space), $\mathfrak{F}_\alpha \subset \mathfrak{G}_\beta$, $\mathfrak{G}_\alpha \subset \mathfrak{F}_\beta$ if $\alpha < \beta$, and both $\bigcup_{\xi<\omega_1} \mathfrak{F}_\xi$ and $\bigcup_{\xi<\omega_1} \mathfrak{G}_\xi$ coincide with the Borel field. In this case, an arbitrary Borel set B is a Baire set, and thus can be classified according to the †Baire class of the characteristic function of B. It can also be classified according to the smallest ordinal number ξ with respect to which $B \in \mathfrak{F}_\xi$ (or $B \in \mathfrak{G}_\xi$).

If S is a subset of a topological space T, then S is regarded as a topological space with the relative topology and so is regarded as a measurable space $(S, \mathfrak{B}(S))$. Hence every subset of $\mathbf{R}^n$ is regarded as a measurable space. S is also a measurable space $(S, \mathfrak{B}(T) \cap S)$, as mentioned above. Since $\mathfrak{B}(S) = \mathfrak{B}(T) \cap S$, no contradiction arises. Let T_λ, $\lambda \in \Lambda$ be a family of topological spaces. Then the topological product T of this family is regarded as a measurable space $(T, \mathfrak{B}(T))$ or $(T, \prod_{\lambda\in\Lambda} \mathfrak{B}(T_\lambda))$. It should be noted that these two σ-algebras may be different. They are the same if T has a countable open base.

Let $(X_i, \mathfrak{B}_i)$, $i = 1, 2$, be measurable spaces. A bijective mapping f from X_1 to X_2 is called a **Borel isomorphic mapping** if $f(\mathfrak{B}_1) \equiv \{f(B_1) \mid B_1 \in \mathfrak{B}_1\} = \mathfrak{B}_2$. If there exists such an f, then $(X_2, \mathfrak{B}_2)$ is said to be **Borel isomorphic** to $(X_1, \mathfrak{B}_1)$. Borel isomorphism is an equivalence relation. A measurable space is called a **standard measurable space** (or a **standard Borel space**) if it is Borel isomorphic to a Borel subset of $\mathbf{R}^1$ (viewed as a measurable space as explained above). A striking fact is that a standard measurable space is Borel isomorphic to one of the following spaces: $\{1, 2, \ldots, n\}$ $(n = 1, 2, \ldots)$, $\mathbf{N}$, $[0, 1]$. A measurable space which is Borel isomorphic to an †analytic subset of $\mathbf{R}^1$ is called an **analytic measurable space**. Every complete separable metric space, viewed as a measurable space, is standard. More generally, every †Luzin space is a standard measurable space. Hence the spaces of †distributions, $\mathscr{D}'$ and $\mathscr{S}'$, are standard measurable spaces, being Luzin spaces. Every †Suslin space, viewed as a measurable space, is analytic.

D. Measure

A real-valued set function m defined on a finitely additive class $\mathfrak{M}$ on a space X is called a **finitely additive measure** (or **Jordan measure**) if it satisfies the following two conditions: (I) $0 \leqslant m(E) \leqslant \infty$, $m(\varnothing) = 0$; (II$_0$) $A, B \in \mathfrak{M}$, $A \cap B = \varnothing$ imply $m(A \cup B) = m(A) + m(B)$.

A set function μ defined on a completely additive class $\mathfrak{B}$ in a space X is called a **measure (completely additive measure** or **σ-additive**

measure) on $\mathfrak{B}$ (or on X) if it satisfies the following two conditions: (I) $0 \leqslant \mu(E) \leqslant \infty$, $\mu(\varnothing) = 0$; (II) $E_n \in \mathfrak{B}$ $(n = 1, 2, \ldots)$, $E_j \cap E_k = \varnothing$ $(j \neq k)$ imply $\mu(\bigcup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} \mu(E_n)$ (**complete additivity** or **σ-additivity** or **countable additivity**). The triple $(X, \mathfrak{B}, \mu)$ or the pair (X, μ), where X is a space, $\mathfrak{B}$ a completely additive class of subsets of X, and μ a measure, is called a **measure space**. We call μ or $(X, \mathfrak{B}, \mu)$ **finite** or **bounded** if $\mu(X) < \infty$ and **σ-finite** if there exists a sequence $\{X_n\}$ satisfying $\mu(X_n) < \infty$ and $\bigcup_{n=1}^{\infty} X_n = X$. A set $A \in \mathfrak{B}$ is called **atomic** if (i) $\mu(A) > 0$ and (ii) $B \in \mathfrak{B}$ and $B \subset A$ imply that $\mu(B) = 0$ or $\mu(B) = \mu(A)$. A measure space $(X, \mathfrak{B}, \mu)$ is called **nonatomic** if no element of $\mathfrak{B}$ is atomic. The simplest bounded measure is given by defining $m(A) = 1$ if $a \in A$ and $m(A) = 0$ if $a \notin A$, where a is a given point in X. Such a measure is called the Dirac **δ-measure**. A measure m with $m(X) = 1$ is called a †probability measure (→ 342 Probability Theory).

Equalities appearing in (II_0) and (II) include the possibility that both sides equal ∞. In measure theory and integration theory, addition and multiplication involving $\pm\infty$ are carried out as follows (a denotes a real number): $(\pm\infty) + a = a + (\pm\infty) = \pm\infty$; $(\pm\infty) - a = \pm\infty$, $a - (\pm\infty) = \mp\infty$; $(\pm\infty) + (\pm\infty) = \pm\infty$; $(\pm\infty) - (\mp\infty) = \pm\infty$; $a \cdot (\pm\infty) = (\pm\infty) \cdot a = \pm\infty$ if $a > 0$; $a \cdot (\pm\infty) = (\pm\infty) \cdot a = \mp\infty$ if $a < 0$; $(+\infty) \cdot (+\infty) = +\infty$. (Sometimes it is further agreed to put $0 \cdot (\pm\infty) = (\pm\infty) \cdot 0 = 0$.)

A $\mathfrak{B}$-measurable set in a measure space $(X, \mathfrak{B}, \mu)$ is also called **μ-measurable** (or simply **measurable**). For a sequence $\{A_n\}$ of μ-measurable sets the following conditions hold: (i) $\mu(\liminf_{n\to\infty} A_n) \leqslant \liminf_{n\to\infty} \mu(A_n)$; (ii) if $\mu(\bigcup_{n=n_0}^{\infty} A_n) < +\infty$ for some n_0, then $\mu(\limsup_{n\to\infty} A_n) \geqslant \limsup_{n\to\infty} \mu(A_n)$; and (iii) if $\lim_{n\to\infty} A_n$ exists and the hypothesis of (ii) holds, then $\mu(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} \mu(A_n)$. For a finitely additive measure μ on a completely additive class $\mathfrak{B}$ to be a completely additive space $(X, \mathfrak{B}, \mu)$. If the set of all points for which a property **P** fails to hold is a null set, then we say that **P** holds **almost everywhere (a.e.** or **at almost all points)**. Sometimes we use the expression "almost **P**" to describe the same situation.

E. Construction of Measures

A set function $\mu^*(A)$ defined for every subset of X is called a **Carathéodory outer measure** (or **outer measure**) on X if it has the following three properties: (i) $0 \leqslant \mu^*(A) \leqslant \infty$, $\mu^*(\varnothing) = 0$; (ii) $A \subset B \Rightarrow \mu^*(A) \leqslant \mu^*(B)$; (iii) $\mu^*(\bigcup_{n=1}^{\infty} A_n) \leqslant \sum_{n=1}^{\infty} \mu^*(A_n)$. Because of (iii), the inequality $\mu^*(B) \leqslant \mu^*(B \cap A) + \mu^*(B \cap A^c)$ always holds; if the equality holds for every B, then A is called **measurable with respect to** μ^*. It follows from this definition that a set A satisfying $\mu^*(A) = 0$ is always measurable with respect to μ^*. The class $\mathfrak{B}$ of all sets measurable with respect to μ^* is a completely additive class, and if we define $\mu(A) = \mu^*(A)$ for A belonging to $\mathfrak{B}$, then $\mu(A)$ gives a complete measure on $\mathfrak{B}$. This measure is called the **Carathéodory measure** induced by μ^* or the **generalized Lebesgue measure**.

A finitely additive measure m defined on a finitely additive class $\mathfrak{M}$ is called **completely additive** when $A_n \in \mathfrak{M}$, $A_j \cap A_k = \varnothing$ $(j \neq k)$, $\bigcup_{n=1}^{\infty} A_n \in \mathfrak{M}$ imply $m(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} m(A_n)$. If by means of such an m we define $\mu^*(A)$ for an arbitrary $A \subset X$ to be the infimum of all possible values $\sum_{n=1}^{\infty} m(A_n)$, where $A \subset \bigcup_{n=1}^{\infty} A_n$ $(A_n \in \mathfrak{M})$, then μ^* gives a Carathéodory outer measure. From this μ^* a measure μ on a completely additive class $\mathfrak{B}$ is induced as described above, and we have $\mathfrak{B} \supset \mathfrak{M}$ and $\mu(A) = m(A)$ for $A \in \mathfrak{M}$. Therefore m can be extended to a measure μ on $\sigma[\mathfrak{M}]$ (**E. Hopf's extension theorem**). In particular, if X is a countable union of sets of finite measure, then this extension is unique.

$x_k \leqslant b_k, k=1,2,\ldots,n\}$ $(-\infty \leqslant a_k < b_k \leqslant \infty)$ is defined by $m(I)=\prod_{k=1}^{n}(b_k-a_k)$. Let $\mathfrak{M}_0$ be the collection of all sets that can be represented as the finite union of disjoint left-open intervals, and for such expression $A=\bigcup_{j=1}^{r} I_j$ in $\mathfrak{M}_0$ define $m(A)=\sum_{j=1}^{r} m(I_j)$. Let $\mathfrak{M}=\mathfrak{M}_0 \cup \{\varnothing\}$ and $m(\varnothing)=0$. Then m gives a finitely additive measure on $\mathfrak{M}$ that is completely additive. Therefore m determines an outer measure μ^*, which in turn determines a measure μ. This μ^* is called the **Lebesgue outer measure** (or simply **outer measure**). Sets that are measurable with respect to μ^* are said to be **Lebesgue measurable** (or simply **measurable**), and the measure μ is called the (n-dimensional) **Lebesgue measure** (or simply **measure**). Every interval is measurable, and its measure coincides with its volume. Open sets, closed sets, and Borel sets are all measurable. More generally, suppose that an outer measure μ^* defined on a metric space X with a metric d satisfies the condition: If $d(A,B)>0$, then $\mu^*(A\cup B)=\mu^*(A)+\mu^*(B)$. Then every closed subset of X is μ^*-measurable, and therefore so is every Borel subset. Here $d(A,B)=\inf\{d(a,b)\mid a\in A, b\in B\}$ denotes the distance between the two sets A and B. A measure μ defined on the class of all Borel subsets of a topological space X is called a **Borel measure**. The cardinality of the set of all Lebesgue measurable subsets of $\mathbf{R}^n$ is $2^{\mathfrak{c}}$, while the cardinality of the class of Borel sets is $\mathfrak{c}$ (here $\mathfrak{c}$ is the cardinal number of the †continuum, i.e., the cardinal number of $\mathbf{R}$). Therefore there exists a Lebesgue measurable set that is not a Borel set. It follows immediately from the definition that the Lebesgue measure is K-regular, so for every Lebesgue measurable set A we can find an †F_σ-set B and a †G_δ-set C such that $B\subset A\subset C$ and $\mu(C-B)=0$.

Historically, for a bounded subset A of $\mathbf{R}^n$, C. Jordan defined $\overline{m}(A)$ to be the infimum of all possible values $m(B)$, where $B\in\mathfrak{M}$ and $B\supset A$, and $\underline{m}(A)$ to be supremum of all possible values $m(B)$, where $B\in\mathfrak{M}$ and $B\subset A$. He called $\overline{m}(A)$ the **outer volume** of A and $\underline{m}(A)$ the inner volume of A (in the case of $\mathbf{R}^2$, the **outer area** and **inner area**, respectively). When $\overline{m}(A)=\underline{m}(A)$, A is called **Jordan measurable**, and this common value is defined to be the **Jordan measure (Jordan content)** of A. Jordan measure is only finitely additive, and was found to be unsatisfactory in many respects. It was Lebesgue who modified this notion and introduced completely additive measures. Jordan measurable sets are always Lebesgue measurable.

Using the †axiom of choice, a set that is not Lebesgue measurable can be constructed (G. Vitali). For example, a set obtained by choosing exactly one element from each coset of the additive group of all rationals in the additive group of the reals is not measurable [3, pp. 67–70].

H. Product Measure

When two σ-finite complete measures μ_X and μ_Y are defined on completely additive classes $\mathfrak{B}_X$ and $\mathfrak{B}_Y$ on X and Y, respectively, an element C belonging to the smallest finitely additive class $\mathfrak{K}$ in the Cartesian product that contains $\{A\times B\mid A\in\mathfrak{B}_X, B\in\mathfrak{B}_Y\}$ can be represented as a finite disjoint union $C=\bigcup_{j=1}^{n}(A_j\times B_j)$ $(A_j\in\mathfrak{B}_X, B_j\in\mathfrak{B}_Y)$. If we define $v(C)=\sum_{j=1}^{n}\mu_X(A_j)\mu_Y(B_j)$ (here we agree to put $0\cdot\infty=0$), then this value is independent of the way the set C is represented, and v defines a completely additive measure on $\mathfrak{K}$. By extending v by means of Hopf's extension theorem, we obtain a (complete) measure space, called the **(complete) product measure space** obtained from the measure spaces $(X,\mathfrak{B}_X,\mu_X)$ and $(Y,\mathfrak{B}_Y,\mu_Y)$. The measure obtained in this way on the space $X\times Y$ is called the **product measure** of μ_X and μ_Y and is denoted by $\mu_X\times\mu_Y$. If we denote by $\mathfrak{M}_p$ the class of all Lebesgue measurable subsets of the p-dimensional Euclidean space $\mathbf{R}^p$ and by m_p the p-dimensional Lebesgue measure, then the (complete) product measure space of $(\mathbf{R}^p,\mathfrak{M}_p,m_p)$ and $(\mathbf{R}^q,\mathfrak{M}_q,m_q)$ is $(\mathbf{R}^{(p+q)},\mathfrak{M}_{p+q},m_{p+q})$. The product measure space of any finite number of measure spaces $(X_i,\mathfrak{B}_i,\mu_i)$ $(i=1,\ldots,n)$ is defined similarly.

Let X_λ $(\lambda\in\Lambda)$ be spaces with an index set of arbitrary cardinality. For the product space $X=\prod_{\lambda\in\Lambda}X_\lambda$, an **$n$-cylinder set**, or simply a **cylinder set**, is a set of the form $A\times\prod_{\lambda\notin\{\lambda_1,\ldots,\lambda_n\}}X_\lambda$ $(A\subset X_{\lambda_1}\times\ldots\times X_{\lambda_n})$. If a finitely additive class $\mathfrak{A}_\lambda$ is given for each X_λ, the class of all subsets that can be represented as the union of a finite number of cylinder sets of the form $A_1\times A_2\times\ldots\times A_n\times\prod_{\lambda\notin\{\lambda_1,\ldots,\lambda_n\}}X_\lambda$ $(A_j\in\mathfrak{A}_{\lambda_j}, j=1,2,\ldots,n)$ is a finitely additive class in the space X. When each $\mathfrak{A}_\lambda$ is completely additive, the completely additive class $\mathfrak{A}$ generated by this finitely additive class is called the **product** of the completely additive classes $\mathfrak{A}_\lambda$ and is denoted by $\prod_{\lambda\in\Lambda}\mathfrak{A}_\lambda$. When a measure space $(X_\lambda,\mathfrak{B}_\lambda,\mu_\lambda)$ with $\mu_\lambda(X_\lambda)=1$ is given for $\lambda\in\Lambda$, a measure μ can be defined in the following way on the completely additive class $\mathfrak{B}=\prod_{\lambda\in\Lambda}\mathfrak{B}_\lambda$ in the product space X: To begin with, for a cylinder set of the form $A_1\times\ldots\times A_n\times X'$ (here $A_j\in\mathfrak{B}_{\lambda_j}$ and $X'=\prod_{\lambda\notin\{\lambda_1,\ldots,\lambda_n\}}X_\lambda$), we define $\mu(A_1\times\ldots\times A_n\times X')=\mu_{\lambda_1}(A_1)\ldots\mu_{\lambda_n}(A_n)$. If we extend μ to the finitely additive class $\mathfrak{C}$ consisting of all sets that can be represented as the finite union of such cylinder sets, then this extension gives a completely additive measure on $\mathfrak{C}$, and therefore, by Hopf's extension theorem, there exists a unique extension to a mea-

sure μ on $\mathfrak{B}$, and μ satisfies $\mu(X)=1$. We denote this μ by $\mu=\prod_{\lambda\in\Lambda}\mu_\lambda$.

I. Radon Measure

Let X be a †locally compact Hausdorff space, $\mathfrak{B}$ the topological σ-algebra on X, and $C_0(X)$ the real linear space of all real-valued continuous functions f on X having †compact support (i.e., the closure of the set $\{x \mid f(x)\neq 0\}$ is compact). A (real) †linear functional φ defined on $C_0(X)$ is called a **positive Radon measure** if $\varphi(f)\geqq 0$ whenever $f\geqslant 0$. For such a functional φ there corresponds a Borel measure μ on X for which $\varphi(f)=\int_X f\,d\mu$ holds for every $f\in C_0(X)$. Lebesgue measure m_n is regarded as a positive Radon measure on $\mathbf{R}^n$. If X is σ-compact (i.e., X can be represented as the countable union of compact sets), then the property $\varphi(f)=\int_X f\,d\mu$ for all $f\in C_0(X)$ defines the measure μ uniquely on the class of Borel sets of X. A linear functional on $C_0(X)$ that can be written as the difference of two positive Radon measures is called a **Radon measure**. For a linear functional φ defined on $C_0(X)$ to be a Radon measure, it is necessary and sufficient that for an arbitrary $f\in C_0(X)$, the set $\{\varphi(g) \mid |g|\leqslant|f|, g\in C_0(X)\}$ be bounded. Equivalently, it is necessary and sufficient that the restriction of φ to the subspace of all functions in $C_0(X)$ having their support in a fixed compact subset of X must be continuous with respect to the †topology of uniform convergence. Therefore, if X is compact, an arbitrary continuous linear functional on $C(X)$ is a Radon measure. L. Schwartz investigated Radon measures on spaces that are not locally compact [6].

J. Measurable Functions

When a completely additive class $\mathfrak{B}$ on a space X is given, a function f defined on a $\mathfrak{B}$-measurable set E and taking real (and possibly $\pm\infty$) values is called a **$\mathfrak{B}$-measurable function** on E if for an arbitrary real number α, the set $\{x \mid f(x)>\alpha\}$ is $\mathfrak{B}$-measurable. The condtion $f>\alpha$ may be replaced by $f\geqslant\alpha$, $f<\alpha$, $f\leqslant\alpha$. A function f can also be defined to be $\mathfrak{B}$-measurable if the †inverse image under f of any Borel set is $\mathfrak{B}$-measurable. When f and g are $\mathfrak{B}$-measurable, so are $af+bg$ (a,b constants), $f\cdot g$, f/g, $\max(f,g)$, $\min(f,g)$, $|f|^p$ (p a constant), whenever they are well defined. The superior and inferior limits of a sequence of $\mathfrak{B}$-measurable functions are also $\mathfrak{B}$-measurable. In a complete product measure space of two σ-finite measure spaces, a function $f(x,y)$ may fail to be measurable as a function of two variables even if it is measurable with respect to each of the variables x and y separately. For example, let $\mathfrak{F}$ be the class of all closed subsets F of $\mathbf{R}^2$ satisfying $m_2(F)>0$, $F\subset[0,1]\times[0,1]$. Using the fact that the cardinality of $\mathfrak{F}$ does not exceed the cardinal number of the continuum, we can prove by †transfinite induction that the elements of $\mathfrak{F}$ can be indexed as F_ξ, $\bar{\bar{\xi}}<\mathfrak{c}$, $\bar{\bar{\xi}}$ being the cardinality of ξ, and that for every $F_\xi\in\mathfrak{F}$ we can pick two points $z_\xi=(x_\xi, y_\xi)$, $z'_\xi=(x'_\xi, y'_\xi)$ in such a way that if $F_\xi\neq F_\eta$, then x_ξ, x'_ξ, x_η, x'_η are all distinct and y_ξ, y'_ξ, y_η, y'_η are all distinct. Furthermore, we can prove that the set E consisting of all such z_ξ is not measurable. Therefore, denoting the characteristic function of the set E by $f(x,y)$, $f(x,y)$ is not measurable; but if we fix x (resp. y), then as a function of y (resp. x), $f(x,y)$ is measurable, since $f(x,\cdot)$ (resp. $f(\cdot,y)$) is always 0 except possibly at one point. If $\mathfrak{B}$ is the class of all Lebesgue measurable sets or the class of all Borel sets, then a $\mathfrak{B}$-measurable function is called a **Lebesgue measurable function** (or simply **measurable function**) or a **Borel measurable function**, respectively.

In Euclidean space, the class of all Borel measurable functions coincides with the class of all Baire functions, and an arbitrary Lebesgue measurable function is equal almost everywhere to a Baire function of at most the second class. For a function f that is finite almost everywhere on a Lebesgue measurable set E to be Lebesgue measurable, it is necessary and sufficient that for an arbitrary $\varepsilon>0$ we can find a closed subset F such that $m(E-F)<\varepsilon$ and f is continuous on F (**Luzin's theorem**).

If a sequence $\{f_n\}$ of $\mathfrak{B}$-measurable functions on a measure space $(X,\mathfrak{B},\mu)$ converges almost everywhere to f on a set E with $\mu(E)<\infty$, then for an arbitrary $\varepsilon>0$ we can find a set F ($F\subset E$, $F\in\mathfrak{B}$) such that $\mu(E-F)<\varepsilon$ and f_n converges uniformly on F. If $X=\mathbf{R}^n$ and $\mathfrak{B}$ is either the class of Borel sets or the class of Lebesgue measurable sets, then the set F can be chosen to be a closed set (**Egorov's theorem**).

For a finite measurable function $f(x)$ defined on the real line, there exists a sequence $\{h_n\}$ such that $\lim_{n\to\infty}f(x+h_n)=f(x)$ almost everywhere (H. Auerbach).

The functional equation $f(x+y)=f(x)+f(y)$ has infinitely many nonmeasurable solutions (G. Hamel; → 388 Special Functional Equations).

Let $\mathfrak{A}_i$, $i=1,2$, be σ-algebras on X_i, $i=1,2$, respectively. A mapping $f: X_1\to X_2$ is said to be measurable $\mathfrak{A}_1/\mathfrak{A}_2$ if $f^{-1}(A_2)\in\mathfrak{A}_1$ for every $A_2\in\mathfrak{A}_2$. When $\mathfrak{A}'_2$ generates $\mathfrak{A}_2$, $f: X_1\to X_2$ is measurable $\mathfrak{A}_1/\mathfrak{A}_2$ if $f^{-1}(A'_2)\in\mathfrak{A}_1$ for every $A'_2\in\mathfrak{A}'_2$. For example, a mapping $f:\mathbf{R}^1\to\mathbf{R}^1$ is Borel (Lebesgue) measurable if and only if

f is measurable $\mathfrak{B}^1/\mathfrak{B}^1(\mathfrak{M}_1/\mathfrak{B}^1)$, where $\mathfrak{B}^1$ and $\mathfrak{M}_1$ denote the Borel subsets and the Lebesgue measurable subsets of $\mathbf{R}^1$. Measurability of mappings is preserved by compositions, namely, if $f: X_1 \to X_2$ and $g: X_2 \to X_3$ are measurable $\mathfrak{A}_1/\mathfrak{A}_2$ and $\mathfrak{A}_2/\mathfrak{A}_3$, respectively, then $g \circ f: X_1 \to X_3$ is measurable $\mathfrak{A}_1/\mathfrak{A}_3$. If the projection $\pi_a: \prod_{i \in I} X_i \to X_a$ is measurable, then $\prod_{i \in I} \mathfrak{A}_i/\mathfrak{A}_a$. Since $\pi_i^{-1}(A_i)$ $(i \in I, A_i \in \mathfrak{A}_i)$ generate $\prod_{i \in I} \mathfrak{A}_i$, a mapping $f: X \to \prod_{i \in I} X_i$ is measurable $\mathfrak{A}/\prod_{i \in I} \mathfrak{A}_i$ if $\pi_i \circ f: X \to X_i$ is measurable $\mathfrak{A}_i/\mathfrak{A}$ for every $i \in I$, where $\mathfrak{A}$ is a σ-algebra on X.

K. Image Measures

Let μ be a regular measure on a measurable space $(X, \mathfrak{B}_x)$, i.e., the completion of a measure on $\mathfrak{B}_X$. Then every mapping $f: X \to Y$ measurable $\mathfrak{B}_X/\mathfrak{B}_Y$ induces a σ-algebra on Y:

$$\mathfrak{B} = \{B \subset Y \mid f^{-1}(B) \text{ is } \mu\text{-measurable}\},$$

and a measure on $\mathfrak{B}$:

$$\nu(B) = \mu(f^{-1}(B)).$$

The measure ν is called the **image measure** of μ under the mapping f; this is denoted by μf^{-1} or $f \cdot \mu$. Obviously, $\nu(Y) = \mu(X)$, and ν is complete. Although $\mathfrak{B}$ includes $\mathfrak{B}_Y$ by virtue of the measurability of f, ν is not regular in general. However, ν is regular if μ is a bounded measure and if both $(X, \mathfrak{B}_X)$ and $(Y, \mathfrak{B}_Y)$ are analytic measurable spaces (→ 22 Analytic Sets I). Therefore the image measure of a regular probability measure under a measurable mapping is also a regular probability measure in practically all useful cases.

L. Related Topics

(i) **Integration**. For a nonnegative measurable function $f(x)$ on $(X, \mathfrak{B}, m)$, we can define the integral of $f(x)$ on a set $E \in \mathfrak{B}$, denoted by $\int_E f(x)\mu(dx)$, $\int_E f(x)d\mu(x)$, $\int_E f\,d\mu$, or $\int_E f$. For a real measurable function $f(x)$, the integral is defined to be equal to $\int_E f^+\,d\mu - \int_E f^-\,d\mu$, if at least one of these integrals is finite, where f^+ and f^- are the positive and negative parts of f (→ 221 Integration Theory B).

(ii) **Fubini's theorem for measures**. Let $(X, \mathfrak{B}, \mu)$ be the complete product measure space of $(X_i, \mathfrak{B}_i, \mu_i)$ $(i = 1, 2)$, where both μ_1 and μ_2 are σ-finite and complete. For $E \in \mathfrak{B}$ we define the sections $E(x_1)$ and $E(x_2)$ by

$$E(x_1) = \{x_2 \mid (x_1, x_2) \in E\},$$

$$E(x_2) = \{x_1 \mid (x_1, x_2) \in E\}.$$

Then we have $E(x_1) \in \mathfrak{B}_2$ for almost all $(\mu_1) x_1 \in X_1$, $E(x_2) \in \mathfrak{B}_1$ for almost all $(\mu_2) x_2 \in X_2$, and

$$\mu(E) = \int_{X_1} \mu_2(E(x_1))\mu_1(dx_1)$$

$$= \int_{X_2} \mu_1(E(x_2))\mu_2(dx_2)$$

(→ 221 Integration Theory E).

(iii) **The Radon-Nikodym theorem and the Lebesgue decomposition theorem for measures.** Let μ and ν be σ-finite measures on $(X, \mathfrak{B})$. If $\mu(E) = 0$ implies $\nu(E) = 0$, then ν is said to be **absolutely continuous** with respect to μ, denoted by $\nu \prec \mu$. $\nu \prec \mu$ if and only if ν is expressible in the form $\nu(E) = \int_E f\,d\mu$, where f is a nonnegative $\mathfrak{B}$-measurable function (the **Radon-Nikodym theorem**). Such an f is uniquely determined up to μ-measure 0 and is called the **Radon-Nikodym derivative** of ν with respect to μ, denoted by $d\nu/d\mu$. The concept opposite to absolute continuity is singularity. ν is said to be **singular** with respect to μ if there exists an $E \in \mathfrak{B}$ such that $\nu(E^c) = \mu(E) = 0$. If μ and ν are two arbitrary σ-finite measures, then ν is expressible uniquely as the sum of an absolutely continuous measure ν_1 and a singular measure ν_2 with respect to μ (**Lebesgue decomposition theorem**) (→ 221 Integration Theory D, 380 Set Functions C).

(iv) **Invariant Measures**. Let $(X, \mathfrak{B})$ be a measurable space and G a group of Borel isomorphic mappings from $(X, \mathfrak{B})$ to itself. A measure μ on $(X, \mathfrak{B})$ is said to be **invariant** under G if $\mu(E) = \mu(g^{-1}(E))$ for every $E \in \mathfrak{B}$ and every $g \in G$. For example, the Lebesgue measure m on $(\mathbf{R}^n, \mathfrak{B}^n)$ is invariant under the group of congruent transformations (→ 225 Invariant Measures).

(v) **The Lebesgue-Stieltjes measure**. Let f bc a right continuous monotone increasing function on $\mathbf{R}$. Then there exists a unique measure μ on $(\mathbf{R}, \mathfrak{B}^1)$ such that $\mu((a, b]) = f(b) - f(a)$ for $a < b$. This measure is called the **Lebesgue-Stieltjes measure** induced by f (→ 166 Functions of Bounded Variation B).

(vi) **Baire measurable functions** and **universally measurable functions**. Let $f(x)$ be a real-valued function defined on a topological space X. If for every open set O the inverse image $f^{-1}(O)$ has the †Baire property (resp. is measurable with respect to the completion of any σ-infinite †Borel measure on X), then f is said to be **Baire measurable** (resp. **universally measurable** or **absolutely measurable**). Universally measurable functions can be defined similarly on a measurable space.

(vii) **Disintegration**. Let $(S, \mathfrak{S})$ and $(T, \mathfrak{T})$ be standard measurable spaces, ν a σ-finite measure on $(S, \mathfrak{S})$, and $p: S \to T$ a ν-measurable mapping. If the image measure $\mu = \nu f^{-1}$ is σ-finite, there exists a unique family of measures

$\{v_t, t \in T\}$ on $(S, \mathfrak{S})$ such that (1) $t \mapsto v_t(E)$ is universally measurable for every $E \in \mathfrak{S}$, (2) v_t is concentrated on $p^{-1}(t)$ for almost every t with respect to μ, and (3) the following equality holds:

$$v(E) = \int_T \mu(dt) v_t(E), \; E \in \mathfrak{S}.$$

This expression is called the **disintegration** of v with respect to μ. Every v-measurable function f on S is v_t-measurable for almost every t with respect to μ, and the integral of f is expressible in the form of iterated integrals as follows:

$$\int_S f(s) v(ds) = \int_T \mu(dt) \int_S f(s) v_t(ds).$$

This corresponds to Fubini's theorem on product measures. When v is a probability measure on $(S, \mathfrak{S})$, $p(s)$ is regarded as a $(T, \mathfrak{T})$-valued random variable on $(S, \mathfrak{S}, v)$, and v_t is the conditional probability measure under the condition $p(s) = t$. (→ 342 Probability Theory E. For disintegration of measures on a topological space → [3, ch. 9].)

References

[1] H. Lebesgue, Intégral, longueur, aire, Ann. Mat. Pura Appl., (3) 7 (1902), 231–359.
[2] S. Saks, Theory of the integral, Warsaw, 1937.
[3] P. R. Halmos, Measure theory, Van Nostrand, 1950.
[4] N. Bourbaki, Eléments de mathématique, Intégration Chap. 1–9, Actualités Sci. Ind., Hermann, 1965–1969.
[5] H. L. Royden, Real analysis, Macmillan, second edition, 1963.
[6] W. Rudin, Real and complex analysis, McGraw-Hill, second edition, 1974.
[7] L. Schwartz, Probabilités cylindriques et applications radonifiantes, J. Fac. Sci. Univ. Tokyo, IA, 18 (1971), 139–286.

271 (XX.4) Mechanics

A. Newton's Three Laws of Motion

The study of laws governing the motion of bodies began with the laws of falling bodies, of which the first exact formulation was made by Galileo. But the general relationship between force and acceleration was first described by †Newton, who established **Newton's three laws of motion**; the mechanics based on them is **Newtonian mechanics**. Newton expounded these laws in his famous book *Principia mathematica philosophiae naturalis* (1686–1687), where the law of gravitation and its application, problems of fluid motion, motions of the planets in the solar system, etc., were systematically treated.

Newton's first law. A body continues its state of rest or uniform motion in a straight line unless it is compelled to change that state by external action (i.e., force). This is also called the **law of inertia**.

Newton's second law. The rate of change of momentum is proportional to force and is in the direction in which the force acts. Here, **momentum** is defined as the product $m\mathbf{v}$ of the mass m and the velocity $\mathbf{v}$. This law can be expressed as $d(m\mathbf{v})/dt = \mathbf{F}$, where $\mathbf{F}$ is the force expressed in an appropriately chosen system of units. Since $d\mathbf{v}/dt$ is the acceleration $\mathbf{a}$, the law takes the form $m\mathbf{a} = \mathbf{F}$, when the mass is constant. These equations are called **equations of motion**. The second law is often simply called the **law of motion**.

Newton's third law. When two bodies 1 and 2 in the same system interact, the force exerted on body 1 by body 2 is equal and opposite in direction to that exerted on body 2 by body 1. This law is called the **law of action and reaction**, or simply the **law of reaction**.

Various attempts at a rigorous axiomatization of Newtonian mechanics have been made, beginning with E. Mach's work. At the beginning of the 20th century, it was found that Newtonian mechanics requires modification when bodies travel at speeds approaching the speed of light or when it is applied to physical systems of molecular size or smaller. These modifications led to the establishment of the theory of †relativity and †quantum mechanics. In contrast to these later theories, Newtonian mechanics is called **classical mechanics**.

B. Newton's Law of Gravitation

Kepler discovered the following three laws for the motion of planets around the sun (valid within the accuracy in observation available at the time):

Kepler's first law. The orbit of a planet is an ellipse with the sun at one of its foci.

Kepler's second law. The area swept per unit time by the straight line segment joining the planet and the sun is independent of the position of the planet in its orbit.

Kepler's third law. The square of the period (the time needed for the planet to go around the orbit once) is proportional to the cube of the major axis of the orbit.

From these empirical laws, Newton deduced his **law of universal gravitation**: Between any

pair of point particles, with masses m_1 and m_2 and at a distance r, there arises an attractive force along the line joining the two points; the magnitude of this force is given by $F = Gm_1m_2r^{-2}$, where G is a universal constant (approximately 6.670×10^{-8} dyne cm^2 g^{-2}) called the gravitational constant.

C. Kinetic and Potential Energies

If the force $\mathbf{F}$ on a particle is a function of the position $\mathbf{x}$ of the particle and is the gradient $-\nabla U$ of a time-independent (scalar) function $U(\mathbf{x})$, called the **potential**, then

$$E = \frac{1}{2}m\mathbf{v}^2 + U(\mathbf{x}) \qquad \left(\mathbf{v} = \frac{d\mathbf{x}}{dt}\right)$$

is a constant of the motion of the particle. E is called the **total energy**, while the first and second terms on the right-hand side are called the **kinetic energy** and **potential energy**, respectively.

For a system of n particles at points $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}$ with masses $m_1, \ldots, m_n$, suppose that the force acting on the particle at $\mathbf{x}^{(j)}$ is $-\nabla^{(j)}U$ (the gradient of U relative to $\mathbf{x}^{(j)}$) for a common potential function $U(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)})$. Then

$$E = \sum_{j=1}^{n} \frac{1}{2} m_j \mathbf{v}_j^2 + U(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}) \qquad \left(\mathbf{v}_j = \frac{d\mathbf{x}^{(j)}}{dt}\right)$$

is the constant total energy of the motion. For example, Newton's gravitational force acting among a number of particles can be described by the **Newtonian potential**

$$U = \sum_{i<j} -Gm_im_j|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}|^{-1}.$$

A potential U which is a sum of functions depending only on a pair of coordinates as in the above example is called a **two-body interaction**.

D. Apparent Force

The coordinate system in which Newton's three laws of motion hold is called an **inertial system**. In some cases (e.g., on a rotating sphere such as the earth) it is more convenient to use a moving coordinate system, in which the equation of motion derived from Newton's second law by a coordinate transformation (from an inertial system to the moving coordinate system) takes a form similar to the second law except for an additional **apparent force** to be added to the force in the original equation of motion. For a coordinate system rotating at a constant angular velocity ω (relative to an inertial system), the apparent force consists of the **centrifugal force**, of magnitude $m\omega^2\rho$ (ρ being the distance to the axis of rotation), which pushes the particle away from the axis of rotation (along the perpendicular), and the **Coriolis force** $2m\mathbf{v} \times \omega$ ($\times$ denotes the †vector product), which bends the motion of the particle in a direction perpendicular to both the axis of rotation and the velocity of the particle.

E. Dynamics of Rigid Bodies

A rigid body is defined as a system of particles whose mutual distances are permanently fixed. Like a point particle, it is an ideal concept introduced into mechanics to simplify the theoretical treatment. Actual solid bodies can in most cases be regarded as rigid under the action of forces of ordinary magnitudes. Since a rigid body can be imagined to be made up of an infinite number of particles, the equations of motion for a system of particles can also be applied to it. Thus the motion of a rigid body can be completely determined by the theorems of momentum and angular momentum.

The **momentum** of a rigid body is defined by

$$\mathbf{Q} = \int_K \frac{d\mathbf{r}}{dt} dm,$$

where dm is the mass of the volume element at a point $\mathbf{r}$ of the rigid body K and $d\mathbf{r}/dt$ is its velocity. If the external forces acting on K are denoted by $\mathbf{F}_i$ ($i = 1, 2, \ldots$), we have

$$d\mathbf{Q}/dt = \sum \mathbf{F}_i,$$

which expresses the **theorem of momentum**. If the velocity and acceleration of the **center of gravity (center of mass** or **barycenter)** $\int \mathbf{r}\, dm / \int dm$ of the rigid body are denoted by $\mathbf{V}_G$ and $\mathbf{A}_G$, respectively, and the **mass** $\int dm$ by m, we have $\mathbf{Q} = m\mathbf{V}_G$, and the theorem of momentum becomes $m\,d\mathbf{V}_G/dt = m\mathbf{A}_G = \sum \mathbf{F}_i$.

The **angular momentum** of a rigid body K about an arbitrary point $\mathbf{r}_0$ is defined by

$$\mathbf{H} = \int_K (\mathbf{r} - \mathbf{r}_0) \times \frac{d\mathbf{r}}{dt} dm.$$

If $\mathbf{P}_i$ is the vector from $\mathbf{r}_0$ to the point at which $\mathbf{F}_i$ acts, we have

$$d\mathbf{H}/dt = \sum (\mathbf{P}_i \times \mathbf{F}_i) = \mathbf{G}.$$

This is called the **theorem of angular momentum**.

For the case of a rigid body with one point $\mathbf{r}_0$ fixed, the angular momentum $\mathbf{H}$ and the angular velocity ω are related by

$$H_x = A\omega_x - F\omega_y - E\omega_z,$$

$$H_y = -F\omega_x + B\omega_y - D\omega_z,$$

$$H_z = -E\omega_x - D\omega_y + C\omega_z,$$

where H_x, H_y, H_z; $\omega_x, \omega_y, \omega_z$ are the components of $\mathbf{H}$ and ω in the xyz-coordinate system

fixed in space with its origin at the fixed point $\mathbf{r}_0$, and

$$A=\int(y^2+z^2)\,dm, \quad B=\int(z^2+x^2)\,dm,$$

$$C=\int(x^2+y^2)\,dm, \quad D=\int yz\,dm,$$

$$E=\int zx\,dm, \quad F=\int xy\,dm,$$

with the integrals taken over the whole rigid body. We call A, B, and C the **moments of inertia** about the x-, y-, and z-axes, respectively, and D, E, F the corresponding **products of inertia**. The rotational motion of a rigid body with one axis fixed is completely determined by the theorem of angular momentum. However, for rotation about a fixed point, the theorem is not very convenient to use, because A, B, C, D, E, and F are generally unknown functions of time.

The †quadric

$$Ax^2+By^2+Cz^2+2Dyz+2Ezx+2Fxy=1$$

represents an †ellipsoid with its center at the origin, called the **ellipsoid of inertia**. If the principal axes ξ, η, and ζ are taken as coordinate axes, the equation of the ellipsoid of inertia becomes $A\xi^2+B\eta^2+\Gamma\zeta^2=1$, where A, B, Γ are the moments of inertia about the ξ-, η-, ζ-axes and are called the **principal moments of inertia**, while the ξ-, η-, ζ-axes themselves are called the **principal axes of inertia**. If the components of the angular momentum $\mathbf{H}$ and angular velocity in the direction of the principal axes of inertia are denoted by (H_1, H_2, H_3) and $(\omega_1, \omega_2, \omega_3)$, respectively, then $H_1=A\omega_1$, $H_2=B\omega_2$, $H_3=\Gamma\omega_3$. Furthermore, if the ξ-, η-, ζ-components of the resultant moment of the external forces $\mathbf{G}=\sum(\mathbf{P}_i\times\mathbf{F}_i)$ are denoted by (G_1, G_2, G_3), then $d\mathbf{H}/dt=\mathbf{G}$ becomes

$$A\,d\omega_1/dt=G_1+(B-\Gamma)\omega_2\omega_3,$$

$$B\,d\omega_2/dt=G_2+(\Gamma-A)\omega_3\omega_1,$$

$$\Gamma\,d\omega_3/dt=G_3+(A-B)\omega_1\omega_2.$$

These are called **Euler's differential equations**.

The study of the motion of a rigid body can be reduced mostly to the study of the motion of its ellipsoid of inertia, since the latter is attached to the rigid body. The method of describing the motion of a rigid body by means of its ellipsoid of inertia is known as **Poinsot's representation**. The motions of two bodies having equal ellipsoids of inertia are the same if the external forces acting on them have equal resultant moments, even if their geometric forms are different.

When a rigid body moves under no constraint, the motion of its center of gravity G can be determined by the use of the theorem of momentum. Also, the rotational motion about the center of gravity can be found from $d\mathbf{H}'/dt=\mathbf{G}'$, which is a modification of the theorem of angular momentum. Here $\mathbf{H}'$ and $\mathbf{G}'$ are respectively the angular momentum and moment of external forces about the center of gravity. In this case also, simplification can be achieved by considering the equation in the reference system that coincides with the principal axes of the ellipsoid of inertia with its center at the center of gravity.

F. Analytical Dynamics

Mechanics as originally formulated by Newton was geometric in nature, but later L. Euler, J. L. Lagrange, and others developed the analytical method of treating mechanics that is now called **analytical dynamics**. Lagrange introduced **generalized coordinates** q_j $(j=1,2,\ldots,f$, where f is the number of **degrees of freedom** of the system considered), which uniquely represent the configuration of the dynamical system, and derived **Lagrange's equations of motion**:

$$\frac{d}{dt}\left(\frac{\partial\mathfrak{L}}{\partial\dot{q}_j}\right)-\frac{\partial\mathfrak{L}}{\partial q_j}=0, \qquad j=1,2,\ldots,f,$$

where $\dot{q}_j=dq_j/dt$, and $\mathfrak{L}=T-U$ (T=kinetic energy, U=potential energy) is a function of q_j and $\dot{q}_j$ called the **Lagrangian function**. Later, W. R. Hamilton introduced

$$p_j=\partial T/\partial\dot{q}_j,$$

$$H=\sum p_j\dot{q}_j-\mathfrak{L}=H(p_1,\ldots,p_f;q_1,\ldots,q_f)$$

and transformed the equations to **Hamilton's canonical equations**:

$$\frac{dq_j}{dt}=\frac{\partial H}{\partial p_j}, \qquad \frac{dp_j}{dt}=-\frac{\partial H}{\partial q_j}, \qquad j=1,2,\ldots,f.$$

Here p_j is the **generalized momentum** conjugate to q_j, and q_j, p_j are called **canonical variables**. If the functions representing the configuration of the dynamical system in terms of q_j do not explicitly contain the time t, the **Hamiltonian function** (or **Hamiltonian**) H coincides with the total energy of the system $T+U$.

The transformation $(p,q)\rightarrow(P,Q)$ under which canonical equations preserve their form is called a **canonical transformation**. It is given by

$$p_j=\frac{\partial W}{\partial q_j}, \qquad P_j=-\frac{\partial W}{\partial Q_j}, \qquad K=H+\frac{\partial W}{\partial t},$$

where $W=W(q_1,\ldots,q_f;Q_1,\ldots,Q_f)$ and K is the Hamiltonian of the transformed system. The set of canonical transformations forms a group, called a **group of canonical transforma-**

tions. An infinitesimal transformation is given by

$$dp_j = -\varepsilon\frac{\partial S}{\partial q_j}, \quad dq_j = \varepsilon\frac{\partial S}{\partial p_j},$$

where ε is an infinitesimal constant. Here S is an arbitrary function and is said to be the **generating function** of the infinitesimal transformation. Canonical equations can be interpreted to mean that the variations of p and q during the time interval $\varepsilon = dt$ are the infinitesimal canonical transformations whose generating function is $H(p,q,t)$.

The variation of an arbitrary function $F(p,q)$ under an infinitesimal transformation is given by

$$dF = \varepsilon(F,S),$$

where

$$(u,v) = \sum_j\left(\frac{\partial u}{\partial q_j}\frac{\partial v}{\partial p_j} - \frac{\partial u}{\partial p_j}\frac{\partial v}{\partial q_j}\right) = \sum_j\frac{\partial(u,v)}{\partial(q_j,p_j)}$$

is **Poisson's bracket**. Therefore the time rate of change of a dynamical quantity $F(p,q)$ can be written as

$$dF/dt = (F,H).$$

Thus the function $F(p,q)$ that satisfies $(F,H)=0$ is an †integral of the canonical equations.

If a canonical transformation $(p,q)\rightarrow(P,Q)$ such that $P_j=\alpha_j$, $Q_j=\beta_j$ are constant is found, the motion of the system can be determined by

$$p_j = \frac{\partial W}{\partial q_j}, \quad \beta_j = \frac{\partial W}{\partial \alpha_j},$$

where W is the †complete solution of the **Hamilton-Jacobi differential equation**:

$$\frac{\partial W}{\partial t} + H\left(\frac{\partial W}{\partial q_1}, \ldots, \frac{\partial W}{\partial q_f}; \quad q_1, \ldots, q_f, t\right) = 0$$

(→ 82 Contact Transformations).

G. Theory of Elasticity

(1) General remarks. Suppose that a solid body is deformed elastically by the action of external forces. We may inquire about the magnitudes of deformation, stress, and strain caused by the external forces at each point of the body. The **theory of elasticity** studies this problem, assuming the body to be a continuum and utilizing classical mechanics as a basis, and "endeavours to obtain results which shall be practically important in applications to architecture, engineering and all other useful arts in which the material of construction is solid" [1].

Cartesian coordinates (x,y,z) are employed for defining the 3-dimensional space containing the body. We confine ourselves to the small-displacement theory of elasticity, in which the components of the displacement $\mathbf{u} = (u,v,w)$ of an arbitrary point $P(x,y,z)$ of the body are assumed to be small enough to justify the linearization of the governing differential equations and boundary conditions.

(2) Stress. Consider an infinitesimal rectangular parallelepiped enclosed by the following six surfaces:

$$x=\text{const}, \qquad y=\text{const}, \qquad z=\text{const},$$
$$x+dx=\text{const}, \quad y+dy=\text{const}, \quad z+dz=\text{const},$$

in the body. The **stress** at the point $P(x,y,z)$ is defined as those internal forces per unit area acting on the six surfaces of the parallelepiped. It has nine components, which are usually represented by

$$\begin{pmatrix} \sigma_x & \tau_{yx} & \tau_{zx} \\ \tau_{xy} & \sigma_y & \tau_{zy} \\ \tau_{xz} & \tau_{yz} & \sigma_z \end{pmatrix},$$

where the σ and τ are called **normal** and **shearing** (or **tangential**) **stresses**, respectively. These nine quantities form a tensor called the **stress tensor**. Since it can be shown that $\tau_{xy}=\tau_{yx}$, $\tau_{yz}=\tau_{zy}$, $\tau_{zx}=\tau_{xz}$, the stress tensor is a symmetric tensor. By considering equilibrium conditions of the parallelepiped, the equations of equilibrium are found to be

$$\frac{\partial\sigma_x}{\partial x} + \frac{\partial\tau_{yx}}{\partial y} + \frac{\partial\tau_{zx}}{\partial z} + \bar{X} = 0, \ldots, \ldots,$$

where $\bar{X}, \ldots$ are **body forces** per unit volume.

(3) Strain. The infinitesimal rectangular parallelepiped fixed to the body at the point P before deformation is transformed, after deformation, into an infinitesimal parallelepiped which is no longer rectangular. The **strain** at the point P is defined as the changes caused by the deformation of the parallelepiped: extensions of the three sides and changes from right angles of the three angles formed by the three sides of the parallelepiped. Thus the strain has six components, usually represented by

$$(\varepsilon_x, \varepsilon_y, \varepsilon_z, \gamma_{yz}, \gamma_{zx}, \gamma_{xy}),$$

where ε and γ are called **elongation** and **shearing strains**, respectively. These six quantities, with slight modification, form a symmetric tensor called the **strain tensor**.

(4) Strain-displacement relations. In small-displacement theory, the strain-displacement relations are given in linear form by

$$\varepsilon_x = \frac{\partial u}{\partial x}, \quad \varepsilon_y = \frac{\partial v}{\partial y}, \quad \ldots, \quad \gamma_{xy} = \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}.$$

(5) Stress-strain relations. In the theory of elasticity, stress and strain are assumed to

obey a linear relationship called **Hooke's law**:

$$\{\sigma\}=[A]\{\varepsilon\},$$

where

$$\{\sigma\}^T=[\sigma_x,\sigma_y,\sigma_z,\tau_{yz},\tau_{zx},\tau_{xy}],$$

$$\{\varepsilon\}^T=[\varepsilon_x,\varepsilon_y,\varepsilon_z,\gamma_{yz},\gamma_{zx},\gamma_{xy}],$$

and $[A]$ is a symmetric positive definite matrix. For an isotropic body, the relations are

$$E\varepsilon_x=\sigma_x-\nu(\sigma_y+\sigma_z), \quad \ldots, \quad G\gamma_{xy}=\tau_{xy},$$

where E, ν, and $G=E/2(1+\nu)$ are elastic constants called the **modulus of elasticity in tension** or **Young's modulus, Poisson's ratio**, and the **modulus of elasticity in shear** or the **modulus of rigidity**, respectively.

(6) Boundary conditions. The surface of the body can be divided into two parts with regard to boundary conditions: the part S_1 over which the boundary conditions are prescribed in terms of external forces and the part S_2 over which the boundary conditions are prescribed in terms of displacements. Obviously $\partial V=S_1+S_2$, where ∂V is the whole surface of the body.

(7) Small-displacement theory of elasticity. We have seen that the equations which govern the problem are 3 equations of equilibrium, 6 strain-displacement relations, and 6 stress-strain relations in terms of 15 unknowns, namely, 6 stress components, 6 strain components, and 3 displacement components. Thus our problem is reduced to a boundary value problem in which these 15 field equations are to be solved under the specified boundary conditions. Since all the field equations and boundary conditions are linear with respect to the unknowns under the assumption that the displacements are small, we obtain linear relationships between the external load and resulting deformation of the body; this is the **small-displacement theory of elasticity**. It should be remembered, however, that the assumption of small displacement sets limits to the application of the theory to practical problems.

(8) Variational principles. Several **variational principles** have been formulated in the small-displacement theory of elasticity. These include the principle of minimum potential energy $[\mathbf{u}]$, the generalized principle $[\sigma,\varepsilon,\mathbf{u}]$, the Hellinger-Reissner principle $[\sigma,\mathbf{u}]$, the principle of minimum complementary energy $[\sigma]$, and so forth, where the symbols in the brackets represent independent functions subject to variation. In connection with the aforementioned variational principles, **variational principles with relaxed continuity requirements** have also been formulated by relaxing the continuity requirements imposed on admissible functions.

One of the practical advantages of these variational principles is that they often provide the problem with approximate formulations and approximate methods of solution, among which the **Rayleigh-Ritz method** is well known. Theories of beams, plates, shells, and multicomponent structures are typical examples of such approximate formulations. Recently, these variational principles have been found to provide effective bases for the formulations of the †finite element method.

(9) Notation. Various symbols are used for stress and strain. For example, σ, τ and ε, γ are commonly used in the engineering literature. However, in Love's treatise [6], X_x, Y_x, Z_x and e_{xx}, e_{xy}, e_{xz} are used in place of σ_x, τ_{xy}, τ_{xz} and ε_x, γ_{xy}, γ_{xz}, respectively. Also, various notations are used for elastic constants. E is widely used, while G and ν are less common. Love uses μ and σ for G and ν, respectively. In the engineering literature the reciprocal of Poisson's ratio m $(=1/\nu)$ is used and is called the **Poisson number**.

(10) Finite-displacement theory of elasticity. When the displacement of the body is no longer small (infinitesimal) but is finite, we should abandon small-displacement theory and instead employ **finite-displacement theory**, in which stress and strain are defined in a manner different from that in small-displacement theory, keeping in mind the difference between spatial and material variables. Thus equations of equilibrium, strain-displacement relations, and the boundary conditions on S_1 become nonlinear equations with stress and displacement components as unknowns, although the stress-strain relations remain linear. Thus the problem is reduced to solving a nonlinear boundary value problem, sometimes called a nonlinear elasticity problem. Variational principles have been formulated for finite-displacement theory and are frequently used in the formulation of approximate methods of solution.

When the stress becomes large enough to exceed the so-called **elastic limit**, where the linear stress-strain relationships cease to hold, the theory of elasticity is no longer valid and should be replaced by the **theory of plasticity** [9, 10].

References

[1] E. T. Whittaker, A treatise on the analytical dynamics of particles and rigid bodies, Cambridge Univ. Press, fourth edition, 1927.

[2] A. Sommerfeld, Mechanics, translated by M. O. Stern, Academic Press, 1952. (Original in German, 1943.)
[3] J. B. Marion, Classical dynamics of particles and systems, Academic Press, 1965.
[4] V. I. Arnold, Mathematical methods of classical mechanics, translated by K. Vogtmann and A. Weinstein, Springer, 1978. (Original in Russian, 1974.)
[5] R. Abraham and J. E. Marsden, Foundations of mechanics, Benjamin/Cummings, second edition, 1978.
[6] A. E. H. Love, A treatise on the mathematical theory of elasticity, Cambridge Univ. Press, fourth edition, 1934.
[7] S. Timoshenko and J. N. Goodier, Theory of elasticity, McGraw-Hill, 1951.
[8] A. E. Green and W. Zerna, Theoretical elasticity, Oxford Univ. Press, 1954.
[9] R. Hill, Mathematical theory of plasticity, Oxford Univ. Press, 1950.
[10] K. Washizu, Variational methods in elasticity and plasticity, Pergamon, second edition, 1975.

272 (XI.7) Meromorphic Functions

A. General Remarks

A single-valued †analytic function in a domain D in the complex plane $\mathbf{C}$ is called **meromorphic** in D if it has no singularities other than †poles. A function that is meromorphic in the whole complex plane including the point at infinity is a rational function (**Liouville's theorem**). Specifically, if a function is meromorphic in the domain $\mathbf{C}$, then the function is called simply a **meromorphic function**, and a meromorphic function that is not a rational function is called a **transcendental meromorphic function**. A meromorphic function $f(z)$ can be represented as a quotient of two †entire functions. Let $\{z_k\}$ $(k=1,2,\ldots)$ be poles of $f(z)$, and let $f_k(z)=a_{n_k}^{(k)}/(z-z_k)^{n_k}+\ldots+a_1^{(k)}/(z-z_k)$ denote the †singular parts of $f(z)$ at z_k $(k=1,2,\ldots)$. Then $f(z)$ can also be written in the form

$$f(z)=g(z)+\sum_{k=1}^{\infty}(f_k(z)-p_k(z)),$$

where $g(z)$ is an entire function and the $p_k(z)$ $(k=1,2,\ldots)$ are rational entire functions (**Weierstrass's theorem**). Assume that a sequence $\{z_k\}$ $(k=1,2,\ldots)$ converges only to the point at infinity and that $f_k(1/(z-z_k))$ $(k=1,2,\ldots)$ are rational entire functions of $1/(z-z_k)$ which have no constant terms. Then there exists a meromorphic function of z with $f_k(1/(z-z_k))$ as its †singular part at z_k (**Mittag-Leffler's theorem**).

B. Nevanlinna Theory

The theory of meromorphic functions can be considered an extension of the theory of entire functions. In particular, value distribution theory, originating in Picard's theorem, was studied by many people, and in 1925 R. Nevanlinna published a systematic theory unifying the results obtained until then. This is called **Nevanlinna theory**.

We let $f(z)$ denote a meromorphic function in $|z|<R\leqslant+\infty$, and when we say that $f(z)$ takes on a value, the value may be ∞. For a value α, $n(r,\alpha)$ denotes the number of **α-points** of $f(z)$, i.e., points z with $f(z)=\alpha$, in $|z|\leqslant r<R$, where each α-point is counted with its multiplicity. We set

$$N(r,\alpha)=\int_0^r\frac{n(t,\alpha)-n(0,\alpha)}{t}dt+n(0,\alpha)\log r,$$

$$m(r,\alpha)=\frac{1}{2\pi}\int_0^{2\pi}\log^+\left|\frac{1}{f(re^{i\theta})-\alpha}\right|d\theta$$

if $\alpha\neq\infty$, and

$$N(r,\infty)=\int_0^r\frac{n(t,\infty)-n(0,\infty)}{t}dt+n(0,\infty)\log r,$$

$$m(r,\infty)=\frac{1}{2\pi}\int_0^{2\pi}\log^+|f(re^{i\theta})|\,d\theta$$

if $\alpha=\infty$, where $\log^+a=\max(\log a,0)$ for $a>0$. The functions N and m are called the **counting function** and **proximity function** of $f(z)$, respectively, and $T(r)=T(r,f)=m(r,\infty)+N(r,\infty)$ is the **order function** (or **characteristic function**) of $f(z)$. $T(r)$ is an increasing function of r and a †convex function of $\log r$ and is useful for expressing $f(z)$ as an infinite product, etc.

The following relation holds among $T(r)$, $m(r,\alpha)$, and $N(r,\alpha)$ for any α:

$$T(r)=m(r,\alpha)+N(r,\alpha)+O(1), \tag{1}$$

where $O(1)$ is †Landau's symbol (**Nevanlinna's first fundamental theorem**). By this theorem, if a bounded remainder is disregarded, then $m(r,\alpha)+N(r,\alpha)$ is equal to $T(r)$ for all α. This equality thus demonstrates a beautifully balanced distribution of α-points.

We see that $N(r,\alpha)$ is in a sense the mean value of the number of α-points in $|z|\leqslant r$, and $m(r,\alpha)$ is the mean proximity to α of the value $f(z)$ on $|z|=r$. If the term $\log^+$ in the definition of the proximity function is replaced by the logarithm of the reciprocal of the chordal distance between $f(re^{i\theta})$ and α on the complex sphere, then the remainder term in (1) is

eliminated. Hence the definition of the proximity function is sometimes given in this form.

C. The Order of Meromorphic Functions

For an entire function $f(z)$, the equality

$$\limsup_{r\to\infty}\frac{\log T(r)}{\log r}=\limsup_{r\to\infty}\frac{\log\log M(r)}{\log r}$$

holds, where $M(r)=M(r,f)=\max_{|z|=r}|f(z)|$. Since the right-hand side is the order of $f(z)$ (→ 429 Transcendental Entire Functions), we define the **order (lower order)** ρ of a meromorphic function $f(z)$ by

$$\rho=\limsup_{r\to\infty}\frac{\log T(r)}{\log r}\quad\left(\liminf_{r\to\infty}\frac{\log T(r)}{\log r}\right).$$

The **order** of a meromorphic function in $|z|<R$ is also defined by

$$\rho=\limsup_{r\to R}\frac{\log T(r)}{\log(1/(R-r))}.$$

D. Meromorphic Functions on a Disk

The order function $T(r)$ is bounded if and only if $f(z)$ can be represented as the quotient of two bounded holomorphic functions $h_1(z)$, $h_2(z)$ in $|z|<R$ (Nevanlinna). If $T(r)$ is bounded, $\lim_{r\to R}f(re^{i\theta})$ exists and is finite for every θ, $0\leqslant\theta<2\pi$, except possibly for a set with †linear measure zero (P. Fatou and Nevanlinna). Among functions $f(z)$ such that $\lim_{r\to R}T(r)=\infty$, those satisfying

$$\limsup_{r\to R}\frac{T(r)}{\log(1/(R-r))}=\infty$$

have properties similar to those of transcendental meromorphic functions.

E. Meromorphic Functions in the Whole Finite Plane

Any meromorphic function such that

$$\limsup_{r\to R}\frac{T(r)}{\log r}<K$$

is a rational function. If $f(z)$ is a meromorphic function of order ρ and $\{r_j(\alpha)\}$, $r_j(\alpha)\leqslant r_{j+1}(\alpha)$ $(j=1,2,\ldots)$ is the set of absolute values of α-points, then $\sum_{j=1}^{\infty}(1/r_j(\alpha))^{\rho+\varepsilon}$ converges for any α. Furthermore,

$$f(z)=z^k e^{P(z)}$$

$$\times\lim_{r\to\infty}\prod_{\substack{|a_\mu|\leqslant r\\|b_\nu|\leqslant r}}\frac{\left(1-\dfrac{z}{a_\mu}\right)\exp\left(\dfrac{z}{a_\mu}+\ldots+\dfrac{z^p}{pa_\mu^p}\right)}{\left(1-\dfrac{z}{b_\nu}\right)\exp\left(\dfrac{z}{b_\nu}+\ldots+\dfrac{z^p}{pb_\nu^p}\right)},$$

where p is the smallest integer satisfying $p+1>\rho$, the a_μ and b_ν are the zeros and poles of $f(z)$, respectively, k is an integer, and $P(z)$ is a polynomial of degree at most p (**Hadamard's theorem**).

Let $\alpha_1,\ldots,\alpha_q$ $(q\geqslant 3)$ be distinct values. Then for any meromorphic function in $|z|<R\leqslant\infty$,

$$(q-2)T(r)<\sum_{j=1}^{q}N(r,\alpha_j)-N_1(r)+D(r),\qquad 0\leqslant r<R.$$

Here

$$N_1(r)=\int_0^r\frac{n_1(t)-n_1(0)}{t}dt+n_1(0)\log r,$$

$n_1(r)$ is the number of †multiple points in $|z|\leqslant r$ (a multiple point of order k is counted $k-1$ times), and $D(r)$ is the remainder such that if $R=\infty$, then $D(r)<K(\log T(r)+\log r)$ for some K except possibly for values of r belonging to the union of a countable number of intervals with finite total length, and if $R<\infty$, then $D(r)<K(\log T(r)+\log(1/(R-r)))$ except possibly for the union of a countable number of intervals $\{I_j\}$ with $\sum_j\int_{I_j}d(1/(R-r))<\infty$ (**Nevanlinna's second fundamental theorem**).

Several theorems on value distribution of meromorphic functions can be obtained directly from this theorem. For instance, if $f(z)$ is a transcendental meromorphic function, the equation $f(z)=\alpha$ has an infinite number of roots for every value α except for at most two values called **Picard's exceptional values** (**Picard's theorem**). For a meromorphic function of order ρ, $\lim_{r\to\infty}\sum_{r_j\leqslant r}(r_j(\alpha))^{-\lambda}$ $(\lambda<\rho)$ diverges for every value α except for at most two values (**Borel's theorem**). A value α for which the series converges is called a **Borel exceptional value**. We call $\delta(\alpha)=\delta(\alpha,f)=1-\limsup_{r\to\infty}N(r,\alpha)/T(r)$ the **defect** of f. It always satisfies $0\leqslant\delta(\alpha)\leqslant 1$, and the values with $\delta(\alpha)>0$ are called **Nevanlinna's exceptional values**. The number of values α (may be ∞) with $\delta(\alpha)>0$ is at most countable for any meromorphic function $f(z)$, and $\sum_{i=1}^{\infty}\delta(\alpha_i)\leqslant 2$. There are many studies concerning the values α with $\delta(\alpha)=0$

F. Julia Directions

Among functions that have an essential singularity at the point at infinity and are meromorphic in the whole plane, there are some that possess no †Julia directions. These functions, called **Julia exceptional functions**, are of order 0. A necessary and sufficient condition for $f(z)$ to be a Julia exceptional function is that $f(z)$ can be written in the form $z^m\prod_\mu(1-z/a_\mu)/\prod_\nu(1-z/b_\nu)$ (A. Ostrowski,

1925). Concerning zeros a_μ and poles b_ν of $f(z)$, the following three properties are obtained using the theory of †normal families due to P. Montel: (1) There are constants K_1, K_2, and K_3 independent of r such that $|n(r,\infty)-n(r,0)|<K_1$, $n(2r,\infty)-n(r,\infty)<K_2$, and $n(2r,0)-n(r,0)<K_3$. (2) There are constants K_4 and K_5 such that for any p and q,

$$|a_p|^m \prod_{|a_\mu|<|a_p|}\left|\frac{a_p}{a_\mu}\right| \Big/ \prod_{|b_\nu|<|a_p|}\left|\frac{a_p}{b_\nu}\right| < K_4,$$

$$|b_q|^{-m} \prod_{|b_\nu|<|b_q|}\left|\frac{b_q}{b_\nu}\right| \Big/ \prod_{|a_\mu|<|b_q|}\left|\frac{b_q}{a_\mu}\right| < K_5.$$

(3) There exists an $\varepsilon>0$ satisfying $|a_p/b_q-1|\geqslant\varepsilon>0$ for any p and q.

G. Valiron gave a precise form of Julia directions that corresponds to Borel's theorem (*Acta Math.*, 52 (1928), [3]). Namely, if the order ρ of a meromorphic function $f(z)$ in $|z|<\infty$ is positive and finite, then there exists a direction J defined by $\arg z=\alpha$ such that the zeros $z_\nu(a,\Delta)$ of $f(z)-a$ in any angular domain $\Delta:|\arg z-\alpha|<\delta$ containing J have the property $\sum_\nu |z_\nu(a,\Delta)|^{-(\rho-\varepsilon)}=\infty$ for any $\varepsilon>0$ except for at most two values of a. The direction J is called a **Borel direction**.

G. Relations between Two or More Meromorphic Functions

Borel's unicity theorem can be stated as follows: Let f_j $(j=1,\ldots,n)$ be nonvanishing entire functions satisfying $\sum_{j=1}^n f_j=1$; then for some $(c_1,\ldots,c_n)\neq(0,\ldots,0)$, $\sum_{j=1}^n c_jf_j=0$. This is contained in the following theorem: Let f_j $(j=1,\ldots,n)$ be transcendental entire functions such that $\sum_{j=1}^n f_j=1$; then $\sum_{j=1}^n\delta(0,f_j)\leqslant n-1$. If two meromorphic functions $f_1(z)$, $f_2(z)$ have the same α_j-points for five distinct values α_j $(j=1,\ldots,5)$ (where multiplicity is not taken into account), then they coincide everywhere. If the †Riemann surface of the †algebraic function $w(z)$ defined by a polynomial $P(z,w)=0$ of z, w is of †genus >1, it is impossible to find meromorphic functions $z=f(\zeta)$, $w=g(\zeta)$ that satisfy $P(f(\zeta),g(\zeta))=0$ (†uniformization by meromorphic functions).

H. Asymptotic Values

If a meromorphic function $f(z)\to\alpha$ as $z\to\infty$ along a curve C, the value α and the curve C are called an **asymptotic value** and **asymptotic path**, respectively. Each (Picard's) exceptional value of $f(z)$ is an asymptotic value. For meromorphic functions, no simple relation is known between their order and the number of their asymptotic values. There exists a meromorphic function of order 0 with an infinite number of asymptotic finite values. Some results analogous to those for entire functions are obtained for meromorphic functions by applying the theory of normal families. F. Marty established a systematic theory of normal families of meromorphic functions by using spherical distance.

I. Inverse Functions

Generally, the inverse function of a meromorphic function $w=f(z)$ is infinitely multiple-valued. Let $P(w,w_0)$ be a function element of the inverse function with center at w_0, and let C be an arbitrary curve starting at w_0 and with ω its terminal point. For any domain S containing C, $P(w,w_0)$ can be continued analytically in S up to a point arbitrarily near ω (**Iversen's theorem**). Continue the function element analytically along each half-line starting at its center. Then the set of arguments of half-lines along which the analytic continuation meets a singularity at a finite point is of zero linear measure (**Gross's theorem**).

By considering the inverse image of the suitably cut Riemann surface of the inverse function, the z-plane can be divided into fundamental domains such that each domain is the inverse image of the whole w-plane (with suitable slits removed) and has a boundary each point of which is †accessible from the inside of the domain, and the boundary curves of fundamental domains cluster nowhere in the plane.

For a meromorphic function $f(z)$, the set of functions $z'=\varphi(z)$ defined by $f(z')=f(z)$ (i.e., transformations between points that give $f(z)$ the same value) has the property of a †hypergroup. If $\varphi(z)$ is single-valued, then it is a linear entire function, and if it is finitely multiple-valued, then it is an algebraic function. The †cluster set of the inverse function at a transcendental singularity consists of only one point, ∞, that is, it is an †ordinary singularity. To an analytic continuation along a curve that determines a transcendental singularity of the inverse function there corresponds a curve in the z-plane terminating at ∞. This curve is an asymptotic path of $f(z)$. Namely, the value $f(z)$ tends to the coordinate of the transcendental singularity as $z\to\infty$ along this path. Each asymptotic value of a transcendental meromorphic function $w=f(z)$ corresponds to a transcendental singularity of its inverse function $z=\varphi(w)$, and if we consider two asymptotic paths to be the same if they correspond to the same singularity, then there exists a one-to-one correspondence between the set of asymptotic paths and the set of transcendental singularities of the inverse

function. The inverse function of any meromorphic function of order ρ has at most 2ρ †direct transcendental singularities if $\rho \geqslant 1/2$ and at most 1 such singularity if $\rho < 1/2$ (L. V. Ahlfors).

J. Theory of Covering Surfaces

Ahlfors established the theory of covering surfaces by a metricotopological method and in applying it, obtained Nevanlinna theory and many other results on meromorphic functions.

Let F_r denote the covering surface of the Riemann sphere F_0 with radius 1/2; F_r is the image of $|z| \leqslant r$ under a meromorphic function $w=f(z)$. The area of F_r divided by π, where π is the area of F_0, is given by

$$A(r)=\frac{1}{\pi}\int\int_{|z|<r}\frac{|f'(z)|^2}{(1+|f(z)|^2)^2}\rho\,d\rho\,d\theta,$$

$$z=\rho e^{i\theta},$$

and is called the **mean number of sheets** of F_r. The length of the boundary of F_r is given by

$$L(r)=\int_{|z|=r}\frac{|f'(z)|}{1+|f(z)|^2}|dz|.$$

The relation

$$T(r)=\int\frac{A(r)}{r}dr+O(1)$$

holds (T. Shimizu, Ahlfors).

Consider the Riemann surface of the inverse function of a meromorphic function $w=f(z)$ in $|z|<R\leqslant+\infty$. It has a countable number of components Q_ν over a domain on the w-plane. Let Δ_ν denote the inverse image of Q_ν on the z-plane. If Δ_ν together with its boundary is contained in $|z|<R$, the component Q_ν is called an **island**, and otherwise, a **peninsula**.

Let D be a simply connected domain of the w-plane, $n(r,D)$ be the sum of the sheet numbers of the islands of F_r over D, and $m(r,D)$ be the sum of the areas of the peninsulas of F_r over D divided by the area of D. Then

$$m(r,D)+n(r,D)=A(r)+O(L(r)).$$

Let D_j $(j=1,\ldots,q)$ $(q\geqslant 3)$ be disjoint simply connected domains on the w-plane. Then

$$\sum_{j=1}^{q} n(r,D_j)-\sum_{j=1}^{q} n_1(r,D_j)>(q-2)A(r)-O(L(r)),$$

where $n_1(D)$ is the sum of the orders of branch points in all islands of F_r over D.

For a meromorphic function $f(z)$, $L(r)<A(r)^{1/2+\varepsilon}$, where r satisfies $0\leqslant r<\infty$ except for $r\in\bigcup_j I_j$ for some intervals I_j. Hence in the case where D_j is a point α_j, this inequality yields

$$\sum_{j=1}^{q} n(r,\alpha_j)-\sum_{j=1}^{q} n_1(r,\alpha_j)$$
$$>(q-2)A(r)-O(A(r)^{1/2+\varepsilon})$$

with some exceptional intervals of values of r. This latter important inequality corresponds to Nevanlinna's second fundamental theorem. Let D_j $(j=1,\ldots,q)$ $(q\geqslant 3)$ be disjoint simply connected domains on the w-plane. If every simply connected island over D_j has at least μ_j sheets, then $\sum_{j=1}^{q}(1-(1/\mu_j))\leqslant 2$ (**disk theorem**). It follows from this theorem that given three disjoint disks D_j on the Riemann sphere, there is at least one D_j that has an infinite number of islands over it, and also that given five D_j, there exists at least one D_j that has a 1-sheeted island over it (**Ahlfors's five-disk theorem**). This theorem corresponds to †Bloch's theorem for entire functions. These theorems can also be obtained for meromorphic functions on a disk. Ahlfors established a more important theory by introducing a differential metric.

K. Recent Development

The Nevanlinna brothers raised several important problems, which gave strong motivation for later investigations. The first major breakthrough after World War II was given by A. Gol'dberg in 1956. He gave an example which has infinitely many deficient values (Nevanlinna's exceptional values). W. Hayman proved that $\sum\delta(a,f)^\alpha$ converges for $\alpha>1/3$, and there are examples of meromorphic functions for which the series diverges for $\alpha<1/3$. Finally, A. Weitsman showed that the series converges for $\alpha=1/3$. The second major breakthrough was given by A. Edrei and W. Fuchs in 1959, whose works concern the following Nevanlinna theorem: Let $K(f)$ be

$$\limsup_{r\to\infty}\frac{N(r,0)+N(r,\infty)}{T(r,f)}$$

and $\kappa(\rho)=\inf K(f)$, where inf is taken over all meromorphic function f of order ρ. Then $\kappa(\rho)>0$ if ρ is neither a positive integer nor ∞. Furthermore, Nevanlinna gave a conjecture for an exact value of $\kappa(\rho)$. This conjecture is still open, although several estimates have appeared. In the case of entire functions having only negative zeros the conjecture was positively solved by S. Hellerstein and J. Williamson. Edrei and Fuchs proved that $\kappa(\rho)=1$ for $0\leqslant\rho\leqslant 1/2$ and $\kappa(\rho)=\sin\pi\rho$ for $1/2<\rho\leqslant 1$. They also proved the ellipse theorem: Let $f(z)$ be a transcendental meromorphic function of order ρ $(0\leqslant\rho\leqslant 1)$. Put $u=1-\delta(a,f)$, $v=1-\delta(b,f)$. Then $u, v\in[0,1]$ and $u^2-2uv\cos\pi\rho+v^2\geqslant\sin^2\pi\rho$. If further $u<\cos\pi\rho$, then $v=1$.

For the sum of deficiencies Edrei proved that

$$\sum \delta(a,f) \leqslant \begin{cases} 1-\cos\pi\rho & (0\leqslant\rho\leqslant 1/2) \\ 2-\sin\pi\rho & (1/2<\rho\leqslant 1) \end{cases}$$

unless the number of deficient values is one. It was conjectured that $\sum \delta(a,f)=2$ implies that the order ρ of f is a half integer, the number ν of deficient values is at most 2ρ, and the value of $\delta(a,f)$ is a multiple of $1/\rho$. For the case of entire functions this conjecture is true (A. Pfluger). Weitsman proved that $\nu\leqslant 2\rho$. Other cases remain open. All the above results still hold even if the order is replaced by the lower order. The inverse problem was completely solved by D. Drasin. Many of the above results depend on the concept of Pólya peaks.

N. V. Govorov and V. Petrenko proved independently that

$$\liminf_{r\to\infty}\frac{\log M(r,f)}{T(r,f)}\leqslant\pi\rho \quad \text{for} \quad \rho>1/2$$

for every entire function of order ρ. This was a conjecture made by R. Paley. For $\rho\leqslant 1/2$, an exact upper bound $\pi\rho/\sin\pi\rho$ was given by Valiron. Furthermore, the following result was proved by Edrei and Fuchs:

$$\limsup_{r\to\infty}\frac{N(r,0)}{\log M(r,f)}\geqslant\frac{\sin\pi\rho}{\pi\rho} \quad (0\leqslant\rho<1).$$

L. History

The value distribution theory of meromorphic functions had its inception with the classical Picard theorem. It first appeared as the value distribution theory of entire functions and was developed into a well-organized field by way of the Nevanlinna theory and the Ahlfors theory of covering surfaces. In recent years, emphasis has also been placed on the study of meromorphic functions on open Riemann surfaces (→ 367 Riemann Surfaces). The value distribution of a set of several meromorphic functions was studied first by A. Bloch and developed into the study of **meromorphic curves** by Ahlfors, and H. and J. Weyl [2]. The behavior of meromorphic functions in neighborhoods of general singularities has also been studied. An example of results in that field is the theory of †cluster sets.

References

[1] R. H. Nevanlinna, Eindeutige analytische Funktionen, Springer, 1935, revised edition, 1953; English translation, Analytic functions, Springer, 1970.

[2] H. Weyl and J. Weyl, Meromorphic functions and analytic curves, Princeton Univ. Press, 1943.

[3] G. Valiron, Directions de Borel des fonctions méromorphes, Mémor. Sci. Math., Gauthier-Villars, 1938.

[4] W. K. Hayman, Meromorphic functions, Clarendon Press, 1964.

[5] W. Fuchs, Developments in the classical Nevanlinna theory of meromorphic functions, Bull. Amer. Math. Soc., 73 (1967), 275–291.

[6] A. Edrei and W. Fuchs, The deficiencies of meromorphic functions of order less than one, Duke Math. J., 27 (1960), 233–249.

[7] S. Hellerstein and J. Williamson, Entire functions with negative zeros and a problem of R. Nevanlinna, J. Anal. Math., 22 (1969), 233–267.

[8] A. Edrei, Solution of the deficiency problem for functions of small order, Proc. London Math. Soc., 26 (1973), 435–445.

[9] A. Weitsman, Meromorphic functions with maximal deficiency sum and a conjecture of F. Nevanlinna, Acta Math., 123 (1969), 115–139.

[10] A. Weitsman, A theorem on Nevanlinna deficiencies, Acta Math., 128 (1972), 41–52.

[11] D. Drasin, The inverse problem of Nevanlinna theory, Acta Math., 138 (1977), 83–151.

[12] A. Gol'dberg, Meromorphic functions, J. Soviet Math., 4 (1975), 157–216.

[13] V. Petrenko, Growth of meromorphic functions of finite lower order (in Russian), Izv. Akad. Nauk SSSR, 33 (1969), 414–454.

Also → references to 124 Distribution of Values of Functions of a Complex Variable.

273 (II.17) Metric Spaces

A. General Remarks

The distance between two points $x=(x_1,\ldots,x_n)$ and $y=(y_1,\ldots,y_n)$ in the n-dimensional †Euclidean space $\mathbf{R}^n$ is defined by $\rho(x,y)=\sqrt{(y_1-x_1)^2+\ldots+(y_n-x_n)^2}$. The function $\rho(x,y)$ is nonnegative for every pair (x,y) and has the following properties: (i) $\rho(x,y)=0$ if and only if $x=y$; (ii) $\rho(x,y)=\rho(y,x)$; and (iii) $\rho(x,z)\leqslant\rho(x,y)+\rho(y,z)$ for any three points x, y, z. Property (iii) is called the **triangle inequality**.

B. Definition of Metric Spaces

Abstracting the notion of distance from Euclidean spaces, M. Fréchet defined metric

spaces [1] (1906). A **metric** on a set X is a nonnegative function ρ on $X \times X$ that satisfies (i), (ii), and (iii) of Section A, and a **metric space** (X, ρ), or simply X, is a set X provided with a metric ρ. The members of X are called points, ρ is called the **distance function**, and $\rho(x, y)$ is called the **distance** from x to y. The distance function is sometimes denoted by $d(x, y)$ or $\mathrm{dis}(x, y)$. If (i) is replaced by its weaker form (i') $\rho(x, x)=0$, the function ρ is called a **pseudometric** (or **pseudodistance function**), and X is called a **pseudometric space**.

Examples of metric spaces:

(1) The n-dimensional Euclidean space $\mathbf{R}^n$, in particular the real number system $\mathbf{R}$ with $\rho_0(x, y)=|x-y|$. (2) The †function space $L_p(\Omega)$. (3) The †function space $C(\Omega)$. (4) The †sequence space s, i.e., the space $\mathbf{R}^{\mathbf{N}}$ of all sequences of real numbers with metric $\rho(x, y)=\sum_{n=1}^{\infty} 2^{-n}|x_n-y_n|/(1+|x_n-y_n|)$, where $x=(x_1, x_2, \ldots)$ and $y=(y_1, y_2, \ldots)$. (5) The †sequence space m, i.e., the space of all bounded sequences of real numbers with metric $\rho(x, y)=\sup_n|x_n-y_n|$ for $x=(x_1, x_2, \ldots)$ and $y=(y_1, y_2, \ldots)$. (6) A **Baire zero-dimensional space** $(\Omega^{\mathbf{N}}, \rho)$, where Ω is a set and the distance $\rho(x, y)$ between $x=(x_1, x_2, \ldots)$ and $y=(y_1, y_2, \ldots)$ is equal to the reciprocal of the minimum n such that $x_n \neq y_n$. When the cardinal number τ of Ω is specified, the Baire space is denoted by $B(\tau)$. (7) For any set X, define ρ by setting $\rho(x, x)=0$ and $\rho(x, y)=1$ when $x \neq y$. Then (X, ρ) is a metric space, called a **discrete metric space**. (8) For any set X, define ρ by setting $\rho(x, y)=0$ for any members x and y. Then ρ is a pseudometric, and the resulting space X is called an **indiscrete pseudometric space**.

For a subset M of a metric space X, $\sup\{\rho(x, y) \mid x, y \in M\}$ is called the **diameter** of M (denoted by $d(M)$), and M is said to be **bounded** if its diameter is finite (including $M=\varnothing$). For two subsets A, B of X, $\inf\{\rho(x, y) \mid x \in A, y \in B\}$ is called the **distance** between A and B, denoted by $\rho(A, B)$. We have $\rho(A, B)=\rho(B, A)$. When a family $\mathfrak{M}=\{M_\lambda \mid \lambda \in \Lambda\}$ of subsets of X is a covering of X, i.e., $X=\bigcup_\lambda M_\lambda$, the supremum of the diameters $d(M_\lambda)$ of M_λ in $\mathfrak{M}$, $\sup\{d(M_\lambda) \mid \lambda \in \Lambda\}$, called the **mesh of the covering** $\mathfrak{M}$. For a positive number ε, a covering whose mesh is less than ε is called an ε-**covering**. A metric space X is called **totally bounded** (or **precompact**) (F. Hausdorff, 1927) if for each positive number ε there exists a finite ε-covering of X.

A subset X_1 of a metric space X becomes a metric space if we define its metric ρ_1 by setting $\rho_1(x, y)=\rho(x, y)$ for $x, y \in X_1$, where ρ is the metric of X. The space (X_1, ρ_1) is called a **metric subspace** of (X, ρ). A subset of X is called **totally bounded** (or **precompact**) if it is totally bounded as a metric subspace. Any totally bounded subset is bounded. Conversely, in the Euclidean space $\mathbf{R}^n$ any bounded subset is totally bounded.

A bijection f from a metric space (X_1, ρ_1) onto a metric space (X_2, ρ_2) is called an **isometric mapping** if f preserves the metric, i.e., $\rho_2(f(x), f(y))=\rho_1(x, y)$ for any points $x, y \in X_1$; and X_1 and X_2 are called **isometric** if there is an isometric mapping from X_1 onto X_2.

Let X be a metric space with metric ρ, and left f be an injection from a set Y into X. Then the function $\rho'(y_1, y_2)=\rho(f(y_1), f(y_2))$ $(y_1, y_2 \in Y)$ is a distance function on Y, and with this metric the set Y becomes a metric space called the **metric space induced by** f; f is an isometric mapping from (Y, ρ') onto $(f(Y), \rho)$.

For a finite number of metric spaces (X_1, ρ_1), $\ldots$, (X_n, ρ_n), we can define a metric ρ on their Cartesian product $X=X_1 \times \ldots \times X_n$ by setting

$$\rho(x, y)=\sqrt{\rho_1(x_1, y_1)^2+\ldots+\rho_n(x_n, y_n)^2},$$

for two points $x=(x_1, \ldots, x_n)$, $y=(y_1, \ldots, y_n)$ of X. Thus we obtain a metric space (X, ρ), called the **product metric space** of (X_1, ρ_1), $\ldots$, (X_n, ρ_n). The n-dimensional Euclidean space $\mathbf{R}^n$ is the product metric space of n copies of the real line $(\mathbf{R}, \rho_0)$.

C. Topology for Metric Spaces

For a point x of a metric space (X, ρ) and any positive number ε, the set $U_\varepsilon(x)$ of all points y such that $\rho(x, y)<\varepsilon$ is called the ε-**neighborhood** (or ε-**sphere**) of x. We can introduce a topology for X by taking the family of all ε-neighborhoods as a †base for the neighborhood system (→ 425 Topological Spaces). Then the following five propositions hold, any one of which can be used to define the same topology: (i) A subset O is †open if and only if for any point x in O there is a positive number ε such that the ε-neighborhood of x is contained in O. (ii) A subset F is †closed if and only if any point whose every ε-neighborhood contains at least one point of F is contained in F. (iii) A subset U is a neighborhood of a point x if and only if U contains some ε-neighborhood of x. (iv) A point x is an †interior point of a subset A if and only if some ε-neighborhood of x is contained in A; the interior A^i of A is the set of all such points. (v) A point x is adherent to a subset A if every ε-neighborhood of x contains at least one point of A; the closure $\bar{A}$ of A is the set of all such points, and $x \in \bar{A}$ if and only if $\rho(x, A)=0$.

Every metric space X satisfies the †first countability axiom. A metric space X is a

†Hausdorff space and, more specifically, a †perfectly normal space; it is also †paracompact.

In the same way, we can define a topology for each pseudometric space that satisfies the first countability axiom, but a pseudometric space is not necessarily Hausdorff.

D. Convergence of Sequences

A sequence $\{x_n\}$ of points in a metric space is said to **converge** to a point x (written $\lim_{n\to\infty} x_n = x$) if $\rho(x_n, x)$ tends to zero as $n\to\infty$. The point x is called the **limit** of $\{x_n\}$. This convergence is equivalent to convergence with respect to the topology defined in Section C (→ 87 Convergence). As the first countability axiom is satisfied, we may define the topology by means of convergent sequences of points: the closure $\bar{A}$ of a subset A is the set of all limits of sequences of points in A.

E. Separable Metric Spaces

For a metric space X the following three conditions are equivalent: (i) There exists a countable family $\mathfrak{D}_0$ of open sets of X such that each open set of X is the union of members of $\mathfrak{D}_0$ (†second countability axiom). (ii) X is †separable, that is, X has a countable subset that is †dense in X. (iii) Every open covering of X has a countable subcovering (†Lindelöf space). A metric space with any of these properties is called a **separable metric space**. The sequence space s is separable. Any separable metric space can be isometrically embedded in the sequence space m, i.e., is isometric to a subspace of m (→ 168 Function Spaces B).

F. Compact Metric Spaces

For a metric space X, the following five conditions are equivalent: (i) X is †compact, that is, every open covering of X has a finite subcovering. (ii) X is †countably compact, that is, every countable open covering of X has a finite subcovering. (iii) X is †sequentially compact, that is, any sequence of points in X has a convergent subsequence. (iv) Every nested family $F_1 \supset F_2 \supset \ldots$ of nonempty closed sets of X has a nonempty intersection. (v) Every infinite subset M of X has an accumulation point x, i.e., $x \in \overline{M-\{x\}}$. A metric space satisfying any of these conditions is called a **compact metric space** (M. Fréchet [1]). Every real-valued continuous function defined on a compact metric space has a maximum and a minimum. A metric space is compact if and only if it is totally bounded and complete (→ Section J). Every totally bounded metric space is separable. In particular, every compact metric space is separable.

Let $\mathfrak{U} = \{U_\lambda\}$ be an open covering of a compact metric space X. There exists a positive number δ such that every set with $d(A) < \delta$ is contained in some U_λ. The number δ is called the **Lebesgue number** of the open covering $\mathfrak{U}$.

A subset A of a metric space is said to be **compact** if it is compact as a metric subspace, and A is said to be **relatively compact** if its closure is compact. Bounded closed sets in $\mathbf{R}^n$, in particular closed intervals of real numbers, are compact. For these sets, conditions (i), (iv), and (v) are called the **Heine-Borel theorem** (or **Borel-Lebesgue theorem**), **Cantor's intersection theorem**, and the **Bolzano-Weierstrass theorem**, respectively.

G. Product Spaces of Metric Spaces

Let $(X_1, \rho_1), \ldots, (X_n, \rho_n)$ be metric spaces. Then the Cartesian product $X = X_1 \times \ldots \times X_n$ has distance functions

$$\bar{\rho}_p(x, y) = \{\rho_1(x_1, y_1)^p + \ldots + \rho_n(x_n, y_n)^p\}^{1/p}, \qquad p \geq 1$$

and

$$\bar{\rho}_\infty(x, y) = \max\{\rho_1(x_1, y_1), \ldots, \rho_n(x_n, y_n)\},$$

where $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$. The topology of X induced by each one of these metrics coincides with the product topology. In particular, for the n-dimensional Euclidean space $\mathbf{R}^n$, the metrics $\bar{\rho}_p$ $(p \geq 1)$ and $\bar{\rho}_\infty$ define the same topology.

Let $(X_1, \rho_1), \ldots, (X_n, \rho_n), \ldots$ be a countable number of metric spaces. If we define a metric ρ on the Cartesian product $X = \prod_{n=1}^{\infty} X_n$ by

$$\rho(x, y) = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{\rho_n(x_n, y_n)}{1 + \rho_n(x_n, y_n)},$$

where $x = (x_1, x_2, \ldots)$ and $y = (y_1, y_2, \ldots)$, then the topology defined by ρ is identical with the product topology. For the Cartesian product of an uncountable number of metric spaces, we cannot construct a metric ρ such that the topology induced by ρ agrees with the product topology in general.

H. Uniformity of Metric Spaces

Every metric space X is a †uniform space, for which we may take a countable number of subsets $\{(x, y) \mid \rho(x, y) < 2^{-n}\}$, $n = 1, 2, \ldots$, of $X \times X$ as a base of †uniformity (→ 436 Uniform Spaces).

I. Uniform Continuity

A mapping f from a metric space (X, ρ) into a metric space (Y, σ) is †continuous if for any point x in X and any positive number ε there is a positive number δ such that $f(U_\delta(x)) \subset V_\varepsilon(f(x))$, where $U_\delta(x)$ is a δ-neighborhood for ρ and $V_\varepsilon(y)$ is an ε-neighborhood for σ; that is, $\rho(x, x') < \delta$ implies $\sigma(f(x), f(x')) < \varepsilon$. In this case, we must generally choose δ depending on x and ε. In the special case where we can choose δ depending only on ε, independently of x, we call f **uniformly continuous** in X. (The notion of uniform continuity may be generalized to uniform spaces.) Not every continuous mapping is necessarily uniformly continuous, but every continuous mapping from a compact metric space into a metric space is uniformly continuous.

J. Complete Metric Spaces

A sequence $\{x_n\}$ of points in a metric space (X, ρ) is called a **fundamental sequence** (or **Cauchy sequence**) if $\rho(x_n, x_m) \to 0$ as $n, m \to \infty$. Every convergent sequence is a fundamental sequence, but the converse is not always true. A metric space is called **complete** if every fundamental sequence in the space converges to some point of the space (M. Fréchet [1]). A topological space that is homeomorphic with a complete separable metric space is sometimes called a **Polish space** (→ 22 Analytic Sets I). The metric spaces introduced in examples (1) through (5) of Section B are complete. (In example (3) we must assume that the space Ω is a compact Hausdorff space.) A metric space is compact if and only if it is complete and totally bounded. A †locally compact metric space is homeomorphic to a complete metric space.

For a metric space X, we can construct a complete metric space Y such that there is an isometric mapping φ from X onto a dense subspace X_1 of Y (F. Hausdorff, 1914). Such a pair (Y, φ) is called a **completion** of X. If X has two completions (Y_1, φ_1) and (Y_2, φ_2), then there is an isometric mapping f from Y_1 onto Y_2 with $\varphi_2 = f \circ \varphi_1$. In this sense the completion of X is unique. By identifying X with $\varphi(X)$ when (Y, φ) is the completion of X, any metric space can be regarded as a dense subspace of a complete metric space. For example, the completion of the rational number system $\mathbf{Q}$ is the real number system $\mathbf{R}$. A metric space is totally bounded (= precompact) if and only if its completion is compact.

Baire-Hausdorff theorem: In a complete metric space every set of the †first category is a †boundary set. That is, every set that can be expressed as the union of a countable number of sets whose closures have no interior point has no interior point. In other words, if the union $\bigcup_{n=1}^{\infty} F_n$ of closed sets $F_1, F_2, \ldots$ of X has an interior point, then at least one of the F_n must have an interior point.

K. The Metrization Problem

A topological space X is called **metrizable** if we can introduce a suitable metric for X which induces a topology identical to the original one. A †T_1-space satisfying the second countability axiom is metrizable if and only if it is †regular (**Uryson-Tikhonov theorem**; P. S. Uryson, *Math. Ann.*, 94 (1925), A. Tikhonov, *Math. Ann.*, 95 (1925)). However, a metric space does not necessarily satisfy the second countability axiom. Therefore, the Uryson-Tikhonov theorem does not provide a necessary and sufficient condition for metrizability. The following are some necessary and sufficient conditions for a topological space X to be metrizable:

(1) There exists a nonnegative real-valued function d on $X \times X$ satisfying the first two axioms given in Section A and the following condition: There exists a real-valued function $\varphi(w)$ that converges to zero as $w \to 0$ such that, for any three points x, y, z and any positive number ε, $d(x, y) < \varphi(\varepsilon)$ and $d(y, z) < \varphi(\varepsilon)$ imply $d(x, z) < \varepsilon$ (E. W. Chittenden, *Trans. Amer. Math. Soc.*, 18 (1917)).

(2) X is a T_1-space that has a countable number of open coverings $\mathfrak{M}_1, \mathfrak{M}_2, \ldots$ satisfying the following two conditions: (i) If $U_1, U_2 \in \mathfrak{M}_{n+1}$ have a common point, there is a set $U \in \mathfrak{M}_n$ with $U \supset U_1 \cup U_2$; (ii) for any point x in X, if U_n is any member of $\mathfrak{M}_n$ containing x, the family $\{U_n\}_{n=1,2,\ldots}$ is a base for the neighborhood system of x (P. S. Aleksandrov and Uryson, *C. R. Acad. Sci., Paris*, 177 (1923), and N. Aronszajn). When X is a uniform space, this amounts to saying that X has a metric compatible with the uniform structure if and only if X is a T_1-space and has a countable base of uniformity.

(3) X is a T_1-space that admits a countable number of open coverings $\mathfrak{M}_1, \mathfrak{M}_2, \ldots$ such that $\{S(S(x, \mathfrak{M}_i), \mathfrak{M}_j) \mid i, j = 1, 2, \ldots\}$ is a base for the neighborhood system of x at each point of X, where $S(A, \mathfrak{M})$ is the †star of A relative to $\mathfrak{M}$ (R. L. Moore, *Fund. Math.*, 25 (1935); K. Morita, *Proc. Japan Acad.*, 27 (1951); A. H. Stone, *Pacific J. Math.*, 10 (1960); A. V. Arkhangel'skiĭ, *Dokl. Akad. Nauk SSSR*, 2 (1961)).

(4) X is regular and has a †σ-locally finite open base (J. Nagata, *J. Inst. Polytech. Osaka City Univ.*, 1 (1950); Yu. M. Smirnov, *Uspekhi Mat. Nauk*, 6 (1951)).

(5) X is regular and has a †σ-discrete open base (R. H. Bing, *Canad. J. Math.*, 3 (1951)).

(6) X is a †collectionwise normal Moore space (→ below; Bing, *ibid.*).

(7) X is a †perfect image of a subspace of a Baire's zero-dimensional space (Morita, *Sci. Rep. Tokyo Kyoiku Daigaku*, sec. A, 5 (1955)).

(8) X is a Hausdorff †M-space such that the diagonal is a †G_δ-set in the direct product $X \times X$ (A. Okuyama, *Proc. Japan Acad.* 40 (1964); C. J. R. Borges, *Pacific J. Math.*, 17 (1966); J. Chaber, *Fund. Math.*, 94 (1977)).

A regular space is said to be a Moore space if it has a countable number of open coverings $\mathfrak{M}_i$ such that $\{S(x, \mathfrak{M}_i)\}$ is a base for the neighborhood system of x for any point x. A Moore space is not necessarily metrizable. F. B. Jones (*Bull. Amer. Math. Soc.*, 43 (1937)) proved under the assumption $2^{\aleph_0} < 2^{\aleph_1}$ that every separable normal Moore space is metrizable and asked whether or not every normal Moore space is metrizable. (3) and (6) are partial answers to the question. A normal Moore space is metrizable if it is locally compact and †locally connected (G. Reed and P. Zenor). The existence of a nonmetrizable separable normal Moore space is consistent with and independent of the axioms of the usual ZFC set theory, the †Zermelo-Fraenkel set theory with the †axiom of choice (F. D. Tall). W. G. Fleissner (*Trans. Amer. Math. Soc.*, 273 (1982)) constructed a normal nonmetrizable Moore space assuming an axiom weaker than the continuum hypothesis, while P. J. Nyikos (1980) has proved that every normal space with the first countability axiom is collectionwise normal from the strong axiom of set theory.

In connection with (7) the following result is known. Let f be a †closed continuous mapping from a metric space X onto a topological space Y. Then the following conditions are equivalent: (1) Y is metrizable; (2) For each $y \in Y$ the †boundary $\partial f^{-1}(y)$ of the inverse image is compact; (3) Y satisfies the first countability axiom (Morita and S. Hanai, *Proc. Japan Acad.*, 32 (1956); Stone, *Proc. Amer. Math. Soc.*, 7 (1956); I. A. Vaĭnsteĭn, *Dokl. Akad. Nauk SSSR*, 57 (1947)). In particular, perfect images of metric spaces are metrizable.

For quotient topological spaces of metric spaces → 425 Topological Spaces CC.

References

[1] M. Fréchet, Sur quelques points du calcul fonctionnel, Rend. Circ. Mat. Palermo, 22 (1906), 1–74.
Also → references to 425 Topological Spaces.
For metrization problem see in particular
[2] J. Nagata, Modern general topology, North-Holland, 1968.
[3] R. Engelking, General topology, Polish Scientific Publishers, 1977.
[4] Y. Kodama and K. Nagami, Theory of topological spaces (in Japanese), Iwanami, 1974.
[5] A. V. Arkhangel'skiĭ, Mappings and spaces, Russian Math. Surveys, 21 (1966), 115–162.

274 (XII.16) Microlocal Analysis

A. General Remarks

Let X be an open set of $\mathbf{R}^n$. Then $X \times (\mathbf{R}^n \setminus 0)$ can be identified with $T^*X \setminus 0$, the †cotangent bundle of X minus the zero-section. To every locally integrable function (or distribution or hyperfunction (→ 125 Distributions and Hyperfunctions)) $f(x)$ defined on X, one can assign closed subsets of $X \times (\mathbf{R}^n \setminus 0)$ called the wave front set of f and the singularity spectrum (or analytic wave front set or essential support) of f. The wave front set (resp. the singularity spectrum) of f describes in detail the singularity of f modulo the infinitely differentiable functions (resp. the real analytic functions). One can further associate with f more refined objects defined on $X \times (\mathbf{R}^n \setminus 0)$, such as microfunctions. In many cases one can recover knowledge of the structure of f by analyzing these objects defined on $X \times (\mathbf{R}^n \setminus 0)$. Such an analysis on the cotangent bundle of X is called **microlocal analysis**. Microlocal analysis is particularly successful if f is a solution of a system of linear (pseudo-)differential equations, because in that case one can use various linear transformations, such as differential operators, pseudodifferential operators (→ 345 Pseudodifferential Operators), or microdifferential operators and Fourier integral operators, or quantized contact transformations.

B. Microlocal Analysis for Distributions

Let u be a distribution defined in an open subset X of $\mathbf{R}^n$. The **wave front set** $\mathrm{WF}(u)$ of u is defined as the complement in $X \times (\mathbf{R}^n \setminus 0)$ of the collection of all (x_0, ξ_0) in $X \times (\mathbf{R}^n \setminus 0)$ such that for some neighborhood U of x_0, V of ξ_0 we have for each $\varphi \in C_0^\infty(U)$ and each $N > 0$,

$$\langle u, \varphi \exp(-i\tau x \cdot \xi) \rangle = O(\tau^{-N})$$

as $\tau \to \infty$, uniformly in $\xi \in V$ (L. Hörmander; [1]). $\mathrm{WF}(u)$ is considered to be a subset of

$T^*X \setminus \{0\}$. If $\mathrm{WF}(u)=\varnothing$, then u is a C^∞-function. Let $\pi: X\times(\mathbf{R}^n\setminus 0)\to X$ be the natural projection. If $\pi(\mathrm{WF}(u))$ contains x_0, then u is not a C^∞-function in any small neighborhood of x_0. Thus $\mathrm{WF}(u)$ is the obstruction for u to be infinitely differentiable. Let $p(x,D)$ be a †linear partial differential operator of order m. Assume that $p(x,D)u=f$. Then

$$\mathrm{WF}(f)\subset\mathrm{WF}(u)\subset\mathrm{WF}(f)\cup p_m^{-1}(0),$$

where $p_m:(x,\xi)\to p_m(x,\xi)$ is the †principal symbol of $p(x,D)$. Such a technique of localizing the problem on the cotangent bundle has been used in the form of the estimation of the Fourier transform of f since the advent of the †singular integral operators of A. P. Calderón and A. Zygmund [3,4] (see also S. Mizohata [5,6]). The formula above contains as a special case the classical result that u is a C^∞-function if $p(x,D)$ is †elliptic and if f is a C^∞-function.

In obtaining useful results of microlocal analysis for distributions one often uses Fourier integral operators and pseudodifferential operators (or singular integral operators) (→ 345 Pseudodifferential Operators).

C. Fourier Integral Operators [1,2,7–9]

A **Fourier integral operator** $B: C_0^\infty(\mathbf{R}^n)\to\mathscr{D}'(\mathbf{R}^n)$ is a locally finite sum of linear operators of the type

$$Af(x)=(2\pi)^{-(n+N)/2}\int_{R^{n+N}} a(x,\theta,y)\times\exp(i\varphi(x,\theta,y))f(y)\,dy\,d\theta.$$

Here $a(x,\theta,y)$ is a C^∞-function satisfying the inequality

$$|D_x^\alpha D_\theta^\beta D_y^\gamma a(x,\theta,y)|\leqslant C(1+|\theta|)^{m-\rho|\beta|+(1-\rho)(|\alpha|+|\gamma|)}$$

for some fixed m and ρ, $1\geqslant\rho>1/2$, and any triple of †multi-indices α, β, γ, and $\varphi(x,\theta,y)$ is a real-valued C^∞-function which is homogeneous of degree 1 in θ for $|\theta|>1$. The function φ is called the **phase function** and a the **amplitude function**.

Let $C_\varphi=\{(x,\theta,y)\,|\,d_\theta\varphi(x,\theta,y)=0,\ \theta\neq0\}$ and $W=\{(x,y)\in\mathbf{R}^n\times\mathbf{R}^n\,|\,\exists\theta\neq0$ such that $(x,\theta,y)\in C_\varphi\}$. If $d_{x,\theta,y}\varphi(x,\theta,y)\neq0$ for $\theta\neq0$, then the kernel distribution $k(x,y)$ of A is of class C^∞ outside W. A phase function φ is called nondegenerate if the $d_{x,\theta,y}(\partial\varphi(x,\theta,y)/\partial\theta_j)$, $j=1,2,\ldots,N$, are linearly independent at every point of C_φ. In this case, C_φ is a smooth manifold in $\mathbf{R}^{n+N+n}$ and the mapping $\Phi: C_\varphi\ni(x,\theta,y)\to(x,y,\xi,\eta)$, $\xi=d_x\varphi(x,\theta,y)$, $\eta=d_y\varphi(x,\theta,y)$, is an immersion of C_φ to $T^*(\mathbf{R}^n\times\mathbf{R}^n)\setminus 0$, the cotangent bundle of $\mathbf{R}^n\times\mathbf{R}^n$ minus its zero-section. The image $\Phi C_\varphi=\Lambda_\varphi$ is a **conic Lagrangian manifold**, i.e., the canonical 2-form $\sigma=\sum_j d\xi_j\wedge dx_j-\sum_j d\eta_j\wedge dy_j$ vanishes on Λ_φ and the multiplicative group of positive numbers acts on Λ_φ. Let $\lambda_1,\lambda_2,\ldots,\lambda_{2n}$ be a system of local coordinates in Λ_φ. These, together with $\partial\varphi/\partial\theta_1,\partial\varphi/\partial\theta_2,\ldots,\partial\varphi/\partial\theta_N$, constitute a system of local coordinate functions of $\mathbf{R}^{n+N+n}$ in a neighborhood of C_φ. Let J denote the Jacobian determinant

$$\frac{D\left(\lambda_1,\lambda_2,\ldots\lambda_{2n},\dfrac{\partial\varphi}{\partial\theta_1},\ldots,\dfrac{\partial\varphi}{\partial\theta_N}\right)}{D(x,\theta,y)}.$$

The function $a_{\Lambda_\varphi}=\sqrt{J}\,a|_{C_\varphi}\Phi^{-1}\exp(\pi Mi/4)$ is called the **symbol** of A. Here $a|_{C_\varphi}$ is the restriction of a to C_φ and M is an integer called the **Keller-Maslov index** [10, 11]. The conic Lagrangian manifold $\Lambda_\varphi=\Lambda_\varphi(A)$ and the symbol $a_{\Lambda_\varphi}=a_{\Lambda_\varphi}(A)$ essentially determine the singularity of the †kernel distribution $k(x,y)$ of the Fourier integral operator A. Conversely, given a conic Lagrangian manifold Λ in $T^*(\mathbf{R}^n\times\mathbf{R}^n)\setminus 0$ and a function a_Λ on it, one can construct locally a Fourier integral operator A such that $\Lambda_\varphi(A)=\Lambda$ and $a_{\Lambda_\varphi}(A)=a_\Lambda$. For global construction of such a Fourier integral operator one requires detailed consideration of the Keller-Maslov index. A globally defined Fourier integral operator A with $\Lambda_\varphi(A)=\Lambda$ and $a_{\Lambda_\varphi}(A)=a_\Lambda$ exists if and only if a_Λ is not a function on Λ but a section of the complex line bundle $\Omega_{1/2}\otimes L$, where $\Omega_{1/2}$ is the bundle of square roots of the volume elements of Λ and L is a $\mathbf{Z}_4$ bundle over Λ called the **Maslov bundle**. The factor $\sqrt{J}\exp(\pi Mi/4)$ in the definition of a_{Λ_φ} above appears as the effect of trivialization of the bundle $\Omega_{1/2}\otimes L$. Those Fourier integral operators whose associated conic Lagrangian manifolds are the graphs of †homogeneous canonical transformations of $T^*(\mathbf{R}^n)$ are most frequently used in the theory of linear partial differential equations. Let A be a Fourier integral operator such that $\Lambda_\varphi(A)$ is the graph of a homogeneous canonical transformation χ. Then the adjoint of A is a Fourier integral operator such that the associated conic Lagrangian manifold is the graph of the inverse transformation χ^{-1}. Let A_1 be another such operator; if $\Lambda_\varphi(A_1)$ is the graph of χ_1, then the composed operator A_1A is also a Fourier integral operator and $\Lambda_\varphi(A_1A)$ is the graph of the composed homogeneous canonical transformation $\chi_1\chi$.

Consider the kernel distribution $k(x,y)$ of A. If the phase function φ of A is nondegenerate, then $\mathrm{WF}(k)$ is contained in $\Lambda_\varphi(A)$. Moreover, if the symbol $a_{\Lambda_\varphi}(A)$ does not vanish, then $\mathrm{WF}(k)=\Lambda_\varphi(A)$. Let u be a distribution and A be a Fourier integral operator such that $\Lambda_\varphi(A)$ is the graph of a homogeneous canonical transformation χ. Then $\mathrm{WF}(Au)\subset\chi(\mathrm{WF}(u))$.

A pseudodifferential operator of class $S^m_{\rho,1-\rho}(\mathbf{R}^n)$ is a particular type of Fourier integral operator (→ 345 Pseudodifferential Operators). In fact, a Fourier integral operator A is a pseudodifferential operator of class $S^m_{\rho,1-\rho}(\mathbf{R}^n)$ if and only if $\Lambda_\varphi(A)$ is the graph of the identity mapping of $T^*(\mathbf{R}^n)$. Hence for any Fourier integral operator A, A^*A and AA^* are pseudodifferential operators.

The following theorem is due to Yu. V. Egorov [12]: Let $P(x,D)$ be a pseudodifferential operator of class $S^m_{\rho,1-\rho}(\mathbf{R}^n)$ with the symbol $p(x,\xi)$, and let A be a Fourier integral operator such that the associated conic Lagrangian manifold $\Lambda_\varphi(A)$ is the graph of a homogeneous canonical transformation χ of $T^*(\mathbf{R}^n)$. Then there exists a pseudodifferential operator $Q(x,D)$ with the symbol $q(x,\xi)\in S^m_{\rho,1-\rho}(\mathbf{R}^n)$ such that $P(x,D)A=AQ(x,D)$ and $q(x,\xi)-p(\chi(x,\xi))\in S^{m-2\rho+1}_{\rho,1-\rho}(\mathbf{R}^n)$. Note that $m-2\rho+1<m$.

Assume that $m=1$, $\rho=1$, and that $p_1(x,\xi)$ is a real-valued C^∞-function, homogeneous of degree 1 in ξ for $|\xi|>1$, such that $p(x,\xi)-p_1(x,\xi)\in S^0_{1,0}(\mathbf{R}^n)$ and $d_\xi p_1(x^0,\xi^0)\neq 0$ at (x^0,ξ^0), where $p_1(x^0,\xi^0)=0$. Then one can find a Fourier integral operator A such that the function $q(x,\xi)$ of Egorov's theorem satisfies the relation $q(x,\xi)-\xi_1\in S^0_{1,0}(\mathbf{R}^n)$.

The boundedness of Fourier integral operators in the space $L_2(\mathbf{R}^n)$ (or the spaces $H^s(\mathbf{R}^n)$) has also been studied in several cases. Some sufficient conditions for boundedness can be found in [7, 8, 14–16].

The theory of Fourier integral operators has its origin in the asymptotic representation of solutions of the wave equation, (see, e.g., [17, 18]). Fourier integral operators were first used by G. I. Eskin [7].

D. Essential Support or Analytic Wave Front Set of a Distribution

Inspired by the physical idea introduced by C. Chandler and H. P. Stapp, J. Bros and D. Iagolnitzer introduced the notion of the **essential support** of a distribution, which is a closed subset of $X\times(\mathbf{R}^n\setminus 0)$ [19]. Let u be a distribution defined on an open set X of $\mathbf{R}^n$ and χ be a C^∞-function with compact support around $x_0\in X$ which is locally analytic and different from 0 at x_0. Let $\Sigma_\chi(u)$ be the subset of $\mathbf{R}^n\setminus 0$ of which the complement is defined as follows. A point η is in the complement of $\Sigma_\chi(u)$ if there exist a conic neighborhood U of η, constants α, $\gamma_0>0$, and C_N such that

$$|\langle u,\chi\exp\{-ix'\xi-\gamma|\xi||x-x_0|^2\}\rangle| \leqslant C_N(1+|\xi|)^{-N}e^{-\alpha\gamma|\xi|}$$

for all $\xi\in U$, $0<\gamma<\gamma_0$, and all positive integers N.

The essential support $\Sigma_{x_0}(u)$ of u at x_0, is the limit of $\Sigma_\chi(u)$ when the width $|\operatorname{supp}\chi|$ of the support of χ around x_0 tends to 0. The **essential support** $\Sigma(u)$ of u is the closed subset $\bigcup_{x\in X}\{x\}\times\Sigma_x(u)$ of $X\times\mathbf{R}^n\setminus 0$. $\Sigma(u)$ is the obstruction for u to be real analytic. L. Hörmander [20] also defined the **analytic wave front set** of a distribution, which is also the obstruction for a distribution to be real analytic. The definition of an analytic wave front set is quite different from that of essential support. However they coincide with each other [21]. Moreover, both of them coincide with the singularity spectrum of u if the distribution u is regarded as a hyperfunction.

E. Microlocal Analysis for Hyperfunctions [22]

(1) Microfunctions. Let N be a real analytic manifold of dimension $n+d$ and M its submanifold of codimension d. In what follows, T_MN and T^*_MN denote the normal bundle of M and the conormal bundle supported by M, respectively. Here the **normal bundle** T_MN is defined to be the quotient bundle $TN|_M/TM$ of the tangent bundle and the **conormal bundle** T^*_MN to be the subbundle of the cotangent bundle $T^*N|_M$ that annihilates TM. Identifying N with $\{(x,v)\in TN\mid v=0\}$ or $\{(x,\xi)\in T^*N\mid\xi=0\}$, we define the **tangential sphere bundle** SN and **cotangential sphere bundle** S^*N by $(TN\setminus N)/R^\times_+(=\bigcup_{x\in N}(T_xN\setminus\{0\})/R^\times_+)$ and $(T^*N\setminus N)/R^\times_+(=\bigcup_{x\in N}(T^*_xN\setminus\{0\})/R^\times_+)$, respectively. The **normal sphere bundle** $S_MN=(T_MN\setminus M)/R^\times_+$ and **conormal sphere bundle** $S^*_MN=(T^*_MN\setminus M)/R^\times_+$ are defined in the same manner. In parallel with the algebraic geometry (→ 16 Algebraic Varieties) we define the **real monoidal transform** of N with center M to be the manifold $(N\setminus M)\cup S_MN$ with boundary, in which the center M is blown up to S_MN by the polar coordinates. We denote it by $\widetilde{^MN}$. We mainly use this notion when N is a †complexification X of M, regarding X as a $2n$-dimensional real manifold. In this case we can canonically identify T_MX with $\sqrt{-1}\,TM$, and hence S_MX with $\sqrt{-1}\,SM$. We denote by $x+\sqrt{-1}v0$ the point in S_MX that corresponds to $(x,\sqrt{-1}\,v)$ in $\sqrt{-1}\,SM$ by this identification. An open subset W of $X\setminus M$ is called a **conoidal neighborhood** of a subset U of $\sqrt{-1}\,SM$ if $W\cup\sqrt{-1}\,SM$ is a neighborhood of U in $\widetilde{^MX}$. Let ε, $\tilde\varepsilon$, and $\tilde\tau$ denote respectively the canonical embedding mappings from $X\setminus M$ to X, from $\widetilde{^MX}\setminus\sqrt{-1}\,SM$ to $\widetilde{^MX}$ and from $\sqrt{-1}\,SM$ to $\widetilde{^MX}$. We then define the †sheaves $\tilde{\mathcal{O}}$ and $\mathcal{A}$

by $\tilde{\varepsilon}_*\varepsilon^{-1}\mathcal{O}_X$ and $\tilde{\mathcal{O}}|_{\sqrt{-1}SM}(=\tilde{\tau}^{-1}\tilde{\mathcal{O}}_X)$. Here $\mathcal{O}_X$ denotes the sheaf of germs of holomorphic functions on X. We also define the sheaf $\mathcal{Q}$ on $\sqrt{-1}\ SM$ by $\mathcal{H}^1_{\sqrt{-1}SM}(\tau^{-1}\mathcal{O}_X)$, where τ is the canonical projection from $\widetilde{{}^MX}$ to X and $\mathcal{H}^p_S$ denotes the pth †derived functor of the functor Γ_S of taking the sections with support in S. Then $\mathcal{Q}$ is isomorphic to $\tilde{\mathcal{A}}/\tau^{-1}\mathcal{A}$ for the sheaf $\mathcal{A}$ of real analytic functions on M. The sheaf $\tilde{\mathcal{A}}$ is, so to speak, the sheaf of "boundary values" of holomorphic functions. Actually there exists a canonical mapping b from $\tilde{\mathcal{A}}$ to $\tau^{-1}\mathcal{B}_M$, where $\mathcal{B}_M$ denotes the sheaf of †hyperfunctions on M (→ 125 Distributions and Hyperfunctions). Thus we see that $\mathcal{Q}$ describes the singularities of the boundary value of a holomorphic function. These sheaves $\tilde{\mathcal{A}}$ and $\mathcal{Q}$ are easy to understand intuitively. However, they are defined on $\sqrt{-1}\ SM$, while $\sqrt{-1}\cdot S^*M$ is more important in analysis. Our final goal, namely, the sheaf of microfunctions, is constructed on $\sqrt{-1}\ S^*M$ through cohomological machinery starting from $\mathcal{Q}$. In order to do this, we introduce the disk bundle DM by

$$\{(x+\sqrt{-1}\,v0,\ (x,\sqrt{-1}\,\xi\infty)) \in\sqrt{-1}\ SM\times_M\sqrt{-1}\ S^*M\,|\,\langle v,\xi\rangle\leqslant 0\}.$$

Here $\sqrt{-1}\ SM\times_M\sqrt{-1}\ S^*M$ denotes the †fiber product of $\sqrt{-1}\ SM$ and $\sqrt{-1}\ S^*M$ over M and the symbol $\xi\infty$ is used to emphasize that ξ designates the codirection, which is dual to the infinitesimally small quantity $v0$. The canonical projections from DM to $\sqrt{-1}\ S^*M$ and from $\sqrt{-1}\ SM$ to M are both denoted by τ. Similarly, the projections from DM to $\sqrt{-1}\ SM$ and from $\sqrt{-1}\ S^*M$ to M are denoted by π. We denote by a the antipodal mapping on $\sqrt{-1}\ S^*M$, namely, $a(x,\sqrt{-1}\,\xi\infty)=(x,-\sqrt{-1}\,\xi\infty)$. For a sheaf $\mathcal{F}$ on $\sqrt{-1}\ S^*M$, we also denote $a_*\mathcal{F}(=a^{-1}\mathcal{F})$ by $\mathcal{F}^a$. In the following, $R^j\tau_*$, etc., denotes the jth †right-derived functor of the functor τ_* of taking the direct image of sheaves, etc. (→ 383 Sheaves). Now the sheaf $\mathcal{C}_M$ of **microfunctions** is defined on $\sqrt{-1}\ S^*M$ by $(R^{n-1}\tau_*\pi^{-1}\mathcal{Q})^a\otimes\pi^{-1}\omega_M$, where ω_M denotes the †orientation sheaf of M. Note that $R^j\tau_*\pi^{-1}\mathcal{Q}=0$ holds for $j\neq n-1$.

Remark: Here we have defined the sheaf $\mathcal{C}_M$ of microfunctions on $\sqrt{-1}\ S^*M$. However, it is sometimes more convenient to define the sheaf on $\sqrt{-1}\ T^*M\ (=T^*_MX)$ by the following convention: $\hat{\mathcal{C}}_{M,(x,\sqrt{-1}\xi)}=\mathcal{C}_{M,(x,\sqrt{-1}\xi\infty)}$ if $\xi\neq 0$, and $\mathcal{B}_{M,x}$, if $\xi=0$. Several authors (e.g., [23]) call this sheaf $\hat{\mathcal{C}}_M$ the sheaf of microfunctions and denote it by $\mathcal{C}_M$.

(2) Basic Properties of Microfunctions. The sheaf $\mathcal{C}_M$ defined above has the following properties: (i) The sheaf $\mathcal{C}_M$ is a †flabby sheaf on $\sqrt{-1}\ S^*M$. (ii) For each $(x,\sqrt{-1}\,\xi\infty)$ in $\sqrt{-1}\ S^*M$, there exists a surjective mapping from $\mathcal{B}_{M,x}$ to $\mathcal{C}_{M,(x,\sqrt{-1}\xi\infty)}$. This mapping is denoted by sp. The mapping from $\mathcal{B}_M$ to $\pi_*\mathcal{C}_M$ is also denoted by sp. (iii) We have the following exact sequence: $0\to\mathcal{A}_M\to\mathcal{B}_M\overset{sp}{\to}\pi_*\mathcal{C}_M\to 0$. (iv) $R^k\pi_*\mathcal{C}_M=0$ holds for $k\neq 0$. The exact sequence (iii) shows that the singularities of hyperfunctions are dispersed over $\sqrt{-1}\ S^*M$ and that the dispersed object is described by the sheaf $\mathcal{C}_M$ of microfunctions. For a hyperfunction $f\in\mathcal{B}_M(M)$, we call $sp(f)\ (\in\mathcal{C}(\sqrt{-1}\ S^*M))$ the **spectrum of** f. We denote $\operatorname{supp} sp(f)$ by S.S.f and call it the **singularity spectrum** of f or the **singular spectrum**. It is known [21] that this coincides with the analytic wave front set of f and with the essential support of f if f is a distribution. (v) The following sequence is an exact sequence on $\sqrt{-1}\ SM$: $0\to\tilde{\mathcal{A}}_M\overset{b}{\to}\tau^{-1}\mathcal{B}_M\to\pi_*\tau^{-1}\mathcal{C}_M\to 0$. In the following, a subset A of the $(n-1)$-dimensional sphere S^{n-1} is said to be convex if $\varpi^{-1}(A)\cup\{0\}$ is convex, where ϖ is the canonical projection from $\mathbf{R}^n\setminus\{0\}$ to S^{n-1}, and if $\varpi^{-1}(A)\cup\{0\}$ is convex and includes no straight line, A is said to be **properly convex**. A subset Z of $\sqrt{-1}\ SM$ (resp., $\sqrt{-1}\ S^*M$) is also said to be **(properly) convex** if $\tau^{-1}(x)\cap Z$ (resp., $\pi^{-1}(x)\cap Z$) is so for each x in M. For a subset Z of $\sqrt{-1}\ SM$ its polar set Z° is, by definition, $\{(x,\sqrt{-1}\,\xi_x\infty)\in\sqrt{-1}\ S^*M\,|\,\langle v_x,\xi_x\rangle>0$ holds for each $x+\sqrt{-1}\,v_x0$ in $Z\}$. The polar set Z° of a subset Z of $\sqrt{-1}\cdot S^*M$ is defined in the same way. (vi) Let U be an open subset of $\sqrt{-1}\ SM$ such that $\tau^{-1}(x)\cap U$ is a nonvoid connected set for each x in M. Let V denote $U^{\circ\circ}$. Then we have (a) The restriction mapping $\rho:\Gamma(V;\tilde{\mathcal{A}})\to\Gamma(U;\tilde{\mathcal{A}})$ is a bijection. Here $\Gamma(V;\tilde{\mathcal{A}})$, etc. denotes the space of global sections of $\tilde{\mathcal{A}}$ over V, etc. (b) $0\to\tilde{\mathcal{A}}(U)\to\mathcal{B}(M)\overset{sp}{\to}\mathcal{C}(\sqrt{-1}\ S^*M-U^\circ)$ is an exact sequence. (vii) Let $f(x)$ be a hyperfunction on M. Then the following two statements are equivalent: (a) $sp(f)_{(x_0,\sqrt{-1}\xi_0\infty)}=0$. (b) There exist a finite family of open subsets U_j of $\sqrt{-1}\ SM$ whose polar set U_j° does not contain $(x_0,\sqrt{-1}\,\xi_0\infty)$ and φ_j in $\Gamma(U_j,\tilde{\mathcal{A}})$ such that $f=\sum_j b(\varphi_j)$. Then we say that f is **micro-analytic** at $(x_0,\sqrt{-1}\,\xi_0\infty)$.

(3) Operations on Microfunctions. Let M and N be real analytic manifolds, and let f be a real analytic mapping from N to M. We denote by T^*_NM the kernel of the natural mapping from $N\times_M T^*M$ to T^*N. It is also called a **conormal bundle supported by** N. The associated sphere bundle is denoted by S^*_NM. Denote by ρ and ϖ the natural mappings from $N\times_M\sqrt{-1}\ S^*M\setminus\sqrt{-1}\ S^*_NM$ to $\sqrt{-1}\ S^*N$ and from $N\times_M\sqrt{-1}\ S^*M\setminus\sqrt{-1}\ S^*_NM$ to $\sqrt{-1}\ S^*M$, respectively. (i) Let $\mathcal{B}'_M$ denote the

sheaf $\{u \in \mathscr{B}_M \mid \text{S.S.}\, u \cap \sqrt{-1}\, S_N^* M = \varnothing\}$. Then we have the following two canonical homomorphisms: $f^*: \mathscr{B}'_M \to \mathscr{B}_N$ and $f^*: \rho_! \varpi^{-1} \mathscr{C}_M \to \mathscr{C}_N$. Here and in what follows, for a continuous mapping φ from N to M and a sheaf $\mathscr{F}$ on N, $\varphi_! \mathscr{F}$ denotes the sheaf on M defined by assigning $\{s \in \Gamma(\varphi^{-1}(U); \mathscr{F}) \mid \varphi|_{\text{supp}\, s}: \text{supp}\, s \to U$ is †proper$\}$ to each open subset U of M. These two homomorphisms are consistent. We call each of them a **substitution** (homomorphism) and denote it by $(f^*u)(y) = u(f(y))$. (ii) Let v_M denote the sheaf of †densities. Then there exist the following two canonical homomorphisms: $f_*: f_!(\mathscr{B}_N \otimes_{\mathscr{A}_N} v_N) \to \mathscr{B}_M \otimes_{\mathscr{A}_M} v_M$ and $f_*: \varpi_! \rho^{-1}(\mathscr{C}_N \otimes_{\mathscr{A}_N} v_N) \to \mathscr{C}_M \otimes_{\mathscr{A}_M} v_M$. These two homomorphisms are consistent. We call each an **integration along a fiber** and denote it by $(f_* u)(x) = \int_{f^{-1}(x)} u$. (iii) Using the result in (i), we can define the **product** $u_1 u_2$ **of two hyperfunctions** u_1 and u_2, if $\text{S.S.}\, u_1 \cap (\text{S.S.}\, u_2)^a = \varnothing$. Furthermore, the singularity spectrum of $u_1 u_2$ is contained in $\{(x, \sqrt{-1}(\theta\xi_1 + (1-\theta)\xi_2)\infty) \mid (x, \sqrt{-1}\,\xi_1 \infty) \in \text{S.S.}\, u_1, (x, \sqrt{-1}\,\xi_2 \infty) \in \text{S.S.}\, u_2, 0 \leqslant \theta \leqslant 1\} \cup \text{S.S.}\, u_1 \cup \text{S.S.}\, u_2$.

F. Microdifferential Operators [22]

(1) Microlocal Operators. Let M be a real analytic manifold, and define the sheaf $\mathscr{L}_M$ on $\sqrt{-1}\, S_M^*(M \times M) = \sqrt{-1}\, S^*M$ by $\mathscr{H}^0_{\sqrt{-1} S_M^*(M \times M)}(\mathscr{C}_{M \times M} \otimes v_M)$. A section $K(x, y)\, dy$ of $\mathscr{L}_M$ naturally determines an integral operator $\mathscr{K}: u(y) \to \int K(x, y) u(y)\, dy$. An operator thus obtained is called a **microlocal operator**, because it acts on the sheaf $\mathscr{C}_M$ of microfunctions as a sheaf homomorphism. Usually we identify an operator $\mathscr{K}$ and a kernel function $K(x, y)\, dy$. (i) $\mathscr{L}_M$ is a sheaf of rings by the natural composition. The unit element of $\mathscr{L}_M$ is $\delta(x-y)\, dy$. It acts on $\mathscr{C}_M$ as an identity operator. (ii) Let $K(x, y)\, dy$ be the kernel function of a microlocal operator $\mathscr{K}$ defined near $(x_0, \sqrt{-1}\,\xi_0 \infty)$. Then its **adjoint operator** $\mathscr{K}^*$ is, by definition, the microlocal operator defined near $(x_0, -\sqrt{-1}\,\xi_0 \infty)$ with the kernel function $K(y, x)\, dy$. The operation $*$ is a sheaf isomorphism between $\mathscr{L}_M$ and $\mathscr{L}_M^a$, where a denotes the antipodal mapping (→ Section E).

(2) Microdifferential Operators. A microdifferential operator is an analog of a microlocal operator in the complex domain. By a procedure similar to that used to define the sheaf of microfunctions we first define the sheaf $\mathscr{C}_{Y|X}^\infty$ of **holomorphic microfunctions** for a submanifold Y of X (→ [22, definition 1.1.7 on p. 319], where it is denoted by $\mathscr{C}_{Y|X}$). It follows from the definition that $\mathscr{C}_{Y|X}^\infty$ is supported by $P_Y^*(Y \times X)$, which is identified with $Y \times_X P^*X$. Here and in what follows P^*X, etc., denotes the cotangential projective bundle of X, etc. Then the sheaf $\mathscr{E}_X^\infty$ of **microdifferential operators (of infinite order)** is, by definition, $\mathscr{C}_{X|X \otimes X}^\infty \otimes_{\mathscr{O}_X} \Omega_X$, where Ω_X denotes the sheaf of holomorphic $\dim X$-forms.

Remark. Several notations are used to denote $\mathscr{E}_X^\infty$ in the literature. For example, [22] uses the symbol $\mathscr{P}_X$ and calls it the sheaf of **pseudodifferential operators**. As in the case of microfunctions, some authors use the symbol $\mathscr{E}_X^\infty$ to denote a sheaf on T^*X. In this case $\mathscr{E}_X^\infty|_X$ is, by definition, $\mathscr{D}_X^\infty$, the sheaf of linear differential operators (of infinite order). One should be careful in these notational confusions in referring to papers using microdifferential operators. We note also that the symbols $\mathscr{D}$ and $\mathscr{E}$ have nothing to do with the symbols in distribution theory.

We now list the basic properties of microdifferential operators.

(i) When X is a complexification of a real analytic manifold M, $\mathscr{E}_X^\infty|_{\sqrt{-1}S^*M}$ is a subring of $\mathscr{L}_M$.

(ii) Let Ω be an open subset of P^*X. Using a local coordinate system (x) on X, we define $\hat{\Omega}$ by $\{(x, \xi) \in \mathbf{C}^n \times (\mathbf{C}^n - \{0\}) \mid (x, \xi\infty) \in \Omega\}$. Let $\{p_j(x, \xi)\}_{j \in \mathbf{Z}}$ be a sequence of holomorphic functions on $\hat{\Omega}$ satisfying the following conditions: (1) $p_j(x, \xi)$ is homogeneous of degree j in ξ. (2) For each $\varepsilon > 0$ and each compact subset K of $\hat{\Omega}$, there exists a constant $C_{\varepsilon, K}$ such that $\sup_K |p_j(x, \xi)| \leqslant C_{\varepsilon, K} \varepsilon^j / j!$ $(j \geqslant 0)$ holds. (3) For each compact subset K of $\hat{\Omega}$, there exists a constant R_K such that $\sup_K |p_j(x, \xi)| \leqslant R_K^{-j}(-j)!$ $(j < 0)$ holds. Then there is a one-to-one correspondence between the space of such sequences and the space of sections of $\mathscr{E}_X^\infty$ over Ω.

(iii) A sequence satisfying the conditions in (ii) is called a **symbol sequence**, and the corresponding section of $\mathscr{E}_X^\infty$ is denoted by $\sum_{j \in \mathbf{Z}} p_j(x, D_x)$. If we define a subsheaf $\mathscr{E}_X(m)$ of $\mathscr{E}_X^\infty$ by $\{P = \sum_j p_j(x, D_x) \in \mathscr{E}_X^\infty \mid p_j(x, \xi) = 0\ (j \geqslant m+1)\}$, it is independent of the choice of the local coordinate systems. A microdifferential operator belonging to $\mathscr{E}_X(m)$ is said to be of **order** (at most) m. We denote $\bigcup_m \mathscr{E}_X(m)$ by $\mathscr{E}_X$ and call a section of $\mathscr{E}_X$ a **microdifferential operator of finite order**.

(iv) Let $\Phi_\lambda(z)$ denote $\Gamma(\lambda)/(-z)^\lambda$, where its branch is chosen so that $\Phi_\lambda(-1) = \Gamma(\lambda)$. When $\lambda = 0, -1, -2, \ldots$, we consider its †finite part. Let Ω be a complex neighborhood of $(x_0, \sqrt{-1}\,\xi_0 \infty) \in \sqrt{-1}\, S^*\mathbf{R}^n \cong \mathbf{R}^n \times \sqrt{-1}\, S^{n-1}$. Using a symbol sequence $\{p_j(z, \xi)\}$ on $\hat{\Omega}$, we define a multivalued holomorphic function $K(z, w, \zeta)$ by $\sum_j p_j(z, \zeta) \Phi_{n+j}(\langle z - w, \zeta \rangle)$ and consider its boundary value from the domain

$\mathrm{Re}\langle z-w,\zeta\rangle<0$. We denote the resulting microfunction by

$$\sum_j p_j(x,\sqrt{-1}\,\xi)\Phi_{n+j}(\sqrt{-1}\,(\langle x-y,\xi\rangle+\sqrt{-1}\,0)).$$

Then

$$K(x,y)=(2\pi)^{-n}\int\left\{\sum_j p_j(x,\sqrt{-1}\,\xi)\Phi_{n+j}(\langle x-y,\sqrt{-1}\,\xi\rangle-0)\right\}\omega(\xi)$$

is a well-defined microfunction in a neighborhood of $(x_0,x_0;\sqrt{-1}(\xi_0,-\xi_0)\infty)$ whose support is contained in the antidiagonal set $\Delta^a=\{(x,y;\sqrt{-1}(\xi,\eta)\infty)\in\sqrt{-1}\,S^*(M\times M)\,|\,x=y,\ \xi+\eta=0\}$. Here $\omega(\xi)$ is the volume element of the $(n-1)$-dimensional sphere S^{n-1}. Hence $K(x,y)dy$ defines a microlocal operator. The mapping which associates $K(x,y)dy$ with $\{p_j(x,\xi)\}$ is compatible with the inclusion mapping stated in (i).

(v) By using the †plane wave decomposition of the δ-function due to F. John (→ 125 Distributions and Hyperfunctions CC), we find that microdifferential operators are a natural generalization of †linear differential operators: A linear differential operator corresponds to a symbol sequence $\{p_j(x,\xi)\}_{j\geqslant 0}$, where p_j is a polynomial with respect to ξ.

(vi) (a) Let $P=\sum_j p_j(x,D_x)$ and $Q=\sum_k q_k(x,D_x)$ be microdifferential operators. Then their composition $R=P\circ Q$ is a microdifferential operator with the symbol sequence $\{r_l\}_{l\in Z}$ given by

$$r_l(x,\xi)=\sum_{l=j+k-|\alpha|}\frac{1}{\alpha!}D_\xi^\alpha p_j(x,\xi)D_x^\alpha q_k(x,\xi).$$

Here $D_\xi^\alpha=\partial^{|\alpha|}/\partial\xi_1^{\alpha_1}\ldots\partial\xi_n^{\alpha_n}$ and $\alpha!=\alpha_1!\ldots\alpha_n!$ for the †multi-index $\alpha=(\alpha_1,\ldots,\alpha_n)$. (b) Let $P=\sum_j p_j(x,D_x)$ be a microdifferential operator. Let $\delta(x-y)$ denote the residue class $[1]$ of the left $\mathscr{E}_{X\times X}$-Module $\mathscr{E}_{X\times X}/(\sum_{k=1}^n\mathscr{E}_{X\times X}(x_k-y_k)+\sum_{k=1}^n\mathscr{E}_{X\times X}(\partial/\partial x_k+\partial/\partial y_k))$. ("Module" means sheaf of modules.) Then there exists a unique microdifferential operator $R=\sum_l r_l(y,D_y)$ such that $P(x,D_x)\delta(x-y)=R(y,D_y)\delta(x-y)$. Furthermore, $r_l(x,\xi)$ is given by

$$\sum_{l=j-|\alpha|}\frac{(-1)^j}{\alpha!}D_\xi^\alpha D_x^\alpha p_j(x,\xi).$$

R is called the **adjoint operator** of P and is denoted by P^*. When X is a complexification of the real manifold M, it coincides with the adjoint operator $P^*\in\mathscr{L}_M^a$.

(vii) For a microdifferential operator P in $\mathscr{E}_X(m)$, we define its **principal symbol** $\sigma_m(P)$ by $p_m(x,\xi)$. The principal symbol $\sigma_m(P)$ is independent of the choice of local coordinate system. It gives an isomorphism between $\mathscr{E}_X(m)/\mathscr{E}_X(m-1)$ and $\mathscr{O}_{T^*X}(m)$, the sheaf of holomorphic functions on T^*X which are homogeneous of degree m with respect to ξ.

(viii) (a) Let P be in $\mathscr{E}_X(m)_{(x_0,\xi_0)}$. Assume that $\sigma_m(P)(x_0,\xi_0)\neq 0$. Then its inverse P^{-1} (i.e., $PP^{-1}=P^{-1}P=1$) exists in $\mathscr{E}_X(-m)_{(x_0,\xi_0)}$. (b) Let P and G be in $\mathscr{E}_X(m)_{(x_0,\xi_0)}$, $\mathscr{E}_X(l)_{(x_0,\xi_0)}$, respectively. Suppose that $H^j_{\sigma_l(G)}(\sigma_m(P))(x_0,\xi_0)=0$ $(j=0,\ldots,p-1)$ and that $H^p_{\sigma_l(G)}(\sigma_m(P))(x_0,\xi_0)\neq 0$. (Here $H_f(g)$ is, by definition, the †Poisson bracket $\{f,g\}$ of f and g (→ 82 Contact Transformations)). Then for each S in $\mathscr{E}_{X,(x_0,\xi_0)}$ we can find Q and R in $\mathscr{E}_{X,(x_0,\xi_0)}$ so that $S=QP+R$ with $(\mathrm{ad}\,G)^pR\overset{\mathrm{def}}{=}\overbrace{[G,[G,\ldots}^{p},[G,R],\ldots]=0$ holds. This result is usually referred to as the **Späth-type division theorem (for microdifferential operators)**. In particular, when $G=x_n$ and $(x_0,\xi_0)=(0;1,0,\ldots,0)$, R has the form $\sum_{k=0}^{p-1}R^{(k)}(x,D')D_n^k$. Here $R^{(k)}(x,D')=R^{(k)}(x,D_1,\ldots,D_{n-1})$. As a corollary to this expression we find the following **(Weierstrass-type) preparation theorem (for microdifferential operators)**: Let P be as above, and let $G=x_n$. Then we can find Q and W in $\mathscr{E}_{X,(0;1,0,\ldots,0)}$ such that $P=QW$ with invertible Q and $W=D_n^p+\sum_{k=0}^{p-1}W^{(k)}(x,D')D_n^k$, where $W^{(k)}$ belongs to $\mathscr{E}_X(p-k)$ and $\sigma_{p-k}(W^{(k)})(0;1,0,\ldots,0)=0$.

(ix) **Quantized contact transformation**. (a) Let X be an n-dimensional complex manifold and Ω an open subset of P^*X. Let P_j $(j=1,2,\ldots,n)$ be in $\mathscr{E}_X(1)$ (Ω) and Q_j $(j=1,\ldots,n)$ in $\mathscr{E}_X(0)(\Omega)$. Assume that $[P_j,P_k]=[Q_j,Q_k]=0$ and $[P_j,Q_k]=\delta_{jk}$ hold $(1\leqslant j,k\leqslant n)$. Let φ be the contact transformation from Ω to $P^*\mathbf{C}^n$ defined by $p\mapsto(\sigma_0(Q_1)(p),\ldots,\sigma_0(Q_n)(p),\sigma_1(P_1)(p),\ldots,\sigma_1(P_n)(p))$. Then there exists a unique $\mathbf{C}$-algebra homomorphism $\Phi:\varphi^{-1}\mathscr{E}_{\mathbf{C}^n}\to\mathscr{E}_X|_\Omega$ such that $\Phi(x_j)=Q_j$ and $\Phi(D_j)=P_j$ $(j=1,\ldots,n)$. Furthermore, Φ is an isomorphism, $\Phi\mathscr{E}_{\mathbf{C}^n}(m)=\mathscr{E}_X(m)$ holds, and $\sigma_m(\Phi(R))=\sigma_m(R)\circ\varphi$ holds for R in $\mathscr{E}_{\mathbf{C}^n}(m)$. We call the pair (φ,Φ) a **quantized contact transformation**. In the above situation, the $\mathscr{E}_{X\times\mathbf{C}^n}$-Module

$$\mathscr{M}=\mathscr{E}_{X\times\mathbf{C}^n}\Big/\left(\sum_{j=1}^n\mathscr{E}_{X\times\mathbf{C}^n}(x_j-Q_j)+\sum_{j=1}^n\mathscr{E}_{X\times\mathbf{C}^n}(-D_j-P_j)\right)$$

is a simple holonomic system (→ Section H) whose support is the graph of φ^a. Let u be the canonical generator of $\mathscr{M}$, i.e., the residue class of 1 in $\mathscr{M}$. Then $R^*u=\Phi(R)u$ holds for $R\in\mathscr{E}_{\mathbf{C}^n}$. (b) Conversely, let φ be a contact transformation from a neighborhood of p in P^*X to $P^*\mathbf{C}^n$, and let u be a generator of a simple holonomic system whose support is the graph

of φ^a. Then a **C**-algebra isomorphism Φ: $\varphi^{-1}\mathscr{E}_{\mathbf{C}^n} \to \mathscr{E}_X$ is defined in a neighborhood of p through $R^*u = \Phi(R)u$ ($R \in \mathscr{E}_{\mathbf{C}^n}$), and (φ, Φ) becomes a quantized contact transformation. (c) In particular, let p be a point in $\sqrt{-1}\, S^*M$ and Ω its complex neighborhood. Let (φ, Φ) be a real quantized contact transformation defined on Ω (i.e., φ maps $\sqrt{-1}\, S^*M$ to $\sqrt{-1} \cdot S^*\mathbf{R}^n$). Then, in a neighborhood of $(p, \varphi(p)^a) \in \sqrt{-1}\, S^*(M \times \mathbf{R}^n)$, there exists a microfunction solution $K(y,x) \neq 0$ of the equations $x_j K(y,x) = Q_j K(y,x)$, $-(\partial/\partial x_j)K(y,x) = P_j K(y,x)$ ($j = 1, \dots, n$) ($x \in \mathbf{R}^n$, $y \in M$). Such a microfunction is unique up to constant multiple. The integral operator $\mathscr{K}: v(x) \mapsto \int K(y,x)v(x)\,dx$ gives rise to a sheaf isomorphism between $\varphi^{-1}\mathscr{C}_{\mathbf{R}^n}$ and $\mathscr{C}_M$ in a neighborhood of p. Furthermore, we have $\mathscr{K}(Rv) = \Phi(R)(\mathscr{K}v)$ for $v \in \mathscr{C}_{\mathbf{R}^n}$ and $R \in \mathscr{E}_{\mathbf{C}^n}$. This $\mathscr{K}$ is the counterpart of the Fourier integral operator (→ Section C).

(x) Algebraic properties of $\mathscr{E}_X^\infty$ and $\mathscr{E}_X$. (a) $\mathscr{E}_X$ is †coherent as a left $\mathscr{E}_X$-Module, and its stalk is a †left Noetherian ring. (b) $\mathscr{E}_X^\infty$ is †faithfully flat over $\mathscr{E}_X$. (c) $\mathscr{E}_X$ is †flat over $\mathscr{E}_X(0)$. (d) $\mathscr{E}_X$ is flat over $\pi^{-1}\mathscr{D}_X$.

G. Microdifferential Equations [22, 24]

(1) Background. A system $\mathscr{M}$ of **microdifferential equations** (of finite order) is by definition a †coherent left (or right) $\mathscr{E}_X$-Module, i.e., there exists locally an exact sequence of the form $\mathscr{E}_X^p \to \mathscr{E}_X^q \to \mathscr{M} \to 0$. For a coherent $\mathscr{E}$-Module $\mathscr{M}$, the support $\operatorname{Supp}\mathscr{M}$ of $\mathscr{M}$ is an important geometric object associated with $\mathscr{M}$. It is called the **characteristic variety** of $\mathscr{M}$. For a coherent $\mathscr{D}$-Module $\mathscr{M}$, its characteristic variety is by definition $\operatorname{Supp}(\mathscr{E} \otimes_{\mathscr{D}} \mathscr{M})$. It is often denoted by S.S.$\mathscr{M}$. Since a microdifferential operator is a microlocal operator, the result in (viii) (a) of Section F (3) asserts that $\mathscr{E}xt_{\mathscr{E}}^j(\mathscr{M}, \mathscr{C}_M)$ is supported by the characteristic variety of M intersected with $\sqrt{-1}\, S^*M$. Now, it is known that $V = \operatorname{Supp}\mathscr{M}$ is †involutory (= involutive, in involution) in P^*X, namely, $f|_V = g|_V = 0$ entails $\{f, g\}|_V = 0$ [22, theorem 5.3.2 on p. 453]. One of the most important problems in microlocal analysis is to study how much information V can give concerning the structure of $\mathscr{M}$ itself, and hence that of $\mathscr{E}xt_{\mathscr{E}}^j(\mathscr{M}, \mathscr{C}_M)$. The epoch-making discovery of [22] is that V determines the structure of $\mathscr{E}^\infty \otimes_{\mathscr{E}} \mathscr{M}$ at generic points of V as follows.

(2) Structure Theorems. The fundamental result of [22, theorem 5.3.7 on p. 455] is the following theorem for coherent $\mathscr{E}$-Modules:

Structure theorem 1. Let $\mathscr{M}$ be a coherent $\mathscr{E}_X$-Module satisfying the following conditions: (1) $\mathscr{E}xt_{\mathscr{E}_X}^j(\mathscr{M}, \mathscr{E}_X) = 0$ for $j \neq d$. (2) $V \overset{\text{def}}{=} \operatorname{Supp}\mathscr{M}$ is **regular** at $p \in V$ in the sense that V is nonsingular near p and that $\omega|_V(p) \neq 0$ for the †canonical 1-form ω. Then, through a quantized contact transformation (φ, Φ), $\mathscr{E}^\infty \otimes \mathscr{M}$ is isomorphic to a direct summand of a direct sum of finite copies of **partial de Rham system** $\mathscr{N}_0 = \mathscr{E}_{\mathbf{C}^n}^\infty / (\sum_{j=1}^d \mathscr{E}_{\mathbf{C}^n}^\infty \partial/\partial z_j)$ with $\varphi(p) = (0; 0, \dots, 0, 1) \in P^*\mathbf{C}^n$.

By studying the canonical form of V under real contact transformations, [22] further gives the following structure theorem in a real domain, i.e., in $\sqrt{-1}\, S^*M$ for a real analytic manifold M.

Structure theorem 2. Let X be a complexification of a real analytic manifold M, and let $\mathscr{M}$ be as in structure theorem 1. Let p be a point in $V \cap \sqrt{-1}\, S^*M$. Suppose that the following three conditions are satisfied: (3) $V \cap \bar{V}$ is regular at p. (4) $T_q(V) \cap T_q(\bar{V}) = T_q(V \cap \bar{V})$ holds for each q in $V \cap \bar{V}$. Here $\bar{V}$ denotes the complex conjugate of V (with respect to $\sqrt{-1}\, S^*M$) and $T_q(V)$, etc., denotes the tangent space of V, etc., at q. (5) The generalized Levi form of V is of constant †signature (a, b) near p, where the **generalized Levi form** of V is, by definition, the Hermitian form

$$L(v) = \sum_{j,k} \{p_j, \bar{p}_k\}(x, \sqrt{-1}\,\xi) v_j \bar{v}_k$$

for the p_j such that $V = \bigcap_j p_j^{-1}(0)$. Then $\mathscr{E}^\infty \otimes \mathscr{M}$ is isomorphic to a direct summand of a direct sum of finite copies of the system $\mathscr{E}^\infty \otimes_{\mathscr{E}} \mathscr{N}$ considered in a neighborhood of $(x, \sqrt{-1}\,\xi) = (0, \sqrt{-1}(0, \dots, 0, 1)) \in \sqrt{-1}\, S^*\mathbf{R}^n$, where $\mathscr{N}$ is given by

$$\frac{\partial}{\partial x_j} f = 0 \qquad (j = 1, \dots, r),$$

$$\left(\frac{\partial}{\partial x_{r+2k-1}} + \sqrt{-1}\frac{\partial}{\partial x_{r+2k}}\right) f = 0 \qquad (k = 1, \dots, s),$$

$$\left(\frac{\partial}{\partial x_{r+2s+l}} + \sqrt{-1}\, x_{r+2s+l}\frac{\partial}{\partial x_n}\right) f = 0 \qquad (l = 1, \dots, a),$$

$$\left(\frac{\partial}{\partial x_{r+2s+l}} - \sqrt{-1}\, x_{r+2s+l}\frac{\partial}{\partial x_n}\right) f = 0 \qquad (l = a+1, \dots, a+b).$$

Here, $r = 2\operatorname{codim} V - \operatorname{codim}(V \cap \bar{V})$ and $s = \operatorname{codim}(V \cap \bar{V}) - \operatorname{codim} V - (a+b)$.

The first (resp. second) type of equation in the above are called (partial) **de Rham equations** (resp. (partial) **Cauchy-Riemann equations**). The third and the fourth are called **Lewy-Mizohata equations** (of type (a, b)) after these authors' pioneering works [25, 26]. Thus any system is seen to be microlocally isomorphic to a mixture of these three types of equations, generically speaking. As a corollary to

this, structure theorem 2 clarifies the structure of microfunction solutions of $\mathscr{M}$ as follows:

Structure theorem 3. Let M, X, $\mathscr{M}$, V, and p be as in structure theorem 2. Then $\mathscr{E}xt^j_{\mathscr{E}_X^\infty}(\mathscr{E}_X^\infty \otimes \mathscr{M}, \mathscr{C}_M)=0$ $(j \neq a)$ holds, and the remaining cohomology group $\mathscr{F}=\mathscr{E}xt^a_{\mathscr{E}_X^\infty}(\mathscr{E}_X^\infty \otimes \mathscr{M}, \mathscr{C}_M)$ has the following structure in a neighborhood U of p: There exists an s-dimensional complex manifold Y, a real analytic manifold N, and a †smooth mapping φ from $V \cap U \cap \sqrt{-1}\, S^*M$ to $Y \times \sqrt{-1}\, S^*N$ such that $\mathscr{F}=\varphi^{-1}\mathscr{G}$ for a sheaf $\mathscr{G}$ on $Y \times \sqrt{-1}\cdot S^*N$ that is a direct summand of $\mathscr{H}^N$, with $\mathscr{H}$ being the solution sheaf of the partial Cauchy-Riemann equations associated with Y.

H. Holonomic Systems

A coherent (left) $\mathscr{E}$-Module $\mathscr{M}$ is called **holonomic** if Supp$\mathscr{M}$ is †Lagrangian. A coherent (left) $\mathscr{D}$-Module $\mathscr{M}$ is called holonomic if $\mathscr{E} \otimes_{\mathscr{D}} \mathscr{M}$ is so. Even though the term "holonomic" is currently used, another term, "**maximally overdetermined**," is used to describe the same object in some of the literature, including [22]. The importance of such a system lies in the fact that the space of its microfunction solutions is finite-dimensional [27, 29]. In this sense it resembles an ordinary differential equation. A holonomic system, however, does not satisfy condition (2) of structure theorem 1 of Section G, and its structure is rather complicated. A result which corresponds to structure theorem 1 in Section G is the following: Let V be an involutory submanifold of T^*X, and let $\mathscr{E}_V$ be the subring of $\mathscr{E}_X$ generated by $\{P \in \mathscr{E}_X(1) | \sigma_1(P)|_V=0\}$. Then a coherent $\mathscr{E}_X$-Module $\mathscr{M}$ defined on an open subset Ω of T^*X is said to have **regular singularities** along V if for any point p of Ω, there exist a neighborhood U of p and an $\mathscr{E}_V$-sub-Module $\mathscr{M}_0$ of $\mathscr{M}$ defined on U which is coherent over $\mathscr{E}(0)$, and which generates $\mathscr{M}$ as an $\mathscr{E}_X$-Module. A holonomic system is said to have R.S., which is an abbreviation for regular singularities, if it has R.S. along its support. Then for an arbitrary holonomic $\mathscr{E}$-Module $\mathscr{M}$ we can find a holonomic $\mathscr{E}$-Module $\mathscr{M}_{reg}$ with R.S. such that $\mathscr{E}^\infty \otimes_{\mathscr{E}} \mathscr{M}_{reg} \cong \mathscr{E}^\infty \otimes_{\mathscr{E}} \mathscr{M}$ holds. See [28] for the proof of this striking result and related topics on **holonomic systems with regular singularities**.

An elementary class of holonomic systems is that of **simple holonomic systems**. A holonomic $\mathscr{E}$-Module $\mathscr{M}$ is called simple if there exists a left Ideal $\mathscr{I}$ such that $\mathscr{M}=\mathscr{E}/\mathscr{I}$ and that the symbol Ideal $\{\sigma(P) | P \in \mathscr{I}\}$ coincides with the defining Ideal of Supp$\mathscr{M}$. Let u denote the generator $1 \bmod \mathscr{I}$ of a simple holonomic system $\mathscr{E}/\mathscr{I}$. Suppose that $\Lambda=$Supp$\mathscr{M}$ is nonsingular. Then the **principal symbol** $\sigma_\Lambda(u)$ is defined as follows: For $P=\sum_j p_j(x, D_x) \in \mathscr{E}(m)$, let L_p denote the first-order linear differential operator

$$L_P=H_{p_m(x,\xi)}+\left(p_{m-1}(x,\xi)-\frac{1}{2}\sum_i\frac{\partial^2 p_m(x,\xi)}{\partial x_i \partial \xi_i}\right).$$

Here, H_{p_m} denotes the Hamiltonian vector field defined by $\{p_m, \cdot\}$. Let Ω_Λ and Ω_X denote the sheaf of n-forms of Λ and X, respectively, and let $\Omega_\Lambda^{\otimes 1/2}$ (resp. $\Omega_\Lambda^{\otimes 1/2} \otimes \Omega_X^{\otimes -1/2}$) denote a line bundle L such that $L^{\otimes 2}$ is isomorphic to Ω_Λ (resp. $\Omega_\Lambda \otimes \Omega_X^{\otimes -1}$). Since the line bundles $\Omega_\Lambda^{\otimes 1/2}$ and $\Omega_\Lambda^{\otimes 1/2} \otimes \Omega_X^{\otimes -1/2}$ do not exist globally in general, all the equations among their sections should be understood up to a constant multiplicative factor. When Λ is a purely imaginary Lagrangian submanifold of $\sqrt{-1}\, S^*M$ for a real analytic manifold M, these line bundles can be constructed globally by using the Keller-Maslov index (→ Section C). Let v and ψ denote respectively the first term and the second term in the right-hand side of the above definition of L_P. Then $\tilde{L}_P: \Omega_\Lambda^{\otimes 1/2} \to \Omega_\Lambda^{\otimes 1/2}$ is given by $\tilde{L}_P(s)=(1/2s)L_v(s^2)+\psi s$ for $s \in \Omega_\Lambda^{\otimes 1/2}$, where $L_v(s^2)$ denotes the †Lie derivative of s^2 along v. One can then prove that the system of equations $\tilde{L}_P s=0$ $(P \in \mathscr{I})$ admits locally one and exactly one nonzero solution s in $\Omega_\Lambda^{\otimes 1/2}$ up to a constant factor. Then the principal symbol $\sigma_\Lambda(u)$ of u equals $s \otimes \sqrt{dx} \in \Omega_\Lambda^{\otimes 1/2} \otimes \Omega_X^{\otimes -1/2}$ by definition. The principal symbol $\sigma_\Lambda(u)$ is homogeneous with respect to ξ, and its homogeneous degree is called the **order** of u and is denoted by $\mathrm{ord}_\Lambda(u)$. The microlocal structure of a simple holonomic $\mathscr{E}$-Module with a nonsingular characteristic variety is determined by the order of its generator as follows: (a) Let $\mathscr{E}u$ and $\mathscr{E}v$ denote simple holonomic $\mathscr{E}$-Modules with the same characteristic variety Λ. Then $\mathscr{E}u$ is isomorphic to $\mathscr{E}v$ if and only if $\mathrm{ord}_\Lambda(u)-\mathrm{ord}_\Lambda(v)$ is an integer. (b) Let $\mathscr{E}u$ be a simple holonomic system, and let α denote the order of u. Then through a suitable quantized contact transformation, $\mathscr{E}u$ is isomorphic to $\mathscr{E}_{C^n}w$, where w satisfies

$$\left(z_1\frac{\partial}{\partial z_1}+\left(\alpha+\frac{1}{2}\right)\right)w=0,$$

$$\frac{\partial w}{\partial z_j}=0 \qquad (j=2, \ldots, n).$$

Thus the microlocal structure of a simple holonomic system $\mathscr{M}=\mathscr{E}u$ is fairly simple at nonsingular points of its characteristic variety. Moreover a Hartogs-type theorem for microdifferential equations [28, ch. I, §2] entails that if Supp$\mathscr{M}$ has the form $\Lambda_1 \cup \Lambda_2$ with Lagrangian manifolds Λ_1 and Λ_2 such that

$\mathrm{codim}_{\Lambda_1}(\Lambda_1 \cap \Lambda_2) \geq 2$; then $\mathscr{M}$ has the form $\mathscr{M}_1 \oplus \mathscr{M}_2$ with $\mathrm{supp}\,\mathscr{M}_j = \Lambda_j$ $(j=1,2)$. Hence the case where $\mathrm{codim}_{\Lambda_1}(\Lambda_1 \cap \Lambda_2) = 1$ is important. Suppose Λ_1 and Λ_2 are nonsingular in this case and $T_x\Lambda_1 \neq T_x\Lambda_2$ at a point $x \in \Lambda_1 \cap \Lambda_2$. Then, through a quantized contact transformation (ψ, Φ), the system $\mathscr{M}$ is isomorphic to $\mathscr{E}_{\mathbf{C}^n} w$, with w satisfying the following equations defined near $(0; dz_1 \infty)$:

$$\left(z_1 \frac{\partial}{\partial z_1} + \left(\alpha_1 + \frac{1}{2}\right)\right) w = 0,$$

$$\left(z_2 \frac{\partial}{\partial z_2} + \left(\alpha_1 - \alpha_2 + \frac{1}{2}\right)\right) w = 0,$$

$$\frac{\partial w}{\partial z_j} = 0 \qquad (j = 3, \ldots, n),$$

where $\alpha_j = \mathrm{ord}_{\Lambda_j}(u)$ $(j=1,2)$ with $\psi(\Lambda_1) = \{z_1 = 0, \xi_2 = \ldots = \xi_n = 0\}$ and $\psi(\Lambda_2) = \{z_1 = z_2 = 0, \xi_3 = \ldots = \xi_n = 0\}$. For more general cases and an application → [30].

The theory of holonomic systems and its applications are being studied most intensively, raising the hope of establishing a unified theory of special functions of several variables; → [23] and references cited therein.

I. History

Microlocal analysis means local analysis on the cotangent bundle. It emphasizes the importance of localization in cotangent bundles in analysis, which was pointed out by S. Mizohata [5, 6] immediately after the advent of singular integral operators in the works of A. P. Calderón and A. Zygmund [3, 4]. Since then, localization in the cotangent bundle has been used frequently in the theory of linear partial differential equations. R. T. Seeley [31] proved that the symbol of a singular integral operator is well defined on the cotangent bundle. Works by J. J. Kohn and L. Nirenberg [32] and L. Hörmander (*Comm. Pure Appl. Math.*, 18 (1965)) strengthened the trend of localizing the problem on the cotangent bundle. Although it seems that the term "microlocal analysis" first appeared in the literature in 1973 (T. Kawai, *Astérisque*, 2 and 3 (1973)), the basic part of the theory had been constructed during the period from 1969 to 1972 by M. Sato (*Proc. Intern. Conf. Functional Anal. and Related Topics*, 1969), Yu. V. Egorov [12], Hörmander [8, 20], J. J. Duistermaat and Hörmander (*Acta Math.*, 128 (1972)), M. Kashiwara and Kawai (*Proc. Japan Acad.*, 46 (1970)), and Sato, Kawai, and Kashiwara [22]. Apparently the work of V. P. Maslov [11] had an important influence on the work of Egorov. The most important part of Sato's contribution was the fact that, through the construction of the sheaf of microfunctions he found that the singularities of hyperfunctions can be canonically dispersed over the cotangent bundle (→ Section E) and that a hyperfunction solution u of a linear differential equation $Pu = 0$ is concentrated on the characteristic variety when its singularities are thus dispersed (→ Section G). The last-stated fact was also formulated by Hörmander [8, 20] in the framework of distribution theory. The most important part of the contribution of Egorov [12] was the discovery that one can use an integral transformation introduced by V. I. Eskin [7] to find a transformation of pseudodifferential operators compatible with a homogeneous canonical transformation, i.e., a contact transformation, so that the commutation relations and the orders of the operators can be preserved. Hörmander (*Acta Math.*, 121 (1969)) independently introduced integral transformations of the same type, calling them Fourier integral operators. Egorov [13] and L. Nirenberg and F. Treves [33] successfully used the transformation of operators to study the regularity and existence of solutions. Subsequently, Hörmander [8] elaborated the theory of Fourier integral operators. Kashiwara and Kawai (*Proc. Japan Acad.*, 46 (1970)) observed that a pseudodifferential operator in their sense (now called a microdifferential operator; → Section F) gives rise to a sheaf homomorphism on the sheaf of microfunctions and that the structure of the microfunction solutions of pseudodifferential equations is determined by the principal symbol of the operator in question if it has simple characteristics. Then Sato, Kawai, and Kashiwara [22] succeeded in amalgamating these two theories, namely, the theory of microfunctions and the theory of the transformation of operators. (They called the transformation a quantized contact transformation in [22]). Such an amalgamation was also done independently by Hörmander [8, 20], who introduced the notion of the wave front set for distributions as a substitute for the support of microfunctions. Incidentally, it is noteworthy that C. Chandler and H. P. Stapp (*J. Math. Phys.*, 10 (1969)) and D. Iagolnitzer and Stapp (*Comm. Math. Phys.*, 14 (1969)) obtained a notion similar to the singularity spectrum in a physical context. Their results were later elaborated (around 1971–1973) by J. Bros and Iagolnitzer [19]. With the aid of the above-mentioned amalgamation of the theories, the works of Sato, Kawai, and Kashiwara [22], Hörmander [20], Duistermaat and Hörmander (*Acta Math.*, 121 (1969)), Kawai (*Publ. Res. Inst. Math. Sci.*, 7 (1971–1972)) and K. G. Andersson (*Trans. Amer. Math. Soc.*, 177 (1973)) have clarified the

importance of the bicharacteristic strip, which is a submanifold of a cotangent bundle, not a base manifold, as a carrier of singularities of solutions of pseudodifferential equations (→ Section G); and the works of Sato, Kawai, and Kashiwara [22], Sato (*Actes Congr. Internat. Math. Nice*, 1971) and Kawai (*Publ. Res. Inst. Math. Sci.*, 7 (1971–1972)) revealed the hidden mechanism of the celebrated counterexample of H. Lewy [25]. Among these, the contribution of Sato, Kawai, and Kashiwara [22] was most decisive and fundamental in that it first clarified the structure of a general system of pseudodifferential equations at generic points of the characteristic variety and then derived from it the above-quoted results on the structure of the solutions (→ Section G).

Microlocal analysis has now become one of the most important and basic concepts in the theory of linear partial differential equations and theoretical physics. For recent developments → Hörmander [34] and V. Guillemin, Kashiwara, and Kawai [23] and references cited therein.

References

[1] J. J. Duistermaat, Fourier integral operators, Lecture notes, Courant Inst. Math. Sci., 1973.
[2] V. Guillemin and S. Sternberg, Geometric asymptotics, Amer. Math. Soc. Math. Surveys 14 (1977).
[3] A. P. Calderón and A. Zygmund, Singular integral operators and differential equations, Amer. J. Math., 79 (1957), 901–921.
[4] A. P. Calderón, Uniqueness in the Cauchy problem for partial differential equations, Amer. J. Math., 80 (1958), 16–36.
[5] S. Mizohata, Systèmes hyperboliques, J. Math. Soc. Japan, 11 (1959), 205–233.
[6] S. Mizohata, Note sur le traitement par les opérateurs d'integrale singulière de Cauchy, J. Math. Soc. Japan, 11 (1959), 234–240.
[7] G. I. Eskin, The Cauchy problem for hyperbolic systems in convolutions, Math. USSR-Sb., 3 (1967), 243–277.
[8] L. Hörmander, Fourier integral operators I, Acta Math., 127 (1971), 79–183.
[9] J. Chazarain (ed.), Fourier integral operators and partial differential equations, Lecture notes in math. 459, Springer, 1975.
[10] J. B. Keller, Corrected Bohr Sommerfeld conditions for nonseparable systems, Ann. Phys., 4 (1958), 180–188.
[11] V. P. Maslov, Théorie des perturbations et méthodes asymptotiques, Gauthier-Villars, 1972. (Original in Russian, 1965.)
[12] Yu. V. Egorov, On canonical transformation of pseudodifferential operators (in Russian), Uspekhi Mat. Nauk, 25 (1969), 235–236.
[13] Yu. V. Egorov, Conditions for solvability of pseudodifferential equations, Soviet Math. Dokl., 10 (1969), 1020–1022.
[14] D. Fujiwara, On the boundedness of integral operators with highly oscillatory kernels, Proc. Japan Acad., 51 (1975), 96–99.
[15] H. Kumano-go, A calculus of Fourier integral operators on $\mathbf{R}^n$ and the fundamental solution for an operator of hyperbolic type, Comm. Partial Diff. Eq., 1 (1976), 1–44.
[16] K. Asada and D. Fujiwara, On some oscillatory integral transformations in $L^2(\mathbf{R}^n)$, Japan. J. Math., 4 (1978), 299–361.
[17] A. Sommerfeld, Optics, Lectures on Theoretical Physics, vol. 4, Academic Press, 1964.
[18] P. D. Lax, Asymptotic solution of oscillatory initial value problems, Duke Math. J., 24 (1957), 627–646.
[19] D. Iagolnitzer, Analytic structure of distributions and essential support theory, Structural Analysis and Collision Amplitude, North-Holland, 1976, 295–358.
[20] L. Hörmander, Uniqueness theorem and wave front sets for solution of linear differential equation with analytic coefficients, Comm. Pure Appl. Math., 24 (1971), 671–704.
[21] J. M. Bony, Equivalence des diverses notions de spectre singulier analytique, Séminaire Goulaouic-Schwartz 1976–77, École Polytechnique.
[22] M. Sato, T. Kawai, and M. Kashiwara, Microfunctions and pseudodifferential equations, Lecture notes in math. 287, Springer, 1973, 265–529.
[23] V. Guillemin, M. Kashiwara, and T. Kawai, Seminar on microlocal analysis, Ann. math. studies 93, Princeton Univ. Press, 1979.
[24] J. E. Björk, Rings of differential operators, North-Holland, 1979.
[25] H. Lewy, An example of a smooth linear partial differential equation without solution, Ann. Math., 66 (1957), 155–158.
[26] S. Mizohata, Solutions nulles et solutions nonanalytiques, J. Math. Kyoto Univ., 1 (1962), 271–302.
[27] M. Kashiwara, On the maximally overdetermined system of linear differential equations I, Publ. Res. Inst. Math. Sci., 10 (1974–1975), 563–579.
[28] M. Kashiwara and T. Kawai, On holonomic systems of microdifferential equations III, Systems with regular singularities, Publ. Res. Inst. Math. Sci., 17 (1981), 813–979.
[29] M. Kashiwara and T. Kawai, Finiteness theorem for holonomic systems of microdifferential equations, Proc. Japan Acad., 52 (1976), 341–343.
[30] M. Sato, M. Kashiwara, T. Kimura, and

T. Oshima, Microlocal analysis of prehomogeneous vector spaces, Inventiones Math., 62 (1980), 117–179.
[31] R. T. Seeley, Singular integrals on compact manifolds, Amer. J. Math., 81 (1959), 658–690.
[32] J. J. Kohn and L. Nirenberg, An algebra of pseudodifferential operators, Comm. Pure Appl. Math., 18 (1965), 269–305.
[33] L. Nirenberg and F. Treves, On local solvability of linear partial differential equations II, Comm. Pure Appl. Math., 23 (1970), 459–510.
[34] L. Hörmander, Seminar on singularities of solutions of linear partial differential equations, Ann. math. studies 91, Princeton Univ. Press, 1979.

275 (VII.14) Minimal Submanifolds

A. General Remarks

An immersion f of an m-dimensional manifold M with boundary ∂M (possibly empty) into a Riemannian manifold N is called **minimal** if the †mean curvature vector field H of M with respect to the induced Riemannian metric vanishes identically. Then M is called a **minimal submanifold of** N. This definition comes from the following variational problem: By a smooth †variation of f is meant a smooth mapping $F: I\times M\rightarrow N$, where $I=(-1,1)$, such that each $f_t=F(t,\cdot): M\rightarrow N$ is an immersion, $f_0=f$, and $f_t|\partial M=f|\partial M$ for all $t\in I$. Let dV_t be the volume element of the metric induced by f_t, and set $V(t)=\int_M dV_t$, the volume of M at time t. Then the first variation of the volume is expressed as

$$\left.\frac{dV}{dt}\right|_{t=0}=-m\int_M\left\langle H, F_*\left.\frac{\partial}{\partial t}\right|_{t=0}\right\rangle dV_0,$$

where $\partial/\partial t$ denotes the canonical vector field along the I factor in $I\times M$. Thus the mean curvature vector field H of f vanishes identically if and only if $dV/dt|_{t=0}=0$ for all variations of f. Therefore a minimal submanifold gives an extremal of the volume integral, though neither necessarily minimal nor of the least volume.

In the case $N=\mathbf{R}^n$, an immersion $x: M\rightarrow\mathbf{R}^n$ is viewed as a vector-valued function, and the mean curvature vector field H is expressed as $H=\Delta x/m$, where Δ denotes the †Laplace-Beltrami operator $-g^{ij}\nabla_i\nabla_j$. Thus x is minimal if and only if each component function of x is harmonic. In particular, there is no compact minimal submanifold without boundary in $\mathbf{R}^n$.

The history of the theory of minimal submanifolds goes back to J. L. Lagrange, who studied minimal surfaces in the 3-dimensional Euclidean space $\mathbf{R}^3$. In 1762 he developed his algorithm for the †calculus of variations, which can be applied to higher-dimensional problems and is now known as the †Euler-Lagrange equation. For instance, let D be a domain in the plane $\mathbf{R}^2$ and $z=f(x,y)$, $(x,y)\in D$, the equation of a surface in $\mathbf{R}^3$. As a necessary condition for the surface to have the least area among the surfaces with fixed boundary, Lagrange obtained a †quasilinear elliptic partial differential equation of the second order, called the **minimal surface equation**:

$$(1+z_y^2)z_{xx}-2z_xz_yz_{xy}+(1+z_x^2)z_{yy}=0.$$

Before this, in 1744, L. Euler had found that a †catenoid is a minimal surface. In 1766, J. B. M. C. Meusnier showed that a right †helicoid is a minimal surface. Besides catenoids and helicoids, in 1834, H. F. Scherk found that the surface defined by $z=\log(\cos y)-\log(\cos x)$ is a minimal surface, which is called **Scherk's surface**.

In the latter half of the 19th century, †Plateau's problem (→ Section C; 334 Plateau's Problem) was studied extensively by O. Bonnet, B. Riemann, K. Weierstrass, A. Enneper, G. Darboux, and others. The problem is stated as follows: Given a †Jordan curve Γ in $\mathbf{R}^3$ (or in $\mathbf{R}^n$), find a surface of least area having Γ as its boundary. On the other hand, in 1866, Weierstrass gave a general formula, called the **Weierstrass-Enneper formula** (→ Section B (5)), to express a simply connected minimal surface in terms of a complex analytic function and a meromorphic function with certain properties. The formula allows one to construct a great variety of minimal surfaces by choosing those functions.

The existence of a minimal surface of disk type having a prescribed boundary curve was first obtained in 1930 by J. Douglas and T. Radó independently as a solution to Plateau's problem, admitting singularities. The result was improved by R. Courant for the case of finitely many boundary curves by the method of †Dirichlet integrals. The method was carried out further by C. B. Morrey for the generalized Plateau's problem in a Riemannian manifold (→ Section C (5)). Indeed, the Euclidean space was replaced by any complete Riemannian manifold which is metrically well behaved at infinity. For example, any compact or any †homogeneous Riemannian manifold is in this class.

The existence proof of minimal surfaces cannot in general be applied directly to the case of higher-dimensional minimal submanifolds. The notion of varifolds (→ Section G (2))

and the corresponding generalization of a minimal submanifold, called a minimal variety (→ Section G (2)), were then introduced by F. Almgren, who proved the existence of a minimal variety in a compact Riemannian manifold. It was also proved that a minimal variety is approximated by regular minimal submanifolds. This work has been developed as geometric measure theory by E. R. Reifenberg, W. H. Fleming, H. Federer, F. Almgren, and others (→ Section G).

The study of minimal surfaces given as the graph of a real-valued function of two variables leads to that of solutions of the minimal surface equation. In 1915, S. Bernshteĭn proved that a minimal surface $z=f(x,y)$ defined on the entire plane $\mathbf{R}^2$ must be a plane. Subsequently, this was generalized to the following problem: Is a minimal hypersurface $x_{n+1}=f(x_1,\ldots,x_n)$ in $\mathbf{R}^{n+1}$ defined on the entire space $\mathbf{R}^n$ a hyperplane? The answer turns out to be affirmative for $n\leqslant 7$ and negative for $n\geqslant 8$ (→ Section F (1)).

In general, when a bounded domain D in $\mathbf{R}^n$ and a continuous function φ on its boundary ∂D are given, the problem of finding a minimal hypersurface M defined by the graph of a real-valued function f on $\bar{D}$ with $f|\partial D=\varphi$ gives rise to a typical †Dirichlet problem. The basic questions are those of existence, uniqueness, and regularity of solutions. These were first studied by Radó for $n=2$ and later by L. Bers, R. Finn, H. Jenkins, J. Serrin, R. Osserman, and others (→ Section D (1)).

Minimal submanifolds of a sphere have interesting properties; some are analogous to those in Euclidean space but some are not. Among them, J. Simons, in 1968, gave a differential equation that the norm of the second fundamental form of a minimal submanifold in a sphere should satisfy. He then showed that a totally geodesic submanifold is isolated among the compact minimal submanifolds in the sphere. S. S. Chern, M. P. do Carmo, S. Kobayashi, N. Ejiri, T. Ito, T. Otsuki, N. Wallach, and others have made further contributions to this subject (→ Section F (2)).

B. Minimal Surfaces

(1) **Branched minimal surfaces**. A minimal surface M in $\mathbf{R}^n$ is an immersed surface with vanishing mean curvature vector field. M equipped with an atlas of †isothermal coordinates is viewed as a †Riemann surface. A **branch point** of a †harmonic mapping $f: M\to\mathbf{R}^n$ is a point at which the differential df_p is zero. A harmonic mapping $f: M\to\mathbf{R}^n$ of a Riemann surface M is called a **branched** (or **generalized) minimal immersion** if it is conformal except at the branch points, and the image $f(M)$ is then called a **branched minimal surface**. The solution to Plateau's problem given by Douglas and Radó is a branched minimal surface.

(2) **Maximum principle**. When $n=3$, the following maximum principle for minimal surfaces holds: If M_1 and M_2 are two connected branched minimal surfaces in $\mathbf{R}^3$ such that for a point $p\in M_1\cap M_2$, the surface M_1 locally lies on one side of M_2 near p, then M_1 and M_2 coincide near p.

(3) **Convex hull property**. In general, every branched minimal surface with boundary in $\mathbf{R}^n$ lies in the **convex hull** of its boundary curve, the smallest closed convex set containing the boundary.

(4) **Reflection principle**. If the boundary curve of a branched minimal surface contains a straight line γ, then the surface can be analytically continued as a branched minimal surface by reflection across γ. Based on this principle, the following holds: Let Γ be an analytic Jordan curve in $\mathbf{R}^n$ and $f: M\to\mathbf{R}^n$ a branched minimal immersion with boundary Γ. Then f is analytic up to the boundary, i.e., $f(M)$ is contained in the interior of a larger branched minimal surface (H. Lewy). The smooth version of this theorem was given by S. Hildebrandt: If $f: M\to\mathbf{R}^n$ is a branched minimal immersion with smooth boundary curve, then f is smooth up to the boundary.

(5) **Weierstrass-Enneper formula**. Every simply connected minimal surface in $\mathbf{R}^3$ is represented in the form

$$x_k(\zeta)=\mathrm{Re}\left\{\int_0^\zeta \phi_k(z)\,dz\right\}+c_k, \qquad k=1,2,3,$$

where $\phi_1=f(1-g^2)/2$, $\phi_2=\sqrt{-1}\,f(1+g^2)/2$, $\phi_3=fg$, and c_k is a constant. Here, $g(\zeta)$ is a meromorphic function on a domain D in the complex ζ-plane, and $f(\zeta)$ is an analytic function on D with the property that at each point ζ, where $g(\zeta)$ has a pole of order m, $f(\zeta)$ has a zero of order $2m$, D being either the unit disk or the entire plane. This formula is quite useful for constructing various minimal surfaces. For instance, on setting $f\equiv 1$ and $g(\zeta)=\zeta$, **Enneper's surface** is obtained:

$$(x_1,x_2,x_3)=\left(u-\frac{u^3}{3}+uv^2,\, v-\frac{v^3}{3}+vu^2,\, u^2-v^2\right), \qquad (u,v)\in\mathbf{R}^2.$$

Another feature of the formula is that general theorems about minimal surfaces can be obtained by translating statements about analytic functions into the corresponding minimal-surface ones. For example, a surface $f: M\to\mathbf{R}^3$ is minimal if and only if the †Gauss mapping

$G: M \to S^2$ is **antiholomorphic**, i.e., the complex conjugate of the mapping is holomorphic, when S^2 is viewed as the Riemann sphere with complex structure by the stereographic projection (from the south pole) onto the complex plane. As a consequence, a complete minimal surface in $\mathbf{R}^3$ is a plane or else the normals to the surface are everywhere dense in S^2 (R. Osserman).

(6) **Stability**. A minimal submanifold M is called **stable** if for every compact region on M all the second variations of the volume are positive. For instance, if $M \to \mathbf{R}^n$ is a minimal hypersurface and fv is a variation vector field, v being the unit normal vector field of M and f a smooth function with compact support on M, then the second variation formula for the volume $V(t)$ is given by

$$\left.\frac{d^2 V}{dt^2}\right|_{t=0} = \int_M (|\nabla f|^2 - |A|^2 f^2) dv_0,$$

where $|A|$ denotes the norm of the second fundamental form of M. In particular, when $n=3$, then $-|A|^2/2 = K$, the Gauss curvature of M. Therefore a minimal surface M in $\mathbf{R}^3$ is stable if and only if

$$\int_M (|\nabla f|^2 + 2Kf^2) dV > 0$$

for any smooth function f with compact support on M. It follows that the stability of minimal surface M in $\mathbf{R}^3$ is related to the boundary value problem of a linear elliptic operator $L = \Delta - 2K$, Δ being the Laplacian on M. Namely, a minimal surface M in $\mathbf{R}^3$ is stable if and only if the first eigenvalue $\lambda_1(D)$ of L on any bounded domain D in M is nonnegative.

In connection with the †Gauss mapping, a sufficient condition for a domain D in a minimal surface to be stable was obtained by H. Schwarz in 1885: If a minimal surface M in $\mathbf{R}^3$ has one-to-one Gauss mapping $G: M \to S^2$, then a domain $D \subset M$ such that $G(D)$ is contained in a hemisphere of S^2 is stable. This was generalized by J. L. Barbosa and M. do Carmo as follows: If the area $A(G(D))$ of the spherical image $G(D)$ is less than 2π, then D is stable. This result is sharp in the sense that for any $\varepsilon > 0$ there exists an unstable domain D with $A(G(D)) = 2\pi + \varepsilon$. As an analog to Bernshteĭn's theorem, the following holds: A complete and stable minimal surface in $\mathbf{R}^3$ is a plane. H. Mori has also obtained results in this regard.

As for the existence of unstable minimal surfaces, it is known that if there are two distinct stable minimal surfaces with the same smooth boundary curve Γ in $\mathbf{R}^3$, then there exists an unstable branched minimal surface with Γ as its boundary (M. Morse and C. Tompkins, M. Schiffman).

C. Plateau's Problem (→ 334 Plateau's Problem)

Let Γ be a †rectifiable †Jordan curve in $\mathbf{R}^n$ and $D = \{(x, y) \in \mathbf{R}^2 \mid x^2 + y^2 < 1\}$. Then there exists a continuous mapping $f: \bar{D} \to \mathbf{R}^n$ such that (a) $f \mid \partial D$ maps homeomorphically onto Γ; (b) $f \mid D$ is harmonic and **almost conformal**, i.e., $\langle f_x, f_y \rangle = 0$ and $|f_x| = |f_y|$ in D with $|df| > 0$ except at isolated branch points; (c) the induced area of f is the least among the class of piecewise smooth surfaces bounding Γ with (a). This mapping f is called the **classical solution** or the **Douglas-Radó solution** to Plateau's problem for Γ, which may have singularities called †branch points. The resulting surface S is a branched minimal disk. This theorem establishes the existence of a surface of least area among all surfaces homeomorphic to a disk. It has been generalized by R. Courant for Γ consisting of finitely many rectifiable Jordan curves.

For a branched minimal disk S bounded by smooth Γ in $\mathbf{R}^3$, there is a relation, called the **Gauss-Bonnet-Sasaki-Nitsche formula**, among the †total curvature $\kappa(\Gamma)$ of Γ, the total curvature of S, and the orders of branch points:

$$1 + \sum (m_\alpha - 1) + \sum M_\beta + \frac{1}{2\pi} \int_S |K| dA \leqslant \frac{1}{2\pi} \kappa(\Gamma),$$

where K denotes the Gauss curvature of S, $m_\alpha - 1$ the orders of the interior branch points, and $2M_\beta$ the orders of the boundary branch points, which must be even.

(1) **Regularity**. A minimal disk of least area in $\mathbf{R}^3$ has no boundary branch points when Γ is real analytic (R. Gulliver and F. Leslie) or when Γ is smooth and the total curvature $\kappa(\Gamma)$ is less than 4π (J. C. C. Nitsche). In general, a classical solution for smooth Γ in $\mathbf{R}^3$ cannot have infinitely many branch points. A remarkable fact is that every classical solution to Plateau's problem in $\mathbf{R}^3$ is free of branch points in its interior, i.e., is a regular immersion (R. Osserman).

(2) **Embeddedness**. A classical solution is not necessarily an embedding; it may have self-intersections. For instance, if Γ is knotted in $\mathbf{R}^3$, then every solution must have self-intersections. It is known, however, that immersed minimal disks of least area in $\mathbf{R}^3$ which can self-intersect only in their interiors are embeddings. In particular, if Γ is an **extremal** Jordan curve, i.e., if Γ lies on the boundary of its †convex hull, then the classical solution for Γ is an embedding (F. Tomi and A. J. Tromba; W. H. Meeks and S. T. Yau). If the topological type is not specified, then there always exists an embedded minimal surface. That is, if Γ is the union of any finite collection of disjoint

smooth Jordan curves in $\mathbf{R}^3$, then there exists a compact embedded minimal surface with boundary Γ which is smooth up to the boundary (R. Hardt and L. Simon). However, there exists an unknotted Jordan curve that never bounds an embedded minimal disk.

(3) **Uniqueness.** In general, the classical solutions to Plateau's problem do not have the uniqueness property. Radó was the first who gave a condition on a boundary curve to guarantee the uniqueness of the minimal disk. Namely, if a Jordan curve Γ in $\mathbf{R}^n$ admits a one-to-one orthogonal projection onto a convex curve in a plane $\mathbf{R}^2$ in $\mathbf{R}^n$, then the classical solution to Plateau's problem for Γ is free of branch points and can be expressed as the graph over this plane. When $n=3$, the solution is unique. Another geometric condition on Γ has been given by Nitsche: An analytic Jordan curve Γ in $\mathbf{R}^3$ with total curvature $\kappa(\Gamma)\leqslant 4\pi$ (or a smooth Γ with $\kappa(\Gamma)<4\pi$) bounds a unique immersed minimal disk. Moreover, the generic uniqueness holds: In the space $\mathscr{A}$ of all smooth Jordan curves in $\mathbf{R}^n$ with suitable topology, there exists an open and dense subset $\mathscr{B}$ such that for any Γ in $\mathscr{B}$ there exists a unique area-minimizing minimal disk (F. Morgan and A. J. Tromba).

(4) **Finiteness.** As for the finiteness of the classical solutions to Plateau's problem, several conditions on boundary curves are known. An analytic Jordan curve in $\mathbf{R}^3$ bounds only finitely many minimal disks of least area (F. Tomi). An analytic †extremal Jordan curve in $\mathbf{R}^3$ bounds only finitely many minimal disks with relative minima of area (H. Beeson). A smooth Jordan curve Γ with total curvature $\kappa(\Gamma)\leqslant 6\pi$ bounds only finitely many minimal disks (Nitsche). Moreover, generically, i.e., for an open and dense subset of boundary curves, there are at most finitely many minimal surfaces with given boundary, relative minima or not (R. Böhme and Tromba).

(5) **Generalized Plateau's problem.** C. B. Morrey's setting of the generalized Plateau's problem is as follows: A homotopically trivial rectifiable Jordan curve Γ is given in an n-dimensional Riemannian manifold N. Let D denote a disk in $\mathbf{R}^2$. Find a mapping $f:\bar{D}\rightarrow N$ such that (a) $f\,|\,\partial D$ maps homeomorphically onto Γ, (b) the induced area of f is the least among the class of piecewise smooth surfaces in N bounded by Γ with (a). Obviously when $N=\mathbf{R}^n$, this is the classical Plateau's problem. A solution was given by Morrey under the assumption that N is **homogeneously regular**; i.e., that there exist $0<k<K$ such that for any point $y\in N$ there is a local coordinate system (V,Φ) around y for which $\Phi(V)=\{x\in\mathbf{R}^n\,|\,\|x\|<1\}$ and the Riemannian metric $\sum g_{\alpha\beta}dx^\alpha dx^\beta$ satisfies

$$k\sum(v^\alpha)^2\leqslant\sum g_{\alpha\beta}v^\alpha v^\beta\leqslant K\sum(v^\alpha)^2$$

for any x and $(v_1,\ldots,v_n)\in\mathbf{R}^n$. Any compact or any homogeneous Riemannian manifold is homogeneously regular. Morrey's solution is as follows: If N is a homogeneously regular Riemannian manifold and if Γ is a homotopically trivial rectifiable Jordan curve in N, then there exists a branched minimal immersion f: $D\rightarrow N$ with least area bounded by Γ with (a).

The regularity of Morrey's solution is similar to the classical one. If Γ is smooth, then so is the solution f up to the boundary. If, furthermore, $\dim N=3$, then the solution f is an immersion in its interior. If N and Γ are real analytic and $\dim N=3$, then the solution f is an immersion up to the boundary.

D. Existence Problems of Minimal Submanifolds

(1) **The Dirichlet problem for the minimal surface equation.** When the graph of a (vector-valued) function is a minimal submanifold in some Euclidean space, the function must satisfy the so-called †minimal surface equation. To be precise, let D be a (bounded) domain in $\mathbf{R}^n$, $f:D\rightarrow\mathbf{R}^k$ a (vector-valued) function, and put $F:D\rightarrow\mathbf{R}^{n+k}:F(x)=(x,f(x))$ for $x\in D$. Let M be the graph of $f: M=F(D)$. Then M is minimal if and only if f (or F) satisfies one of the following equations:

if $n=2$ and $k=1$,

$$(1+f_y^2)f_{xx}-2f_xf_yf_{xy}+(1+f_x^2)f_{yy}=0; \qquad (1)$$

if n is arbitrary and $k=1$,

$$\sum_{i=1}^{n}\frac{\partial}{\partial x^i}\left(\frac{1}{W}\frac{\partial f}{\partial x^i}\right)=0 \quad \text{where} \quad W=\sqrt{1+|\nabla f|^2}; \qquad (2)$$

if $n=2$ and k is arbitrary,

$$(1+|f_y|^2)f_{xx}-2(f_x\cdot f_y)f_{xy}+(1+|f_x|^2)f_{yy}=0; \quad (3)$$

and if n and k are arbitrary,

$$\sum_{i,j=1}^{n}g^{ij}\frac{\partial^2 f}{\partial x^i\partial x^j}=0 \quad \text{where} \quad g_{ij}=\frac{\partial F}{\partial x^i}\frac{\partial F}{\partial x^j}, \qquad (4)$$

and (g^{ij}) is the inverse matrix of (g_{ij}).

The basic problem for this class of equations is the †Dirichlet problem. For $n=2$ and $k=1$, the following theorem of Radó and Finn is fundamental: There exists a solution f of the Dirichlet problem corresponding to an arbitrary continuous function φ on the boundary ∂D of D if and only if D is convex. Since the difference of any two solutions of eq. (1) satisfies the maximum principle, there can be at most one solution f of the Dirichlet problem

corresponding to a given continuous function φ on ∂D. As for the removability of isolated singularities, the following is known: Every solution of eq. (1) in $0<x^2+y^2<1$ extends continuously to the origin. The extended function is smooth at the origin and satisfies eq. (1) in the full disk $x^2+y^2<1$ (Finn).

For arbitrary n and $k=1$, namely, for the case of minimal hypersurfaces, the following is known: Let D be a bounded domain in $\mathbf{R}^n$ with smooth boundary. The Dirichlet problem for eq. (2) has a solution for any continuous function φ on ∂D if and only if the mean curvature of ∂D with respect to the inner normal is nonnegative at every point. As for the uniqueness and regularity, the same results as above hold (Jenkins and Serrin).

For $n=2$ and arbitrary k, namely, for the case of minimal surfaces in $\mathbf{R}^{2+k}$, exactly the same result of existence as in the classical case $k=1$ holds. A solution exists for an arbitrary continuous vector-valued function on ∂D if and only if D is convex (Osserman). Though the removability of singularities holds under suitable restrictions, the uniqueness fails in this case.

For the last case, $n>2$ and $k>1$, there are essentially no results on either existence or uniqueness for the Dirichlet problem.

(2) **Existence of minimal surfaces and topology**. For a compact Riemannian manifold N, there are results on the existence of minimal surfaces related to the homotopy groups of N. Let f be a continuous mapping from a †Riemann surface Σ_g of †genus g into N. If the †induced mapping $f_\#:\pi_1(\Sigma_g)\rightarrow\pi_1(N)$ of f on the †fundamental groups is injective, then there exists a branched minimal immersion $h:\Sigma_g\rightarrow N$ such that $h_\#=f_\#$ on $\pi_1(\Sigma_g)$ and the induced area of h is the least among all mappings with the same action on $\pi_1(\Sigma_g)$. If furthermore $\pi_2(N)=0$, then h can be deformed from f continuously (R. Schoen and S. T. Yau). If $\pi_2(N)\neq 0$, then there exists a generating set for $\pi_2(N)$ consisting of branched minimal immersions of spheres that minimize energy and area in their homotopy classes (J. Sacks and K. Uhlenbeck). If, furthermore, $\dim N=3$, then the above minimal immersion h in the homotopy class is an embedding or a two-to-one covering mapping. In the second case, its image is an embedded real projective plane. If h' is another such least-area mapping, then either h' is equal to h up to a conformal reparametrization or the images of h and h' are disjoint (Meeks and Yau).

E. Gauss Mapping of Minimal Surfaces

On a connected, orientable surface M there exist local isothermal coordinates (x,y) at each point. On putting $z=x+\sqrt{-1}\,y$, the metric has the form $ds^2=2F|dz|^2$, and M is viewed as a Riemann surface. A conformal immersion $f:M\rightarrow\mathbf{R}^n$ is minimal if and only if $\Delta f=0$, or equivalently $\partial_z\partial_{\bar{z}}f=0$. In the case $n=3$, the †Gauss mapping $G:M\rightarrow S^2$ of a surface f: $M\rightarrow\mathbf{R}^3$ is defined by assigning to each point $p\in M$ the unit normal vector translated parallel to the origin of $\mathbf{R}^3$. A surface $f:M\rightarrow\mathbf{R}^3$ is minimal if and only if the Gauss mapping is antiholomorphic. In the general case, the Gauss mapping G is defined to be a mapping assigning to each point $p\in M$ the oriented tangent space $f_*T_p(M)\subset\mathbf{R}^n$. Thus G is a mapping from M into the †Grassmann manifold $\tilde{M}_{n,2}(\mathbf{R})=SO(n)/SO(2)\times SO(n-2)$ of oriented planes in $\mathbf{R}^n$, which is naturally diffeomorphic to the †complex quadric Q_{n-2} in the complex projective space $\mathbf{C}P^{n-1}$. An immersion f: $M\rightarrow\mathbf{R}^n$ is minimal if and only if its Gauss mapping is antiholomorphic.

Let $f:M\rightarrow\mathbf{R}^n$ be a complete orientable minimal surface. Let χ be the Euler characteristic and $-\pi C=\int_M K\,dA$, the total curvature of $f(M)$. Assume that the total curvature is finite. Then the following results of Chern and Osserman are fundamental: (a) M is conformally a compact Riemann surface $\overline{M}$ with finite number, say r, of points deleted; (b) C is an even integer and satisfies $C\geqslant 2(r-\chi)=4g+4r-4$, where g is the genus of M ($=$genus of $\overline{M}$); (c) if $f(M)$ does not lie in any proper affine subspace of $\mathbf{R}^n$, then $C\geqslant 4g+r+n-3\geqslant 4g+n-2\geqslant n-2$ (F. Gackstatter); (d) if $f(M)$ is simply connected and nondegenerate, i.e., its image under the Gauss mapping does not lie in a hyperplane of $\mathbf{C}P^{n-1}$, then $C\geqslant 2n-2$ and this inequality is sharp; (e) when $n=3$, C is a multiple of 4, with the minimum value 4 attained only for Enneper's surface and the catenoid; (f) the Gauss mapping G of M extends to a mapping of $\overline{M}$ whose image $G(\overline{M})$ is an algebraic curve in $\mathbf{C}P^{n-1}$ lying in Q_{n-2}; the total curvature of $f(M)$ is equal in absolute value to the area of $G(\overline{M})$, counting multiplicity; (g) $G(M)$ intersects a fixed number of times, say m (counting multiplicity), every hyperplane in $\mathbf{C}P^{n-1}$ except for those hyperplanes containing any of the finite number of points of $G(\overline{M}-M)$; the total curvature of $f(M)$ equals $-2\pi m$.

In particular, if a complete minimal surface in $\mathbf{R}^n$ has k †ends, then the total curvature never exceeds $2\pi(\chi-k)$. Enneper's surface and the catenoid are the only two complete minimal surfaces in $\mathbf{R}^3$ whose Gauss mapping is one-to-one.

If the Gauss mapping of a complete minimal surface of finite total curvature in $\mathbf{R}^3$ omits more than 3 points of S^2, then it is a plane (Osserman). F. Xavier proved that the Gauss mapping of any complete nonflat minimal

surface in $\mathbf{R}^3$ can omit at most 6 points of S^2. It is an open question whether this can be improved to 4 points. The Gauss mapping of Scherk's surface omits exactly 4 points of S^2.

F. Minimal Submanifolds

(1) **Bernshteĭn's problem.** As was mentioned in Section A, a minimal surface in $\mathbf{R}^3$ represented as the graph of a function on the entire plane $\mathbf{R}^2$ is a plane. This is known as **Bernshteĭn's theorem** (1915). This is, however, not true in $\mathbf{R}^4$, namely, there is a minimal surface defined as the graph of a vector-valued function on the entire plane, which is not a plane. The **generalized Bernshteĭn problem** is stated as follows: Let $f:\mathbf{R}^n\rightarrow\mathbf{R}$ be a function satisfying the minimal hypersurface equation

$$\sum_{i=1}^{n}\frac{\partial}{\partial x^i}\left(\frac{1}{W}\frac{\partial f}{\partial x^i}\right)=0,$$

where $W=\sqrt{1+|\nabla f|^2}$. Is the graph of f an affine hyperplane? The answer is affirmative for $n\leqslant 7$ (E. De Giorgi, F. J. Almgren, J. Simons). For $n\geqslant 8$, it is negative (E. Bombieri, De Giorgi, E. Giusti). Though the problem turns out to be negative for $n\geqslant 8$, no concrete counterexample is known, even for $n=8$.

(2) **Minimal submanifolds in the sphere.** Minimal submanifolds in the unit sphere S^n have some special properties. For example, there is no compact minimal submanifold of S^n contained in an open hemisphere. On the other hand, there is no stable minimal closed submanifold in S^n (B. Lawson, Simons). More generally, any closed minimal hypersurface in a Riemannian manifold with positive Ricci curvature is unstable (Simons). As for the existence of minimal surfaces in a sphere, the following theorem of Lawson is general: Any closed surface of any genus, except the real projective plane, can be minimally immersed into the unit sphere S^3. As an analog to Bernshteĭn's theorem, a sphere S^2, which is minimally immersed in S^3, is necessarily totally geodesic. On the other hand, by making use of higher-order second fundamental forms and the relations among them, a complete description of the minimal immersions of $\mathbf{R}^2$ into S^n has been given by K. Kenmotsu.

As a general theorem which gives a necessary and sufficient condition for a Riemannian manifold to be minimally immersed in the unit sphere S^n, the following is fundamental: An isometric immersion x of an m-dimensional Riemannian manifold M into S^n, viewed as a vector-valued function into $\mathbf{R}^{n+1}$, is minimal if and only if $\Delta x=mx$ (T. Takahashi). As its immediate application, any m-dimensional compact homogeneous space whose linear isotropy group is irreducible can be minimally immersed into the n-sphere of curvature λ/m, corresponding to any eigenvalue ($\neq 0$) of the Laplacian, where $n+1$ is the dimension of the eigenspace corresponding to λ (W. Y. Hsiang, Takahashi).

Since for the †symmetric spaces of †rank 1 the eigenvalues and the corresponding eigenspaces of the Laplacian are known, the minimal immersions of such spaces into the unit sphere have been determined as well as the †rigidity of such minimal immersions (N. Wallach). For example, if the m-sphere of constant curvature c is minimally immersed into the unit n-sphere, but not contained in any great hypersphere, then for each nonnegative integer k, $c=m/k(k+m-1)$ and $n+1\leqslant(2k+m-1)\cdot[(k+m-2)!/k!(m-1)!]$. The immersion is rigid if and only if $m=2$ or $k\leqslant 3$. Similar results have been known for the projective spaces over real numbers, complexes, quaternions, or Cayley numbers.

Simons showed that the scalar curvature ρ of a compact minimal submanifold M of dimension m in the unit $(m+p)$-sphere is not greater than $m(m-1)$. Furthermore, if $\rho\geqslant m(m-1)-mp/(2p-1)$, then either $\rho=m(m-1)$ or $\rho=m(m-1)-mp/(2p-1)$. In the former case, M is totally geodesic and therefore is the unit sphere S^m. In the latter case, M is either a hypersurface of the unit $(m+1)$-sphere which is isometric to the product $S^k(\sqrt{k/m})\times S^{m-k}(\sqrt{(m-k)/m})$ of spheres of radius $\sqrt{k/m}$ and $\sqrt{(m-k)/m}$, respectively, called the **generalized Clifford torus** or the **Veronese surface** in S^4. Here the **Clifford torus** is the torus $S^1(\sqrt{1/2})\times S^1(\sqrt{1/2})\subset S^3\subset\mathbf{R}^4$, and the Veronese surface is defined as follows: Let (x,y,z) be the natural coordinate system in $\mathbf{R}^3$ and (u^1,u^2,u^3,u^4,u^5) that in $\mathbf{R}^5$. The mapping defined by

$$u^1=yz/\sqrt{3},\quad u^2=zx/\sqrt{3},\quad u^3=xy/\sqrt{3},$$

$$u^4=(x^2-y^2)/2\sqrt{3},\quad u^5=(x^2+y^2-2z^2)/6$$

gives an isometric immersion of $S^2(\sqrt{3})$ into S^4. Two points (x,y,z) and $(-x,-y,-z)$ of $S^2(\sqrt{3})$ are mapped into the same point of S^4, and thus the mapping defines an embedding of the real projective plane $\mathbf{R}P^2$ into S^4. This embedded real projective plane in S^4 is called the Veronese surface.

T. Otsuki proved the following: Let M be a complete minimal hypersurface immersed in the unit $(n+1)$-sphere with two principal curvatures. If their multiplicities are m and $n-m\geqslant 2$, then M is congruent to the generalized Clifford torus $S^m(\sqrt{m/n})\times S^{n-m}(\sqrt{(n-m)/n})\subset S^{n+1}\subset\mathbf{R}^{n+2}$. If one of their multiplicities is 1, then M is a hypersurface of S^{n+1} in $\mathbf{R}^{n+2}=\mathbf{R}^n\times\mathbf{R}^2$ whose orthogonal projection into $\mathbf{R}^2$

is a curve of which the †support function $x(t)$ is a solution of the following nonlinear differential equation of the second order:

$$nx(1-x^2)(d^2x/dt^2)+(dx/dt)^2+(1-x^2)(nx^2-1)$$
$$=0.$$

Furthermore, there are countably many compact minimal hypersurfaces immersed but not embedded in S^{n+1}. Only $S^{n-1}(\sqrt{(n-1)/n})\times S^1(\sqrt{1/n})$ is minimally embedded in S^{n+1}. This just corresponds to the trivial solution $x(t)\equiv 1/\sqrt{n}$ of the above equation.

(3) **Minimal submanifolds in Riemannian manifolds.** In general, it is difficult to determine all the complete stable minimal submanifolds in a given Riemannian manifold. If, however, some curvature conditions are given to the Riemannian manifold, then a classification has been given: Let N be a complete 3-dimensional Riemannian manifold with nonnegative scalar curvature ρ, and let M be a complete, stable, and orientable minimal surface in N.

(a) If M is compact, then either M is conformally equivalent to the †Riemann sphere S^2 or else it is a totally geodesic flat torus. Furthermore, if $\rho>0$, then the latter case never occurs.

(b) If M is not compact, then M is conformally equivalent to the complex plane $\mathbf{C}$ or a cylinder (D. Fischer Colbrie; R. Schoen).

As a generalization of the procedure for generating periodic minimal surfaces in $\mathbf{R}^3$ with octahedral or tetrahedral symmetry (Schwarz, 1867), T. Nagano and B. Smyth gave a construction procedure to generate periodic minimal surfaces in $\mathbf{R}^n$ or n-tori with symmetry corresponding to any †Weyl group of rank n.

G. Minimal Varieties

Recent progress in the study of Plateau's problem in higher dimensions is closely related to the point of view of geometric measure theory and †the calculus of variations.

(1) **Integral currents.** Let T be a †current of degree k, or simply a k-current, defined on an open set U of $\mathbf{R}^n$, and let ∂T be the boundary of T, the exterior derivative of T. For simplicity, in the following only currents with compact support are considered.

The **mass** $M(T)$ of a k-current T is the dual norm $M(T)=\sup\{T(\varphi)\mid M(\varphi)\leqslant 1\}$ of the **comass** $M(\varphi)$ on k-forms, which is defined by $M(\varphi)=\sup\{\|\varphi(x)\|\mid x\in U\}$ with $\|\varphi(x)\|=\sup\{\langle\varphi(x),v_1\wedge\ldots\wedge v_k\rangle\mid v_1,\ldots,v_k$ are orthonormal vectors at $x\in U\}$. It follows that $M(T)$ coincides with k-dimensional volume when T is a k-current defined by integration on a k-dimensional simplex. We say that T is **normal** if $M(T)$ and $M(\partial T)$ are finite. Normal currents form a †Banach space with norm $N(T)=M(T)+M(\partial T)$.

A current T is called **rectifiable** if it can be approximated in the mass norm by currents of type $f_*\gamma$, where γ is a finite polyhedral chain with integer coefficients, $f:\operatorname{supp}\gamma\rightarrow U$ is a †Lipshitz mapping of the support of γ into U, and $f_*\gamma$ is the current defined by means of $f_*\gamma(\varphi)=\gamma(f^*\varphi)$. If both T and ∂T are rectifiable, T is called an **integral current**. Integral currents give the appropriate notion of generalized manifolds with boundary in the study of higher-dimensional Plateau problems.

One of the fundamental properties of integral currents is the compactness, stated as follows: Given a compact subset $K\subset U$ and a number $c>0$, the set of integral currents T such that $\operatorname{supp}T\subset K$ and $N(T)\leqslant c$ is sequentially compact in the †weak topology. Since the mass $M(T)$ is †lower semicontinuous in the weak topology, it follows that Plateau's problem can always be solved in the space of integral currents (→ Section C; 334 Plateau's Problem). The question arises: To what extent is this a satisfactory solution to the problem? For codimension 1, it is known from the work of Federer and others that an integral $(n-1)$-current of least mass in $\mathbf{R}^n$ is nonsingular in codimension <7. In particular, in $\mathbf{R}^n$, $n\leqslant 7$, every integral $(n-1)$-current of least mass is an analytic manifold. In general codimensions, it is known that the set of regular points is dense (Reifenberg, Almgren, and others).

The definition of rectifiable and integral currents carries over to those on a Riemannian manifold M in a natural way. Then the space of integral currents $I_*(M)$ on M with the boundary operator ∂, for which $\partial^2=0$, form a chain complex. It is then a fundamental theorem, due to Federer and Fleming, in homological integration theory that there is a natural isomorphism $H_*(I_*(M))\cong H_*(M;\mathbf{Z})$ of the homology of the complex of integral currents with the integral singular homology groups of M. From this it follows that if M is a compact Riemannian manifold M, then each class $\alpha\in H_p(I_*(M))\cong H_p(M;\mathbf{Z})$ contains an integral current of least mass among all integral currents in α.

For a homology class of codimension 1, it has also been proved by Almgren that if M is a compact Riemannian manifold of dimension $\leqslant 7$, then for each $\alpha\in H_{n-1}(M;\mathbf{Z})$ there exists a finite collection of mutually disjoint, compact oriented minimal hypersurfaces $S_1,\ldots,S_l$ embedded in M and integers $n_1,\ldots,n_l$ such that the integral current $\sum_{j=1}^{l} n_j s_j$ represents α. On the other hand, it has been recently proved

that every compact Riemannian manifold of dimension $\leqslant 6$ contains a nonempty closed embedded minimal hypersurface (J. T. Pitts).

(2) **Varifolds.** The theory of integral currents developed by Federer, Fleming, and others yields reasonable spaces for purposes of the calculus of variations. However, it is not entirely feasible to use these in the study of the actual soap films that occur in physical experiments. It turns out that we should work in a more set-theoretic fashion and give up the notion of orientability and the boundary operator. A convenient theory for describing these physical experiments is the theory of varifolds developed by Almgren.

A k-dimensional **varifold** on a Riemannian manifold M is a †Radon measure on the bundle $G_k(M)$ of unoriented tangent k-planes of M. For simplicity, only varifolds with compact support are considered. These are regarded as continuous linear functionals on the space $C(G_k(M))$ of continuous functions on $G_k(M)$. Thus a k-dimensional submanifold S of finite volume embedded in M determines a varifold V_S by integration:

$$V_S(f)=\int_S f(T_xS)\,d\mathscr{H}^k(x), \qquad f\in C(G_k(M)),$$

where $\mathscr{H}^k$ denotes the k-dimensional †Hausdorff measure on M. More generally, to any rectifiable current there corresponds an underlying varifold obtained by neglecting orientations.

Let $V_k(M)$ denote the space of k-dimensional varifolds on M. Given a varifold $V\in V_k(M)$, the **mass** $M(V)$ is defined to be the V-measure of the total space, i.e., $M(V)=V(1)$. Let $f: M\rightarrow M$ be a C^1-mapping. Then f induces naturally a mapping f_* of $V_k(M)$ into itself. In particular, if X is a smooth vector field on M with associated flow f_t, then there is a smooth function $M(t)=M(f_{t_*}V)$. A varifold $V\in V_k(M)$ is then called a k-dimensional **minimal variety** in M if the first variation $(d/dt)M(t)|_{t=0}=0$ for all smooth vector fields on M. For example, minimal submanifolds, complex analytic subvarieties of Kähler manifolds and integral currents of least mass are minimal varieties.

An appropriate notion of rectifiable and integral varifolds is defined, and an analog of the compactness theorem can be obtained. As a consequence, it was proved by Almgren using †Morse theory methods that if M is an n-dimensional compact Riemannian manifold, then for each p, $1\leqslant p\leqslant n-1$, there exists at least one minimal variety of dimension p. As for the regularity of minimal varieties, it is known that if V is a k-dimensional variety, then in the support of V there is a relatively open dense subset that is a regular minimal submanifold of dimension k (W. K. Allard).

H. Maximal Hypersurfaces in Minkowski Space

Let L^{n+1} be a Minkowski space, i.e., $L^{n+1}=\{(x_1,\dots,x_n,t)\,|\,(x_1,\dots,x_n)\in\mathbf{R}^n, t\in\mathbf{R}\}$ with the †Lorentzian metric $\sum_{i=1}^n (dx_i)^2-(dt)^2$. Let M be a †spacelike hypersurface of L^{n+1} so that the induced metric on M is Riemannian. If the mean curvature vector field H of M, defined in the same way as in the Riemannian case, vanishes identically, then M is called **maximal**. In contrast to the Riemannian case, by the variation in the normal direction of H, the volume increases, provided $H\neq 0$. However, the equation describing a maximal hypersurface is similar to that for the Riemannian case (→ Sections D (1), F (1)). Indeed, let D be a domain in $\mathbf{R}^n$ and $f: D\rightarrow\mathbf{R}$ a smooth function. Then the graph of f in L^{n+1} is a maximal hypersurface if and only if f satisfies the following quasilinear partial differential equation of the second order:

$$(1-|\nabla f|^2)\sum_{i=1}^n\frac{\partial^2 f}{\partial x_i^2}+\sum_{i,j=1}^n\frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j}\frac{\partial^2 f}{\partial x_i\partial x_j}=0.$$

This is †elliptic for $|\nabla f|<1$, since M is spacelike. E. Calabi proposed a problem similar to Bernshteĭn's in 1968. Contrary to Bernshteĭn's case, the answer is affirmative for any n. Namely, a maximal hypersurface in L^{n+1} defined as the graph of a function on the entire space $\mathbf{R}^n$ is a hyperplane in $\mathbf{R}^{n+1}$. More generally, a maximal hypersurface, which is a closed set in L^{n+1}, is a hyperplane (S. Y. Cheng and S. T. Yau). It is also known that any maximal hypersurface in L^{n+1} is stable.

References

[1] F. J. Almgren, Plateau's problem, An invitation to varifold geometry, Benjamin, 1966.
[2] R. Courant, Dirichlet's principle, conformal mapping, and minimal surfaces, Interscience, 1950.
[3] H. Federer, Geometric measure theory, Springer, 1969.
[4] H. Federer, Colloquium lectures on geometric measure theory, Bull. Amer. Math. Soc., 84 (1978) 291–338.
[5] D. Hoffman and R. Osserman, The geometry of the generalized Gauss map, Mem. Amer. Math. Soc., 236 (1980).
[6] H. B. Lawson, Minimal varieties, Proc. Symp. Pure Math., 27 (1975), 143–175.
[7] H. B. Lawson, Lectures on minimal submanifolds I, Publish or Perish, 1980.
[8] C. B. Morrey, Multiple integrals in the calculus of variations, Springer, 1966.

[9] J. C. C. Nitsche, On new results in the theory of minimal surfaces, Bull. Amer. Math. Soc., 71 (1965), 195–270.
[10] J. C. C. Nitsche, Vorlesungen über Minimalflächen, Springer, 1975.
[11] R. Osserman, Minimal varieties, Bull. Amer. Math. Soc., 75 (1969), 1092–1120.
[12] R. Osserman, A survey of minimal surfaces, Van Nostrand, 1969.
[13] T. Radó, On the problem of Plateau, Springer, 1933.
[14] A. J. Tromba, On the number of simply connected minimal surfaces spanning a given curve, Mem. Amer. Math. Soc., 194 (1977).
[15] Minimal submanifolds and geodesics, Proc. Japan-US Seminar, Tokyo, 1977, North-Holland, 1979.
[16] S. Nishikawa, On maximal spacelike hypersurfaces in a Lorentzian manifold, Nagoya Math. J., 95 (1984), 117–124.
[17] O. Kobayashi, Maximal surfaces with conelike singularities, J. Math. Soc. Japan, 36 (1984), 609–617.

276 (I.6) Model Theory

A. Language

Every mathematical theory has an appropriate language. To determine a language for a theory means to determine a language for the related mathematical system. Such a language consists of the following symbols (the symbols given here are examples of only one notational system).

(1) Symbols that express logical concepts (†logical symbols): $\forall$, $\exists$, $\neg$, $\wedge$, $\vee$, $\rightarrow$;

(2) †free variables: $a_0, a_1, a_2, \ldots$;

(3) †bound variables: $x_0, x_1, x_2, \ldots$;

(4) symbols that denote individual objects (individual constants): $c_0, c_1, c_2, \ldots, c_\alpha, \ldots$;

(5) †function symbols: $f_0, f_1, f_2, \ldots, f_\alpha, \ldots$;

(6) †predicate symbols: $P_0, P_1, P_2, \ldots, P_\alpha, \ldots$.

The †cardinalities of the sets of symbols in (4), (5), and (6) are arbitrary, except that there must be at least one predicate symbol. It is assumed that each set of symbols is †well ordered. Also, it is understood that to each f_j in (5) there corresponds a positive integer i_j, while to each P_j there corresponds a nonnegative integer (these integers are called the number of arguments of f_j and P_j, respectively).

In practice, other kinds of languages are also dealt with. One example is a system with infinitely long expressions, which permits †transfinite ordinals for the numbers of arguments of f_j and P_j, and which has the extended concepts $\bigwedge_{\alpha<\beta}$, $\bigvee_{\alpha<\beta}$, $\exists x_1 \ldots \exists x_\alpha \ldots$ for $\alpha<\beta$, and $\forall x_1 \forall x_2 \ldots \forall x_\alpha \ldots$ for $\alpha<\beta$, where α and β are transfinite ordinals. Another language includes variables of higher †type as well as $\forall$ and $\exists$ over those variables. Free and bound variables may not be distinguished typographically. In that case a variable not bound by $\forall$ or $\exists$ in a †formula is called **free**. (The notion of a formula is defined later.) To simplify this discussion, however, we restrict ourselves to the †first-order predicate language with a typographic distinction between free and bound variables. We also assume that there are only a countable number of variables, and hence we use natural numbers as subscripts.

Set $L_1=\{\text{logical symbols}\}$, $L_2=\{a_0,a_1,a_2,\ldots\}$, $L_3=\{x_0,x_1,x_2,\ldots\}$, $L_4=\{c_0,c_1,c_2,\ldots\}$, $L_5=\{f_0,f_1,f_2,\ldots\}$, $L_6=\{P_0,P_1,P_2,\ldots\}$, $L=\langle L_1,L_2,L_3,L_4,L_5,L_6\rangle$. To determine a language L is to specify such a list L. Since $\forall$, $\exists$, $\neg$, $\wedge$, $\vee$, $\rightarrow$, are normally used for L_1, $a_0, a_1, a_2, \ldots$ for L_2, and $x_0, x_1, x_2, \ldots$ for L_3, we may assume that these are fixed, and hence to determine a language is to determine $\langle L_4,L_5,L_6\rangle$. We take $\langle L_1,L_2,L_3\rangle$ just described and assume an arbitrary but fixed $L=\langle L_4,L_5,L_6\rangle$. First we define the notions term of L (or L-term) and formula of L (or L-formula).

Definition of the **terms** of L (L-terms): (1) Each free variable a_j is an L-term. (2) Each individual constant c_j of L is an L-term. (3) If f_j is a function symbol of L, i_j is the number of arguments of f_j, and each of $t_1, \ldots, t_{i_j}$ is an L-term, then $f_j(t_1, \ldots, t_{i_j})$ is also an L-term. (4) The L-terms are only those constructed by (1)–(3).

A term that does not contain a free variable is called a **closed term**.

Definition of the **formulas** of L (or L-formulas): (1) Let P_j be a predicate symbol of L and i_j be the corresponding natural number. If each of $t_1, \ldots, t_{i_j}$ is an L-term, then $P_j(t_1, \ldots, t_{i_j})$ is an L-formula. This type of formula is called a **prime formula** (or **atomic formula**). (2) If A and B are L-formulas, then each of $\neg(A)$, $(A)\wedge(B)$, $(A)\vee(B)$, and $(A)\rightarrow(B)$ is an L-formula. (3) Let F be an L-formula and x_i be a bound variable that does not occur in F. Then an expression obtained by putting ( ) around F, replacing some occurrences in F of a free variable, say a_j, by x_i, and prefixing $\forall x_i$ or $\exists x_i$ is an L-formula. (4) The L-formulas are only those constructed by (1)–(3).

A formula that has no occurrence of a free variable is called a **closed formula**. The parentheses used in the formation of a formula may be omitted if no ambiguity arises thereby.

B. Structures

Let L be a specific language as described in the previous section. Then $\mathfrak{M}=[M:\rho;\sigma;\tau]$ defined by (1)–(4) below is called a **structure** for L (or L-structure).

(1) M is a nonempty set. (M is called the **universe** of $\mathfrak{M}$.)

(2) ρ is a mapping from L_4 into M.

(3) Let $L_5^i=\{f_j \mid$ the number of arguments of f_j is $i\}$. Then $L_5=L_5^1\cup L_5^2\cup\ldots\cup L_5^i\cup\ldots$ provides a partition of L_5. Let $\mathfrak{F}_i$ be the set of all mappings from $M^i=M\times\ldots\times M$ (i times) into M and σ_i be a mapping from L_5^i into $\mathfrak{F}_i$. We define σ for an arbitrary f of L_5 by $\sigma(f)=\sigma_i(f)$, where i is the number of arguments of f. Then σ is obviously a mapping from L_5 into $\bigcup_{i=1}^{\infty}\mathfrak{F}_i$.

(4) Decompose L_6 into $L_6^0\cup L_6^1\ldots\cup L_6^i\cup\ldots$ as in (3). Let P_i be the set of all subsets of M^i and τ_i be a mapping from L_6^i into P_i, where P_0 is the set $\{M,\varnothing\}$ ($\varnothing$ is the empty set). Then τ is defined for every i and for an arbitrary P of L_6^i by $\tau(P)=\tau_i(P)$.

If we denote $\rho(c)$ by $\bar{c}$, $\sigma(f)$ by $\bar{f}$, and $\tau(P)$ by $\bar{P}$, then we may understand that ρ is represented by $\bar{c}_0,\bar{c}_1,\ldots,\sigma$ is represented by $\bar{f}_0$, $\bar{f}_1,\ldots,$and τ is represented by $\bar{P}_0,\bar{P}_1,\ldots.$ Therefore $\mathfrak{M}$ is normally expressed as

$$\mathfrak{M}=[M:\bar{c}_0,\bar{c}_1,\ldots;\bar{f}_0,\bar{f}_1,\ldots;\bar{P}_0,\bar{P}_1,\ldots].$$

C. Satisfiability

We fix not only a language L but also a structure $\mathfrak{M}$ for L. Then the property that an L-formula is **satisfiable** is defined by the following procedure:

Let $\mathfrak{m},\mathfrak{n},\ldots$ stand for †sequences of the elements of M, say $(m_0,m_1,\ldots),(n_0,n_1,\ldots),\ldots,$ called $\mathfrak{M}$-sequences. We write $\mathfrak{m}\overset{i}{=}\mathfrak{n}$ to indicate that each entry of $\mathfrak{m}$ except the ith one is equal to the corresponding entry of $\mathfrak{n}$. Using these concepts, the value of an L-term at an $\mathfrak{M}$-sequence $\mathfrak{m}$, denoted by $t[\mathfrak{m}]$, is defined as follows:

(1) If t is a free variable a_j, then $t[\mathfrak{m}]=m_j$.

(2) If t is an individual constant c_j, then $t[\mathfrak{m}]=\bar{c}_j$.

(3) If t is of the form $f_j(t_1,\ldots,t_i)$, then $t[\mathfrak{m}]=\bar{f}_j(t_1[\mathfrak{m}],\ldots,t_i[\mathfrak{m}])$. If t is an L-term, then evidently $t[\mathfrak{m}]$ is an element of M.

Based on this definition of $t[\mathfrak{m}]$, the relation A is satisfiable by $\mathfrak{m}$ in $\mathfrak{M}$, denoted by $\mathfrak{M},\mathfrak{m}\models A$, is defined for an arbitrary L-formula A and an arbitrary $\mathfrak{M}$-sequence $\mathfrak{m}$ as follows:

(1) $\mathfrak{M},\mathfrak{m}\models P_j(t_1,\ldots,t_i)\Leftrightarrow\langle t_i[\mathfrak{m}],\ldots,t_i[\mathfrak{m}]\rangle\in\bar{P}_j$.

(2) $\mathfrak{M},\mathfrak{m}\models\neg B\Leftrightarrow\mathfrak{M},\mathfrak{m}\models B$ is false.

(3) $\mathfrak{M},\mathfrak{m}\models B\wedge C\Leftrightarrow\mathfrak{M},\mathfrak{m}\models B$ and $\mathfrak{M},\mathfrak{m}\models C$.

(4) $\mathfrak{M},\mathfrak{m}\models B\vee C\Leftrightarrow\mathfrak{M},\mathfrak{m}\models B$ or $\mathfrak{M},\mathfrak{m}\models C$.

(5) $\mathfrak{M},\mathfrak{m}\models B\rightarrow C\Leftrightarrow\mathfrak{M},\mathfrak{m}\models B$ implies $\mathfrak{M},\mathfrak{m}\models C$.

(6) $\mathfrak{M},\mathfrak{m}\models\forall x_jF(x_j)\Leftrightarrow\mathfrak{M},\mathfrak{n}\models F(a_i)$ for an arbitrary $\mathfrak{n}$ that satisfies $\mathfrak{m}\overset{i}{=}\mathfrak{n}$, where a_i has the least index among the free variables that do not occur in $F(x_j)$.

(7) $\mathfrak{M},\mathfrak{m}\models\exists x_jF(x_j)\Leftrightarrow$there exists an $\mathfrak{n}$ such that $\mathfrak{n}\overset{i}{=}\mathfrak{m}$ and $\mathfrak{M},\mathfrak{n}\models F(a_i)$, where a_i satisfies the same condition as in (6).

Following are some consequences of this definition.

(1) For an arbitrary L-formula A and an arbitrary $\mathfrak{M}$-sequence $\mathfrak{m}$, exactly one of $\mathfrak{M},\mathfrak{m}\models A$ and $\mathfrak{M},\mathfrak{m}\models\neg A$ holds.

(2) Let $a_{j_1},\ldots,a_{j_i}$ include all free variables that occur in A, and let $\mathfrak{m}$ and $\mathfrak{n}$ be $\mathfrak{M}$-sequences for which $m_{j_1}=n_{j_1},\ldots,m_{j_i}=n_{j_i}$. Then $\mathfrak{M},\mathfrak{m}\models A$ and $\mathfrak{M},\mathfrak{n}\models A$ are equivalent. So we may write $\mathfrak{M}\models A\begin{bmatrix}a_{j_1},\ldots,a_{j_i}\\ m_{j_1},\ldots,m_{j_i}\end{bmatrix}$ instead of $\mathfrak{M},\mathfrak{m}\models A$, for any formula A whose free variables are among $a_{j_1},\ldots,a_{j_i}$, and any $\mathfrak{M}$-sequence $\mathfrak{m}$.

(3) If A is a closed formula, then for an arbitrary pair of sequences $\mathfrak{m}$ and $\mathfrak{n}$, $\mathfrak{M},\mathfrak{m}\models A$ and $\mathfrak{M},\mathfrak{n}\models A$ are equivalent. Therefore, for a closed formula A, we may express the statement "for some (or, equivalently, for all) $\mathfrak{M}$-sequence $\mathfrak{m}$, $\mathfrak{M},\mathfrak{m}\models A$ holds" by $\mathfrak{M}\models A$.

(4) Let a_i be an arbitrary variable that does not occur in $\forall xF(x)$ or $\exists xF(x)$. Then $\mathfrak{M},\mathfrak{m}\models\forall xF(x)$ is equivalent to $\mathfrak{M},\mathfrak{n}\models F(a_i)$ for an arbitrary $\mathfrak{n}$ such that $\mathfrak{n}\overset{i}{=}\mathfrak{m}$. Likewise, $\mathfrak{M},\mathfrak{m}\models\exists xF(x)$ is equivalent to the statement that there is an $\mathfrak{n}$ such that $\mathfrak{n}\overset{i}{=}\mathfrak{m}$ and $\mathfrak{M},\mathfrak{n}\models F(a_i)$.

D. Models

Here again we fix a language L. Let A be a closed L-formula and $\mathfrak{M}$ an L-structure. If $\mathfrak{M}\models A$, then $\mathfrak{M}$ is called a **model** of A. Furthermore, if $\Gamma=\{A_0,A_1,\ldots.\}$ is an arbitrary set of closed formulas and $\mathfrak{M}\models A_i$ for all A_i in Γ, then the structure $\mathfrak{M}$ is called a **model** of Γ.

(1) **Consistency**. Consider a logical system whose language is L. If there is a model of the set of all provable closed formulas of the system, then the system is †consistent. In particular, the †first-ordered predicate calculus is consistent.

(2) **Completeness**. A logical system is said to be **complete** if every closed formula that is satisfied in every structure is provable in the

system. In particular, the first-order predicate calculus is complete.

K. Gödel proved (2). Later L. Henkin gave an alternative proof whose essential idea contributed to proving the following proposition: If a set Γ of closed L-formulas is consistent, then there is a model of Γ. Henkin also introduced a (nonstandard) second-order semantics, relative to which the †second-order predicate calculus is complete. This can be shown by extending Henkin's technique (for the first order) to the second-order language.

(3) Here we extend the language slightly by adding the second-order free predicate variables $\alpha_1^n, \alpha_2^n, \ldots, \alpha_i^n, \ldots$ $(n=1,2,\ldots)$ and the second-order bound predicate variables $\varphi_1^n, \varphi_2^n, \ldots, \varphi_i^n, \ldots$ $(n=1,2,\ldots)$, where n indicates the number of arguments of a variable. Otherwise the definition of the language is the same as for the case of the first-order predicate calculus. For simplicity, however, we assume that there are no individual constants, function symbols, or predicate symbols.

The structure is defined as follows: Put $\mathfrak{M}=[M:S_1,S_2,\ldots,S_n,\ldots]$, where M is a nonempty set and S_n is a set of subsets of $M\times\ldots\times M$ (n times). An $\mathfrak{M}$-sequence $\mathfrak{m}$ is defined as before, and $\mathfrak{s}_n$ denotes $(s_1^n,s_2^n,\ldots,s_i^n,\ldots)$, where each s_i^n is a member of S_n. The concept of satisfiability is defined as follows:

$\mathfrak{M},(\mathfrak{m},\mathfrak{s}_1,\ldots,\mathfrak{s}_n,\ldots)\models\alpha_j^n(x_{i_1},\ldots,x_{i_n})\Leftrightarrow(m_{i_1},\ldots,m_{i_n})\in s_j^n$.

$\mathfrak{M},(\mathfrak{m},\mathfrak{s}_1,\ldots,\mathfrak{s}_n,\ldots)\models\forall\varphi_k^n A(\varphi_k^n)\Leftrightarrow$ for an arbitrary $\mathfrak{s}_n'$ for which $\mathfrak{s}_n'\overset{j}{=}\mathfrak{s}_n$, $\mathfrak{M},(\mathfrak{m},\mathfrak{s}_1,\ldots,\mathfrak{s}_n',\ldots)\models A(\alpha_j^n)$, where α_j^n has the smallest index among the free predicate variables that do not occur in $A(\varphi_k^n)$.

$\mathfrak{M},(\mathfrak{m},\mathfrak{s}_1,\ldots,\mathfrak{s}_n,\ldots)\models\exists\varphi_k^n A(\varphi_k^n)\Leftrightarrow$ there exists an $\mathfrak{s}_n'$ such that $\mathfrak{s}_n'\overset{j}{=}\mathfrak{s}_n$ and $\mathfrak{M},(\mathfrak{m},\mathfrak{s}_1,\ldots,\mathfrak{s}_r,\ldots)\models A(\alpha_j^n)$, where α_j^n satisfies the same condition as in the previous clause.

Satisfiability for other cases is defined as for first-order predicate language. A structure $\mathfrak{M}$ is called **normal** if all axioms of the second-order predicate calculus are true in $\mathfrak{M}$.

Completeness of the second-order predicate calculus: Every closed formula that is satisfiable in all normal structures is provable in the second-order predicate calculus.

(4) Let the cardinality of L_4 be τ, and Γ be an arbitrary set of closed L-formulas. If Γ has a model, then Γ has a model of cardinality max $(\tau,\aleph_0)$. This follows from Henkin's method. Historically, however, it was first proved by Th. Skolem and L. Löwenheim for a special case, and was later generalized by A. I. Mal'tsev and A. Robinson.

(5) The following results are all due to A. Tarski and R. L. Vaught.

Definition 1. Two L-structures $\mathfrak{M}$ and $\mathfrak{N}$ are said to be **elementarily (arithmetically) equivalent** if for an arbitrary closed L-formula A, $\mathfrak{M}\models A\Leftrightarrow\mathfrak{N}\models A$.

Definition 2. Let

$$\mathfrak{M}=[M:q_0,q_1,\ldots;g_0,g_1,\ldots;Q_0,Q_1,\ldots]$$

and

$$\mathfrak{N}=[N:r_0,r_1,\ldots;h_0,h_1,\ldots;R_0,R_1,\ldots]$$

be two structures. $\mathfrak{M}$ is an **elementary extension** of $\mathfrak{N}$ if the following two conditions are satisfied: (i) $M\supset N$; $q_j=r_j$ $(j=0,1,\ldots)$; the restriction of g_j to N is identical to h_j $(j=0,1,\ldots)$; the restriction of Q_j to N is identical to R_j $(j=0,1,\ldots)$. (If this condition holds, then $\mathfrak{M}$ is said to be an extension of $\mathfrak{N}$.) (ii) For an arbitrary L-formula A and an arbitrary $\mathfrak{N}$-sequence $\mathfrak{n}$, if $\mathfrak{N},\mathfrak{n}\models A$ then $\mathfrak{M},\mathfrak{n}\models A$.

Theorem 1. Let $\mathfrak{M}$ be an extension of $\mathfrak{N}$. A necessary and sufficient condition for $\mathfrak{M}$ to be an elementary extension of $\mathfrak{N}$ is that for an arbitrary L-formula of the form $\exists xF(x)$ and an arbitrary $\mathfrak{N}$-sequence $\mathfrak{n}$, if $\mathfrak{M},\mathfrak{n}\models\exists xF(x)$, then there is some element n of N such that for the $\mathfrak{M}$-sequence $\mathfrak{m}$ for which $\mathfrak{m}\overset{i}{=}\mathfrak{n}$ and $m_i=n$, $\mathfrak{M},\mathfrak{m}\models F(a_i)$, where a_i is an arbitrary free variable that does not occur in $F(x)$.

Theorem 2. Here we place a condition on L that each set of symbols be at most countable and arranged in the ω-type (→ 312 Ordinal Numbers). Let the cardinality of the universe M of $\mathfrak{M}$ be an infinite cardinal $\mathfrak{a}$, M' be a subset of M of cardinality $\mathfrak{c}$, and $\mathfrak{b}$ be an infinite cardinal that satisfies $\mathfrak{c}\leqslant\mathfrak{b}\leqslant\mathfrak{a}$. Then there exists an L-structure $\mathfrak{N}$ whose universe N has cardinality $\mathfrak{b}$ and such that $M'\subset N$ and $\mathfrak{M}$ is an elementary extension of $\mathfrak{N}$.

Theorem 3. Suppose that L satisfies the same condition as in Theorem 2. Let the cardinality of the universe M of $\mathfrak{M}$ be $\mathfrak{a}$ ($\mathfrak{a}$ is an infinite cardinal) and $\mathfrak{b}$ be a cardinal for which $\mathfrak{a}\leqslant\mathfrak{b}$. Then there exists an L-structure $\mathfrak{N}$ that is a proper elementary extension of $\mathfrak{M}$ and whose universe has cardinality $\mathfrak{b}$.

E. Ultraproducts

Assume that for a set of L-structures Σ and a set of indices I, there is a mapping θ from I onto Σ. If α is a member of I, $\mathfrak{M}$ is a member of Σ, and $\theta(\alpha)=\mathfrak{M}$, then $\mathfrak{M}$ may be denoted by $\mathfrak{M}^\alpha$. It should be noted that there may be more than one α corresponding to the same structure. If D is a †maximal filter of I and $\mathfrak{M}^\alpha$

is expressed as

$$\mathfrak{M}^\alpha = [M^\alpha : \bar{c}^\alpha, \ldots : \bar{f}^\alpha, \ldots : \bar{P}^\alpha, \ldots],$$

then $\prod_{\alpha\in I} M^\alpha$ is defined by

$$\prod_{\alpha\in I} M^\alpha = \{\varphi \mid \varphi \text{ is a mapping from } I \text{ into } \bigcup_{\alpha\in I} M^\alpha, \text{ where } \varphi(\alpha)\in M^\alpha\}.$$

For any two elements φ and ψ of $\prod_{\alpha\in I} M^\alpha$, $\varphi \overset{D}{=} \psi$ is defined by

$$\varphi \overset{D}{=} \psi \Leftrightarrow \{\alpha \mid \varphi(\alpha) = \psi(\alpha)\} \in D.$$

Then $\varphi \overset{D}{=} \psi$ is an equivalence relation between the elements of $\prod_{\alpha\in I} M^\alpha$. Furthermore, the set $\prod_{\alpha\in I} M^\alpha$ partitioned by $\overset{D}{=}$ is expressed by $\prod_{\alpha\in I} M^\alpha/D$, and each element m of $\prod_{\alpha\in I} M^\alpha/D$ is expressed by $m = [\varphi]$, where φ is a representing element of m.

Next we define an operator that produces a new structure from Σ. Put $M = \prod_{\alpha\in I} M^\alpha/D$. For an individual constant c of L, let $\bar{c} = [\varphi]$, where $\varphi(\alpha) = \bar{c}^\alpha$ for every α. For an n-ary function f of L and arbitrary elements $m_1 = [\varphi_1], \ldots, m_n = [\varphi_n]$ of M, define $\bar{f}(m_1, \ldots, m_n) = [\psi]$, where $\psi(\alpha) = \bar{f}^\alpha(\varphi_1(\alpha), \ldots, \varphi_n(\alpha))$ for every α. For an n-ary predicate P of L, define $\langle m_1, \ldots, m_n\rangle \in \bar{P}$ by

$$\langle m_1, \ldots, m_n\rangle \in \bar{P}$$
$$\Leftrightarrow \{\alpha \mid \langle \varphi_1(\alpha), \ldots, \varphi_n(\alpha)\rangle \in P^\alpha\} \in D.$$

According to these definitions, put

$$\mathfrak{M} = [M : \bar{c}, \ldots, \bar{f}, \ldots, \bar{P}, \ldots],$$

and denote it by $\prod_{\alpha\in I} \mathfrak{M}^\alpha/D$, called the **ultraproduct** of $\{\mathfrak{M}^\alpha\}_{\alpha\in I}$ (with respect to D). $\mathfrak{M}$ is an L-structure.

Fundamental Theorem of Ultraproducts. Let $\mathfrak{M} = \prod_{\alpha\in I} \mathfrak{M}^\alpha/D$ be the ultraproduct of $\{\mathfrak{M}^\alpha\}_{\alpha\in I}$, $\mathfrak{m} = (m_1, m_2, \ldots)$ be an $\mathfrak{M}$-sequence, φ_i be a representing element of m_i, and A be an arbitrary formula. Then $\mathfrak{M}, \mathfrak{m} \models A \Leftrightarrow \{\alpha \mid \mathfrak{M}^\alpha, (\varphi_1(\alpha), \varphi_2(\alpha), \ldots) \models A\} \in D$.

By using this fundamental theorem we have the following result: Suppose that Γ is a set of closed formulas in L such that every finite subset of Γ has a model. Let I be the set of all the finite subsets of Γ. For each $\alpha \in I$ and each $A \in \Gamma$, let $\mathfrak{M}^\alpha$ be a model of α and $\hat{A}$ the set of all the finite sets in I which contain A as a member. Let $F = \{\hat{A} \mid A \in \Gamma\}$. Since F has the finite intersection property, there is a maximal filter D such that $D \supseteq F$. Let $\mathfrak{M}$ be the ultraproduct $\prod_{\alpha\in I} \mathfrak{M}^\alpha/D$. Since $\hat{A}$ is a subset of the set $\{\alpha \mid \mathfrak{M}^\alpha \models A\}$ and $\hat{A}$ belongs to D, the set $\{\alpha \mid \mathfrak{M}^\alpha \models A\}$ belongs to D, for each $A \in \Gamma$. Hence we have that M is a model of Γ, by the foregoing fundamental theorem of ultraproducts. Therefore, we have the following theorem.

Compactness Theorem. A set Γ of closed formulas has a model if and only if every finite subset of it has a model.

In the case where all the structures $\mathfrak{M}^\alpha$ coincide with the single structure $\mathfrak{N}$, the ultraproduct of $\{\mathfrak{M}^\alpha\}_{\alpha\in I}$ (with respect to D) may be written $\mathfrak{N}^I/D$ and called the **ultrapower** of $\mathfrak{N}$ (with respect to D).

Let

$$\mathfrak{M} = [M; q_0, q_1, \ldots; g_0, g_1, \ldots; Q_0, Q_1, \ldots]$$

and

$$\mathfrak{N} = [N; r_0, r_1, \ldots; h_0, h_1, \ldots; R_0, R_1, \ldots]$$

be two structures. $\mathfrak{M}$ and $\mathfrak{N}$ are said to be **isomorphic** if there is a bijection f from M to N such that the following three conditions hold: (i) $f(q_0) = r_0, f(q_1) = r_1, \ldots$. (ii) The sequences $g_0, g_1, \ldots$ and $h_0, h_1, \ldots$ are of the same type and $f(g_j(a_1, \ldots, a_n)) = h_j(f(a_1), \ldots, f(a_n))$ holds for every n-tuple $a_1, \ldots, a_n$ in M. (iii) The sequences $Q_0, Q_1, \ldots$ and $R_0, R_1, \ldots$ are of the same type and $R_j = \{\langle f(a_1), \ldots, f(a_n)\rangle \mid \langle a_1, \ldots, a_n\rangle \in Q_j\}$.

Let j be the function from N to N^I/D defined by $j(a) = [\varphi_a]$ for each $a \in N$, where φ_a is the constant function from I to N such that $\varphi_a(\alpha) = a$ for each $\alpha \in I$. Let $\mathfrak{M}$ be the substructure of $\mathfrak{N}^I/D$ whose universe is the range of j. Then j is an isomorphism of $\mathfrak{N}$ to $\mathfrak{M}$. In the following we identify a and $j(a)$ for each $a \in N$. Then $\mathfrak{N}$ is an elementary substructure of $\mathfrak{N}^I/D$ by the fundamental theorem of ultraproducts.

If $\mathfrak{M}$ and $\mathfrak{N}$ are isomorphic, then $\mathfrak{M}$ and $\mathfrak{N}$ are elementarily equivalent. By using this fact and the fundamental theorem of ultraproducts, we have the following result. Let $\mathfrak{M}$ and $\mathfrak{N}$ be two structures. If there is a nonempty set I and a maximal filter D on I such that $\mathfrak{M}^I/D$ and $\mathfrak{N}^I/D$ are isomorphic, then $\mathfrak{M}$ and $\mathfrak{N}$ are elementarily equivalent. H. J. Keisler proved the converse of this proposition by using the G.C.H. (generalized continuum hypothesis), and later S. Shelah proved it without the G.C.H. **Keisler-Shelah isomorphism theorem**: Let $\mathfrak{M}$ and $\mathfrak{N}$ be two structures. Then $\mathfrak{M}$ and $\mathfrak{N}$ are elementarily equivalent if and only if there is a nonempty set I and a maximal filter D on I such that $\mathfrak{M}^I/D$ and $\mathfrak{N}^I/D$ are isomorphic.

The ultraproduct operation has various applications in number theory, algebraic geometry, and analysis. Here we give an example due to J. Ax and S. Kochen. Let P be the set of prime numbers. Let $\mathbf{Q}_p$ and $Z_p((t))$ be the field of p-adic numbers and the field of formal

power series over $Z_p=\{0,1,\dots,p-1\}$ for each p in P, respectively. **Ax-Kochen isomorphism theorem**: Suppose that D is a nonprincipal maximal filter on P. Then $\prod_{p\in P}\mathbf{Q}_p/D$ and $\prod_{p\in P}Z_p((t))/D$ are isomorphic.

As an immediate consequence of this theorem, we have the following partial solution of Artin's conjecture on Diophantine equations. **Theorem**: For each positive integer d, there exists a finite set Y of primes such that every homogeneous polynomial $f(t_1,\dots,t_n)$ of degree d over $\mathbf{Q}_p$, with $n>d^2$, has a nontrivial zero in $\mathbf{Q}_p$ for every $p\notin Y$ (→ 118 Diophantine Equations).

We give another example in nonstandard analysis. A. Robinson developed the general theory of nonstandard analysis in [10]. Here we explain a theorem due to A. R. Bernstein. Let X be a nonempty set and $U(X)$ be the smallest transitive set (i.e., $a\in b$ and $b\in U(X)$ imply $a\in U(X)$) which has X as a member and is closed under the following operations: pairing, union, power set, and subset operation (i.e., $a\in U(X)$ and $b\subset a$ imply $b\in U(X)$). Let L be the first-order predicate logic with equality whose set of nonlogical constants consists of a binary predicate symbol $\in$ and individual constant symbols c_a for $a\in U(X)$ (→ 411 Symbolic Logic F). Then the first-order structure $\mathfrak{N}=[U(X);\,a(a\in U(X));\,E]$ is an L-structure, where E is the $\in$ relation on the set $U(X)$. Let $\mathfrak{M}=[U(X)^I/D:a(a\in U(X));\,E^I/D]$ be the ultrapower $\mathfrak{N}^I/D$ of $\mathfrak{N}$ with respect to a nonprincipal ultrafilter D on a set I. For each $a\in U(X)$, let a^* be the set of all elements $[\varphi]$ in $U(X)^I/D$ such that $\{i\in I\mid\varphi(i)\in a\}\in D$. Then a is a proper subset of a^* if a is infinite. Since $\mathfrak{N}$ and $\mathfrak{M}$ are elementarily equivalent, these two sets a and a^* have common first-order properties in the following sense: for each formula $\Phi(a_0)$ in L,

$$(\forall b\in a)(\mathfrak{N}\models\Phi[{}^{a_0}_{b}])\Leftrightarrow(\forall b\in a^*)(\mathfrak{M}\models\Phi[{}^{a_0}_{b}]).$$

From this it follows that if r is a relation on a set a in $\mathfrak{N}$, then r^* is a relation on the set a^*; and if f is a mapping from a to b in $\mathfrak{N}$, then f^* is a mapping from a^* to b^*. Hence a^* is a mathematical object which greatly resembles a. By using this type of resemblance between a and a^* we have the following result.

Let H be a Hilbert space over the complex number field $\mathbf{C}$ such that $\dim(H)=\omega$ and let T be a bounded linear operator on H. Let $X=H\cup\mathbf{C}$ and consider the first-order structure $\mathfrak{M}$ as above. Since $\mathbf{R}$ (the set of all real numbers) and $\mathbf{N}$ (the set of all natural numbers) are infinite sets which belong to $U(X)$, $\mathbf{R}^*$ and $\mathbf{N}^*$ have elements which do not belong to $\mathbf{R}$ and $\mathbf{N}$, respectively. Such elements are called **nonstandard real numbers** and **nonstandard natural numbers**, respectively. By the fundamental theorem of ultraproducts we can conclude that there are many nonstandard real numbers α such that $0<{}^*\alpha<{}^*a$ in $\mathbf{R}^*$ for any $a\in\mathbf{R}$. Such a nonstandard real number α is called an **infinitesimal real number**. Since the norm operator $\|\ \ \|$ is a mapping from H to $\mathbf{R}$, $\|\ \ \|^*$ is a mapping from H^* to $\mathbf{R}^*$. If $\|x\|^*$ ($x\in H^*$) is infinitesimal, then x is said to be **infinitesimal** in H^*. Let S be the set of all linear subspaces of H. For a linear subspace K of H^* that is contained in S^* let K° be the set of elements $x\in H$ such that $x-x_0$ is infinitesimal for some x_0 in K. Then K° is a closed linear subspace of H. Let $\mathfrak{e}=\{e_i\}_{i\in\mathbf{N}}$ be an orthonormal basis of the Hilbert space H; $\mathfrak{e}$ can be considered as a mapping from $\mathbf{N}$ to H, and hence $\mathfrak{e}^*=\{e_j\}_{j\in\mathbf{N}^*}$ is a mapping from $\mathbf{N}^*$ to H^*. For each $j\in\mathbf{N}^*$, let H_j be the linear subspace of H^* spanned by $\{e_k\mid k\leqslant j\}$. For a given bounded linear operator T on H, T^* is a linear operator on H^*. We define $T_j=P_jT^*P_j$, where P_j is the projection from H^* to H_j. Since $\dim(H_j)=j$ is a (nonstandard) natural number, there exists a tower $J_{0j}\subset J_{1j}\subset\ldots\subset J_{jj}=H_j$ of closed, T_j-invariant linear subspaces of H_j such that $\dim(J_{kj})=k(k\leqslant j)$. Then $J_{kj}{}^\circ$ is a closed, T-invariant linear subspace of H. If there is a polynomial $p(x)$ such that $p(T)$ is a compact operator, then we get a nonstandard natural number j such that $J_{kj}{}^\circ$ is a proper subspace of H for some $k\leqslant j$. This gives the following result, which is an affirmative solution of a problem of K. Smith and P. R. Halmos. **Theorem** (Bernstein [4]): Let T be a bounded linear operator on an infinite-dimensional Hilbert space H over the complex numbers and let $p(x)\neq 0$ be a polynomial with complex coefficients such that $p(T)$ is compact. Then T leaves invariant at least one closed linear subspace of H other than H or $\{0\}$.

F. Categoricity in Powers

Let Γ be a set of closed formulas in a first-order language L which has a designated binary predicate symbol P_0. In the following, we assume that the interpretation $\bar{P}_0$ of P_0 by $\mathfrak{M}$ is the equality relation on the universe of $\mathfrak{M}$ for every L-structure $\mathfrak{M}$. Γ is said to be **categorical** if all the models of Γ are isomorphic. By Theorem 3 in Section D, any Γ having a model of infinite cardinality is not categorical. Hence, there exists no interesting Γ which is categorical. Therefore, we consider the weaker notion of **categoricity in powers**. Let κ be an infinite cardinal and $n(\Gamma,\kappa)$ be the number of nonisomorphic models of Γ of cardinality κ. Then Γ is said to be categorical in κ if $n(\Gamma,\kappa)=1$, i.e., if all the models of Γ of cardinality κ are isomorphic. There exist many

interesting Γ's which are categorical in κ for some κ. For example, the set of axioms of algebraically closed fields of characteristic 0 is categorical in $\aleph_1$, and the set of axioms of dense linear orderings without endpoints is categorical in $\aleph_0$. With respect to this notion, J. Łoś conjectured that if Γ is categorical in κ for some $\kappa > \overline{\overline{L}}$ (the cardinality of L), then Γ is categorical in κ for all $\kappa > \overline{\overline{L}}$. This conjecture was solved affirmatively by M. Morley in the case $\overline{\overline{L}} = \aleph_0$, and later by S. Shelah in the general case. **Theorem (1)**: Let Γ be a set of closed formulas in L. Then Γ is categorical in κ for some $\kappa > \overline{\overline{L}}$ if and only if Γ is categorical in κ for all $\kappa > L$. We also have the following interesting theorem, due to J. T. Baldwin and A. H. Lachlan. **Theorem (2)**: Let Γ be a set of closed formulas in L such that $L = \aleph_0$. If Γ is categorical in $\aleph_1$, then $n(\Gamma, \aleph_0) = 1$ or $\aleph_0$.

As for $n(\Gamma, \kappa)$ there are two famous conjectures due to R. Vaught and others. A set Γ of closed formulas in L is called **complete** in L if it has a model and, for any closed formula A in L, either $A \in \Gamma$ or $\neg A \in \Gamma$.

Conjecture 1. There is no complete set Γ of closed formulas in L such that $\aleph_0 < n(\Gamma, \aleph_0) < 2^{\aleph_0}$.

Conjecture 2. There is no finite set Γ of closed formulas in L such that $n(\Gamma, \aleph_0) = \aleph_0$ and $n(\Gamma, \aleph_1) = 1$.

G. Omitting Type Theorem

By a_0-formulas (or a-formulas), we mean formulas which have no free variables except a_0. Suppose that $\mathfrak{M}$ is an L-structure and Σ is a set of a-formulas in L. Then, $\mathfrak{M}$ realizes Σ if there is an element m of the universe of $\mathfrak{M}$ such that $\mathfrak{M} \models A\left[\begin{smallmatrix} a \\ m \end{smallmatrix}\right]$ for all A in Σ, and $\mathfrak{M}$ omits Σ if $\mathfrak{M}$ does not realize Σ. For example, if L is a first-order language such that $L_4 = \{c_0, c_1\}$, $L_5 = \{f_0, f_1\}$, $L_6 = \{P_0, P_1\}$, where f_0, f_1, P_0, P_1 are all binary, and $\mathfrak{N}$ is the L-structure $[\mathfrak{N}; 0, 1; +, x; =, <]$, where $\mathfrak{N}$ is the set of natural numbers, and $\Sigma = \{P_1(\overline{n}, a) \mid n = 0, 1, 2, \ldots\}$, where $\overline{0} = c_0$, $\overline{1} = f_0(\overline{0}, c_1), \ldots, \overline{n+1} = f_0(\overline{n}, c_1), \ldots$, then clearly $\mathfrak{N}$ omits Σ. Also, if $\mathfrak{M}$ is an L-structure, m is an element of the universe of $\mathfrak{M}$, and $\Sigma_m = \{A \mid A$ is an a-formula such that $\mathfrak{M} \models A\left[\begin{smallmatrix} a \\ m \end{smallmatrix}\right]\}$, then clearly $\mathfrak{M}$ realizes Σ_m, where Σ_m is called the type of m in $\mathfrak{M}$. Suppose that Γ is a set of closed formulas in L. Then, by the completeness theorem of L, we can easily see that Γ has a model realizing Σ if and only if $\Gamma^{\cup}\Sigma$ is consistent. On the other hand, it is rather difficult to obtain a necessary and sufficient condition for Γ to have a model omitting Σ. The following is a sufficient condition: Γ is said to locally omit Σ if there is no a-formula A such that $\Gamma^{\cup}\{A\}$ is consistent and, for any formula B in Σ, $\Gamma^{\cup}\{A, \neg B\}$ is not consistent.

Theorem. Suppose that Γ is a consistent set of closed formulas in a countable language L, and Σ is a set of a_0-formulas in L. If Γ locally omits Σ, then Γ has a countable model which omits Σ. Also, if Γ has a model of power greater than $\beth_\xi$ omitting Σ for each $\xi < \omega_1$, then Γ has a model omitting Σ in each infinite power, where $\beth_\xi$ is defined by $\beth_0 = \aleph_0$, $\beth_{\xi+1} = 2^{\beth_\xi}$, $\beth_\sigma = \sup_{\xi<\sigma} \beth_\xi$ if σ is a limit ordinal.

References

[1] J. Ax and S. Kochen, Diophantine problems over local fields I, Amer. J. Math., 87 (1965), 605–630.
[2] J. Ax and S. Kochen, Diophantine problems over local fields II, Amer. J. Math., 87 (1965), 630–648.
[3] J. Ax and S. Kochen, Diophantine problems over local fields III, Ann. Math., (2) 83 (1966), 437–456.
[4] A. R. Bernstein, Non-standard analysis, Studies in Model Theory 8, M. D. Morley (ed.), Mathematical Association of America, 1973, 35–58.
[5] C. C. Chang and H. J. Keisler, Model theory, North-Holland, 1973.
[6] S. C. Kleene, Introduction to metamathematics, Van Nostrand, 1952.
[7] M. Machover and J. Hirschfelt, Lectures on non-standard analysis, Lecture notes in math. 94, Springer, 1969.
[8] M. Morley, Categoricity in powers, Trans. Am. Math. Soc., 114 (1965), 514–538.
[9] M. Morley, Omitting types of elements, The Theory of Models, J. W. Addison et al. (eds.), North-Holland, 1965, 265–273.
[10] A. Robinson, Non-standard analysis, North-Holland, 1966.
[11] G. E. Sacks, Saturated model theory, Benjamin, 1972.
[12] S. Shelah, Classification theory, North-Holland, 1978.

277 (III.23) Modules

A. General Remarks

In this article, we consider mainly modules with operator domain (Section C), in particular modules over a †ring. Modules over a

field are linear spaces (→ 256 Linear Spaces). Modules over a commutative ring are important in algebraic geometry (→ 16 Algebraic Varieties, 67 Commutative Rings, 284 Noetherian Rings). The theory of modules over a †group ring can be identified with the theory of linear representations of a group (→ 362 Representations). Modules without operator domain may be regarded as modules over the ring **Z** of rational integers, and the theory of finitely generated Abelian groups can be generalized to the theory of modules over a †principal ideal domain (→ 52 Categories and Functors, 200 Homological Algebra).

B. Modules

A **module** (without operator domain) is a †commutative group M whose law of composition is written additively: $a+b=b+a$ $(a, b \in M)$; the †identity element is denoted by 0, and the inverse element of a by $-a$. Every subgroup H of M is a normal subgroup. For any $a \in M$, the left and right cosets of H containing the element a are identical: $H+a=a+H$ (→ 190 Groups A).

In the set N^M of all mappings of a set M to a module N, we define an addition by the sums of values: $(f+g)(x)=f(x)+g(x)$. Then N^M forms a module. The set $\mathrm{Hom}(M, N)$ of all homomorphisms of a module M to a module N forms a subgroup of the module N^M, called the **module of homomorphisms** of M to N. The composite of homomorphisms is a homomorphism. Hence the set $\mathrm{Hom}(M, M)=\mathscr{E}(M)$ of all endomorphisms of M forms a †ring with respect to the addition and the multiplication defined by composition; this is called the **endomorphism ring** of M. The †unity element of $\mathscr{E}(M)$ is the identity mapping of M, and the †invertible elements of $\mathscr{E}(M)$ are the automorphisms of M.

Let $\{x_\lambda\}_{\lambda \in \Lambda}$ be a family of elements in a module M. The sum $\sum_{\lambda \in \Lambda} x_\lambda$ is well defined if $x_\lambda=0$ $(\lambda \in \Lambda)$ except for a finite number of λ. For any family $\{N_\lambda\}_{\lambda \in \Lambda}$ of subsets of M, $\sum_{\lambda \in \Lambda} N_\lambda$ denotes the set of all elements of the form $\sum_{\lambda \in \Lambda} x_\lambda$ $(x_\lambda \in N_\lambda)$, where $x_\lambda=0$ except for a finite number of λ. If all the N_λ are subgroups of M, then $N=\sum_{\lambda \in \Lambda} N_\lambda$ is also a subgroup, called the **sum** of $\{N_\lambda\}_{\lambda \in \Lambda}$. If every element of N can be written uniquely in the form $\sum_{\lambda \in \Lambda} x_\lambda$ $(x_\lambda \in N_\lambda)$, N is called the **direct sum** of $\{N_\lambda\}_{\lambda \in \Lambda}$. When the N_λ are subgroups, this is equivalent to the condition that $N_\lambda \cap \sum_{\lambda \neq \mu \in \Lambda} N_\mu=\{0\}$ $(\lambda \in \Lambda)$.

C. Modules with Operator Domain

Suppose that we are given a set A and a module M. If with each pair of elements $a \in A$ and $x \in M$ there is associated a unique element $ax \in M$ satisfying the condition (1) $a(x+y)=ax+ay$ $(a \in A; x, y \in M)$, we say that A is an **operator domain** of M and M is a **module with operator domain** A (**module over** A or A-**module**) (→ 190 Groups E). The mapping $A \times M \rightarrow M$ given by $(a, x) \rightarrow ax$ is called the †**operation** of A on M. Any $a \in A$ induces an endomorphism $a_M: x \rightarrow ax$ of M as a module (not as an A-module). To give the structure of an A-module to a module M amounts to giving a mapping $A \rightarrow \mathscr{E}(M)$ $(a \rightarrow a_M)$.

If N is a subgroup of an A-module M such that $ax \in N$ for any $a \in A$ and $x \in N$, then N forms an A-module, called an A-**submodule** (or **allowed submodule**) of M. If $\{N_\lambda\}_{\lambda \in \Lambda}$ is a family of A-submodules of an A-module M, then the intersection $\bigcap_{\lambda \in \Lambda} N_\lambda$ and the sum $\sum_{\lambda \in \Lambda} N_\lambda$ are both A-submodules of M.

Let R be an †equivalence relation in an A-module M such that if $a \in A$ and xRy, then $axRay$. Then R is said to be **compatible with the operation** of A. In this case, an operation of A is induced on the quotient set M/R. Moreover, if R is compatible with the addition, namely, xRx' and yRy' imply $(x+y)R(x'+y')$, then M/R forms an A-module, called a **factor** A-**module** of M. The equivalence class N containing 0 is an A-submodule of M, and M/R coincides with M/N.

D. Modules over a Group or a Ring

If a (multiplicative) group structure is given to the operator domain A of a module M, we always assume (in addition to condition (1) in Section C) that the following two conditions are satisfied: (2) $(ab)x=a(bx)$; (3) $1x=x$ $(a, b \in A, x \in M)$.

If a ring structure is given to A, we always assume (besides conditions (1) and (2)) that the following condition holds: (4) $(a+b)x=ax+bx$ $(a, b \in A, x \in M)$. This means that the mapping $A \rightarrow \mathscr{E}(M)$ $(a \rightarrow a_M)$ is a †ring homomorphism. If the ring A has unity element 1, and $1_M=$ identity mapping (namely, condition (3) holds), then the A-module M is called **unitary**. We consider only unitary A-modules. Any module M can be regarded as a **Z**-module or as an $\mathscr{E}(M)$-module.

When M is a module over a ring A, an element of A is called a **scalar**, A itself is called the **ring of scalars** (**basic ring** or **ground ring**), and the operation $A \times M \rightarrow M$ is called the **scalar multiplication**. The elements ax $(a \in A)$ are called **scalar multiples** of x, and the totality of these elements is denoted by Ax. Let $\{x_\lambda\}_{\lambda \in \Lambda}$ be a family of elements in M. An element of the form $\sum_{\lambda \in \Lambda} a_\lambda x_\lambda$, where the a_λ are elements of A and equal to 0 except for a finite number

of λ, is called a linear combination of $\{x_\lambda\}_{\lambda\in\Lambda}$. The set N of linear combinations of $\{x_\lambda\}_{\lambda\in\Lambda}$ is the smallest A-submodule of M containing all the x_λ ($\lambda\in\Lambda$) and is equal to the sum $\sum_{\lambda\in\Lambda}Ax_\lambda$. The A-module N is said to be **generated** by $\{x_\lambda\}_{\lambda\in\Lambda}$, and $\{x_\lambda\}_{\lambda\in\Lambda}$ is called a **system of generators** of N. A module having a finite number of generators is said to be **finitely generated (of finite type** or simply **finite**). The module Ax generated by a single element x is called **monomial.** If A is a †field (which may be noncommutative), an A-module is a linear space over A (→ 256 Linear Spaces).

Let a be an element of an A-module M. If there exists a nonzero divisor λ of A such that $\lambda a=0$, then a is called a **torsion element** of M. We say M is a **torsion A-module** if every element of M is a torsion element, and M is **torsion free** if M has no torsion element other than 0. An element a of M is called **divisible** if for any nonzero divisor $\lambda\in A$ there exists an element $b\in M$ such that $a=\lambda b$. M is called a **divisible A-module** if every element of M is divisible.

Strictly speaking, the A-modules we have considered so far are called **left A-modules.** If we define the operation of A on M by xa ($a\in A, x\in M$) instead of ax and modify conditions (1)–(4) appropriately (in particular, condition (2) becomes $x(ab)=(xa)b$), then M is called a **right A-module.** If A° is a group or ring anti-isomorphic to A, then a left A-module can be naturally identified with a right A°-module. If A is a commutative group or ring, we can disregard the distinction between left and right A-modules.

Let A and B be groups or rings. Sometimes we consider an A-module structure and a B-module structure simultaneously on the same module M. If the operations of A and B commute with each other, namely, $a(bx)=b(ax)$ ($a\in A, b\in B, x\in M$), it is convenient to put one of the operations to the right. If M has a left A-module structure and a right B-module structure, satisfying condition (5) $(ax)b=a(xb)$, then M is called an **A-B-bimodule.** If G is a group and K is a commutative ring, the G-K-bimodule structure is equivalent to the left $K[G]$-module structure, where $K[G]$ is the †group ring.

E. Operator Homomorphisms

A homomorphism $f:M\to N$ of A-modules M and N such that $f(ax)=af(x)$ ($a\in A, x\in M$) is called an **A-homomorphism (operator homomorphism** or **allowed homomorphism).** If A is a ring, f is also called an **A-linear mapping.** Regarding A as an A-module, an A-linear mapping $M\to A$ is called a **linear form** on M. The composite of A-homomorphisms is an A-homomorphism.

Let $f:M\to L$ be an A-homomorphism of A-modules. The A-submodule $\operatorname{Im} f=f(M)$ of L is called the **image** of f, and the A-submodule $\operatorname{Ker} f=\{x\mid x\in M, f(x)=0\}$ of M is called the **kernel** of f. $\operatorname{Coim} f=M/\operatorname{Ker} f$ is called the **coimage** of f, and $\operatorname{Coker} f=L/\operatorname{Im} f$ the **cokernel** of f. The binary relation xRy on M defined by $f(x)=f(y)$ ($x,y\in M$) coincides with the equivalence relation defined by $x-y\in N=\operatorname{Ker} f$ ($x,y\in M$), and the mapping f induces an A-isomorphism $\bar f: M/N\to f(M)$.

A sequence of A-homomorphisms of A-modules M_n ($n\in\mathbf{Z}$)

$$\ldots\to M_{n-1}\xrightarrow{f_{n-1}}M_n\xrightarrow{f_n}M_{n+1}\to\ldots$$

is called an **exact sequence** if $\operatorname{Im} f_{n-1}=\operatorname{Ker} f_n$ for all n. The A-module $\{0\}$ is denoted by 0. Exactness of $0\to N\xrightarrow{f}M$ or $M\xrightarrow{g}L\to 0$ means that the mapping $f:N\to M$ is injective or the mapping $g:M\to L$ is surjective, respectively. In an †inductive (†projective) system $\{M_\lambda, f_{\lambda\mu}\}$ of A-modules, where every $f_{\lambda\mu}: M_\lambda\to M_\mu$ $\lambda<\mu$ ($\lambda>\mu$) is an A-homomorphism, the limit $M=\varinjlim M_\lambda$ ($\varprojlim M_\lambda$) is also an A-module. If $0\to L_\lambda\to M_\lambda\to N_\lambda\to 0$ is exact for every λ and

$$\begin{array}{ccccc} L_\lambda & \to & M_\lambda & \to & N_\lambda \\ \downarrow & & \downarrow & & \downarrow \\ L_\mu & \to & M_\mu & \to & N_\mu \end{array}$$

is a †commutative diagram, then $0\to\varinjlim L_\lambda\to\varinjlim M_\lambda\to\varinjlim N_\lambda\to 0$ is also exact. For the projective limit, however, we can only state the exactness of $0\to\varprojlim L_\lambda\to\varprojlim M_\lambda\to\varprojlim N_\lambda$.

The set of all A-homomorphisms of an A-module M to an A-module N, denoted by $\operatorname{Hom}_A(M,N)$, is a subgroup of the module $\operatorname{Hom}(M,N)$, and is called the **module of A-homomorphisms.** The set $\operatorname{Hom}_A(M,M)=\mathscr{E}_A(M)$ of all A-endomorphisms of an A-module M forms a †subring of the ring $\mathscr{E}(M)$ and coincides with the set of all elements commuting with any a_M ($a\in A$). We denote by $GL(M)$ the group of all †invertible elements in $\mathscr{E}_A(M)$. If A is a commutative ring, $\operatorname{Hom}_A(M,N)$ can be regarded as an A-module by defining $(af)(x)=af(x)$, namely, $af=a_N\circ f$. In particular, $\mathscr{E}_A(M)$ is an †associative algebra over A. If M is an A-B-bimodule, $\operatorname{Hom}_A(M,N)$ forms a left B-module by $(bf)(x)=f(xb)$. If N is an A-B-bimodule, $\operatorname{Hom}_A(M,N)$ forms a right B-module by $(fb)(x)=f(x)b$.

F. Direct Products and Direct Sums

In the Cartesian product $P=\prod_{\lambda\in\Lambda}M_\lambda$ of a family $\{M_\lambda\}_{\lambda\in\Lambda}$ of A-modules, we define addition and an A-operation as follows: $\{x_\lambda\}+\{y_\lambda\}=\{x_\lambda+y_\lambda\}$, $a\{x_\lambda\}=\{ax_\lambda\}$. Then P forms

an A-module. We call $\prod_{\lambda\in\Lambda} M_\lambda$ the **direct product** of modules $\{M_\lambda\}_{\lambda\in\Lambda}$. The **canonical projection** assigning x_λ to $\{x_\lambda\}$ is denoted by $p_\lambda: P\rightarrow M_\lambda$. Suppose that an A-module M and A-homomorphisms $f_\lambda: M\rightarrow M_\lambda (\lambda\in\Lambda)$ are given. Then there exists a unique A-homomorphism $f: M\rightarrow P$ such that $p_\lambda\circ f=f_\lambda$ $(\lambda\in\Lambda)$; f is given by $f(x)=\{f_\lambda(x)\}$.

In the direct product $\prod_{\lambda\in\Lambda} M_\lambda$, the set S of all elements whose components x_λ are equal to 0 except for a finite number of λ is called the **direct sum** of modules $\{M_\lambda\}_{\lambda\in\Lambda}$ and is denoted by $\sum_{\lambda\in\Lambda} M_\lambda$ (or $\coprod_{\lambda\in\Lambda} M_\lambda$ or $\bigoplus_{\lambda\in\Lambda} M_\lambda$). The **canonical injection** assigning $\{\ldots,0,x_\lambda,0,\ldots\}\in S$ to $x_\lambda\in M_\lambda$ is denoted by $j_\lambda: M_\lambda\rightarrow S$. If an A-module M and A-homomorphisms $f_\lambda: M_\lambda\rightarrow M$ are given, then there exists a unique A-homomorphism $f: S\rightarrow M$ such that $f\circ j_\lambda=f_\lambda$ $(\lambda\in\Lambda)$, defined by $f(\{x_\lambda\})=\sum_{\lambda\in\Lambda} f_\lambda(x_\lambda)$. When M is an A-module and $\{N_\lambda\}_{\lambda\in\Lambda}$ is a family of A-submodules of M, the A-homomorphism $f: \sum_{\lambda\in\Lambda} N_\lambda\rightarrow M$ defined by $f(\{x_\lambda\})=\sum_{\lambda\in\Lambda} x_\lambda$ is an A-isomorphism if and only if M is the direct sum of $\{N_\lambda\}$.

If $M_\lambda=M$ for all $\lambda\in\Lambda$, $\prod_{\lambda\in\Lambda} M_\lambda$ and $\sum_{\lambda\in\Lambda} M_\lambda$ are denoted by M^Λ and $M^{(\Lambda)}$, respectively. M^Λ can be regarded as the set of all mappings of Λ to M. The direct product $M_1\times\ldots\times M_n$ and direct sum $M_1\oplus\ldots\oplus M_n$ of a finite number of A-modules $M_1,\ldots,M_n$, can be identified with each other and, if $M_i=M$ $(1\leqslant i\leqslant n)$, we simply denote it by M^n.

G. Free Modules

Let A be a ring. A family $\{x_\lambda\}_{\lambda\in\Lambda}$ of elements in an A-module M is called **linearly independent** if $\sum_{\lambda\in\Lambda} a_\lambda x_\lambda=0$ $(a_\lambda\in A)$ implies $a_\lambda=0$ for all $\lambda\in\Lambda$. This is equivalent to saying that the mapping $A^{(\Lambda)}\rightarrow M$ that assigns $\sum_{\lambda\in\Lambda} a_\lambda x_\lambda\in M$ to $\{a_\lambda\}$ is injective. A linearly independent family $\{x_\lambda\}_{\lambda\in\Lambda}$ generating M is called a **basis** of M. A family $\{x_\lambda\}_{\lambda\in\Lambda}$ is a basis if and only if every element of M can be written uniquely in the form $\sum_{\lambda\in\Lambda} a_\lambda x_\lambda$ $(a_\lambda\in A)$.

An A-module that has a basis is called a **free module** over A. If A is a field (which may be noncommutative), every A-module is a free module (→ 256 Linear Spaces). The †cardinality of a basis of a free module M over A depends only on M if A is a field (which may be noncommutative) or a commutative ring; this number is called the **rank** (or **dimension**) of M. Any submodule of a free module over a †principal ideal domain is a free module.

H. Simple Modules and Semisimple Modules

An A-module M is called **simple** if $M\neq 0$ and M has no A-submodules except M and 0. If M and N are simple A-modules, any A-homomorphism of M to N is an isomorphism or the **zero homomorphism** (i.e., one which sends every element of M to 0) (**Schur's lemma**). If an A-module M is the sum of a family $\{M_\lambda\}_{\lambda\in\Lambda}$ of simple submodules, M is the direct sum of a suitable subfamily $\{M_{\lambda'}\}_{\lambda'\in\Lambda'}$ $(\Lambda'\subset\Lambda)$. In this case, M is called **semisimple** (or **completely reducible**).

If an A-module M can be decomposed into the direct sum of A-submodules N and N', then N' is called a **complementary submodule** of N. An A-module M is semisimple if and only if every A-submodule of M has a complementary submodule. Let A be a ring. Then the A-module A is semisimple if and only if every A-module is semisimple. In this case A is called a †semisimple ring (→ 368 Rings G). Every simple module over a semisimple ring A is A-isomorphic to a †minimal left ideal of A.

I. Chain Conditions

The set of all A-submodules of an A-module M forms an †ordered set under the inclusion relation. An A-module is called a **Noetherian module** if the ordered set satisfies the †maximal condition and an **Artinian module** if it satisfies the †minimal condition (→ 311 Ordering C).

Let N be an A-submodule of an A-module M. Then M is Noetherian (Artinian) if and only if N and M/N are both Noetherian (Artinian). A ring A is called a †left Noetherian ring (†left Artinian ring) if A is Noetherian (Artinian) as a left A-module, and similarly for right Noetherian and Artinian rings. Every finitely generated module over a Noetherian (Artinian) ring is Noetherian (Artinian). Over an arbitrary ring A, a module M is Noetherian if and only if every A-submodule of M is finitely generated.

A finite sequence $\{M_i\}_{0\leqslant i\leqslant r}$ of A-submodules of an A-module M is called a †Jordan-Hölder sequence if $M=M_0$, $M_i\supset M_{i+1}$, $M_r=\{0\}$, and the M_i/M_{i+1} $(0\leqslant i<r)$ are simple. If such a sequence exists, M is said to be **of finite length**. The number r, called the **length** of M, depends only on M. The quotient modules M_i/M_{i+1} $(0\leqslant i<r)$ are uniquely determined by M up to A-isomorphism and permutation of the indices (C. Jordan and O. Hölder). An A-module M is of finite length if and only if M is Noetherian and Artinian. A semisimple A-module is of finite length if and only if it is finitely generated.

An A-module M is called **indecomposable** if M cannot be decomposed into the direct sum of two A-submodules different from M and $\{0\}$. Any A-module of finite length can be decomposed into the direct sum of a finite

sequence $N_1, \ldots, N_n$ of indecomposable A-submodules different from $\{0\}$. The direct summands N_i $(1 \leqslant i \leqslant n)$ are unique up to A-isomorphism and permutation of the indices (W. Krull, R. Remak, and O. Schmidt).

J. Tensor Products

Let A be a ring. Given a right A-module M and a left A-module N, we construct a module $M \otimes_A N$ (called the **tensor product** of M and N) and a canonical mapping $\varphi: M \times N \rightarrow M \otimes_A N$ as follows. Let F be a free $\mathbf{Z}$-module (free Abelian additive group) generated by $M \times N$, and R be the subgroup generated by the elements of the forms $(x+x', y)-(x, y)-(x', y)$, $(x, y+y')-(x, y)-(x, y')$, $(xa, y)-(x, ay)$ $(x, x' \in M, y, y' \in N, a \in A)$. We define $M \otimes_A N = F/R$, and call the natural projection φ. If we denote $\varphi(x, y)$ by $x \otimes y$, then we have $(x_1 + x_2) \otimes y = x_1 \otimes y + x_2 \otimes y$, $x \otimes (y_1 + y_2) = x \otimes y_1 + x \otimes y_2$, and $(xa) \otimes y = x \otimes (ay)$. Any element of $M \otimes_A N$ is written in the form $\sum x_i \otimes y_i$ $(x_i \in M, y_i \in N)$.

The tensor product $M \otimes_A N$ of M and N and the canonical mapping $\varphi: M \times N \rightarrow M \otimes_A N$ can be characterized as follows: For a module L, a mapping $f: M \times N \rightarrow L$ is called **biadditive** if the conditions $f(x+x', y) = f(x, y)+f(x', y)$, $f(x, y+y') = f(x, y)+f(x, y')$ hold. A biadditive mapping f satisfying the condition $f(xa, y) = f(x, ay)$ is called an A-**balanced mapping**. Then we have (i) the canonical mapping $\varphi: M \times N \rightarrow M \otimes_A N$ is A-balanced; and (ii) for any module L and any A-balanced mapping $f: M \times N \rightarrow L$, there exists a unique homomorphism $f_*: M \otimes_A N \rightarrow L$ such that $f(x, y) = f_*(x \otimes y)$ $(x \in M, y \in N)$.

A right (left) A-module can be regarded as a left (right) A°-module, where A° is the ring anti-isomorphic to A. In this sense, we have $M \otimes_A N \cong N \otimes_{A^\circ} M$.

Let A be a commutative ring. For A-modules M, N, and L, a mapping $f: M \times N \rightarrow L$ is called a **bilinear mapping** if f is biadditive and satisfies $f(ax, y) = f(x, ay) = af(x, y)$ $(a \in A, x \in M, y \in N)$. The set $\mathfrak{L}(M, N; L)$ of all bilinear mappings $M \times N \rightarrow L$ forms A-submodule of the A-module $L^{M \times N}$. A bilinear mapping $M \times N \rightarrow A$ is called a **bilinear form** on $M \times N$. The tensor product $M \otimes_A N$ becomes an A-module if we define $a(x \otimes y) = (ax) \otimes y$ $(= x \otimes (ay))$, and the canonical mapping $M \times N \rightarrow M \otimes_A N$ is bilinear. For any A-module L and bilinear mapping $f: M \times N \rightarrow L$, there exists a unique A-linear mapping $f_*: M \otimes_A N \rightarrow L$ satisfying $f(x, y) = f_*(x \otimes y)$. By this correspondence $f \leftrightarrow f_*$, we get an A-isomorphism $\mathfrak{L}(M, N; L) \cong \operatorname{Hom}_A(M \otimes_A N, L)$. If A is a field, $M \otimes_A N$ coincides with the tensor product $M \otimes N$ as a linear space (→ 256 Linear Spaces H, I).

In general, let M be a B-A-bimodule and N be a left A-module. Then $M \otimes_A N$ becomes a left B-module if we define $b(x \otimes y) = (bx) \otimes y$. Let N be an A-B-bimodule and M be a right A-module. Then $M \otimes_A N$ becomes a right B-module if we define $(x \otimes y)b = x \otimes (yb)$. In particular, we have $A \otimes_A N \cong N$, $M \otimes_A A \cong M$.

Let M, M' be right A-modules and N, N' be left A-modules. For A-homomorphisms $f: M \rightarrow M'$ and $g: N \rightarrow N'$, there exists a unique homomorphism $h: M \otimes_A N \rightarrow M' \otimes_A N'$ satisfying $h(x \otimes y) = f(x) \otimes g(y)$; h is called the **tensor product** of f, g and is denoted by $f \otimes g$. We give here some simple examples (also → Section L).

Examples. (1) Let M, N be free modules (linear spaces, for example) over a commutative ring A. If $\{x_i\}_{i \in I}$ and $\{y_j\}_{j \in J}$ are bases of M and N, respectively, $M \otimes_A N$ is also a free module with a basis $\{x_i \otimes y_j\}_{i \in I, j \in J}$. If the dimensions $\dim M$, $\dim N$ are finite, $\dim M \otimes_A N = \dim M \dim N$.

(2) For an †ideal $\mathfrak{a}$ of a commutative ring A, the †factor ring $M = A/\mathfrak{a}$ can be regarded as an A-module, and we have $M \otimes_A N \cong N/\mathfrak{a}N$. For instance, $(\mathbf{Z}/m\mathbf{Z}) \otimes_{\mathbf{Z}} (\mathbf{Z}/n\mathbf{Z}) \cong \mathbf{Z}/(m, n)\mathbf{Z}$, where (m, n) denotes the greatest common divisor of m and n.

K. Hom and ⊗

We continue to consider modules over a ring A. Concerning the direct sum and product, we have

$$\operatorname{Hom}_A\left(\sum_\lambda M_\lambda, \prod_\mu N_\mu\right) \cong \prod_{\lambda, \mu} \operatorname{Hom}_A(M_\lambda, N_\mu)$$

and

$$\left(\sum_\lambda M_\lambda\right) \otimes_A \left(\sum_\mu N_\mu\right) \cong \sum_{\lambda, \mu} (M_\lambda \otimes_A N_\mu).$$

Concerning projective and inductive limits we have

$$\operatorname{Hom}_A\left(\varinjlim M_\lambda, \varprojlim N_\mu\right) \cong \varprojlim \operatorname{Hom}_A(M_\lambda, N_\mu)$$

and

$$\left(\varinjlim M_\lambda\right) \otimes_A \left(\varinjlim N_\mu\right) = \varinjlim (M_\lambda \otimes_A N_\mu).$$

An A-homomorphism $f: M \rightarrow M'$ induces a homomorphism $\operatorname{Hom}_A(M', N) \rightarrow \operatorname{Hom}_A(M, N)$ by the assignment $g \rightarrow g \circ f$. An exact sequence $M' \rightarrow M \rightarrow M'' \rightarrow 0$ gives rise to the exact sequence

$$0 \rightarrow \operatorname{Hom}_A(M'', N) \rightarrow \operatorname{Hom}_A(M, N) \rightarrow \operatorname{Hom}_A(M', N). \quad (1)$$

An A-homomorphism $f: N \to N'$ induces a homomorphism $\mathrm{Hom}_A(M, N) \to \mathrm{Hom}_A(M, N')$ by the assignments $g \to f \circ g$, and an exact sequence $0 \to N' \to N \to N''$ gives rise to the exact sequence

$$0 \to \mathrm{Hom}_A(M, N') \to \mathrm{Hom}_A(M, N) \to \mathrm{Hom}_A(M, N''). \qquad (2)$$

Let M be a right A-module and N', N, N'' be left A-modules. An A-homomorphism f: $N \to N'$ induces the homomorphism $1_M \otimes f$: $M \otimes_A N \to M \otimes_A N'$, and an exact sequence $N' \to N \to N'' \to 0$ gives rise to the exact sequence

$$M \otimes_A N' \to M \otimes_A N \to M \otimes_A N'' \to 0. \qquad (3)$$

Exchanging left and right, we obtain similar results (→ 52 Categories and Functors B; 200 Homological Algebra).

Let Q be an A-module. If for any exact sequence of A-modules

$$0 \to M' \to M \to M'' \to 0, \qquad (4)$$

the induced sequence

$$0 \to \mathrm{Hom}_A(M'', Q) \to \mathrm{Hom}_A(M, Q) \to \mathrm{Hom}_A(M', Q) \to 0 \qquad (5)$$

is exact, then Q is called an **injective A-module**. This is equivalent to the condition that if M' is an A-submodule of an A-module M, then any A-homomorphism $M' \to Q$ can be extended to an A-homomorphism $M \to Q$. Any A-module is an A-submodule of some injective A-module, and any injective A-module is a divisible A-module. If A is a †Dedekind domain, any divisible A-module is an injective A-module.

Let P be an A-module. If for any exact sequence (4), the induced sequence

$$0 \to \mathrm{Hom}_A(P, M') \to \mathrm{Hom}_A(P, M) \to \mathrm{Hom}_A(P, M'') \to 0 \qquad (6)$$

is exact, then P is called a **projective A-module**. This is equivalent to the condition that for any surjective A-homomorphism $g: M \to M''$ and any A-homomorphism $f: P \to M''$, there exists an A-homomorphism $h: P \to M$ satisfying $g \circ h = f$. Any A-module is a factor A-module of some projective A-module. A projective A-module has no torsion element. A free A-module is a projective A-module. In general, an A-module is a projective A-module if and only if it is a direct summand of a free A-module.

Let R be a right A-module, If for any exact sequence (4), the induced sequence

$$0 \to R \otimes_A M' \to R \otimes_A M \to R \otimes_A M'' \to 0 \qquad (7)$$

is exact, then R is called a **flat A-module**. Any projective A-module is a flat A-module. A flat A-module R is called **faithfully flat** if $R \otimes_A M = \{0\}$ implies $M = \{0\}$. A flat right A-module R is faithfully flat if and only if $R \neq R\mathfrak{A}$ for any left ideal $\mathfrak{A}$ ($\neq A$) of A. Let A be a †principal ideal domain. Then an A-module R is flat if and only if R has no torsion element, and R is faithfully flat if and only if R has no torsion element and $R \neq Rp$ for any †prime element p of A. We have the following important examples:

(1) For a commutative ring A and its multiplicatively closed subset S, the †ring of quotients A_S is flat as an A-module. However, A_S is not faithfully flat. For instance, the field of rational numbers $\mathbf{Q}$ is not faithfully flat as a $\mathbf{Z}$-module.

(2) Let A be a †semilocal ring and $\bar{A}$ be its completion. Then $\bar{A}$ is faithfully flat as an A-module (→ 284 Noetherian Rings; also [1, 7]).

In the exact sequence (4), if $\mathrm{Im}\,\varphi = \mathrm{Ker}\,\psi$ is a direct summand of the A-module M, we say that (4) **splits**. Then (5), (6), and (7) are exact for any A-modules Q, P, R. The exact sequence (4) splits if M' is injective or M'' is projective.

By ${}_AM$, M_A, and ${}_AM_B$, we mean that M is a left A-module, a right A-module, and an A-B-bimodule, respectively. As already stated, ${}_AM_B$ and ${}_AN$ imply ${}_B(\mathrm{Hom}_A(M, N))$, and ${}_AM$ and ${}_AN_B$ imply $(\mathrm{Hom}_A(M, N))_B$. Similarly ${}_BM_A$ and N_A imply $(\mathrm{Hom}_A(M, N))_B$, and M_A and ${}_BN_A$ imply ${}_B(\mathrm{Hom}_A(M, N))$. Furthermore, ${}_BM_A$ and ${}_AN$ imply ${}_B(M \otimes_A N)$, and M_A and ${}_AN_B$ imply $(M \otimes_A N)_B$.

With ${}_BL_A$, ${}_AM$, and ${}_BN$, we have

$$\mathrm{Hom}_A(M, \mathrm{Hom}_B(L, N)) \cong \mathrm{Hom}_B(L \otimes_A M, N). \qquad (8)$$

Similarly, for ${}_AM_B$, L_A, and N_B, we have

$$\mathrm{Hom}_A(L, \mathrm{Hom}_B(M, N)) \cong \mathrm{Hom}_B(L \otimes_A M, N). \qquad (8')$$

If B is a commutative ring, (8) and (8′) are B-isomorphisms. Furthermore, with L_A, ${}_AM_B$, and ${}_BN$, we have

$$(L \otimes_A M) \otimes_B N \cong L \otimes_A (M \otimes_B N). \qquad (9)$$

We denote by M^* the set $\mathrm{Hom}_A(M, A)$ of all linear forms on an A-module M. Then ${}_AM$ implies M^*_A, and M_A implies ${}_AM^*$; the A-module M^* is called the **dual module** of M. A as a left A-module is dual to A as a right A-module, and vice versa. For a family of A-modules $\{M_\lambda\}_{\lambda \in \Lambda}$, we have a canonical correspondence $(\sum_{\lambda \in \Lambda} M_\lambda)^* \cong \prod_{\lambda \in \Lambda} M_\lambda^*$. From this, we have a canonical isomorphism $(M^*)^* \cong M$ for any finitely generated projective A-module M. Many facts concerning this †duality are similar to those valid for linear spaces (→ 256 Linear Spaces G).

Let A be a commutative ring. Letting $A = B = N$ in (8) and (8′), we have the canonical A-

isomorphisms

$$\mathrm{Hom}_A(M,L^*)\cong\mathrm{Hom}_A(L,M^*)$$
$$\cong(L\otimes_A M)^*=\mathfrak{L}(L,M;A),$$

namely, any bilinear form on $L\times M$ is represented by a linear mapping $M\to L^*$ or $L\to M^*$.

L. Extension and Restriction of a Basic Ring

Fix a ring homomorphism $\rho: A\to B$. We regard B as a B-A-bimodule by defining a right operation of A on B by $b\cdot a=b\rho(a)$ $(a\in A, b\in B)$. This bimodule is denoted by B_ρ.

For every left A-module M, we construct the left B-module $\rho^*(M)=B_\rho\otimes_A M$, which is called the **scalar extension** of M by ρ. Every A-homomorphism of A-modules $f: M\to M'$ induces the B-homomorphism $\rho^*(f)=1_B\otimes f:\rho^*(M)\to\rho^*(M')$.

For every left B-module N, we construct the left A-module $\rho_*(N)=\mathrm{Hom}_B(B_\rho,N)$, which is called the **scalar restriction** (or **scalar change**) of N by ρ. By the assignment $h\to h(1)$, we have a module isomorphism $\mathrm{Hom}_B(B_\rho,N)=N$. We identify $\rho_*(N)$ with N under this isomorphism. The operation of A on N is then given as $a\cdot y=\rho(a)y$ $(a\in A, y\in N)$. If A is a subring of B and ρ is the canonical injection, then the operation of A on $\rho_*(N)$ is the restriction of the operation of B on N. Every B-homomorphism of B-modules $f: N\to N'$ induces the A-homomorphism $\rho_*(f):\rho_*(N)\to\rho_*(N')$. For any left A-module M and left B-module N, an A-linear mapping $f: M\to\rho_*(N)=N$ is called a **semilinear mapping** with respect to ρ. This means that f is an additive homomorphism satisfying $f(ax)=\rho(a)f(x)$ $(a\in A, x\in M)$.

The extension and the restriction of a basic ring are related by the canonical isomorphism $\mathrm{Hom}_A(M,\rho_*(N))\cong\mathrm{Hom}_B(\rho^*(M),N)$ for an A-module M and a B-module N (→ equation (8')). An element α of the left-hand side and an element β of the right-hand side are associated by the relation $\alpha(x)=\beta(1\otimes x)$ $(x\in M)$ (→ 52 Categories and Functors).

Let A and B be commutative rings. Then for A-modules M and M', we have the canonical B-linear mapping $\rho^*: B\otimes_A\mathrm{Hom}_A(M,M')\to\mathrm{Hom}_B(B\otimes_A M, B\otimes_A M')$, which is a B-isomorphism if M is a finitely generated free (or more generally, projective) A-module. Using the notation ρ^*, we have $\rho^*(\mathrm{Hom}_A(M,M'))\cong\mathrm{Hom}_B(\rho^*(M),\rho^*(M'))$.

We now give some examples where the basic rings are noncommutative. Let G be a group and H its subgroup. Let ρ denote the homomorphism of group rings $K[H]\to K[G]$ induced by the canonical injection $H\to G$, where K is a commutative ring. For any $K[H]$-module M, $\rho^*(M)$ is called the **induced module** of M. The representation of G associated with $\rho^*(M)$ is the †induced representation of the representation of H associated with M. Next, we fix a group G and consider a homomorphism $\rho: K[G]\to\bar{K}[G]$ induced by a homomorphism of commutative rings $\sigma: K\to\bar{K}$. If $K=\bar{K}$ and σ is an automorphism, then the representation associated with the "scalar extension" $\rho^*(M)$ of a $K[G]$-module M is the †conjugate representation to the representation associated with M. If $\bar{K}=K/\mathfrak{A}$ ($\mathfrak{A}$ is an ideal of K) and σ is the canonical projection, then the representation over $\bar{K}$ associated with the scalar extension $\rho^*(M)$ of a $K[G]$-module M is the **reduction modulo** $\mathfrak{A}$ of the representation over K associated with M, and $\rho^*(M)$ is canonically isomorphic to $M/\mathfrak{A}M$. Furthermore, the †localization and the †completion can also be treated under the formulation of scalar extension (→ 67 Commutative Rings G, 284 Noetherian Rings B).

References

[1] M. Nagata, Local rings, Interscience, 1962.
[2] H. P. Cartan and S. Eilenberg, Homological algebra, Princeton Univ. Press, 1956.
[3] D. G. Northcott, An introduction to homological algebra, Cambridge Univ. Press, 1960.
[4] S. MacLane, Homology, Springer, 1963.
[5] C. Chevalley, Fundamental concepts of algebra, Academic Press, 1956.
[6] N. Bourbaki, Eléments de mathématique, Algèbre, ch. 2, Actualités Sci. Ind., 1236b, Hermann, third edition, 1962.
[7] N. Bourbaki, Eléments de mathématique, Algèbre commutative, ch. 1, 2, Actualités Sci. Ind., 1290a, Hermann, 1961.
[8] C. W. Curtis and I. Reiner, Representation theory of finite groups and associative algebras, Interscience, 1962.
[9] R. Godement, Cours d'algèbre, Hermann, 1963.
[10] S. Lang, Algebra, Addison-Wesley, 1965.
[11] S. T. Hu, Elements of modern algebra, Holden-Day, 1965.

278 (XIII.23) Monge-Ampère Equations

A. Monge-Ampère Equations

A **Monge-Ampère differential equation** is a second-order partial differential equation of the form

$$Hr+2Ks+Lt+M+N(rt-s^2)=0, \qquad (1)$$

where H, K, L, M, and N are functions of x, y, z, p, and q, and r, s, t, p, and q represent the partial derivatives

$$r=\frac{\partial^2 z}{\partial x^2},\quad s=\frac{\partial^2 z}{\partial x\partial y},\quad t=\frac{\partial^2 z}{\partial y^2},$$

$$p=\frac{\partial z}{\partial x},\quad q=\frac{\partial z}{\partial y}.$$

The characteristic manifolds are integrals of a system of differential equations defined as follows:

Case (i) $N\neq 0$.

$$N\,dp+L\,dx+\lambda_1\,dy=0,$$
$$N\,dq+\lambda_2\,dx+H\,dy=0,$$
$$dz-p\,dx-q\,dy=0, \tag{2}$$
$$N\,dp+L\,dx+\lambda_2\,dy=0,$$
$$N\,dq+\lambda_1\,dx+H\,dy=0,$$
$$dz-p\,dx-q\,dy=0, \tag{3}$$

where λ_1 and λ_2 are the two roots of the equation $\lambda^2+2K\lambda+HL-MN=0$.

Case (ii) $N=0$, $H\neq 0$.

$$dy=\lambda_1\,dx,\qquad H\,dp+H\lambda_2\,dq+M\,dx=0,$$
$$dz-p\,dx-q\,dy=0, \tag{4}$$
$$dy=\lambda_2\,dx,\qquad H\,dp+H\lambda_1\,dq+M\,dx=0,$$
$$dz-p\,dx-q\,dy=0, \tag{5}$$

where λ_1 and λ_2 are the two roots of the equation $H\lambda^2-2K\lambda+L=0$.

Case (iii) $N=0$, $H=0$, $L\neq 0$.

$$dx=0,\qquad M\,dy+2K\,dp+L\,dq=0,$$
$$dz-p\,dx-q\,dy=0, \tag{6}$$
$$2K\,dy-L\,dx=0,\qquad M\,dy+L\,dq=0,$$
$$dz-p\,dx-q\,dy=0. \tag{7}$$

Case (iv) $N=0$, $H=0$, $L=0$.

$$dx=0,\qquad 2K\,dp+M\,dy=0,$$
$$dz-p\,dx-q\,dy=0, \tag{8}$$
$$dy=0,\qquad 2K\,dq+M\,dx=0,$$
$$dz-p\,dx-q\,dy=0. \tag{9}$$

A manifold $x(\lambda)$, $y(\lambda)$, $z(\lambda)$, $p(\lambda)$, $q(\lambda)$ that satisfies the system (2), (3) of differential equations for case (i), (4), (5) for case (ii), (6), (7) for case (iii), or (8), (9) for case (iv) is a characteristic manifold of equation (1).

The following result is known concerning Monge-Ampère equations: The union of surface elements of an integral surface of (1) is generated in two ways by characteristic manifolds depending on one parameter. Conversely, if a manifold, each element of which is composed of a point of a surface S and the tangent plane at that point, is generated by a family of characteristic manifolds depending on one parameter, then the surface S is an integral surface of (1).

B. Intermediate Integrals

If a relation $dV(x,y,z,p,q)=0$ is a consequence of a system of differential equations of characteristic manifolds, for example, in case (i) when V satisfies

$$N(\partial V/\partial x+p\partial V/\partial z)-L\partial V/\partial p-\lambda_2\partial V/\partial q=0,$$
$$N(\partial V/\partial y+q\partial V/\partial z)-\lambda_1\partial V/\partial p-H\partial V/\partial q=0,$$

$V(x,y,z,p,q)=c$ (c an arbitrary constant) is called an **integral** of the system of differential equations. (i) If $V=c$ is an integral of a system of differential equations of characteristic manifolds, the solution $z(x,y)$ of $V=c$ considered anew as a partial differential equation of the first order is a solution of (1). Conversely, if every solution of $V=c$ (excepting †singular ones) satisfies equation (1), $V=c$ is an integral of a system of differential equations of characteristic manifolds. (ii) If $u(x,y,z,p,q)$ and $v(x,y,z,p,q)$ are two integrals of a system of differential equations of characteristic manifolds, then for any arbitrary function φ, $\varphi(u,v)$ is also an integral. Thus, as a consequence of (i), every solution $z(x,y)$ of $\varphi(u,v)=0$ is also a solution of (1). The converse is also true, namely, for every solution z of (1) we can find a function φ such that $z(x,y)$ is also a solution of $\varphi(u,v)=0$. The relation $\varphi(u,v)=0$ is called an **intermediate integral** of (1). Sometimes an integral of a system of differential equations of characteristic manifolds is also called an intermediate integral. If each of the two systems of differential equations defining the characteristic manifolds has an intermediate integral, then the two intermediate integrals $\varphi(u,v)=0$ and $\psi(u,v)=0$ form a †complete system of partial differential equations of the first order. Integrating this complete system, the †general solution of equation (1) is obtained.

References

[1] E. Goursat, Cours d'analyse mathématique III, Gauthier-Villars, fourth edition, 1927.
[2] E. Goursat, Leçons sur l'intégration des équations aux dérivées partielles du second ordre à deux variables indépendantes I, Hermann, 1896.

279 (VII.16) Morse Theory

A. General Remarks

We are interested in smooth functions on a smooth manifold of dimension m that have only the simplest critical points. Here, "smooth" means differentiable of class C^∞. Such a function $f: M \rightarrow \mathbf{R}$ enables us to investigate the topology of M. The decomposition of M into level sets of f contains a lot of topological information on M. For instance, it is shown that M is †homotopy equivalent to a †CW complex that is determined by critical points of f; furthermore the †Euler characteristic of M can be computed by means of f, because the critical points are related to the †homology groups of M. These types of investigations, called **Morse theory**, were originated by H. Poincaré [1] and G. D. Birkhoff [2] and then developed into the form we see today by M. Morse [3]. An excellent exposition of this theory has been given by J. Milnor [4]. R. S. Palais [5] has extended the theory to Hilbert manifolds. Morse theory has been fruitfully applied to differential topology and differential geometry.

B. Critical Points of Functions on Manifolds

For a point $p \in M$, M_p is by definition the tangent space to M at p. Let f be a smooth real-valued function on M. A point $p \in M$ is called a **critical point** of f if the induced mapping $f_*: M_p \rightarrow \mathbf{R}_{f(p)}$ is zero at p. For every †local chart $(x_1, \ldots, x_m)$ around a critical point p of f with $p=(0, \ldots, 0)$,

$$\frac{\partial f}{\partial x_1}(0)=\ldots=\frac{\partial f}{\partial x_m}(0)=0.$$

The value $f(p)$ is then called a **critical value** of f. If the matrix $(\partial^2 f/\partial x_i \partial x_j(0))$ is invertible, then the point is called a **nondegenerate critical point**, and if the matrix is not invertible, then the point is said to be **degenerate**. These notions are independent of the choice of local charts around p. The above matrix is called the **Hessian** of f at p. The nullity and index of the Hessian of f at p are called the **nullity** and **index** of f at p, respectively. The function f has a local minimum at a nondegenerate critical point of index 0, and a local maximum at a nondegenerate critical point of index n.

A smooth function f on M is called a **Morse function** if it satisfies: (A1) Every critical point of f is nondegenerate. The **Morse lemma** states that if p is a nondegenerate critical point of f and if the index of f at p is λ, then there exists a local chart $(y_1, \ldots, y_m)$ around p with $p=(0, \ldots, 0)$ such that f is expressed as

$$f(y)=f(p)-(y_1)^2-\ldots-(y_\lambda)^2+(y_{\lambda+1})^2+\ldots+(y_m)^2.$$

C. The Existence of Morse Functions

Let M be a compact smooth m-dimensional manifold without boundary. The existence of Morse functions on M is shown by †embedding M into $\mathbf{R}^n$ for a sufficiently large n. For convenience, M will be here identified with its embedded image. At each point $p \in M$ the set ν_p of all unit normals forms an $(n-m-1)$-dimensional unit sphere, and $\nu(M)=\bigcup_{p \in M} \nu_p$ is a smooth $(n-1)$-dimensional compact manifold. Let S be the unit hypersphere in $\mathbf{R}^n$ around the origin. The Gauss normal mapping $\varphi: \nu(M) \rightarrow S$ sends each point $(p, v(p)) \in \nu(M)$ to $v \in S$ via a canonical parallel translation of $\mathbf{R}^n$. Since φ is smooth, the †Sard theorem implies that the set E of all critical values of φ has measure zero on S. Thus φ_* at every $(p, v(p)) \in \varphi^{-1}(v)$ has maximal rank if v is taken in $S-E$. For every fixed $v \in S-E$ let $h_v: M \rightarrow \mathbf{R}$ be defined by $h_v(x)=\langle v, x\rangle, x \in M$, where $\langle\ ,\ \rangle$ denotes the canonical inner product of $\mathbf{R}^n$. Then h_v is a Morse function. To see this let $d\Sigma$, dv, and dM be volume elements of S, ν, and M, respectively. Then $dv \wedge dM$ is a volume element of $\nu(M)$, and the function $G: \nu(M) \rightarrow \mathbf{R}$ is obtained by the relation $\varphi^* d\Sigma=G(x, v(x))\, dv \wedge dM$. $G(x, v(x))$ is called the **Lipschitz-Killing curvature** at $v(x) \in \nu_x$. If $A(v(x))$ denotes the †second fundamental form of M with respect to $\nu(x)$, then $G(x, v(x))=(-1)^m \det(A(v(x)))$ (Chern and Lashof [6]). Every critical point $p \in M$ of h_v for $v \in S-E$ has a normal $v(p)$ to M at p, and the Hessian of h_v is given by $A(v(p))$. Thus $G(p, v(p)) \neq 0$ ensures the nondegeneracy at p [6].

Another approach to the construction of Morse functions is to find for a given embedding of M into $\mathbf{R}^n$ an open dense set $U \subset \mathbf{R}^n$ such that the Euclidean distance d_x from any fixed point x on U to points on M has no degenerate critical points on M. In this case $d_x^{-1}((-\infty, a])$ is compact for all $a \in R$. It is also possible to construct Morse functions on compact manifolds with boundary as well as on noncompact manifolds. But a striking result of A. Phillips states that there exists a Morse function on every noncompact manifold which has no critical points [7].

For a pair of smooth manifolds M, N let $C^\infty(M, N)$ be the set of all smooth mappings from M to N. The **weak** (or **compact-open**) C^∞

topology on $C^\infty(M, N)$ is generated by the sets defined as follows. Let $f \in C^\infty(M, N)$, and let (U, φ) and (V, ψ) be charts of M and N, respectively. Let $K \subset U$ be a compact set such that $f(K) \subset V$. For $\varepsilon \in (0, \infty)$, a weak †subbasic neighborhood $N^\infty(f; (U, \varphi), (V, \psi), K, \varepsilon)$ of f is defined to be the set of all smooth mappings $g \in C^\infty(M, N)$ such that (1) $g(K) \subset V$, (2) $\|D^k(\varphi \circ f \circ \psi^{-1})(x) - D^k(\varphi \circ g \circ \psi^{-1})(x)\| < \varepsilon$, $x \in \varphi(K)$, $k = 1, 2, \ldots$, where D^k denotes the kth derivative with respect to the coordinates. The weak C^∞ topology is thus defined on $C^\infty(M, N)$, and the space with this topology is denoted by $C^\infty_w(M, N)$. It follows from the construction of the embedding in the Whitney embedding theorem that the set of all smooth embeddings of compact M into $\mathbf{R}^n$ is dense in $C^\infty_w(M, \mathbf{R}^n)$ if $n > 2 \dim M$. Moreover, the set of all Morse functions on a compact M forms an open dense set in $C^\infty_w(M, \mathbf{R})$ (Nagano [8], Hirsh [9], Auslander and MacKenzie [10]).

D. Decomposition of M by a Morse Function

To decompose M into levels of a Morse function we now consider the process of †attaching a handle. Let M be a compact manifold with †boundary ∂M. Let D^s be the s-disk, and let $g: (\partial D^s) \times D^{m-s} \to \partial M$ be an †embedding. Then a †manifold $X(M; g; s)$ with a handle attached by g is defined as the quotient set obtained from the disjoint union $M \cup (D^s \times D^{m-s})$ by identifying points in $\partial D^s \times D^{m-s}$ and their images under g and equipped with a natural differentiable structure. Similarly, if $g_i: (\partial D^{s_i}_i) \times D^{m-s_i}_i \to \partial M$ $(i = 1, \ldots, k)$ are embeddings with disjoint images, we can define the †handlebody $X(M; g_1, \ldots, g_k; s_1, \ldots, s_k)$ [11].

For an $f \in C^\infty(M, R)$ and for $a, b \in f(M)$ with $a < b$ let $M^a = \{x \in M \mid f(x) \leqslant a\}$ and $M^b_a = \{x \in M \mid a \leqslant f(x) \leqslant b\}$. M^a_a is called a **level set** of f. For a Morse function f on a compact manifold M without boundary, the following fundamental facts are known [4]: (1) The **first fundamental theorem**. If M^b_a contains no critical points, then M^b_a is diffeomorphic to $M^a_a \times [a, b]$; (2) The **second fundamental theorem**. Let $c \in (a, b)$ be a unique critical value in $[a, b]$, and let $p_1, \ldots, p_k \in M^c_c$ be critical points of f with p_i having index λ_i. Then M^b is diffeomorphic to $X(M^a; f_1, \ldots, f_k, \lambda_1, \ldots, \lambda_k)$ for suitable embeddings $f_1, \ldots, f_k$. It follows from the existence of Morse functions and from (1) and (2) that every compact manifold can be obtained by successively attaching handles to a disk (→ 114 Differential Topology). Furthermore, if M admits a Morse function with only two critical points, then M is homeomorphic to a sphere (Reeb [12]).

Concerning the homotopy types of M and M^a we have the following: (3) If $c \in f(M)$ is a critical value of f with points $p_1, \ldots, p_k$ with p_i having index λ_i and if $M^{c+\varepsilon}_{c-\varepsilon}$ is compact for $\varepsilon > 0$ and contains no critical points other than $p_1, \ldots, p_k$, then $M^{c+\varepsilon}$ has the same homotopy type as $M^{c-\varepsilon} \cup e^{\lambda_1} \cup \ldots \cup e^{\lambda_k}$, where e^{λ_j} $(j = 1, \ldots, k)$ are λ_j-cells. (4) Assume that a Morse function f on a (not necessarily compact) M satisfies: (A2) M^a is compact for all $a \in \mathbf{R}$. Then by choosing a sequence $a_1 < a_2 < \ldots$ of regular values of f and applying (3) to each $M^{a_{i+1}}_{a_i}$, we see that M has the same homotopy type as that of a CW complex ζ, where the number of λ-cells belonging to ζ is the same as that of critical points of index λ of f.

Moreover, we have the following Morse inequalities: (5) Let f be a Morse function on a compact M. If M_λ is the number of critical points of f on M of index λ, and if R_λ is a λ-dimensional †Betti number of M, then

$$M_0 \geqslant R_0,$$

$$M_1 - M_0 \geqslant R_1 - R_0,$$

$$M_\lambda - M_{\lambda-1} + \ldots + (-1)^\lambda M_0$$

$$\geqslant R_\lambda - R_{\lambda-1} + \ldots + (-1)^\lambda R_0, \qquad 1 < \lambda < n-1,$$

$$M_n - M_{n-1} + \ldots + (-1)^n M_0$$

$$= R_n - R_{n-1} + \ldots + (-1)^n R_0.$$

In particular, we have $M_k \geqslant R_k$ for all k. By using these facts S. Smale obtained an affirmative solution of the †Poincaré conjecture in high dimensions [11, 13].

The concept of critical manifolds in the sense of Bott [14] is stated as follows: Let M be a compact manifold embedded into an open set $U \subset \mathbf{R}^k$. Let $f: U \to \mathbf{R}$ be a smooth function. M is called a **nondegenerate critical manifold** of f on U if (1) all points of M are critical points of f and (2) the nullity of all $x \in M$ is equal to $\dim M$. When M is such a nondegenerate critical manifold of f on U, f is constant on M, and the index λ of f at x is well defined and is the same at all points on M. The †Poincaré polynomial is expressed as $P(M; t) = \sum t^k \dim H^k(M)$. The Morse polynomial is defined by $\mathfrak{M}(f; t) = \sum_N t^{\lambda_N} P(N; t)$, where N runs over all critical manifolds of f and λ_N is the index of N. Under certain conditions for orientability along each nondegenerate critical manifold of f, we have again the **Morse inequality**

$$\mathfrak{M}(f; t) \geqslant P(M; t).$$

E. Morse Theory on Hilbert Manifolds

Let M be a †Hilbert manifold, and let $f: M \to \mathbf{R}$ be a smooth function. The Hessian $\partial^2 f_p$ of f at a critical point of f is a symmetric bilinear

form on M_p given by

$$\partial^2 f_p(u,v) = d^2(f\circ\varphi^{-1})(d\varphi_p(u), d\varphi_p(v)), \qquad u,v\in M_p,$$

where φ is a coordinate mapping of a local chart around p. The **index** (**coindex**) of f at a critical point $p\in M$ of f is defined to be the supremum of dimensions of subspaces of M_p on which $\partial^2 f_p$ is negative (positive) definite. The self-adjoint bounded operator A which represents $\partial^2 f_p$ is given by $\partial^2 f_p(u,v) = \langle u, A(v)\rangle_p$. If A is invertible, then f is said to be **nondegenerate** at p. These notions do not depend on the choice of local charts around p.

The Morse lemma has been generalized to a Riemannian Hilbert manifold M by R. Palais and S. Smale [15] as follows. Let f be a smooth function defined on M. If $x\in M$ is a nondegenerate critical point of f then there exist a chart φ around x and a projection P such that for y near x

$$f(y) = f(x) + \|P\varphi(y)\|^2 - \|(I-P)\varphi(y)\|^2.$$

The above result has been extended to critical manifolds [16]. A connected submanifold N of a Hilbert manifold M is called a **nondegenerate critical manifold** of $f: M\to\mathbf{R}$ if (1) every point $p\in M$ is a critical point of f and (2) for each $p\in M$ there exists a closed subspace E_p of M_p such that $M_p = N_p\oplus E_p$ and the Hessian form restricted to E_p is nondegenerate. Let $\nu = (\pi, E, N)$ be a smooth Hilbert-space bundle over a compact connected manifold N, and let $\langle\ ,\ \rangle$ be a Riemannian structure for ν. Assume that a smooth function $f: E\to\mathbf{R}$ has the zero section of ν as a nondegenerate critical manifold. Then there exist a tubular neighborhood $\nu_\varepsilon = \{v\in E \mid \|v\| < \varepsilon\}$ of the zero section in ν and a fiber-preserving diffeomorphism $\psi: \nu_\varepsilon\to\psi(\nu_\varepsilon)$ and an orthogonal bundle projection P such that for $v\in\nu_\varepsilon$

$$f\circ\psi(v) = f(N) + \|Pv\|^2 - \|(I-P)v\|^2.$$

Let us assume that (B) M is a complete Riemannian manifold, and (C) (**Palais-Smale condition**) if f is bounded on a set $S\subset M$ and if the norm $\|\nabla f\|$ of the †gradient vector of f has infimum 0 on S, then there is a critical point of f on the closure $\bar{S}$ of S. Since M is not locally compact, condition (C) is required. In order to prove the first and second fundamental theorems of Morse theory it is necessary that the integral curves of ∇f exist, and (C) ensures their existence. Namely, under conditions (A1), (B), and (C), one of the following two facts follows for any $b\in f(M)$ and for any regular point $p\in M^b$: (1) The integral curve $c:[0,r]\to M^b$ of ∇f with $c(0)=p$ exists, and $f\circ c(r) = b$ holds for some $r\in[0,\infty)$; (2) the integral curve $c:[0,\infty)\to M^b$ of ∇f exists and $\lim_{t\to\infty} c(t)$ exists such that the limit point $c(\infty) = \lim_{t\to\infty} c(t)\in M^b$ is a critical point of f. These facts imply that the first fundamental theorem of Morse theory holds if M_a^b does not contain any critical point of f, and also the second fundamental theorem holds if M_a^b has only nondegenerate critical points on M_c^c for some $c\in(a,b)$. Furthermore, assume that $f: M\to\mathbf{R}$ and M satisfies (A1), (B), and (C), and let a, $b\in f(M)$ be regular values with $a<b$. If for each nonnegative integer λ, R_λ is the λth Betti number of M_a^b, and if M_λ is the number of critical points of index λ of f in M_a^b, then the **Morse inequality** holds:

$$M_0 \geqslant R_0,$$

$$M_1 - M_0 \geqslant R_1 - R_0,$$

$$\ldots,$$

$$\sum_{k=0}^{\lambda} (-1)^{\lambda-k} M_k \geqslant \sum_{k=0}^{\lambda} (-1)^{\lambda-k} R_k,$$

$$\ldots,$$

$$\sum_{k=0}^{\infty} (-1)^k M_k = \sum_{k=0}^{\infty} (-1)^k R_k.$$

In particular, $M_\lambda \geqslant R_\lambda$ holds for all λ. If f is bounded below, then the λth Betti number R_λ^* of M^b and the number M_λ^* of all critical points of f with index λ in M^b satisfy $M_\lambda^* \geqslant R_\lambda^*$ for all λ.

F. Morse Theory of Path Spaces

Morse theory on Hilbert manifolds applies to the energy functions on Riemannian Hilbert manifolds which consist of all H_1-curves on a compact Riemannian manifold M, and the theory is useful for proving the existence of closed geodesics on M.

Let M be a smooth manifold, and let $I = [0,1]$. Let $H_1(I,M)$ be the set of all continuous curves $\sigma: I\to M$ such that for each local chart (V,φ) of M, $\varphi\circ\sigma$ is absolutely continuous and $\|(\varphi\circ\sigma)'\|$ is locally square integrable. In particular, if $M = \mathbf{R}^m$, then $H_1(I,\mathbf{R}^m)$ is a Hilbert space with the inner product

$$(\sigma,\rho)_1 := (\sigma(0),\rho(0)) + \int_I (\sigma(t),\rho(t))\,dt, \qquad \sigma,\rho\in H_1(I,\mathbf{R}^m).$$

For each $\sigma\in H_1(I,M)$, set $H_1(I,M)_\sigma = \{X\in H_1(I,TM) \mid X(t)\in M_{\sigma(t)}, t\in I\}$, where TM is by definition the †tangent bundle over M. For any fixed pair of points $p,q\in M$, let $\Omega(M;p,q) = \{\sigma\in H_1(I,M) \mid \sigma(0)=p, \sigma(1)=q\}$, and for each $\sigma\in\Omega(M;p,q)$, let $\Omega(M;p,q)_\sigma = \{X\in H_1(I,M)_\sigma \mid X(0)=0\in M_p, X(1)=0\in M_q\}$. Then $H_1(I,M)_\sigma$ forms a vector space, and $\Omega(M;p,q)_\sigma$ is a subspace of $H_1(I,M)_\sigma$. M can be embedded into $\mathbf{R}^n$ for sufficiently large n by the Whitney

embedding theorem [17]. Then we have the following facts [17]: (1) $H_1(I,M)=\{\sigma\in H_1(I,\mathbf{R}^n)\mid\sigma(I)\subset M\}$, (2) $H_1(I,M)$ is a closed submanifold of $H_1(I,\mathbf{R}^n)$, (3) $\Omega(M;p,q)$ is a closed submanifold of $H_1(I,M)$, (4) the tangent space to $H_1(I,M)$ at each point $\sigma\in H_1(I,M)$ is $H_1(I,M)_\sigma$, (5) the tangent space to $\Omega(M;p,q)$ at each point $\sigma\in\Omega(M;p,q)$ is $\Omega(M;p,q)_\sigma$. Here M is identified with the embedded image in $\mathbf{R}^n$. Thus $H_1(I,M)$ and $\Omega(M;p,q)$ carry the structure of a Hilbert manifold, and they are independent of the choice of embeddings [17].

The energy functions on $H_1(I,M)$ and $\Omega(M;p,q)$ play important roles in the development of Morse theory. M is now assumed to be a complete Riemannian manifold. It follows from the Nash isometric embedding theorem (→ 365 Riemannian Submanifolds B) that M is isometrically embedded into $\mathbf{R}^{n'}$ for a sufficiently large n'. $H_1(I,\mathbf{R}^{n'})$ carries a complete Riemannian metric in a natural way, and $H_1(I,M)$ becomes a complete Riemannian manifold with the metric induced from $H_1(I,\mathbf{R}^{N'})$. Similarly $\Omega(M;p,q)$ admits a complete Riemannian metric. Then the **energy function** $E:H_1(I,M)\to\mathbf{R}$ is defined to be

$$E(\sigma):=\frac{1}{2}\int_I\langle\sigma',\sigma'\rangle\,dt,$$

where $\langle\ ,\ \rangle$ is the inner product induced from the Riemannian structure of M. Then the Palais-Smale condition (C) is satisfied for E on $H_1(I,M)$. That is, if a sequence $\{\sigma_k\}$ on $H_1(I,M)$ satisfies: (1) $\{E(\sigma_k)\}$ is bounded above; and (2) $\nabla E(\sigma_k)\to 0$ as $k\to\infty$, then there is a subsequence $\{\sigma_{k'}\}$ of $\{\sigma_k\}$ such that $\{\sigma_{k'}\}$ converges to a critical point $\sigma\in H_1(I,M)$ of E. Also, condition (C) is fulfilled for the energy function on $\Omega(M;p,q)$.

For a fixed pair of points $p,q\in M$, let $\Omega=\Omega(M,p,q)$. Let $\sigma\in\Omega$ be a critical point of E. A tangent vector W to Ω at $\sigma:I\to M$ with $\sigma(0)=p$, $\sigma(1)=q$ is a †piecewise differentiable vector field along σ such that $W(0)=0\in M_p$ and $W(1)=0\in M_q$. A **proper variation** $\alpha:(-\varepsilon,\varepsilon)\times I\to M$ along σ which is associated with W satisfies the following: (1) $\alpha(0,t)=\sigma(t)$, $t\in I$; (2) there exists a finite partition $0=t_0<t_1<\ldots<t_k=1$ of I such that $\alpha|(-\varepsilon,\varepsilon)\times[t_{i-1},t_i]$ is differentiable for all $i=1,\ldots,\kappa$; (3) $\alpha(u,0)=p$ and $\alpha(u,1)=q$ hold for all $u\in(-\varepsilon,\varepsilon)$; and (4) $\partial\alpha(0,t)/\partial u=W(t)$, $t\in I$. It follows from the first variation formula (→ 178 Geodesics A) that $\sigma\in\Omega$ is a critical point of E if and only if σ is a geodesic on M. Let W_1, $W_2\in\Omega_\sigma$, and let U be an open set in $\mathbf{R}^2$ around the origin. Then a **proper variation** $\alpha:U\times I\to M$ along σ that is associated with W_1 and W_2 satisfies the following: (1) $\alpha(0,0,t)=\sigma(t)$, $t\in I$; (2) there exists a finite partition $0=t_0<t_1<\ldots<t_k=1$ of I such that $\alpha|U\times[t_{i-1},t_i]$ is differentiable for all $i=1,\ldots,k$; (3) $\alpha(u_1,u_2,0)=p$, $\alpha(u_1,u_2,1)=q$ for all $(u_1,u_2)\in U$; and (4) $\partial\alpha(0,0,t)/\partial u_1=W_1(t)$, $\partial\alpha(0,0,t)/\partial u_2=W_2(t)$, $t\in I$. Then the **Hessian** E_{**} of E at σ is given by

$$E_{**}(W_1,W_2)=\frac{1}{2}\frac{\partial^2E(\bar\alpha(u_1,u_2))}{\partial u_1\partial u_2}\bigg|_{(0,0)},$$

where $\bar\alpha(u_1,u_2)\in\Omega$ is by definition the curve $\bar\alpha(u_1,u_2)(t)=\alpha(u_1,u_2,t)$. The second variation formula (→ 178 Geodesics A) then gives

$$E_{**}(W_1,W_2)=-\sum\langle W_2(t),\Delta_tW_1'(t)\rangle-\int_I\langle W_2,W_1''+R(W_1,\sigma')\sigma'\rangle\,dt,$$

where $\Delta_tW_1'(t)=W_1'(t+0)-W_1'(t-0)$. We have the **Morse index theorem**, which states: (1) The †null space of E at a critical point σ is the linear space spanned by all Jacobi fields (→ 178 Geodesics A) along σ that vanish at 0 and 1; (2) if $\{\sigma(s_1),\sigma(s_2),\ldots,\sigma(s_l)\}$ $(0<s_1<s_2<\ldots<s_l<1)$ is the set of all points conjugate to $\sigma(0)$ along σ and if λ_i is the multiplicity of the conjugate point $\sigma(s_i)$, then the index of E at σ is equal to $\lambda_1+\ldots+\lambda_l$. It follows from the Sard theorem together with the differentiability of the exponential mapping on M that except for a set of measure zero in $M\times M$, p, q can be chosen so that p, q is not a conjugate pair along any geodesic in Ω. Then all critical points of E are nondegenerate, and for any $c>0$, $\Omega^c=\{\omega\in\Omega\mid E(\omega)=c\}$ contains at most finitely many nondegenerate critical points with finite indices.

G. Existence of Closed Geodesics

Let M be a compact Riemannian manifold. By replacing I with the circle S^1, we consider a Hilbert manifold $\Lambda(M)=H_1(S^1,M)$. $\Lambda(M)$ carries the structure of a complete Riemannian manifold. Every point $p\in M$ is naturally embedded in $\Lambda(M)$ as a point curve, and M is a totally geodesic submanifold of $\Lambda(M)$. A point $\sigma\in\Lambda(M)$ is a critical point of the energy function $E:\Lambda(M)\to\mathbf{R}$ if and only if σ is a closed geodesic of M. It is known that E satisfies the Palais-Smale condition (C). The index and nullity of E at a critical point σ is finite. Since M is compact, the fundamental length $\sqrt{c}$ of M is positive and $\Lambda^c=\{\sigma\in\Lambda(M)\mid E(\sigma)\leqslant c\}$ is a †deformation retract of $M\subset\Lambda(M)$ by means of the deformation along the integral curves of $-\nabla E$.

If M is not simply connected, then each nontrivial element of the †fundamental group of M represents a class of homotopic curves in

which there is a closed geodesic whose length realizes the infimum of all these curves. If M is simply connected, then there is a minimum integer k for which $\pi_k(M)\neq 0$. Clearly, $\pi_{k-1}(\Lambda(M))\approx\pi_k(M)\neq 0$. Suppose that E has no positive critical value. Then $\Lambda(M)$ is a deformation retract of M, a contradiction. Therefore there exists at least one closed geodesic on every compact Riemannian manifold (Lyusternik and Fet [18]).

Proof of the existence of many closed geodesics on M is difficult due to the following: (1) There is a continuous $O(2)$ action on $\Lambda(M)$ that assigns $\sigma\in\Lambda(M)$ and $\alpha\in O(2)$ to the curve $t\to\sigma(t+\alpha)$, $t\in S^1$; (2) for each integer k and for each critical point $\sigma\in\Lambda(M)$, the curve $t\to\sigma(kt)$ is a critical point with energy $k^2E(\sigma)$.

A remarkable result has been obtained by Lyusternik and Shnirel'man [19], which states that there are at least three closed geodesics on every simply connected compact manifold of dimension 2. Fet then proved that there exist at least two closed geodesics on a compact manifold if all critical points of E on $\Lambda(M)$ are nondegenerate [20]. By developing a precise argument concerning the Morse lemma around an isolated degenerate critical point $\sigma\in\Lambda(M)$ of E, Gromoll and Meyer have proved that there exist infinitely many closed geodesics if the sequence of Betti numbers $\{b_i(\Lambda(M))\}$ with respect to any field is unbounded [21]. If $A:M\to M$ is a certain isometry, then there are also infinitely many A-invariant closed geodesics if the Betti numbers of the space of A-invariant H_1-curves are not bounded [22, 23]. By investigating the Z_2-cohomology of $\Lambda(M)$ of compact symmetric spaces, Ziller has proved that if M has the same homotopy type as that of a symmetric space of rank $\geqslant 2$, then M has infinitely many closed geodesics [24].

Many attempts have been made by W. Klingenberg and others to prove the existence of infinitely many geometrically distinct closed geodesics on every compact Riemannian manifold [25].

References

[1] H. Poincaré, Sur les lignes géodésiques des surfaces convexes, Trans. Amer. Math. Soc., 6 (1905), 237–274. (Oeuvres, Gauthier-Villars, 1953, vol. 6, 38–85.)
[2] G. D. Birkhoff, Dynamical systems, Amer. Math. Soc. Colloq. Publ., 1927.
[3] M. Morse, Calculus of variations in the large, Amer. Math. Soc. Colloq. Publ., 1934.
[4] J. Milnor, Morse theory, Ann. math. studies 51, Princeton Univ. Press, 1963.
[5] R. S. Palais, Morse theory on Hilbert manifolds, Topology, 2 (1963), 299–340.
[6] S. S. Chern and R. Lashof, On the total curvature of immersed manifolds, Amer. J. Math., 79 (1959), 306–318.
[7] A. Phillips, Submersions of open manifolds, Topology, 6 (1967), 171–206.
[8] T. Nagano, Variational theory in the large (in Japanese), Kyôritu, 1971.
[9] M. W. Hirsch, Immersions of manifolds, Trans. Amer. Math. Soc., 93 (1959), 242–276.
[10] L. Auslander and R. E. MacKenzie, Introduction to differentiable manifolds, McGraw-Hill, 1963.
[11] S. Smale, Generalized Poincaré's conjecture in dimensions greater than four, Ann. Math., (2) 74 (1961), 391–406.
[12] G. Reeb, Sur certains propriétés topologiques des variétés feuilletees, Actualités Sci. Ind., 11 (1952), 91–154.
[13] J. Milnor, Lectures on the h-cobordism theorem, Princeton Univ. Press, 1965.
[14] R. Bott, Nondegenerate critical manifolds, Ann. Math., (2) 60 (1954), 248–261.
[15] R. Palais and S. Smale, A generalized Morse theory, Bull. Amer. Math. Soc., 70 (1964), 165–172.
[16] W. Meyer, Kritische Mannigfaltigkeiten in Hilbertmannigfaltigkeiten, Math. Ann., 170 (1967), 45–66.
[17] J. T. Schwarz, Nonlinear functional analysis, Gordon & Breach, 1969.
[18] L. Lyusternik and A. I. Fet, Variational problems on closed manifolds (in Russian), Dokl. Akad. Nauk SSSR, 81 (1951), 17–18.
[19] L. Lyusternik and L. Shnirel'man, Méthodes topologiques dans les problèms variations, Actualités Sci. Ind., Hermann, 1934.
[20] A. I. Fet, Variation problems on closed manifolds, Amer. Math. Soc. Transl., ser. 1, 6 (1962), 147–206. (Original in Russian, 1952.)
[21] D. Gromoll and W. Meyer, Periodic geodesics on compact Riemannian manifolds, J. Differential Geom., 3 (1969), 493–510.
[22] K. Grove, Condition (C) for the energy integral on certain path spaces and applications to the theory of geodesics, J. Differential Geom., 8 (1973), 207–223.
[23] M. Tanaka, On the existence of infinitely many isometry-invariant geodesics, J. Differential Geometry, 17 (1982) 171–184.
[24] W. Ziller, Geschlossene Geodätische auf global symmetrischen und homogen Räume, Bonner Math. Schr., 85, 1976.
[25] W. Klingenberg, Lectures on closed geodesics, Grundlagen der math. Wiss. 230, Springer, 1978.

280 (XVIII.9) Multivariate Analysis

A. General Remarks

Multivariate analysis consists of methods of statistical analysis of multivariate observations represented by a collection of points in a finite-dimensional Euclidean space $\mathbf{R}^p$. With the development of powerful computers, multivariate techniques are beginning to be utilized in many fields of science and technology.

B. The Multivariate Linear Model

The multivariate linear model is an immediate extension of the univariate linear model. Suppose that $\mathbf{X}=(\mathbf{X}^{(1)}\ldots\mathbf{X}^{(n)})$ denotes the $p\times n$ matrix of n observations of p-dimensional data. Suppose that it can be expressed as

$$\mathbf{X}=BZ+\mathbf{U}, \qquad (1)$$

where B is a $p\times m$ matrix of unknown parameters, Z is a known $m\times n$ matrix of independent variables, and $\mathbf{U}$ is a $p\times n$ matrix of errors. We assume that (i) the †expectations of the elements of $\mathbf{U}$ are zero, that is, the $p\times n$ matrix $E(\mathbf{U})=0$, and call the relation (1) a **multivariate linear model**. We usually assume further that (ii) the column vectors $\mathbf{U}^{(i)}$, $i=1,\ldots,n$, of $\mathbf{U}$ are independent and identically distributed, and that (iii) $\mathbf{U}^{(i)}$ is distributed according to a multivariate †normal distribution with †covariance matrix Σ. Analogous to the univariate case, the †least squares estimator $\hat{\mathbf{B}}$ of B is defined to be the $p\times m$ matrix that minimizes

$$\operatorname{tr}(\mathbf{X}-BZ)(\mathbf{X}-BZ)'$$

and is given explicitly by

$$\hat{\mathbf{B}}=\mathbf{X}Z'(ZZ')^{-1} \quad \text{when} \quad |ZZ'|\neq 0.$$

Here the symbol $'$ means the transpose of a matrix. Also, an unbiased estimator of Σ is given by

$$\hat{\Sigma}=\mathbf{Q}_e/(n-m), \qquad \mathbf{Q}_e=\mathbf{X}\mathbf{X}'-\hat{\mathbf{B}}Z\mathbf{X}'.$$

$\hat{\mathbf{B}}$ is an unbiased estimator of B under the assumption (i) above and is the †best linear unbiased estimator under (i) and (ii), while $\hat{\Sigma}$ is unbiased when (i) and (ii) are assumed. Under the assumptions (i)–(iii), $\hat{\mathbf{B}}$ and $\hat{\Sigma}$ form a set of complete †sufficient statistics; hence they are †uniformly minimum variance unbiased estimators. Also under (i)–(iii), elements of $\hat{\mathbf{B}}$ are normally distributed, and their covariance can be expressed by

$$\Sigma\otimes M \quad (\otimes \text{ denotes the †Kronecker product}),$$

where $M=(ZZ')^{-1}$. Applying †Cochran's theorem for the multivariate case, $\mathbf{Q}_e$ is shown to be distributed according to a †Wishart distribution with $n-m$ degrees of freedom.

To test the hypothesis $B=B_0$ under (i)–(iii), we put $\mathbf{Q}_B=(\hat{\mathbf{B}}-B_0)ZZ'(\hat{\mathbf{B}}-B_0)'$ and have $(\mathbf{X}-B_0Z)(\mathbf{X}-B_0Z)'=\mathbf{Q}_B+\mathbf{Q}_e$, where $\mathbf{Q}_B$ and $\mathbf{Q}_e$ are independently distributed. The distribution of $\mathbf{Q}_B$ is a Wishart distribution with m degrees of freedom when the hypothesis is true, and a †noncentral Wishart distribution when $B\neq B_0$. Based on this fact, several procedures have been proposed. If we require the invariance of procedures with respect to linear transformations of the coordinates of p-dimensional vectors, the roots $\lambda_1,\ldots,\lambda_p$ of the †characteristic equation $|\mathbf{Q}_B-\lambda\mathbf{Q}_e|=0$ form a †maximal invariant statistic; hence the testing procedures should be defined in terms of these roots (→ 396 Statistic I). Also, the consideration of †power leads to procedures that reject the hypothesis when these roots are large. Commonly used test statistics are (1) the †likelihood ratio test $W=|\mathbf{Q}_e|/|\mathbf{Q}_B+\mathbf{Q}_e|=\prod_i(1+\lambda_i)^{-1}$ (S. S. Wilks); (2) $\operatorname{tr}\mathbf{Q}_e^{-1}\mathbf{Q}_B=\sum\lambda_i$ (D. N. Lawley and H. Hotelling); (3) $\max\lambda_i$ (S. N. Roy); (4) $\operatorname{tr}\mathbf{Q}_B(\mathbf{Q}_B+\mathbf{Q}_e)^{-1}=\sum\lambda_i(1+\lambda_i)^{-1}$ (K. C. S. Pillai). Pillai's trace test (4) is locally the most powerful invariant; Wilks's likelihood ratio test (1) has the maximum Bahadur efficiency. The power functions of the tests (1)–(3) have the monotonicity property, namely, they are monotonically nondecreasing with respect to each eigenvalue of $\Omega=\Sigma^{-1}(B-B_0)ZZ'(B-B_0)'$, the matrix of noncentrality parameters for $\mathbf{Q}_B$. The monotonicity for Pillai's test (4) is known to hold only for restricted cases where the critical value c for the acceptance region $\operatorname{tr}\mathbf{Q}_B(\mathbf{B}_B+\mathbf{Q}_e)^{-1}\leqslant c$ should satisfy $0\leqslant c\leqslant 1$. All the tests (1)–(4) are unbiased. With respect to 0–1 loss, the tests (1) and (4) are †admissible Bayes and the tests (2) and (3) are †admissible. Tests based on $\min\lambda_i$ are inadmissible.

Small-sample distributions of these statistics are complicated but when $n\to\infty$, $n\log W$, $n\operatorname{tr}\mathbf{Q}_B\mathbf{Q}_e^{-1}$, and $n\operatorname{tr}\mathbf{Q}_B(\mathbf{Q}_B+\mathbf{Q}_e)^{-1}$ are asymptotically distributed according to a chi-square distribution with pm degrees of freedom under the null hypothesis. Even under the alternative hypothesis that the matrix of noncentrality parameters $\Omega=o(1)$ for large n, the asymptotic distributions remain the same. When $\Omega=O(1)$, they are noncentral chi-square distributions of pm degrees of freedom and noncentrality parameter $\delta=\operatorname{tr}\Omega$. If $\Omega=O(n)$, they are normal distributions, namely, $-\sqrt{n}(\log W+\log|I+\theta|)$, $\sqrt{n}(\operatorname{tr}\mathbf{Q}_B\mathbf{Q}_e^{-1}-\operatorname{tr}\theta)$ and $\sqrt{n}(\operatorname{tr}\mathbf{Q}_B(\mathbf{Q}_B+\mathbf{Q}_e)^{-1}-\operatorname{tr}\theta(I+\theta)^{-1})$ for $\theta=\lim\Omega/n$ have asymptotically normal distributions of zero means and variances given by $2\operatorname{tr}(I-(I+\theta)^{-2})$, $2\operatorname{tr}(2\theta+\theta^2)$, and

$2\operatorname{tr}((I+\theta)^{-2}-(I+\theta)^{-4})$, respectively. As a special case, if $m=1$ there exists only one nonzero λ, and the procedures in this paragraph all coincide and are equivalent to one based on $T^2=M^{-1}(\hat{\mathbf{B}}-B_0)'\hat{\Sigma}^{-1}(\hat{\mathbf{B}}-B_0)$. It is known that under the hypothesis, $(n-p)T^2/(n-1)p$ is distributed according to an †F-distribution with $(p, n-p)$ degrees of freedom. When $p=2$ (resp. $m=2$), $(n-m-1)(1-\sqrt{W})/(m\sqrt{W})$ (resp. $(n-p-1)(1-\sqrt{W})/(p\sqrt{W})$) is distributed according to an F-distribution with degrees of freedom $(2m, 2(n-m-1))$ (resp. $(2p, 2(n-p-1))$). Simultaneous †confidence regions of B can be derived from the testing procedures in this paragraph, that is,

$$\operatorname{tr}\mathbf{Q}_e^{-1}(B-\hat{\mathbf{B}})ZZ'(B-\hat{\mathbf{B}})'<c.$$

Moreover, when the matrix B is decomposed as $B=(B_1 \vdots B_2)$, where B_1 and B_2 are a $p\times q$ matrix and a $p\times(m-q)$ matrix, respectively, and the hypothesis to be tested is of the form $B_1=0$, the test procedures can be obtained as follows: Decompose Z as

$$Z=\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix},$$

where Z_1 is a $q\times n$ matrix and Z_2 is an $(m-q)\times n$ matrix, and put

$$\hat{\mathbf{B}}_2^*=\mathbf{X}Z_2'(Z_2Z_2')^{-1},$$

$$\mathbf{Q}^*=\mathbf{X}\mathbf{X}'-\hat{\mathbf{B}}_2^*Z_2\mathbf{X}', \qquad \mathbf{Q}_{B_1}=\mathbf{Q}^*-\mathbf{Q}_e.$$

Then $\mathbf{Q}_{B_1}$ and $\mathbf{Q}_e$ are independent, $\mathbf{Q}_{B_1}$ is distributed according to a Wishart distribution with q degrees of freedom when the hypothesis is true, and we can apply the procedures in the previous paragraph, simply replacing $\mathbf{Q}_B$ by $\mathbf{Q}_{B_1}$.

Such a procedure is called **multivariate analysis of variance** (or **MANOVA**, for short). Various standard situations can be treated in this way (after some linear transformation of variables, if necessary). Some examples are (1) $\mathbf{X}=(\mathbf{X}^{(1)}\ldots\mathbf{X}^{(n)})$, where the $\mathbf{X}^{(i)}$ $(i=1,\ldots,n)$ are distributed independently according to a p-dimensional normal distribution $N(\boldsymbol{\mu},\Sigma)$. We can express $\mathbf{X}$ as $\mathbf{X}=\boldsymbol{\mu}\mathbf{1}'+\mathbf{U}$, and the estimators are given by $\hat{\boldsymbol{\mu}}=\bar{\mathbf{X}}=\mathbf{X}\mathbf{1}/n$, $\hat{\Sigma}=(\mathbf{X}-\bar{\mathbf{X}}\mathbf{1}')(\mathbf{X}-\bar{\mathbf{X}}\mathbf{1}')'/(n-1)$. In this case, we obtain a test for the hypothesis $\boldsymbol{\mu}=\boldsymbol{\mu}_0$ based on **Hotelling's** T^2 statistic, i.e., the test with a †critical region of the form

$$T^2=n(\bar{\mathbf{X}}-\boldsymbol{\mu}_0)'\hat{\Sigma}^{-1}(\bar{\mathbf{X}}-\boldsymbol{\mu}_0)>c.$$

(2) Suppose that $p\times n_i$ matrices $\mathbf{X}_i$, $i=1,\ldots,k$, are samples of size n_i from p-dimensional normal distributions $N(\boldsymbol{\mu}_i,\Sigma)$ with common covariance matrix Σ. The tests for the hypothesis $\boldsymbol{\mu}_1=\ldots=\boldsymbol{\mu}_k$ are obtained from the following observation: Let $\mathbf{Q}_e=\sum_i(\mathbf{X}_i-\bar{\mathbf{X}}_i\mathbf{1}')(\mathbf{X}_i-\bar{\mathbf{X}}_i\mathbf{1}')'$, where $\bar{\mathbf{X}}_i=\mathbf{X}_i\mathbf{1}/n_i$, $\mathbf{Q}_\mu=\sum n_i(\bar{\mathbf{X}}_i-\bar{\mathbf{X}})(\bar{\mathbf{X}}_i-\bar{\mathbf{X}})'$, $\bar{\mathbf{X}}=\sum n_i\bar{\mathbf{X}}_i/\sum n_i$. Then $\sum(\mathbf{X}_i-\bar{\mathbf{X}}\mathbf{1}')(\mathbf{X}_i-\bar{\mathbf{X}}\mathbf{1}')'=\mathbf{Q}_\mu+\mathbf{Q}_e$. We call $\mathbf{Q}_e$ the **matrix of the sum of squares within classes**, and $\mathbf{Q}_\mu$ the **matrix of the sum of squares between classes**. The latter is distributed according to a Wishart distribution when the hypothesis is true.
(3) Suppose that $\mathbf{X}_{ij}$ are p-dimensional vectors, and that

$$\mathbf{X}_{ij}=\boldsymbol{\mu}+\boldsymbol{\alpha}_i+\boldsymbol{\beta}_j+\mathbf{U}_{ij}, \qquad i=1,\ldots,m, \quad j=1,\ldots,n,$$

where $\boldsymbol{\mu}$, $\boldsymbol{\alpha}_i$, $\boldsymbol{\beta}_j$ are p-dimensional constant vectors such that $\sum\boldsymbol{\alpha}_i=0$ and $\sum\boldsymbol{\beta}_j=0$, and the $\mathbf{U}_{ij}$ are independently distributed according to a p-dimensional normal distribution $N(\mathbf{0},\Sigma)$. We set

$$\mathbf{Q}_\alpha=n\sum(\bar{\mathbf{X}}_{i.}-\bar{\mathbf{X}})(\bar{\mathbf{X}}_{i.}-\bar{\mathbf{X}})',$$

$$\mathbf{Q}_\beta=m\sum(\bar{\mathbf{X}}_{.j}-\bar{\mathbf{X}})(\bar{\mathbf{X}}_{.j}-\bar{\mathbf{X}})',$$

$$\mathbf{Q}_e=\sum\sum(\mathbf{X}_{ij}-\bar{\mathbf{X}}_{i.}-\bar{\mathbf{X}}_{.j}+\bar{\mathbf{X}})\times(\mathbf{X}_{ij}-\bar{\mathbf{X}}_{i.}-\bar{\mathbf{X}}_{.j}+\bar{\mathbf{X}})',$$

where

$$\bar{\mathbf{X}}_{i.}=\sum_j\mathbf{X}_{ij}/n, \qquad \bar{\mathbf{X}}_{.j}=\sum_i\mathbf{X}_{ij}/m,$$

$$\bar{\mathbf{X}}=\sum\sum\mathbf{X}_{ij}/mn.$$

Then we have $\sum\sum(\mathbf{X}_{ij}-\bar{\mathbf{X}})(\mathbf{X}_{ij}-\bar{\mathbf{X}})'=\mathbf{Q}_\alpha+\mathbf{Q}_\beta+\mathbf{Q}_e$, and $\mathbf{Q}_\alpha$, $\mathbf{Q}_\beta$, $\mathbf{Q}_e$ are distributed independently according to (noncentral) Wishart distributions with degrees of freedom $m-1$, $n-1$, and $(n-1)(m-1)$, respectively. The tests for the hypothesis $\boldsymbol{\alpha}_i=0$ $(i=1,\ldots,m)$ or $\boldsymbol{\beta}_j=0$ $(j=1,\ldots,n)$ are obtained from these matrices.

C. Tests for Covariance Matrices

Let $\mathbf{X}_{ij}(p\times 1)$, $j=1,\ldots,N_i$, be a random sample from p-variate normal distribution $N(\boldsymbol{\mu}_i,\Sigma_i)$ for $i=1,\ldots,k$. For testing the hypothesis H_0: $\Sigma_1=\ldots=\Sigma_k$ with unknown mean vectors $\boldsymbol{\mu}_i$ against all alternatives, the likelihood ratio statistic is given by

$$\frac{N^{N/2}}{\prod N_i^{N_i/2}}\cdot\frac{\prod|\mathbf{S}_i|^{N_i/2}}{|\sum\mathbf{S}_i|^{N/2}}, \qquad N=N_1+\ldots+N_k, \tag{2}$$

where $\mathbf{S}_i=\sum_{j=1}^{N_i}(\mathbf{X}_{ij}-\bar{\mathbf{X}}_i)(\mathbf{X}_{ij}-\bar{\mathbf{X}}_i)'$ for $\bar{\mathbf{X}}_i=\sum_{j=1}^{N_i}\mathbf{X}_{ij}/N_i$. If we replace the sample size N_i by the degrees of freedom $n_i=N_i-1$ and N by $n=N-k$ in (2), the modified likelihood ratio test is unbiased for general p and k [16]. For $k=1$, the hypothesis specifies $H_0:\Sigma_1=\Sigma_0$ (a given positive definite matrix) and the likelihood ratio statistic is given by $|S|^{N/2}\operatorname{etr}(-\Sigma_0^{-1}S/2)$ $(\operatorname{etr}(\alpha)=\exp(\operatorname{tr}\alpha))$. Again replacing N by $n=N-1$ yields an unbiased test. Moreover the power function of this modified likelihood ratio test depends only on the eigenvalues of $\Sigma_1\Sigma_0^{-1}$ and is increasing with respect to the absolute deviation of each

eigenvalue from 1, that is, $|\mathrm{ch}_i \Sigma_1 \Sigma_0^{-1} - 1|$. For $k=2$, it is conjectured that the modified likelihood ratio test has a power function monotonically increasing with respect to $|\mathrm{ch}_i \Sigma_1 \Sigma_2^{-1} - 1|$. This conjecture has not yet been proved; we know only that the power is increasing if the maximum eigenvalue $\mathrm{ch}_1 \Sigma_1 \Sigma_2^{-1}$ increases from 1 or the minimum eigenvalue $\mathrm{ch}_p \Sigma_1 \Sigma_2^{-1}$ decreases from 1.

D. Estimation of Mean Vector and Covariance Matrix

Let $\mathbf{X}(p \times 1)$ have p-variate normal distribution $N(\boldsymbol{\mu}, I)$. For estimating the unknown mean, take the sum of squared errors S as a loss $L(\boldsymbol{\mu}, d) = \|\boldsymbol{\mu} - d\|^2$. Since each component $\mathbf{X}_i$ of $\mathbf{X}$ is distributed independently according to $N(\mu_i, 1)$ for $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)'$, it is natural to suppose that $\mathbf{X}$ is a good estimator. In fact $\mathbf{X}$ is a minimax estimator for $\boldsymbol{\mu}$. However, Stein showed that $\mathbf{X}$ is admissible for $p=1$ and 2 but inadmissible for $p \geqslant 3$. James and Stein showed that the estimator

$$\left(1 - \frac{p-2}{\|\mathbf{X}\|^2}\right)\mathbf{X} \tag{3}$$

dominates $\mathbf{X}$ when $p \geqslant 3$. This estimator is further dominated by Stein's positive part estimator $(1-(p-2)/\|\mathbf{X}\|^2)_+ \mathbf{X}$, where $(a)_+$ means a if a is nonnegative and zero if a is negative. The class of estimators $\mathbf{X} + \nabla \log f(\mathbf{X})$, where $\nabla = (\partial/\partial x_1, \ldots, \partial/\partial x_p)'$ for $\mathbf{X}' = (x_1, \ldots, x_p)$ are all minimax for $\boldsymbol{\mu}$ dominating $\mathbf{X}$, if $f(\mathbf{X}) \geqslant 0$ and $\sqrt{f(\mathbf{X})}$ is superharmonic $(\nabla^2 \sqrt{f(\mathbf{X})} \leqslant 0)$, satisfying $E[\|\nabla \log f(\mathbf{X})\|^2] < \infty$ and $E[\Sigma|\partial^2 f(\mathbf{X})/\partial x_i^2|/f(\mathbf{X})] < \infty$. Putting $f(\mathbf{X}) = \|\mathbf{X}\|^{-(p-2)}$ yields the James-Stein estimator (3). For this problem a class of monotone estimators is essentially complete, where an estimator $d(\mathbf{X})$ is called monotone if $d(\mathbf{X}) \leqslant d(\mathbf{Y})$ whenever $\mathbf{X} \leqslant \mathbf{Y}$ (defined componentwise). Stein's positive-part estimator is not monotone and is still inadmissible [8].

Let $\mathbf{X}_i(p \times 1)$, $i = 1, \ldots, n$, be a random sample taken from the p-variate normal distribution $N(\mathbf{0}, \Sigma)$. Then the maximum likelihood estimator for Σ is given by $\mathbf{S}/n$, where $\mathbf{S} = \Sigma_{i=1}^n \mathbf{X}_i \mathbf{X}_i'$. $\mathbf{S}$ has Wishart distribution $W_p(n, \Sigma)$. To study the estimation problem of Σ, take two loss functions $L_1(\Sigma, d) = \mathrm{tr}\, d\Sigma^{-1} - \log|d\Sigma^{-1}| - p$ and $L_2(\Sigma, d) = (1/2)\mathrm{tr}(d\Sigma^{-1} - I)^2$. Under L_1-loss, the best scalar multiple of $\mathbf{S}$ is given by $\mathbf{S}/n$, and under L_2-loss it is given by $\mathbf{S}/(n+p+1)$. The James-Stein minimax estimator for L_1-loss is given by $d_1(\mathbf{S}) = K\Delta K'$, where K is the lower triangular matrix with positive diagonal element for which $\mathbf{S} = KK'$ and $\Delta = \mathrm{diag}(\Delta_1, \ldots, \Delta_p)$ with $\Delta_i = 1/(n+p+1-2i)$. A minimax estimator for L_2-loss is given by $d_2(\mathbf{S}) = K\Delta K'$ with the same lower triangular matrix K, but Δ is the solution of the linear equation $A\Delta = b$, where A is given by

$$A = \begin{bmatrix} (n+p-1)(n+p+1) & n+p-3 & \cdots & n-p+1 \\ n+p+3 & (n+p-3)(n+p-1) & \cdots & n-p+1 \\ \cdots & \cdots & \cdots & \cdots \\ n-p+1 & n-p+1 & \cdots & (n-p+1)(n-p+3) \end{bmatrix}$$

and $b' = (n+p-1, n+p-3, \ldots, n-p+1)$. The minimax estimators $d_1(\mathbf{S})$ and $d_2(S)$ dominate the best scalar multiples of $\mathbf{S}$. However, they are inadmissible. Let P be a permutation matrix, then the estimator $\Sigma_p P' d_i(P\mathbf{S}P')P/p!$ dominates $d_i(\mathbf{S})$ because of the convexity of the loss function L_i for $i=1$ and 2. The estimator $h_1(\mathbf{S}) = (\mathbf{S} + ub(u)C)/n$, where $u = 1/\mathrm{tr}\, C\mathbf{S}^{-1}$ and C is a fixed positive-definite matrix, dominates $\mathbf{S}/n$ under L_1-loss if $0 \leqslant b(u) \leqslant 2(p-1)/n$ and $b(\cdot)$ is nonincreasing. Under L_2-loss $(\mathbf{S} + bC/\mathrm{tr}\, C\mathbf{S}^{-1})/(n+p+1)$ dominates $\mathbf{S}/(n+p+1)$, if $0 \leqslant b \leqslant 2(p-1)/(n-p+3)$ [7]. The Haff estimator $h_1(\mathbf{S})$ is dominated by the James-Stein estimator $d_1(\mathbf{S})$ if $p \geqslant 6$. The estimators $(\mathbf{S} + b(\mathrm{tr}\, C\mathbf{S}^{-1}/\mathrm{tr}\, C\mathbf{S}^{-2})C)/n$ for $0 \leqslant b \leqslant 2(p-1)/n$ under L_1-loss also dominate $\mathbf{S}/n$ and are not dominated by $d_1(\mathbf{S})$. Similar results hold under L_2-loss.

E. Correlation among Variables

In order to represent the interrelationships among the p variates, the population correlation matrix $\mathrm{P} = \{\rho_{ij}\}$ is used. Also, we can calculate the population †multiple correlation coefficient of the ith component $\mathbf{X}_i$ and all the rest of the variables by

$$\rho_{i \cdot 1 \ldots (i) \ldots p} = \sqrt{1 - \mathrm{P}/\mathrm{P}_{ii}},$$

and the †partial correlation coefficient of $\mathbf{X}_i$ and $\mathbf{X}_j$, given all others, by

$$\rho_{ij \cdot 1 \ldots (i) \ldots (j) \ldots p} = -\mathrm{P}_{ij}/\sqrt{\mathrm{P}_{ii}\mathrm{P}_{jj}},$$

where P_{ij} are the cofactors of P (→ 397 Statistical Data Analysis).

The sample correlation matrix is calculated from the data, and the **sample multiple correlation** and the **sample partial correlation coefficients** are calculated from the sample correlation matrix in exactly the same way as the population coefficients are calculated from the population correlation matrix. When $\mathbf{X} = (\mathbf{X}^{(1)} \ldots \mathbf{X}^{(n)})$ is a sample of size n from a multivariate normal population, the sampling distributions of $R_{i \cdot 1 \ldots (i) \ldots p}$ and $R_{ij \cdot 1 \ldots (i) \ldots (j) \ldots p}$ are known (→ 374 Sampling Distributions).

The determinant of the covariance matrix $|\Sigma|$ (or $|\mathbf{S}|$), called the **(sample) generalized variance**, is a measure of the dispersion of a p-dimensional distribution. The distance of two distributions with mean vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, respectively, and with common variance Σ is often expressed by

$$\delta=(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\Sigma^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2),$$

which is called the **Mahalanobis generalized distance**.

When the data consists of $(p+q)$-dimensional vectors $\begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix}$ with $q<p$, the interrelation of $\mathbf{X}$ and $\mathbf{Y}$ as a whole can be expressed in the following way: Let the covariance matrix be partitioned as

$$\Sigma=\begin{pmatrix}\Sigma_{\mathrm{XX}} & \Sigma_{\mathrm{YX}}\\ \Sigma_{\mathrm{XY}} & \Sigma_{\mathrm{YY}}\end{pmatrix}$$

and the nonzero roots of the equation $|\rho\Sigma_{\mathrm{YY}}-\Sigma_{\mathrm{YX}}\Sigma_{\mathrm{XX}}^{-1}\Sigma_{\mathrm{XY}}|=0$ be $\rho_1,\ldots,\rho_q$. Then $\rho_1^{1/2},\ldots,\rho_q^{1/2}$, called the **canonical correlation coefficients**, are the maximal invariant parameters with respect to linear transformation of $\mathbf{X}$ and $\mathbf{Y}$. Also, if we denote the eigenvector corresponding to a root ρ_i by $\boldsymbol{\eta}_i$, i.e.,

$$(\rho_i\Sigma_{\mathrm{YY}})\boldsymbol{\eta}_i=(\Sigma_{\mathrm{YX}}\Sigma_{\mathrm{XX}}^{-1}\Sigma_{\mathrm{XY}})\boldsymbol{\eta}_i,$$

the linear function $\boldsymbol{\eta}_i'\mathbf{Y}$ and $\boldsymbol{\eta}_i'\Sigma_{\mathrm{YX}}\Sigma_{\mathrm{XX}}^{-1}\mathbf{X}$ are called the **canonical variates**.

F. Principal Components

An important problem in multivariate analysis is to express the variations of many variables by a small number of indices. **Principal component analysis** is a technique of dealing with this problem.

Let $\mathbf{X}$ be a $p\times n$ matrix with n column vectors of p-dimensional data. A linear transform of $\mathbf{X}$, $\mathbf{T}=A\mathbf{X}$ (A an $r\times p$ matrix, $r<p$, $\mathbf{T}$ an $r\times n$ matrix) is called the **principal component** if A is chosen so as to maximize the sum of the squares of the sample multiple correlation coefficients of each of the row vectors of $\mathbf{X}$ to those of $\mathbf{T}$, namely, if A is an $r\times p$ matrix formed by the r eigenvectors of the sample correlation matrix of $\mathbf{X}$ corresponding to the r largest eigenvalues or any nonsingular linear transformation of them. This is a characterization of principal components in terms of correlation optimality.

The principal component $\mathbf{T}=\mathbf{AX}$ is also characterized by the information-loss optimality in that all eigenvalues of $(\mathbf{X}-C\mathbf{Y}-b\mathbf{1}_n')(\mathbf{X}-C\mathbf{Y}b\mathbf{1}_n')'$ are simultaneously minimized subject to the condition: C is a $p\times r$ matrix, b is a $p\times 1$ vector and $\mathbf{Y}$ is an $r\times n$ matrix. The solution is given by $C\mathbf{Y}+b\mathbf{1}_n'=A'\mathbf{T}+\bar{x}\mathbf{1}_n'$, where $\bar{x}=\mathbf{X}\mathbf{1}_n/n$. The variation optimality of the principal component is given by maximizing simultaneously all the eigenvalues of the matrix $C'(\mathbf{X}-\bar{\mathbf{X}}\mathbf{1}_n')\cdot(\mathbf{X}-\bar{\mathbf{X}}\mathbf{1}_n')'C$, subject to the condition that C is a $p\times r$ matrix such that $C'C=I_r$. The solution is given by $C=A'$, namely, $C'\mathbf{X}=\mathbf{T}$ [15]. If all the correlations between the components of $\mathbf{X}$ are positive, the largest eigenvalue of the covariance matrix of $\mathbf{X}$ is simple and positive. All the coefficients of the first principal component (components of the eigenvector) can be taken to be positive (Perron-Frobenius theorem).

When we assume normality, the eigenvalues of the sample correlation matrix $\mathbf{R}$ are the †maximum likelihood estimators of the eigenvalues of the population correlation matrix, and their sampling distributions can be obtained. A hypothesis relevant to principal component analysis is, for example, that the smallest $p-r$ eigenvalues of the correlation matrix are equal, which can be tested by the statistic

$$R_{p-r}=|\mathbf{R}|/(\lambda_1\ldots\lambda_r((p-\lambda_1-\ldots-\lambda_r)/(p-r))^{p-r}),$$

where $\lambda_1,\ldots,\lambda_r$ are the r largest eigenvalues of $\mathbf{R}$. Under the hypothesis, $-c\log R_{p-r}$ (c a constant) is asymptotically distributed according to a chi-square distribution when $n\to\infty$.

Variations of principal component analysis can be obtained by taking the eigenvectors of the covariance matrix of the raw data or of a multiple of it by some weight matrix.

G. Factor Analysis

Factor analysis is closely related to principal component analysis. We assume a model

$$\mathbf{X}=BF+\mathbf{U},$$

where B and F are unknown $p\times r$ and $r\times n$ matrices of constants ($p>r$) and $\mathbf{U}$ is a $p\times n$ matrix of independent errors. F is called the matrix of **factor scores** and B the matrix of **factor loadings**. We assume $FF'=nI$. If $E(\mathbf{UU}')=n\Phi$ is known, then by applying the least squares principle, we can determine B and F so as to minimize the trace of $(\mathbf{X}-BF)'\Phi^{-1}(\mathbf{X}-BF)$. Then B is obtained by taking the r eigenvectors of $\Phi^{-1}\mathbf{XX}'$ corresponding to the r largest eigenvalues. When Φ is diagonal but unknown, we can solve the simultaneous equation for B and Φ, whose solutions are the matrix $\hat{\mathbf{B}}$ with columns equal to eigenvectors of $\hat{\boldsymbol{\Phi}}^{-1}\mathbf{XX}'$ and the diagonal matrix $\hat{\boldsymbol{\Phi}}$ with elements equal to the diagonal elements of $\mathbf{XX}'/n-\hat{\mathbf{B}}\hat{\mathbf{B}}'$.

If we assume that $\mathbf{U}$ is normal, the procedure in the preceding paragraph for the case

when Φ is known is equivalent to the maximum likelihood method. When Φ is unknown, we can assume further that the columns in F are also multivariate normal vectors distributed independently of $\mathbf{U}$, which implies that the columns of $\mathbf{X}$ are also normal vectors with the covariance matrix $\Sigma = BB' + \Phi$. B and Φ are estimated from the sample covariance matrix, and the solutions of the simultaneous equation for B and Φ gives the maximum likelihood estimators. However, there is the so-called **identification** problem of determining whether for given Σ and r the decomposition $\Sigma = BB' + \Phi$ is unique. This problem has not yet been completely settled. If the solution $\hat{\mathbf{\Phi}}$ is positive definite and $\hat{\mathbf{B}}\hat{\mathbf{B}}'$ is positive semidefinite, it is called proper. When Φ is not positive definite, it is called a Haywood case, and when $\hat{\mathbf{B}}\hat{\mathbf{B}}'$ is not positive semidefinite it is called a complex case. Sometimes iterative procedures lead to improper solutions.

H. Canonical Correlation Analysis

Canonical correlation analysis can also be used for descriptive purposes. Sample canonical variates have various descriptive implications. Suppose that $\eta_1'\mathbf{Y}$ and $\xi_1'\mathbf{X}$ are the first canonical variates corresponding to the largest canonical correlation $\rho_1^{1/2}$. Then $\rho_1^{1/2}$ is the largest possible correlation between a linear function of $\mathbf{X}$ and a linear function of $\mathbf{Y}$ and is actually equal to the correlation of $\eta_1'\mathbf{Y}$ and $\xi_1'\mathbf{X}$. Similarly, the second canonical correlation is equal to the largest possible correlation between linear functions in $\mathbf{X}$ and in $\mathbf{Y}$ which are orthogonal to $\xi_1'\mathbf{X}$ and $\eta_1'\mathbf{Y}$, respectively, and so forth.

As a second interpretation of the canonical variates, we consider the linear regression model

$$\mathbf{Y} = B\mathbf{X} + \mathbf{U},$$

where $\mathbf{Y}$, B, $\mathbf{U}$ are $q \times n$, $q \times p$, $q \times n$ matrices, respectively, and rank $B = r < q$. Then there exist an $r \times p$ matrix A and a $q \times r$ matrix C satisfying $C'\Sigma^{-1}C = I$ such that $B = CA$, where $E(\mathbf{U}\mathbf{U}') = n\Sigma$. Putting $\mathbf{T} = A\mathbf{X}$, we get a regression model $\mathbf{Y} = C\mathbf{T} + \mathbf{U}$, regarding $\mathbf{T}$ as a matrix of regressor variables. If Σ is known, least squares considerations lead to minimizing $\operatorname{tr}(\mathbf{Y} - B\mathbf{X})'\Sigma^{-1}(\mathbf{Y} - B\mathbf{X})$ with the condition rank $B = r$. The resulting row vectors of A consist of the r eigenvectors corresponding to the r largest eigenvalues of the matrix $\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}\Sigma^{-1}\mathbf{S}_{yx}$ and column vectors of C consist of the r eigenvectors corresponding to the r largest eigenvalues of the matrix $\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}\Sigma^{-1}$. If Σ is replaced by $\mathbf{S}_{yy}$, $\mathbf{T} = A\mathbf{X}$ and $\mathbf{Z} = C'\mathbf{S}_{yy}^{-1}\mathbf{Y}$ are equal to the matrices of canonical variates. If we assume that $\mathbf{U}$ is normal, this procedure is equivalent to the maximum likelihood method. It should be remarked that although the model here is not symmetric in $\mathbf{X}$ and $\mathbf{Y}$, the results are symmetric in $\mathbf{X}$ and $\mathbf{Y}$, and therefore they will be the same if the roles of $\mathbf{X}$ and $\mathbf{Y}$ are interchanged in this model.

I. Linear Discriminants and Problems of Classification

Let $p \times n_i$ matrices $\mathbf{X}_i$ $(i = 1, \ldots, k)$ be the set of observations for k distinct populations with a common covariance matrix. We determine a vector $\mathbf{a}$ such that $\mathbf{T}_i = \mathbf{a}'\mathbf{X}_i$ reveals the differences of the k populations as much as possible, or, more precisely, so that the ratio of the sum of squares between classes of $\mathbf{T}$ to the sum of squares within classes is maximized. If the matrices of the sums of squares between and within classes are $\mathbf{Q}_b$ and $\mathbf{Q}_w$, respectively, the ratio is equal to

$$l = \mathbf{a}'\mathbf{Q}_b\mathbf{a}/\mathbf{a}'\mathbf{Q}_w\mathbf{a},$$

which is maximized when $\mathbf{a}$ is equal to the eigenvector of $\mathbf{Q}_w^{-1}\mathbf{Q}_b$ corresponding to the largest eigenvalue. The linear function $t = \mathbf{a}'\mathbf{X}$ is called the **linear discriminant function**. When $k = 2$, $\mathbf{a}$ is given by $\mathbf{a} = \mathbf{Q}_w^{-1}(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)$, where $\bar{\mathbf{X}}_1$ and $\bar{\mathbf{X}}_2$ are sample mean vectors. When $k > 2$, we let A be the matrix formed by the r eigenvectors corresponding to the first r largest eigenvalues of $\mathbf{Q}_w^{-1}\mathbf{Q}_b$, and set $\mathbf{T}_i = A\mathbf{X}_i$. From this we can construct the r-dimensional discriminant function. These functions can be used to locate the k populations in r-dimensional space, and also to decide to which population a new observation belongs. For the latter problem we can also construct k quadratic functions $s_i = (\mathbf{X} - \bar{\mathbf{X}}_i)'\mathbf{Q}_w^{-1}(\mathbf{X} - \bar{\mathbf{X}}_i)$, where $\bar{\mathbf{X}}_i$ is the sample mean vector of the ith population, and $\mathbf{X}$ a new observation. Then we can decide whether $\mathbf{X}$ belongs to the population corresponding to the minimum s_i. Such a method is called a **classification procedure**.

J. Discrete Multivariate Analysis

Let $\mathbf{X}_{ijk}$ be an observed frequency on three characteristics, each belonging to ijkth class for $1 \leqslant i \leqslant I$, $1 \leqslant j \leqslant J$ and $1 \leqslant k \leqslant K$. Assume that $\mathbf{X}_{ijk}$ is a sample from multinomial distribution having p_{ijk} as a probability of occurrence for an (i, j, k) cell, where $p_{\ldots} = \sum p_{ijk} = 1$. The multinomial observation $\mathbf{X}_{ijk}$ with probability p_{ijk} is called an $I \times J \times K$ contingency table. If we further assume that

$$\log p_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha_{ij} + \beta_{jk} + \gamma_{ki} + \delta_{ijk}$$

with the restrictions on parameters $\sum_i \alpha_i = \sum_i \alpha_{ij} = \sum_j \alpha_{ij} = 0$, similarly on β's and γ's and $\sum_i \delta_{ijk} = \sum_j \delta_{ijk} = \sum_k \delta_{ijk} = 0$, we say that a saturated log-linear model is given. Here the number of observations and the number of parameters are equal, and no errors can be estimated. The parameter δ_{ijk} is called the three-factor effect, or equivalently, the second-order interaction. Similarly, α_{ij}, β_{jk}, γ_{ki} are called the two-factor effects or the first-order interactions. Finally α_i, β_j, γ_k are called the main effects. A simple model is obtained when all the first- and second-order interactions vanish. This is equivalent to the independence model: $p_{ijk} = p_{i\cdot\cdot} p_{\cdot j \cdot} p_{\cdot\cdot k}$ and the maximum likelihood estimators for μ, α_i, β_j, γ_k are obtained from $\hat{p}_{ijk} = \mathbf{X}_{i\cdot\cdot}\mathbf{X}_{\cdot j\cdot}\mathbf{X}_{\cdot\cdot k}/n^3$, where $n = \mathbf{X}_{\cdot\cdot\cdot}$. A nontrivial model is obtained by putting all the second-order interactions equal to zero. The likelihood equations for this model are given by

$$np_{ij\cdot} = \mathbf{X}_{ij\cdot}, \quad np_{i\cdot k} = \mathbf{X}_{i\cdot k}, \quad np_{\cdot jk} = \mathbf{X}_{\cdot jk}. \tag{4}$$

Bartlett first described a solution of (4) when $I = J = K = 2$. For a $2 \times 2 \times 2$ table, the solution can be expressed by $\hat{p}_{ijk} = (x_{ijk} \pm \theta)/n$ because of the constraints for $\mathbf{X}_{ij\cdot}$, $\mathbf{X}_{i\cdot k}$, and $\mathbf{X}_{\cdot jk}$. Putting $\delta_{111} = \theta$ for $\hat{p}_{ijk}$ yields a cubic equation for θ:

$$(\mathbf{X}_{111} + \theta)(\mathbf{X}_{122} + \theta)(\mathbf{X}_{221} + \theta)(\mathbf{X}_{212} + \theta)$$
$$= (\mathbf{X}_{121} - \theta)(\mathbf{X}_{112} - \theta)(\mathbf{X}_{211} - \theta)(\mathbf{X}_{222} - \theta).$$

For a general $I \times J \times K$ table, the equations (4) have a unique solution within the no-three-factor effect model if there exists $q_{ijk} > 0$ satisfying (4). The unique solution maximizes the likelihood. To solve the likelihood equations (4), standard iterative procedures, such as the Newton-Raphson method, can be applied. However, the following iterative scaling method of Deming and Stephan is more useful:

$$np_{ijk}^{(3m+1)} = p_{ijk}^{(3m)} \frac{\mathbf{X}_{ij\cdot}}{p_{ij\cdot}^{(3m)}},$$

$$np_{ijk}^{(3m+2)} = p_{ijk}^{(3m+1)} \frac{\mathbf{X}_{i\cdot k}}{p_{i\cdot k}^{(3m+1)}},$$

$$np_{ijk}^{(3m+3)} = p_{ijk}^{(3m+2)} \frac{\mathbf{X}_{\cdot jk}}{p_{\cdot jk}^{(3m+2)}}. \tag{5}$$

The first iteration adjusts $p_{ijk}^{(3m)}$ by fitting proportionally with respect to k so that $np_{ij\cdot}^{(3m+1)}$ is equal to $\mathbf{X}_{ij\cdot}$, and similarly for the second and third iterations. Starting with any initial values satisfying the first-order interaction model, the iterative scaling method (5) converges to the unique solution of (4) as $m \to \infty$.

K. Other Problems

The sampling distributions associated with the procedures discussed here are usually very complicated, and often only asymptotic properties are known (→ 374 Sampling Distributions). Nonparametric rank analogs of many multivariate techniques for normal distribution are found in [18]. Robustness of the distributions of test statistics or of latent roots is investigated in [3, 10, 14]. Multivariate data analysis is found in [5].

References

[1] T. W. Anderson, An introduction to multivariate statistical analysis, Wiley, second edition, 1984.
[2] Y. M. M. Bishop, S. E. Fienberg, and P. W. Holland, Discrete multivariate analysis, MIT Press, 1976.
[3] Y. Fujikoshi, Asymptotic expansions for the distributions of the sample roots under nonnormality, Biometrika, 67 (1980), 45–51.
[4] N. C. Giri, Multivariate statistical inference, Academic Press, 1977.
[5] R. Gnanadesikan, Methods for statistical data analysis of multivariate observations, Wiley, 1977.
[6] S. J. Haberman, The analysis of frequency data, Univ. of Chicago Press, 1974.
[7] L. R. Haff, Empirical Bayes estimation of the multivariate normal covariance matrix, Ann. Statist., 8 (1980), 586–597.
[8] I. Hashimoto, Estimation for distributions with monotone likelihood ratio: Case of vector-valued parameter, Commun. Statist., A10 (1981), 1901–1913.
[9] W. James and C. Stein, Estimation with quadratic loss, Proc. Fourth Berkeley Symp. Math. Statist. and Prob., 1 (1961), 361–379.
[10] T. Kariya, Robustness of multivariate tests, Ann. Statist., 9 (1981), 1267–1275.
[11] M. G. Kendall, Multivariate analysis, Griffin, 1975.
[12] A. M. Kshirsagar, Multivariate analysis, Dekker, 1972.
[13] D. N. Lawley and A. E. Maxwell, Factor analysis as a statistical method, Butterworth, 1963.
[14] R. J. Muirhead and C. M. Waternaux, Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations, Biometrika, 67 (1980), 31–43.
[15] M. Okamoto, Optimality of principal components, Multivariate Analysis II, P. R. Krishnaiah (ed.), Academic Press, 1968, 673–685.
[16] M. D. Perlman, Unbiasedness of the likelihood ratio tests for equality of several covariance matrices and equality of several multivariate normal populations, Ann. Statist., 8 (1980), 247–263.

[17] M. D. Perlman and I. Olkin, Unbiasedness of invariant tests for MANOVA and other multivariate problems, Ann. Statist., 8 (1980), 1326–1341.
[18] M. L. Puri and P. K. Sen, Nonparametric methods in multivariate analysis, Wiley, 1971.
[19] C. R. Rao, Linear statistical inference and its applications, Wiley, 1973.
[20] R. Schwartz, Admissible tests in multivariate analysis of variance, Ann. Math. Statist., 38 (1967), 698–710.
[21] M. S. Srivastava and C. G. Khatri, An introduction to multivariate statistics, North-Holland, 1979.
[22] C. Stein, Estimation of the mean of a multivariate normal distribution, Ann. Statist., 9 (1981), 1135–1151.

N

281 (XIX.5) Network Flow Problems

A. Introduction

The **network flow problem** is a special kind of mathematical-programming problem (→ 255 Linear Programming, 264 Mathematical Programming, 292 Nonlinear Programming), where the variables of the objective function and the constraints are all defined in terms of a graph (→ 186 Graph Theory). Owing to its special structure, the mathematical properties of the network flow problem as well as the solution algorithms have been investigated in detail. Network flow problems have a variety of useful applications in fields such as transportation, scheduling, and resource allocation, and in operations research in general, so that they now constitute an important class of mathematical programming. By the term "network" (*Netzwerk* in German, *réseau* in French, *set'* in Russian) we usually mean a graph with some physical attributes attached to its edges and vertices.

B. Basic Form of the Problem

Let $G=(V,E,\partial^+,\partial^-)$ be a graph with vertex set V, edge set E, and incidence relations ∂^+, $\partial^-:E\to V$, and let $\mathbf{R}$ be the field of real numbers. (Most of the statements in the following are valid if we take for $\mathbf{R}$ an ordered field or ordered additive group more general than the field of real numbers.) Furthermore, we regard the collection Ξ of all functions $\xi:E\to\mathbf{R}$ as a vector space of dimension $|E|$, denoting $\xi(e_\kappa)$ by ξ^κ if $E=\{e_1,\dots,e_n\}$. Similarly, the collection Ω of all functions ω: $V\to\mathbf{R}$ is a vector space of dimension $|V|$. If we define the mapping $\delta^\pm:V\to 2^E$ by $\delta^\pm v=\{e\mid\partial^\pm e=v\}$ (→ 186 Graph Theory), a linear mapping $\partial:\Xi\to\Omega$ is naturally introduced through the relation $(\partial\xi)(v)=\sum_{e\in\delta^+v}\xi(e)-\sum_{e\in\delta^-v}\xi(e)$. A vector $\xi\in\Xi$ is called a **flow** on G if $\partial\xi=0$. The dual space Ξ^* of Ξ and the dual Ω^* of Ω are identified with the collection of functions $\eta:E\to R$ and that of $\zeta:V\to R$, respectively, under the obvious correspondence. Then the linear mapping $\delta:\Omega^*\to\Xi^*$, which is contragredient to ∂, is defined by means of the relation $(\delta\zeta)(e)=\zeta(\partial^+e)-\zeta(\partial^-e)$. A vector $\eta\in\Xi^*$, which is the image of ζ under δ, i.e., which satisfies $\eta=\delta\zeta$, is called a **tension** on G, and ζ is called the **potential** on G corresponding to the tension η. (Sometimes, $\xi(e)$ is called a flow **in** edge e, $\eta(e)$ a tension **across** e, and $\zeta(v)$ a potential **at** vertex v.)

A continuous curve C ($\subset\mathbf{R}\times\mathbf{R}$) on the Euclidean plane $\mathbf{R}\times\mathbf{R}$ is said to be **monotone** if $(x_1-x_2)(y_1-y_2)\geqslant 0$ for any (x_1,y_1), $(x_2,y_2)\in C$. A monotone curve C is called a **characteristic curve** if its projection to each coordinate axis, i.e., $C_x\equiv\{x\mid(x,y)\in C\}$ and $C_y\equiv\{y\mid(x,y)\in C\}$, is a closed interval. For a characteristic curve C, two convex functions, $\varphi(x)=\int_{x_0}^{x}(y-y_0)dx$ and $\psi(y)=\int_{y_0}^{y}(x-x_0)dy$, are defined, where (x_0,y_0) is a point fixed on C and the integrals are taken along $(x,y)\in C$. (It is understood that $\varphi(x)=\infty$ if $x\notin C_x$ and $\psi(y)=\infty$ if $y\notin C_y$.) These two functions defined for a fixed C are conjugate to each other in Fenchel's sense, and they satisfy $\varphi(x)+\psi(y)\geqslant(x-x_0)(y-y_0)$ for any $(x,y)\in\mathbf{R}\times\mathbf{R}$, where the inequality reduced to an equality if and only if $(x,y)\in C$ (→ 88 Convex Analysis, 292 Nonlinear Programming).

A network N in network flow theory is a graph G to each edge e_κ whose edge set $E=\{e_1,\dots,e_n\}$ is given a characteristic curve C^κ. On a network N, the following three problems, PI, PII, and PIII, are defined. PI: Find a flow ξ that minimizes $\Phi(\xi)\equiv\sum_{\kappa=1}^{n}\varphi_\kappa(\xi^\kappa)$ under the constraints $\xi^\kappa\equiv\xi(e_\kappa)\in C_x^\kappa$ $(\kappa=1,\dots,n)$. PII: Find a tension η which minimizes $\Psi(\eta)\equiv\sum_{\kappa=1}^{n}\varphi^\kappa(\eta_\kappa)$ under the constraints $\eta_\kappa\equiv\eta(e_\kappa)\in C_y^\kappa$ $(\kappa=1,\dots,n)$. PIII: Find a pair (ξ,η) of a flow ξ and a tension η such that $(\xi^\kappa,\eta_\kappa)\in C^\kappa$ for all $\kappa=1,\dots,n$. (With respect to PI or PII, a flow or a tension satisfying the "constraints" is ordinarily called a **feasible** flow or tension, respectively.) Then, as a special case of the (Karush-)Kuhn-Tucker theorem and the duality theorem in nonlinear programming, we have the following theorems (A) and (B).

(A) For a given network N, one and only one of the following four alternatives is the case:

(i) There is a feasible flow in PI, and at the same time, there is a feasible tension in PII. In this case, both PI and PII have a solution, and, for any solution $\hat{\xi}$ of PI and any solution $\hat{\eta}$ of PII, the pair $(\hat{\xi},\hat{\eta})$ is a solution of PIII; and, conversely, for any solution $(\hat{\xi},\hat{\eta})$ of PIII, the flow $\hat{\xi}$ is a solution of PI and the tension $\hat{\eta}$ is a solution of PII.

(ii) There is a feasible tension in PII, whereas there is no feasible flow in PI. In this case, $\Phi(\xi)$ of PII is not bounded below for feasible tensions ξ.

(iii) There is a feasible tension in PII, whereas there is no feasible flow in PI. In this case, $\Psi(\eta)$ of PII is not bounded below for feasible tensions η.

(iv) There is neither a feasible flow in PI nor a feasible tension in PII.

(B) Let $C_x^\kappa=[b^\kappa,c^\kappa]$ ($b^\kappa\leqslant c^\kappa$; b^κ can be $-\infty$, and c^κ, ∞) and $C_y^\kappa=[d_\kappa,f_\kappa]$ (d_κ can be $-\infty$, and f_κ, ∞). Then a necessary and sufficient condition for PI to have a feasible flow is that

$\sum_1 c^\kappa - \sum_2 b^\kappa \geqslant 0$ for every cutset (cocycle or cocircuit) ω of G, where the summation $\sum_1$ (resp. $\sum_2$) is taken over all the edges e_κ ($\in E$) that lie in ω in the positive (resp. negative) direction. Similarly, a necessary and sufficient condition for PII to have a feasible tension is that $\sum_1 f_\kappa - \sum_2 d_\kappa \geqslant 0$ for every tieset (cycle or circuit) θ of G, where $\sum_1$ (resp. $\sum_2$) is taken over all the edges e_κ that lie in θ in the positive (resp. negative) direction.

C. Shortest Paths, Maximum Flows, and Minimum-Cost Flows

We choose one of the edges of G, say e_1, as the **reference edge** and assign it the parametric characteristic curve $C^1\langle\alpha\rangle = \{(x,y) \mid y = x + \alpha\}$. Then if there is a feasible flow on the network obtained from G by contracting (i.e., short-circuiting) e_1, and if at the same time there is a feasible tension on the network obtained from G by deleting (i.e., open-circuiting) e_1, then problem PIII has a solution $(\hat{\xi}\langle\alpha\rangle, \hat{\eta}\langle\alpha\rangle)$ for every real α, and $\hat{\xi}^1\langle\alpha\rangle$ and $\hat{\eta}_1\langle\alpha\rangle$ are uniquely determined for each α. The problem of determining these parametric solutions is the **two-terminal problem** for the **two-terminal network** N_1 (which is obtained from G by deleting e_1) with the vertex $\partial^- e_1$ as the **entrance** (or **source**) and the vertex $\partial^+ e_1$ as the **exit** (or **sink**). The curve $C = \{(\hat{\xi}^1\langle\alpha\rangle, \hat{\eta}_1\langle\alpha\rangle) \mid \alpha \in \mathbf{R}\}$ enjoys the properties of a characteristic curve, and is called the **two-terminal characteristic** of N_1 with respect to the entrance $\partial^- e_1$ and exit $\partial^+ e_1$. A two-terminal network for which only the projections to the x-axis $C^\kappa_x = [b^\kappa, c^\kappa]$ of the edge characteristics are specified ($\kappa = 2, \ldots, n$) is called a **capacitated network**, and the **maximum-flow problem** for a capacitated network N_1 can be mathematically formulated as the problem of determining the projection to the x-axis $C_x = [b, c]$ of the two-terminal characteristics of N_1. For the maximum-flow problem, the relation $c = \min\{\sum'_1 c^\kappa - \sum'_2 b^\kappa\}$ holds, where the minimum is taken over all the cutsets that contain e_1 in the negative direction and where $\sum'_1$ (resp. $\sum'_2$) denotes the summation over all the edges, except e_1, lying in a cutset in the positive (resp. negative) direction. This relation is called the **maximum-flow minimum-cut theorem**. (A similar relation holds also for b.)

Similarly, or dually, the problem of determining the y-projection $C_y = [d, f]$ of the two-terminal characteristic of a two-terminal network N_1 for which the y-projections $C^\kappa_y = [d_\kappa, f_\kappa]$ of the characteristics are specified to the nonreference edges e_κ ($\kappa = 2, \ldots, n$) is a network flow formulation of the **shortest-path problem**. For the shortest-path problem, the relation $f = \min\{\sum'_1 f_k - \sum'_2 d_\kappa\}$, called the **maximum-separation minimum-distance theorem**, holds, where the minimum is taken over all the tiesets that contain e_1 in the positive direction and where $\sum'_1$ (resp. $\sum'_2$) denotes the summation over all the edges, except e_1, lying in a tieset in the positive (resp. negative) direction.

The **minimum-cost flow problem** is to determine the two-terminal characteristic of N_1 when all the C^κ ($\kappa \neq 1$) are of staircase form.

A number of algorithms exist for time complexity ($\rightarrow$ 71 Complexity of Computations) $O(|V|^3)$ for the shortest-path problem [4]; the algorithm proposed by E. W. Dijkstra [7] for the case $d_\kappa \leqslant 0 \leqslant f_\kappa$ for all $\kappa (\neq 1)$ is of complexity $O(|V|^2)$. The algorithm of A. V. Karzanov [8, 9] for the maximum-flow problem is of complexity $O(|V|^3)$. The minimum-cost flow problem can be solved by alternately solving the subproblems of the shortest-path type and those of the maximum-flow type, but no algorithm of time complexity polynomial in $|E|$ and $|V|$ has been found.

D. Transportation and Scheduling

Let us make the edges of a graph $G = (V, E, \partial^+, \partial^-)$ correspond to the transportation routes and the flow to the stream of a commodity; impose a **capacity constraint** of the form $0 \leqslant \xi^\kappa \leqslant c^\kappa$ on the flow ξ^κ in each edge e_κ, and assume a **cost** function of the form $\varphi_\kappa(\xi^\kappa) = f_\kappa \xi^\kappa$ for each edge. (c^κ is a constant called the **capacity** of edge e_κ, and f_κ is a constant called the **unit cost** of edge e_κ.) Furthermore, let us specify a subset V_1 ($\subset V$) of vertices as the set of entrances and another subset V_2 ($\subset V$) as the set of exits, where $V_1 \cap V_2 = \varnothing$, and prescribe the amount of inflow $q(v)$ to each entrance $v \in V_1$ and the amount of outflow $q(v)$ from each exit $v \in V_2$, where we must have $\sum_{v \in V_1} q(v) = \sum_{v \in V_2} q(v)$. Planning a transportation plan that satisfies all the above-prescribed conditions and that minimizes the total cost $\sum_{\kappa \in E} \varphi_\kappa(\xi^\kappa)$ can be reduced to finding a minimum-cost flow on the extended network $\tilde{G} = (\tilde{V}, \tilde{E}, \tilde{\partial}^+, \tilde{\partial}^-)$, defined as follows, such that the flow in the reference edge is to be maximum: $\tilde{V} = V \cup \{s\} \cup \{t\}$ ($s, t \notin V$), $\tilde{E} = E \cup E_1 \cup E_2 \cup \{e_0\}$ ($E_1 \cap E = E_2 \cap E = \varnothing$, $e_0 \notin E \cup E_1 \cup E_2$; e_0 is the reference edge, $\tilde{\partial}^+ e_0 = t$, $\tilde{\partial}^- e_0 = s$; $E_1 = \{e \mid \tilde{\partial}^+ e = s, \tilde{\partial}^- e = v, v \in V_1\}$, $C(e) = \{(0,y) \mid y \leqslant 0\} \cup \{(x,0) \mid 0 \leqslant x \leqslant q(\tilde{\partial}^- e)\} \cup \{(q(v), y) \mid 0 \leqslant y\}$, where $\tilde{\partial}^- e = v \in V_1$; $E_2 = \{e \mid \tilde{\partial}^+ e = v, \tilde{\partial}^- e = t, v \in V_2\}$, $C(e) = \{(0,y) \mid y \leqslant 0\} \cup \{(x,0) \mid 0 \leqslant x \leqslant q(v)\} \cup \{(q(v), y) \mid 0 \leqslant y\}$, where $\tilde{\partial}^+ e = v \in V_2$), $\tilde{\partial}^\pm|_E = \partial^\pm$, $C(e_\kappa) = \{(0,y) \mid y \leqslant f_\kappa\} \cup \{(x, f_\kappa) \mid 0 \leqslant x \leqslant c^\kappa\} \cup \{(c^\kappa, y) \mid y \geqslant f_\kappa\}$ for $e_\kappa \in E$.

For the project-scheduling problem with an

acyclic graph $G=(V,E,\partial^+,\partial^-)$ as the **arrow diagram** for which the **start node** $s\in V$ and the **completion node** are specified, we make an edge $e(\in E)$ correspond to a **job** (or **activity**), and a vertex correspond to the **event** that all the jobs corresponding to $\delta^- v$ have been finished, whereas those corresponding to $\delta^+ v$ are ready to commence. Then we interpret the negative of the tension of an edge as the time (i.e., duration) spent on the corresponding job and the potential at a vertex as the time (i.e., instant) of the corresponding event taking place. Furthermore, we assume that to each edge e_κ (or the corresponding job) are given the **normal duration** $-d_\kappa(>0)$, the **crash duration** $-f_\kappa(>0, <-d_\kappa)$, and the unit cost $c^\kappa(>0)$ we have to pay in order to decrease the job time by one unit, where the job time necessarily lies between the normal and the crash duration. Under the above-listed specifications, we consider the relation between the total project time (i.e., the duration from the event corresponding to the start node to that corresponding to the completion node) and the extra cost to be paid for decreasing the project time below the normal one. This problem can be reduced to the two-terminal characteristic problem for the following network $\tilde{G}$: $\tilde{G}$ is obtained from G by adding the reference edge e_0 $(\tilde{\partial}^+ e_0=t, \tilde{\partial}^- e_0=s)$; $C^\kappa=\{(0,y)\mid y\leqslant d_\kappa\}\cup\{(x,d_\kappa)\mid 0\leqslant x\leqslant c^\kappa\}\cup\{(c^\kappa,y)\mid d_\kappa\leqslant y\leqslant f_\kappa\}\cup\{(x,f_\kappa)\mid c^\kappa\leqslant x\}$.

E. Applications to Combinatorial Optimization

Many problems in combinatorial optimization can be reduced to network flow problems. The problem of finding a maximum matching on a bipartite graph G (→ 186 Graph Theory) is reduced to the maximum-flow problem for the graph representing the transportation problem on G with one of the two vertex sets as the entrance set and the other as the exit set, where all the edges have a unit capacity and the amount of inflow/outflow to/from each entrance/exit vertex is equal to unity (the cost being irrelevant). (The existence of an integer solution to this kind of maximum-flow problem is proved constructively on the basis of the solution algorithm.) The maximum-flow minimum-cut theorem in this particular case can be stated as follows: The maximum cardinality of matchings on a bipartite graph is equal to the minimum cardinality of vertex subsets which cover all the edges (this is known as the **König-Egerváry theorem**). The Dilworth theorem for a partially ordered set, the criterion for the existence of a graph with prescribed vertex degrees, etc., are known to be reducible to network flow problems [2].

F. Generalizations

The network flow problem may be generalized in various directions. Replacing a graph by a matroid (→ 66 Combinatorics) or considering stronger conditions on the feasibility of flows and tensions are natural extensions [4, 6, 10, 11]. Another extension is to consider several kinds of flow, instead of a single kind, that simultaneously affect the capacities of edges. This latter problem is called the **multicommodity flow problem** in contrast to the **single-commodity flow problem** that was treated above [5]. It can be said that any extension of the network flow problem aims at a mathematical model that has wider application without losing the advantage of having simple effective solution algorithms.

References

[1] G. J. Minty, Monotone networks, Proc. Roy. Soc. London, A257 (1960), 194–212.
[2] L. R. Ford, Jr., and D. R. Fulkerson, Flows in networks, Princeton Univ. Press, 1962.
[3] C. Berge and A. Ghouila-Houri, Programmes, jeux et réseaux de transport, Dunod, 1962.
[4] M. Iri, Network flow, transportation and scheduling—Theory and algorithms, Academic Press, 1969.
[5] T. C. Hu, Integer programming and network flows, Addison-Wesley, 1969.
[6] E. L. Lawler, Combinatorial optimization—Networks and matroids, Holt, Rinehart and Winston, 1976.
[7] E. W. Dijkstra, A note on two problems in connexion with graphs, Numerische Math., 1 (1959), 269–271.
[8] E. A. Dinits, Algorithm solutions to maximum network flow problems by power analysis (in Russian), Dokl. Akad. Nauk SSSR, 194 (1970), 754–757.
[9] A. V. Karzanov, Calculation of maximum network flow by power methods (in Russian), Dokl. Akad. Nauk SSSR, 215 (1974), 49–52.
[10] M. Iri and S. Fujishige, Use of matroids in operations research, circuits and systems theory, Intern. J. Syst. Sci., 12 (1981), 27–54.
[11] C. P. Bruter, Eléments de la théorie des matroïdes, Lecture notes in math. 387, Springer, 1974.

282 (XX.17) Networks

A. Linear Graphs (→ 186 Graph Theory)

A **linear graph** (or simply **graph**) is an object composed of (i) a finite set $\{B_\kappa\}$ $(\kappa=1,\dots,n)$ of elements called **branches** (or **edges**), (ii) a finite set $\{N_a\}$ $(a=1,\dots,m)$ of elements called **nodes** (or **vertices**), and (iii) an **incidence relation** between branches and nodes represented by a function $[B_\kappa:N_a]$ from $\{B_\kappa\}\times\{N_a\}$ to $\{0,1,-1\}$ such that for every κ, there exist exactly one N_a with $[B_\kappa:N_a]=1$ and exactly one N_a with $[B_\kappa:N_a]=-1$, with all the other $[B_\kappa:N_a]$ equal to 0. (Intuitively, a branch B_κ starts from the node N_a with $[B_\kappa:N_a]=1$ and ends at the node N_b with $[B_\kappa:N_b]=-1$.) In terms of topology, a (linear) graph is a 1-dimensional finite †simplicial complex (→ 70 Complexes, 186 Graph Theory). A **network** in the wide sense is a linear graph whose branches and nodes are endowed with some physical properties.

B. Networks

A **contact network**, one of the simplest kinds of networks, is an abstraction of a circuit whose branches correspond to contact points of relays and switches that are allowed to take a finite number of physical states, e.g., the two states "on" and "off." The theory of contact networks is developed by means of †Boolean algebra and is applied to switching networks, such as telephone exchange networks and the logical networks of digital computers.

In most cases, to the branches B_κ of a network two kinds of real quantities i_κ and E_κ are assigned (which may be variables or functions of time) satisfying the conditions:

(i) $\sum_{\kappa=1}^{n}[B_\kappa:N_a]i_\kappa=0 \quad (a=1,\dots,m);$

(ii) there exist E_a such that

$$\sum_{a=1}^{m}[B_\kappa:N_a]E_a=E_\kappa \quad (\kappa=1,\dots,n).$$

In the case of **electric networks**, where i_κ, E_κ, and E_a are the current in branch B_κ, the voltage across branch B_κ, and the potential at node N_a, respectively, these conditions are known as **Kirchhoff's laws**.

The **network flow problem**, which has a number of practically important applications in †operations research (e.g., transportation problems, project-scheduling problems) and is a special case of †mathematical programming, can be formulated as the problem of minimizing $\sum_{\kappa=1}^{m}f_\kappa(i_\kappa)$ under condition (i) (or minimizing $\sum_{\kappa=1}^{n}f_\kappa(E_\kappa)$ under condition (ii)), where for each branch B_κ, f_κ is a given †convex function defined on a given interval $[a_\kappa,b_\kappa]$.

C. Electric Networks

Since there has been a great deal of research on electric networks, "network" often means "electric network." We call a branch in which the current is a given function of time a current source, and one across which the voltage is a given function of time a voltage source. A branch that is either a current source or a voltage source is called a **source branch**. A network with M source branches is called an **M-port network**. If the currents i_κ in and the voltages E_κ across the non-source branches (n' in number) are related by

$$E_\kappa=\sum_{\lambda=1}^{n'}z_{\kappa\lambda}i_\lambda \quad (\kappa=1,\dots,n')$$

or

$$i_\kappa=\sum_{\lambda=1}^{n'}y_{\kappa\lambda}E_\lambda \quad (\kappa=1,\dots,n'),$$

where the $z_{\kappa\lambda}$ or $y_{\kappa\lambda}$ are linear †integrodifferential operators, the network is said to be **linear**. If $z_{\lambda\kappa}=z_{\kappa\lambda}$ or $y_{\kappa\lambda}=y_{\lambda\kappa}$, it is said to be **reciprocal** or **bilateral**; if the $z_{\kappa\lambda}$ or $y_{\kappa\lambda}$ are invariant under the change of the origin of time, it is said to be **time-invariant**; and if a linear time-invariant network satisfies the condition

$$\int_{-\infty}^{t}\sum_{B_\kappa\in S}i_\kappa(\tau)E_\kappa(\tau)\,d\tau\leqslant 0$$

for every t and for every choice of functions of time for i_κ or E_κ associated with the source branches B_κ in S, provided that the current-voltage relations for non-source branches are satisfied, then the network is said to be **passive**. Under certain nonsingularity conditions, for the currents in and the voltages across the source branches (denoted by I_κ and e_κ) of a linear M-port network, we have the relations

$$e_\kappa=\sum_{B_\lambda\in S}Z_{\kappa\lambda}I_\lambda, \qquad I_\kappa=\sum_{B_\lambda\in S}Y_{\kappa\lambda}e_\lambda \qquad (B_\kappa\in S),$$

where the matrices $Z_{\kappa\lambda}$ and $Y_{\kappa\lambda}$ of linear integrodifferential operators are called the **port-impedance matrix** and the **port-admittance matrix** of the network, respectively. **Analysis** determines $Z_{\kappa\lambda}$, $Y_{\kappa\lambda}$ from a given linear graph and given $z_{\kappa\lambda}$, $y_{\kappa\lambda}$, while **synthesis** finds a network (i.e., $z_{\kappa\lambda}$ or $y_{\kappa\lambda}$, as well as a linear graph) when part of $Z_{\kappa\lambda}$, $Y_{\kappa\lambda}$, or some relations to be satisfied by them are given. (In synthesis, the $z_{\kappa\lambda}$ or $y_{\kappa\lambda}$ are usually confined to some special class.) In analysis as well as synthesis we usu-

ally deal with the †Laplace transforms $\bar{z}_{\kappa\lambda}(s)$, $\bar{y}_{\kappa\lambda}(s)$, $\bar{Z}_{\kappa\lambda}(s)$, $\bar{Y}_{\kappa\lambda}(s)$ instead of $z_{\kappa\lambda}$, $y_{\kappa\lambda}$, $Z_{\kappa\lambda}$, $Y_{\kappa\lambda}$ themselves, where the characteristics $\bar{Z}_{\kappa\lambda}(s)$, $\bar{Y}_{\kappa\lambda}(s)$ of a network are determined by the topological properties of its linear graph and the properties of $\bar{z}_{\kappa\lambda}(s)$, $\bar{y}_{\kappa\lambda}(s)$ as functions of the complex variable s. (If ω is the angular frequency, $s = i\omega$.)

The following fundamental facts are known in regard to analysis and synthesis:

(1) If $\bar{Z}_{\kappa\lambda}(s)$, $\bar{Y}_{\kappa\lambda}(s)$ are rational functions of s holomorphic on the open right half-plane, a necessary and sufficient condition for a network to be passive is that for arbitrary real numbers ξ_κ,

$$\sum_{B_\kappa, B_\lambda \in S} \xi_\kappa \xi_\lambda \bar{Z}_{\kappa\lambda}(s) \quad \text{or} \quad \sum_{B_\kappa, B_\lambda \in S} \xi_\kappa \xi_\lambda \bar{Y}_{\kappa\lambda}(s)$$

is a **positive real function** of s, i.e., a function whose value is real when s is real and lies on the right half-plane when s lies on the right half-plane.

(2) Every passive one-port network can be synthesized by using a finite number of three kinds of branches, i.e., positive resistors ($E_\kappa = R_\kappa i_\kappa$, $R_\kappa > 0$), positive capacitors ($i_\kappa = C_\kappa dE_\kappa/dt$, $C_\kappa > 0$), and positive inductors ($E_\kappa = L_\kappa di_\kappa/dt$, $L_\kappa > 0$).

(3) Every passive M-port network ($M \geqslant 2$) can be synthesized by using, in addition to the three kinds of branches mentioned in (2), ideal transformers (an ideal transformer is a pair of branches (B_κ, B_λ) such that $i_\kappa = n i_\lambda$, $E_\lambda = n E_\kappa$, and n = real number) and ideal gyrators (an ideal gyrator is a pair of branches (B_κ, B_λ) such that $i_\kappa = E_\lambda$, $i_\lambda = -E_\kappa$); ideal gyrators are not needed, however, to synthesize a reciprocal network.

However, very little is known about the synthesis of passive M-port networks without using ideal transformers and gyrators. Topological methods are expected to be powerful for such synthesis problems.

The linear graph structure of a network loses its significance if ideal transformers are admitted.

References

[1] K. Kondo (ed.), RAAG memoirs of the unifying study of basic problems in engineering and the physical sciences by means of geometry, Association for Science Documents Information, Tokyo, I, 1955; II, 1958; III, 1962.
[2] W. Cauer, Synthesis of linear communication networks, McGraw-Hill, second edition, 1958. (Original in German, 1941.)
[3] J. E. Storer, Passive network synthesis, McGraw-Hill, 1957.
[4] L. Weinberg, Network analysis and synthesis, McGraw-Hill, 1962.
[5] O. Brune, Synthesis of a finite two-terminal network whose driving-point impedance is a prescribed function of frequency, J. Math. Phys., 10 (1931), 191–236.
[6] Y. Oono and K. Yasuura, Synthesis of finite passive $2n$-terminal networks with prescribed scattering matrices, Mem. Fac. Eng. Kyushu Univ., 14 (1954), 125–177.
[7] S. Seshu and M. B. Reed, Linear graphs and electrical networks, Addison-Wesley, 1961.
[8] W. H. Kim and R. T. Chien, Topological analysis and synthesis of communication networks, Columbia Univ. Press, 1962.
[9] C. Berge and A. Ghouila-Houri, Programming, games and transportation networks, Methuen, 1965. (Original in French, 1962.)
[10] L. R. Ford, Jr., and D. R. Fulkerson, Flows in networks, Princeton Univ. Press, 1962.
[11] R. W. Newcomb, Linear multiport synthesis, McGraw-Hill, 1966.

283 (XXI.37) Newton, Isaac

Sir Isaac Newton (December 25, 1642–March 20, 1727), the English mathematician and physicist, was born into a family of farmers in Woolsthorpe, Lincolnshire. In 1661, he entered Cambridge University, where he was greatly influenced by the professor who was teaching geometry, I. Barrow, and where he began research in Kepler's optics and Descartes's geometry.

In 1665, he discovered the †binomial theorem, and in the same year, during a stay at his birthplace to escape the plague, he began work on his three great discoveries—the spectral decomposition of light, the universal law of gravity, and differential and integral calculus. He returned to Cambridge University in 1667, and in the following year invented the reflecting telescope and proposed his theory of light particles. During this period, he succeeded Barrow as professor and lectured on optics. At the same time, he probed deeper into the calculus. Guided by Barrow's insight that differentiation and integration were inverse operations and also by his own research on infinite series, Newton obtained the †fundamental theorem of calculus. Leibniz obtained the same theorem a little later, and a struggle resulted between the two over priority. The two discoveries were independent, but because Leibniz's notation was superior, the later development of calculus owes more to him. The dynamic elucidation of the heliocentric theory was accomplished in Newton's main

work, *Principia mathematica philosophiae naturalis* (1686–1687), in which Kepler's law on the movement of the planets, Galileo's theory of movement, and Huygens's theory of oscillation were unified into the three laws of Newtonian dynamics. These natural laws, which deal with all dynamic phenomena in the universe, are the most superlative realization of Descartes's concept of exploring the mathematical structure of nature; they had an essential influence on the later development of the natural sciences. The style of writing is similar to that in Euclid's *Stoicheia*. In the *Principia*, Newton also sets forth his philosophical position.

In 1695, Newton moved to London and became engrossed in theology. He was appointed Master of the Mint and was president of the Royal Society from 1703 until his death. While he is sometimes said to have divorced himself from science, many of his notes on geometry date from this time.

References

[1] D. T. Whiteside (ed.), Sir Isaac Newton, Mathematical works, Johnson Reprint, I, 1964; II, 1969.
[2] D. T. Whiteside (ed.), Sir Isaac Newton, Mathematical papers, Cambridge Univ. Press, I, 1967; II, 1968; III, 1969; IV, 1971; V, 1972; VI, 1974; VII, 1976; VIII, 1981.
[3] I. Newton, Philosophiae naturalis principia mathematica, London, 1687; English translation, Mathematical principles of natural philosophy, trans. by A. Motte in 1729, Univ. of California Press, 1934.
[4] D. Brewster, Memoirs of the life, writings, and discoveries of Sir Isaac Newton, Constable, 1855.

284 (III.12) Noetherian Rings

A. General Remarks

In this article, we mean by ring a commutative ring with unity element. Thus a **Noetherian ring** is a commutative ring with unity element that satisfies the †maximum condition for its ideals; if it is also an †integral domain, then it is called a **Noetherian integral domain** or **Noetherian domain** (for right and left Noetherian rings → 368 Rings F).

A ring is Noetherian if and only if every prime ideal of the ring has a †finite basis (**Cohen's theorem**). A ring is Noetherian if it is generated by a finite number of elements over a Noetherian ring (**Hilbert's basis theorem**). The following three conditions for a ring R are equivalent: (i) R is an **Artinian ring**, that is, it satisfies the †minimum condition for its ideals (for right and left Artinian rings → 368 Rings F). (ii) R is a Noetherian ring and every prime ideal of R is a maximal ideal. (iii) There exist a finite number of Noetherian rings $R_i (i=1, \ldots, n)$ whose †maximal ideals are †nilpotent such that R is the direct sum of the $R_i (i=1, \ldots, n)$. We say that the **restricted minimum condition** holds in a ring R if $R/\mathfrak{a}$ is an Artinian ring for every nonzero ideal $\mathfrak{a}$ of R; the latter condition is satisfied if and only if R is either an Artinian ring or a Noetherian domain of †Krull dimension 1. Every ideal of a Noetherian ring R can be expressed as the intersection of a finite number of †primary ideals. Given a ring R and an R-module M, a submodule P of M is said to be a **primary submodule** of M if every element a of R that is a **zero divisor with respect to** M/P (i.e., there exists an $m \in M/P$ such that $m \neq 0$ and $am=0$) is **nilpotent with respect to** M/P (i.e., there exists a natural number n such that $a^n(M/P)=0$).

The property of ideals in Noetherian rings stated above can be generalized to the case of Noetherian modules: If an R-module M is a †Noetherian module, then every submodule of M can be expressed as the intersection of a finite number of primary submodules.

Let R be a Noetherian ring, $\mathfrak{a}$ an ideal of R, M a finite R-module, and N and N' submodules of M. Then we have (i) the **Artin-Rees lemma**: There exists a natural number r such that for all $n>r$, $\mathfrak{a}^n N \cap N' = \mathfrak{a}^{n-r} \cdot (\mathfrak{a}^r N \cap N')$. (ii) **Krull's intersection theorem**: $\bigcap_{n=1}^{\infty} \mathfrak{a}^n M = \{m \in M \mid \exists a \in \mathfrak{a}$ such that $(1-a)m=0\}$ (hence, in particular, if $\mathfrak{m}$ is the †Jacobson radical or R, then $\bigcap_{n=1}^{\infty} \mathfrak{m}^n M = \{0\}$. (iii) **Krull's altitude theorem**: If $\mathfrak{a}$ is generated by s elements and $\mathfrak{p}$ is a †minimal prime divisor of $\mathfrak{a}$, then the height of $\mathfrak{p} \leq s$.

B. Topology Defined by an Ideal

Let R be a ring, $\mathfrak{a}$ an ideal of R, and M an R-module. Then the **$\mathfrak{a}$-adic topology** of M is defined to be the topology on M such that $\{\mathfrak{a}^n M \mid n=1,2, \ldots\}$ is a †base for the neighborhood system of zero. In particular, let R be Noetherian, M a finite R-module, and N a submodule of M. Then by the Artin-Rees lemma, the $\mathfrak{a}$-adic topology of N coincides with the topology on N as a subspace of M with the $\mathfrak{a}$-adic topology. Returning to the general case, M is a †T_0-space (under the $\mathfrak{a}$-adic topology) if and only if $\bigcap_{n=1}^{\infty} \mathfrak{a}^n M = \{0\}$,

or, in other words, if and only if M is a †metric space, where the †distance $d(a,b)$ between points a, b in M is defined to be $\inf\{2^{-n} \mid a-b \in \mathfrak{a}^n M\}$. Moreover, if N is a submodule of M, then M/N is a T_0-space if and only if N is a closed subset of M, that is, if and only if $\bigcap_{n=1}^{\infty}(N+\mathfrak{a}^n M)=N$. A sequence $\{a_n\}=(a_1, a_2, \ldots, a_n, \ldots)$ is called a **Cauchy sequence** (under the $\mathfrak{a}$-adic topology) if $\forall n \exists N \forall r \forall s$ (with each of them a natural number) $a_{N+r}-a_{N+s} \in \mathfrak{a}^n M$; for this it is sufficient that $\forall n \exists N \forall r\ a_{N+r+1}-a_{N+r} \in \mathfrak{a}^n M$. If this sequence converges to zero (i.e., $\forall n \exists N \forall r\ a_{N+r} \in \mathfrak{a}^n M$), it is called a **null sequence**. The set $\mathfrak{M}$ of all Cauchy sequences in M becomes an R-module if we define their sum and multiplication by an element of R by $\{a_n\}+\{b_n\}=\{a_n+b_n\}$, $c\{a_n\}=\{ca_n\}$. Then the set $\mathfrak{N}$ of all null sequences is a submodule of $\mathfrak{M}$. An element m of M is identified with the sequence $(m, m, \ldots, m, \ldots)$, and we regard M as a submodule of $\mathfrak{M}$. Then the R-module $\bar{M}=\mathfrak{M}/\mathfrak{N}$ is the †completion of $M/(\bigcap_n \mathfrak{a}^n M)$ (as a metric space under the $\mathfrak{a}$-adic topology). $\bar{M}$ is called the **$\mathfrak{a}$-adic completion** of M. If $\mathfrak{a}$ has a finite basis, then the topology of $\bar{M}$ (as the completion) coincides with its $\mathfrak{a}$-adic topology. If $M=R$, we define a multiplication in $\mathfrak{M}$ by $\{a_n\}\{b_n\}=\{a_n b_n\}$. In this case, $\mathfrak{M}$ is a ring in which $\mathfrak{N}$ is an ideal, and hence the completion $\bar{R}=\bar{M}$ is a ring. If $\mathfrak{a}=\sum_{i=1}^{r} a_i R$, then considering the †ring of formal power series $\tilde{R}=R[[x_1, \ldots, x_r]]$ and its ideal $\tilde{\mathfrak{n}}=\bigcap_{n=1}^{\infty}(\sum_{i=1}^{r}(a_i-x_i)\tilde{R}+\mathfrak{a}^n\tilde{R})$, we have $\bar{R} \cong \tilde{R}/\tilde{\mathfrak{n}}$.

C. Zariski Rings

For a Noetherian ring R and an ideal $\mathfrak{a}$ of R, every element b of R such that $1-b \in \mathfrak{a}$ has an inverse in R if and only if every ideal of R is a closed subset of R under the $\mathfrak{a}$-adic topology. When this condition is satisfied, the ring R with the $\mathfrak{a}$-adic topology is called a **Zariski ring**; we often express this by saying that $(R, \mathfrak{a})$ is a Zariski ring. A Zariski ring is called **complete** if it is a complete topological space. The completion $\bar{R}$ of the Zariski ring $(R, \mathfrak{a})$ has the $\mathfrak{a}\bar{R}$-adic topology, and $(\bar{R}, \mathfrak{a}\bar{R})$ is a Zariski ring. Furthermore, (i) $\bar{R}$ is a †faithfully flat R-module; and (ii) when N is a submodule of a finite R-module M, then N is a closed subspace of M (under their $\mathfrak{a}$-adic topologies), and their completions are identified with $N \otimes_R \bar{R}$, $M \otimes_R \bar{R}$.

D. Local Rings

Suppose that R is a Noetherian ring having only a finite number of maximal ideals and J is the Jacobson radical of R. Then the Zariski ring (R, J) is called a **semilocal ring**. Furthermore, if R has only one maximal ideal, then (R, J) is called a **local ring**. A ring that has only a finite number of maximal ideals is called a **quasisemilocal ring**; if it has only one maximal ideal, it is called a **quasilocal ring**. (In some literature, the terms *local ring* and *semilocal ring* are used under weaker conditions; in the weakest case, quasisemilocal rings and quasi-local rings are simply called semilocal rings and local rings, respectively, and local rings and semilocal rings in our sense are called **Noetherian local rings** and **Noetherian semi-local rings**, respectively.)

Assume that R is a semilocal ring with maximal ideals $\mathfrak{m}_1, \ldots, \mathfrak{m}_n$ and the Jacobson radical $J=\mathfrak{m}_1 \cap \ldots \cap \mathfrak{m}_n$. For every finite R-module M, we introduce the J-adic topology as its natural topology. The completion $\bar{R}$ of R is a semilocal ring with maximal ideals $\mathfrak{m}_1\bar{R}, \ldots, \mathfrak{m}_n\bar{R}$ and is naturally isomorphic to the direct sum of the completions of the local rings $R_{\mathfrak{m}_i}$ $(i=1, \ldots, n)$. Since R is a Zariski ring, (1) $\bar{R}$ is faithfully flat; (2) submodules of M are closed subsets of M; and (3) the completion of M is identified with $M \otimes_R \bar{R}$. If $(R, \mathfrak{m})$ is a **complete local ring** (i.e., a local ring and a complete Zariski ring at the same time), then R contains a subring I with the following properties: (i) I is a complete local ring, and $I/(\mathfrak{m} \cap I)=R/\mathfrak{m}$; and (ii) for the †characteristic p of $R/\mathfrak{m}$ (p is either zero or a prime number), $\mathfrak{m} \cap I=pI$. Therefore, if $\mathfrak{m}$ is generated by n elements, then R is a homomorphic image of the ring of formal power series in n variables over I. This theorem is called the **structure theorem of complete local rings**, and I is called a **coefficient ring** of R. If R contains a field, then I is a field, called a **coefficient field** of R. A complete local ring is a †Hensel ring.

When $(R, \mathfrak{m})$ is a local ring and $\sum_{i=1}^{r} x_i R$ is $\mathfrak{m}$-primary, then we have an inequality $r \geqslant$ (†Krull dimension of R); if the equality holds, then we say that $x_1, \ldots, x_r$ form a **system of parameters** of R. Furthermore, if $\mathfrak{m}=\sum_{i=1}^{r} x_i R$, then we say that $x_1, \ldots, x_r$ form a **regular system of parameters** of R. A local ring that has a regular system of parameters is called a **regular local ring** (cf. †Jacobian criterion). A regular local ring is a †unique factorization domain. Let d be the Krull dimension of a local ring $(R, \mathfrak{m})$. Then R is a regular local ring if and only if one of the following holds: (1) every R-module has finite †homological dimension, (2) every R-module has homological dimension of at most d, or (3) the homological dimension of $R/\mathfrak{m}$ (as an R-module) is finite (and actually coincides with d). A Noetherian ring R' is called a **regular ring** if $R'_{\mathfrak{p}}$ is

a regular local ring for every prime ideal $\mathfrak{p}'$. A regular local ring is a regular ring.

Consider a local ring $(R, \mathfrak{m})$ and an $\mathfrak{m}$-primary ideal $\mathfrak{q}$. Then the length $l(n)$ of $R/\mathfrak{q}^n$ (as an R-module) is a function of n. For a sufficiently large n, the length $l(n)$ can be expressed as a polynomial $f(n)$ in n with rational coefficients. The degree of $f(n)$ coincides with the Krull dimension d of R. The **multiplicity** of $\mathfrak{q}$ is (coefficient of n^d in $f(n)) \times (d!)$. If $x_1, \ldots, x_d$ form a system of parameters, then (multiplicity of $\sum_{i=1}^{d} x_i R) \leqslant$ (length of $R/\sum x_i R$); if the equality holds, then we call $x_1, \ldots, x_d$ a **distinct system of parameters**. A local ring that has a distinct system of parameters is called a **Macaulay local ring**. A local ring is a Macaulay local ring if and only if one of the following holds: (1) every system of parameters is a distinct system of parameters, or (2) if an ideal $\mathfrak{a}$ of height s is generated by s elements, then every †prime divisor of $\mathfrak{a}$ is of height s. A regular local ring is a Macaulay local ring.

The notion of multiplicity can be also defined in general Noetherian rings [4]. Let R be a Noetherian ring. If $R_{\mathfrak{m}}$ is a Macaulay local ring for every maximal ideal $\mathfrak{m}$, then R is called a **locally Macaulay ring**. Furthermore, if height $\mathfrak{m}$ = Krull dim R for every maximal ideal $\mathfrak{m}$, then R is called a **Macaulay ring**. If R is a locally Macaulay ring, then the polynomial ring in a finite number of variables over R is also a locally Macaulay ring. In general, an ideal $\mathfrak{a}$ is called an **unmixed ideal** (or **pure ideal**) if the height of every prime divisor of $\mathfrak{a}$ coincides with height $\mathfrak{a}$; otherwise, $\mathfrak{a}$ is called a **mixed ideal**. Thus if R is a locally Macaulay ring, an ideal $\mathfrak{a}$ of the polynomial ring $R[x_1, \ldots, x_n]$ over R is generated by r elements, and height $\mathfrak{a} = r$, then $\mathfrak{a}$ is unmixed (**unmixedness theorem**).

If the completion of a local ring R is a †normal ring, then we say that R is **analytically normal**. If the completion of a semilocal ring R has no nilpotent element except zero, then we say that R is **analytically unramified**. A semilocal integral domain R which is a ring of quotients of a finitely generated ring over a field is analytically unramified. If R is a normal local ring, then R is analytically normal (O. Zariski).

Let $(R, \mathfrak{m})$ be a local ring, and let $\mathfrak{q}$ be an $\mathfrak{m}$-primary ideal. Set $F_i = \mathfrak{q}^i/\mathfrak{q}^{i+1}$ $(i = 0, 1, 2, \ldots,$ $\mathfrak{q}^0 = R)$. Let $a = a' \pmod{\mathfrak{q}^{i+1}} \in F_i$ and $b = b'$ $\pmod{\mathfrak{q}^{j+1}} \in F_j$. We put $ab = a'b' \pmod{\mathfrak{q}^{i+j+1}} \in F_{i+j}$. Then the direct sum of modules $F = \sum_{i=0}^{\infty} F_i$ becomes a graded ring generated by F_1 over F_0, in which F_i is the module of homogeneous elements of degree i. F, called the **form ring** (or **associated graded ring**) of R with respect to $\mathfrak{q}$, plays an important role in the theory of local rings, particularly in the theory of multiplicity.

E. Chains of Prime Ideals

Let R be a Noetherian ring with prime ideals $\mathfrak{p}$, $\mathfrak{q}$ such that $\mathfrak{p} \subset \mathfrak{q}$. Consider the length n of a chain of prime ideals $\mathfrak{p} = \mathfrak{p}_0 \subsetneqq \mathfrak{p}_1 \subsetneqq \cdots \subsetneqq \mathfrak{p}_n = \mathfrak{q}$ which cannot be refined any more. It is not true in general that n is uniquely determined by $\mathfrak{p}$ and $\mathfrak{q}$ (M. Nagata). However, n is uniquely determined for a rather large class of Noetherian rings, for instance the rings that are homomorphic images of locally Macaulay rings and, in particular, finitely generated rings over a †Dedekind domain.

F. Integral Closures

Let R be a Noetherian integral domain with the field of quotients k, let K be a finite algebraic extension of k, and let $\tilde{R}$ be the †integral closure of R in K. Then (i) If R is of Krull dimension 1, then for an arbitrary ring R' such that $R \subset R' \subset K$ and for every nonzero ideal $\mathfrak{a}'$ of R', the quotient $R'/\mathfrak{a}'$ is a finite $R/(\mathfrak{a}' \cap R)$-module. In particular, R' is a Noetherian domain satisfying the restricted minimum condition. (ii) If R is of Krull dimension 2, then $\tilde{R}$ is Noetherian. (iii) In general, $\tilde{R}$ is a †Krull ring, and for an arbitrary prime ideal $\mathfrak{p}$ of R there are only a finite number of prime ideals $\tilde{\mathfrak{p}}$ of R such that $\mathfrak{p} = \tilde{\mathfrak{p}} \cap R$. For any of such $\tilde{\mathfrak{p}}$, the field of quotients of $\tilde{R}/\tilde{\mathfrak{p}}$ is a finite algebraic extension of that of $R/\mathfrak{p}$. Result (i) is called the **Krull-Akizuki theorem**. We say that R satisfies the **finiteness condition for integral extensions** if $\tilde{R}$ is a finite R-module for any choice of K. A Noetherian ring R is called a **pseudogeometric ring**, or a universally **Japanese ring**, if $R/\mathfrak{p}$ satisfies the finiteness condition for integral extensions for every prime ideal $\mathfrak{p}$. A ring is pseudogeometric if it is generated by a finite number of elements over a pseudogeometric ring.

G. History

J. W. R. Dedekind first introduced the concept of ideals in the theory of integers. The main objects studied in ring theory were subrings of number fields or function fields until M. Sono (*Mem. Coll. Sci. Univ. Kyoto*, 2 (1917), 3 (1918–1919)) originated an abstract study of Dedekind domains, which was followed by E. Noether (*Math. Ann.*, 83 (1921), 96 (1926)), who originated the theory of Noetherian rings.

W. Krull made further important contributions to the development of the theory of Noetherian rings and general commutative rings [1]. Many other authors, including E. Artin, Y. Akizuki, and S. Mori, also contributed to the theory. The theory of local rings was originated by Krull (*J. Reine Angew. Math.*, 179 (1938)) and developed by C. Chevalley (*Ann. Math.*, 44 (1943)), I. S. Cohen (*Trans. Amer. Math. Soc.*, 59 (1946)), and Zariski (*Ann. Inst. Fourier*, 2 (1950)), and later by many authors, including P. Samuel, Nagata, M. Auslander, D. A. Buchsbaum, and J.-P. Serre [4]. The theory of Noetherian rings is applied to algebraic geometry.

References

[1] W. Krull, Idealtheorie, Erg. Math., Springer, second edition, 1968.
[2] B. L. van der Waerden, Algebra I, II, Springer, 1966–1967.
[3] O. Zariski and P. Samuel, Commutative algebra I, II, Van Nostrand, 1958–1960.
[4] M. Nagata, Local rings, Wiley, 1962 (Krieger, 1975).
[5] N. Bourbaki, Eléments de mathématique, Algèbre, ch. 1, 8, Actualités Sci. Ind., 1144b, 1261a, Hermann, 1964, 1958.
[6] N. Bourbaki, Eléments de mathématique, Algèbre commutative, ch. 1–7, Actualités Sci. Ind., Hermann; ch. 1, 2, 1290a, 1961; ch. 3, 4, 1293a, 1967; ch. 5, 6, 1308, 1964; ch. 7, 1314, 1965.
[7] J.-P. Serre, Algèbre locale, multiplicités, Lecture notes in math. 11, Springer, 1965.
[8] D. G. Northcott, Lessons on rings, modules and multiplicities, Cambridge Univ. Press, 1968.
[9] H. Matsumura, Commutative algebra, Benjamin, 1970; second edition, 1980.

285 (VI.16) Non-Euclidean Geometry

A. History

The validity of the fifth postulate of Euclid's *Elements*, the †axiom of parallels, has been a subject of argument ever since it was formulated (→ 139 Euclidean Geometry). At the beginning of the 18th century, G. Saccheri tried to prove the postulate by assuming the validity of other axioms. Under the hypothesis that the axiom does not hold, he deduced various extraordinary results. Although he was mistaken in thinking that he had obtained a contradiction, his work is regarded as a forerunner to the study of **non-Euclidean geometry**.

At the beginning of the 19th century, N. I. Lobachevskiĭ and J. Bolyai opened up the impasse by establishing a geometry based on postulates that contradict the fifth postulate. This geometry is called **hyperbolic geometry** or **Lobachevskiĭ's non-Euclidean geometry**. Actually, a similar idea had been conceived by C. F. Gauss, but he refrained from publishing it because of likely misunderstanding by a public still strongly influenced by I. Kant's philosophy. On the other hand, B. Riemann constructed so-called **elliptic geometry** (or **Riemann's non-Euclidean geometry**), which is different from both Euclidean and hyperbolic geometry. Euclidean geometry (including the theory of similarity) is sometimes called **parabolic geometry**. In general, a space satisfying axioms that contradict Euclid's postulates is called a **non-Euclidean space**.

Around the turn of the 20th century, A. Cayley, F. Klein, and H. Poincaré constructed models of non-Euclidean spaces that are subsets of Euclidean spaces, and E. Beltrami constructed a differential geometric model. By means of these models, it was established that non-Euclidean geometries are consistent as long as there is no inconsistency in the underlying Euclidean geometries. On the other hand, D. Hilbert established a complete system of axioms for Euclidean geometry and showed, by constructing non-Euclidean models, that the axiom of parallels is independent of the other axioms (→ 155 Foundations of Geometry). The logical foundation of non-Euclidean geometries was thus clarified. Moreover, A. Einstein showed in his †theory of relativity that actual space-time does not satisfy Euclidean axioms. Together with Euclidean spaces, non-Euclidean spaces are often used as fundamental models both in the problem of †space forms and in the theory of †symmetric spaces.

B. Axiomatic Considerations

Hilbert's system of axioms of plane Euclidean geometry consists of axioms of incidence, order, congruence, parallels, and continuity (→ 155 Foundations of Geometry). Specifically, the axiom of parallels is stated as follows: Suppose we are given a straight line and a point in a plane. If the straight line does not contain the point, then there exists only one line through the point that does not intersect the line.

The system of axioms of hyperbolic geometry is obtained by replacing the axiom of parallels by the following: Suppose we are given a straight line and a point in a plane. If the line does not contain the point, then there exist at least two lines that pass through the point without intersecting the line. (The other four groups of axioms are unaltered.) In this case, if l is a given line that does not contain a given point C, then there exist exactly two lines **parallel** to l that pass through C. We denote them by XY and $X'Y'$, and they are characterized as follows (Fig. 1): Any line that passes through C and lies in $\angle X'CY$ necessarily intersects l; by contrast, neither the two lines XY, $X'Y'$ nor any line in $\angle XCX'$ intersects l. Euclidean geometry can be considered as a "limit" of this geometry where the lines XY and $X'Y'$ coincide.

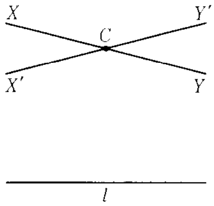


Fig. 1

In elliptic geometry, the axiom of parallels is replaced by the following: Suppose we are given a straight line and a point in a plane. If the line does not contain the point, then any line passing through the point intersects the line. In this case, the lines are closed curves, and the axioms of order must be modified. Specifically, in Euclidean geometry, the axioms of order are based on the notion of a point A lying between points B and C, where A, B, C are distinct points on a line. In elliptic geometry, however, to define order we utilize the notion of a pair A, C of points separating another pair B, D (and vice versa), where A, B, C, D are distinct points on a line. The axioms of order are modified accordingly.

The sum of inner angles of a triangle is smaller or greater than two right angles according as we use the axioms of hyperbolic or elliptic geometry.

C. The Projective-Geometric Point of View

We take †projective coordinates in an n-dimensional, real †projective space $\mathbf{P}^n$ and consider a quadric hypersurface defined by $Q: ax_0^2 + x_1^2 + \ldots + x_n^2 = 0$, $a \neq 0$ (→ 90 Coordinates; 343 Projective Geometry). We denote by G the group consisting of the totality of projective transformations of $\mathbf{P}^n$ that leave Q invariant. We call Q the **absolute** and call G the **group of congruent transformations**. When $a<0$, then this Q is a real quadric hypersurface. In this case, there exists a domain H^n (the totality of points inside Q) whose boundary coincides with Q, and the group G acts †transitively on H^n. The pair $\{G, H^n\}$ provides a model of hyperbolic geometry, and the n-dimensional **hyperbolic space** H^n is homeomorphic to an n-dimensional open cell. Points of H^n, points on Q, and points outside Q are called **ordinary points**, **points at infinity**, and **ultrainfinite points** (or **ideal points**), respectively. Two lines on H^n are said to be parallel if they intersect on the absolute Q. Next, when $a>0$, the absolute Q is an imaginary quadric hypersurface, and the group G acts transitively on $\mathbf{P}^n$. The pair $\{G, \mathbf{P}^n\}$ provides a model of elliptic geometry. The n-dimensional **elliptic space** P^n is homeomorphic to n-dimensional real projective space; hence it is compact. In elliptic geometry, any two distinct lines in a plane necessarily intersect at a point. The above models $\{G, \mathbf{P}^n\}$ of non-Euclidean geometries are called **Klein's models**.

Let $\{G, H^n\}$ be a Klein's model, A, B distinct points in P^n, l the line containing A, B, and I, J be two points where the line l meets Q (Fig. 2). If we denote by (A, B, I, J) the †anharmonic ratio of these four points, then the **non-Euclidean distance** ρ between the points A and B is given by $\rho = a\log(A, B, I, J)$, a constant. Next, let l, g be lines in H^n intersecting at a point D. In the plane determined by l and g, we draw two imaginary tangents u, v to the absolute through D. Denoting by (l, g, u, v) the anharmonic ratio of these four lines, the **non-Euclidean angle** θ between the lines l and g is given by $\theta = (1/2i)\log(l, g, u, v)$, $i = \sqrt{-1}$. Generally, let O be a point in the projective space $\mathbf{P}^n$, and denote by p the †polar of O with respect to the absolute. By counting the polar p doubly, it can be regarded as a quadric hypersurface, which will be denoted by S_0. A quadric hypersurface S of H^n is called a **non-Euclidean hypersphere** if it belongs to the †pencil of quadric hypersurfaces determined by Q and S_0. S is called a **proper hypersphere**, a **limiting hypersphere**, or an **equidistant hypersurface** according as the center O is an ordi-

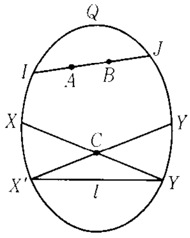


Fig. 2

nary point, a point at infinity, or an ultrainfinite point, respectively. In the case of elliptic geometry $\{G, \mathbf{P}^n\}$, the distance ρ between two points A and B is given by $\rho=(a/i)\cdot\log(A,B,I,J)$ as I, J are imaginary points. Moreover, we may get parabolic geometry as the "limit" $(a\to\infty)$ of the geometry given by Klein's models.

D. The Conformal-Geometric Point of View

Let S^n be an n-dimensional †conformal space. If we take suitable $(n+2)$-hyperspherical coordinates in S^n, the space S^n can be realized as a quadric hypersurface $x_1^2+x_2^2+\ldots+x_n^2-2x_0x_\infty=0$ in $(n+1)$-dimensional real projective space $\mathbf{P}^{n+1}$. A point in $\mathbf{P}^{n+1}$ represents a hypersphere in S^n (→ 76 Conformal Geometry; 90 Coordinates). We denote by $\tilde{G}$ the group consisting of the totality of †Möbius transformations leaving invariant a (real, point, or imaginary) hypersphere Q. When Q is a real hypersphere, the space S^n is divided into Q and the two open cells H^n, H^n_*. If we denote by G the totality of transformations of the group $\tilde{G}$ that do not interchange H^n and H^n_*, then G is a subgroup of index 2 of $\tilde{G}$. In this case, each of the pairs $\{G, H^n\}$ and $\{G, H^n_*\}$ provides a model of hyperbolic geometry. When Q is a point hypersphere, the space E^n obtained from S^n by omitting the point Q is homeomorphic to an open cell, and the pair $\{\tilde{G}, E^n\}$ provides a model of parabolic geometry. On the other hand, when Q is an imaginary hypersphere, $\tilde{G}$ is isomorphic to the †orthogonal group $O(n+1)$. We call the pair $\{\tilde{G}, S^n\}$ a **spherical geometry** and S^n an n-dimensional **spherical space**. In this case two points x and x' are called equivalent if x' is the image of x by symmetry with respect to Q. The space P^n obtained from S^n by identifying equivalent points is homeomorphic to the n-dimensional real projective space. If we denote by G the group obtained from $\tilde{G}$ by making its actions effective on P^n, then G is the factor group of $\tilde{G}$ by a cyclic group $\mathbf{Z}_2=\mathbf{Z}/2\mathbf{Z}$. The pair $\{G, P^n\}$ provides a model of elliptic geometry. These models are called **Poincaré's models**. They were introduced as a result of research on †automorphic functions in the case $n=2$.

In Poincaré's model, every straight line is represented either by a circle orthogonal to Q or by a circle passing through Q according as Q is a (real or imaginary) hypersphere or a point hypersphere. In spherical geometry, however, straight lines are usually called great circles, and two distinct great circles lying on a 2-dimensional sphere necessarily intersect at two points that are symmetric with respect to Q. Also, in Poincaré's model the distance between two points A and B is defined as before, by making use of the anharmonic ratio of four points A, B, I, J on a circle (Fig. 3) (→ 74 Complex Numbers G).

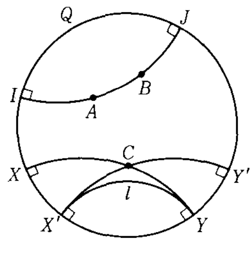


Fig. 3

E. The Differential-Geometric Point of View

An n-dimensional space M of †constant curvature is by definition a †Riemannian manifold whose line element ds is given by

$$ds^2=\frac{dx_1^2+dx_2^2+\ldots+dx_n^2}{\left(1+\frac{K}{4}(x_1^2+x_2^2+\ldots+x_n^2)\right)}$$

with respect to appropriate local coordinates, where K is a constant called the †sectional curvature (→ 364 Riemannian Manifolds). According as K is positive, zero, or negative, M can be considered locally as an elliptic space, Euclidean space, or hyperbolic space, respectively. In this case, lines are †geodesics of M, and the non-Euclidean distance and non-Euclidean angle are those defined in the Riemannian manifold. When $n=2$, a †simply connected and †complete space of positive constant curvature is †embedded in 3-dimensional Euclidean space as a sphere, and a space of negative constant curvature is †locally isometric to a **pseudosphere** (Fig. 4), which is a surface of revolution obtained by rotating a †tractrix around its asymptote. †Complete n-dimensional spaces of constant curvature $(n\geqslant 2)$ are called **space forms**. A simply connected space form is necessarily one of spherical space, Euclidean space, or hyperbolic space. Each of these is a †universal covering manifold of a general connected space form with a curvature of the same sign, and the group of †covering transformations is isomor-

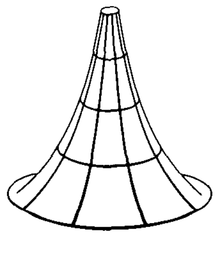


Fig. 4

phic to a †discontinuous subgroup of the group of congruent transformations.

Each of the spaces, Euclidean, non-Euclidean, and spherical, is a †homogeneous space on which the corresponding group of congruent transformations acts transitively. Actually, each of these spaces has the structure of a †symmetric Riemannian homogeneous space of rank 1.

References

[1] D. M. Y. Sommerville, The elements of non-Euclidean geometry, Bell, 1914.
[2] F. Klein, Vorlesungen über nicht-Euklidische Geometrie, Springer, 1928 (Chelsea, 1960).
[3] E. Cartan, Leçons sur la géométrie des espaces de Riemann, Gauthier-Villars, second edition, 1963.
[4] H. S. M. Coxeter, Non-Euclidean geometry, Univ. of Toronto Press, fifth edition, 1965.
[5] F. Engel and P. Stäckel, Urkunden zur Geschichte der nicht-Euklidischen Geometrie I, II, Teubner, 1898, 1913.

286 (XII.21) Nonlinear Functional Analysis

A. General Remarks

At present, the theory of nonlinear problems is still not unified, and many individual results obtained for specific classes of problems are stated in the languages of the corresponding fields of study. However, there are some fundamental facts and methods of a general nature concerning nonlinear problems, which may be referred to as the subject matter of **nonlinear functional analysis**.

B. Iterative Methods

Let X be a †Banach space. Consider a nonlinear mapping $G: X \to X$ and the equation

$$Gx = 0 \quad (x \in X). \tag{1}$$

Set $F = I - G$. Then (1) can be written as

$$x = Fx. \tag{2}$$

F satisfies the **Lipschitz condition** if there exists a constant α such that

$$\|Fx - Fy\| \leqslant \alpha \|x - y\| \tag{3}$$

for all x, y in X. In particular, the mapping F is said to be **nonexpansive** if $0 < \alpha \leqslant 1$. If $0 < \alpha < 1$, then F is called a **contraction**. (Sometimes a nonexpansive mapping is also called a contraction.) A contraction F satisfies the **contraction principle**: F has a unique fixed point x_0, and the **iteration**

$$x_{n+1} = Fx_n \quad (n = 1, 2, \ldots)$$

with an arbitrary initial element x_1 always converges to x_0 [1–3]. Similar results hold for a contraction that is defined only on a †ball and leaves the ball invariant. This leads to †Newton's iterative process and the †implicit function theorem.

C. Methods of Monotonicity

By definition, a nonlinear mapping G from a †Hilbert space H to H is a **monotone** or **accretive operator** if

$$\mathrm{Re}(Gx - Gy, x - y) \geqslant 0 \quad (x, y \in H).$$

G. Minty [4] proved that if $G: H \to H$ is monotone and continuous, then $\lambda I + G$ is a mapping onto H for any $\lambda > 0$, and its inverse $(\lambda I + G)^{-1}$ is nonexpansive. He has also shown that in the hypothesis of the theorem, we can replace the continuity requirement for G by maximality of G within the class of accretive operators that are possibly multivalued. Various developments of Minty's ideas, including generalization of his results to Banach spaces and applications to partial differential equations, have been obtained by F. Browder (*Amer. Math. Soc. Proc. Symposia in Appl. Math.*, 17 (1965)), J. Leray and J. L. Lions (*Bull. Soc. Math. France*, 93 (1965)), J. L. Lions [5], W. Strauss, H. Brézis (*Amer. Math. Soc. Proc. Symposia in Pure Math.*, 18 (1970)), and others.

A mapping A is said to be **dissipative** if $-A$ is accretive. Dissipative mappings play a central role in the theory of nonexpansive semigroups (→ Section X).

D. Topological Methods

In the geometric study of ordinary differential equations [6] some familiar theorems of topology and †differential topology have been strong tools, e.g., †Brouwer's fixed-point theorem is utilized to establish the existence of periodic solutions. However, in order to deal with nonlinear partial differential equations we have to generalize these theorems to infinite-dimensional cases. For example, a fixed-point theorem in an infinite-dimensional space was first obtained by G. D. Birkhoff and O. D.

Kellogg (*Trans. Amer. Math. Soc.*, 23 (1922), 95–115). The theory of the degree of mappings was generalized to the case of Banach spaces by J. Leray and J. Schauder [33] for the class of mappings of the form $I-F$, where F is a **compact** continuous mapping, i.e., the image by F of any bounded set is relatively compact. Let D be an open bounded set in a Banach space X and ∂D be its boundary. Let $F:\bar{D}\rightarrow X$ be a continuous compact mapping and Φ denote $I-F$. If a point p in X does not belong to $\Phi(\partial D)$, then we can define the **Leray-Schauder degree** $d(\Phi, p, D)$ of Φ relative to p [1–3]. The Leray-Schauder degree $d(\Phi, p, D)$ is an integer with the following properties: (i) $d(I, p, D)=1$ if $p\in D$. If $p\notin\Phi(D)$, then $d(\Phi, p, D)=0$. (ii) (Homotopy invariance) $d(\Phi, p, D)$ depends only on the compact homotopy class of $\Phi:\partial D\rightarrow X\setminus\{p\}$. More precisely, let $K:[0,1]\times\partial D\rightarrow X$ be a continuous compact mapping such that $x+K(t,x)\neq p$ for any $t\in[0,1]$ and any $x\in\partial D$. Let $\Phi_0(x)=x+K(0,x)$ and $\Phi_1(x)=x+K(1,x)$. Then $d(\Phi_0,p,D)=d(\Phi_1,p,D)$. (iii) If p and p' are in the same component of $X\setminus\Phi(\partial D)$, then $d(\Phi,p,D)=d(\Phi,p',D)$. (iv) (Continuity) $d(\Phi,p,D)$ is a continuous locally constant function of Φ (with respect to uniform convergence) and of $p\in X\setminus\Phi(\partial D)$. (v) (Domain decomposition) If D is the union of finite number of open disjoint sets D_j $(j=1,2,\ldots,N)$ with $\partial D_j\subset\partial D$ and $\Phi(x)\neq p$ on $\bigcup_{j=1}^{N}\partial D_j$, then $d(\Phi,p,D)=\sum_{j=1}^{N}d(\Phi,p,D_j)$. (vi) (Excision) If Δ is a closed subset of D on which $\Phi(x)\neq p$, then $d(\Phi,p,D)=d(\Phi,p,D\setminus\Delta)$. (vii) (Cartesian product formula) If $X=X_1\oplus X_2$ with $D_i\subset X_i$, $\Phi=(\Phi_1,\Phi_2)$ with $\Phi_i:\bar{D}_i\rightarrow X_i$ $(i=1,2)$, $D=D_1\times D_2$, and $p=(p_1,p_2)$, then $d(\Phi,p,D)=d(\Phi_1,p_1,D_1)\times d(\Phi_2,p_2,D_2)$, provided that the right-hand side is well defined.

The degree of mapping of Φ is also defined for some proper Fredholm mapping Φ (→ Section E; K. D. Elworthy and A. J. Tromba [38]; [32]).

Brouwer's fixed-point theorem (→ 153 Fixed-Point Theorems) is generalized to **Schauder's fixed-point theorem**: A compact mapping F of a closed bounded convex set K in a Banach space X into itself has a fixed point. Using the Leray-Schauder degree theory, one has the **Leray-Schauder fixed-point theorem**: Let D be a bounded open set of a Banach space X containing the origin O. Let $F(x,t):\bar{D}\times[0,1]\rightarrow X$ be a compact mapping such that $F(x,0)\equiv 0$. Suppose that $F(x,t)\neq x$ for any $x\in\partial D$ and $t\in[0,1]$. Then the compact mapping $F(x,1)$ has a fixed point x in D [1–3].

Other homotopy invariants, such as †homotopy groups and †cohomotopy groups, are also used in nonlinear functional analysis [1, 3]. For instance we have the following theorem of Shvarts (1964) [3]: Let X and Y be two Banach spaces. Let $D=\{x\in X \mid \|x\|<1\}$ be the unit ball and $\partial D=\{x\in X \mid \|x\|=1\}$ be the unit sphere of X. Suppose L is a fixed continuous linear †Fredholm operator from X to Y of index $p\geqslant 0$. Let PL be the set of compact perturbation of L mapping ∂D into $Y\setminus 0$, i.e., $PL=\{\Phi=L+K \mid K$ is a continuous compact mapping of ∂D to Y such that $\Phi(x)=Lx+K(x)\neq 0$ for $x\in\partial D\}$. Two mappings Φ_0 and Φ_1 in PL are said to belong to the same **compact homotopy class** on ∂D if there exists a continuous compact mapping $h:[0,1]\times\partial D\rightarrow Y$ such that $Lx+h(t,x)\neq 0$ for x in ∂D, $\Phi_0(x)=Lx+h(0,x)$, and $\Phi_1(x)=Lx+h(1,x)$.

Shvarts's theorem: Let L be a fixed continuous linear Fredholm mapping from X to Y of index $p\geqslant 0$. Then the compact homotopy classes on ∂D of PL are in one-to-one correspondence with the elements of the pth stable homotopy group $\pi_{n+p}(S^n)$ $(n\geqslant p+1)$ (→ 202 Homotopy Theory H).

Warning: The topological structure of an infinite-dimensional Hilbert or Banach space is quite different from that of a finite-dimensional Euclidean space. For instance, let X be a Hilbert space of infinite dimension, and let $D=\{x\in X \mid \|x\|<1\}$ be its unit ball and $\partial D=\{x\in X \mid \|x\|=1\}$ be the unit sphere of X. S. Kakutani (*Proc. Imp. Acad. Tokyo*, 19 (1943)) gave a fixed-point free continuous mapping of $\bar{D}$ into itself if X is separable [1–3]. Thus naive generalization of the Brouwer fixed-point theorem is no longer true in infinite-dimensional spaces. V. Klee and C. Bessaga [17] proved that the unit sphere ∂D is C^∞-diffeomorphic to X for an arbitrary Hilbert space X of infinite dimension. N. H. Kuiper proved that the group of invertible continuous linear operators on X is contractible if X is separable [3]. All this is in striking contrast to the well-known facts for finite-dimensional spaces (→ 202 Homotopy Theory, 427 Topology of Lie Groups and Homogeneous Spaces). This is the reason why compactness assumptions are made in the theorems mentioned above.

E. Calculus in Banach (or Locally Convex) Spaces

When one considers a nonlinear operator, it often happens that the domain and the range are neither linear spaces nor their open subsets. The domain might be a space of all smooth mappings of a compact manifold into another, and so might be the range. Such spaces have no linear structure, and hence linearity or semilinearity do not make sense in

general. The concept of infinite-dimensional manifolds is therefore introduced of necessity in nonlinear functional analysis.

Definition of differentiable mappings. Let E and F be real Banach spaces and let $L(E,F)$ $(=L(E)$ if $E=F)$ be the Banach space of all bounded linear operators with uniform operator norm. Let U be an open subset of E and x a point of U. A mapping (= nonlinear operator) f of U into F is called **Gâteaux differentiable** at x if $\lim_{t\to 0} t^{-1}(f(x+ty)-f(x))=df(x,y)$ exists for any $y\in E$. $df(x,y)$ is called the **Gâteaux derivative** of f at x. f is called **Fréchet differentiable** at x if there exists a linear operator $A\in L(E,F)$ such that $\lim_{y\to 0}\|f(x+y)-f(x)-Ay\|/\|y\|=0$. A is called the **Fréchet derivative** of f at x and is denoted by $df(x)$ or $f'(x)$. f is Fréchet differentiable in U if and only if it is Gateaux differentiable, $df(x,y)$ is linear in y, and $\sup_{y\neq 0}\|df(x,y)\|/\|y\|$ is bounded [10].

Let U be an open subset of E. A mapping f of U into F is said to be of **class** C^0 if it is continuous and to be of **class** C^1 if it is Fréchet-differentiable at each point $x\in U$ and the **differential** $df(x)\in L(E,F)$ is continuous as a mapping of U into $L(E,F)$. The differential $df(x)$ is also called the **linearized operator**. If the mapping $df: U\to L(E,F)$ is of class C^{r-1}, then f is said to be of **class** C^r. $d(d^{r-1}f)(x)$ is written as $d^rf(x)$, and called the r**th differential** at x. $d^rf(x)$ is an r-linear, bounded, symmetric operator of $E\times\ldots\times E$ (r times) into F. f is said to be of **class** C^∞ if f is of class C^r for every r. For an open subset V of F, $f: U\to V$ is called a C^r **diffeomorphism** if f is a bijection and both f and f^{-1} are of class C^r.

A C^r mapping f of U into F is called a **Fredholm mapping (Fredholm map)** if $df(x)\in L(E,F)$ is a linear †Fredholm operator for every $x\in U$. Since $\operatorname{Ind} df(x)$ is constant if U is connected, that integer $\operatorname{Ind} df(x)$ is called the **index** of f.

F. Taylor's Theorem and Its Converse

Let $f: U\to F$ be of class C^r $(r\geqslant 1)$. A generalized Taylor theorem claims that f can be approximated by a polynomial mapping: Let $x\in U$ and $y\in E$ be sufficiently close to 0 so that $x+ty\in U$ for $0\leqslant t\leqslant 1$. Then

$$f(x+y)=\sum_{k=0}^{r}\frac{1}{k!}d^kf(x)(y,\ldots,y)+\int_0^1\frac{(1-t)^{r-1}}{(r-1)!}\times\{d^rf(x+ty)-d^rf(x)\}(y,\ldots,y)dt.$$

Let $L^k_{\mathrm{sym}}(E,F)$ be the Banach space of all k-linear, bounded, symmetric operators of $E\times\ldots\times E$ into F with the uniform topology. If for every k, $0\leqslant k\leqslant r$, there exists a continuous mapping $\varphi_k: U\to L^k_{\mathrm{sym}}(E,F)$ such that

$$f(x+y)=\sum_{k=0}^{r}\frac{1}{k!}\varphi_k(x)(y,\ldots,y)+o(\|y\|^r)$$

at every $x\in U$ and y sufficiently close to 0, then $f: U\to F$ is of class C^r, and $d^kf(x)=\varphi_k(x)$.

G. The Implicit Function Theorem

Using the notation above, let $f: U\to V$ be of class C^r, $r\geqslant 1$, and assume that $0\in U$, $0\in V$, and $f(0)=0$, where 0 is the origin of E or F. Suppose there is an $A\in L(F,E)$, called the **right inverse** of $df(0)$, such that $df(0)A=1_F$ (the identity). Then the following assertions hold: (i) The image of F under A, AF, is a closed subspace of E, and $E=\operatorname{Ker} df(0)\oplus AF$. (ii) There are neighborhoods U_1, U_2, V' of the zeros of $\operatorname{Ker} df(0)$, AF, F, respectively, such that $U_1\oplus U_2\subset U$ and such that the mapping g: $U_1\oplus U_2\to U_1\oplus V'$ defined by $g(u,v)=(u,f(u,v))$ is a C^r-diffeomorphism. Therefore, denoting the inverse of g by $h=(h_1,h_2)$, we have $h_1(u,w)\equiv u$ and $f(u,h_2(u,w))\equiv w$. The latter means that the nonlinear equation $f(u,v)=w$ can be solved with respect to v.

H. Existence and Uniqueness of Integral Curves

Using the notation above, let f be a C^r mapping $(r\geqslant 1)$ of U into E. Since $U\times E$ is the †tangent bundle of U, $(x,f(x))$ can be regarded as a C^r tangent vector field on U. The equation of †integral curves is $(d/dt)x(t)=f(x(t))$. A local existence and uniqueness of solutions is stated as follows: For an arbitrarily fixed $x\in U$, there are $\varepsilon>0$ and an open neighborhood W of x such that there exists uniquely a C^r mapping h of $W\times(-\varepsilon,\varepsilon)$ into U satisfying $(d/dt)\,h(w,t)=f(h(w,t))$ and $h(w,0)\equiv w$.

Using this fact, one can prove the **Frobenius theorem**: Let E' be a closed linear subspace of E with a direct summand E'', and let $f: U\to L(E',E'')$ be a C^r mapping $(r\geqslant 1)$ such that $f(0)=0$. To each $x\in U$ one associates a closed linear subspace $D_x=\{(v,f(x)v)\mid v\in E'\}$. The disjoint union $D=\bigcup_{x\in U}D_x$ can be regarded as a subbundle of the tangent bundle $U\times E$. A mapping $\tilde{u}$ of U into E is called a **cross section** of D if $\tilde{u}(x)\in D_x$ for every $x\in U$. D is called **involutive** if for any two C^1 cross sections $\tilde{u}$, $\tilde{v}$ of D, the Lie bracket product $[\tilde{u},\tilde{v}]$ defined by $[\tilde{u},\tilde{v}](x)=d\tilde{u}(x)(\tilde{v}(x))-d\tilde{v}(x)(\tilde{u}(x))$ is again a cross section of D. Now suppose D is an involutive subbundle of $U\times E$. Then for an arbitrarily fixed $x\in U$, there are a neighborhood W

of x and a C^r diffeomorphism h of W onto a neighborhood of 0 of E such that $dh(x)(D_x) \equiv E'$. This fact shows that an involutive subbundle can be trivialized by a suitable change of local coordinate systems.

I. Local Theories on Locally Convex Spaces

All local theories mentioned in Sections E–H are constructed on Banach spaces. However, it is important in concrete applications to construct these theories on a wider class of locally convex topological linear spaces.

Let E, F be †locally convex topological linear spaces, and let U be an open subset of E. A mapping $f: U \rightarrow F$ is said to be of class C^0 if it is continuous. f is said to be of class C^r if f is of class C^{r-1} and the following is fulfilled: For every $x \in U$, there is an r-linear continuous symmetric mapping $d^r f(x)$ of $E \times \ldots \times E$ into F such that $d^r f: U \times E \times \ldots \times E \rightarrow F$ is continuous. If we put

$$F(y) = f(x+y) - f(x) - df(x)(y) - \ldots - \frac{1}{r!} d^r f(x)(y, \ldots, y)$$

for every y sufficiently close to $0 \in E$, then the mapping G defined by

$$G(t,y) = \begin{cases} F(ty)/t^r, & t \neq 0, \\ 0, & t = 0, \end{cases} \quad (t \in \mathbf{R})$$

is continuous at $(0,0) \in \mathbf{R} \times E$. The definitions of C^∞ mappings and C^r diffeomorphisms are given as in Section E.

J. Implicit Function Theorems in Locally Convex Spaces

The implicit function theorem does not hold in general locally convex spaces. However, since it is useful for nonlinear problems, several sufficient conditions are presently being studied. The following are some of them. We assume that $f: U \rightarrow F$ is a C^r mapping ($r \geqslant 1$) such that $f(0) = 0$ and that $df(0): E \rightarrow F$ has a continuous right inverse.

(i) If $\dim F < \infty$, the implicit function theorem holds just as in Section G.

(ii) An implicit function theorem that can be applied to nonlinear elliptic differential operators can be restated as follows. Suppose E, F are †projective limit of families of Banach spaces $\{E^k, k \geqslant d\}$, $\{F^k, k \geqslant d\}$ and $U = E \cap U^d$, where U^d is an open subset of E^d. If $f: U \rightarrow F$ can be extended to a C^r mapping ($r \geqslant 2$) of $E^k \cap U^d$ into F^k for every $k \geqslant d$, if f satisfies the following inequalities, called a linear estimate with respect to $|\ |_k$:

$$\begin{aligned} |df(x+y)(v)|_k &\leqslant C(|y|_k|v|_d + |y|_d|v|_k) \\ &\quad + P_k(|y|_{k-1})|v|_{k-1}, \quad k > d, \\ |d^2 f(x+y)(u,v)|_k &\leqslant C(|y|_k|u|_d|v|_d + |y|_d|u|_k|v|_d \\ &\quad + |y|_d|u|_d|v|_k) + P_k(|y|_{k-1})|u|_{k-1}|v|_{k-1}, \quad k > d, \end{aligned} \tag{4}$$

where $|\ |_k$ is the norm in E^k or F^k, C is a positive constant independent of k, and P_k is a polynomial with positive coefficients, and if a right inverse A of $df(0)$ satisfies the inequality of Gårding type,

$$|Au|_k < C'|u|_k + D_k|u|_{k-1}, \quad k > d, \tag{5}$$

where $C' > 0$ independent of k and $D_k > 0$, then the implicit function theorem holds just as in Section G, and the obtained mapping h satisfies the same inequalities as (4) [11].

(iii) **Nash-Moser implicit function theorem.** Though linear estimates such as (4) hold for many differential operators, the second inequality (5) is sometimes out of order, especially if f is a nonlinear hyperbolic operator. However, one can often obtain instead of (5) a weaker inequality:

$$|Au|_k \leqslant C'|u|_{k+s} + D_k|u|_{k+s-1}, \quad s > 0.$$

J. Nash [37] and J. Moser [12] approximated such an operator A by some smoothing operators and proved an implicit function theorem under a certain additional condition. The Nash-Moser implicit function theorem was successfully applied to many difficult problems, e.g., the embedding problem of Riemannian manifolds [37], the small divisor problem of celestial mechanics [36], free boundary problems (e.g., L. Hörmander *Arch. Rational Mech. Anal.*, 62 (1976), 1–52), and other problems (e.g., S. Klainerman, *Comm. Pure Appl. Math.*, 33 (1980), 43–101; M. Kuranishi, *Amer. Math. Soc. Proc. Symposia in Pure* Math., 30 (1977), 97–105).

(iv) **Analytic implicit function theorem.** In cases (ii) and (iii), the spaces E, F were given as projective limits of Banach spaces. On the contrary, H. Jacobowitz considered the case where E, F are †inductive limits of Banach spaces. For instance, the space E of the smooth functions can be approximated by a family of Banach spaces $\{E_\varepsilon | \varepsilon > 0\}$ of all real analytic functions with ε as the radius of convergence. Under this circumstance, certain conditions for f and the right inverse of $df(0)$ yield an implicit function theorem [13].

(v) Mather's implicit function theorem [14]. The difficulty of implicit function theorems in Fréchet spaces is concentrated in the following fact: Even if $df(0)$ has a right inverse and even if x is sufficiently close to 0, $df(x)$ may not have

a right inverse. If the implicit function theorem holds, such a phenomenon should not happen. In cases (ii)–(iv), several functional-analytic conditions exclude this pathological phenomenon. In some special cases, these conditions can be replaced by algebraic ones. J. Mather, using his division theorem, proved an implicit function theorem that was applied to the theory of singularities.

K. Infinite-Dimensional Manifolds

A Hausdorff space M is called a C^r **Banach manifold** modeled on a Banach space E if the following conditions are satisfied: (a) M is covered by a family of open subsets $\{U_\alpha\}_{\alpha\in A}$. For each U_α there are an open subset V_α of E and a homeomorphism ψ_α of U_α onto V_α. Such a pair is called a local coordinate system or a local chart of M. (b) If $U_\alpha \cap U_\beta \neq \varnothing$, $\psi_\alpha\psi_\beta^{-1}$ is a C^r-diffeomorphism of $\psi_\beta(U_\alpha \cap U_\beta)$ onto $\psi_\alpha(U_\alpha \cap U_\beta)$. (c) The index set A is maximal among those that satisfy (a) and (b).

If M (resp. M') is a C^r Banach manifold modeled on E (resp. E'), then so is $M \times M'$ modeled on $E \oplus E'$. A mapping $f: M \rightarrow M'$ is said to be of class C^r if f expressed through local coordinate systems is of class C^r.

Suppose M is covered by a family of local charts $\{(U_\alpha, \psi_\alpha)\}_{\alpha\in A}$. On the disjoint union $\bigcup_{\alpha\in A} U_\alpha \times E$, define an equivalence relation $\sim$ as follows: For $(x,u)\in U_\alpha\times E$, $(y,v)\in U_\beta \times E$, $(x,u)\sim(y,v)$ if and only if $x=y$ and $d(\psi_\alpha\psi_\beta^{-1})(x)v=u$. The set of equivalence classes is called the **tangent bundle** of M, which will be denoted by T_M. There is another definition of the tangent bundle which uses the ring of C^r functions on M. However, since E is not reflexive in general, the latter gives us a different vector bundle. T_M is a C^{r-1} Banach manifold modeled on $E\oplus E$. The correspondence which sends $(x,u)\in\bigcup_{\alpha\in A} U_\alpha\times E$ to x induces a C^{r-1} mapping of T_M onto M, called the **projection** of T_M, denoted by π.

A topological group G is called a **Banach-Lie group** if G is a C^∞ Banach manifold and the group operation $(g,h)\rightarrow g^{-1}h$ is a C^∞ mapping of $G\times G$ into G. If E is a Banach space, then $GL(E)$, the group of all invertible bounded linear operators, is a Banach-Lie group under the uniform topology. If E is a Hilbert space, then the group of all unitary operators is also a Banach-Lie group under the same topology.

The concepts of manifolds and Lie groups are similarly defined when the model space is a Hilbert space or a Fréchet space. These are called a **Hilbert manifold** and a **Fréchet manifold**, respectively. For some differential topologies on separable Hilbert manifolds $\rightarrow$ 279 Morse Theory E.

L. Structures on Infinite-Dimensional Manifolds

Suppose M is a C^{r+1} Hilbert manifold modeled on E. At each $x\in M$, the tangent space $T_xM=\pi^{-1}(x)$ is a Hilbert space, linear-homeomorphic to E. M is called a C^r **Riemannian manifold** if there is defined an inner product $\langle u,v\rangle_x$ on each T_xM such that $\langle\cdot,\cdot\rangle_x$ is of class C^r with respect to x. Existence of such a structure is ensured by using a partition of unity if M is paracompact.

Let M be a C^r Banach manifold ($r\geqslant 1$) modeled on a Banach space E. Each tangent space T_xM is linear homeomorphic to E. M is called a C^r **Finsler manifold** if there is a norm $|u|_x$ defined on each T_xM such that $|\ |_x$ is continuous with respect to x. A paracompact C^1 Banach manifold can have a C^1 Finsler structure.

Let M be a C^{r+2} Fréchet manifold ($r\geqslant 0$) and $C^{r+1}(M)$, $\Gamma^{r+1}(T_M)$ the spaces of all C^{r+1} functions and of all C^{r+1} vector fields on M, respectively. A bilinear mapping ∇ of $\Gamma^{r+1}(T_M)\times\Gamma^{r+1}(T_M)$ into $\Gamma^r(T_M)$ is called an **affine connection** on M if ∇ satisfies $\nabla_{f\tilde u}\tilde v=f\nabla_{\tilde u}\tilde v$, $\nabla_{\tilde u}f\tilde v=(\tilde u f)\tilde v+f\nabla_{\tilde u}\tilde v$ for every $\tilde u,\tilde v\in\Gamma^{r+1}(T_M)$, $f\in C^{r+1}(M)$. For an affine connection ∇, $T(\tilde u,\tilde v)=\nabla_{\tilde u}\tilde v-\nabla_{\tilde v}\tilde u-[\tilde u,\tilde v]$ is called the **torsion tensor**, and $R(\tilde u,\tilde v)=\nabla_{\tilde u}\nabla_{\tilde v}-\nabla_{\tilde v}\nabla_{\tilde u}-\nabla_{[\tilde u,\tilde v]}$ is called the **curvature tensor** of ∇. If M is a Riemannian manifold, then there exists a unique affine connection without torsion which leaves the Riemannian inner product parallel.

M. Local Linearization Theorems

Let M be a C^{r+1} Banach manifold ($r\geqslant 1$) modeled on E. Let $\tilde u$ be a C^r vector field on M such that $\tilde u(x)\neq 0$ at $x\in M$. Then there are a neighborhood U_x of x and a C^r diffeomorphism ψ of U_x onto an open subset V of E such that $d\psi\tilde u(\psi^{-1}(y))\equiv(y,v)$, $v\in E$, for every $y\in V$, where v does not depend on y.

N. Morse Lemma

Let M be a C^{r+2} Hilbert manifold and f be an $\mathbf{R}$-valued C^{r+2} function. Suppose x is a critical point of f, i.e., $df(x)=0$. x is called a **nondegenerate critical point** if $d^2f(x)$ is a nondegenerate bilinear form. For such x there are a neighborhood U_x of x and a C^r diffeomorphism ψ of U_x onto an open neighborhood of 0 of the model space E such that $\psi(x)=0$, and $f(\psi^{-1}(y))=|Py|^2-|(1-P)y|^2$, where P is an orthogonal

projection in E. $i_x = \dim(1-P)E$ $(0 \leqslant i_x \leqslant \infty)$ is called the **index** of the critical point x of f.

O. Submanifolds

A subset N of a C^r Banach manifold M modeled on E is called a C^r **submanifold** if at each point $x \in N$ there are a neighborhood U_x of x and a C^r-diffeomorphism ψ of U_x onto an open neighborhood V of 0 of E such that $\psi(x) = 0$ and $\psi(U_x \cap N) = V \cap F$, where F is a closed linear subspace of E. There are some other definitions of submanifolds. One of them requires in addition that F be a direct summand of E, and another uses instead of $U_x \cap N$ its connected component containing x. In the latter definition, a submanifold is not necessarily locally closed.

P. Sard-Smale Theorem

Although it is not easy to define nontrivial measures on infinite-dimensional manifolds, the concept "almost everywhere" can sometimes be replaced by that of residual sets. A subset of a topological space is called a †residual set if it contains an intersection of countably many open dense subsets. A residual set in a complete metric space or in a †Baire space is dense. S. Smale [15] extended †Sard's theorem to infinite-dimensional manifolds as follows: Let M, N be C^r Banach manifolds and $f: M \to N$ be a C^r Fredholm mapping. If M is separable and $r > \max\{0, \mathrm{Ind}(df(x))\}$ for each $x \in M$, then $R_f = N - f(C)$ is a residual subset of N, where C is the set of critical points of f, i.e., a point x where $df(x)$ is not surjective.

Q. Calculus of Variations and Infinite-Dimensional Manifolds

Many problems in the calculus of variations can be understood as problems seeking critical points of functions defined on infinite-dimensional manifolds. R. Palais and S. Smale set up the following Condition C and fixed a category of functions where the critical points can be chased through gradient-like vector fields [6].

Palais-Smale Condition C. Let f be a C^1 function on a C^1 Finsler manifold M. If S is any subset of M on which f is bounded but on which $|df(x)|$ is not bounded away from 0, then there is a critical point of f adherent to S.

In general, it is not easy to examine Condition C for a concrete f. However, many concrete problems, where the Euler equations are nonlinear elliptic, satisfy Condition C.

Morse theory. Let M be a C^∞ complete Riemannian manifold and f a C^∞ function bounded below satisfying Condition C and having only nondegenerate critical points. Then, using the Morse lemma, one can make a †handlebody decomposition of M by the same method as in the case of finite-dimensional manifolds (→ 279 Morse Theory).

Lyusternik-Shnirel'man theory. This theory, constructed on finite-dimensional manifolds, can be extended naturally to Finsler manifolds. Let M be a complete C^2 Finsler manifold and f a C^2 function satisfying Condition C and bounded below. Then f has at least $\mathrm{cat}(M)$ critical points, where $\mathrm{cat}(M) = m$ means that M can be covered by m closed contractible subsets of M but not by $m-1$ ones. If there is no such integer, then we set $\mathrm{cat}(M) = \infty$.

Both Morse theory and Lyusternik-Shnirel'man theory have been successfully employed in the global theory of the calculus of variations.

R. Bifurcation Theory

Bifurcation theory concerns itself with the structure of the zeros of the functional equation of w with a parameter λ:

$$G(\lambda, w) = 0. \tag{6}$$

In general, the state w satisfying (6) represents the equilibrium (time-independent or stationary) solution of the †evolution equation

$$w_t = G(\lambda, w) \quad \text{and} \quad w(0) = w_0. \tag{7}$$

Here the evolution equation itself stems from a mathematical model describing natural phenomena, $w = w(t)$ stands for the state at time t, and λ is the set of parameters representing the physical environment. For example, in the †Navier-Stokes equation appearing in fluid dynamics, $w(t)$ represents the unknown velocity field at time t and λ is the †Reynolds number. It is important to study bifurcation phenomena because they typically accompany the transition to instability of the state when some characteristic parameter passes through a certain value, called a **critical value**.

Let X, Y, and Λ be real Banach spaces, and let $G(\lambda, w)$ be a mapping from $\Lambda \times X$ to Y. Suppose that there exists a mapping $\tilde{w}(\lambda): \Lambda \to X$ satisfying $G(\lambda, \tilde{w}(\lambda)) = 0$. One calls $(\lambda, \tilde{w}(\lambda))$ a trivial solution of (6).

$(\lambda_0, \tilde{w}(\lambda_0))$ is called a **bifurcation point** of $G(\lambda, w)$ (with respect to the trivial solution) if in any neighborhood of $(\lambda_0, \tilde{w}(\lambda_0))$ there exists a nontrivial solution of (6). (In general, there may appear another type of solution that is not connected with the trivial solution [20]).

Assume that $G(\lambda, w)$ is of class C^1 in some neighborhood of $(\lambda_0, \tilde{w}(\lambda_0))$ in $\Lambda \times X$. Then it follows from the †implicit function theorem that $(\lambda_0, \tilde{w}(\lambda_0))$ is not a bifurcation point if $G_w(\lambda_0, \tilde{w}(\lambda_0))$, the †Fréchet derivative of G with respect to w, is nonsingular.

S. The Principle of Linearized Stability

Closely tied to the phenomenon of bifurcation is the property of **stability**. Suppose that the dynamics of a physical system are governed by (7). Let $w(t; w_0)$ denote the solution of (7). An equilibrium solution $\tilde{w} = \tilde{w}(\lambda)$ is called **stable** if for any $\varepsilon > 0$ there exists a $\delta > 0$ such that $\|w(t; w_0) - \tilde{w}\| < \varepsilon$ for all $t > 0$ whenever $\|w_0 - \tilde{w}\| < \delta$. Furthermore, $\tilde{w}$ is said to be **asymptotically stable** if, in addition, $w(t; w_0) \to \tilde{w}$ as $t \to \infty$.

By the **principle of linearized stability** we mean that the stability of an equilibrium solution $\tilde{w}$ is determined formally by the †spectrum of the linearized operator $G_w(\lambda, \tilde{w}(\lambda))$. Assume that $G_w(\lambda, \tilde{w}(\lambda))$ has only a †point spectrum. Then (i) if $\tilde{w}(\lambda)$ is stable, the spectrum of $G_w(\lambda, \tilde{w}(\lambda))$ is contained in $\{z \in \mathbf{C} \mid \operatorname{Re} z \leqslant 0\}$, and (ii) if the spectrum of $G_w(\lambda, \tilde{w}(\lambda))$ is contained in $\{z \in \mathbf{C} \mid \operatorname{Re} z < 0\}$, $\tilde{w}$ is asymptotically stable.

Suppose that as λ crosses a certain value λ_0, one or more eigenvalues of $G_w(\lambda, \tilde{w}(\lambda))$ cross the imaginary axis from the left to the right half-plane, where $\tilde{w}(\lambda)$ is the known equilibrium solution. This is precisely the situation when $\tilde{w}(\lambda)$ becomes unstable. For notational simplicity, we put $u = w - \tilde{w}(\lambda)$ and $F(\lambda, u) = G(\lambda, u + \tilde{w}(\lambda))$.

T. Bifurcation from Simple Eigenvalues

Take $\Lambda = \mathbf{R}$. Let $F \in C^2(\mathbf{R} \times X, Y)$ be such that $F(\lambda, 0) = 0$ for any real λ. Set $L_0 = F_u(\lambda_0, 0)$, $L_1 = F_{\lambda,u}(\lambda_0, 0)$, and suppose that (i) $\operatorname{Ker}(L_0)$ is spanned by $u_0 \neq 0$; (ii) $\operatorname{codim} R(L_0) = 1$; and (iii) $L_1 u_0 \notin R(L_0)$, where $R(L_0)$ denotes the †range of L_0. Then there exists a C^1-curve $(\lambda, \psi): (-\delta, \delta) \to \mathbf{R} \times X$ defined on some interval $(-\delta, \delta)$ such that $\lambda(0) = \lambda_0$, $\psi(0) = 0$, and $F(\lambda(s), s(u_0 + \psi(s))) = 0$ for any $s \in (-\delta, \delta)$. Moreover, in a neighborhood of $(\lambda_0, 0)$ any zero of F either lies on this curve or is a trivial solution (M. G. Crandall and P. H. Rabinowitz, *J. Functional Anal.*, 8 (1971)).

In the above case, there exist only three possible situations of the curves (λ, φ) and $(\lambda, 0)$, called subcritical, supercritical, and transcritical bifurcations. In the third case, there occurs the so-called **exchange of stability** [19–21].

U. Bifurcation of Periodic Solutions (Hopf Bifurcation Theorem)

If $u(\lambda)$ loses stability by virtue of a pair of complex conjugate eigenvalues crossing the imaginary axis, then under suitable conditions one can prove the existence of bifurcating time-periodic solutions of (7). Rewrite equation (7) as

$$u_t + L_0 u + g(\lambda, u) = 0. \tag{8}$$

Suppose (i) $L_0: D(L_0) \subset X \to X$ is a densely defined linear operator on X such that $-L_0$ generates a strongly continuous semigroup on X, which is holomorphic on $X^{\mathbf{C}}$ (= the complexification of X). L_0 has compact resolvent; i is a simple eigenvalue of L_0, and $ni \notin \sigma(L_0)$, the spectrum of L_0, for $n = 0, 2, 3, \ldots$. As a consequence of (i), if $r > -\operatorname{Re} \lambda$ for any $\lambda \in \sigma(L_0)$, then the fractional power $(rI + L_0)^\alpha$ for $\alpha \geqslant 0$ is well defined. Because their domains are independent of r, one can set $X_\alpha = D((rI + L_0)^\alpha)$, which are Banach spaces under the norm $\|u\|_\alpha = \|(rI + L_0)^\alpha u\|$. Suppose (ii) there exist an $\alpha \in [0, 1)$ and a neighborhood $\mathcal{O}$ of $(0, 0)$ in $\mathbf{R} \times X$ such that $g \in C^2(\mathcal{O}, X_\alpha)$, where $C^k(\mathcal{O}, X_\alpha)$ denotes the space of all X_α-valued C^k functions defined on $\mathcal{O}$. Moreover $g(\lambda, 0) = 0$ if $(\lambda, 0) \in \mathcal{O}$, and $g_u(0, 0) = 0$. (iii) Let $\beta = \beta(\lambda)$ be a continuously differentiable function defined in a neighborhood of 0 such that $\beta(\lambda) \in \sigma(L_0 + g_u(\lambda, 0))$ and $\beta(0) = i$. Suppose that $\operatorname{Re} \beta'(0) \neq 0$. If assumptions (i), (ii) and (iii) are satisfied, then there exist a positive δ and continuously differentiable functions $(\rho, \lambda, u): (-\delta, \delta) \to \mathbf{R}^2 \times C^0(\mathbf{R}, X_\alpha)$ such that (a) for $0 < |s| < \delta$, $u(s)$ is a $2\pi\rho(s)$-periodic solution of period $2\pi\rho(s)$ of (8) corresponding to $\lambda = \lambda(s)$; (b) $\rho(0) = 1$, $\lambda(0) = 0$, $u(0) = 0$, and $u(s) \neq 0$ if $s \neq 0$; and (c) any $2\pi\rho$-periodic solution of (8) in $C^0_{2\pi\rho}(\mathbf{R}, X_\alpha)$ (= the space of $2\pi\rho$-periodic continuous functions with value in X_α) with $|\rho - 1|$, $|\lambda|$ and $\|u\|$ sufficiently small is of the above form for some $|s| < \delta$ up to a translation of the real line. Moreover, if $g \in C^{k+1}(\mathcal{O}, X_\alpha)$, then the functions ρ, λ, u are of class C^k.

V. The Lyapunov-Schmidt Procedure

Suppose that $L(\lambda) = F_u(\lambda, 0)$ has at $\lambda = 0$ an n-fold eigenvalue at the origin, i.e., $\dim \operatorname{Ker} L(0) = n$, and assume that $X \subset Y$. Let $\operatorname{Ker} L(0)$ be spanned by $\varphi_1, \ldots, \varphi_n$, and let P be the projection onto this linear subspace that commutes with $L(0)$. Then P must take the form $Pu = \sum_{j=1}^n \langle u, \varphi_j^* \rangle \varphi_j$, where the $\varphi_j^* \in Y^* \subset X^*$ are null functions of the adjoint operator $L(0)^*$ and $\langle \varphi_j, \varphi_k^* \rangle = \delta_{jk}$. P is a linear operator from Y to Y; hence it can be regarded as a mapping from X to X as well, and $Q = I - P$ is a projec-

tion onto the range of $L(0)$ in Y. Using P and Q, the equation $F(\lambda, u)=0$ can be decomposed into the system of equations

$$QF(\lambda, v+\psi)=0 \quad \text{and} \quad PF(\lambda, v+\psi)=0, \tag{9}$$

where $v=Pu$ and $\psi=Qu$. One solves the first equation for $\psi=\psi(\lambda, v)$ by the implicit function theorem, and then, substituting this into the second, one has the **bifurcation equation**

$$\tilde{F}(\lambda, v)=PF(\lambda, v+\psi(\lambda, v))=0. \tag{10}$$

Solutions of the bifurcation equation are in one-to-one correspondence with solutions of the original system sufficiently close to the bifurcation point.

There exists another method of reducing an infinite-dimensional problem to a finite-dimensional one, called the **center manifold theorem** [21, 39].

W. A Global Result

The following global result is due to P. H. Rabinowitz (*J. Functional Anal.*, 7 (1971)). Assume that $F(\lambda, u)=u-\lambda Lu+H(\lambda, u)$, where L is a compact linear operator and $H: \mathbf{R}\times X\to X$ is a compact mapping with $H(\lambda, u)=o(\|u\|)$ at 0 uniformly on bounded λ-intervals. Then, if μ^{-1} is an eigenvalue of L of odd multiplicity, $(\mu, 0)$ is a bifurcation point for F with respect to the trivial solution. Moreover, the closure of the set of nontrivial zeros of F contains a component that meets $(\mu, 0)$ and either is unbounded in $\mathbf{R}\times X$ or meets $(\hat{\mu}, 0)$, where $\mu\neq\hat{\mu}$ and $\hat{\mu}^{-1}$ is an eigenvalue of L.

The beginning of bifurcation theory seems to be in the celebrated work of H. Poincaré (*Acta Math.*, 7 (1885)).

X. Abstract Cauchy Problems

Suppose that we are given an **abstract Cauchy problem**

$$\frac{du}{dt}=Au \qquad (t>0), \tag{11}$$

$$u(+0)=a \tag{12}$$

in a Banach space X, where $a\in X$ and the nonlinear operator A is assumed, for simplicity, to be independent of t. If the domain $D(A)$ of A coincides with X and A is †Lipschitz continuous, we can reduce the abstract Cauchy problem to the †integral equation of Volterra type

$$u(t)=a+\int_0^t Au(s)\,ds,$$

and by applying the iteration procedure we can easily show that the abstract Cauchy problem has a uniquely determined solution. A similar treatment works for a singular but mildly nonlinear A of the form $A=L+N$ if the linear operator L is the generator of a †strongly continuous semigroup e^{tL} in the sense of Hille-Yosida theory (→ 378 Semigroups of Operators and Evolution Equations) and the nonlinear mapping N is Lipschitz continuous. In this case, we reduce the problem to the integral equation

$$u(t)=e^{tL}a+\int_0^t e^{(t-s)L}Nu(s)\,ds.$$

Here, if we merely have to assure ourselves of the local existence of the solution, then N can merely be locally Lipschitz continuous; and in this case, N can even be singular to some extent if e^{tL} is †holomorphic in t [22, 23]. A typical application of this procedure was made by P. Sobolevskiĭ, T. Kato, and H. Fujita to the Navier-Stokes equation (→ 204 Hydrodynamical Equations; 205 Hydrodynamics) to construct regular solutions [22].

†Galerkin's method (→ 304 Numerical Solution of Partial Differential Equations) is sometimes quite convenient for obtaining (†weak) solutions of (11) and (12) [5]. Again, typical applications were made to the Navier-Stokes equations by E. Hopf and others [5, 24, 25]. To prove the convergence of approximate solutions constructed by Galerkin's method and subject to so-called energy estimates, we often make use of Aubin's compactness theorem concerning vector-valued functions [5].

Remarkable developments have taken place since 1964 for the case where A is dissipative, i.e., $-A$ is accretive. F. Browder [26] proved that if X is a Hilbert space and A is dissipative, then the mere continuity of A is sufficient for the †well-posedness of (11) and (12). Then Y. Kōmura [27] brought about a crucial advance by showing that if A is a (possibly multivalued and) maximal dissipative operator with $D(A)$ dense in the Hilbert space X, then (11) and (12) are uniquely solvable for any $a\in D(A)$. Furthermore, he founded the theory of **nonlinear semigroups of operators** by establishing a †nonlinear version of the Hille-Yosida theory for semigroups of †nonexpansive operators in Hilbert spaces [28]. Subsequent developments and applications were made in various directions by T. Kato (*J. Math. Soc. Japan*, 19 (1967)), M. Crandall and A. Pazy (*J. Functional Anal.*, 3 (1969)), H. Brézis and A. Pazy (*J. Functional Anal.*, 6 (1970)), S. Oharu (*J. Math. Soc. Japan*, 22 (1970)), M. Crandall and T. Liggett [29], Y. Konishi (*Proc. Japan Acad.*, 47 (1971)), and others. While Kōmura's original proof was based on the **Yosida approximation** $A(I-\varepsilon A)^{-1}(\varepsilon\to 0)$ for A, it is proved in [29] that in any Banach space a dissipative operator A that is maxi-

mal in a certain sense generates a nonlinear semigroup T_t by the **exponential formula**

$$T_t x = \lim_{n\to\infty}\left(I-\frac{t}{n}A\right)^{-n} x. \qquad (13)$$

The scope of applications of the generating theorem (13) can be seen, e.g., in B. K. Quinn (*Comm. Pure Appl. Math.*, 24 (1971)), M. G. Crandall (*Israel J. Math.*, 12 (1972)), Y. Konishi (*Proc. Japan Acad.*, 48 (1972) and *J. Math. Soc. Japan*, 25 (1973)), and S. Aizawa (*Hiroshima Math. J.*, 6 (1976)).

Y. Nonlinear Semigroups in Banach Lattices

Let X be a Banach lattice (→ 310 Ordered Linear Spaces). An operator $A: X\supset D(A)\to X$ is said to be **dispersive** if for all x, $y\in D(A)$,

$$\|(x-y)^+\| \leqslant \|(x-y-\lambda(Ax-Ay))^+\| \qquad (^\forall\lambda>0).$$

When a dispersive operator A satisfies the range condition $R(I-\lambda A)\supset \overline{D(A)}$ for any $\lambda>0$, it generates an **order-preserving semigroup** $T_t=e^{tA}$ on $\overline{D(A)}$: $\|(T_tx-T_ty)^+\|\leqslant\|(x-y)^+\|$ for $t>0$. We have therefore the preservation of order: $x\leqslant y$ implies $T_tx\leqslant T_ty$. We can prove in particular that the order of initial data is inherited by the solutions of a nonlinear heat equation (Y. Konishi).

Remark. Various pathological phenomena arise when we do not restrict the form of A in the abstract Cauchy problem (11), (12). We cite merely the "blowing up" of solutions of Cauchy problems for †nonlinear heat equations (→ 291 Nonlinear Problems) and the nonlinear wave propagations described by nonlinear †Schrödinger equations (J. B. Baillon, T. Cazenave, and M. Figueira).

Z. Abstract Cauchy-Kovalevskaya Theorem in a Scale of Banach Spaces

T. Yamanaka (*Comment. Math. Univ. St. Paul*, 9 (1960)), L. V. Obsyannikov (*Soviet Math. Doklady*, 6 (1965) and 12 (1971)), F. Treves (*Trans. Amer. Math. Soc.*, 150 (1970)), L. Nirenberg (*J. Differential Geometry*, 6 (1972)), and T. Nishida [30] discussed abstract treatments of classical Cauchy-Kovalevskaya theorem for partial differential equations: Let $S=\{B_\rho\}_{\rho>0}$ be a collection of Banach spaces depending on the real parameter $\rho>0$. Let $\|u\|_\rho$ denote the norm of an element $u\in B_\rho$. The collection S is called a **scale of Banach spaces** if, for any ρ and $\rho'<\rho$, $B_\rho\subset B_{\rho'}$ and $\|u\|_{\rho'}\leqslant\|u\|_\rho$ for any u in B_ρ. Consider in S the initial value problem of the form

$$\frac{du}{dt}=F(u(t),t), \quad |t|<\delta, \quad \text{and} \quad u(0)=0. \qquad (14)$$

Assume the following conditions on F: (i) For some numbers $R>0$, $\eta>0$, $\rho_0>0$ and every pair of numbers ρ, ρ' such that $0<\rho'<\rho<\rho_0$, $(u,t)\to F(u,t)$ is a continuous mapping of $\{u\in B_\rho \mid \|u\|_\rho<R\}\times\{t \mid |t|<\eta\}$ into $B_{\rho'}$. (ii) For any $\rho'<\rho<\rho_0$ and all u, $v\in B_\rho$ with $\|u\|_\rho<R$, $\|v\|_\rho<R$, and for any t, $|t|<\eta$, F satisfies $\|F(u,t)-F(v,t)\|_{\rho'}<C\|u-v\|_\rho/(\rho-\rho')$, where C is a constant independent of t, u, v, ρ, or ρ'. (iii) $F(0,t)$ is a continuous function of t, $|t|<\eta$, with values in B_ρ for every $\rho<\rho_0$ and satisfies, with a fixed constant K, $\|F(0,t)\|_\rho\leqslant K/(\rho_0-\rho)$, $0\leqslant\rho<\rho_0$.

Abstract Cauchy-Kovalevskaya theorem. Under the preceding hypotheses there is a positive constant M such that there exists a unique function $u(t)$ which for every positive $\rho<\rho_0$ and $|t|<M(\rho_0-\rho)$ is a continuously differentiable function of t with values in B_ρ, $\|u(t)\|_\rho<R$, and satisfies (14) (→ also [31]).

These results cover a theorem of M. Nagumo (*Japan. J. Math.*, 18 (1948)), which generalizes the classical Cauchy-Kovalevskaya theorem.

References

[1] L. Nirenberg, Topics in nonlinear functional analysis, Lecture notes, New York Univ., 1973–1974.

[2] J. T. Schwartz, Nonlinear functional analysis, Gordon & Breach, 1969.

[3] M. Berger, Nonlinearity and functional analysis, Academic Press, 1977.

[4] G. J. Minty, Monotone (nonlinear) operators in Hilbert space, Duke Math. J., 29 (1962), 341–346.

[5] J. L. Lions, Quelques méthodes de resolution des problèmes aux limites nonlinéaires, Dunod, 1969.

[6] J. K. Hale, Ordinary differential equations, Interscience, 1969.

[7] M. A. Krasnosel'skiĭ, Topological methods in the theory of nonlinear integral equations, Pergamon, 1964. (Original in Russian, 1956.)

[8] A. Friedman, Partial differential equations of parabolic type, Prentice-Hall, 1964.

[9] O. A. Ladyzhenskaya and N. N. Ural'tseva, Linear and quasilinear elliptic equations, Academic Press, 1968. (Original in Russian, 1956.)

[10] E. Hille and R. S. Phillips, Functional analysis and semigroups, Amer. Math. Soc. Coll. Publ., 31 (1957).

[11] H. Omori, Infinite-dimensional Lie transformation groups, Lecture notes in math. 427, Springer, 1974.

[12] J. Moser, A new technique for construction of solutions of nonlinear differential equa-

tions, Proc. Nat. Acad. Sci. US, 47 (1961), 1824–1831.
[13] H. Jacobowitz, Implicit function theorems and isometric embeddings, Ann. Math., 95 (1972), 191–225.
[14] J. Mather, Stability of C^∞ mappings I, Ann. Math., 87 (1968), 89–104.
[15] S. Smale, An infinite-dimensional version of Sard's theorem, Amer. J. Math., 87 (1965), 861–866.
[16] R. Palais and S. Smale, A generalized Morse theory, Bull. Amer. Math. Soc., 70 (1964), 165–172.
[17] C. Bessaga, Every infinite-dimensional Hilbert space is diffeomorphic with its unit sphere, Bull. Acad. Sci. Polon., 14 (1966), 27–31.
[18] R. Palais, Lusternik-Schnirelman theory on Banach manifolds, Topology, 5 (1966), 115–132.
[19] M. G. Crandall and P. H. Rabinowitz, Mathematical theory of bifurcation, Bifurcation Phenomena in Mathematical Physics and Related Topics, C. Bardos and D. Bessis (eds.), Reidel, 1980.
[20] J. E. Marsden, Qualitative methods in bifurcation theory, Bull. Amer. Math. Soc., 84 (1978), 1125–1148.
[21] D. H. Sattinger, Bifurcation and symmetry breaking in applied mathematics, Bull. Amer. Math. Soc., new ser., 3 (1980), 779–819.
[22] T. Kato, Nonlinear evolution equations in Banach spaces, Amer. Math. Soc. Proc. Symposia in Appl. Math., 17 (1965), 50–67.
[23] P. E. Sobolevskiĭ, Equations of parabolic type in a Banach space, Amer. Math. Soc. Transl., (2) 49 (1966), 1–62. (Original in Russian, 1961.)
[24] E. Hopf, Uber die Anfangswertaufgabe für die hydrodynamischen Grundgleichungen, Math. Nachr., 4 (1951), 213–231.
[25] O. A. Ladyzhenskaya, The mathematical theory of viscous incompressible flow, Gordon & Breach, 1963. (Original in Russian, 1961).
[26] F. Browder, Nonlinear equations of evolution, Ann. Math., 80 (1964), 485–523.
[27] Y. Kōmura, Nonlinear semigroups in Hilbert space, J. Math. Soc. Japan, 19 (1967), 493–507.
[28] Y. Kōmura, Differentiability of nonlinear semigroups, J. Math. Soc. Japan, 21 (1969), 375–402.
[29] M. G. Crandall and T. M. Liggett, Generation of semigroups of nonlinear transformations on general Banach spaces, Amer. J. Math., 93 (1971), 265–298.
[30] T. Nishida, A note on a theorem of Nirenberg, J. Differential Geometry, 12 (1977), 629–633.
[31] T. Kano and T. Nishida, Sur les ondes de surface de l'eau avec une justification mathématique des équations des ondes en eau peu profonde, J. Math. Kyoto Univ., 19 (1979), 335–370.
[32] Yu. G. Borisovich, V. G. Zvyagin, and Yu. I. Sapronov, Nonlinear Fredholm maps and the Leray-Schauder theory, Russian Math. Surveys, 32 (1977), 1–54. (Original in Russian, 1977.)
[33] J. Leray and J. Schauder, Topologie et équation fonctionnelles, Ann. Sci. Ecole Norm. Sup., 51 (1934), 45–78.
[34] L. Lyusternik and L. Shnirel'man, Methode topologique dans les problemes variationnels, Herman, 1934. (Original in Russian, 1930.)
[35] M. Morse, Calculus of variations in the large, Amer. Math. Soc. Colloq. Publ., 18 (1934).
[36] J. Moser, A rapidly convergent iteration method and nonlinear partial differential equations I, II, Ann. Scuola Norm. Pisa, 20 (1966), 265–315, 499–535.
[37] J. Nash, The embedding problem for Riemannian manifolds, Ann. Math., 63 (1965), 20–63.
[38] K. D. Elworthy and A. J. Tromba, Differential structures and Fredholm maps, Amer. Math. Soc. Proc. Symposia in Pure Math., 15 (1970), 45–94.
[39] J. Carr, Applications of centre manifold theory, Springer, 1981.

287 (XX.33) Nonlinear Lattice Dynamics

A. Lattice Dynamics

In order to elucidate certain characteristic features of nonlinear waves, **one-dimensional lattice** models have been studied. Around 1953, E. Fermi et al. performed computer experiments on nonlinear lattices to verify a generally accepted belief that nonlinear coupling between the †normal modes of harmonic oscillators would lead to complete energy sharing between these modes. To their surprise, their nonlinear lattices yielded very little energy sharing at all; on the contrary, the interactions resulted in the recurrence of the initial state. These results were later interpreted in terms of solitons (→ 387 Solitons), i.e., nonlinear waves that preserve identity despite mutual interaction.

The equations of motion for a uniform 1-dimensional chain of particles of mass m with

nearest-neighbor interaction can be written as

$$m\frac{d^2Q_n}{dt^2} = -\varphi'(Q_n - Q_{n-1}) + \varphi'(Q_{n+1} - Q_n)$$

$$(n = \dots, 1, 2, \dots).$$

It was shown that a lattice with a nonlinear interaction of the form

$$\varphi(r) = e^{-r} + r + \text{const.}$$

admits solutions in closed form. Later, it was shown that this lattice (the **exponential lattice** or the **Toda lattice**) is a completely **integrable system**.

It is convenient to introduce s_n, which is the generalized momentum canonically conjugate to the mutual displacement $r_n = Q_{n+1} - Q_n$. For a lattice with exponential interaction,

$$e^{-r_n} - 1 = ds_n/dt,$$

and if we introduce

$$S_n = \int^t s_n\, dt,$$

then the equations of motion can be written as

$$\log(1 + d^2S_n/dt^2) = S_{n+1} + S_{n-1} - 2S_n,$$

and the displacements are given by

$$Q_n = S_n - S_{n+1}.$$

We have a solution

$$S_n = \log\{1 + e^{2(\alpha n + \beta t + \delta)}\},$$

where α and δ are arbitrary constants and $\beta = \pm\sinh\alpha$. The associated wave

$$e^{-r_n} - 1 = \beta^2\operatorname{sech}^2(\alpha n + \beta t + \delta)$$

represents a solitary wave or soliton.

The multisoliton (N-soliton) solution is given by

$$S_n = \log\det\Psi_n,$$

where Ψ_n is an $N\times N$ matrix whose elements are

$$(\Psi_n)_{jk} = \delta_{jk} + c_j c_k \frac{(Z_j Z_k)^{n+1}}{1 - Z_j Z_k} e^{(\beta_j + \beta_k)t}$$

with

$$Z_j = \pm e^{-\alpha_j},$$

$$\beta_j = \mp\sinh\alpha_j.$$

Asymptotically the wave reduces as $t\to\mp\infty$ to an assembly of solitons

$$e^{-r_n} - 1 = \sum_{j=1}^{N} \beta_j^2\operatorname{sech}^2(\alpha_j n + \beta_j t + \delta_j^{\mp}).$$

B. Conserved Quantities

The equations of motion for a periodic exponential lattice of N particles can be written in a **Lax representation** $dL/dt = BL - LB$, where L and B are $N\times N$ matrices with the elements

$$L_{nn} = b_n, \qquad L_{n,n+1} = L_{n+1,n} = a_n,$$

$$L_{1,N} = L_{N,1} = a_N,$$

$$B_{n,n+1} = -B_{n+1,n} = -a_n,$$

$$B_{1,N} = -B_{N,1} = a_N$$

(the other elements of L and B are all zero), with

$$a_n = \tfrac{1}{2}e^{-(Q_{n+1} - Q_n)/2},$$

$$b_n = \tfrac{1}{2}P_n \qquad (P_n = dQ_n/dt).$$

The eigenvalues λ of L can be shown to be independent of time, and so the motion of the lattice is a spectrum-preserving deformation. Now, if we define $\{I_i\}$ by

$$\det(\lambda 1 - L) = \lambda^N + \lambda^{N-1}I_1 + \dots + \lambda I_{N-1} + I_N,$$

the $n\{I_l\}$ are polynomials of a_n and b_n. These are constants of motion that were discovered independently by M. Hénon and H. Flaschka. Thus the lattice has N conserved quantities; I_1 is related to the total momentum and I_2 to the total energy, but higher-index conserved quantities have no physical interpretation.

C. Method of Integration

Let L and B be the infinite matrices obtained from the foregoing ones in the limit $N\to\infty$. The eigenvalues λ of the equation

$$L\varphi = \lambda\varphi$$

are independent of time, and the time evolution of φ is given by the equation

$$d\varphi/dt = B\varphi.$$

If the motion in the lattice is restricted to a finite region, we can clearly speak of the scattering of the wave φ due to the deformation in the lattice. For a given initial motion $Q_n(0)$ and $P_n(0)$, or $L(0)$, we calculate the initial scattering data of asymptotic form $\varphi\sim z^n$ $(n\to\infty)$. The **scattering data** consist of the reflection coefficient $R(z)$, the bound state eigenvalue $\lambda_j = -(z_j + z_j^{-1})/2$ ($\lambda_j < -1$ or $\lambda_j > 1$), and the coefficient c_j of the normalized bound state eigenfunction of asymptotic form $c_j z_j^n$ for $n\to\infty$.

From the initial data and the equations of motion for $n\to+\infty$, we get the scattering data at a later time t. In effect, we construct the kernel

$$F(m) = \frac{1}{2\pi i}\oint R(z,0)e^{-(z-z^{-1})t}z^{m-1}\,dz$$
$$+ \sum_j c_j^2(0)e^{-(z_j - z_j^{-1})t}z_j^m$$

of the discrete integral equation (**Gel'fand-Levitan-Morchenko equation**)

$$\kappa(n,m)+F(n+m)+\sum_{n'=n+1}^{\infty}\kappa(n,n')F(n'+m)=0,\qquad m\geqslant n+1.$$

After solving this equation for $\kappa(n,m)$, we calculate $K(n,n)$, given by

$$\frac{1}{[K(n,n)]^2}=1+F(2n)+\sum_{n'=n+1}^{\infty}\kappa(n,n')F(n'+n).$$

Then the initial value problem is solved in the form

$$e^{-(Q_n-Q_{n-1})}=\left[\frac{K(n,n)}{K(n-1,n-1)}\right]^2.$$

The solution can be given as $dQ_n/dt=s_n-s_{n+1}$, with

$$s_n=\kappa(n-1,n).$$

The simplest case $R(z)=0$ yields the multi-soliton solution.

For the periodic case also, eigenvalues of the equations $L\varphi=\mu\varphi$ and $d\varphi/dt=B\varphi$, under suitable boundary conditions and for certain initial data, give sufficient information to construct a solution to the initial value problem. Such a method of obtaining a general solution for the periodic lattice was developed by E. Date and S. Tanaka, and independently by M. Kac and P. Van Moerbeke. Following Date and Tanaka, the solution can be written in terms of the multivariable theta function, or the Riemann theta function ϑ, as

$$P_n=\frac{d}{dt}\log\frac{\vartheta(\alpha n+\beta t+\delta)}{\vartheta(\alpha(n+1)+\beta t+\delta)}+\text{const.},$$

where α, β, and δ are certain vector constants.

There has been much activity recently toward interpreting the integrability of the Toda lattice in terms of Lie algebras (B. Kostant [9]).

References

[1] E. Fermi, J. Pasta, and S. Ulam, Studies of non-linear problems, Los Alamos Report LA-1940 (1955); Collected papers of Enrico Fermi, vol. II, Univ. of Chicago Press, 1965, 978–988.
[2] M. Toda, Vibration of a chain with nonlinear interaction, J. Phys. Soc. Japan, 22 (1967), 431–436.
[3] M. Toda, Studies of a nonlinear lattice, Phys. Rep., 18C (1975), 1–123.
[4] M. Hénon, Integrals of the Toda lattice, Phys. Rev., B9 (1974), 1921–1923.
[5] H. Flaschka, On the Toda lattice. I, Existence of integrals, Phys. Rev., B9 (1974), 1924–1925; II, Inverse-scattering solution, Prog. Theor. Phys., 51 (1974), 703–716.
[6] E. Date and S. Tanaka, Analogue of inverse scattering theory for the discrete Hill's equation and exact solutions for the periodic Toda lattice, Prog. Theor. Phys., 55 (1976), 457–465.
[7] M. Kac and P. Van Moerbeke, A complete solution of the periodic Toda problem, Proc. Nat. Acad. Sci. US, 72 (1975), 2879–2880.
[8] M. Toda, Theory of nonlinear lattices, Springer, 1981. (Original in Japanese, 1978.)
[9] B. Kostant, The solution to a generalized Toda lattice and representation theory, Advances in Math., 34 (1979), 195–338.

288 (XIII.10) Nonlinear Ordinary Differential Equations (Global Theory)

A. General Remarks

Many well-known functions (with the notable exception of the †Γ-function), such as the exponential, trigonometric, †elliptic, and †automorphic functions, satisfy ordinary differential equations of simple forms. For the purpose of finding new transcendental functions, P. Painlevé initiated the systematic study of the equation

$$F(x,y,y',\ldots,y^{(n)})=0 \tag{1}$$

in the complex domain. To investigate the solution in its whole domain of definition, he assumed that F is a polynomial in $y, y', \ldots, y^{(n)}$ whose coefficients are analytic in x. Such an equation is called an **algebraic differential equation**. If F is linear in $y^{(n)}$, then (1) is written as

$$y^{(n)}=\frac{P(x,y,y',\ldots,y^{(n-1)})}{Q(x,y,y',\ldots,y^{(n-1)})}, \tag{2}$$

where P and Q are polynomials in $y, y', \ldots, y^{(n-1)}$ with coefficients that are analytic functions of x. Equation (2) is called a **rational differential equation**.

If F is linear in $y, y', \ldots, y^{(n)}$, i.e., (1) is a linear differential equation, and if the coefficient of $y^{(n)}$ is 1, then singular points of solutions are situated at the singular points of the coefficients (→ 254 Linear Ordinary Differential Equations (Local Theory)). If F is not linear, then singular points of solutions of (1) are divided into two categories, one consisting of those points whose positions are determined by the equation itself and are independent of individual solutions, and the other consisting of those points whose positions depend on the

choice of particular solutions. In other words, the singularities of the first category appear independently of the choice of arbitrary constants involved in the general solution, while those of the second category depend on the choice of the arbitrary constants. The former are called **fixed singularities** and the latter **movable singularities**. The linear differential equation (1) has fixed singularities only, which are situated at the singularities of the coefficients. In the same way, †branch points of solutions can be classified into two kinds, **fixed branch points** and **movable branch points**.

B. Algebraic Differential Equations of the First Order

Consider the equation

$$y'=\frac{P(x,y)}{Q(x,y)}, \tag{3}$$

where P and Q are relatively prime polynomials in x, y. The fixed singular points of (3) are defined to be points ξ, ξ' with the following properties: (i) $Q(\xi,y)\equiv 0$. (ii) $Q(\xi',y)\not\equiv 0$, $P(\xi',y)=Q(\xi',y)=0$ have a root $y=\eta'$. (iii) If, in (ii), we substitute $1/z$ for y and if the same relation (ii) holds for $x=\xi$ and $z=0$, we count such a value ξ_1 as a fixed singular point. (iv) We transform equation (3) by setting $x=1/t$. If the value $t=0$ satisfies (i) or (ii) for this transformed equation, we count ∞ as a singular point ξ or ξ' accordingly. The points ξ, ξ' are, in general, †transcendental singularities of solutions, and the points ξ' cannot be †essential singularities of solutions but may be †ordinary transcendental singularities. A singular point of a solution different from ξ and ξ' is an †algebraic singular point, and for any point distinct from ξ and ξ', equation (3) admits a solution with an algebraic singularity at this point. A necessary and sufficient condition that (3) has no movable branch point is that (3) be a †Riccati equation.

Consider the algebraic differential equation of the first order

$$F(x,y,y')=0. \tag{4}$$

After defining the fixed singular points ξ and ξ' of (4), where the algebraic function of x, y defined by (4) has bad singularities, Painlevé proved that movable singularities of solutions are algebraic. Before Painlevé's work, L. Fuchs gave a necessary and sufficient condition that (4) has no movable branch points, and then H. Poincaré showed that if this condition is satisfied, then (4) is either reducible to the Riccati equation if $g=0$, integrable by the use of elliptic functions if $g=1$, or algebraically integrable if $g>1$, where, with x the independent variable, g denotes the †genus of the algebraic curve defined by (4). Painlevé found that there were gaps in the proofs of Fuchs and Poincaré and completed these by proving his theorem and the following one: Let $\varphi(x,y_0,x_0)$ be the solution of (4) satisfying the initial condition $y(x_0)=y_0$. Let $\bar{x}$, $\bar{x}_0$ be points different from ξ and ξ', and let L be a curve connecting $\bar{x}_0$ to $\bar{x}$ and not passing through any ξ or ξ'. If we denote by $\varphi_L(\bar{x},y_0,\bar{x}_0)$ the value at $x=\bar{x}$ of the branch obtained by continuing $\varphi(x,y_0,\bar{x}_0)$ analytically in a neighborhood of L and regard $\varphi_L(\bar{x},y_0,\bar{x}_0)$ as a function of y_0, then $\varphi_L(\bar{x},y_0,\bar{x}_0)$ coincides, in a neighborhood of every point $y_0=b$, with several branches of an †algebroid function of y_0. Painlevé studied the case when the general solution is finitely many-valued and gave a condition for $\varphi(x,y_0,x_0)$ to be an algebraic function of y_0.

When equation (4) does not contain x explicitly, no movable branch points appear if and only if all solutions are single-valued, and then the solutions are expressible in terms of rational, exponential, and elliptic functions. Such an equation is called a **Briot-Bouquet differential equation**.

J. Malmquist proved, by using P. Boutroux's method of studying the behavior of solutions in the neighborhood of a fixed singularity, that if equation (4) admits at least one solution that has an essential singularity and is finitely many-valued and free from movable branch points around this singularity, then (4) is an equation without movable branch points. If (4) admits a solution that is a finitely many-valued transcendental function, then an algebraic transformation may be applied to (4) so that it will become an equation without movable branch points. It is an immediate consequence of the first assertion that if (3) admits a solution that has an essential singularity and is finitely many-valued and free from movable branch points around the singularity, then (3) is a Riccati equation.

Later, equations (3) and (4) were studied by M. Hukuhara, K. Yosida, T. Satō, T. Kimura, and T. Matuda. The following results are due to Kimura. If a solution $\varphi(x)$ of (3) has an essential singularity at $x=\xi$, then, in an arbitrary neighborhood of ξ, $\varphi(x)$ assumes every value with the exception of the roots of $P(\xi,y)=0$. If (3) is not a Riccati equation, it is determined by a finite number of algebraic processes whether or not (3) admits a solution that has an essential singularity at $x=\xi$ and has no movable branch point around ξ. If (3) admits such a solution, then the singularity is a †logarithmic branch point. For an essential singularity of a solution there exists, in general, a direction similar to a †Julia's direction, which was investigated by Hukuhara and

Kimura. Matuda studied in detail the behavior of solutions as x tends to ξ along a half-line and concluded that, except for some special cases, any solution tends to a certain value as x tends to ξ along a half-line. To obtain algebraic solutions C. Briot and J. C. Bouquet devised a method similar to †Puiseux expansion in the theory of algebraic functions. Hukuhara improved their method and succeeded in reducing (3) to several differential equations of standard forms in a neighborhood of $x=\xi$. This enables us to apply the local theory to the global study. Hukuhara's method was used by Kimura and Matuda to obtain the results discussed in this paragraph.

C. Algebraic Differential Equations of the Second Order

For second-order algebraic differential equations we can pose the same problems as for first-order equations: When do these equations have single-valued or finitely many-valued general solutions? What new transcendental functions are needed to integrate such equations? These problems, studied by E. Picard and Painlevé, are difficult because of the existence of movable transcendental singularities. However, Painlevé succeeded in determining rational differential equations of the second order without movable branch points. Such equations, with the exception of those that are integrated by the use of solutions of the first-order and linear differential equations, can be transformed by rational transformations into one of the following six differential equations:

(I) $y''=6y^2+x,$

(II) $y''=2y^3+xy+\alpha,$

(III) $y''=\dfrac{y'^2}{y}-\dfrac{y'}{x}+\dfrac{\alpha y^2+\beta}{x}+\gamma y^3+\dfrac{\delta}{y},$

(IV) $y''=\dfrac{y'^2}{2y}+\dfrac{3y^3}{2}+4xy^2+2(x^2-\alpha)y+\dfrac{\beta}{y},$

(V) $$y''=y'^2\left(\frac{1}{2y}+\frac{1}{y-1}\right)-\frac{y'}{x}+\frac{(y-1)^2}{x^2}\left(\alpha y+\frac{\beta}{y}\right)+\frac{\gamma y}{x}+\frac{\delta y(y+1)}{y-1},$$

(VI) $$y''=\frac{y'^2}{2}\left(\frac{1}{y}+\frac{1}{y-1}+\frac{1}{y-x}\right)-y'\left(\frac{1}{x}+\frac{1}{x-1}+\frac{1}{y-x}\right)+\frac{y(y-1)(y-x)}{x^2(x-1)^2}\times\left\{\alpha+\frac{\beta x}{y^2}+\frac{\gamma(x-1)}{(y-1)^2}+\frac{\delta x(x-1)}{(y-x)^2}\right\},$$

where α, β, γ, and δ are constants. These equations and their solutions are called **Painlevé equations** and **Painlevé transcendental functions**, respectively. Equation (VI) was discovered by B. Gambier, who found an omission in Painlevé's calculations. All solutions of (I) are single-valued, and their properties were investigated by Boutroux. The solutions of (VI) have, in general, logarithmic branch points at $x=0$, 1, ∞ and were studied by R. Garnier.

The case when the equation is of degree 2 with respect to y'' was studied by Malmquist and F. Tricomi.

The following facts are known concerning movable transcendental singularities of the rational equation $y''=P(x,y,y')/(Q(x,y,y')$, where P and Q are relatively prime polynomials (Kimura). Let p, q be the degrees of P and Q with respect to y'. If $p>q+2$, then any solution $\varphi(x)$ admits no movable essential singularities, but its derivative $\varphi'(x)$ may admit such singularities. If $p>q+2$ and Q is not decomposable as $Q_1(x,y)Q_2(x,y,y')$, then neither φ nor φ' admits movable essential singularities. If $p\leqslant q+2$, then both φ and φ' may have movable essential singularities. If $p\leqslant q+2$ and Q is not decomposable as above, then φ' has no movable essential singularities. If, for a solution $\varphi(x)$, $x=a$ is a movable essential singularity of $\varphi(x)$ or $\varphi'(x)$, the $\varphi(x)$ or $\varphi'(x)$ assumes all values other than a finite number of exceptional values in an arbitrary neighborhood of $x=a$. If $p>q+2$, then every solution possesses †Iversen's property, and hence the set of movable singularities is not a continuum.

D. Higher-Order Equations and Other Equations

Painlevé's method of obtaining the second-order equations without movable branch points is applicable to higher-order equations. The determination of third-order equations without movable branch points was attempted by Painlevé, J. Chazy, and Garnier by the use of this method, but is not yet complete. Chazy studied in detail an equation of the form

$$y'''=\frac{(1-1/n)y''^2}{y'}+b(y)y'y''+c(y)y'^3$$

and showed that when $n=-2$ and $b(y)=0$, †Fuchsian and †Kleinian functions are obtained as solutions (→ 32 Automorphic Functions).

R. Fuchs, a son of L. Fuchs, derived equation (VI), at almost the same time as Gambier, from the study of †monodromy groups. He showed that the monodromy group of the

equation

$$\frac{1}{z}\frac{d^2z}{dt^2}=\left(\frac{\alpha}{t^2}+\frac{\beta}{(t-1)^2}+\frac{\gamma}{(t-x)^2}+\frac{\delta}{t(t-1)}+\frac{3}{4(t-y)^2}+\frac{a}{t(t-1)(t-x)}+\frac{b}{t(t-1)(t-y)}\right)$$

remains invariant as the singularity x varies if and only if α, β, γ, and δ remain constant; the singularity y, considered as a function of x, satisfies equation (VI); and a and b are rational functions of x and y and y'. Investigating a second-order linear differential equation of Fuchsian type with †regular singularities 0, 1, $\infty, x_1, \ldots, x_n, y_1, \ldots, y_n$ where $y_1, \ldots, y_n$ are apparent fixed singularities, Garnier was led, under the hypothesis that the monodromy group of this equation remains invariant as $x_1, \ldots, x_n$ vary, to a †completely integrable system of partial differential equations and showed that a symmetric function of $y_1, \ldots, y_n$, considered as a function of any one of $x_1, \ldots, x_n$, satisfies an equation without movable branch points (→ 253 Linear Ordinary Differential Equations (Global Theory)).

For nonalgebraic equations, movable branch points appear even in the first-order case. Kimura obtained a sufficient condition that for an equation $F(x,y,y')=0$, where F is a polynomial of y' with meromorphic coefficients in x, y, every solution has Iversen's property in a domain of the complex plane.

It was shown by O. Hölder that the Γ-function satisfies no algebraic differential equation. Also, if a function meromorphic in the unit circle is of †order ∞, then it satisfies no algebraic differential equation.

References

[1] P. Painlevé, Oeuvres de Paul Painlevé, 1, 2, 3, Centre Nat. Recherche Sci. France, 1972, 1974, 1975.
[2] P. Boutroux, Leçons sur les fonctions définies par les équations différentielles du premier ordre, Gauthier-Villars, 1908.
[3] E. L. Ince, Ordinary differential equations, Longmans-Green, 1927.
[4] L. Bieberbach, Theorie der gewöhnlichen Differentialgleichungen, Springer, second edition, 1965.
[5] M. Hukuhara, T. Kimura, and T. Matuda, Equations différentielles ordinaires du premier ordre dans le champ complexe, Publ. Math. Soc. Japan, 1961.
[6] J. Malmquist, Sur les fonctions à un nombre fini de branches définies par les équations différentielles du premier ordre, Acta Math., 36 (1913), 297–343; 74 (1941), 175–196.
[7] E. Hille, Ordinary differential equations in the complex domain, Wiley, 1976.

289 (XIII.9) Nonlinear Ordinary Differential Equations (Local Theory)

A. General Remarks

Consider a system of n differential equations

$$dy_j/dx=f_j(x,y_1,\ldots,y_n),\qquad j=1,\ldots,n,\tag{1}$$

where the f_j are analytic functions of $x, y_1, \ldots, y_n$. To simplify the notation, we use vector notation $\mathbf{y}$ instead of $(y_1, \ldots, y_n)$. If all the f_j are holomorphic at a point $(x,\mathbf{y})=(a,\mathbf{b})$, there exists one and only one solution $\mathbf{y}(x)$ for (1) such that $\mathbf{y}\to\mathbf{b}$ as $x\to a$ and $\mathbf{y}(x)$ is holomorphic at $x=a$. We say that a point $(a,\mathbf{b})$ is a **singular point** of the system (1) if it is a singular point of f_j for some j. The well-known †Cauchy's existence theorem can no longer be applied to the case when $(a,\mathbf{b})$ is a singular point of system (1). In this case, the following three problems arise naturally: (i) to determine whether solutions $\mathbf{y}(x)$ such that $\mathbf{y}(x)\to\mathbf{b}$ as $x\to a$ exist, and if they exist, to determine the number of independent solutions; (ii) to construct analytic expressions for solutions $\mathbf{y}(x)$ such that $\mathbf{y}(x)\to\mathbf{b}$ as $x\to a$ or, in a slightly more general way, analytic expressions for bounded solutions $\mathbf{y}(x)$ such that the values of $(x,\mathbf{y}(x))$ stay in a neighborhood of the singular point $(a,\mathbf{b})$; (iii) to investigate the properties of these solutions. These three problems are called **local problems**, since only those solutions in a neighborhood of the singular point $(a,\mathbf{b})$ are considered. However, even when $n=1$, the study of local problems is very difficult except for the case of singular points of particular types at which the functions f_j are meromorphic.

When $n>1$, the problem becomes even harder; research on this case lags that for $n=1$. In the subsequent discussion we assume without loss of generality that $a=0$, $\mathbf{b}=\mathbf{0}$.

B. The Case of a Single Equation

Consider the equation

$$dy/dx=Y(x,y)/X(x,y),\tag{2}$$

where X and Y are holomorphic functions of (x,y) at $(0,0)$ $x, y\in\mathbf{C}$. When $X(0,0)=0$ and

$Y(0,0)\neq 0$, we can rewrite equation (2) in the form $dx/dy = X(x,y)/Y(x,y)$, and we see that (i) if $X(0,y)\not\equiv 0$, equation (2) has one and only one solution $y(x)$ which is algebraic at $x=0$ and tends to 0 as $x\to 0$; (ii) if $X(0,y)\equiv 0$, there is no solution of equation (2) such that $y\to 0$ as $x\to 0$.

The case where $X(0,0)=0$ and $Y(0,0)=0$ was first studied by C. A. A. Briot and J. C. Bouquet. In order to obtain algebraic solutions they introduced a method similar to †Puiseux expansion in the theory of algebraic functions. A. R. Forsyth and J. Malmquist studied a problem of reduction for equation (2) by using the Briot-Bouquet method. The theory of reduction was completed by M. Hukuhara, who divided a neighborhood of $(0,0)$ into a finite number of subdomains in such a way that (i) the union of these subdomains covers the given neighborhood of $(0,0)$ completely; (ii) in each of these subdomains equation (2) takes one of eight canonical forms. Hukuhara investigated the properties of the solutions for these canonical forms. Among them, the following two are well known as the **Briot-Bouquet differential equations**:

$$x\,dy/dx = f(x,y), \qquad f(0,0)=0, \tag{3}$$

$$x^{\sigma+1}\,dy/dx = f(x,y),$$
$$f(0,0)=0, \qquad \sigma\geqslant 1 \text{ is an integer.} \tag{4}$$

C. Systems of Differential Equations

When y is a vector with components (y_j), equations (3) and (4) can be written as

$$x\,dy_j/dx = f_j(x, y_1, \ldots, y_n), \qquad j=1,\ldots,n, \tag{5}$$

$$x^{\sigma+1}\,dy_j/dx = f_j(x, y_1, \ldots, y_n), \qquad j=1,\ldots,n. \tag{6}$$

After Briot and Bouquet, the singular points of these types were studied by many authors, including Poincaré, E. Picard, H. Dulac, Malmquist, W. J. Trjitzinski, and Hukuhara. A method of constructing solutions of these equations consists of two parts: (i) formally transforming the given equations into simplified or reduced equations of the simplest possible form by applying a formal transformation of the type

$$y = \sum p_{kl}x^k z^l \tag{7}$$

or

$$y_j = \sum p_{k_0k_1\ldots k_n} x^{k_0} z_1^{k_1}\ldots z_n^{k_n}, \qquad j=1,2,\ldots,n; \tag{8}$$

(ii) verifying convergence or the validity of †asymptotic expansions for the **formal solutions** of the given equations, which are obtained by substituting bounded solutions of the equations satisfied by z or z_j into (7) or (8). By studying these analytic expressions, the properties of the solutions can be clarified.

D. Properties of Solutions of Briot-Bouquet Differential Equations

For equation (3), the character of the simplified equation depends on the value of $\lambda = f_y(0,0)$. We have the following four cases: (i) λ is neither 0 nor negative. A suitable formal transformation (7) changes (3) to

$$x\,dz/dx = \lambda z + bx^{\lambda}.$$

In particular, if λ is not equal to a positive integer, then $b=0$. The double power series (7) is uniformly convergent. There exists a function $\varphi(x,z)$ of (x,z) holomorphic at $(0,0)$ such that $y=\varphi(x, x^{\lambda}(b\log x + C))$ is a general solution of (3) with an integration constant C. (ii) $\lambda=0$. The equation satisfied by z takes the form

$$x\,dz/dx = z^{m+1}(b + b'z^m), \qquad m\geqslant 1.$$

If $b=0$, b' necessarily vanishes and $x=0$ is a holomorphic point of (3). A general solution is given by

$$Z(x) = (C - mb\log x)^{-1/m}, \qquad b\neq 0, \qquad b'=0,$$

and

$$Z(x) = \left(\frac{b'}{b}\mathfrak{a}\left(\frac{mb^2}{b'}\log\frac{1}{x} + C\right)\right)^{-1/m}, \qquad bb'\neq 0.$$

Here, $\zeta = \mathfrak{a}(t)$ is the branch (of the inverse function of $\zeta - \log\zeta = t$) such that $\mathfrak{a}(t) - t\log t - t \to 0$ as $t\to\infty$. Then there exists a holomorphic function $\varphi(x,z)$ of (x,z) for $|x|<\delta$, $|m\arg z + \arg b| < 3\pi/2 - \varepsilon$, $|z|<\Delta$ such that $\varphi(x, Z(x))$ is a general solution of (3). (iii) λ is a negative rational number $-\mu/\nu$. The equation satisfied by z is written as

$$\nu x\,dz/dx = z(-\mu + b(x^{\mu}z^{\nu})^m + b'(x^{\mu}z^{\nu})^{2m}).$$

A general solution has the form

$$Z(x) = x^{\lambda}(C - mb\log x)^{-1/m\nu}, \qquad b'=0,$$

and

$$Z(x) = x^{\lambda}\left(\frac{b'}{b}\mathfrak{a}\left(\frac{\nu mb^2}{b'}\log\frac{1}{x} + C\right)\right)^{-1/m\nu},$$
$$bb'\neq 0.$$

In this case, there exists a holomorphic function $\varphi(x,z)$ of (x,z) for $|m\mu\arg x + m\nu\arg z + \arg b \mp \pi/2| < \pi - \varepsilon$, $|x|<\delta$, $|z|<\Delta$ such that $\varphi(x, Z(x))$ is a general solution of (3). M. Iwano expressed this solution in the form $\psi((x^{\mu}Z(x)^{\nu})^m, x, Z(x))$, where $\psi(w,x,z)$ is holomorphic for $|\arg w \mp \frac{\pi}{2}| < \pi - \varepsilon$, $0<|w|<\delta$, $|x|<\Delta$, $|z|<\Delta$. In particular, if $b=b'=0$, $\varphi(x,z)$ is holomorphic at $(0,0)$. (iv) λ is a negative

irrational number. The equation in z has the form

$x\,dz/dx=\lambda z.$

The formal transformation (7) may either diverge or converge. Dulac proved that if (7) diverges and if there exists a solution $y(x)$ such that $y(x)\to 0$ as $x\to 0$ along a suitable path L, then $|x^\alpha y(x)^\beta \arg x|\to\infty$ and $|x^\alpha y(x)^\beta \arg y(x)|\to\infty$ as $x\to 0$, $x\in L$ for any α and β. However, the existence of such a solution is not yet verified. C. L. Siegel proved that if λ satisfies certain inequalities, the formal transformation (7) is divergent. An example such that (7) is divergent was first given by Dulac. A very simple example for such a case was given by Y. Sibuya.

In the case $f(0,0)=0$ and $\lambda=f_y(0,0)\neq 0$ for (4), the most complete result was obtained by Hukuhara, namely: (i) A suitable formal transformation (7) reduces equation (4) to

$x^{\sigma+1}dz/dx=z(\alpha_0+\alpha_1 x+\ldots+\alpha_0 x^\sigma), \qquad \alpha_0=\lambda.$

A general solution is given by $Z(x)=C\cdot x^{a_\sigma}e^{\Lambda(x)}$, where $\Lambda(x)$ is a polynomial in $1/x$ of degree σ. (ii) A general solution of (4) is expressed by a uniformly convergent power series of the form $\sum\varphi_k(x)Z(x)^k$, where the $\varphi_k(x)$ are holomorphic functions of x for a certain sectorial neighborhood of $x=0$ and have †asymptotic expansions $\sum p_{jk}x^k$ as $x\to 0$.

Assume $f_j(0,0,\ldots,0)=0$ for equations (5). Put $\lambda_{jk}=\partial f_j/\partial y_k(0,0,\ldots,0)$. Denote by $\lambda_1,\ldots,\lambda_n$ the eigenvalues of an $n\times n$ matrix with elements $\{\lambda_{jk}\}$. Then (i) if an angle ω can be chosen so that for some $m\leqslant n$, all of $|\omega|$, $|\arg\lambda_1-\omega|,\ldots,|\arg\lambda_m-\omega|$ are less than $\pi/2$, equations (5) possess solutions that are expressed as uniformly convergent $(m+1)$-tuple power series of x, $Z_1(x),\ldots,Z_m(x)$:

$\sum p_{jk_0k_1\ldots k_m}x^{k_0}Z_1^{k_1}\ldots Z_m^{k_m}.$

Here the $Z_k(x)$ are general solutions of the simplified equations and have the expression

$$Z_k(x)=x^{\lambda_k}(C_k+\text{a polynomial of } C_1,\ldots,C_{k-1},\log x), \qquad k=1,2,\ldots,m.$$

(ii) Moreover, Iwano extended the result of (i) as follows: When there exists one and only one zero among the other $n-m$ eigenvalues, equations (5) have solutions, depending on $m+1$ arbitrary constants, that are expressed as $(m+1)$-tuple uniformly convergent power series

$\sum p_{jk_0k_1\ldots k_m}(Z_{m+1}(x))x^{k_0}Z_1(x)^{k_1}\ldots Z_m(x)^{k_m}.$

Here the coefficients $p_{jk_0k_1\ldots k_m}(z_{m+1})$ are holomorphic functions of z_{m+1} in a sectorial neighborhood of $z_{m+1}=0$ and admit asymptotic expansions in powers of z_{m+1} as $z_{m+1}\to 0$. In this case, the functions $Z_1(x),\ldots,Z_{m+1}(x)$ cannot generally be integrated by quadratures, but as $Z_{m+1}(x)\to 0$, the power series expansions coincide with the expressions obtained in (i). (iii) In the case when the Jacobian matrix (λ_{jk}) is the zero matrix, Iwano constructed, under additional assumptions, a convergent analytic expression of a general solution for equations (5).

Let the right-hand side of (6) be holomorphic at $(0,0,\ldots,0)$. In 1939 Trjitzinski proved the existence of solutions that admit asymptotic expansions in powers of n arbitrary constants. In 1940 and 1941, Malmquist proved, under strong conditions, the existence of solutions that are expressed as uniformly convergent power series of $Z_\alpha(x),\ldots,Z_\beta(x)$: $\sum p_{jk_\alpha\ldots k_\beta}(x)\,Z_\alpha(x)^{k_\alpha}\ldots Z_\beta(x)^{k_\beta}$. Here the $Z_k(x)$ are polynomials of x and $\log x$ of the form

$$Z_k(x)=e^{\Lambda_k(x)}x^{\lambda_x}(C_k+\text{a polynomial of } C_{k-1},\ldots,C_\alpha,\log x), \qquad k=\alpha,\ldots,\beta,$$

where the coefficients admit asymptotic expansions in powers of x. Trjitzinski's result is contained in Malmquist's result as a special case. On the other hand, under much weaker assumptions than Malmquist's, Hukuhara solved a problem on the formal simplification of equations (6) and formal solutions. Iwano improved Hukuhara's result on formal solutions and discussed the convergence of formal solutions under weaker conditions than Malmquist's (*Ann. Mat. Pura Appl.*, 1957, 1959). The $p_{jk_\alpha\ldots k_\beta}(x)$ are holomorphic functions of x in a sectorial neighborhood of $x=0$ and admit asymptotic expansions in powers of x as $x\to 0$. The angle of the sector in which the asymptotic expansions are valid is largest for Iwano's method.

If $\sum\sigma_j<n$, equations of the form

$x^{\sigma_j}dy_j/dx=f_j(x,y_1,\ldots,y_n), \qquad j=1,\ldots,n,$

possess at least $n-\sum\sigma_j$ solutions that are holomorphic for $|x|<\delta$ (R. W. Bass, *Amer. J. Math.*, 77 (1955)). This result is analogous to those obtained by O. Perron, F. Lettenmyer, and Hukuhara and Iwano in the linear case (→ 254 Linear Ordinary Differential Equations (Local Theory)).

E. Singular Perturbations

The terms on the right-hand side of the nonlinear differential equations

$$\varepsilon^{\sigma_j}dy_j/dx=f_j(x,y_1,\ldots,y_n,\varepsilon), \qquad j=1,2,\ldots,n, \tag{9}$$

are holomorphic functions of (x,y,ε) for $|x|<a$, $\|y\|<b$, $0<|\varepsilon|<c$, $|\arg\varepsilon|<d$, and admit uniformly convergent expansions in powers of y with coefficients asymptotically developable in

powers of ε. The σ_j are nonnegative integers. W. R. Wasow, W. A. Harris, Sibuya, and Iwano and T. Saito discussed problems on constructing asymptotic or convergent expansions for bounded solutions that are dependent on several arbitrary constants.

In equations (9), the f_j and $\partial f_j/\partial y_k$ are continuous functions of (x, y, ε) for $-\infty < x < +\infty$, $\|y\| < b$, $|\varepsilon| < c$, and periodic functions of period T with respect to x. Moreover, assume that a system of degenerate algebraic equations

$$0 = f_j(x, y_1, \dots, y_n, 0), \quad j = 1, \dots, n, \tag{10}$$

has a periodic solution $y_j = p_j(x)$ of period T for $-\infty < x < +\infty$. Then if equations (9) have periodic solutions $y_j = p_j(x, \varepsilon)$ of period T such that $y_j \to p_j(x)$ as $\varepsilon \to 0$, the $p_j(x, \varepsilon)$ are called **singular perturbations** of $p_j(x)$ for equations (9). Concerning this problem, see I. M. Volk (*Prikl. Mat. Mekh. SSSR*, 10 (1946)). The work of Wasow (1950) on a single equation

$$\varepsilon^\sigma y^{(n)} = f(x, y, y', \dots, y^{(m)}, \varepsilon), \quad \sigma > 0, \quad n > m \geqslant 0, \tag{11}$$

is remarkable.

References

[1] H. Dulac, Points singuliers des équations différentielles, Mémor. Sci. Math., 61, Gauthier-Villars, 1934.
[2] E. Picard, Traité d'analyse III, Gauthier-Villars, 1896.
[3] E. L. Ince, Ordinary differential equations, Dover reprint, 1956.
[4] M. Hukuhara, T. Kimura, and T. Matuda, Equations différentielles ordinaires du premier ordre dans le champ complex, Math. Soc. Japan, 1961.
[5] E. Hille, Ordinary differential equations in the complex domain, Wiley, 1976.

290 (XIII.11) Nonlinear Oscillation

A. General Remarks

By **nonlinear oscillation** we usually mean oscillation described by periodic or †almost periodic solutions of nonlinear ordinary differential equations. The theory of nonlinear oscillation is sometimes called **nonlinear mechanics**. In connection with oscillations in †dynamical systems and electrical circuits, the theory of nonlinear oscillation has been studied intensively in the Soviet Union under the direction of N. M. Krylov and N. N. Bogolyubov since around 1930, and after World War II research in this field became active in Western countries also.

Written as first-order systems, the differential equations in this theory take one of the forms

$$d\mathbf{x}/dt = \mathbf{X}(\mathbf{x}) \tag{1}$$

or

$$d\mathbf{x}/dt = \mathbf{X}(\mathbf{x}, t), \tag{2}$$

where $\mathbf{x}$ is a vector and t is a scalar. A differential equation of the form (1) is said to be **autonomous**. In a differential equation of the form (2), $\mathbf{X}(\mathbf{x}, t)$ is usually assumed to be periodic or almost periodic in t. In the former case, the differential equation (2) is said to be **periodic**, and in the latter case, **almost periodic**. Oscillations in physical systems are described mostly by periodic or almost periodic differential equations; therefore it is a principal problem in the theory of nonlinear oscillation to find a periodic or almost periodic solution of these differential equations. However, an oscillation described by a solution of a differential equation can be actually realized only when the solution is †stable (→ 394 Stability) under a small variation of the initial value. Therefore it is important to investigate the stability of periodic or almost periodic solutions. In view of the fact that an actual phenomenon may be only approximately described by mathematical equations, sometimes it is necessary to require a certain stability of the system so that solutions of a periodic or almost periodic equation stay stable under a small variation of the equation itself. Such stability is called **structural stability**, the investigation of which is also important.

It may happen that a differential equation possesses neither a periodic solution nor an almost periodic solution, but that it has an almost periodic †integral manifold (i.e., a manifold $\mathbf{x} = \mathbf{f}(t, \theta)$ in $t\mathbf{x}$-space, where θ is a parameter, such that $\mathbf{f}(t, \theta)$ is periodic or almost periodic in t and periodic in θ) containing the †trajectory of the differential equation passing through an arbitrary point of the manifold. In this case, we can consider that a solution corresponding to a trajectory lying on the manifold describes an oscillation. Therefore it is important to find a periodic or almost periodic integral manifold of a differential equation and to investigate the stability of such a manifold.

The methods used most frequently in research on nonlinear oscillation are: (i) geometric methods, (ii) analytic methods, and (iii) numerical methods.

B. Linear Oscillations

Let S_1 and S_1^* be the spaces of all ω-periodic solutions of the ω-periodic linear system

$$d\mathbf{x}/dt = A(t)\mathbf{x}, \qquad A(t+\omega) = A(t), \tag{3}$$

and its adjoint system, respectively. Then the space P_ω of continuous ω-periodic functions $\mathbf{R}\to\mathbf{R}^n$ has direct sum decompositions $P_\omega = S_1 + S_2$ and $P_\omega = S_1^* + S_2^*$ so that $(\mathbf{x}, \mathbf{y}) = 0$ for every $\mathbf{x} \in S_1$, $\mathbf{y} \in S_2$ or $\mathbf{x} \in S_1^*$, $\mathbf{y} \in S_2^*$, where $(\mathbf{x}, \mathbf{y}) = (1/\omega)\int_0^\omega {}^t\mathbf{x}(s)\mathbf{y}(s)\,ds$. Let $\{\xi^1, \ldots, \xi^m\}$ and $\{\eta^1, \ldots, \eta^m\}$ (m may be 0) be bases of S_1 and S_1^* orthonormal with respect to $(\cdot,\cdot)$. Then for $\mathbf{p} \in P_\omega$ there is a unique ω-periodic solution $\mathbf{x}$ of

$$d\mathbf{x}/dt = A(t)\mathbf{x} + \mathbf{p}(t) \tag{4}$$

belonging to S_2 under the condition $(\eta^k, \mathbf{p}) = 0$ ($k = 1, \ldots, m$), which is represented in the form $\mathbf{x} = G[\mathbf{p}]$ by a bounded linear operator $G: P_\omega \to P_\omega$. Also, if (3) and hence its adjoint system have no nontrivial ω-periodic solution, then (4) always has a unique ω-periodic solution given by $\mathbf{x} = G[\mathbf{p}]$, and this is true even if the ω-periodicity is replaced by almost periodicity. The almost periodic system (3) is said to be **regular** if (4) has almost periodic solution for an arbitrary almost periodic function $\mathbf{p}(t)$. A necessary and sufficient condition for (3) to be regular is that (3) induce an **exponential dichotomy**, that is, the solution space S of (3) have a direct sum decomposition $S = S_- + S_+$ such that $|\mathbf{x}(t)| \leqslant M e^{-\gamma|t-s|}|\mathbf{x}(s)|$ holds for $-\infty < s \leqslant t < \infty$ if $\mathbf{x} \in S_+$ and for $-\infty < t \leqslant s < \infty$ if $\mathbf{x} \in \mathbf{S}_-$, where M and γ are positive constants. An autonomous (resp. periodic) system (3) is regular if and only if no †characteristic roots (resp. characteristic exponents) have zero real parts.

C. Geometric Methods

Geometric methods are used frequently for finding a periodic solution in an autonomous or periodic case. In the autonomous case (1), a periodic solution describes a closed orbit (t is a parameter of a curve) in $\mathbf{x}$-space, which is usually called a **phase space**. The geometric method is used to show the existence and the stability of a closed orbit by investigating geometrically the behavior of orbits in the phase space. In such an approach, the properties of †critical points and †limit sets (→ 126 Dynamical Systems) are utilized frequently. This method is effective especially for 2-dimensional cases on the basis of the †Poincaré-Bendixson theorem, and various results are given for a generalized **Liénard's** (or **Duffing's**) **differential equation** $\ddot{x} + f(x)\dot{x} + g(x) = 0$, which includes the **van der Pol differential equation** $\ddot{x} - \lambda(1 - x^2)\dot{x} + x = 0$ [1, 2].

In the periodic case (2), let $\omega (>0)$ be a period of $\mathbf{X}(\mathbf{x}, t)$ with respect to t and $\mathbf{x} = \varphi(t, \alpha)$ be a solution of (2) such that $\varphi(0, \alpha) = \alpha$. Then a periodic solution of (2) is given by $\mathbf{x} = \varphi(t, \alpha_0)$, where α_0 satisfies $\varphi(\omega, \alpha_0) = \alpha_0$. Thus the existence of a periodic solution can be shown by investigating geometrically the existence of a †fixed point of the mapping $\mathbf{x} \to \mathbf{x}' = \varphi(\omega, \mathbf{x})$ in the phase space. In such an approach, †Brouwer's fixed-point theorem is utilized frequently. In the mapping $\mathbf{x} \to \mathbf{x}'$, it may happen that there is no fixed point but that there exists an †invariant manifold. In this case, we get a periodic integral manifold.

Geometric methods give information on qualitative properties, but usually not on quantitative properties like the shape of an oscillation. Thus these methods are in general not sufficient for the analysis of the phenomena met in practice.

D. Analytic Methods

At present analytic methods are used most frequently in the study of nonlinear oscillations because, in comparison with geometric methods, they enable us to get many quantitative results in addition to the qualitative ones. However, these methods are usually efficient only for **weakly nonlinear differential equations**, that is, differential equations differing only slightly from linear differential equations (general nonlinear differential equations are called sometimes **strongly nonlinear differential equations**). In this sense, analytic methods are all †perturbation methods in the wider sense [3], and the variety of the methods lies in the form of perturbation and the method of calculation. To make use of analytic methods, we always reduce the given differential equation to a differential equation of the form

$$\dot{x} = A\mathbf{x} + \varepsilon\mathbf{X}(\mathbf{x}, t, \varepsilon). \tag{5}$$

Here ε is a parameter with small absolute value, A is a matrix of the form $A = \mathrm{diag}(O_p, B)$, where O_p is a $p \times p$ zero matrix and B is a matrix whose eigenvalues have all nonzero real parts, and $\mathbf{X}(\mathbf{x}, t, \varepsilon)$ is periodic or almost periodic in t.

(i) When $\mathbf{X}(\mathbf{x}, t, \varepsilon)$ is periodic in t, Poincaré's perturbation method is used frequently.

(ii) In practical problems, we frequently meet the case $A = 0$, that is, the case where (5) is of the form

$$\dot{\mathbf{x}} = \varepsilon\mathbf{X}(\mathbf{x}, t, \varepsilon). \tag{6}$$

In this case, the **average** $\mathbf{X}_0(\mathbf{x},\varepsilon)=\lim_{T\to 0}(1/T)\int_0^T \mathbf{X}(\mathbf{x},t,\varepsilon)\,dt$ exists. If $d\mathbf{x}/dt=\mathbf{X}_0(\mathbf{x},0)$ has a periodic solution $\xi(t)$ and the related variational linear system is regular, then (6) has an (almost) periodic integral manifold $\mathbf{x}=\mathbf{f}_\varepsilon(t,\theta)$ such that $\mathbf{f}_\varepsilon(t,\theta)\to\xi(\theta)$ uniformly as $\varepsilon\to 0$. The **method of averaging** based on this fact was devised by Bogolyubov and Mitropol'skiĭ [4] and is used frequently. The equation

$$\ddot{\mathbf{x}}+\omega^2 x=\varepsilon f(x,\dot{x},t,\varepsilon) \tag{7}$$

is one of the examples that can be reduced to an equation of the form (6). Equation (7) appears frequently in practical problems, and hence various convenient techniques are devised, such as the **method of linearization**, the **asymptotic method**, the **method of harmonic balance** [4], etc., through which we can apply the method of averaging directly to the given equation (7).

(iii) Consider an ω-periodic system

$$d\mathbf{x}/dt=A(t)\mathbf{x}+\mathbf{f}(\mathbf{x},t). \tag{8}$$

For any $\mathbf{x}(t)\in P_\omega$, $d\mathbf{x}/dt=A(t)\mathbf{x}+\mathbf{f}(\mathbf{x}(t),t)$ is of the form (4) and it has an ω-periodic solution $\sum_{k=1}^m a_k\xi^k+G[N\mathbf{x}]$, or

$$T\mathbf{x}=\sum_{k=1}^m a_k\xi^k+G\left[N\mathbf{x}-\sum_{k=1}^m(\eta^k,N\mathbf{x})\eta^k\right] \tag{9}$$

under the condition $(\eta^k,N\mathbf{x})=0$ $(k=1,\ldots,m)$ as in Section B, where $N:P_\omega\to P_\omega$ is defined by $N\mathbf{x}=\mathbf{f}(\mathbf{x}(\cdot),\cdot)$, and the ω-periodic solutions of (8) correspond to the fixed points of the transformation $T:P_\omega\to P_\omega$ induced by (9), where the a_k in (9) are given by $a_k=(\xi^k,\mathbf{x})$, as is required when $\mathbf{x}$ is a fixed point. This is the **alternative** or **bifurcation method** [6–9], which is extendable to the case of almost periodic systems under the regularity of (3) [10]. In order to obtain a fixed point of T, various kinds of fixed point theorems can be utilized.

(iv) In Section C and (iii) above, the search for a fixed point of an appropriate mapping is a principal device. However, the choice of a suitable domain for the mapping is a crucial problem, and usually an a priori bound for solutions is looked for. The concept of stability and hence Lyapunov's second method are effective in such situations [11].

E. Numerical Methods

Numerical methods are used for obtaining explicit forms of the oscillations. They are convenient in practical applications, since they can be used efficiently whether or not the nonlinearity of the system is weak. For an autonomous or periodic system, one can utilize the following methods as efficient means of computing periodic solutions: **Newton's iterative method**, the **finite-element method**, the **Lindstedt-Poincaré method**, the **method of multiple scales**, the **method of harmonic balance**, the **Galerkin method**, etc. [12–15].

For weakly nonlinear cases, analytic methods (that is, perturbation methods) are very efficient. However, when a parameter is fixed beforehand, it is not easy to know whether the conclusion obtained by the perturbation method is valid for the given value. For strongly nonlinear cases, it is very difficult to analyze the problems by analytic methods, and at present hardly any efficient methods exist. Numerical methods will therefore become exceedingly important in research on nonlinear oscillations. These are now of tolerable efficiency and reliability, due to progress in high-speed machine computation.

F. Nonstationary Oscillations

Physically speaking, an oscillation is a stationary state. However, when a system contains a parameter varying slowly with time, the oscillation also varies slowly with time (for example, when the length of a pendulum varies slowly, the amplitude of the pendulum also varies slowly). Such variation of oscillations in the course of time is represented by $d\mathbf{x}/dt=\varepsilon\mathbf{f}(\mathbf{x},t,\varepsilon t,\varepsilon)$, where $\mathbf{f}(\mathbf{x},t,s,\varepsilon)$ is (almost) periodic in t, s. Such cases provide important problems in the theory of nonlinear oscillations; one such problem has been posed by Mitropol'skiĭ [16] under the name **nonstationary oscillations** and has been investigated by means of the method of multiple scales [15].

References

[1] G. Sansone and R. Conti, Nonlinear differential equations, Pergamon, 1964. (Original in Italian, 1956.)
[2] S. Lefschetz, Differential equations: Geometric theory, Interscience, 1959.
[3] G. E. O. Giacaglia, Perturbation methods in non-linear systems, Springer, 1972.
[4] N. N. Bogolyubov and Yu. A. Mitropol'skiĭ, Asymptotic methods in the theory of nonlinear oscillations, Gordon & Breach, 1961. (Original in Russian, 1958.)
[5] N. Minorsky, Nonlinear oscillations, Van Nostrand, 1962.
[6] J. K. Hale, Ordinary differential equations, Wiley, 1969.
[7] M. A. Krasnosel'skiĭ, Translation along trajectories of differential equations, Amer. Math. Soc., 1968. (Original in Russian, 1966.)

[8] L. Cesari, R. Kannan, and J. D. Schuur, Nonlinear functional analysis and differential equations, Dekker, 1976.
[9] N. Rouche and J. Mawhin, Equations differentielles ordinaires II, Masson, 1973.
[10] M. A. Krasnosel'skiĭ, V. Sh. Burd, and Yu. S. Kolesov, Nonlinear almost periodic oscillations, Wiley, 1973. (Original in Russian, 1970.)
[11] T. Yoshizawa, Stability theory and the existence of periodic solutions and almost periodic solutions, Springer, 1975.
[12] C. Hayashi, Nonlinear oscillations in physical systems, McGraw-Hill, 1964.
[13] A. Andronov, A. Vitt, and S. Khaikin, Theory of oscillations, Addison-Wesley, 1966. (Original in Russian, 1959.)
[14] M. Urabe, Nonlinear autonomous oscillations, Academic Press, 1967.
[15] A. H. Nayfeh and D. T. Mook, Nonlinear oscillations, Wiley, 1979.
[16] Yu. A. Mitropol'skiĭ, Problems of the asymptotic theory of nonstationary vibrations, Daniel, 1965. (Original in Russian, 1964.)

291 (XIII.12) Nonlinear Problems

A. General Remarks

Nonlinear problems deal with nonlinear mappings or operators and the related equations. Until recently it was customary to consider nonlinear problems as belonging to applied mathematics and the physical sciences. However, nonlinear problems now belong to modern mathematics. Many phenomena in mathematical physics are essentially described by nonlinear equations, e.g., the motions of several particles or of viscous or compressible fluids (→ 420 Three-Body Problem, 204 Hydrodynamical Equations). Some of these equations are approximated by linear equations only when the variables appearing in the equations are restricted to very small domains; they are treated by perturbation methods when the variables stay in comparatively small domains. If these requirements cannot be met and we have to deal with equations in which the variation of the variables are not negligible, nonlinear problems certainly arise.

The methods of solution of nonlinear problems are not as powerful or general as those for linear differential equations. For instance, the †principle of superposition of solutions does not hold for nonlinear problems, and therefore Fourier methods are no longer applicable. Indeed, some nonlinear problems can be dealt with only by means of very particular or ad hoc devices.

B. Methods Used in Nonlinear Problems

Consider a nonlinear equation

$$G(x)=0, \qquad (1)$$

where G is a nonlinear mapping or operator of a subset S of a linear space X into itself. If we put $G=I-F$ (I is the identity), equation (1) becomes

$$x=F(x). \qquad (2)$$

Then a solution of (2) is a fixed point of F. Therefore fixed-point theorems of various kinds are useful for solving (2) (→ 286 Nonlinear Functional Analysis).

Let $x^{(0)}\in S$, and suppose that we can define $x^{(k)}$, $k=1,2,\ldots$, by

$$x^{(k)}=F(x^{(k-1)}), \qquad k=1,2,\ldots.$$

If F is continuous and the sequence $x^{(k)}$ converges to a point $x\in S$, then x is a fixed point of (2). Such a method of constructing an approximate sequence by iteration is called the **iterative method**. †Newton's iterative process, given by

$$x^{(k)}=x^{(k-1)}-[G'(x^{(k-1)})]^{-1}G(x^{(k-1)}),$$

is one such method, where G' denotes the †Fréchet derivative of G.

If X is an infinite-dimensional space, many concepts and methods of the theory of functional analysis can be used for a number of nonlinear problems (→ 286 Nonlinear Functional Analysis).

We note that there exist nonlinear transformations that change nonlinear equations into linear ones. For example, the †hodograph method, which is often applied in hydrodynamics, consists of reducing a system of †quasilinear partial differential equations of the form

$$A_i(u,v)u_x+B_i(u,v)u_y+C_i(u,v)v_x+D_i(u,v)v_y=0 \qquad (i=1,2)$$

to linear differential equations

$$A_iy_v-B_ix_v-C_iy_u+D_ix_u=0 \qquad (i=1,2)$$

by means of the †hodograph transformation, which changes the independent variables from x, y to u, v (→ 205 Hydrodynamics).

C. Nonlinear Algebraic and Transcendental Equations

Consider a system of equations

$$f_i(x_1,\ldots,x_n)=0 \qquad (i=1,\ldots,n). \qquad (3)$$

Newton's iterative process can be applied as follows. Starting with a point $x^{(0)}=(x_1^{(0)},\ldots,x_n^{(0)})$ lying near the desired solution, we define $x^{(k)}=(x_1^{(k)},\ldots,x_n^{(k)})$, $k=1,2,\ldots$, by solving the system of equations

$$\sum_{i=1}^{n}\frac{\partial f_j}{\partial x_i}(x^{(k-1)})(x_j^{(k)}-x_j^{(k-1)})+f_j(x^{(k-1)})=0 \qquad (j=1,\ldots,n).$$

Under some conditions the iteration $\{x^{(k)}\}$ converges to the solution (→ 301 Numerical Solution of Algebraic Equations).

If the system (3) is a real one, (3) is equivalent to $\sum f_i^2=0$. It is clear that a method of obtaining a minimum of a function $f(x_1,\ldots,x_n)$ is applicable to solving $\sum f_i^2=0$. Taking a point $x^{(0)}$, we define

$$x^k=x^{(k-1)}-\lambda_{k-1}\nabla f(x^{(k-1)}), \qquad k=1,2,\ldots,$$

where $\nabla f=(\partial f/\partial x_1,\ldots,\partial f/\partial x_n)$, and let λ_{k-1} satisfy

$$f(x^{(k-1)}-\lambda_{k-1}\nabla f(x^{(k-1)}))\leqslant f(x^{(k-1)}-\lambda\nabla f(x^{(k-1)})) \qquad (\lambda\geqslant 0).$$

Under suitable assumptions, a subsequence $x^{(k_j)}$ converges to x such that $\nabla f(x)=0$ and $f(x^{(k_j)})$ decreases monotonically to $f(x)$ [1].

Let f be a continuous mapping from a domain D of $\mathbf{R}^n$ or $\mathbf{C}^n$ into itself. Then for any $x^{(0)}\in D$, the iteration $x^{(k)}$ can always be defined by $x^{(k)}=f(x^{(k-1)})$. The problem of the behavior of $x^{(k)}$ is an interesting and important one in pure mathematics; recent work in the physical sciences is yielding many new concepts related to this problem (→ 126 Dynamical Systems, 433 Turbulence and Chaos).

D. Nonlinear Differential Equations

Consider a nonlinear system of ordinary differential equations

$$dx/dt=f(t,x) \qquad (x\in\mathbf{R}^n).$$

The initial value problem with initial condition $x(\tau)=\xi$ is equivalent to the problem of solving the nonlinear integral equation

$$x(t)=\xi+\int_\tau^t f(s,x(s))\,ds. \qquad (4)$$

The solution of (4) is a fixed point of the operator $T:\varphi(t)\mapsto\xi+\int_\tau^t f(s,\varphi(s))\,ds$ defined for a suitable function space. Therefore fixed-point theorems are applicable to T, and the iterative method for T is called the †method of successive approximation. Also the method of the †Cauchy polygon is useful (→ 316 Ordinary Differential Equations (Initial Value Problems)). These ideas are extended to an abstract Cauchy problem,

$$du/dt=A(t)u \quad (t>0), \qquad u(+0)=a,$$

in a Banach space X, where $a\in X$ and $A(t)$ is a nonlinear operator (→ 286 Nonlinear Functional Analysis).

For extensive studies of nonlinear ordinary and partial differential equations → 314 Ordinary Differential Equations (Asymptotic Behavior of Solutions), 290 Nonlinear Oscillation, 394 Stability, 321 Partial Differential Equations (Initial Value Problems), 323 Partial Differential Equations of Elliptic Type, 325 Partial Differential Equations of Hyperbolic Type.

Nonlinear differential equations of special types appear in many fields of pure and applied mathematics, e.g., the Monge-Ampère equation and the equation for minimal surfaces in differential geometry (→ 183 Global Analysis, 275 Minimal Submanifolds), and the Toda lattice equation and the Korteweg–de Vries equation in mathematical physics (→ 287 Nonlinear Lattice Dynamics, 387 Solitons).

E. Nonlinear Problems of Control Systems

The basic equation for a †control system in which the state of the controlled object can be represented by an n-vector x is given by the following system of differential equations [3]:

$$\dot{x}=Ax-\varphi(\sigma)b, \qquad \dot{\xi}=\varphi(\sigma), \qquad \sigma=c'x-\gamma\xi, \qquad (5)$$

where $\dot{x}$ and $\dot{\xi}$ stand for dx/dt and $d\xi/dt$, respectively, ξ is a scalar function representing the control, b, c are constant n-vectors, γ is a constant number, c' is the transpose of c, and A is a constant $n\times n$ matrix whose characteristic roots have negative real parts. Furthermore, we assume that when the control has no effect upon the system, x is determined by $\dot{x}=Ax$. The quantities under consideration are all real. Finally, the function $\varphi=\varphi(\sigma)$ is a scalar function characteristic of the control mechanism. Generally, φ is nonlinear in σ, and hence equation (5) is nonlinear. Normally we assume that φ has the following properties: (i) $\varphi(\sigma)$ is a real-valued continuous function on $(-\infty,\infty)$ with $\varphi(0)=0$ and $\sigma\varphi(\sigma)>0$ for $\sigma\neq 0$; (ii) $\int_0^{\pm\infty}\varphi(\sigma)\,d\sigma=+\infty$. We say that (5) is **absolutely stable** if

$$x(t)\to 0, \qquad \xi(t)\to 0 \qquad \text{as} \qquad t\to+\infty$$

for any choice of φ subject to (i) and (ii) and for every solution of (5). In the study of control systems an important problem is to obtain a necessary and sufficient condition for the system to be absolutely stable. In this connection, we have the following result, due to M. V.

Popov: (5) is absolutely stable if there exists a nonnegative q such that

$$\mathrm{Re}\{(1+i\omega q)(c'(i\omega I-A)^{-1}b)\}+q\gamma\geqslant 0 \qquad (6)$$

for any real ω, where I is the $n\times n$ identity matrix. Conversely, if the absolute stability of (5) is given by means of the †Lyapunov function (→ 394 Stability)

$$V(x,\sigma)=x'Bx+\alpha\sigma^2+\alpha f'x+\beta\int_0^\sigma \varphi(\sigma)\,d\sigma,$$

where B is a constant matrix and f is a constant vector, then there exists a nonnegative q for which (6) is satisfied for all real ω.

F. Nonlinear Equations in Applied Mathematics

Some examples of nonlinear problems are given here (→ also 205 Hydrodynamics; 318 Oscillations).

(1) The nonlinear †differential-difference equation

$$du(t)/dt=(a-u(t-1))u(t)$$

is called the **Cherwell-Wright differential equation** [4]. Given the initial condition $u(t)=g(t)$ ($0\leqslant t\leqslant 1$, $g(t)$ is a given continuous function), its solution is uniquely determined for $0\leqslant t<\infty$. If $a\leqslant 0$ and $g(1)>0$, then $u(t)\to 0$ as $t\to\infty$; if $a=0$ and $g(1)<0$, then $u(t)\to-\infty$ as $t\to\infty$. For $a>0$, $u(t)\to-\infty$ as $t\to\infty$ if $g(1)<0$, while $u(t)$ either approaches a monotonically ($0<a\leqslant 1/e$) or oscillates (boundedly) around a ($a>1/e$) if $g(1)>0$. In particular, we have damping oscillations for $a\leqslant 3/2$, while oscillatory solutions without damping appear for $a>\pi/2$.

(2) If we regard a star as a gas sphere and assume the polytropic relation $p=K\rho^\gamma$ (K and γ are constants) between the pressure p and the density ρ at each point inside the star, then we have a differential equation of the second order that determines the density distribution. This equation constitutes the basis of the classical theory of the internal structure of the stars and is called **Emden's differential equation** or the **polytropic differential equation**. It reads:

$$(1/\xi^2)d(\xi^2\,d\theta/d\xi)/d\xi=-\theta^n,$$

where $\gamma=1+1/n$, $\rho=\lambda\theta^n$, $r=\alpha\xi$, $\alpha=((n+1)K\lambda^{\gamma-2}/(4\pi G))^{1/2}$, with r the distance from the center, G the universal gravitational constant, and λ an arbitrary constant. The solution that satisfies the conditions $\theta=1$ and $d\theta/d\xi=1$ at $\xi=0$ is called the **Lane-Emden function** of index n. Emden's equation is invariant under the transformation $\xi\to A\xi$, $\theta\to A^{-2/(n-1)}\theta$ (with A an arbitrary constant). Although Emden's equation can be reduced to an equation of the first order, it has not been possible to solve it analytically except for the cases $n=0$, 1, and 5. For certain values of n between 0.5 and 6, Emden gave numerical solutions, which were refined later by Green, D. H. Sadler, and D. C. Miller [5].

(3) **Caianiello's differential equations**, which describe the state of a network of neurons [6], are

$$x_i(t+\tau)=Y\left[\sum_j\sum_r a_{ij}^{(r)}x_j(t-r\tau)-\theta_i\right],$$

where the function $x_i(t)$ represents the state of the ith neuron at the time t and takes only the values 0 and 1. The state of the system is to be considered at the discrete times $t=0, \tau, 2\tau, \ldots$. The (real) coefficient $a_{ij}^{(r)}$ represents the weight of the effect of hysteresis in the relay process from the cell j to the cell i. The nonnegative integer θ_i is the threshold value of the cell i. $Y[x]$ is the unit step function that is equal to 1 for $x\geqslant 0$ and vanishes for $x<0$.

(4) The following **Hodgkin-Huxley differential equation** arises in the study of conduction and excitation in nerve systems [7]:

$$\frac{\partial^2 V}{\partial x^2}=\frac{2r_0}{R_0}\left(C_0\frac{\partial V}{\partial t}+g_1m^3h(V-V_1)+g_2n^4(V-V_2)+g_3(V-V_3)\right),$$

$$\frac{\partial m}{\partial t}=-(\alpha_1(V)+\beta_1(V))m+\alpha_1(V),$$

$$\frac{\partial h}{\partial t}=-(\alpha_2(V)+\beta_2(V))h+\alpha_2(V),$$

$$\frac{\partial n}{\partial t}=-(\alpha_3(V)+\beta_3(V))n+\alpha_3(V),$$

where α_i, β_i ($1\leqslant i\leqslant 3$) are given functions of V, and r_0, R_0, C_0, g_i, V_i ($1\leqslant i\leqslant 3$) are constants. The unknown function V is sought in the domain $0<x<\infty$, $0<t<\infty$, while the initial values of V, m, h, n, and the boundary value of V are given at $t=0$ and $x=0$, respectively.

References

[1] T. L. Saaty and J. Bram, Nonlinear mathematics, McGraw-Hill, 1964 (Dover, 1981).
[2] T. L. Saaty, Modern nonlinear equations, McGraw-Hill, 1967 (Dover, 1981).
[3] S. Lefschetz, Stability of nonlinear control systems, Academic Press, 1965.
[4] S. Kakutani and L. Markus, On the nonlinear difference-differential equation $y'(t)=(A-By(t-\tau))y(t)$, Contributions to the theory of nonlinear oscillations, Ann. Math. Studies, Princeton Univ. Press, 1958, vol. 4, 1–18.

[5] R. Bellman, Stability theory of differential equations, McGraw-Hill, 1953.
[6] E. R. Caianiello, Outline of a theory of thought-process and thinking machines, J. Theoret. Biol., 1 (1961), 204–235.
[7] A. L. Hodgkin and A. F. Huxley, A quantitative description of membrane current and its application to conduction and excitation in nerve, J. Physiol. (London), 117 (1952), 500–544.

292 (XIX.4) Nonlinear Programming

A. Problems

A nonlinear programming problem is a type of mathematical programming problem where it is required to minimize or maximize a nonlinear function $\theta(x)$ of n-vector variable x defined in a closed connected set X^0 with a set of linear or nonlinear constraints. Minimization or maximization of a continuously differentiable function $\theta(x)$ under the equality condition expressed by a set of continuously differentiable functions has been traditionally dealt with by the method of Lagrange multipliers. Hence a typical nonlinear programming problem is usually formulated as follows.

(NLP) Minimize $\theta(x)$ under the condition $\bar{x} \in X^0 \subset \mathbf{R}^n$ and $g_i(x) \leqslant 0$ for $i = 1, 2, \ldots, m$.

Or, equivalently, determine the set of all x such that $\theta(\bar{x}) = \min_{x \in C} \theta(x)$, where $C = \{x \mid x \in X^0 \text{ and } g(x) \leqslant 0\}$.

Here we need only consider minimization, since maximization problems can be converted to minimization problems by virtue of the obvious relation $\max \theta = -\min(-\theta)$.

The set C is known as the **feasible region** or the **constraint set**, and $\bar{x}$ is called an **optimal solution** or simply a solution. In many nonlinear programming problems X^0 is $\mathbf{R}^n$. If $X^0 = \mathbf{R}^n$ and θ and g are linear functions on $\mathbf{R}^n$, then the problem becomes a linear programming problem (→ 255 Linear Programming). The problem of minimizing a quadratic function subject to linear constraints is called the **quadratic programming problem** (→ 349 Quadratic Programming). If X^0 is a convex set, θ is convex (or concave), and the g_i are convex on X^0, then the minimization (or maximization) problem is known as a **convex** (or **concave**) **programming problem**. Convex and concave functions are important in nonlinear programming, because they admit reasonably straightforward sufficient conditions for optimality and also because they constitute the only important class of functions for which necessary optimality conditions can be given without the differentiability condition. Subject to suitable modifications, the method of Lagrange multipliers can also be applied to the solution of nonlinear programming problems.

The **Lagrangian function** ψ associated with the minimization problem (NLP) is defined by

$$\psi(x, u) = \theta(x) + u'g(x),$$

where $u = (u_1, \ldots, u_m)$ and u' denotes the vector of Lagrange multipliers.

A pair $(\bar{x}, \bar{u})$ is called a **saddle point** of $\psi(x, u)$, provided $\bar{x} \in X^0$, $\bar{u} \subset \mathbf{R}^m$, $\bar{u} \geqslant 0$, and $\psi(\bar{x}, u) \leqslant \psi(\bar{x}, \bar{u}) \leqslant \psi(x, \bar{u})$ for all $x \in X^0$ and all $u \in \mathbf{R}^m$ such that $u \geqslant 0$. It follows easily that:

(1) (H. Uzawa [18]) If $(\bar{x}, \bar{u})$ is a saddle point of $\psi(x, u)$, then $\bar{x}$ is an optimal solution.

(2) (Kuhn and Tucker [11]) Assume that X^0 is open and convex, and that θ and g are differentiable and convex. If there exists a pair $(\bar{x}, \bar{u})$ such that

$$\nabla\theta(\bar{x}) + \bar{u}'\nabla g(\bar{x}) = 0, \quad \bar{x} \in C, \quad \bar{u}'g(\bar{x}) = 0, \quad \bar{u} \geqslant 0$$

(where $\nabla\theta(x)$ and $\nabla g(x)$ denote, respectively, the gradient vector of θ at x and the Jacobian matrix of g at x), then $\bar{x}$ is an optimal solution.

B. Necessary Conditions for Optimality

In Section A sufficient conditions for optimality were given. Some necessary conditions are also known.

(3) When $\theta(x)$ is continuously differentiable and all the g_i are linear, that is, when the condition is expressed as $Ax \leqslant b$ and $x \geqslant 0$, where A is an $m \times n$ matrix and b an m-vector, $\bar{x}$ is an optimal solution only if $\bar{x}$ is an optimal solution of the following linear programming problem.

(LP^0) Minimize $v'x$ under the condition that $Ax \leqslant b$, $x \geqslant 0$, where $v = \nabla\theta(x)$.

By applying the duality theorem of linear programming, it can be proved that there exists a vector $\bar{u} \geqslant 0$ such that

$$\nabla\varphi(\bar{x}, \bar{u}) = \nabla\theta(\bar{x}) + \bar{u}'A = 0.$$

(4) (F. John [9]) Assume that X^0 is open, and that θ and g are differentiable. Then, if $\bar{x}$ is a solution of the problem (NLP), there exist $u_0 \geqslant 0$ and $\bar{u} \in \mathbf{R}^m$ such that

$$\bar{u}_0\nabla\theta(\bar{x}) + \bar{u}'\nabla g(\bar{x}) = 0,$$

$$g(\bar{x}) \leqslant 0, \quad \bar{u}'g(\bar{x}) = 0, \quad \bar{u} \geqslant 0.$$

Now we consider the following: The vector-valued function g is said to satisfy **Guignard's constraint qualification** at an inner point x of X^0 if any vector y satisfying the linear inequalities $\nabla g_i(x) \cdot y \leqslant 0$ for $i \in I = \{i \mid g_i(x) = 0\}$ is in the convex hull spanned by the vectors tangent to the set C at x.

(5) (Guignard [7]) Let X^0 be an open set, let $\bar{x}$ be an optimal solution of (NLP), let θ and g be differentiable at $\bar{x}$, and assume that g satisfies the Guignard constraint qualification. Then there exists $\bar{u} \in \mathbf{R}^m$ such that $\nabla\theta(\bar{x}) + \bar{u}'\nabla g(\bar{x}) = 0$, $g(x) \leqslant 0$, and $\bar{u}'g(\bar{x}) = 0$, $\bar{u} \geqslant 0$.

Guignard's constraint qualification is satisfied if neither of the following conditions hold: (L) Vectors $g_i(x)$ for $i \in I$ are linearly independent.

(S) (**Slater's constraint qualification**) The g_i are convex and X^0 is a convex set; and there exists a vector x such that $g_i(x) < 0$ for all i. With convexity of θ and g_i, differentiability is not required:

(6) (Kuhn and Tucker [11]) Let X^0 be a convex set, let θ and g be convex on X^0, and assume that g satisfies Slater's constraint qualification. If $\bar{x}$ is an optimal solution, then there exists $\bar{u} \in \mathbf{R}^m$, $\bar{u} \geqslant 0$, such that $\bar{u}'g(\bar{x}) = 0$ and $(\bar{x}, \bar{u})$ is a saddle point of $\psi(x, v) = \theta(x) + u'g(x)$.

C. Sensitivity Analysis

Now consider the following class of problems.

(Pc) Minimize the function $\theta(x)$ under the condition that $g(x) \leqslant c$, where c is a real m-vector. We denote by $\bar{x}_c$ the set of the solutions of the problem (Pc) and denote $\theta_c^* = \theta(\bar{x}_c)$, $\bar{x}_c \in \bar{X}_c$. Suppose that Guignard's condition is satisfied for each c and that the set of Lagrange multipliers Λ_c is nonempty. Then for any real vector a, we have

$$\inf a'u_c \leqslant \lim_{t \to 0} \frac{1}{t}(\theta_{c+t_a}^* - \theta_c^*) \leqslant \sup a'u_c, \quad u_c \in \Lambda_c.$$

Therefore, if the Lagrange multiplier is uniquely determined for some c, then u_c represents the vector of the rates of increments of the objective function to the small increments of the components of the constraints vector. Hence the components of the Lagrange-multiplier vector are called the **imputed prices** or **shadow costs** of the constraints; these have important economic implications, especially when the objective function is expressed in terms of money or profits.

More generally, we can consider the following class of problems.

(GPc): Minimize the function $\theta(x, c)$ under the condition that $g(x, c) \leqslant 0$, where c is a real parameter.

Denote by $\bar{x}(c)$ and $u(c)$ the solution and the Lagrange multiplier of (GPc) corresponding to the parameter c, respectively, and let $\theta^*(c) = \theta(\bar{x}(c))$. Then under a set of regularity conditions we have

$$\nabla\theta^*(c) = \nabla_c\theta(x(c), c) + u(c)'\nabla_c g(x(c), c),$$

where ∇_c denotes the gradient of θ or g with respect to the parameter.

D. Duality

A **duality theorem** in mathematical programming is the statement of a certain relationship between two problems. This relationship has the following two aspects: (i) one problem is a constrained minimization problem and the other is a constrained maximization problem; (ii) the existence of a solution to one of these problems ensures the existence of a solution to the other, and in this case their respective values are equal.

Let $\psi(x, u)$ be the Lagrangian form of (NLP), and define $\omega(u) = \inf_{x \in X^0} \psi(x, u)$. Then we can state the following two problems.

(P) (primary problem) Miminize $\theta(x)$ under the condition that $x \in X^0$ and $g(x) \leqslant 0$

(D) (dual problem) Maximize $\omega(u)$ under the condition $u \geqslant 0$.

If $(\bar{x}, \bar{u})$ gives a saddle point of $\psi(x, u)$, then

$$\inf\theta(x) = \theta(\bar{x}) = \sup\omega(u) = \omega(\bar{u}) = \psi(\bar{x}, \bar{u}).$$

The dual problem can be formulated alternatively as follows.

($\tilde{\mathrm{D}}$) Maximize $\psi(x, u) = \theta(x) + u'g(x)$ subject to $(x, u) \in Y = \{(x, u) \mid x \in X^0, u \in R^m, u \geqslant 0, \nabla_x\psi(x, u) = 0\}$ (where $\nabla_x\psi(x, u)$ denotes the vector whose components are the partial derivatives $\partial\psi(x, u)/\partial x_i$ for $i = 1, \ldots, n$).

There are a number of duality theorems related to problems (P) and ($\tilde{\mathrm{D}}$); two such theorems are as follows:

(1) (P. Wolfe [20]) Suppose that X^0 is open and convex, that θ and g are differentiable and convex, and that g satisfies the Kuhn-Tucker constraint qualification. Then, if $\bar{x}$ is a solution of (P), there exists a $\bar{u} \in \mathbf{R}^m$ such that $(\bar{x}, \bar{u})$ is a solution of ($\tilde{\mathrm{D}}$) and $\theta(\bar{x}) = \psi(\bar{x}, \bar{u})$.

(2) (O. L. Mangasarian and J. Ponstein [13]) Suppose that X^0 is open and convex, that θ and g are differentiable and convex, and that $(\hat{x}, \hat{u})$ is a solution of ($\tilde{\mathrm{D}}$). If $\psi(x, \hat{u})$ is strictly convex in some neighborhood of $\hat{x}$, then $\hat{x}$ is a solution of (P) and $\theta(\hat{x}) = \psi(\hat{x}, \hat{u})$.

The above two problems (P) and (D) are not symmetric. The notion of symmetric duality was introduced by G. B. Dantzig, E. Eisenberg, and R. W. Cottle:

Primary: Minimize
$F(x, u) \equiv K(x, u) - u'\nabla_u K(x, u)$,
subject to the constraints
$\nabla_u K(x, u) \leqslant 0$, $x \geqslant 0$, and $u \geqslant 0$;

Dual: Maximize
$G(x, u) \equiv K(x, u) - x'\nabla_x K(x, u)$,
subject to the constraints
$\nabla_x K(x, u) \geqslant 0$, $x \geqslant 0$, and $u \geqslant 0$,

where K is continuously differentiable in $(x,u)\in\mathbf{R}^n\times\mathbf{R}^m$.

(3) Dantzig et al. proved [6] the existence of a common optimal solution $(\bar{x},\bar{u})$ to both the primary and dual problems, provided (i) an optimal solution $(\bar{x},\bar{u})$ to the primary problem exists, (ii) K is convex in x for each u and concave in u for each x, and (iii) K is twice differentiable and the matrix of second partials $(\partial^2K/\partial u^i\partial u^j)$ is negative definite at $(\bar{x},\bar{u})$.

Rockafellar [15] gave another expression of the duality relation: Define $F(x,y)=\theta(x)$ if $g(x)\leqslant y$ and $=\infty$ otherwise, and denote $\varphi(y)=\inf_{x\in X^0}F(x,y)$. Then $\theta(x)=\varphi(0)$. For any nonlinear function $\varphi(y)$, the conjugate $\varphi^*(\eta)$ is defined by

$$\varphi^*(\eta)=\sup_y(\eta'y-\varphi(y)).$$

(4) (Rockafellar [15]) $\sup\omega(u)=\varphi^{**}(0)=\operatorname{clco}\varphi(0)$, where $\operatorname{clco}\varphi(y)$ denotes the closed convex hull or maximum convex minorant of $\varphi(y)$ defined by $\operatorname{clco}\varphi(y)=\sup_l 1'y\{1'z\leqslant\varphi(z)$ for all $z\}$. It follows from the above that $\inf\theta(x)=\sup\omega(u)$ if and only if $\varphi(0)=\operatorname{clco}\varphi(0)$, which holds true if $\varphi(y)$ is convex.

Further forms of the duality theorem hold for linear or quadratic programming problems (→ 255 Linear Programming, 349 Quadratic Programming).

E. Algorithms

In a limited class of problems, i.e., when the objective function is quadratic and the constraints are linear (→ 349 Quadratic Programming) the optimal solution can be obtained by solving a system of linear equations by the simplex method or other algorithms; but in most nonlinear programming the solution is calculated by some kind of iterative procedure. Note that even when the constraints are given as equalities and all the functions involved are continuously differentiable, so that the optimal solution is explicitly given as a solution of a set of simultaneous equations, we usually require some iteration procedure, such as the Newton-Raphson algorithm, to obtain numerically the solution with preassigned accuracy; and the iteration is not always easy if the functions are sufficiently complex.

Several iterative procedures for solving nonlinear programming problems have been proposed. Since the simplex method is a powerful tool in linear programming, one type of approach is to obtain an approximately optimal solution by approximating the objective and the constraint functions by piecewise linear functions and then applying linear programming techniques to get the approximately optimal solution within each region. Another is to change the constrained problem to a nonconstrained one by introducing a sufficiently large number M, called the penalty, and then maximizing

$$\theta_M(x)=\theta(x)+M\sum_j\max(g_i(x),0)$$

without the constraint. This is called the **penalty method**.

The third and most generally applicable technique is the **gradient method**, of which several variations are known:

(i) **Arrow-Hurwicz-Uzawa gradient method** [4]. Concave or convex programming problems can be solved by finding a saddle point of the Lagrangian function $\psi(x,u)$. Let $\varphi(x,u)$ be strictly concave and of class C^2 in n-vector $x\geqslant0$ and convex and of class C^2 in m-vector $u\geqslant0$ and possess a saddle point $(\bar{x},\bar{u})$. To approach a saddle point of $\varphi(x,u)$ it is natural to devise a gradient process of the form

$$\frac{dx_i}{dt}=\frac{\partial\varphi}{\partial x_i},\qquad\frac{du_j}{dt}=-\frac{\partial\varphi}{\partial u_j}.\tag{1}$$

To keep the variables in the positive orthant, we need to modify (1), and we consider the following system of differential equations:

$$\frac{dx_i}{dt}=\begin{cases}0 & \text{if } x_i=0 \text{ and } \dfrac{\partial\varphi}{\partial x_i}<0\\ \dfrac{\partial\varphi}{\partial x_i} & \text{otherwise}\end{cases}\qquad(i=1,\ldots,n),$$

$$\frac{du_j}{dt}=\begin{cases}0 & \text{if } u_j=0 \text{ and } \dfrac{\partial\varphi}{\partial u_j}>0\\ -\dfrac{\partial\varphi}{\partial u_j} & \text{otherwise}\end{cases}\qquad(j=1,\ldots,m).$$

Under certain regularity hypotheses, there exists a unique solution $(x(t),u(t))$ of the system with any initial point (x^0,u^0), and the x component $x(t)$ of the solution converges to $\bar{x}$ as $t\to\infty$.

Applying the above results to the Lagrangian function $\psi(x,u)$ we can solve the concave or convex programming problem.

(ii) **Rosen's gradient projection method** [16]. If a point x^0 of the feasible region does not give a solution for the minimization problem (P), then we look for a feasible point with a lower function value by proceeding from x^0 in the direction of the gradient of the function $-\theta(x)$. The method fails if x^0 is a boundary point and if the gradient vector points toward the exterior of the feasible region. Rosen's method [16] is to project the gradient onto the boundary of the feasible region and then proceed in the direction of this projection. In this manner, we remain on the boundary of the feasible region.

(iii) **Methods of feasible directions**. These were first described by G. Zoutendijk [21].

Consider the problem of minimizing $\theta(x)$ subject to the constraint $x \in S \subset \mathbf{R}^n$, where S is a closed, connected set satisfying certain regularity conditions and θ is a continuously differentiable function of the n-vector x, such that, for some α, the set $\{x \in S \mid \theta(x) \leqslant \alpha\}$ is bounded and nonempty.

A method of feasible directions is any recipe for solving this problem by proceeding along the following lines: (1) Start with some $x^0 \in S$ such that $\theta(x^0) < \sup \alpha$. (2) Pass from the kth iteration point x^k to x^{k+1} by first determining a direction s^k in x^k such that the ray $x^k + \lambda s^k$ lies in S for all sufficiently small $\lambda > 0$. (3) Then determine the step length λ_k, thus obtaining the $(k+1)$st iterate $x^{k+1} = x^k + \lambda_k s^k$. (4) Repeat this procedure until some prescribed stopping condition is satisfied.

There are many methods of determining the s^k, and in most cases the λ_k are then determined by solving a one-dimensional minimum problem along the direction so obtained. Zoutendijk has unified the various possible methods of feasible directions and the relevant normalization rules that yield the optimal directions s^k, and has investigated this subject in detail from the viewpoint of computational technique [21, 22].

F. Generalizations

The extension of the Kuhn-Tucker theory to linear topological spaces is due to L. Hurwicz [8]. Let $\mathscr{X}$ be a linear space, $\mathscr{Y}$, $\mathscr{Z}$ be linear topological spaces, P_Y, P_Z the nonnegativity cones of $\mathscr{Y}$, $\mathscr{Z}$, respectively, which are closed convex cones containing inner points, D a convex set in $\mathscr{X}$, and F, G concave mappings (→ 88 Convex Analysis A) from D into $\mathscr{Y}$, $\mathscr{Z}$, respectively, such that $G(D)$ contains an inner point of P_Z. If $F(X)$ attains its maximal point when $X = X_0$ and X_0 satisfies $G(X_0) \geqslant 0$, $X_0 \in D$, then there exist $Y_0^* \geqslant 0$, $Z_0^* \geqslant 0$ such that $\Phi(X, Z^*) = Y_0^*(F(X)) + Z_0^*(G(X))$ has a saddle point at (X_0, Z_0^*). The condition that P_Y and P_Z have inner points can be weakened to cover the cases of (l_p), (L_p), (s), (S) (Hurwicz and Uzawa). P. P. Varaiya [19] considered the following nonlinear programming problem in Banach space.

(B) Maximize $f(x)$ subject to $x \in A$, $g(x) \in A_Y$, where X, Y are real Banach spaces, $x \in X$, $g: X \to Y$ is a Fréchet differentiable mapping, f is a real-valued differentiable function, A is a subset of X, and A_Y is a convex set in Y.

The main results are similar to the Kuhn-Tucker necessary conditions. Varaiya also exhibited a saddle value problem related to (B), when A_Y is a closed convex cone. L. W. Neustadt [14] investigated nonlinear programming problems in linear vector spaces and gave an application to the theory of optimal control. The main results are: (i) Kuhn-Tucker type conditions which are both necessary and sufficient for optimality, (ii) a duality theory for obtaining multipliers in the generalized Kuhn-Tucker conditions, and (iii) an application to optimal control theory.

References

[1] J. Abadie (ed.), Nonlinear programming, North-Holland, 1967.
[2] J. Abadie (ed.), Integer and nonlinear programming, North-Holland, 1970.
[3] K. J. Arrow, L. Hurwicz, and H. Uzawa (eds.), Studies in linear and nonlinear programming, Stanford Univ. Press, 1958.
[4] K. J. Arrow and L. Hurwicz, Gradient method for concave programming I, in [3, pp. 117–126]; H. Uzawa, Gradient method for concave programming II, in [3, pp. 127–132].
[5] E. M. Beale, On minimizing a convex function subject to linear inequalities, J. Roy. Statist. Soc., (B) 17 (1955), 173–184.
[6] G. B. Dantzig, E. Eisenberg, and R. W. Cottle, Symmetric dual nonlinear programs, Pacific J. Math., 15 (1965), 809–812.
[7] M. Guignard, Generalized Kuhn-Tucker conditions for mathematical programming problems in a Banach space, SIAM J. Control, 7 (1969), 232–241.
[8] L. Hurwicz, Programming in linear spaces, in [3, pp. 38–102].
[9] F. John, Extremum problems with inequalities with subsidiary conditions, Studies and Essays: Courant Anniversay Volume, Interscience, 1948, 187–204.
[10] H. P. Künzi and W. Krelle, Nichtlineare Programmierung, Springer, 1962.
[11] H. W. Kuhn and A. W. Tucker, Nonlinear programming, Proc. 2nd Berkeley Symp. Math. Statist. Prob., 1951, 481–492.
[12] O. L. Mangasarian, Nonlinear programming, McGraw-Hill, 1969.
[13] O. L. Mangasarian and J. Ponstein, Minimax and duality in nonlinear programming, J. Math. Anal. Appl., 11 (1965), 504–518.
[14] L. W. Neustadt, Sufficiency conditions and a duality theory for mathematical programming problems in arbitrary linear spaces, in [17, pp. 323–348].
[15] R. T. Rockafellar, Augmented Lagrange multiplier functions and duality in nonconvex programming, SIAM J. Control, 12 (1974), 293–322.
[16] J. B. Rosen, The gradient projection method for nonlinear programming, J. Soc.

Indust. Appl. Math., 8 (1960), 181–217; 9 (1961), 514–553.
[17] J. B. Rosen, O. L. Mangasarian, and K. Ritter (eds.), Nonlinear programming, Academic Press, 1970.
[18] H. Uzawa, The Kuhn-Tucker theorem in concave programming, in [3, pp. 32–37].
[19] P. P. Varaiya, Nonlinear programming in Banach space, SIAM J. Appl. Math., 15 (1967), 284–293.
[20] P. Wolfe, A duality theorem tor nonlinear programming, Quart. Appl. Math., 19 (1961), 239–244.
[21] G. Zoutendijk, Methods of feasible directions, Elsevier, 1960.
[22] G. Zoutendijk, Nonlinear programming, computational methods, in [2, pp. 37–86].
[23] G. Zoutendijk, Some algorithms based on the principle of feasible directions, in [17, pp. 93–121].

293 (I.7) Nonstandard Analysis

A. General Remarks

Nonstandard analysis is a new field of research that has branched off from model theory and that provides a powerful method applicable in almost all fields of mathematical science.

About 1960, A. Robinson successfully used a nonstandard model of the real number field **R** to justify Leibniz-type infinitesimal calculus. After this, using higher-order logic, he developed a stronger theory of nonstandard analysis, and applied it to other mathematical fields [1].

In this article, we adopt a first-order logic over a universe. In the final section we present a theory called nonstandard set theory; this is a conservative extension of Zermelo-Fraenkel set theory with the axiom of choice.

B. Axioms for Nonstandard Analysis

A nonempty set U is called a universe if the following four conditions are satisfied:

(a) $x \in U, y \in x \Rightarrow y \in U$;

(b) $x, y \in U \Rightarrow \{x, y\} \in U$;

(c) $x \in U \Rightarrow \bigcup x \in U$;

(d) $x \in U \Rightarrow \mathrm{P}(x) \in U$.

In what follows, U will be a universe containing the field **R** of real numbers.

We construct a language $\mathscr{L}$ describing mathematics in U. The alphabet consists of the following symbols: (1) countably many variables. (2) constants corresponding to all elements in U. (3) predicate symbols $=$ and $\in$. (4) logical symbols $\neg$, $\wedge$, $\vee$, $\rightarrow$, $\exists$, $\forall$. (5) auxiliary symbols [,]. The last two symbols will be omitted where there is no danger of confusion.

Definition of formulas. (1) If t and s are terms (variables or constants), then $t = s$ and $t \in s$ are formulas (atomic formulas). (2) If ϕ and ψ are formulas, then $\neg\phi$, $\phi\wedge\psi$, $\phi\vee\psi$, $\phi\rightarrow\psi$ are formulas. (3) If ϕ is a formula and x is a variable, then $\exists x[\phi]$ and $\forall x[\phi]$ are formulas. (4) The formulas are those that can be constructed by the above procedure.

A formulas ϕ containing at most n free variables is called an n-ary formula. In this case, we often write $\phi(x_1, \ldots, x_n)$ for ϕ. A sentence is a 0-ary formula. We write $\models\phi$ if a sentence ϕ is true under the usual interpretation.

Consider a quadruple $(U, {}^*U, {}^*\!\in, *)$ where ${}^*U$ is a set, ${}^*\!\in$ is a binary relation on ${}^*U$, and $*$ is a mapping: $a \mapsto {}^*a$ from U to ${}^*U$. An element α of ${}^*U$ is called **standard** if $\alpha = {}^*a$ for some $a \in U$, and **nonstandard** if not. Adjoin to $\mathscr{L}$ constants representing all nonstandard elements of ${}^*U$. Then we have a language ${}^*\mathscr{L}$ for ${}^*U$. A sentence in $\mathscr{L}$ is a sentence in ${}^*\mathscr{L}$. Taking ${}^*U$ as the scope of quantifiers, we can interpret every ${}^*\mathscr{L}$-sentence in ${}^*U$. We write ${}^*\!\models\phi$ if an ${}^*\mathscr{L}$-sentence ϕ is true under this interpretation.

Axiom 1 (transfer): For every sentence ϕ in $\mathscr{L}$, we have

$$\models\phi \Leftrightarrow {}^*\!\models\phi.$$

Definition: Let $\phi(x, y)$ be a binary formula in $\mathscr{L}$ (resp. in ${}^*\mathscr{L}$). We say that $\phi(x, y)$ is **concurrent** in U (resp. in ${}^*U$), if, for every finite number of elements $a_1, \ldots, a_n$ in U (resp. in ${}^*U$), there exists an element b in U (resp. in ${}^*U$) such that $\models\phi(a_i, b)$ (resp. ${}^*\!\models\phi(a_i, b)$) for $1 \leqslant i \leqslant n$.

Axiom 2 (enlargement): If a binary formula $\phi(x, y)$ in $\mathscr{L}$ is concurrent in U, there exists an element β in ${}^*U$ such that ${}^*\!\models\phi(a, \beta)$ for all a in U.

A quadruple $(U, {}^*U, {}^*\!\in, *)$ or simply ${}^*U$ is called an **enlargement** of U if it satisfies Axioms 1 and 2.

These axioms are strong enough to develop the basic theory of nonstandard analysis, but sometimes we require a stronger axiom. Let κ be an infinite cardinal.

Axiom 3 (κ-saturation): Let $\phi(x, y)$ be a binary ${}^*\mathscr{L}$-formula concurrent in ${}^*U$. Then, for every subset A of ${}^*U$ with cardinality at most κ, there exists an element β of ${}^*U$ such that ${}^*\!\models\phi(\alpha, \beta)$ for all $\alpha \in A$.

${}^*U$ is called a **κ-saturated model** if it satisfies Axioms 1 and 3, and a κ-saturated enlarge-

ment if it satisfies Axioms 1, 2, and 3. If κ is not less than the cardinality of U, then a κ-saturated model is necessarily an enlargement.

In the remainder of this article, $^*U$ will be an enlargement of U, if the contrary is not explicitly mentioned.

For α in $^*U$, $\hat{\alpha}$ is the set of $\xi \in {}^*U$ such that $\xi \, {}^*\!\!\in \alpha$: $\hat{\alpha} = \{\xi \in {}^*U \mid \xi \, {}^*\!\!\in \alpha\}$. For simplicity, we write $\in$ for $^*\!\!\in$ and identify $\hat{\alpha}$ with α. Under this identification, a subset A of $^*U$ is called **internal** if it belongs to $^*U$ and **external** if not.

Axiom 1 and 2 imply the following results.

(1) There exist infinite hypernatural numbers, namely, elements of $^*\mathbf{N}$ bigger than all $^*n$, $n \in \mathbf{N}$.

(2) There are positive infinitesimal hyperreal numbers.

(3) For every set A in U, there exists a hyperfinite internal subset of $^*A$ containing all $^*a$, $a \in A$. Here, Γ is called **hyperfinite** if we have $^*\!\models \phi(\Gamma)$, where $\phi(x)$ is a formula that says "x is a finite set."

(4) Let A be a set in U. Then $^*A$ has nonstandard elements if and only if A is an infinite set.

(5) If a family $\mathscr{F}$ of sets in U has the finite intersection property, then the family $\{^*A \mid A \in \mathscr{F}\}$ in $^*U$ has a nonempty intersection.

(6) Let α: $^*\mathbf{N} \rightarrow {}^*\mathbf{R}$ be an internal hypersequence. If $\alpha(^*n)$ is infinitesimal for every $n \in \mathbf{N}$, then there exists an infinite hypernatural number λ such that $\alpha(\nu)$ is infinitesimal for every hypernatural number ν less than λ.

If we moreover assume Axiom 3, we have the following results.

(7) The cardinality of an internal set is either finite or more than κ.

(8) If a family $\mathscr{F}$ of internal sets in $^*U$ has the finite intersection property and cardinality κ at most, $\mathscr{F}$ has a nonempty intersection.

(9) Introduce the order topology in $^*\mathbf{R}$. Then every subset of $^*\mathbf{R}$ with cardinality at most κ is bounded and discrete.

(10) Let α, β be internal sets and C a subset of α with cardinality at most κ. Then every mapping from C to β can be prolonged to an internal mapping from α to β.

(10′) In particular, every external sequence $\mathbf{N} \rightarrow \beta$ can be prolonged to an internal sequence $^*\mathbf{N} \rightarrow \beta$. This follows from the assumption of countable saturation. This fact plays an essential role in nonstandard probability theory, which is now in the process of rapid development (→ Section D (3)).

C. Construction of Ultrapower Models

Let I be an infinite set. A mapping α from I to U is a family of elements in U with indices in I. So we write $\langle \alpha(i) \rangle_{i \in I}$ or simply $\langle \alpha(i) \rangle$ for α.

Let $\mathscr{F}$ be a nonprincipal ultrafilter on I and define an equivalence relation $\sim_{\mathscr{F}}$ on the set U^I of all mappings from I to U:

$$\langle \alpha(i) \rangle \sim_{\mathscr{F}} \langle \beta(i) \rangle \Leftrightarrow \{i \in I \mid \alpha(i) = \beta(i)\} \in \mathscr{F}.$$

Denote by $^*U$ the quotient set of U under the relation $\sim_{\mathscr{F}}$. The class of $\langle \alpha(i) \rangle_{i \in I}$ will be denoted by $[\alpha(i)]_{i \in I}$ or simply $[\alpha(i)]$.

Define a binary relation $^*\!\!\in$ on $^*U$:

$$[\alpha(i)] \, {}^*\!\!\in [\beta(i)] \Leftrightarrow \{i \in I \mid \alpha(i) \in \beta(i)\} \in \mathscr{F}.$$

Let $*$ be the diagonal mapping from U to $^*U$ and consider the quadruple $(U, {}^*U, {}^*\!\!\in, *)$.

Theorem (Łoś): The foregoing quadruple satisfies Axiom 1. Moreover, $^*U$ is a countably saturated model.

In particular, let I be the set of all finite subsets of U. For $i \in I$, put $\mu(i) = \{j \in I \mid i \subset j\}$. Then the family of sets $\mathscr{B} = \{\mu(i) \mid i \in I\}$ has the finite intersection property, and therefore there exists an ultrafilter $\mathscr{F}$ on I including $\mathscr{B}$ (use †Zorn's lemma). If we construct $^*U$ from the ultrafilter $\mathscr{F}$, then $^*U$ is a countably saturated enlargement of U [1].

It is not easy to construct a κ-saturated enlargement for an arbitrary cardinal κ. We have two ways: to use a κ-good ultrafilter (the proof of its existence is difficult) [2, 3] or to use an ultralimit (iteration of ultrapowers) [4, 16].

D. Applications

(1) Infinitesimal Calculus. A hyperreal number α (element of $^*\mathbf{R}$) is called **infinitesimal** if $|\alpha|$ is smaller than every positive real number. If $\alpha - \beta$ is infinitesimal, α and β are said to be infinitely close to each other; this is written $\alpha \approx \beta$. If $1/\alpha$ is not infinitesimal, α is called finite. Every finite hyperreal number α is infinitely close to a real number a (completeness of $\mathbf{R}$). We call a the **standard part** of α and write $\mathrm{st}(\alpha)$. Every finite hyperinteger is an integer.

Let f be a real-valued function on a real interval I. Then $^*f$ is a hyperreal-valued function on the hyperreal interval $^*I$. The function f is continuous if and only if $f(x) \approx {}^*f(\eta)$ for every $x \in I$, $\eta \in {}^*I$ with $x \approx \eta$. The function f is uniformly continuous if $^*f(\xi) \approx {}^*f(\eta)$ for every $\xi, \eta \in {}^*I$ with $\xi \approx \eta$.

Let δx be a variable ranging over nonzero infinitesimals. The function $\delta f = {}^*f(a + \delta x) - f(a)$ of δx will be called the infinitesimal increment of f at a. f is differentiable at a if and only if the quotient $\delta f/\delta x$ is of infinitesimal variation. The common standard part of $\delta f/\delta x$ is the derivative $f'(a)$.

The higher-order differential $\delta^n f$ is also justified as a higher-order infinitesimal difference, with the standard part of $\delta^n f/\delta x^n$ being

$f^{(n)}(a)$. We can define the Riemann integral as the standard part of a Riemann hypersum with respect to a hyperfinite partition of infinitesimal width.

This type of reformulation permits us to rewrite the whole of calculus, and many teachers are trying to adapt this theory to elementary calculus [5, 6].

(2) Topological Spaces. Let X be a topological space in U and let a be a point of X. The intersection of $*A$, A varying over the neighborhoods of a is called the **monad** of a and is denoted by Mon(a). There exist hyperneighborhoods of a contained in Mon(a) (infinitesimal neighborhoods). The topology is determined by the system of monads $\langle \mathrm{Mon}(a) \rangle_{a \in X}$. We can thus rewrite the theory of topological spaces. For example, X is Hausdorff if and only if $\mathrm{Mon}(a) \cap \mathrm{Mon}(b) = \varnothing$ for every $a \neq b$ in X. X is compact if and only if every point of $*X$ belongs to the monad of some element in X. This characterization is very useful; using it, Robinson and Bernstein were able to solve the invariant subspace problem for a special class of operators on a Hilbert space [7]. We can also construct a Haar measure very naturally and simply [2].

(3) Measures and Probability Theory. Let $(X, \mathbf{B}, m)$ be a measure space. Then there exist a hyperfinite subset Γ of $*X$ and a positive hyperreal-valued internal function φ such that we have

$$\int_X f\,dm = \sum_{\xi \in \Gamma} {}^*f(\xi)\varphi(\xi)$$

for every integrable function f [8]. The right-hand side is a hyperfinite sum. It follows that the measure m can be extended to a finitely additive measure defined over all subsets of X.

If in particular every measurable finite set is of measure 0, then there exist a hyperfinite subset Γ of $*X$ including X and a hypernatural number ρ such that we have

$$\int_X f\,dm \approx \frac{1}{\rho} \sum_{\xi \in \Gamma} {}^*f(\xi)$$

for every integrable function f [8]. If $m(X) = 1$, then ρ can be taken as the hypercardinality of Γ. Here, the right-hand side is nothing but the mean value of $*f$ on a hyperfinite set. This idea leads us to a simple description of probability theory.

The above method is rather formal; but P. Loeb [9] has pioneered a new approach in probability theory by constructing an external measure space from a hyperfinitely additive internal measure space $(X, \mathscr{A}, \nu)$ in $*U$. In the following, $*U$ is supposed to be countably saturated.

Let $\sigma(\mathscr{A})$ be the external countably additive algebra of subsets of X generated by $\mathscr{A}$. Then the mapping $\mathscr{A} \mapsto \mathrm{st}(\nu(\mathscr{A}))$ from $\mathscr{A}$ to $\mathbf{R}$ can be extended to a measure on $\sigma(\mathscr{A})$. The completion of this measure space is denoted $(X, L(\mathscr{A}), L(\nu))$. In cases where X is hyperfinite, $\mathscr{A}$ is the totality of internal subsets of X, and $\nu(X) = 1$, then the space $(X, L(\mathscr{A}), L(\nu))$ is called a **Loeb space** and $L(\nu)$ a **Loeb measure**.

Every Radon probability space can be represented as the image of a Loeb space by a measure-preserving mapping. Therefore the probability theory on a Radon probability space can be reduced to that on a Loeb space.

For example, for †Lebesgue measure on the interval [0,1], take an infinite hypernatural number λ and put $X = \{\mu/\lambda \mid \mu \in {}^*\mathbf{N}, 0 \leqslant \mu \leqslant \lambda - 1\}$. If we assign $1/\lambda$ to every point of X, we have an internal probability measure ν on X. Then the mapping st: $x \mapsto \mathrm{st}(x)$ from X to [0,1] serves as a measure-preserving mapping from the Loeb space $(X, L(\mathscr{A}), L(\nu))$ onto the Lebesgue measure space on [0,1].

The notion of **lifting** plays a key role in probability theory on Loeb spaces. Let f be a real-valued function on X, and F an internal hyperreal-valued function on X. F is called a lifting of f if we have $f(x) = \mathrm{st}(F(x))$ almost everywhere with respect to $L(\nu)$. Hence f is measurable if and only if it has a lifting.

Shuttling between an internal probability space and its Loeb space by lifting and standard-part mapping, we can develop, simply and partially hyperfinitely, probability theory on the Loeb space. Among others, Anderson [10, 11] and Keisler [12] applied this method quite successfully to Brownian motions, Itô integrals, stochastic differential equations, etc. [17].

E. Nonstandard Set Theory

In our formulation in **B**, we must construct $*U$ from a fixed universe U. Nelson [13], Hrbacek [14] and Čuda [15] independently invented theories that nonstandardize the whole of set theory. In these theories, there is only one real number field $\mathbf{R}$, and $\mathbf{R}$ already contains infinitesimal numbers. Compare this to Robinson-type infinitesimal analysis, where infinitesimal numbers are introduced as an ad hoc tool. The new theories may demand a reflection on mathematical description in the natural sciences.

In the following, we outline (intuitively) Hrbacek's theory, as strengthened and improved by Kawai [16].

We start from ZFC, the Zermelo-Fraenkel set theory plus the axiom of choice. The language of Kawai's theory NST is that of ZFC

plus two constants S and I. We understand S as the totality of standard sets and I as the totality of internal sets.

Let ϕ be a formula of ZFC, that is, a formula of NST without S and I. We write ${}^S\phi$ (resp. ${}^I\phi$), the formula of NST obtained by restricting the scope of variables in ϕ to S (resp. I).

The mathematical axioms and axiom schemes of NST are substantially as follows.

(1) If ϕ is an axiom of ZFC, then ${}^S\phi$ is an axiom of NST. In other words, all the axioms of ZFC are valid in the universe of standard sets.

(2) In the universe of all sets, all the axioms of ZFC are valid except the regularity axiom.

(3) Every standard set is internal. Every element of an internal set is internal.

(4) (transfer) For every n-ary formula $\phi(x_1, \ldots, x_n)$ of ZFC, we have

$$\forall x_1 \in S \ldots \forall x_n \in S[{}^S\phi(x_1, \ldots, x_n) \leftrightarrow {}^I\phi(x_1, \ldots, x_n)].$$

(5) (saturation) For every set of size at most that of S, the scheme corresponding to Axiom 3 (κ-saturation) in Section B holds. Therefore the scheme corresponding to Axiom 2 (enlargement) holds also.

(6) (standardization) For every set a included in a standard set, there exists a standard set b such that we have $\forall x \in S[x \in a \leftrightarrow x \in b]$.

This completes the description of NST.

Theorem (Nelson-Hrbacek-Kawai): NST is a conservative extension of ZFC. That is, let ϕ be a sentence of ZFC. If ${}^S\phi$ can be proved in NST, then ϕ can be proved in ZFC.

Corollary: If ZFC is consistent, so is NST.

References

[1] A. Robinson, Non-standard analysis, North-Holland, 1966; second edition, 1974.
[2] M. Saito, Ultraproducts and non-standard analysis (in Japanese), Tokyo Tosho, 1976.
[3] C. C. Chang and H. J. Keisler, Model theory, North-Holland, 1973.
[4] K. D. Stroyan and W. A. J. Luxemburg, Introduction to the theory of infinitesimals, Academic Press, 1976.
[5] H. J. Keisler, Elementary calculus, Prindle, Weber & Schmidt, 1976.
[6] H. J. Keisler, Foundations of infinitesimal calculus, Prindle, Weber & Schmidt, 1977.
[7] M. Davis, Applied nonstandard analysis, Wiley, 1976.
[8] M. Saito, On the non-standard representation of linear mappings from a function space, Comm. Math. Univ. Sancti Pauli, 26 (1977), 165–185.
[9] P. A. Loeb, Conversion from nonstandard to standard measure spaces and applications in probability theory, Trans. Amer. Math. Soc., 211 (1975), 113–122.
[10] R. M. Anderson, A nonstandard representation for Brownian motion and Itô integration, Israel J. Math., 25 (1976), 15–46.
[11] R. M. Anderson, Star-finite probability theory, Dissertation, Yale Univ., 1977.
[12] H. J. Keisler, An infinitesimal approach to stochastic analysis, Mem. Amer. Math. Soc. 297 (1984).
[13] E. Nelson, Internal set theory, Bull. Amer. Math. Soc., 83 (1977), 1165–1198.
[14] K. Hrbacek, Axiomatic foundations for nonstandard analysis, Fund. Math., 98 (1978), 1–19.
[15] K. Čuda, A nonstandard set theory, Comm. Math. Univ. Carolinae, 17 (1976), 647–663.
[16] T. Kawai, Nonstandard analysis by axiomatic method, Southeast Asian Conference on Logic, C.-T. Chong and M. J. Wicks (eds.), Elsevier, 1983.
[17] N. J. Cutland, Nonstandard measure theory and its applications, Bull. London Math. Soc., 15 (1983), 529–589.

294 (II.9) Numbers

A. General Remarks

From counting, a primitive mental activity, came the **natural numbers** (N) (→ Section B), which serve to denote the number of items or the order in which these items are arranged. We can extend this concept to define, step by step, the †integers (**Z**), †rational numbers (**Q**), †real numbers (**R**), and †complex numbers (**C**) (→ 355 Real Numbers; 74 Complex Numbers).

The extensions up to the rationals are carried out to attain a domain within which the operations of addition, subtraction, multiplication, and division, namely, the **four arithmetic operations** (or **rational operations**) can be performed indefinitely, with division by 0 the only exception. To develop the theory of natural numbers there are two well-known methods, one being G. Peano's system of axioms [4], which will be stated below, and the other being R. Dedekind's set-theoretic treatment [5]. The domain of rational numbers can be extended, taking continuity into consideration, to that of real numbers by several methods of which the best known are those of the Dedekind cut (1872) [8], which will be described below, and of G. Cantor's

fundamental sequences (1872) [9]. There is also a way based on infinite series, given by K. Weierstrass in his lectures (1859–1860). Though in the domain of real numbers (1) the four arithmetic operations can be performed indefinitely and (2) an order relation for magnitude is defined, the equation $x^2+1=0$ has no root. Introducing numbers expressible as $a+ib$ ($i=\sqrt{-1}$), it is possible to solve every equation of the second order. Such numbers, once called imaginary, have been used since the days of G. Cardano [6] in the 16th century. L. Euler also made good use of complex numbers as convenient tools in many calculations and obtained among other things his formula $\exp i\theta=\cos\theta+i\sin\theta$. Indeed, the notation $i=\sqrt{-1}$ was used for the first time by Euler (1777). Furthermore, C. F. Gauss, giving imaginary numbers the name *complex numbers*, showed that any algebraic equation with numerical coefficients always has roots in the domain of complex numbers. The discovery of the geometric representation of complex numbers by several mathematicians in the late 18th and early 19th centuries and their use in many applications have made complex numbers indispensable in mathematics.

Though there are extensions of complex numbers, such as †Hamilton's quaternions or †Cayley numbers, it is generally accepted that when we speak of a *number* we usually mean a complex number.

B. Natural Numbers

Peano, basing his system on a specific natural number, 1, and a function such that to each natural number x corresponds a natural number $x+1$ (hereafter denoted by x' and called the **successor** of x), formulated the fundamental properties of the set $\mathbf{N}$ of natural numbers in the following five axioms, called the **Peano postulates**: (1) $1\in\mathbf{N}$; (2) if $x\in\mathbf{N}$, then $x'\in\mathbf{N}$; (3) if $x\in\mathbf{N}$, then $x'\neq 1$; (4) if $x'=y'$ ($x,y\in\mathbf{N}$), then $x=y$; and (5) if a set M satisfies the two conditions: $1\in M$, and $x\in M$ implies $x'\in M$, then $\mathbf{N}\subset M$. Since these postulates determine $\mathbf{N}$ uniquely up to isomorphism, they can be regarded as a definition of $\mathbf{N}$. Elements of $\mathbf{N}$ are called **natural numbers**.

Owing to Peano's fifth postulate, regarding a certain property $P(n)$ for natural numbers n we can deduce that $P(n)$ is true for every n if we prove both of the following conditions: (i) $P(1)$ is true; and (ii) for any natural number k, if $P(k)$ is true, then $P(k+1)$ is true. Such reasoning is called **mathematical induction** (or **complete induction**). Accordingly, Peano's fifth postulate is called the **axiom of mathematical induction. Double mathematical induction**, a generalization of mathematical induction, is as follows: to show that the property $P(m,n)$ is true for every pair of natural numbers m and n, we have only to show that (iii) $P(m,1)$ and $P(1,n)$ are true for every m and every n; and (iv) for every pair of natural numbers k and l, if $P(k+1,l)$ and $P(k,l+1)$ are true, then $P(k+1, l+1)$ is true. This axiom can be formulated in several other ways and can be generalized further to n-tuple mathematical inductions ($n=2,3,4,\ldots$), generically called **multiple mathematical inductions**.

Assume for a set M that a mapping f from the †Cartesian product $\mathbf{N}\times M$ into M is given. Then a mapping φ from $\mathbf{N}$ into M such that (v) $\varphi(1)=a$; and (vi) $\varphi(x')=f(x,\varphi(x))$ ($x\in\mathbf{N}$) exists and is unique. Defining φ by (v) and (vi) is called the **definition of φ by mathematical induction**.

In particular, given a natural number a, the mapping $\varphi:\mathbf{N}\to\mathbf{N}$ defined by (vii) $\varphi(1)=a'$; and (viii) $\varphi(x')=\varphi(x)'$ is called **addition** by a. We shall write $\varphi(b)=a+b$, whence $x'=x+1$. Addition thus defined obeys the following laws: $a+b=b+a$ (**commutative law**); $(a+b)+c=a+(b+c)$ (**associative law**). Peano's postulates are thus equivalent to the following: (1′) $1\in\mathbf{N}$; (2′) for each pair $a,b\in\mathbf{N}$, $a+b\in\mathbf{N}$ is defined so that addition obeys the commutative and associative laws; (3′) for each pair of natural numbers a and b, one and only one of the following three relations holds: $a=b+c$ ($c\in\mathbf{N}$); $a=b$; $a+c=b$ ($c\in\mathbf{N}$); (4′) the same as (5) (mathematical induction). From (1′)–(4′) follows the **cancellation law**: $a+c=b+c\Leftrightarrow a=b$. Define $a\geqslant b$ if and only if $a=b$ or $a=b+c$ ($a,b,c\in\mathbf{N}$). Then, from (3′), $\mathbf{N}$ becomes a †totally ordered set and $a\geqslant b\Leftrightarrow a+c\geqslant b+c$.

For each $a\in\mathbf{N}$, the mapping $\varphi:\mathbf{N}\to\mathbf{N}$ defined by $\varphi(1)=a$ and $\varphi(x')=\varphi(x)+a$ is called **multiplication** by a, and we write $\varphi(b)=ab$ (or $a\cdot b$). Multiplication obeys the following laws: $ab=ba$ (**commutative law**); $(ab)c=a(bc)$ (**associative law**); $a(b+c)=ab+ac$, $(a+b)c=ac+bc$ (**distributive laws**); and $ac=bc\Leftrightarrow a=b$ (**cancellation law**). The statement $a\cdot 1=1\cdot a=a$ also holds.

Natural numbers, which have been introduced thus far as †ordinal numbers, also have the properties of †cardinal numbers. Denoting $\{1,2,\ldots,n\}=M_n$, we have $\overline{\overline{M_n}}=\overline{\overline{M_m}}\Leftrightarrow m=n$, $\overline{\overline{M_m}}+\overline{\overline{M_n}}=\overline{\overline{M_{m+n}}}$, $\overline{\overline{M_m}}\times\overline{\overline{M_n}}=\overline{\overline{M_{mn}}}$ (→ 49 Cardinal Numbers).

C. Integers

Introducing new numbers which are not in $\mathbf{N}=\{1,2,\ldots\}$, represented by the notations $0,-1,-2,\ldots,-n,\ldots$, we write $\mathbf{Z}=\{\ldots,$

$-n, \ldots, -2, -1, 0, 1, 2, \ldots, n, \ldots\}$. An element of $\mathbf{Z}$ is called an **integer** (or **rational integer**).

Algebraically we can construct $\mathbf{Z}$ from $\mathbf{N}$ as follows: Let the set of all †ordered pairs (k, l) of natural numbers k, l be $M = \mathbf{N} \times \mathbf{N}$, and define in M an †equivalence relation $(k, l) \sim (m, n)$ by $k+n=m+l$. Let the equivalence class of (k, l) be $K(k, l)$, and construct the †quotient space, $M/\sim = M^*$. Then the mapping φ: $\mathbf{Z} \rightarrow M^*$ defined by $\varphi(n) = K(k+n, k)$, $\varphi(0) = K(k, k)$, $\varphi(-n) = K(k, k+n)$ is bijective. Setting $K(k, l) + K(m, n) = K(k+m, l+n)$, addition can be defined in M^* (and accordingly in $\mathbf{Z}$) which is an extension of that in $\mathbf{N}$. Since $K(k, l) - K(m, n) = K(k+n, l+m)$, subtraction can be defined in $\mathbf{Z}$. This makes $\mathbf{Z}$ an †Abelian group with respect to addition. An order relation in $\mathbf{Z}$ is defined by $K(k, l) \geqslant K(m, n) \Leftrightarrow k+n \geqslant m+l$, which makes $\mathbf{Z}$ a totally ordered set. This order relation is an extension of that in $\mathbf{N}$. In particular, $\mathbf{N} = \{a \in \mathbf{Z} \mid a > 0\}$. Furthermore, setting $K(k, l) \times K(m, n) = K(km + ln, kn + lm)$, multiplication in $\mathbf{Z}$ can be defined. It is an extension of that in $\mathbf{N}$ and obeys commutative, associative, and distributive laws. Also, for each a, $b \in \mathbf{Z}$, we have $ab = 0 \Leftrightarrow (a = 0$ or $b = 0)$. Thus, $\mathbf{Z}$ becomes an †integral domain.

D. Rational Numbers

Let P be the set of all ordered pairs (a, b) of integers a, b with $b \neq 0$, and define in P an equivalence relation $(a, b) \sim (c, d)$ by $ad = bc$. Each equivalence class determined by this relation is called a **rational number**. Denoting by $L(a, b)$ the equivalence class to which (a, b) belongs, we can define the sum $x + y$, the difference $x - y$, the product xy, and the quotient x/y of rational numbers $x = L(a, b)$, $y = L(c, d)$, as in the cases of addition and multiplication of integers, in the following way: $x + y = L(ad + bc, bd)$, $x - y = L(ad - bc, bd)$, $xy = L(ac, bd)$, $x/y = L(ad, bc)$, where the quotient is defined only when $c \neq 0$. Thus the set $\mathbf{Q}$ of all rational numbers becomes a †field.

For the same reason that we have identified integers of a special type with natural numbers, we now identify a rational number expressible as $L(a, 1)$ with an integer a. Henceforth, any rational number $L(a, b)$ $(b \neq 0)$ can be expressed in the form of a quotient a/b $(b \neq 0)$ of integers a and b.

From $L(a, b) = L(ca, cb)$ $(c \neq 0)$, it is always possible to assume $b > 0$ in the representation $L(a, b)$ of a rational number x. For any two rational numbers $x = L(a, b)$, $y = L(c, d)$ with $b > 0$, $d > 0$, define an ordering in $\mathbf{Q}$ by $x \geqslant y \Leftrightarrow ad \geqslant bc$, which is an extension of the ordering of integers. Thus (i) $\mathbf{Q}$ becomes totally ordered, and we have (ii) $x \geqslant y \Rightarrow x + z \geqslant y + z$ and (iii) $x \geqslant y$ and $z \geqslant 0 \Rightarrow xz \geqslant yz$. The rational x is called **positive** if $x > 0$ and **negative** if $x < 0$.

E. Real Numbers

Two typical methods of constructing real numbers from rational numbers are those of Dedekind and of Cantor.

Dedekind's Theory of Real Numbers. We call a pair (A_1, A_2) of subsets A_1, A_2 of the set $\mathbf{Q}$ of all rational numbers a **cut** of $\mathbf{Q}$ if they satisfy the following conditions: (i) $A_1 \neq \varnothing$, $A_2 \neq \varnothing$; (ii) $\mathbf{Q} = A_1 \cup A_2$; (iii) $a_1 \in A_1$, $a_2 \in A_2 \Rightarrow a_1 < a_2$. Then the following three cases are distinguished: there is (i) a maximum in A_1 with no minimum in A_2; (ii) no maximum in A_1 with a minimum in A_2; or (iii) no maximum in A_1 with no minimum in A_2. A cut with either condition (i) or (iii) is called a **real number** (in the sense of Dedekind); condition (ii) can be converted to (i). The set of all real numbers is denoted hereafter by $\mathbf{R}$ and each real number by α or β or .... A real number with property (i) is called a **rational real number**, and a real number with property (iii) an **irrational real number**. Any rational real number is uniquely determined by the maximum a^* of A_1, and the mapping: $a \rightarrow a^* = (A_1, A_2)$ from the set $\mathbf{Q}$ of rational numbers onto the set $\mathbf{Q}^*$ of rational real numbers is bijective.

I. For real numbers $\alpha = (A_1, A_2)$ and $\beta = (B_1, B_2)$ we define $\alpha \leqslant \beta$ if and only if $A_1 \subset B_1$. By this ordering $\leqslant$, $\mathbf{R}$ becomes a totally ordered set.

II. For real numbers $\alpha = (A_1, A_2)$ and $\beta = (B_1, B_2)$, put $C_2 = \{a + b \mid a \in A_2, b \in B_2\}$ and $C_1 = \mathbf{R} - C_2$; then $(C_1, C_2) = \gamma$ is a real number. Define the sum $\alpha + \beta$ by setting $\alpha + \beta = \gamma$. Addition thus defined obeys commutative and associative laws, and $\mathbf{R}$ becomes an †Abelian group with 0^* as its zero element. Furthermore, for real numbers $\alpha = (A_1, A_2)$ and $\beta = (B_1, B_2)$ with $0^* \leqslant \alpha$, $0^* \leqslant \beta$, put $D_2 = \{ab \mid a \in A_2, b \in B_2\}$, $D_1 = \mathbf{R} - D_2$; then $(D_1, D_2) = \delta$ becomes a real number. Define the product $\alpha\beta$ by setting $\alpha\beta = \delta$. According as $0^* > \alpha$, $0^* \leqslant \beta$; $0^* \leqslant \alpha$, $0^* > \beta$; or $0^* > \alpha$, $0^* > \beta$, define $\alpha\beta = -((-\alpha)\beta)$; $\alpha\beta = -(\alpha(-\beta))$; and $\alpha\beta = (-\alpha)(-\beta)$, respectively. Multiplication thus defined obeys the commutative, associative, and distributive laws, and $\mathbf{R}$ becomes a field with 1^* as its unity element.

III. For ordering and arithmetic operations, we have (1) $\alpha \geqslant \beta \Rightarrow \alpha + \gamma \geqslant \beta + \gamma$; and (2) $\alpha \geqslant \beta$, $\gamma \geqslant 0^* \Rightarrow \alpha\gamma \geqslant \beta\gamma$.

By letting each rational number a corre-

spond to a rationally real number a^*, we can set up a bijection between the set **Q** of all rational numbers and **Q***. Furthermore, in this correspondence, the sum, product, 0, and 1 of **Q** are mapped to the sum, product, zero element, and unity element of **Q***, respectively; in addition, the ordering is preserved. Thus **Q** and **Q*** are isomorphic with respect to both arithmetic operations and ordering. We shall hereafter identify the element a^* with a. Accordingly we call rational real numbers simply **rational numbers** and similarly irrational real numbers simply **irrational numbers**.

IV. Similarly as for **Q**, we can define a cut (A_1, A_2) of the set of all real numbers **R** (precisely, by the conditions $A_1 \neq \varnothing$, $A_2 \neq \varnothing$; $\mathbf{R} = A_1 \cup A_2$; $d_1 \in A_1, d_2 \in A_2 \Rightarrow d_1 < d_2$). For each cut (A_1, A_2) of **R**, either there is a maximum in A_1 and no minimum in A_2 or no maximum in A_1 and a minimum in A_2 (Dedekind). This property is called the **connectedness** (or **continuity) of real numbers**.

By using definitions (I)–(IV), any real number α can be represented (i) as the †least upper bound of some set A of rational numbers: $\alpha = \sup A$; and also (ii) as the limit of a sequence $\{a_n\}$ of rational numbers: $\alpha = \lim a_n$.

Cantor's Theory of Real Numbers. A sequence $\{a_n\}$ of rational numbers is called a **fundamental sequence** (or **Cauchy sequence**) if it satisfies the following condition: for each positive rational number ε, there exists a natural number n_0 such that $-\varepsilon < a_n - a_m < \varepsilon$ for every a_m and a_n with $m > n_0$ and $n > n_0$. For two Cauchy sequences $\{a_n\}$ and $\{b_n\}$, we can write $\{a_n\} \sim \{b_n\}$ if $\{a_1, b_1, a_2, b_2, \ldots, a_n, b_n, \ldots\}$ is again a Cauchy sequence. The relation $\sim$ is an equivalence relation. Let **R**′ be the set of all equivalence classes obtained by classifying, with respect to $\sim$, the set of all Cauchy sequences, and call an element of **R**′ (an equivalence class) a **real number** (in the sense of Cantor). We shall write $[\{a_n\}]$ hereafter for the equivalence class of $\{a_n\}$. In particular, $\{a_n\}$ with $a_n = a$ $(n = 1, 2, \ldots)$ being a Cauchy sequence, we denote a real number $[\{a_n\}]$ by a^{**}. An element of **R**′ of this type is called a **rational real number** (in the sense of Cantor), while one not of this type is called an **irrational real number** (in the sense of Cantor).

For each pair of real numbers $\alpha = [\{a_n\}]$ and $\beta = [\{b_n\}]$, their sum and product are uniquely defined by $\alpha + \beta = [\{a_n + b_n\}]$ and $\alpha\beta = [\{a_n b_n\}]$, where both $\{a_n + b_n\}$ and $\{a_n b_n\}$ are shown to be Cauchy sequences. Further, for α and β, define $\alpha < \beta$ if $a_n < b_n$ hold for all n larger than some number n_0. As regards arithmetic operations and ordering, **R**′ has the properties (I)–(IV) of Dedekind's theory. Further, any Cauchy sequence of which the terms all are real numbers has a limit in **R**′ (**completeness of real numbers**).

The sets of real numbers obtained by the above two methods are isomorphic with respect to arithmetic operations and ordering (→ 355 Real Numbers).

F. Complex Numbers

To construct complex numbers from real ones, several methods have been devised, among which the one stated below is due to W. R. Hamilton.

An ordered pair (a, b) of real numbers a and b will be called a **complex number**. Arithmetic operations among complex numbers are defined as follows: $(a,b)+(c,d)=(a+c,b+d)$, $(a,b)-(c,d)=(a-c,b-d)$, $(a,b)\cdot(c,d)=(ac-bd,bc+ad)$, and

$$\frac{(a,b)}{(c,d)}=\left(\frac{ac+bd}{c^2+d^2},\frac{bc-ad}{c^2+d^2}\right),\qquad (c,d)\neq(0,0).$$

According to these definitions, the set **C** of all complex numbers becomes a †field with $(0,0)$ and $(1,0)$ as its †zero element and †unity element, respectively. $\mathbf{R}^* = \{(a,0) \mid a \in \mathbf{R}\}$ is a †subfield of **C**, and the mapping $\varphi: \mathbf{R} \to \mathbf{R}^*$ defined by $\varphi(a) = (a, 0)$ proves to be a field †isomorphism. Thus, identifying the element $(a, 0)$ of **R*** with the element a of **R**, **C** can be regarded as an †overfield of **R** ($\mathbf{R} \subset \mathbf{C}$). Hence, the zero element of **C** is the real number 0 and the unity element is the real number 1. The complex number $(0, 1)$ is called the **imaginary unit** and denoted by i. Thus we can write $a + bi$ as usual for a complex number (a, b) (→ 74 Complex Numbers).

References

[1] O. Perron, Irrationalzahlen, de Gruyter, second edition, 1929 (Chelsea, 1948).

[2] E. G. H. Landau, Grundlagen der Analysis, Akademische Verlag., 1934; English translation, Foundations of analysis, Chelsea, 1960.

[3] N. Bourbaki, Eléménts de mathématique, I. Théorie des ensembles, ch. 3, Actualites Sci. Ind., 1243b, second edition, 1967; III. Topologie générale, ch. 4, 1143c, third edition, 1960; ch. 8, 1235b, third edition, 1963; English translation, Theory of sets, Addison-Wesley, 1968; General topology, pt. I, Addison-Wesley, 1966.

[4] G. Peano, Arithmetices principia nova methodo exposita, Bocca, 1889 (Opere scelte II, Cremona, 1958).

[5] R. Dedekind, Was sind und was sollen die Zahlen? Vieweg, 1887 (Gesammelte mathematische Werke 3, Vieweg, 1932); English translation, Essays on the theory of numbers, Open Court, 1901.

[6] G. Cardano, Artis magnae sive de regulis algebraicis, Nuremberg, 1545.
[7] W. R. Hamilton, Lectures on quaternions, Hodges and Smith, 1853.
[8] R. Dedekind, Stetigkeit und irrationale Zahlen, Vieweg, 1872 (Gesammelte mathematische Werke 3, Vieweg, 1932); English translation, Essays on the theory of numbers, Open Court, 1901.
[9] G. Cantor, Über die Ausdehnung eines Satzes aus der Theorie der trigonometrischen Reihen, Math. Ann., 5 (1872), 123–132 (Gesammelte Abhandlungen, Springer, 1932).

295 (V.4) Number-Theoretic Functions

A. Recurrent Sequences

A (complex-valued) function that has the set of nonnegative integers (or the set **N** of natural numbers) as its †domain is called a **number-theoretic** (or **arithmetic**) **function**. Thus it can be regarded as a †sequence of numbers. We first consider recurrent sequences. Let $f(x_0, x_1, \ldots, x_{r-1})$ be a complex-valued function of r variables. Put $u_0 = a_0$, $u_1 = a_1, \ldots, u_{r-1} = a_{r-1}$, and successively define $u_{i+r} = f(u_i, u_{i+1}, \ldots, u_{i+r-1})$ $(i = 0, 1, 2, \ldots)$. The sequence $\{u_n\}$ thus defined is called a **recurrent sequence of order** r determined by the initial values $a_0, a_1, \ldots, a_{r-1}$ and the function f. In particular, when the defining function f is given by $\sum_{i=0}^{r-1} b_i x_i$, $\{u_n\}$ is called a **linear recurrent sequence**. The **Fibonacci sequence** is a special linear recurrent sequence with initial values a_0, a_1 and the defining function $x_0 + x_1$. Let $\alpha = (1+\sqrt{5})/2$, $\beta = (1-\sqrt{5})/2$ be two roots of $1 + x = x^2$, and let c_1 and c_2 be determined by $c_1 + c_2 = a_0$ and $c_1\alpha + c_2\beta = a_1$. Then the Fibonacci sequence with the initial values a_0, a_1 is given by putting $u_n = c_1\alpha^n + c_2\beta^n$ $(n = 0, 1, 2, \ldots)$. If we put $a_0 = 1$, $a_1 = 1$, then we obtain **Binet's formula**:

$$u_n = \frac{1}{\sqrt{5}}\left\{\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+1}\right\}.$$

B. Multiplicative Functions

A number-theoretic function f with domain **N** is said to be **multiplicative** if $f(1) = 1$, $f(mn) = f(m)f(n)$ for $(m, n) = 1$, and to be **completely multiplicative** if $f(1) = 1$, $f(mn) = f(m)f(n)$ for any m and n ($\in$**N**). Similarly, f is said to be **additive** if $f(1) = 0$, $f(mn) = f(m) + f(n)$ for $(m, n) = 1$, and to be **completely additive** if $f(1) = 0$ and $f(mn) = f(m) + f(n)$ for any m and n ($\in$**N**). If f is multiplicative and $\sum_{n=1}^{\infty} f(n)$ is absolutely convergent, then $\sum_{\nu=0}^{\infty} f(p^\nu)$ is absolutely convergent for any prime p. Moreover, we have

$$\sum_{n=1}^{\infty} f(n) = \prod_p \left(\sum_{\nu=0}^{\infty} f(p^\nu)\right),$$

where the infinite product on the right-hand side is also absolutely convergent. Furthermore, if $f(n)$ is completely multiplicative, then we have

$$\sum_{n=1}^{\infty} f(n) = \prod_p (1 - f(p))^{-1}.$$

In particular, for $f(n) = n^{-s}$, which is completely multiplicative (with $s = \sigma + it$ a complex variable), we get the †Euler product formula for the †Riemann zeta function:

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s} = \prod_p (1 - p^{-s})^{-1}$$

for $\sigma > 1$.

C. Convolutions

If f and g are number-theoretic functions, the **convolution** $f * g$ is defined by

$$(f * g)(n) = \sum_{d|n} f(d) g(n/d),$$

with the summation carried over all divisors d of n. For any number-theoretic functions f, g, h, we have $f * g = g * f$ and $(f * g) * h = f * (g * h)$. If f and g are multiplicative number-theoretic functions, $f * g$ is again a multiplicative number-theoretic function.

The **Möbius function** $\mu(n)$ is defined as follows: $\mu(1) = 1$, $\mu(n) = 0$ if n is divisible by the square of a prime, and $\mu(n) = (-1)^r$ if n is the product of r distinct primes. It can easily be proved that μ is multiplicative and that

$$\sum_{d|n} \mu(d) = \begin{cases} 1 & (n = 1), \\ 0 & (n > 1). \end{cases}$$

Let e and ρ be functions defined by $e(1) = 1$, $e(n) = 0$ $(n > 1)$, and $\rho(n) = 1$ for every n. Then e is the identity element for the convolution $*$, and $\mu * \rho = \rho * \mu = e$. It follows that $f * \rho = g$ is equivalent to $f = g * \mu$, that is,

$$g(n) = \sum_{d|n} f(d)$$

implies that

$$f(n) = \sum_{d|n} \mu(d) g(n/d),$$

and vice versa. We call the latter the **Möbius inversion formula**. Similarly, for complex-valued functions F, G defined on $[1, +\infty)$,

$$G(x) = \sum_{n \leqslant x} F(x/n)$$

is equivalent to

$$F(x)=\sum_{n\leqslant x}\mu(n)G(x/n).$$

Let $\varphi(n)$ be the number of integers m not greater than n and such that $(m,n)=1$. The function $\varphi(n)$, called the **Euler function**, is multiplicative, and we have

$$\varphi(n)=n\prod_{p|n}\left(1-\frac{1}{p}\right)=n\sum_{d|n}\frac{\mu(d)}{d},$$

and $\varphi(p^n)=p^n-p^{n-1}$ for every prime p. Let ν be the function defined by $\nu(n)=n$ for every n. Then we have $\mu*\nu=\varphi$ and $\nu=\rho*\varphi$, and hence $\sum_{d|n}\varphi(d)=n$.

The **generalized divisor function** $d_k(n)$ is the number of ways of expressing n as a product of k factors. Thus we have

$$d_k(n)=\sum_{n_1n_2\ldots n_k=n}1=\sum_{m|n}d_{k-1}(m),$$

and $d_k=\rho*\ldots*\rho$ (k factors). Therefore, d_k is multiplicative. For simplicity, we write $d(n)$ instead of $d_2(n)$. Then $d(n)$ is the number of divisors of n, and we call it the **divisor function**. For example, $d(12)=6$. We denote by $\sigma_\alpha(n)$ the sum of αth powers of divisors of n. If $n=\prod_{p|n}p^l$, then we have

$$\sigma_\alpha(n)=\sum_{d|n}d^\alpha=\prod_{p|n}\frac{p^{\alpha(l+1)}-1}{p^\alpha-1},$$

and σ_α is also multiplicative. We write $\sigma_1(n)=\sigma(n)$. A number n satisfying $\sigma(n)=2n$ is said to be a †perfect number. Let $n=\prod_{p|n}p^{l_p}$. The functions $\omega(n)=\sum_{p|n}1$ (the number of distinct prime factors of n) and $\Omega(n)=\sum_{p|n}l_p$ (the number of prime factors of n) are often used, the former being additive and the latter completely additive.

D. Residue Characters and Gaussian Sums

Let k be a positive integer and χ be a completely multiplicative function such that $\chi(n)=0$ for $(n,k)>1$ and $\chi(n_1)=\chi(n_2)$ for $n_1\equiv n_2 \pmod k$. The function χ is called a **Dirichlet character** (or **residue character**) with modulus k (or modulo k). There exist $\varphi(k)$ distinct characters modulo k for a given k. The character satisfying $\chi(n)=1$ for every n coprime to k is called the **principal character** modulo k, and we denote it by χ_0. It is easily proved that

$$\sum_{n=0}^{k-1}\chi(n)=\begin{cases}\varphi(k) & (\chi=\chi_0),\\ 0 & (\chi\neq\chi_0),\end{cases}$$

$$\sum_{\chi}\chi(n)=\begin{cases}\varphi(k) & (n\equiv 1 \pmod k),\\ 0 & (n\not\equiv 1 \pmod k).\end{cases}$$

Suppose that $k=k_1k_2\ldots k_r$ $((k_i,k_j)=1,\ i\neq j)$. Then there exist r characters χ_i modulo k_i permitting the unique decomposition $\chi=\chi_1\chi_2\cdots\chi_r$.

Let χ be a character modulo k and $\rho=\exp(2\pi i/k)$. The **Gaussian sum** modulo k is defined by

$$G(a,\chi)=\sum_{n\,(\mathrm{mod}\,k)}\chi(n)\rho^{an},$$

where n runs over a complete system modulo m (→ 297 Number Theory, Elementary, G). Hence if $a\equiv b \pmod k$, then $G(a,\chi)=G(b,\chi)$. Suppose that $k=k_1k_2\ldots k_r$ $((k_i,k_j)=1,\ i\neq j)$ and $a_i\equiv a \pmod{k_i}$. Then we have

$$G(a,\chi)=\varepsilon G(a_1,\chi_1)G(a_2,\chi_2)\ldots G(a_r,\chi_r),$$

$$\varepsilon=\chi_1(k/k_1)\chi_2(k/k_2)\ldots\chi_r(k/k_r),$$

where $\chi=\chi_1\chi_2\ldots\chi_r$ is the decomposition of χ mentioned above.

Let k_0 be a divisor of k. If $\chi(n)=1$ whenever $(n,k)=1$ and $n\equiv 1 \pmod{k_0}$, then we say that χ is defined mod k_0. The least positive integer $f=f(\chi)$ modulo to which χ is defined is called the **conductor** of χ. If the conductor of χ is k itself, then χ is said to be a **primitive character** modulo k. When χ has the decomposition $\chi=\chi_1\ldots\chi_r$, the conductor of χ is also decomposed as $f(\chi)=f(\chi_1)f(\chi_2)\ldots f(\chi_r)$. If D is a square-free integer, then the †discriminant d of the quadratic number field $\mathbf{Q}(\sqrt{D})$ (where $\mathbf{Q}$ is the rational number field) is either D ($D\equiv 1 \pmod 4$) or $4D$ ($D\equiv 2,3 \pmod 4$). The integers d so represented are called the **fundamental discriminants**. The †Kronecker symbol (d/n) (→ 347 Quadratic Fields) is defined only for such fundamental discriminants. Z. Suetuna and A. Z. Walfisz (1936) proved that if $\chi(n)$ is a real primitive character modulo k, then we necessarily have one of the following cases: (i) $k=p_1p_2\ldots p_r$; (ii) $k=4p_1p_2\ldots p_r$; or (iii) $k=8p_1p_2\ldots p_r$ (with the p_i distinct odd primes). In case (i),

$$\chi(n)=\begin{cases}(k/n) & (k\equiv 1 \pmod 4),\\ (-k/n) & (k\equiv -1 \pmod 4);\end{cases}$$

in case (ii),

$$\chi(n)=\begin{cases}(-k/n) & (k/4\equiv 1 \pmod 4),\\ (k/n) & (k/4\equiv -1 \pmod 4);\end{cases}$$

and in case (iii), $\chi(n)$ is either (k/n) or $(-k/n)$.

If χ is a primitive character modulo k, then $G(a,\chi)=\bar{\chi}(a)G(1,\chi)$, $G(1,\chi)\bar{G}(1,\chi)=k$. In particular, if χ is a real primitive character, then

$$G(1,\chi)=\begin{cases}\sqrt{k} & (\chi(-1)=1),\\ i\sqrt{k} & (\chi(-1)=-1).\end{cases}$$

Sometimes $S(a,k)=\sum_{n\,(\mathrm{mod}\,k)}\rho^{an^2}$ is called the Gaussian sum, where $\rho=\exp(2\pi i/k)$. If $(a,k_i)=1$ $(i=1,2)$ and $(k_1,k_2)=1$, then $S(a,k_1k_2)=S(ak_2,k_1)S(ak_1,k_2)$. If $(a,2)=1$, then

$$S(a,2^r)=\begin{cases}0 & (r=1),\\ (1+i^a)2^{r/2} & (r\text{ even}),\\ 2^{\pi ia/4}\cdot 2^{(r+1)/2} & (r\text{ odd}>1).\end{cases}$$

Let p be an odd prime and $(a,p)=1$. Then we have

$$S(a,p^r)=\begin{cases}(a/p)S(1,p) & (r=1),\\ p^{r/2} & (r \text{ even}),\\ p^{(r-1)/2}S(a,p) & (r \text{ odd}>1),\end{cases}$$

where (a/p) is the †Legendre symbol (→ 297 Number Theory, Elementary, H). The well-known **Gauss formula** is stated as follows:

$$S(1,p)=\begin{cases}\sqrt{p} & (p\equiv 1 \pmod 4),\\ i\sqrt{p} & (p\equiv 3 \pmod 4)).\end{cases}$$

Various proofs of this formula have been given by many authors [5].

Let

$$G(a,\chi_0)=\sum_{h \pmod k}{}' \exp(2\pi iah/k),$$

called the **Ramanujan sum**, be denoted by $c_k(a)$. The sum $\sum'_{h \pmod k}$ means that h runs through a reduced residue system modulo m (→ 297 Number Theory, Elementary, G). It follows that $c_k(1)=\mu(k)$ and

$$c_k(a)=\sum_{d|(k,a)}\mu\left(\frac{k}{d}\right)d.$$

E. Analytic Methods

Let f be a number-theoretic function. One of the problems in analytic number theory is to estimate $\sum_{n\leqslant k}f(n)$ or to expand it into series. We assume that $f(x)$ and $g(x)$ are real-valued functions defined for $x\geqslant 1$ and $g(x)$ is of class C^1. Then we have

$$\sum_{n\leqslant x}f(n)g(n)=F(x)g(x)-\int_1^x F(t)g'(t)\,dt,$$

where $F(x)=\sum_{n\leqslant x}f(n)$. If we take $f(x)=1$, $g(x)=\log x$ or $1/x$ in this formula, then we obtain

$$\sum_{n\leqslant x}\log n = x\log x - x + O(\log x)$$

or

$$\sum_{n\leqslant x}n^{-1}=\log x + C + O(x^{-1})$$

(where C is the †Euler constant), respectively.

We now construct from the function f the Dirichlet series

$$F(s)=\sum_{n=1}^{\infty}f(n)n^{-s}$$

or the power series

$$P(x)=\sum_{n=1}^{\infty}f(n)x^n.$$

There are called the **generating functions** of f. The consideration of these functions makes possible the application of function-theoretic methods to the problem in this section. Sometimes more complicated functions are used.

We mainly consider the generating function represented by the Dirichlet series. The generating functions of ρ, μ, d, χ (→ Sections C, D) are $\zeta(s)=\sum_{n=1}^{\infty}n^{-s}$, $\sum_{n=1}^{\infty}\mu(n)n^{-s}$, $\sum_{n=1}^{\infty}d(n)n^{-s}$, $L(s,\chi)=\sum_{n=1}^{\infty}\chi(n)n^{-s}$, respectively. The †abscissa of absolute convergence of each of these Dirichlet series is 1. The function $L(s,\chi)$ is called the †Dirichlet L-function (→ 123 Distribution of Prime Numbers; 450 Zeta Functions). When $k=1$ and $\chi=\chi_0$, $L(s,\chi)$ is precisely $\zeta(s)$.

Let $F(s)$, $G(s)$, and $H(s)$ be the generating functions of f, g, and $h=f*g$. Let f and g be multiplicative, so that h is also multiplicative. Moreover, if $F(s)$ and $G(s)$ are absolutely convergent for $\sigma\geqslant\sigma_0$, then $H(s)$ is also absolutely convergent for $\sigma\geqslant\sigma_0$, and $F(s)G(s)=H(s)$ for $\sigma\geqslant\sigma_0$. It follows from this that $\sum_{n=1}^{\infty}\mu(n)n^{-s}=\zeta^{-1}(s)$, $\sum_{n=1}^{\infty}\mu(n)\chi(n)n^{-s}=L(s,\chi)^{-1}$, $\sum_{n=1}^{\infty}d_k(n)n^{-s}=\zeta(s)^k$, and $\sum_{n=1}^{\infty}d^2(n)n^{-s}=\zeta^4(s)/\zeta(2s)$. The last result was given by S. Ramanujan. More generally, $\sum_{n=1}^{\infty}d^r(n)n^{-s}=\zeta^{2^r}(s)\varphi(s)$, where $\varphi(s)$ is an analytic function for $\sigma>1/2$. By utilizing analytic methods, we can deduce from this that $\sum_{n\leqslant x}d^r(n)\sim x(c_1(\log x)^{2^r-1}+\ldots+c_{2^r})$ (B. M. Wilson, 1923) (→ 123 Distribution of Prime Numbers; 242 Lattice-Point Problems).

There are many formulas known as the **Euler summation formula**. Among them the following one is convenient to use (E. Landau, L. J. Mordell, H. Davenport): Let $f_0(x)=x-[x]-1/2$ (when x is not an integer), $=0$ (when x is an integer). We successively construct continuous functions $f_1(x), f_2(x), \ldots$ of period 1 such that $f_r'(x)=f_{r-1}(x)$ $(r\geqslant 1)$ if x is not an integer, and $\int_0^1 f_r(x)\,dx=0$. If $F(x)$ is of class C^h on $[a,b]$, then using these auxiliary functions, we have

$$\sum_{a\leqslant m\leqslant b}{}' F(m)=\int_a^b F(x)\,dx + \sum_{r=0}^{h-1}(-1)^r(f_r(a)F^{(r)}(a) - f_r(b)F^{(r)}(b)) + (-1)^{h-1}\int_a^b f_{h-1}(x)F^{(h)}(x)\,dx,$$

where $\sum'$ means that the term corresponding to $m=a$ or $m=b$ in the sum is to be replaced by $F(a)/2$ or $F(b)/2$ whenever a or b is an integer. Since the †Fourier series of $f_0(x)$ is

$$-\sum_{n=1}^{\infty}\frac{\sin(2\pi nx)}{n\pi}$$

(which is convergent), $f_r(x)$ can be expressed by

$$-\sum_{n=-\infty}^{\infty}{}' \frac{\exp(2\pi inx)}{(2\pi in)^{r+1}},$$

where $\sum'$ means that the term with $n=0$ is omitted and this sum actually means

$$\sum_{n=1}^{\infty}\left(\frac{\exp(2\pi inx)}{(2\pi in)^{r+1}}+\frac{\exp(-2\pi inx)}{(-2\pi in)^{r+1}}\right).$$

Hence if $h=1$, we have

$$\sum_{a\leqslant n\leqslant b}{}' F(n)=\int_a^b F(x)dx-\frac{1}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\int_a^b F'(x)\sin(2\pi nx)dx.$$

For instance, if we put $h=1$, $a=1$, $b=N$, $F(x)=x^{-s}$ and let N tend to ∞, then we have the formula

$$\zeta(s)=\sum_{n=1}^{\infty} n^{-s}=\frac{1}{s-1}+\frac{1}{2}-s\int_1^{\infty}\frac{f_0(x)}{x^{s+1}}dx\quad\text{for}\quad\sigma>1.$$

Utilizing the following integration by parts, the integral on the right-hand side can be extended so as to become holomorphic in s in the whole complex plane:

$$\int_1^{\infty}\frac{f_0(x)}{x^{s+1}}dx=-f_1(1)+(s+1)\int_1^{\infty}\frac{f_1(x)}{x^{s+2}}dx=\ldots.$$

Probabilistic considerations are also used for the study of various number-theoretic functions. If $f(n)$ is $\omega(n)$ or $\Omega(n)$, then we have $\sum_{n\leqslant x}f(n)=x\log\log x+cx+o(x)$. Therefore the average order of $\omega(n)$ or $\Omega(n)$ is estimated as $\log\log n$. Let $A(x;\alpha,\beta)$ be the number of n satisfying $n\leqslant x$ and $\log\log n+\alpha\sqrt{\log\log n}<f(n)<\log\log n+\beta\sqrt{\log\log n}$. Then

$$\lim_{x\to\infty}\frac{A(x;\alpha,\beta)}{x}=\frac{1}{\sqrt{2\pi}}\int_{\alpha}^{\beta}e^{-u^2/2}du.$$

For $f(n)=\omega(n)$, the result was proved by P. Erdös and M. Kac (1940) by using the †central limit theorem and V. Brun's †sieve method. Further general formulas were obtained by M. Tanaka (1955). Further development and the present state of probabilistic number theory can be seen in [7, 8].

Finally, we mention the well-known estimation formulas. Let ε be an arbitrary positive number and n be a positive integer. Then $d(n)=O(n^{\varepsilon})$, where O (†Landau's symbol) depends on ε. We know that $\limsup_{n\to\infty}(\log d(n))\cdot(\log\log n)/\log n=\log 2$ (S. Wigert, 1907), and $\liminf_{n\to\infty}\varphi(n)\cdot(\log\log n)/n=e^{-C}$. The result $\omega(n)=O(\log n/\log\log n)$ is often used. Let χ be a primitive character modulo k, and let $S_m=\sum_{n=1}^{m}\chi(n)$. We can prove that $|S_m|<\sqrt{k}\log k$ for all m (I. Schur, 1918) and $S_m=\Omega(\sqrt{k}\log\log k)$ (R. Paley, 1932). This formula, with the symbol Ω, means that there exist infinitely many k satisfying $|S_m|>c\sqrt{k}\log\log k$ by taking m and χ suitably, with c any positive constant. Let χ be a character modulo k, and χ^0 be a primitive character associated with χ. Then we have $L(s,\chi)=L(s,\chi^0)\prod_{p|k}(1-\chi^0(p)p^{-s})$. It follows that

$$\sum_{m=1}^{n}\chi(m)=\sum_{d;d|k,d\nmid f}\mu(d)\chi^0(d)\sum_{m\leqslant n/d}\chi^0(m)$$

if the conductor of χ^0 is f. A. G. Postnikov investigated (1956) the sum of characters. The least integer that is not a quadratic residue modulo p does not exceed $p^{1/2\sqrt{e}}(\log p)^2$, where p is a sufficiently large prime (I. M. Vinogradov, 1926). The least integer that is a †primitive root modulo p does not exceed $2^{m+1}\sqrt{p}$, where m is the number of distinct prime divisors of $p-1$, with p a prime (L. K. Hua, 1942). D. A. Burgess (1962) deals with the latter results. We conclude with Artin's conjecture: If w is a given square-free integer, then there exist infinitely many primes p such that w is a primitive root modulo p. In 1967, C. Hooley proved this conjecture subject to the assumption that the general †Riemann hypothesis holds for †Dedekind zeta functions over certain †Kummer extensions of $\mathbf{Q}$. He also obtained the asymptotic formula for the number of such primes not exceeding x. Recently the †sieve method has been widely applied to the various investigations in the theory of number theoretic functions (→ 123 Distribution of Prime Numbers E).

References

[1] R. G. Ayoub, An introduction to the analytic theory of numbers, Amer. Math. Soc. Math. Surveys, 1963.
[2] A. O. Gel'fond and Yu. V. Linnik, Elementary methods in analytic number theory, Rand McNally, 1965. (Original in Russian, 1962.)
[3] G. H. Hardy and E. M. Wright, An introduction to the theory of numbers, Clarendon Press, fourth edition, 1965.
[4] H. Hasse, Vorlesungen über Zahlentheorie, Springer, 1950.
[5] E. G. H. Landau, Vorlesungen über Zahlentheorie I, Hirzel, 1927 (Chelsea, 1946).
[6] W. J. Leveque, Topics in number theory, Addison-Wesley, 1956.
[7] J. Kubilius, Probabilistic methods in the theory of numbers, Amer. Math. Soc. Transl. Math. Monographs, 1964. (Original in Russian, 1962.)
[8] P. D. T. A. Elliott, Probabilistic number theory I, II, Springer, 1979.

296 (V.1) Number Theory

A. History

Simple and curious relations among integers were discovered and admired from antiquity. For example, we have the relation $3^2+4^2=5^2$, which has geometric meaning concerning right triangles. The Pythagoreans (→ 187 Greek Mathematics) sought similar relations. They were also interested in †perfect numbers (numbers equal to the sum of their divisors, such as $28=1+2+4+7+14$). Modern arithmetic inherits from the Greeks the proof of the existence of an infinite number of primes, the †Euclidean algorithm to obtain the greatest common divisor of two integers (both given in Euclid's *Elements*), and Eratosthenes' †sieve for finding primes. In the 3rd century A.D., Diophantus of Alexandria discovered a method of solving indeterminate equations of degrees 1 and 2; this marked the origin of **Diophantine analysis**. The ancient Chinese also knew how to solve equations of the first degree in some special cases. Arithmetic also developed in India from an early period; its relation to Greek mathematics is not yet entirely clear. In the 12th century, the Indian mathematician Bhâskara knew how to solve †Pell equations using a method much like Lagrange's.

Interest in integers was reborn in Europe in the 17th century. During this period, Bachet de Méziriac (1581–1638) rediscovered the solution of the †Diophantine equation of the first degree and published it in his famous book on mathematical recreations [1]. Primes of the form 2^p-1, closely related to perfect numbers and called †Mersenne numbers, attracted considerable interest. †Fermat, sometimes called the father of number theory, announced numerous results without giving proofs; the most famous among them is the so-called Fermat's last theorem (→ 145 Fermat's Problem). Another famous conjecture of his is that every integer is expressible as the sum of at most n **n-gonal numbers**, i.e., numbers of the form $k(k-1)n/2-k(k-2)$, $k\in\mathbf{N}$. This was proved by Gauss (for the case $n=3$), Jacobi ($n=4$), and Cauchy (for the general case).

In the 18th century, remarkable progress was made by Euler and Lagrange. The second part of Euler's *Algebra* [2] contains rich results of miscellaneous sorts in the field. Lagrange developed the theory of †continued fractions and applied it to arithmetic. Toward the end of the 18th century, Legendre compiled his comprehensive book [3], from whose title originates the term *number theory*.

Gauss's *Disquisitiones* [4] appeared at about the same time as Legendre's book. The theoretical arithmetic of today originates from this work of Gauss. The book includes the theories of †quadratic residues, †quadratic forms, and **cyclotomy** (i.e., arithmetic theory of the roots of unity in the field of complex numbers), all of which appeared as well-developed theories of a remarkably high standard. The work was received with more respect than comprehension. Dirichlet made a lifelong effort to popularize the *Disquisitiones*; he also applied analytical methods to compute the †class number of quadratic forms, thus giving number theory a new direction, the analytic theory of numbers. Gauss treated only binary quadratic forms; Eisenstein, Minkowski, and Siegel generalized the theory to the case of n variables. The algebraic theory of numbers has its origin in Gauss's paper on biquadratic residues. (→ 14 Algebraic Number Fields; 59 Class Field Theory; 118 Diophantine Equations; 182 Geometry of Numbers; 297 Number Theory, Elementary; 347 Quadratic Fields; 430 Transcendental Numbers.)

B. Analytic Methods

Analytic methods are sometimes used to solve arithmetic problems. For example, Legendre conjectured that any arithmetic progression of integers a, $a+d$, $a+2d,\ldots$ contains an infinite number of primes if a and d are relatively prime. The conjecture was first proved by Dirichlet in 1837; in the proof he used an †L-function. Recently, proofs that do not use L-functions have been obtained, although these proofs are still not purely arithmetical. Analysis is indispensable in the formulation of some arithmetic problems. For example, let $x\in\mathbf{R}$ and $\pi(x)$ be the number of primes not exceeding x. Euclid's *Elements* already give the result $\pi(x)\to\infty$ as $x\to\infty$, but to describe the behavior of $\pi(x)$ as $x\to\infty$, notions of analysis are needed. Gauss conjectured that

$$\lim_{x\to\infty}\frac{\pi(x)}{x/\log x}=1.$$

This †prime number theorem was first proved in 1896, using the results of the theory of functions of a complex variable; more elementary proofs have been obtained in recent years.

Analysis is sometimes needed to solve or simplify certain problems in number theory. The branch of mathematics treating such problems is called **analytic number theory**. The question of the extent to which analysis is really needed in dealing with such problems is itself an interesting one.

In the 20th century, analytic number theory has made rapid progress. The problem of distribution of primes has been generalized to the case of algebraic number fields. †Additive number theory dealing with †Waring's problem, †Goldbach's problem, and other problems has developed and formed a new field. We also have **geometric number theory**, which deals with †lattice-point problems. (→ 4 Additive Number Theory; 123 Distribution of Prime Numbers; 242 Lattice-Point Problems; 295 Number-Theoretic Functions; 328 Partitions of Numbers; 450 Zeta Functions.)

References

[1] C. G. Bachet de Méziriac, Problèmes plaisants et délectables qui se font par les numbres, Lyon, 1612.
[2] L. Euler, Vollständige Anleitung zur Algebra, Petersburg, 1770.
[3] A. M. Legendre, Essai sur la théorie des nombres, Paris, 1798.
[4] K. F. Gauss, Disquisitiones arithmeticae, Leipzig, 1801; English translation, Yale Univ. Press, 1966.
[5] M. B. Cantor, Vorlesungen über Geschichte der Mathematik, Teubner, 1894–1908.
[6] L. E. Dickson, History of the theory of numbers I, II, III, Carnegie Institution of Washington, 1919–1923 (Chelsea, 1952).
[7] P. G. L. Dirichlet, Vorlesungen über Zahlentheorie, herausgegeben und mit Zusätzen versehen von R. Dedekind, Braunschweig, fourth edition, 1894.
[8] J. LeVeque, Reviews of papers in number theory, vols. 1–6, Amer. Math. Soc., 1974.

297 (V.2) Number Theory, Elementary

A. The Euclidean Algorithm

We denote the set of natural numbers 1, 2, 3, ... by **N** and the set of rational integers 0, ±1, ±2, ... by **Z**. Evidently **Z** is an ordered †commutative ring and also an †integral domain with respect to ordinary addition and multiplication. For any $a \in \mathbf{Z}$ and $b \in \mathbf{N}$, there exists a unique pair of integers q and r such that $a = qb + r$ $(0 \leqslant r < b)$ (**division algorithm**); q is called the **quotient** and r the **remainder** of the division of a by b. When the remainder r is zero, we say that a is a **multiple** of b, b is a **divisor** of a, and a is **divisible** by b. We denote this relation by $b \mid a$. After a finite number of divisions $a = q_1 b + r_1$, $b = q_2 r_1 + r_2$, $r_1 = q_3 r_2 + r_3, \ldots$ $(b > r_1 > r_2 \ldots > 0)$, we reach an equation of the form $r_k = q_{k+2} r_{k+1}$. The remainder $r = r_{k+1}$, uniquely determined in this manner by a and b, is the **greatest common divisor (G. C. D.)** of a and b. It can be expressed as $r = ax + by$ with integers x and y. This method of obtaining the G. C. D. is called the **Euclidean algorithm**. The greatest common divisor of a and b is denoted by (a, b). If an integer c divides both a and b, then $c \mid (a, b)$. When $(a, b) = 1$, we say that a and b are **relatively prime**, and there are integral solutions x, y of the equation $ax + by = 1$. Given a pair of positive integers a and b, there exists a unique positive integer c that divides any common multiple of a and b, called the **least common multiple (L. C. M.)** of a and b. If d and l denote the G. C. D. and L. C. M. of a and b, respectively, we have $ab = dl$.

B. Prime Numbers

An integer p is called a **prime number** if p is larger than 1 and has no positive divisors other than 1 and itself. A positive integer is called a **composite number** if it has positive divisors other than 1 and itself. The following method of selecting the prime numbers from the sequence 2, 3, 4, 5, ..., known to the Greeks, is called **Eratosthenes' sieve**: First we discard multiples of two, thus reaching the second prime number, 3. Then we discard multiples of three and reach the next prime, 5. Continuing this process, we find the sequence of primes 2, 3, 5, 7, 11, ... by sifting out all multiples.

C. Decomposition into Primes Factors

Every integer $n > 1$ can be uniquely expressed as the product of primes; the resulting decomposition can be written as $n = p^\alpha q^\beta r^\gamma \ldots$ with distinct prime factors $p, q, r, \ldots$ and corresponding exponents $\alpha, \beta, \gamma, \ldots$ uniquely determined by n. This theorem is called the **fundamental theorem of elementary number theory**.

D. Perfect Numbers

We denote by $\sigma(n)$ the sum of positive divisors of n (including 1 and n itself). According as $\sigma(n)$ is greater than, equal to, or less than $2n$, we call n an **abundant number, perfect number**, or **deficient number**. An even number is perfect if and only if it can be represented as $n = 2^{e-1} \cdot (2^e - 1)$ with prime $2^e - 1$ (L. Euler). The existence of an odd perfect number still constitutes an open question. It is well known that, if n is odd and perfect, then n must be of the

form $n=p_0^{a_0}p_1^{a_1}\dots p_t^{a_t}$, where $p_0\equiv a_0\equiv 1$ (mod 4), and a_i is even for $i>0$. Recently it has been proved that t must be $\geqslant 7$ (P. H. Hagis, Jr., *Math. Comp.*, 35 (1980)). Many necessary conditions for an odd number to be perfect have been given and repeatedly improved. It has been recently proved that there exists no odd perfect number less than 10^{50} (Hagis, *Math. Comp.*, 27 (1973)).

Two numbers m and n are said to form a pair of **amicable numbers** if $\sigma(n)-n=m$ and $\sigma(m)-m=n$ (e.g., $m=220$ and $n=284$). Euler listed 61 such pairs. The following two numbers m, n, both having 152 digits, constitute the largest amicable pair currently known:

$$m=3^4\cdot 5\cdot 11\cdot 5281^{19}\cdot 29\cdot 89(2\cdot 1291\cdot 5281^{19}-1),$$

$$n=3^4\cdot 5\cdot 11\cdot 5281^{19}(2^3\cdot 3^3\cdot 5^2\cdot 1291\cdot 5281^{19}-1)$$

(H. J. J. te Riele, *Math. Comp.*, 28 (1974)).

E. Lionnet considered the numbers n such that the product $\prod_{d\mid n} d$ is equal to n^2 and called such numbers **perfect numbers of the second kind**, e.g., $n=p^3$, $n=pp'$, where p and p' are unequal prime numbers. It is also known that there are numbers n such that $\sigma(n)/n$ is an integer. For example, $2^5\cdot 3\cdot 7$, $2^{15}\cdot 3^5\cdot 5^2\cdot 7^2\cdot 11\cdot 13\cdot 17\cdot 19\cdot 31\cdot 43\cdot 257$ have this property.

E. Mersenne Numbers

A number of the form 2^e-1, where e is a prime, is called a **Mersenne number**. For the number 2^e-1 to be prime, it is necessary but not sufficient that e be prime. If a Mersenne number is a prime, it is called a **Mersenne prime**. It is not known whether there are infinitely many Mersenne primes; it has been verified, however, that for $e<44500$, there are 27 cases of e such that 2^e-1 is a Mersenne prime; $e=2$, 3, 5, 7, 13, 17, 19, 31, 61, 89, 107, 127, 521, 607, 1279, 2203, 2281, 3217, 4253, 4423, 9689, 9941, 11213, 21701, 23209, and 44497. Until the 18th century, the verifications for $e\leqslant 31$ were done by direct calculation. For $61\leqslant e\leqslant 127$, the "Lucas test" was utilized to execute the computation. The remaining cases were calculated by means of electronic computers. The number $2^{44497}-1$, which has 13395 digits, is at present the largest known prime.

F. Fermat Numbers

Numbers of the form $2^{2^\nu}+1$ are called **Fermat numbers**. For a number $p=2^e+1$ to be a prime, it is necessary that e be a power of 2. Fermat conjectured that numbers of the form $2^{2^\nu}+1$ are all primes. In fact, for $\nu=0$, 1, 2, 3, 4, the corresponding $p=3$, 5, 17, 257, 65537 are primes; however, $2^{2^5}+1$ is divisible by 641. It is not yet known whether there exist Fermat primes other than these five primes. For further details $\rightarrow$ [10]. Fermat numbers are closely connected with the problem of †geometric construction of regular polygons.

G. Congruence

Let m be a positive integer. Two integers a and b are said to be **congruent modulo** m if their difference is divisible by m; we denote this relation by the **congruence** $a\equiv b$ (mod m) (or simply by $a\equiv b$ (m)) and call m the **modulus** of this congruence. Congruence modulo m is an †equivalence relation, which is compatible with the ring operations and classifies $\mathbf{Z}$ into m classes. We thus obtain the †residue class ring $\mathbf{Z}/m\mathbf{Z}$ with m elements. If p is a prime, then $\mathbf{Z}/p\mathbf{Z}$ is a field which is isomorphic to the †prime field with †characteristic p. A complete system of representatives of the quotient set $\mathbf{Z}/m\mathbf{Z}$ is called a **complete residue system modulo** m. On the other hand, a set of $\varphi(m)$ elements n_i (φ is †Euler's function) such that $n_i\not\equiv n_j$ (m) $(i\neq j)$ and $(n_i, m)=1$ is called a **reduced residue system modulo** m. The set of all residue classes represented by a reduced residue system modulo m forms a multiplicative Abelian group of order $\varphi(m)$; we denote this group by $(\mathbf{Z}/m\mathbf{Z})^*$. If $(a,m)=1$, then $a^{\varphi(m)}\equiv 1$ (mod m). If p is a prime and $(a,p)=1$, then $a^{p-1}\equiv 1$ (mod p) (**Fermat's theorem**), since $\varphi(p)=p-1$. When m is 2, 4, p^k or $2p^k$ ($p\neq 2$; $k=1,2,\dots$), $(\mathbf{Z}/m\mathbf{Z})^*$ is a †cyclic group, whose †generator g is called a **primitive root modulo** m ($\rightarrow$ Appendix B, Table 1). For any a prime to m and generator g of $(\mathbf{Z}/m\mathbf{Z})^*$, there exists a unique number μ ($1\leqslant\mu\leqslant\varphi(m)$) such that $a\equiv g^\mu$ (mod m). We call μ the **index** of a with respect to the basis g, and write $\mu=\mathbf{Ind}_g a$ ($\rightarrow$ Appendix B, Table 2). The group $(\mathbf{Z}/2^k\mathbf{Z})^*$ ($k\geqslant 3$) is Abelian of type $(2, 2^{k-2})$, a basis of which is formed by the residue classes represented by -1 and 5. From this follows $(p-1)!\equiv -1$ (mod p) (**Wilson's theorem**). Generally, if $m=\prod_{i=1}^r m_i$, $(m_i, m_j)=1$ for $i\neq j$, then we have $\mathbf{Z}/m\mathbf{Z}\cong\mathbf{Z}/m_1\mathbf{Z}+\dots+\mathbf{Z}/m_r\mathbf{Z}$ (direct sum) and $(\mathbf{Z}/m\mathbf{Z})^*\cong(\mathbf{Z}/m_1\mathbf{Z})^*\times\dots\times(\mathbf{Z}/m_r\mathbf{Z})^*$ (direct product).

The congruence $ax\equiv b$ (mod m) with $(a,m)=d$ is solvable if and only if $d\mid b$, and when it is solvable, the solution is unique modulo m/d. The simultaneous congruences $x\equiv a_i$ (mod m_i) $(i=1,2,\dots,k)$ are solvable if and only if $a_i\equiv a_j$ (mod(m_i, m_j)) $(i,j=1,\dots,k)$, and when they are solvable, the solution is unique modulo the

greatest common multiple of $m_1, \dots, m_k$. In particular, when $(m_i, m_j) = 1$ $(i \neq j)$, the solution is unique modulo $m_1 m_2 \dots m_k$ (**Chinese remainder theorem**). If $m = p_1^{e_1} \dots p_k^{e_k}$ is the factorization of m, solving the congruence $f(x) \equiv 0$ $(\mathrm{mod}\, m)$, where $f(x)$ is a polynomial with integral coefficients, can be reduced to solving $f(x) \equiv 0$ $(\mathrm{mod}\, p_i^{\alpha_i})$ $(i = 1, \dots, k)$. Also, solving a quadratic congruence can be reduced to solving a linear congruence and a congruence of the form $x^2 \equiv a$ $(\mathrm{mod}\, m)$.

H. Quadratic Residues

When the congruence $x^2 \equiv a$ $(\mathrm{mod}\, m)$, where $(a, m) = 1$, is solvable, a is said to be a **quadratic residue** modulo m; otherwise, a is said to be a **quadratic nonresidue** modulo m. The following two conditions are necessary and sufficient for a to be a quadratic residue modulo m: (i) a is a quadratic residue with respect to each of the prime factors p $(\neq 2)$; and (ii) $a \equiv 1$ $(\mathrm{mod}\, 4)$ or $a \equiv 1$ $(\mathrm{mod}\, 8)$ according as $4 | m$ or $8 | m$.

Given a prime number p and integer a prime to p, the **Legendre symbol** (a/p) is by definition 1 or -1 according as a is a quadratic residue modulo p or not. The value of this symbol is determined by a $(\mathrm{mod}\, p)$, and we have $(ab/p) = (a/p)(b/p)$. Hence the symbol determines a †character of $(\mathbf{Z}/p\mathbf{Z})^*$ of order 2. Furthermore, the congruence $(a/p) \equiv a^{(p-1)/2}$ $(\mathrm{mod}\, p)$ holds (**Euler's criterion**).

I. Reciprocity Law

For odd primes p and q $(p \neq q)$, the formulas

$$(p/q)(q/p) = (-1)^{((p-1)/2)((q-1)/2)},$$

$$(-1/p) = (-1)^{(p-1)/2}, \quad (2/p) = (-1)^{(p^2-1)/8}$$

are called the **law of quadratic reciprocity of the Legendre symbol**, the **first complementary law**, and the **second complementary law**, respectively. These laws were conjectured by Euler and first proved by Gauss, who gave seven different proofs. We now have more than fifty different proofs of these laws. P. Bachmann (*Niedere Zahlentheorie* I (1902)) lists the 47 different proofs of the laws known at the time. Gauss's first proof was elementary; his second proof used quadratic forms. The latter has been reformulated utilizing the theory of quadratic fields [4]. The fourth proof used †Gaussian sums, and the sixth proof used algebraic congruences with integral coefficients. His seventh proof, contained in his posthumous works, used congruences of higher degree [4]. His fourth, sixth, and seventh proofs are related to the arithmetic of †cyclotomic fields [4]. The third and fifth proofs are the most elementary and simple. They are based on Gauss's lemma: Let $r_1, r_2, \dots, r_{(p-1)/2}$ be the residues of divisions of $1a, 2a, \dots, (p-1)a/2$ by an odd prime p, and let n be the number of these residues that are greater than $p/2$. Then we have $(a/p) = (-1)^n$. T. Takagi gave a simplified exposition of the third proof using geometric figures (1904). The same method was rediscovered by G. Frobenius (1914).

When m is an odd integer such that $m = \pm \prod_i p_i^{e_i}$, $(m, n) = 1$, we call $(n/m) = \prod_i (n/p_i)^{e_i}$ **Jacobi's symbol**. If m has no square factor, it is a character of $(\mathbf{Z}/m\mathbf{Z})^*$. If we put $\mathrm{sgn}\, m = +1$ or -1 according as $m > 0$ or $m < 0$, we have the following **law of quadratic reciprocity of Jacobi's symbol** and its **complementary laws**:

If m and n are relatively prime odd numbers, then

$$(m/n)(n/m)$$

$$= (-1)^{((m-1)/2)((n-1)/2) + ((\mathrm{sgn}\, m - 1)/2)((\mathrm{sgn}\, n - 1)/2)},$$

$$(-1/n) = (-1)^{(n-1)/2 + (\mathrm{sgn}\, n - 1)/2},$$

$$(2/n) = (-1)^{(n^*-1)/4},$$

where

$$n^* = (-1)^{(n-1)/2} n.$$

Furthermore, †Kronecker's symbol, another generalization of the Legendre symbol, is used in the theory of quadratic number fields (→ 347 Quadratic Fields).

References

[1] C. F. Gauss, Werke 2, Königliche Gesellschaft der Wissenschaften, Göttingen, 1863.
[2] C. F. Gauss, Sechs Beweise des Fundamentaltheorems über quadratische Reste, Ostwald's Klassiker Nr. 122, Wilhelm Engelmann, 1901.
[3] L. E. Dickson, History of the theory of numbers I, Carnegie Institution of Washington, 1919 (Chelsea, 1952).
[4] P. G. L. Dirichlet, Vorlesungen über Zahlentheorie, Herausgegeben und mit Zusätzen versehen von R. Dedekind, Vieweg, fourth edition, 1894 (Chelsea, 1969).
[5] H. Hasse, Vorlesungen über Zahlentheorie, Springer, 1950.
[6] E. G. H. Landau, Vorlesungen über Zahlentheorie I, Hirzel, 1927 (Chelsea, 1946).
[7] W. J. LeVeque, Topics in number theory, Addison-Wesley, 1956.
[8] H. Rademacher, Lectures on elementary number theory, Blaisdell, 1964.
[9] Z. I. Borevich and I. R. Shafarevich, Number theory, Academic Press, 1966. (Original in Russian, 1964.)

[10] W. Sierpiński, Elementary theory of numbers, Hafner, 1964.

298 (XV.6) Numerical Computation of Eigenvalues

A. General Remarks

Numerical computation of †eigenvalues and †eigenvectors of a matrix provides a basic technique for the numerical solution of various eigenvalue problems. Roughly speaking, there are two kinds of method. In methods of the first kind, one determines the †characteristic polynomial $p(\lambda)=\det(\lambda I-A)$ (where I is the unit matrix) of A (or to give an algorithm to calculate the value of $p(\lambda)$ for an arbitrary λ), then solves the algebraic equation $p(\lambda)=0$ numerically to obtain the eigenvalues λ_μ ($\mu=1,2,\dots$) (→ 301 Numerical Solution of Algebraic Equations), and finally to one determines the eigenvectors x_μ by means of the equations $(\lambda_\mu I-A)x_\mu=0$. In methods of the second kind, one obtains eigenvalues and eigenvectors directly without resorting to the solution of an algebraic equation, relying instead upon repeated application of similarity transformations. (The power method does not fit into this classification, being a different approach altogether; → Section C.) In particular, a real symmetric or complex Hermitian matrix is reduced approximately to a diagonal matrix. In this article we denote a given square matrix of order n by $A=(a_{ij})$ ($i,j=1,\dots,n$).

B. The Jacobi Method

The Jacobi method is an iterative technique for determining all the eigenvalues and eigenvectors of a real symmetric matrix [7]. Before the advent of high-speed computers it was not considered practical, but at present it is one of the most compact and elegant methods. The following algorithm can be extended to Hermitian matrices by replacing †orthogonal transformations by suitable †unitary transformations.

Roughly, the method transforms a given matrix $A=(a_{ij})$ ($a_{ji}=a_{ij}$; $i,j=1,\dots,n$) into a diagonal one by repeated application of 2-dimensional rotations of the reference axes. We first put $A^{(0)}=A$, $U^{(0)}=I$, and compute $A^{(1)}, A^{(2)},\dots, U^{(1)}, U^{(2)},\dots$ successively as follows:

(1) Select an off-diagonal element of $A^{(\iota)}=(a_{ij}^{(\iota)})$ with the maximum absolute value and denote it by $a_{pq}^{(\iota)}$.

(2) Compute

$$\tan\theta=\frac{2a_{pq}^{(\iota)}\operatorname{sgn}(a_{pp}^{(\iota)}-a_{qq}^{(\iota)})}{|a_{pp}^{(\iota)}-a_{qq}^{(\iota)}|+\sqrt{(a_{pp}^{(\iota)}-a_{qq}^{(\iota)})^2+4(a_{pq}^{(\iota)})^2}}$$

(where $\operatorname{sgn}x=1$, 0, or -1 according as $x>0$, $=0$, or <0),

$$\cos\theta=(1+\tan^2\theta)^{-1/2},\quad\text{and}$$

$$\sin\theta=\tan\theta\cdot\cos\theta;$$

and form $T^{(\iota)}=(t_{ij}^{(\iota)})$, where $t_{pp}^{(\iota)}=t_{qq}^{(\iota)}=\cos\theta$, $t_{ii}^{(\iota)}=1$ for $i\neq p,q$, $-t_{pq}^{(\iota)}=t_{qp}^{(\iota)}=\sin\theta$, and $t_{ij}^{(\iota)}=0$ for all other (i,j).

(3) Determine $A^{(\iota+1)}$ and $U^{(\iota+1)}$ by $A^{(\iota+1)}=T^{(\iota)\prime}A^{(\iota)}T^{(\iota)}$ ($'$ denotes the transpose) and $U^{(\iota+1)}=U^{(\iota)}T^{(\iota)}$. In this process, $T^{(\iota)}$ represents an orthogonal transformation (rotation) in the plane spanned by the pth and qth coordinate axes such that $a_{pq}^{(\iota+1)}=a_{qp}^{(\iota+1)}$ is nullified. If we put $N(B)=\sum_{i,j}b_{ij}^2$ and $M(B)=\sum_{i\neq j}b_{ij}^2$, then $N(B)$ is invariant under an orthogonal transformation, so that $N(A^{(\iota)})=N(A)$. Furthermore, since $a_{ii}^{(\iota+1)}=a_{ii}^{(\iota)}$ ($i\neq p,q$) and $(a_{pp}^{(\iota+1)})^2+(a_{qq}^{(\iota+1)})^2=(a_{pp}^{(\iota)})^2+(a_{qq}^{(\iota)})^2+2(a_{pq}^{(\iota)})^2$, we have $M(A^{(\iota+1)})=M(A^{(\iota)})-2(a_{pq}^{(\iota)})^2$. Since $a_{pq}^{(\iota)}$ has the maximum absolute value among all the off-diagonal elements, we have $(a_{pq}^{(\iota)})^2\geqslant M(A^{(\iota)})/(n^2-n)$. Therefore

$$\begin{aligned}M(A^{(\iota+1)})&\leqslant(1-2/(n^2-n))M(A^{(\iota)})\\&\leqslant(1-2/(n^2-n))^{\iota+1}M(A)\\&<M(A)\exp(-2(\iota+1)/(n^2-n)).\end{aligned}$$

It has been proved a fortiori [13] that, after $M(A^{(\iota)})$ comes down to below a certain threshold value, the convergence of the iteration process becomes quadratic, i.e., there is a number c determined by the order n of A and the arrangement of the eigenvalues of A such that $M(A^{(\iota+n(n-1)/2)})<c(M(A^{(\iota)}))^2$. Since the set of eigenvalues of $A^{(\iota)}$ coincides with that of A and the eigenvalues of an arbitrary symmetric matrix B can be made to correspond in a one-to-one way to its diagonal elements in such a way that the difference between an eigenvalue and the corresponding diagonal element is not greater than $M(B)^{1/2}$, $a_{ii}^{(\iota)}$ tends to λ_i ($i=1,\dots,n$) as ι tends to infinity. Moreover, as ι tends to infinity, each column vector of $U^{(\iota)}=(u_{ij}^{(\iota)})$ tends to the corresponding eigenvector, in the sense that $\sum_{k=1}^n a_{ik}u_{kj}^{(\iota)}-\lambda_j u_j^{(\iota)}\to0$.

The number of arithmetic computations required to obtain $A^{(\iota+1)}$ and $U^{(\iota+1)}$ from $A^{(\iota)}$ and $U^{(\iota)}$ is at most proportional to n, so that for a given $\varepsilon>0$, the arithmetic required to reduce $\max_i|a_{ii}^{(\iota)}-\lambda_i|$ below $\varepsilon(M(A))^{1/2}$ is at most proportional to n^3 (because ι is at most

proportional to n^2). On the other hand, the search for an off-diagonal element of $A^{(\iota)}$ with the maximum absolute value, if it is done by simply comparing all the elements, will require effort proportional to n^2, so that the amount of work required by the searching process is proportional to n^4. To bypass this searching process, the **cyclic Jacobi method** and the **threshold Jacobi method** are often used. The former method adopts as $a_{pq}^{(\iota)}$ the off-diagonal element for which $q>p$ and $\iota=(p-1)(n-p/2)+(q-p)$, that is, $a_{12}^{(1)}, a_{13}^{(2)}, \ldots, a_{1n}^{(n-1)}, a_{23}^{(n)}, a_{24}^{(n+1)}, \ldots$ are adopted in this sequence. The latter method adopts as $a_{pq}^{(\iota)}$ off-diagonal elements in a sequence similar to the one above as long as they exceed a given threshold value; but if an element is less than that threshold value, then the next element in the sequence is a candidate for adoption as $a_{pq}^{(\iota)}$, where the threshold value is made to decrease gradually as the iteration process proceeds. However, the search for an element with the maximum absolute value can be done more effectively by taking account of the fact that only the matrix elements lying in rows p and q and in columns p and q change their values when we transform $A^{(\iota)}$ into $A^{(\iota+1)}$. In fact, we can record for each row the value as well as the position of the (off-diagonal) element with the maximum absolute value in that row. By so doing, the effort of searching for an off-diagonal element with the maximum absolute value can be reduced to something proportional to n on the average.

C. The Power Method

The **power method** is suitable for obtaining only the eigenvalue of maximum absolute value [6]. Let us assume that the eigenvalues $\lambda_1, \ldots, \lambda_n$ of A are arranged so that $|\lambda_1|>|\lambda_2|\geqslant|\lambda_3|\geqslant\ldots\geqslant|\lambda_n|$, with λ_1 real, and denote by y_1 the left eigenvector corresponding to λ_1 (which means $y_1(\lambda_1 I-A)=0$). Starting from an arbitrary (real) vector $x^{(0)}$ such that $(y_1, x^{(0)})\neq 0$ and $\chi_{i_0}^{(0)}=1$ for a prescribed i_0, we compute $\theta^{(0)}, \theta^{(1)}, \ldots$ and $x^{(1)}, x^{(2)}, \ldots$ by $Ax^{(\iota)}=\theta^{(\iota)}x^{(\iota+1)}$ $(\iota=0,1,2,\ldots; x_{i_0}^{(\iota+1)}=1)$. Then we have $\lim_{\iota\to\infty}\theta^{(\iota)}=\lambda_1$ and $\lim_{\iota\to\infty}x^{(\iota)}=x_1$ (the eigenvector corresponding to the eigenvalue λ_1). The rate of convergence depends on $|\lambda_1/\lambda_2|$ if the †elementary divisor of A corresponding to λ_1 is linear, but in the case of a nonlinear elementary divisor, the convergence is too slow for practical purposes. If $\lambda_1\neq\lambda_2$ and $|\lambda_1|=|\lambda_2|>|\lambda_3|\geqslant\ldots\geqslant|\lambda_n|$, then the sequences of $\theta^{(\iota)}$ and $x^{(\iota)}$ computed by the formulas above do not converge but in general oscillate. However, from $\theta^{(\iota)}, \theta^{(\iota+1)}, x^{(\iota)}, x^{(\iota+1)}, x^{(\iota+2)}$ for a sufficiently large ι, we can obtain approximate eigenvalues λ_1 and λ_2 as the two roots of the quadratic equation in λ:

$$\begin{vmatrix} \chi_i^{(\iota)} & \chi_j^{(\iota)} & \theta^{(\iota)}\theta^{(\iota+1)} \\ \chi_i^{(\iota+1)} & \chi_j^{(\iota+1)} & \theta^{(\iota)}\lambda \\ \chi_i^{(\iota+2)} & \chi_j^{(\iota+2)} & \lambda^2 \end{vmatrix}=0 \qquad (i \text{ and } j \text{ arbitrary, } i\neq j).$$

The corresponding eigenvectors are given by $x_1=\lambda_2 x^{(\iota+1)}-x^{(\iota+2)}$ and $x_2=\lambda_1 x^{(\iota+1)}-x^{(\iota+2)}$. Complex conjugate pairs of eigenvalues can be dealt with in this manner. This is useful also when $|\lambda_1|\fallingdotseq|\lambda_2|$. The extension to the case of more than two eigenvalues with the same maximum absolute value is obvious.

In order to determine the remaining eigenvalues by the power method we have to combine it with the deflation or transformation of matrices, as discussed later in this section and in the following section. The amount of computation depends on the arrangement of the eigenvalues of A and the required accuracy. We note that the multiplication of a matrix by a vector requires an amount of computation proportional to n^2.

Improvement in the approximations can be incorporated into the power method. If $v=x_\mu+O(\varepsilon)$ is an approximation to the eigenvector x_μ corresponding to an eigenvalue λ_μ of a real symmetric matrix A, then the Rayleigh quotient $\lambda_R=(v, Av)/(v, v)$ affords a good approximation to λ_μ. In fact, we have $|\lambda_\mu-\lambda_R|=O(\varepsilon^2)$.

If λ_i $(i=1,\ldots,n)$ is an eigenvalue of A and x_i the corresponding eigenvector, then $P(\lambda_i)$ $(i=1,\ldots,n)$ is an eigenvalue of $P(A)$ and x_i the corresponding eigenvector, where $P(\xi)$ is a polynomial in ξ and ξ^{-1}. This fact can be utilized to transform the magnitudes of eigenvalues to accelerate the convergence of the power method, to separate eigenvalues with the same absolute value, or to obtain intermediate eigenvalues. The choice $P(\lambda)=\lambda-c$ or $(\lambda-c)^{-1}$ is particularly useful, where c is an appropriate constant. In fact, an efficient algorithm known as **inverse iteration** exists for computing an approximate eigenvector when a good approximation to an eigenvalue is known. Given a trial eigenvector x corresponding to a computed eigenvalue c, one computes an improved approximate eigenvector y by solving $(A-cI)y=x$ or $y=(A-cI)^{-1}x$.

Aitken's δ^2-method is also efficient in accelerating the convergence of the power method.

When an eigenvalue-eigenvector pair is known, another eigenvalue-eigenvector pair can be computed using the process known as **deflation**. If an eigenvalue λ_μ and the corresponding eigenvector x_μ (and also the corresponding left eigenvector ${}^t y_\mu$ if necessary) of A are known, it is possible to "subtract" them

from A to get a problem containing only the remaining eigenvalues. Such deflations are often used in combination with the power method. The following are two examples of deflation methods.

(1) Assuming that x_μ and y_μ are normalized in such a way that $(y_\mu, x_\mu)=1$, form $B=A-\lambda_\mu x_\mu y'_\mu$. Then B has the same set of eigenvalues and eigenvectors as A except for λ_μ. The eigenvalue and the eigenvector of B corresponding to λ_μ of A are 0 and x_μ, respectively. This kind of deflation process can be generalized to the case of nonlinear elementary divisors, but that becomes somewhat more complicated.

(2) After normalizing x_μ so that its nth component $x_{\mu n}$ is equal to 1, form $B=(b_{ij}): b_{ij}=a_{ij}-x_{\mu i}a_{nj}$ $(i,j=1,\ldots,n-1)$. Then B has the same set of eigenvalues as A except for λ_μ. If w_k is the eigenvector corresponding to the eigenvalue λ_k of B, then the corresponding eigenvector x_k of A is given by $x_{ki}=w_{ki}+d_k x_{\mu i}$ $(i=1,\ldots,n-1)$, $x_{kn}=d_k$, where d_k is determined from $\sum_{i=1}^{n-1} n_i w_{ki}=(\lambda_k-\lambda_\mu)d_k$ if $\lambda_k \neq \lambda_\mu$. If $\lambda_k=\lambda_\mu$ and $r \equiv \sum_{i=1}^{n} a_{ni}w_{ki}=0$, then we can put $d_k=0$. If $\lambda_k=\lambda_\mu$ and $r\neq 0$, A has a nonlinear elementary divisor for $\lambda_k=\lambda_\mu$, and the x_k defined by $x_{ki}=w_{ki}/r$ and $x_{kn}=0$ is a generalized eigenvector of A in the sense that $Ax_k=\lambda_\mu x_k+x_\mu$.

D. Transformation of Matrices

There are a number of methods of transforming a given matrix A by means of a suitable similarity transformation $A\to B=S^{-1}AS$ into another matrix B for which it is easier to solve the eigenvalue problem. The **Givens method** [9] transforms a symmetric matrix A into a **tridiagonal matrix** B (i.e., a matrix such that $b_{ij}=0$ for $|i-j|\geqslant 2$) by means of a matrix S which is the product of 2-dimensional rotation matrices. The **Householder method** [10] also transforms a symmetric A into a tridiagonal B by means of an orthogonal matrix S of special type, i.e., a reflection matrix $I\text{-}2uu^*$ ($u^*u=1$, $u^*=$conjugate transpose of u), and the **Lanczos method** [8] transforms a general A into a tridiagonal B. To general matrices the following methods are also applicable: (1) The **Danilevskiĭ method** [1] transforms A into its companion matrix B by repeated application of elimination operations. Here, row interchanges can be combined to increase numerical stability [14]. (2) The **Hessenberg method** [3] transforms A into a B such that $b_{ij}=0$ for $i-j\geqslant 2$ with a triangular S. (3) The Givens method [9] transforms A into a B of the same form as in (2) by repeated application of 2-dimensional rotations. This method now tends to give way to the Householder method or to the elimination method with row interchange. All these methods require an amount of computation proportional to n^3. In general, special treatment is necessary for the case of multiple eigenvalues. As an example, we explain the Givens method for general matrices. Let $N=(n-1)(n-2)/2$. For $\iota=0,1,\ldots,N-1$, choose $(p,q)=(3,2),(4,2),\ldots,(n,2);(4,3),(5,3),\ldots,(n,3);\ldots;(n-1,n-2),(n,n-2);(n,n-1)$ in this order. Using $T^{(\iota)}$ of the same form as in the Jacobi method, and setting $\tan\theta=a^{(\iota)}_{p,q-1}/a^{(\iota)}_{q,q-1}$ in this case, calculate $A^{(0)}=A$, $U^{(0)}=I$, $A^{(\iota+1)}=T^{(\iota)\prime}A^{(\iota)}T^{(\iota)}$, $U^{(\iota+1)}=U^{(\iota)}T^{(\iota)}$ $(\iota=0,1,\ldots,N-1)$, and put $B=A^{(N)}$. Then we have $b_{ij}=0$ for $i-j\geqslant 2$. We can solve the eigenvalue problem for this simplified B and then retransform the eigenvectors of B thus obtained into those of A by means of $U^{(N)}$. It should be noted that the method of bisection based on †Sturm's theorem is effectively used to solve the characteristic equation of a real tridiagonal matrix [14]. This method is remarkably stable numerically. It is used when the number of eigenvalues to be computed is small relative to the order of the given matrix. If all eigenvalues are desired, an alternative method such as the QR method (→ Section F) is recommended.

E. The Lanczos Method

A more detailed exposition of the Lanczos method is now given. Let A be a given real matrix of order n. Pick two vectors c_1 and $\tilde{c}_1$. Determine recursively the vectors c_{i+1} and $\tilde{c}_{i+1}$, $i=1,\ldots,n$, that satisfy the following conditions: (i) $\gamma_{i+1}c_{i+1}=Ac_i-\alpha_i c_i-\beta_i c_{i-1}\equiv b_{i+1}$, $\tilde{\gamma}_{i+1}\tilde{c}_{i+1}=A'\tilde{c}_i-\tilde{\alpha}_i\tilde{c}_i-\tilde{\beta}_i\tilde{c}_{i-1}\equiv\tilde{b}_{i+1}$, $i=2,\ldots,n$; (ii) c_{i+1} is orthogonal to $\tilde{c}_{i-1}$ and $\tilde{c}_i$; (iii) $\tilde{c}_{i+1}$ is orthogonal to c_{i-1} and c_i, where α_i, β_i, $\tilde{\alpha}_i$, and $\tilde{\beta}_i$ are scalars, and where γ_i and $\tilde{\gamma}_i$ are normalizing scalars. Actually, $\alpha_i=\tilde{c}'_iAc_i/\tilde{c}'_ic_i=\tilde{\alpha}_i$, $\beta_i=\gamma_i c'_i\tilde{c}_i/\tilde{c}'_{i-1}c_{i-1}$, $\tilde{\beta}_i=\tilde{\gamma}_i\tilde{c}'_ic_i/c'_{i-1}\tilde{c}_{i-1}$, where $c'_i\tilde{c}_i$, $i=1,\ldots,n$, are assumed nonzero. It can be shown that c_{i+1} is orthogonal to every $\tilde{c}_j$, $1\leqslant j\leqslant i$, and that $\tilde{c}_{i+1}$ is orthogonal to every c_j, $1\leqslant j\leqslant i$ $(c_{n+1}=\tilde{c}_{n+1}=0)$. The conclusion is that the given matrix A is similar to the tridiagonal matrix $H=(h_{ij})$, where $h_{ii}=\alpha_i$ $(i=1,\ldots,n)$, $h_{i,i+1}=\beta_{i+1}$, $h_{i+1,i}=\gamma_{i+1}$ $(i=2,\ldots,n)$. In fact, if C denotes the matrix whose jth column equals c_j $(j=1,\ldots,n)$, one obtains $C^{-1}AC=H$. Thus the eigenvalues of A are identical to those of H.

In principle, the Lanczos method transforms a given matrix to a tridiagonal matrix similar to it if every $c'_i\tilde{c}_i$ is nonzero $(i=1,\ldots,n)$. If $c'_i\tilde{c}_i$ does vanish for a certain i, one selects c_1 and $\tilde{c}_1$ again and restarts. On the other hand, if $b_{i+1}=0$ one chooses an arbitrary vector c_{i+1} orthogonal to $\tilde{c}_1,\ldots,\tilde{c}_i$ and if $\tilde{b}_{i+1}=0$ one selects an arbitrary vector $\tilde{c}_{i+1}$ orthogonal to

$c_1, \dots, c_i$. In the actual numerical computation, the distinction between zero vectors and nonzero vectors is usually blurred by rounding errors.

In the application of the Lanczos method one often observes the loss of orthogonality $c_i' \tilde{c}_j = 0$ $(i \neq j)$. This usually results from cancellation errors in the computation of b_{i+1} and $\tilde{b}_{i+1}$. If the orthogonality is lost, $C^{-1}AC$ may significantly deviate from a tridiagonal matrix. As a practical means for preserving the indicated orthogonality one can reorthogonalize the vectors c_i and $\tilde{c}_i$. Indeed, one can add to the computed c_{i+1} a linear combination of $c_1, \dots, c_i$ so that the sum is orthogonal to $\tilde{c}_1, \dots, \tilde{c}_i$, then take the sum as c_{i+1} after properly normalizing. A similar process applies to $\tilde{c}_{i+1}$. The Lanczos method is often applied in double precision. It has been suggested that one first reduces the given matrix to an upper Hessenberg form by using the Hessenberg method with row interchange before applying the Lanczos method [15].

A further remark on the Lanczos method is in order. Recall that γ_i is determined from Ac_i, c_{i-1}, α_{i-1}, and β_{i-1}; β_i from c_i, $\tilde{c}_i$, c_{i-1}, $\tilde{c}_{i-1}$, and $\tilde{\gamma}_i$; and α_i from Ac_i, c_i, and $\tilde{c}_i$. This shows that the Lanczos method applied to a sparse matrix (a matrix whose elements are almost all zero) requires only a memory proportional to n. The Householder method, on the other hand, requires memory proportional to n^2 even for a sparse matrix.

For a real symmetric matrix one can modify the Lanczos method for a general matrix so that C is orthogonal and H is real, symmetric, and tridiagonal (the details are omitted).

F. The QR Method

The QR method was discovered independently by J. G. F. Francis and by V. N. Kublanovskaya in 1961 [15]. The method has been improved and extended considerably since then. In the usual application to matrix eigenvalue problems, the QR method provides the most efficient iterative process for finding *all* eigenvalues of a given matrix. The matrix is reduced by means of a similarity transformation to Hessenberg form or to tridiagonal form before application of the QR method; the reduction process may be effected by the Householder method or by the elimination method with row interchange. (If the given matrix is a complex matrix, the latter is preferable.) The reason is that one step of the QR process applied to a full matrix is prohibitively expensive, requiring a number of operations proportional to n^3, while one step of the QR process applied to a Hessenberg matrix requires a number of operations proportional to n^2.

We now describe one of the most useful versions of the QR method. Let $A = A_0$ be a given matrix, and define a sequence of matrices $A_1, A_2, \dots$ obtained from A_0 as follows. At the ith $(i = 0, 1, \dots)$ step, choose an appropriate constant s_i, called an origin shift, according to the process described below, and decompose $A_i - s_i I$ into the product of a unitary matrix Q_i and an upper Hessenberg matrix R_i: $A_i - s_i I = Q_i R_i$. The matrix A_{i+1} is then defined by $A_{i+1} = R_i Q_i + s_i I = Q_i^* A_i Q_i$. Hence A_{i+1} is similar to A_i. The QR method is closely related to the power method and to the inverse power method. We describe this relationship in order to obtain an insight into the nature of the QR method. To this end, we first state a well-known theorem: Let A be diagonalizable, and let its eigenvalues λ_i $(i = 1, \dots, n)$ have distinct moduli, say $|\lambda_1| > |\lambda_2| > \dots > |\lambda_n|$. Let $X^{-1}AX = \mathrm{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$, and let X have an LU-decomposition $X = LU$, where L and U are, respectively, lower triangular matrix and an upper triangular matrix. Then the QR algorithm without an origin shift ($s_i = 0$, $i = 0, 1, 2, \dots$) behaves as follows: $a_{pq}^{(i)} \to 0$ $(p > q)$; $a_{pp}^{(i)} \to \lambda_p$; $a_{pq}^{(i)}$ oscillates $(p < q)$, $p, q = 1, 2, \dots, n$, $i \to \infty$, where $a_{pq}^{(i)}$ denotes the (p, q)th element of A_i. In other words, A_i approaches an upper triangular matrix as $i \to \infty$ so that the diagonal elements of A_i converge to the eigenvalues of A.

The relationship between the QR algorithm and the power method is given by $(Q_0 Q_1 \dots Q_i)(R_i R_{i-1} \dots R_0) = (A - s_0 I)(A - s_1 I) \dots (A - s_i I)$. If $s_0 = s_1 = \dots = s_i = 0$, then the right-hand side reduces to A^{i+1}, and since the product $R_i \dots R_0$ is upper triangular, the first column of $\tilde{Q}_i \equiv Q_0 Q_1 \dots Q_i$ equals a scalar multiple of $A^{i+1} e_1$, where $e_1 = (1, 0, 0, \dots, 0)'$. Therefore, by the power method, the first column of $\tilde{Q}_i$ converges to an eigenvector corresponding to the eigenvalue λ_1 of A that has the largest modulus, under a certain fairly mild condition. Since $A_{i+1} = \tilde{Q}_i^* A \tilde{Q}_i$, $A_{i+1} e_1 = \tilde{Q}_i^* A \tilde{Q}_i e_1 \approx \lambda_1 e_1$ for large i. In other words, the first column of A_i converges to the vector $(\lambda_1, 0, 0, \dots, 0)'$ as $i \to \infty$.

The relationship of the QR method to the inverse power method and to the Rayleigh quotient is now explained. From the definition of the QR algorithm with an origin shift, we have $Q_i = [(A - s_i I)^{-1}]^* R_i^*$. Since R_i^* is a lower triangular matrix, the nth column of Q_i equals $Q_i e_n = [(A_i - s_i I)^{-1}]^* R_i^* e_n = (A - s_i I)^{-1} \cdot$(a scalar multiple of e_n), where $e_n = (0, 0, \dots, 0, 1)'$. This last process of obtaining the last column of Q_i represents a process known as the inverse power method. If s_i is close to an eigenvalue of

A_i (and hence to an eigenvalue of A), the last column of A_i gives a good approximate eigenvector of A corresponding to the eigenvalue under consideration. Now, if x is a given column vector such that $x^*x=1$, the value of λ which minimizes $(Ax-\lambda x)^*(Ax-\lambda x)$ is given by the Rayleigh quotient x^*Ax. If $x(x^*x=1)$ happens to be an eigenvector of A, the Rayleigh quotient equals an eigenvalue of A. If we take $x=e_n$, the corresponding Rayleigh quotient equals a_{nn}. Therefore if we take the (n,n)th element of A_i as the origin shift s_i, then s_i can be regarded as the best approximation to an eigenvalue of A_i (hence of A) when e_n is taken as an approximate eigenvector of A_i, in the sense that $\lambda=s_i$ minimizes the functional $(A_ie_n-\lambda e_n)^*(A_ie_n-\lambda e_n)$.

Under the same condition as in the preceeding theorem, the rate of convergence of the QR method with an origin shift is given as follows: $a_{pq}^{(i)}$ $(n \geqslant p \geqslant q \geqslant 1)$ is asymptotically proportional to $(\lambda_p/\lambda_q)^i$ for $s_i=0$ $(i=0,1,\ldots)$ (no origin shift); and with the origin shift s_i, the behavior of $a_{pq}^{(i)}$ at the ith step is determined by $(\lambda_p-s_i)/(\lambda_q-s_i)$. If $s_i \fallingdotseq \lambda_n$ ($=$the eigenvalue of A with the least modulus), each element in the nth row of A_{i+1} except $a_{nn}^{(i+1)}$ exhibits rapid decrease in modulus. A natural and practical choice of the origin shift s_i is given as follows: (i) for a real tridiagonal matrix, s_i is taken to be that eigenvalue of the 2×2 matrix situated at the lower right corner of A_i that is closer to $a_{nn}^{(i)}$ [14]; (ii) for a real upper Hessenberg matrix, the two eigenvalues of the 2×2 matrix situated at the lower right corner of A_i as s_i and s_{i+1} [14]; (iii) for a complex upper Hessenberg matrix, s_i is chosen as in (i) [17, COMQR].

The QR method with origin shift would eventually make each element in the nth row, except $a_{nn}^{(i)}$, smaller in modulus than a prescribed positive number. At this stage of iteration, $a_{nn}^{(i)}$ is taken as an approximate eigenvalue of A. The QR method is then applied anew to the $(n-1)\times(n-1)$ matrix obtained from A_i by deleting the nth row and the nth column of A_i, and another approximate eigenvalue of A is obtained. The method proceeds similarly until all the eigenvalues of A are computed. For maximum accuracy the eigenvalues of the given matrix should be computed in the order of increasing modulus. A word of caution is in order. When the given matrix A has elements of greatly varing modulus, rearrangement of elements of A may be necessary before applying the QR method with an explicit origin shift, where A_i-s_iI $(i=0,1,\ldots)$ is explicitly computed.

The reader is referred to [14–16] for details of the QR method.

G. Generalized Eigenvalue Problem $Ax=\lambda Bx$

An eigenvalue problem of the type $Ax=\lambda Bx$ is called a **generalized eigenvalue problem** and is often encountered in applied mathematics. A necessary and sufficient condition for λ to be an eigenvalue is $\det(A-\lambda B)=0$. If B^{-1} exists, then $Ax=\lambda Bx$ is equivalent to $B^{-1}Ax=\lambda x$ and hence has n eigenvalues, where n is the order of A. If B^{-1} does not exist, the eigenvalue problem $Ax=\lambda Bx$ has at most $n-1$ eigenvalues. We restrict ourselves to one of the most important cases: that where A is real and symmetric and B is real, symmetric, and positive definite. In this case one could solve the problem by reformulating it as $B^{-1}Ax=\lambda x$, where $B^{-1}A$ is explicitly computed. However, $B^{-1}A$ is not in general symmetric. Moreover, when B has eigenvalues of widely different moduli, the elements of B^{-1} may have widely different moduli as well, which would in turn make the computation of those eigenvalues of $B^{-1}A$ of smaller moduli difficult. An efficient and numerically stable method is known which obviates the aforementioned difficulty by exploiting the symmetry of A and B. An outline of this procedure is now given. Since B is positive definite, a lower triangular matrix L exists such that $LL'=B$. This is called the Cholesky decomposition of B. The elements of L can be computed by equating the corresponding elements in $LL'=B$. The eigenvalue problem $Ax=\lambda Bx$ is now equivalent to $L^{-1}A(L')^{-1}y=\lambda y$ with $y=L'x$, where the matrix $L^{-1}A(L')^{-1}=P$ is computed in two stages by solving $LX=A$ and $PL'=X$. In this last equation it is enough to compute the upper right half of P, because P is symmetric while X is not generally symmetric. The upper right half of P can be computed by equating the corresponding elements in $PL'=X$. The eigenvalues λ of $Ax=\lambda Bx$ are given by the eigenvalues of the matrix P. Other types of eigenvalue problems, such as $y'AB=\lambda y'$, $BAy=\lambda y$, and $x'BA=\lambda x'$, often appear in practice, where A is real and symmetric and B is real, symmetric, and positive definite. By using the Cholesky decomposition of B, one can reduce any one of these eigenvalue problems to the ordinary eigenvalue problem for a real symmetric matrix [4].

If A and B are general matrices in $Ax=\lambda Bx$, the following method is known to be effective [18]. First, reduce B to an upper triangular matrix by applying $n-1$ Householder transformations from the left. This reduces the eigenvalue problem $Ax=\lambda Bx$ to the case where B is upper triangular. Next, apply to A a sequence of plane rotations of Givens type from the left, thereby reducing the eigenvalue problem $Ax=\lambda Bx$ to the case where A is an

upper Hessenberg matrix and B is an upper triangular matrix. Then apply the QR method to $B^{-1}Ax=\lambda x$ without explicitly computing $B^{-1}A$ to reduce the eigenvalue problem $Ax=\lambda Bx$ to the case where A and B are both approximately upper triangular. The eigenvalues of $Ax=\lambda Bx$ are then easily computed as ratios of corresponding diagonal elements. This method is called the **QZ method** [18].

A collection of about 50 excellent FORTRAN subroutines for various types of matrix eigenvalue problems is contained in [18]. These subroutines are in most part translations from ALGOL procedures given in [14].

References

[1] H. Wayland, Expansion of determinantal equations into polynomial form, Quart. Appl. Math., 2 (1945), 277–306.
[2] P. S. Dwyer, Linear computation, Wiley, 1951.
[3] R. Zurmühl, Matrizen, Springer, 1950.
[4] L. Collatz, Eigenwertaufgaben mit technischen Anwendungen, Akademische Verlag., 1949.
[5] R. V. Mises and H. Geiringer, Praktische Verfahren der Gleichungsauflösung, Z. Angew. Math. Mech., 9 (1929), 58–77, 152–164.
[6] E. Bodewig, Matrix calculus, Interscience, 1956.
[7] C. G. J. Jacobi, Über ein leichtes Verfahren, die in der Theorie der Säkularstörungen vorkommenden Gleichungen numerisch aufzulösen, J. Reine Angew. Math., 30 (1846), 51–95.
[8] C. Lanczos, An iteration method for the solution of the eigenvalue problem of linear differential and integral operators, J. Res. Nat. Bur. Standards, 45 (1950), 255–282.
[9] W. Givens, Computation of plane unitary rotations transforming a general matrix to triangular form, SIAM J., 6 (1958), 26–50.
[10] J. H. Wilkinson, Householder's method for symmetric matrices, Numer. Math., 4 (1963), 354–361.
[11] P. A. White, The computation of eigenvalues and eigenvectors of a matrix, SIAM J., 6 (1958), 393–437.
[12] S. H. Crandall, Engineering analysis, McGraw-Hill, 1956.
[13] A. Schönhage, Zur quadratischen Konvergenz des Jacobi-Verfahrens, Numer. Math., 6 (1964), 410–412.
[14] J. H. Wilkinson and C. Reinsch, Handbook for automatic computation II, Linear algebra, Springer, 1971.
[15] J. H. Wilkinson, The algebraic eigenvalue problem, Clarendon Press, 1965.
[16] G. W. Stewart, Introduction to matrix computations, Academic Press, 1973.
[17] B. T. Smith, J. M. Boyle, J. J. Dongarra, B. S. Garbow, Y. Ikebe, V. C. Klema, and C. B. Moler, Matrix eigensystem routines—EISPACK guide, Springer, second edition, 1976.
[18] C. B. Moler and G. W. Stewart, An algorithm for generalized matrix eigenvalue problems, SIAM J. Numer. Anal., 10 (1973), 241–256.

299 (XV.7) Numerical Integration

A. Interpolatory Integration Formulas

Numerical integration is a method of finding an approximate numerical value of a definite integral of a given function $f(x)$. Usually the integral $\int_a^b f(x)dx$ or $\int_a^b f(x)w(x)dx$ ($w(x)$ is the **weight function**) is approximated by a linear combination $\sum_{i=1}^n W_i f(x_i)$ of the values of the integrand at the points $x_1, x_2, \dots, x_n$. Integration formulas are divided into two groups, the interpolatory formulas and the formulas based on variable transformation.

In order to obtain an **interpolatory formula**, we interpolate over the integrand $f(x)$ at n points $x_1, x_2, \dots, x_n$ by means of the †Lagrange interpolation polynomial of degree $n-1$, and then integrate the polynomial over $[a,b]$. Depending on the selection of the points x_i and the weights W_i, we have several kinds of formulas.

(1) Newton-Cotes Formulas. We assume $w(x)$ = constant and $x_i = x_0 + ih$ ($i=0,1,\dots,n$). The weights W_i are so determined that the value of the integral can be calculated accurately if the integrand $f(x)$ is a polynomial whose degree does not exceed n. There formulas are called the **Newton-Cotes formulas**: $\int_{x_0}^{x_1} f(x)dx = (f_0 + f_1)h/2$ for $n=1$ (**trapezoidal rule**), $\int_{x_0}^{x_2} f(x)dx = (f_0+4f_1+f_2)h/3$ for $n=2$ (**Simpson's 1/3 rule**), and $\int_{x_0}^{x_3} f(x)dx = (f_0+3f_1+3f_2+f_3)3h/8$ for $n=3$ (**Simpson's 3/8 rule**). The truncation errors of these three formulas are given by $h^3 f^{(2)}(\xi)/12$, $h^5 f^{(4)}(\xi)/90$, $3h^5 f^{(4)}(\xi)/80$, respectively, where ξ is a number in the interval of integration and $f^{(i)}$ denotes the ith derivative of f (differentiability is assumed). For an even n, the polynomial of degree $n+1$ can also be integrated accurately by these formulas.

In Table 1 the coefficients A and B_i of the Newton-Cotes formulas $hA\sum_{i=0}^n B_i f(x_i)$, $h=(x_n-x_0)/n$, and the coefficients C of the error

Table 1

n	A	B_0	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	C
1	1/2	1	1								−1/12
2	1/3	1	4	1							−1/90
3	3/8	1	3	3	1						−3/80
4	4/45	7	32	12	32	7					−8/945
5	5/288	19	75	50	50	75	19				−275/12096
6	1/140	41	216	27	272	27	216	41			−9/1400
7	7/17280	751	3577	1323	2989	2989	1323	3577	751		−8183/518400
8	4/14175	989	5888	−928	10496	−4540	10496	−928	5888	989	−2368/467775

term $Ch^{p+1}f^{(p)}(\xi)$, where $p=n+2$ if n is even and $p=n+1$ if n is odd, are given.

When the interval $[a,b]$ of integration is large, we usually divide it into small subintervals and apply formulas for small n for each part rather than formulas for large n for the whole interval. For example, if the interval is divided into m equal subintervals, we get the following formula by applying the trapezoidal rule for each subinterval:

$$\int_a^b f(x)\,dx = h((f_0+f_m)/2+(f_1+f_2+\ldots+f_{m-1})),$$

where $x_0=a$, $x_m=b$, $h=(b-a)/m$, with truncation error $(b-a)^3 f^{(2)}/(12m^2)$. (Here, as in the rest of the article, $f^{(i)}$ stands for $f^{(i)}(\xi)$, with differentiability assumed as before.) By applying Simpson's 1/3 rule we obtain the formula

$$\int_a^b f(x)\,dx$$

$$=\frac{2}{3}h((f_0+f_{2m})/2+2(f_1+f_3+\ldots+f_{2m-1})$$

$$+(f_2+f_4+\ldots+f_{2m-2})),$$

where $x_0=a$, $x_{2m}=b$, $h=(b-a)/2m$, with truncation error $(b-a)^5 f^{(4)}/180(2m)^4$.

For the integral of a periodic analytic function over a single period the trapezoidal rule with equally spaced points gives the best result asymptotically, as the number of points tends to infinity.

Newton-Cotes formulas can also be obtained by integrating over the interval $[a,b]$ the interpolation polynomial for equally spaced points. The formulas mentioned so far are called **closed formulas** since they use values at the two endpoints of the interval. We can also use **open formulas**, which do not use values at the ends. For example, we have $\int_{x_0}^{x_4} f(x)\,dx = 4h(2f_1-f_2+2f_3)/3$ with truncation error $14h^5 f^{(4)}/45$. Open formulas are useful for numerical solution of differential equations. There are also formulas that include values outside the interval, for example, $(-f_{-1}+13f_0+13f_1-f_2)h/24$ for $\int_{x_0}^{x_1} f(x)\,dx$. Applying these formulas to the m subintervals of $[a,b]$, we obtain a trapezoidal formula with the correction term $h(f_1-f_{-1})/24+h(f_m-f_{m+1})/24=h(\Delta f_0+\Delta f_{-1})/24-h(\Delta f_{m-1}+\Delta f_m)/24$, where Δf_i means $f_{i+1}-f_i$.

(2) Chebyshev Formulas. The **Chebyshev formulas** are a family of integration formulas in which all the weights W_i of $\sum_{i=1}^n W_i f(x_i)$ are equal, while the abscissas x_i are chosen so that the integral can be evaluated exactly when $f(x)$ is an arbitrary polynomial whose degree does not exceed n. When $\int_{-1}^1 f(x)\,dx \fallingdotseq W\sum_{i=1}^n f(x_i)$, it is easy to see that $W=2/n$ since the right-hand side must be equal to the left-hand side when $f(x)\equiv 1$. It is known that the abscissas x_i for $n\leqslant 7$ and $n=9$ are real, while for $n=8$ and $n\geqslant 10$ at least one of the abscissas becomes complex. It is easy to see that Chebyshev formulas are interpolatory.

(3) Gauss Formulas. In the **Gauss formulas** both the weights W_i and the abscissas x_i are chosen so that we obtain the accurate value of the integral when the integrand is any polynomial whose degree does not exceed $2n-1$. If we put $\Pi(x)=\prod_{i=1}^n(x-x_i)$, an arbitrary polynomial of degree $2n-1$ can be expressed in the form

$$f(x)=\sum_{k=1}^{n}\frac{\Pi(x)}{(x-x_i)\Pi'(x_k)}f_k+\Pi(x)\sum_{k=0}^{n-1}a_k x^k,$$

where $f(x_k)=f_k$ and the first term is the †Lagrange interpolation polynomial. By the assumption that the integral of $f(x)$ with weight $w(x)$ equals $\sum_{k=1}^n W_k f_k$, we obtain $W_k=\int_a^b(w(x)\Pi(x)/(x-x_k)\Pi'(x_k))\,dx$ and the relations

$$\int_a^b w(x)\Pi(x)x^k\,dx=0,\qquad k=0,1,\ldots,n-1.$$

Accordingly, the abscissas x_i are determined as the roots of the polynomial $\Pi(x)$ of degree n that is orthogonal to x^k ($k=0,1,\ldots,n-1$) with respect to the weight function $w(x)$.

The following are typical examples of integration formulas of Gaussian type.

(i) **Gauss integration formulas** (in the narrow sense). For $w(x)=1$ and the interval $[-1,1]$, $\Pi(x)$ is the †Legendre polynomial $P_n(x)=(1/2^n n!)d^n(x^2-1)^n/dx^n$. The error is $(n!)^4 2^{2n+1} f^{(2n)}/(2n+1)((2n)!)^3$.

(ii) **Gauss-Laguerre formulas.** For the weight $w(x)=\exp(-x)$ and the interval $[0,\infty)$, $\Pi(x)$ is the †Laguerre polynomial $L_n(x)=(\exp x)d^n(x^n\exp(-x))/dx^n$.

(iii) **Gauss-Hermite formulas.** For the weight function $w(x)=\exp(-x^2)$ and the interval $(-\infty,\infty)$, $\Pi(x)$ is the †Hermite polynomial $H_n(x)=(-1)^n\exp x^2\cdot d^n\exp(-x^2)/dx^n$.

(iv) **Gauss-Chebyshev formulas.** For the weight function $w(x)=(1-x^2)^{-1/2}$ and the interval $[-1,1]$, we use the Chebyshev polynomial $T_n(x)=2^{-(n-1)}\cos(n\arccos x)$. In this case, the W_i are all equal to π/n.

From the definition of W_k we see that the Gauss formulas are interpolatory.

(4) Clenshaw-Curtis Formulas. Although the Gauss formula is in general more accurate than the Newton-Cotes formula with the same number of points of interpolation, the points of interpolation for a Gauss formula of any order are distinct from those of any other order except the point zero, which appears in all formulas of odd order. The **Clenshaw-Curtis formulas** are interpolatory, with the points chosen so that the distribution of the points is similar to that of the Gauss formula and such that, in proceeding from a computation of order n to that of order $2n$, all the function values evaluated in the former computation be used in the latter. For $w(x)=1$, the interval $[-1,1]$, and even n, the points x_k and the weights W_k are given by

$$x_k=\cos\frac{k\pi}{n},\qquad k=0,1,\ldots,n;$$

$$W_0=W_n=\frac{1}{n^2-1};$$

$$W_s=W_{n-s}=\frac{4}{n}\sum_{j=0}^{n/2}{}''\frac{1}{1-4j^2}\cos\frac{2\pi js}{n},$$

$$s=1,2,\ldots,\frac{n}{2}.$$

Σ'' means that the first and the last terms in the sum are to be halved. There are some different types of Clenshaw-Curtis formulas depending on the selection of the points x_k [1].

B. Integration Formulas Based on Variable Transformation

If the integrand has some singularity at the endpoint, any interpolatory formula based on interpolation with a polynomial does not give a good result. In such cases **integration formulas based on variable transformation** are quite effective.

(1) IMT Formula. The †Euler-Maclaurin formula is given by

$$\int_a^b f(x)\,dx=h\left\{\frac{1}{2}f(a)+\sum_{k=1}^{n-1}f(a+kh)+\frac{1}{2}f(b)\right\}$$
$$-\sum_{r=1}^{m}\frac{h^{2r}B_{2r}}{(2r)!}\{f^{(2r-1)}(b)$$
$$-f^{(2r-1)}(a)\}+R_m,$$

$$h=\frac{b-a}{n},$$

$$R_m=\frac{h^{2m+1}}{(2m)!}\int_0^1 B_{2m}(t)\left\{\sum_{k=0}^{n-1}f^{(2m)}(a+kh+ht)\right\}dt,$$

where the $B_n(t)$ are †Bernoulli polynomials of degree n and the B_n are †Bernoulli numbers ($B_0=1, B_1=-1/2, B_2=1/6, B_3=0, B_4=-1/30,\ldots$). This formula suggests that if the higher derivatives of the integrand vanish at the both endpoints, the error of the trapezoidal rule with equally spaced points becomes very small. The **IMT formula** is based on the idea of transforming the variable x of $\int_0^1 f(x)\,dx$ in such a way that all the derivatives of the new integrand vanish at both endpoints by taking $x=\varphi(t)$, $\varphi(t)=K^{-1}\int_0^t\psi(\tau)\,d\tau$, $K=\int_0^1\psi(\tau)\,d\tau$, $\psi(\tau)=\exp(-\tau^{-1}-(1-\tau)^{-1})$. Then the trapezoidal rule with $h=1/n$ is applied to the transformed integral to obtain $(1/Kn)\sum_{j=1}^{n-1}\psi(j/n)f(\varphi(j/n))$ [1]. The asymptotic expression of the error for the IMT formula is proportional to $\exp(-C\sqrt{n})$ with a positive constant C.

(2) Double Exponential Formula. The trapezoidal rule with equally spaced points applied to the integral of an analytic function over $(-\infty,\infty)$ gives in general a result of high accuracy. The **double exponential formula** is based on the idea of transforming $\int_{-1}^1 f(x)\,dx$ to $\int_{-\infty}^{\infty}f(\varphi(t))\varphi'(t)\,dt$ with $x=\varphi(t)=\tanh(\frac{1}{2}\pi\sinh t)$ and applying the trapezoidal rule with a mesh size h, which results in $h\sum_{n=-\infty}^{\infty}f(\varphi(nh))\varphi'(nh)$ [4]. The name of the double exponential formula is attributed to the decay of $\varphi'(t)$ at $t\to\pm\infty$, which is approximately proportional to $\exp(-C\exp|t|)$ with a positive constant C. The transformations $x=\exp(\pi\sinh t)$ and $x=\sinh(\frac{1}{2}\pi\sinh t)$ give the double exponential formulas for the infinite integrals $\int_0^\infty f(x)\,dx$ and $\int_{-\infty}^{\infty}f(x)\,dx$, respectively. In the actual computation the infinite summation is truncated at appropriate upper and lower bounds. The asymptotic expression of the error for the double exponential formula in terms of the number N of the sampling points actually used is proportional to $\exp(-CN/\log N)$ with a positive constant C. The IMT formula and the double exponential

formula are robust against the singularities at the endpoints.

C. Automatic Integration

By an **automatic integration scheme** we mean a computer program for numerical integration of $\int_a^b f(x)dx$ in which the user gives the limits of integration a and b, a subroutine for computing $f(x)$, and an error tolerance ε. Then the program gives a value of the integral which is expected to be correct within the tolerance ε. Usually in an automatic integration scheme the mesh size is halved until the desired accuracy is attained. Automatic integration schemes are classified into two groups, nonadaptive schemes and adaptive schemes. In a **nonadaptive scheme**, the sequence of integration points is chosen according to some fixed rule independent of the shape of the integrand. Newton-Cotes formulas, Clenshaw-Curtis formulas, IMT formulas, and double exponential formulas can be used as base formulas for nonadaptive schemes.

From the historical point of view, **Romberg integration** should be mentioned; it is a kind of nonadaptive automatic integrator. Consider an integral $I=\int_a^b f(x)dx$. Divide the interval $[a,b]$ of integration into 2^k subintervals and apply the trapezoidal rule with the mesh size $h=(b-a)/2^k$, which we denote by $T_0^{(k)}$. Then, starting from the values obtained for $T_0^{(k)}$, $k=0,1,\ldots$, we compute the sequence

$$T_m^{(k)}=\frac{4^m T_{m-1}^{(k+1)}-T_{m-1}^{(k)}}{4^m-1}, \qquad m=1,2,\ldots.$$

From the †Euler-Maclaurin formula the asymptotic error for the trapezoidal rule $T_0^{(k)}$ can be expressed as $I-T_0^{(k)}=c_1h^2+c_2h^4+\ldots+c_mh^{2m}+\ldots$, where $c_m=\text{const}\times(f^{(2m-1)}(b)-f^{(2m-1)}(a))$ does not depend on h. If we compute $T_1^{(k)}=(4T_0^{(k+1)}-T_0^{(k)})/3$ using the values $T_0^{(k+1)}$ and $T_0^{(k)}$, then we see that the asymptotic error expression of $T_1^{(k)}$ becomes $I-T_1^{(k)}=c_2'h^4+c_3'h^6+\ldots+c_m'h^{2m}+\ldots$. Romberg integration is based on the idea of eliminating the term with h^{2m} in the expression of the error by successive application of $T_m^{(k)}$. This is an application of Richardson's extrapolation procedure.

In the **adaptive scheme**, the points are chosen in a way that depends on the shape of the integrand. The Newton-Cotes formula of order 8, for example, is used as the base formula for the adaptive scheme.

D. Approximate Multiple Integration

If a region is a product region, such as a rectangular parallelepiped, a **product rule** obtained by forming the product of 1-dimensional formulas is useful. For integrals over a cube or a $sphere there are monomial rules which are exact for a certain family of monomials. These rules can be used for integrals of dimension lower than 5 or 6. For integrals of higher dimension only methods based on sampling make sense.

E. Numerical Differentiation

In order to find the numerical value of the derivative $f_p'=f'(x_p)$ at a point x_p from the tabulated values $f_k=f(x_k)$, we usually use the derivative of the †Lagrange interpolation formula. This gives

$$f_p'=\sum_{k\neq p}\frac{\Pi'(x_p)}{(x_p-x_k)\Pi'(x_k)}f_p+\sum_{k\neq p}\frac{1}{x_p-x_k}f_k,$$

where $\Pi(x)=\prod_{k=1}^n(x-x_k)$.

When we compute the derivative of a function which can be evaluated at any point in a given interval, the approximation

$$f'(x)=\frac{f(x+h)-f(x-h)}{2h}$$

is useful; similarly, we can use

$$f''(x)=\frac{f(x+h)-2f(x)+f(x-h)}{h^2}.$$

It must be noted that, as h tends to zero, the difference $f(x+h)-f(x-h)$ comes to contain fewer significant digits, so that it is meaningless to carry out $\{f(x+h)-f(x-h)\}/2h$ beyond a reasonable value of h.

References

[1] P. J. Davis and P. Rabinowitz, Methods of numerical integration, Academic Press, 1975.
[2] P. J. Davis and I. Polonsky, Numerical interpolation, differentiation and integration, Handbook of Mathematical Functions, M. Abramowitz and I. A. Stegun (eds.), Nat. Bur. Stand. Appl. Math. Ser., 55 (1964), 875–924.
[3] V. I. Krylov, Approximate calculation of integrals, translated by A. H. Stroud, Macmillan, 1962.
[4] H. Takahasi and M. Mori, Double exponential formulas for numerical integration, Publ. Res. Inst. Math. Sci., 9 (1974), 721–741.
[5] J. R. Rice, A meta-algorithm for adaptive quadrature, J. ACM, 22 (1975), 61–82.
[6] A. H. Stroud and D. H. Secrest, Gaussian quadrature formulas, Prentice-Hall, 1966.
[7] A. H. Stroud, Approximate calculation of multiple integrals, Prentice-Hall, 1971.

300 (XV.1) Numerical Methods

In the earlier history of mathematics, the development of methods of numerical calculation was one of the main purposes of research. Until the beginning of this century, logarithmic calculation played a central role in numerical calculation, and the main topic of this field was to make tables of values of functions. The digital electronic computer (→ 75 Computers), which made its first appearance in the 1940s and has been developing at an exponential rate, has caused drastic changes in numerical technique. Problem solving by numerical methods has now become one of the fundamental means of research in the physical sciences and engineering, and also in the social sciences and humanities. In this article we give examples of the changes in numerical methods brought about by the availability of digital computers and portable calculators.

Computers may be effectively utilized for calculating individual values of functions. This has led to the reexamination and, in some cases, modification of approximate formulas for evaluating functions (→ 142 Evaluation of Functions). For familiar functions, such as †logarithmic, †exponential, and †trigonometric functions, tables have been almost completely replaced by function keys on electronic calculators. Microprogramming algorithms for obtaining values of these functions have also been devised [1]. A complex function-theoretic error-estimation method for use with numerical integration formulas is given in [2]. In this method, graphical outputs of the computer are utilized. For problems in which the existence and uniqueness of solutions have been established, as for algebraic equations and ordinary differential equations, fairly good numerical calculation methods and their error estimates have been established (→ 301 Numerical Solution of Algebraic Equations; 303 Numerical Solution of Ordinary Differential Equations; [3,4]). A method for estimating arithmetic †accumulation errors for operations involving finite numbers of digits has been systematized (→ 138 Error Analysis; [5]).

It is not unusual nowadays for linear equations with 10,000 or more unknowns to be solved. As new computing systems, such as the virtual memory system and the vector operation system, come into existence, new algorithms are examined; a numerical method that is optimal for today's technology may well be suboptimal for the next generation of computers. General-purpose program packages of linear problems, including eigenvalue problems, have been developed (→ 298 Numerical Computation of Eigenvalues; [6]).

Partial differential equations seem to have become more familiar because of the visualization of their numerical solutions in graphical computer outputs. In [7], which appeared much earlier than computers, partial difference equations for the fundamental linear problems in mathematical physics were discussed with a suggested variational treatment (→ 304 Numerical Solution of Partial Differential Equations; [8]). The †finite element method, which started as a calculation technique in structural mechanics and is based on the †calculus of variations, is widely accepted as an efficient approximation method for partial differential equations [9, 10]. The term †simulation is used often to describe procedures in which partial differential equations describing time-dependent phenomena are discretized; the resulting difference system can be solved for long periods of time [11].

Numerical analysis has heretofore long been considered to be the only numerical method (for error analysis in particular), and has been carried out mainly by means of the discretization of equations. Nowadays, however, **mathematical modeling**, taking into account both the phenomena to be described and the capabilities of the computers to be used, has become an important numerical method.

References

[1] J. S. Walther, A unified algorithm for elementary functions, Conference Proceedings, Spring Joint Computer Conference, AFIPS 38 (1971), AFIPS Press, 1971, 379–385.
[2] H. Takahashi and M. Mori, Error estimation in the numerical integration of analytic functions, Rep. Computing Centre, Univ. of Tokyo, 3 (1970), 41–108.
[3] P. Henrici, Elements of numerical analysis, Wiley, 1974.
[4] P. Henrici, Discrete variable methods in ordinary differential equations, Wiley, second edition, 1968.
[5] J. H. Wilkinson, Rounding errors in algebraic processes, HMSO, 1963.
[6] J. H. Wilkinson and C. Reinsch (eds.), Handbook for automatic computation 2, pt. II, Linear algebra, Springer, 1971.
[7] R. Courant, K. Frienrichs, and H. Lewy, On the partial difference equations of mathematical physics, IBM J. Res. Develop., 11 (1967), 215–234. (Original in German, 1928.)
[8] F. John, Lectures on advanced numerical analysis, Gordon & Breach, 1967.
[9] G. Strang and G. J. Fix, An analysis of the finite element method, Prentice-Hall, 1973.

[10] P. G. Ciarlet, The finite element method for elliptic problems, North-Holland, 1978.
[11] R. J. Roache, Computational fluid dynamics, Hermosa, 1972.

301 (XV.5) Numerical Solution of Algebraic Equations

A. General Remarks

Methods for finding roots of an equation $f(x)=0$ (where $f(x)$ is not necessarily a polynomial, but is assumed to be a function with some regularity) by numerical calculation can in general be divided into the following two types: (i) The first has as its goal the finding of good approximate values of the roots; examples are the Bernoulli method (Section J) and the Graeffe method (Section N). (ii) The second improves the accuracy of estimates of the roots; an example is the Newton-Raphson method (Section D). The methods belonging to (i) and (ii) can be used separately. However, in (i) convergence of the approximations may be excessively slow when a pair of nearly equal roots is present, and the size of numbers involved in the calculation may grow exponentially as we proceed through the iterations. In (ii) convergence of the approximate values is not assured unless the initial approximate value is suitably chosen. Accordingly, when an electronic computer is utilized, it is advisable to combine both types of method. Because of the development of computers, solutions with global convergence have become important [1, 2, 15, 17, 21].

B. Successive Substitutions

When an equation $f(x)=0$ is transformed into $x=F(x)$ and the roots are obtained by iterative calculation of $x_{i+1}=F(x_i)$ $(i=0,1,2,\dots)$, sufficient conditions for its convergence are as follows: Let one of the roots of the equation be α. Then $x_{i+1}-\alpha=F(x_i)-\alpha$, and therefore $(x_{i+1}-\alpha)/(x_i-\alpha)=F'(\xi)$ $(x_i<\xi<\alpha$ or $x_i>\xi>\alpha)$. Accordingly, α converges monotonically if $0<F'(\xi)<1$, while it converges with oscillation if $0>F'(\xi)>-1$. If $|F'(\xi)|>1$ and $F^{-1}(x)$ denotes the inverse function of $F(x)$, the iteration $x_{i+1}=F^{-1}(x_i)$ converges. To define the speed of convergence by iterative processes the following notion of order is utilized: When $\lim_{i\to\infty}x_i=\alpha$, the speed of convergence of the sequence $\{x_i\}$ is said to be **of the kth order** if $\lim_{i\to\infty}(x_{i+1}-\alpha)/(x_i-\alpha)^k=c\neq 0$. Necessary and sufficient conditions for the speed of convergence of $\{x_i\}$ to be of the kth order are $\alpha=F(\alpha)$, $F'(\alpha)=F''(\alpha)=\dots=F^{(k-1)}(\alpha)=0$, $F^{(k)}(\alpha)\neq 0$.

The main iterative processes used in numerical calculation are described in the following sections.

C. Regula Falsi

Regula falsi (or the **method of false position**) is a process of obtaining the real root α of an equation $f(x)=0$ by approaching the root from both sides. The calculation procedure is as follows: Assume that $f(x_p)>0$ and $f(x_q)<0$, where x_p and x_q are approximate values of α such that α lies between them. A new approximate value $\bar{x}$ is then obtained from $\bar{x}=(x_qf(x_p)-x_pf(x_q))/(f(x_p)-f(x_q))$. If $f(\bar{x})>0$, then $\bar{x}\to x_p$, and if $f(\bar{x})<0$, then $\bar{x}\to x_q$, and the procedure is repeated (the symbol $\to$ means replacement). The conditions and speed of convergence of regula falsi are as follows: Let $F(x)=(x_qf(x)-xf(x_q))/(f(x)-f(x_q))$; then $F'(\alpha)=(f(x_q)+(\alpha-x_q)f'(\alpha))/f(x_q)$. If f' and f'' exist and are continuous near α, then $f(x_q)=f(\alpha)+(x_q-\alpha)f'(\alpha)+(1/2)(x_q-\alpha)^2f''(\xi)$, where $\alpha>\xi>x_q$ or $\alpha<\xi<x_q$. Accordingly, assume that $F'(\alpha)=(1/2)(x_q-\alpha)^2f''(\xi)/f(x_q)\neq 0$. Then $|F'(x)|<1$ can be satisfied if x and x_q are near α. Therefore, if the initial value is appropriate, the convergence is of the first order. In regula falsi, only calculation of $f(x)$ is necessary, while that of $f'(x)$ is unnecessary. Furthermore, if $f(x)=0$ has nearby real roots, i.e., $\alpha, \beta, \dots$ close to each other and $f'(x)$ small near the roots, there can be no mistake such as the neighboring root β being obtained in the process of finding the root α by this method. This method belongs to the inverse linear interpolation method. This kind of inverse interpolation includes Muller's method [3], which uses †Lagrange's interpolation formula, the Torii-Miyakoda method [4], which uses †Hermite's interpolation formula, and Whittaker's method [5], which uses †Stirling's interpolation formula. These methods converge more rapidly but have more complicated formulas than inverse linear interpolation. The **Sturm method** (in which the interval where the roots exist is narrowed by †Sturm's theorem) and the **Horner method** (which obtains the decimal part digit by digit), both used to obtain the real roots of a high-order algebraic equation $f(x)=0$, are also of this type.

D. The Newton-Raphson Method

In obtaining the real roots of an equation $f(x)=0$, the **Newton-Raphson method** (or the New-

ton iterative process), which converges rapidly, is used when $f'(x) \neq 0$ is computable. Let x_i be a sufficiently close ith approximation of the root α; then $x_{i+1}=x_i-f(x_i)/f'(x_i)$ is closer than x_i to the true solution. The process is repeated until $|x_{i+1}-x_i|$ is small enough. Conditions and the speed of convergence are as follows [6, 7]: Let $F(x)=x-f(x)/f'(x)$; then $F'(x)=f(x)f''(x)/(f'(x))^2$ and $F'(\alpha)=0$. Accordingly, $F''(\alpha)\neq 0$ if $f'(\alpha)\neq 0$ and $f''(\alpha)\neq 0$, and the convergence is of the second order if it is possible to determine an appropriate approximate value x_i that is close to α and satisfies $|F'(x_i)|<1$. In particular, if $f(x_0)f'(x_0)\neq 0$, $h_0=-f(x_0)/f'(x_0)$, $|f''(x)|\leqslant M$, and $|f'(x_0)|\geqslant 2|h_0|M$, then every Newton approximation x_i starting with x_0 is contained in the interval $I=[x_1-|h_0|, x_1+|h_0|]$, the equation has only one root α in I, and $x_i \to \alpha$. Besides,

$$|\alpha-x_{i+1}|\leqslant M|x_i-x_{i-1}|^2/2|f'(x_i)|, \qquad i=1,2,\ldots.$$

With regard to convergence and evaluation of the error, including roundoff error in practical computations, M. Urabe's studies should be consulted [8, 9]. In general, the convergence is of the third order in the Newton-Raphson method, which uses not only $f'(x)$ but also $f''(x)$ [10].

When $f'(x)=0$, we must assume $f''(x)\neq 0$ and use the value $f''(x)$. With regard to this case, the study by S. Hitotumatu should be consulted [11]. In addition, W. Kizner (*SIAM J. Appl. Math.*, 12 (1964)) reported an iterative process in which the convergence is of the fifth order without using any derived function higher in order than $f''(x)$. The essential part of his method lies in a numerical integration of the integral part by the †Runge-Kutta formula, based on the fact that if x_1 is the first approximate value of a root $\bar{x}$ of $f(x)=0$, then

$$\bar{x}=\int_{f(x_1)}^{0}\frac{dx}{df}df+x_1.$$

There may be several iterative processes for solving $f(x)=0$, even if the order k of the convergence is fixed. For example, in obtaining the positive root of $f(x)=x^2-a=0$ $(a>0)$, i.e., the square root $\alpha=a^{1/2}$, both iterative processes $x_{i+1}=(x_i+a/x_i)/2$ and $x_{i+1}=2x_i^3/(3x_i^2-a)$ give convergence of the second order. In obtaining the real root of $f(x)=x^3-a=0$, i.e., the cube root $\alpha=a^{1/3}$, $x_{i+1}=x_i+(a/x_i^2-x_i)/3$ converges of the second order, while $x_{i+1}=x_i/2+(a+a/2)/(2x_i^2+a/x_i)$ converges of the third order.

The Newton-Raphson method is applicable also to holomorphic functions of complex variables.

Take simultaneous equations of two variables, $f(x,y)=0$ and $g(x,y)=0$. If it is possible to transform the equations into $x=\varphi(x,y)$ and $y=\psi(x,y)$, and if $|\partial\varphi/\partial x|+|\partial\psi/\partial x|<1$, $|\partial\varphi/\partial y|+|\partial\psi/\partial y|<1$, then the iterative processes $x_{i+1}=\varphi(x_i,y_i)$ and $y_{i+1}=\psi(x_i,y_i)$ are applicable. In the Newton-Raphson method, let the corrections to the ith approximate values x_i and y_i be Δx_i and Δy_i, and solve

$$f_x(x_i,y_i)\Delta x_i+f_y(x_i,y_i)\Delta y_i=-f(x_i,y_i),$$

$$g_x(x_i,y_i)\Delta x_i+g_y(x_i,y_i)\Delta y_i=-g(x_i,y_i).$$

Then we can take $x_{i+1}=x_i+\Delta x_i$ and $y_{i+1}=y_i+\Delta y_i$ as the next approximate values. This process is repeated until both $|\Delta x_i|$ and $|\Delta y_i|$ are sufficiently small. More generally, the Newton-Raphson method for solving a system of n nonlinear equations $f_i(x_1,\ldots,x_n)=0$, $i=1,2,\ldots,n$, is defined by

$$x^{(k+1)}=x^{(k)}-J(x^{(k)})^{-1}F(x^{(k)}), \qquad k=0,1,2,\ldots,$$

where $x^{(k)}=(x_1^{(k)},\ldots,x_n^{(k)})^t$, $F(x)=(f_1(x_1,\ldots,x_n),\ldots,f_n(x_1,\ldots,x_n))^t$, $J(x)=(\partial f_i(x_1,\ldots,x_n)/\partial x_j)$ (the †Jacobian matrix of F), and where $x^{(0)}$ is chosen appropriately [23].

E. The Bairstow Method

Corresponding to a pair of complex roots $\alpha\pm i\beta$ of an algebraic equation $f(x)=a_0x^n+a_1x^{n-1}+\ldots+a_n=0$ with real coefficients, there is a real quadratic factor $x^2+p^*x+q^*$. The **Bairstow method** (or **Hitchcock method**) obtains the coefficients p^* and q^* of this quadratic factor.

To do this, first choose an appropriate quadratic factor x^2+px+q as a candidate. Then through synthetic division by the quadratic factor, b_j and c_j are computed by means of the formulas

$$b_j=a_j-pb_{j-1}-qb_{j-2}, \qquad j=0,1,\ldots,n,$$

and

$$c_j=b_j-pc_{j-1}-qc_{j-2}, \qquad j=0,1,\ldots,n-1,$$

where

$$b_{-1}=b_{-2}=c_{-1}=c_{-2}=0.$$

By solving the simultaneous equations

$$c_{n-2}\Delta p+c_{n-3}\Delta q=b_{n-1},$$

$$\bar{c}_{n-1}\Delta p+c_{n-2}\Delta q=b_n,$$

where

$$\bar{c}_{n-1}=c_{n-1}-b_{n-1},$$

the quantities Δp and Δq are obtained. Then taking $\tilde{p}=p+\Delta p$, $\tilde{q}=q+\Delta q$, we have $x^2+\tilde{p}x+\tilde{q}$ as a new candidate. This operation is repeated until Δp and Δq are sufficiently small. Since this method corresponds to the Newton-Raphson method in the case of two variables, and, accordingly, the convergence is of the

second order, a choice of suitable initial values of p and q leads to rapid convergence. The key to this method lies in choosing p and q so that $R_1(p,q)=0$ and $R_2(p,q)=0$, where R_1 and R_2 are such that R_1x+R_2 is the remainder of $f(x)$ divided by the trial quadratic factor x^2+px+q. This method was generalized by A. A. Grau (*SIAM J. Appl. Math.*, 11 (1963)). Namely, when $f(x)=(x^2+px+q)g(x)+r(x)$ with $r(x)=r_1x^{k+1}+r_2x^k$, the functions r_1 and r_2 can be used instead of R_1 and R_2. The process corresponds to the Bairstow method when $k=0$, and to the McAuley method (*SIAM J. Appl. Math.*, 10 (1962)) when $k=n-2$.

F. The Durand-Kerner Method

The **Durand-Kerner (DK) method** [12] for solving an algebraic equation $f(z)=z^n+a_1z^{n-1}+\ldots+a_n=0$ $(a_n\neq 0)$ with complex coefficients is an iterative method defined by

$$z_i^{(k+1)}=z_i^{(k)}-\frac{f(z_i^{(k)})}{\prod_{j=1,j\neq 1}^{n}(z_i^{(k)}-z_j^{(k)})},$$
$$i=1,2,\ldots,n,\quad k=0,1,2,\ldots. \quad (1)$$

A feature of this method is that it can determine simultaneously all the roots of $f(z)=0$. Let $\varphi_m(z_1,\ldots,z_n)$, $m=1,2,\ldots,n$, be the mth elementary symmetric functions with respect to $z_1,\ldots,z_n$:

$$\varphi_m(z_1,\ldots,z_n)\quad\sum_{i_1<i_2<\ldots<i_m} z_{i_1}z_{i_2}\ldots z_{i_m}.$$

Set $f_m(z)=(-1)^m\varphi_m(z_1,\ldots,z_n)-a_m$. Then $\alpha=(\alpha_1,\ldots,\alpha_n)^t$, a set of the roots of $f(z)=0$, is a solution of a system of the n equations $f_m(z)=0$, $m=1,2,\ldots,n$. I. O. Kerner [12] showed that the DK method is the Newton-Raphson method applied to the system of nonlinear equations $f_m(z)=0$, $m=1,2,\ldots,n$. Therefore the speed of convergence of (1) is of the second order, provided that it converges. The initial values $z_1^{(0)},\ldots,z_n^{(0)}$ for the DK method are usually chosen as follows: Let $g(z)=f(z-a_1/n)=z^n+c_2z^{n-2}+\ldots+c_n$ and $h(z)=z^n-|c_2|z^{n-2}-\ldots-|c_n|$. If $(c_2,\ldots,c_n)\neq(0,\ldots,0)$ $(n\geqslant 2)$, then it can be shown that $h(z)=0$ has exactly one positive root r, and all the roots of $f(z)=0$ lie in the disk $|z+a_1/n|\leqslant r$. Now, with a positive $r_0\geqslant r$ and $\theta=\pi/(2n)$, put

$$z_i^{(0)}=-\frac{a_1}{n}+r_0\exp\left[\left(\frac{2(i-1)\pi}{n}+\theta\right)\sqrt{-1}\right],$$
$$i=1,2,\ldots,n. \quad (2)$$

Such a choice was proposed by O. Aberth [13]. Hence the process (1) together with (2) can be called the **Durand-Kerner-Aberth (DKA) method**. It is shown for the DKA method that, if r_0 is large enough, then

$$z_i^{(k)}+\frac{a_1}{n}=\left(1-\frac{1}{n}\right)^k\left(z_i^{(0)}+\frac{a_1}{n}\right),\qquad i=1,2,\ldots,n,$$

hold nearly for the first several steps. Thus the DKA method has a certain kind of global convergence property that renders it one of the most powerful methods for solving algebraic equations.

A variant of the Durand-Kerner method is the Ehrlich-Aberth method [13, 14]; it is defined for $i=1, 2, \ldots, n$, $k=0, 1, 2, \ldots$ by

$$z_i^{(k+1)}=z_i^{(k)}-\frac{f(z_i^{(k)})}{f'(z_i^{(k)})-f(z_i^{(k)})\sum_{j=1,j\neq i}^{n}1/(z_i^{(k)}-z_j^{(k)})}.$$

The speed of convergence of this method is of the third order. Further variants of the Durand-Kerner and the Ehrlich-Aberth methods have been proposed by M. Iri et al. [15] and A. W. M. Nourein [16].

G. The Dejon-Nickel Method

Let $f(z)$ be a polynomial of degree n with complex coefficients, and take z_0 such that $f(z_0)\neq 0$. Then we can write

$$f(z_0+h)=f(z_0)\{1+b_ih^i+b_{i+1}h^{i+1}+\ldots+b_nh^n\}\qquad (b_i\neq 0)$$
$$=f(z_0)\{1+b_ih^i(1+\theta)\},\qquad \theta=o(h).$$

Now, choose h sufficiently small so that $|b_ih^i|<1$, $|\theta|<1$, and $\arg(b_ih^i)=\pi$. Then $|1+b_ih^i|=1-|b_ih^i|$ and $|b_ih^i\theta|<|b_ih^i|$ so that $|f(z_0+h)|\leqslant|f(z_0)|\{|1+b_ih^i|+|b_ih^i\theta|\}<|f(z_0)|$. Hence, if z_0 is chosen so that $\min|f(z)|=|f(z_0)|$, then we must have $f(z_0)=0$. This is the outline of Cauchy's existence proof for the roots of algebraic equations. The **Dejon-Nickel method** [17] is a method in which h is chosen as follows:

$$h=(-1/b_k)^{1/k},$$

where

$$\min\{|1/b_j|^{1/j}\mid b_j\neq 0, i\leqslant j\leqslant n\}=|1/b_k|^{1/k}\qquad(=r,\text{ say}).$$

If several such k exist, take the smallest one. The branch of the multivalued function $(-1/b_k)^{1/k}$ is chosen such that $z^{1/k}$ is positive for z positive. If $|f(z_0+h)|\leqslant(1-\varepsilon)|f(z_0)|$ with a preassigned constant ε such that $0<\varepsilon<1$, then put $z_1=z_0+h$. If the inequality does not hold, then for each integer $m\geqslant 1$ choose the smallest integer $l=l(m)$ such that $\max\{|b_j|(r/2^m)^j\mid i\leqslant j\leqslant n\}=|b_l|(r/2^m)^l$, and put $\tilde{z}_m=z_0+(r/2^m)(-|b_l|/b_l)^{1/l}$. Find the smallest integer $m\geqslant 1$ such that $|f(z_0)|\leqslant(1-2^{-1}(r/2^m)^l|b_l|)|f(z_0)|$, and put

$z_1 = \tilde{z}_m$. By continuing this process, a sequence $\{z_j\}$ is constructed such that $|f(z_0)| > |f(z_1)| > |f(z_2)| > \ldots$. It converges toward some root of the equation $f(z) = 0$. S. Hirano has proposed a similar method.

H. Methods for Finding Good First Estimates of Roots

Some of the principal methods for obtaining good first approximate values of the roots are given in the following sections I–N.

I. Matrix Methods

The problem of solving an algebraic equation $f(z) = z^n + a_1 z^{n-1} + \ldots + a_n = 0$ with complex coefficients is equivalent to that of finding the eigenvalues of the **companion matrix**

$$A = \begin{bmatrix} -a_1 & -a_2 & -a_3 & \ldots & -a_{n-1} & -a_n \\ 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \ldots & 1 & 0 \end{bmatrix}.$$

Therefore numerical methods for solving the eigenvalue problems for nonsymmetric matrices are applicable (→ 298 Numerical Computation of Eigenvalues). However, it should be remarked that this might be inefficient because most of the elements of the matrix A are zero [18].

J. The Bernoulli Method

In the **Bernoulli method** the iterative formulas $S_1 = -a_1$, $S_k = -(a_1 S_{k-1} + a_2 S_{k-2} + \ldots + a_{k-1} S_1 + k a_k)$ $(k = 2, 3, \ldots, n)$, $S_k = -a_1 S_{k-1} - a_2 S_{k-2} - \ldots - a_n S_{k-n}$ $(k = n+1, n+2, \ldots)$ are repeatedly applied to an algebraic equation $f(x) = x^n + a_1 x^{n-1} + a_2 x^{n-2} + \ldots + a_n = 0$. When the roots of $f(x) = 0$ are $\alpha_1, \alpha_2, \ldots, \alpha_n$ $(|\alpha_1| \geqslant |\alpha_2| \geqslant \ldots \geqslant |\alpha_n|)$, $S_k / S_{k-1} \to \alpha_1$ if α_1 is a real and simple root and $|\alpha_1| > |\alpha_2|$. When α_1, α_2 are complex roots, we put $\alpha_1 = Re^{i\theta}$, $\alpha_2 = Re^{-i\theta}$. If $|\alpha_3| < R$, then we have

$$\frac{S_k^2 - S_{k+1} S_{k-1}}{S_{k-1}^2 - S_k S_{k-2}} \to R^2,$$

$$\frac{S_k S_{k-1} - S_{k+1} S_{k-2}}{S_{k-1}^2 - S_k S_{k-2}} \to 2R\cos\theta,$$

and hence the two roots of $x^2 - (2R\cos\theta)x + R^2 = 0$ are α_1 and α_2. S_k is the sum of the kth powers of the n roots of the equation, $\alpha_1, \alpha_2, \ldots, \alpha_n$. C. Lanczos used $xf'(x)/f(x) = n + S_1/x + S_2/x^2 + \ldots + S_k/x^k + \ldots$ to compute S_k [10, pp. 26–30]. Let the eigenvalues of the companion matrix A be $\lambda_1, \lambda_2, \ldots, \lambda_n$ $(|\lambda_1| \geqslant |\lambda_2| \geqslant \ldots \geqslant |\lambda_n|)$. Then the Bernoulli method is the same as the †power method for obtaining the maximum characteristic value λ_1 (→ 298 Numerical Computation of Eigenvalues C).

K. The Lehmer Method

The **Lehmer method** [9] is a general method for finding the n roots of an algebraic equation with complex coefficients

$$f(z) = a_0 + a_1 z + a_2 z^2 + \ldots + a_n z^n = 0$$

in the complex plane by repeating the following procedure: First, draw a circle whose center is at the origin and whose radius is $R = r2^{-\theta}$, where r is an arbitrary given number and θ an arbitrary given integer. Then, utilizing the Lehmer theorem, observe whether a root α of $f(z) = 0$ lies inside the circle. If there is no such root, replace R with $2R$, whereas if such a root exists, replace R with $R/2$. Continue this process until an R for which a root α exists within the annulus $R < |z| < 2R$ is obtained.

Second, draw circles with the centers $\beta_k = (5R/3)\exp(i2\pi k/8)$ $(k = 0, 1, \ldots, 7)$ and common radius $\mu = (5R/3)/2$ and cover the annulus obtained by the first step. Then find out which of the circles has the root α in its interior. If α is in the interior of the circle for $k = j$, the origin is shifted to β_j and the operation is started again from the first step. If R_1 satisfying $R_1 < |z - \beta_j| < 2R_1$ is obtained by the first step, we have $R_1 \leqslant (5/12)R$. Therefore, when the first step is repeated N times, the root α is confined in a small circle whose radius is smaller than $2(5/12)^N R$. Then the center β of this small circle gives a good approximate value of α.

L. The Downhill Method

The downhill method is a method for obtaining the extreme values of a function of many variables. It is also applicable in obtaining approximate solutions of a system of equations. Let us consider the **downhill method** in the case of two variables. The problem of obtaining the real roots of simultaneous equations $f(x, y) = 0$ and $g(x, y) = 0$ can be reduced to the problem of obtaining the coordinates (α, β) which give the extreme value 0 of $\Phi(x, y)$, where $\Phi(x, y) = f^2 + g^2$. The values of $\Phi(x, y)$ are calculated at 3^2 points obtained by the combination of $x = x_1, x_1 \pm h$; $y = y_1, y_1 \pm h$, where (x_1, y_1) are arbitrary approximate values of (α, β) constituting the centers of those sets of points, and h is the given step size. Utilizing the values of the function at the 3^2 points,

$\Phi(x, y)$ is approximated by a quadratic surface

$$b_0+b_1x+b_2y+b_{11}(3x^2-2)+b_{22}(3y^2-2)+b_{12}xy.$$

The values of $b_0, b_1, b_2, b_{11}, b_{22}, b_{12}$ are calculated by the †method of least squares, and the center (x_1^*, y_1^*) of the approximate quadratic surface is obtained. Then the first approximations x_1 and y_1 are replaced by the second approximations $x_1+x_1^*$ and $y_1+y_1^*$. This process is repeated while the step size h is made appropriately smaller. This method is an improvement on the method of successive experimental planning used by G. E. P. Box and K. B. Wilson (1954) to obtain optimum conditions in the exploration of response curved surfaces.

By using another minimization technique, J. A. Grant and G. D. Hitchins [21] gave an ever-convergent algorithm for determining initial approximations to the roots of algebraic equations with real coefficients. The practical implementation of this method is given in J. A. Grant and G. D. Hitchins (*Comput. J.*, 18 (1975)).

M. Continuation Methods

Let $F(x)=(f_1(x), \ldots, f_n(x))^t=0$ $(x=(x_1, \ldots, x_n))^t$ be a system of n equations. Suppose that no reasonable approximation for a solution exists. Then, take arbitrary $x^{(0)}$ and define a one-parameter family of equations

$$H(x,t)\equiv F(x)-(1-t)F(x^{(0)})=0, \qquad 0\leqslant t\leqslant 1. \tag{3}$$

Suppose that for each t the equation has a solution $x(t)$ which depends continuously on t. Observe then that $x(0)=x^{(0)}$ and $x(1)$ is a solution for the equation $F(x)=0$. Partition the interval $[0, 1]$ by the points $0=t_0<t_1<\ldots<t_N=1$. First, solve $H(x, t_1)=0$ by some iterative method using $x^{(0)}$ as a first approximation. Let $x^{(1)}$ be a solution thus obtained. Next, solve $H(x, t_2)=0$ by some iterative method using $x^{(1)}$ as a first approximation, and so on. Finally, solve $H(x, t_N)=0$ by some iterative method which uses $x^{(N-1)}$ as a first approximation. Then $x^{(N)}$, a solution thus obtained, can be used as a first approximation in an iterative method applied to the equation $F(x)=0$. This method is called a **continuation method** [22–24].

As another approach, differentiate (3) with respect to t. Then $J(x(t))x'(t)+F(x^{(0)})=0$, where $J(x(t))$ is the Jacobian matrix of F, evaluated at $x=x(t)$. Hence $x(t)$ is the solution of the system of ordinary differential equations

$$x'(t)=-J(x(t))^{-1}F(x^{(0)}), \qquad 0\leqslant t\leqslant 1,$$

subject to the initial condition $x(0)=x^{(0)}$, provided that $J(x(t))$ is nonsingular. Therefore numerical solution at $t=1$ gives a good approximation for a solution of $F(x)=0$. This is called **Davidenko's method of differentiation with respect to a parameter**. These methods can be used to find initial approximations of the Durand-Kerner and the Ehrlich-Aberth methods. They are also applicable to a single algebraic equation $f(x)=0$, which is the special case $n=1$.

N. Other Methods

To obtain the first approximate values of the roots of an algebraic equation, the Graeffe method has been used, in which the roots of the equation are separated by successively forming an equation whose roots are the squares of the roots of the preceding equation. In computer calculations, however, the other methods described in previous sections are more convenient than the Graeffe method.

The **Lanczos method** [10] is a method in which $g(x)$, an approximate function of $f(x)$, is obtained by the process of approximating functions, and the roots of $g(x)=0$ are taken as approximate values of the roots of $f(x)=0$. The Garside-Jarratt-Mack method [25] is a modification of the Lanczos method and approximates $f(x)/f'(x)$ by a rational function $g(x)=(x-a)/(b+cx)$.

O. Error Bounds for Computed Solutions

Let $z_1, \ldots, z_n$ be computed solutions of an algebraic equation $f(z)=0$ which were obtained by some method. If $z_1, \ldots, z_n$ are distinct, then the following result due to B. T. Smith [26] is quite useful for estimating the errors of z_i: Let

$$\Gamma_i: |z-z_i|\leqslant\gamma_i\equiv\frac{n|f(z_i)|}{\prod_{j=1,j\neq i}^{n}|z_i-z_j|}.$$

Then the union of the disks Γ_i contains all the roots $\alpha_1, \ldots, \alpha_n$ of $f(z)=0$. Any connected component of $\bigcup_{i=1}^{m}\Gamma_i$, which consists of just m disks Γ_i, contains exactly m roots of $f(z)=0$. Hence if $\Gamma_i\cap\Gamma_j=\varnothing$ for every $j\neq i$, then $|\alpha_i-z_i|\leqslant\gamma_i$. Smith obtained a more general result for the case where $z_1, \ldots, z_n$ are not necessarily distinct.

References

[1] J. F. Traub, A class of global convergent iteration functions for solution of polynomial equations, Math. Comp., 20 (1966), 113–138.
[2] B. Dejon and P. Henrici (eds.), Construc-

tive aspects of the fundamental theorem of algebra, Wiley-Interscience, 1969.
[3] D. E. Muller, A method for solving algebraic equations using an automatic computer, Math. Tables Aids Comput., 10 (1956), 208–215.
[4] T. Torii and T. Miyakoda, A practical root-finding algorithm based on the cubic Hermite interpolation, Information Processing in Japan, 13 (1973), 143–148.
[5] E. T. Whittaker and G. Robinson, Calculus of observations, Van Nostrand, 1924.
[6] L. B. Rall, Computational solution of nonlinear operator equations, Krieger, 1979.
[7] A. M. Ostrowski, Solution of equations in Euclidean and Banach spaces, Academic Press, 1973.
[8] M. Urabe, Error estimation in numerical solution of equations by iteration process, J. Sci. Hiroshima Univ., ser. A., 26 (1962), 77–91.
[9] M. Urabe, Component-wise error analysis of iterative methods practiced on a floating-point system, Mem. Fac. Sci. Kyushu Univ., 27 (1973), 23–64.
[10] C. Lanczos, Applied analysis, Prentice-Hall, 1956.
[11] S. Hitotumatu, A method of successive approximation based on the expansion of second order, Math. Japonicae, 7 (1962), 31–50.
[12] I. O. Kerner, Ein Gesamtschrittverfahren zur Berechnung der Nullstellen von Polynomen, Numer. Math., 8 (1966), 290–294.
[13] O. Aberth, Iteration methods for finding all zeros of a polynomial simultaneously, Math. Comp., 27 (1973), 339–344.
[14] L. W. Ehrlich, A modified Newton method for polynomials, Comm. ACM, 10 (1967), 107–108.
[15] M. Iri, H. Yamashita, T. Terano, and H. Ono, An algebraic-equation solver with global convergence property, Publ. Res. Inst. Math. Sci., 339 (1978), 43–69.
[16] A. W. M. Nourein, An improvement on two iteration methods for simultaneous determination of the zeros of a polynomial, Int. J. Computer Math., 6 (1977), 241–252.
[17] B. Dejon and K. Nickel, A never failing, fast convergent root-finding algorithm, Constructive Aspects of the Fundamental Theorem of Algebra, B. Dejon and P. Henrici (eds.), Wiley-Interscience, 1969, 1–35.
[18] D. M. Young and R. T. Gregory, A survey of numerical mathematics I, Addison-Wesley, 1972.
[19] D. H. Lehmer, A machine method for solving polynomial equations, J. ACM, 8 (1961), 151–162.
[20] O. L. Rasmussen, Solution of polynomial equations by the method of D. H. Lehmer, BIT, 4 (1964), 250–260.
[21] J. A. Grant and G. D. Hitchins, An always convergent minimization technique for the solution of polynomial equations, J. Inst. Math. Appl., 8 (1971), 122–129.
[22] J. M. Ortega, Numerical analysis, a second course, Academic Press, 1972.
[23] J. M. Ortega and W. C. Rheinboldt, Iterative solution of nonlinear equations in several variables, Academic Press, 1970.
[24] R. Wait, The numerical solution of algebraic equations, Wiley, 1979.
[25] G. R. Garside, P. Jarratt, and C. Mack, A new method for solving polynomial equations, Comput. J., 11 (1968), 87–89.
[26] B. T. Smith, Error bounds for zeros of a polynomial based upon Gerschgorin's theorems, J. ACM, 17 (1970), 661–674.
[27] A. S. Householder, The numerical treatment of a single nonlinear equation, McGraw-Hill, 1970.
[28] J. H. Wilkinson, Rounding errors in algebraic processes, HMSO, 1963.
[29] J. H. Wilkinson, Algebraic eigenvalue problem, Oxford Univ. Press, 1965.
[30] P. Rabinowitz (ed.), Numerical methods for nonlinear algebraic equations, Gordon & Breach, 1970.

302 (XV.4) Numerical Solution of Linear Equations

A. Condition of Linear Systems

The solution of the system of linear algebraic equations

$$\sum_{j=1}^{n} a_{ij}x_j = b_i; \qquad i=1,\dots,n, \quad a_{ij}, b_i \text{ real}, \tag{1}$$

which may be written in the matrix form

$$Ax=b; \quad A=(a_{ij}), \qquad x=(x_i), \quad b=(b_i), \tag{1'}$$

is expressed as quotients of determinants by †Cramer's rule. In practice, however, this form of the solution is of little value for numerical computation, because the direct evaluation of the determinants involves $(n+1)!(n-1)$ multiplications which, even for a moderate-sized system, amount to a prohibitive number of arithmetic operations to be executed, even with modern high-speed computers; besides, it requires high-precision calculation. There are a variety of practical methods of solving efficiently the system (1) with finite-precision calculation. These numerical methods are generally divided into two classes: direct methods and iterative methods.

Whatever method is used, inherent difficulties are encountered when the solution of the system is unstable. The instability is usually measured by the **condition number** of the coefficient matrix, which is defined by

$$\operatorname{cond} A \equiv \sigma_1/\sigma_n \quad (= \|A\|\,\|A^{-1}\| \geqslant 1), \tag{2}$$

where the σ_i ($\sigma_1 \geqslant \sigma_2 \geqslant \ldots \geqslant \sigma_n \geqslant 0$) are the nonnegative square roots of the †eigenvalues of A^tA (called the **singular values** of A), $\|A\| = \max\{\|Ax\|/\|x\| : x \neq 0\}$, and $\|x\| = \sqrt{x^t x}$. This definition is valid also for an $m \times n$ matrix A ($m \geqslant n$). The condition number $\operatorname{cond} A$ satisfies

$$\frac{\|\delta x\|}{\|x\|} \leqslant \frac{\operatorname{cond} A}{1 - \dfrac{\|\delta A\|}{\|A\|}\operatorname{cond} A}\left(\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta b\|}{\|b\|}\right),$$

where $Ax = b$, $(A + \delta A)(x + \delta x) = b + \delta b$, and $\|\delta A\|\,\|A^{-1}\| < 1$. As the condition number increases, solution processes become more susceptible to errors. If $\operatorname{cond} A$ is large, the system is called **ill-conditioned**.

B. Direct Methods

A direct method is one that yields an exact solution in a finite number of arithmetic operations if they are performed without roundoff error. Among the existing direct methods, the one known as **Gaussian elimination** with pivoting, which is based on systematic elimination of unknowns of the equation (1), is found to be the best with respect to time or accuracy. Its dependability was re-established by means of **backward error analysis** (→ 138 Error Analysis; [1–3]). The method generates successive vectors $b^{(k)} = (b_i^k)$ and matrices $A^{(k)} = (a_{ij}^k)$, typically of the form

$$A^{(2)} = \begin{pmatrix} * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & * & * & * \end{pmatrix}$$

when $n = 5$ and $k = 2$. At the kth stage, a pivotal element $a_{ij}^{k-1} \neq 0$ ($i, j \geqslant k$) is chosen in one way or another. Then the ith and kth rows and the jth and kth columns of $A^{(k-1)}$ are interchanged so that a_{ij}^{k-1} becomes a_{kk}^k. The ith and kth elements of $b^{(k-1)}$ are also interchanged. The **pivot** a_{kk}^k is then used to eliminate all the nonzero entries in its column below the diagonal as follows:

$$a_{ij}^k = a_{ij}^{k-1}, \quad i = 1, \ldots, k, \; j = 1, \ldots, n;$$

$$b_i^k = b_i^{k-1}, \quad i = 1, \ldots, k;$$

$$a_{ij}^k = a_{ij}^{k-1} - (a_{ik}^{k-1}/a_{kk}^k)a_{kj}^k, \qquad i = k+1, \ldots, n, \; j = k, \ldots, n;$$

$$b_i^k = b_i^{k-1} - (a_{ik}^{k-1}/a_{kk}^k)b_k^k, \qquad i = k+1, \ldots n.$$

Starting with $A^{(0)} = A$ and $b^{(0)} = b$, a system of linear equations with an upper-triangular coefficient matrix,

$$A^{(n-1)}x^* = b^{(n-1)},$$

is produced at the stage $k = n-1$, where x^* is a permutation of x, caused by interchanging columns. This part of the process is called **forward elimination**. If $a_{ij}^{k-1} = 0$ for all $i, j \geqslant k$ at some stage, then A is a singular matrix of †rank $k-1$ and the system (1) admits infinitely many solutions if $b_i^{k-1} = 0$ for $i = k, \ldots, n$, and no solution otherwise. If this is not the case, A is a nonsingular matrix with $\det A = a_{11}^{n-1} a_{22}^{n-1} \ldots a_{nn}^{n-1}$, and the solution of (1) is given by

$$x_i^* = \left(b_i^{n-1} - \sum_{k=i+1}^{n} a_{ik}^{n-1} x_k^*\right)\Big/ a_{ii}^{n-1}$$

for $i = n, n-1, \ldots, 1$, which is called **back substitution**. Taking the pivot to be an element of the largest absolute value among column elements a_{ik}^{k-1} ($i \geqslant k$) at each stage is called the **partial pivoting** strategy. In **complete pivoting**, the pivot is taken to be an element of the largest absolute value among a_{ij}^{k-1} ($i, j \geqslant k$). These pivotings are introduced to prevent loss of accuracy due to rapid growth of elements of successive $A^{(k)}$. Although a smaller bound for the growth factor is obtained for complete pivoting, in practice partial pivoting appears to be entirely adequate.

If rows or columns are not interchanged in the process, forward elimination effectively produces a factorization of A into the product of a lower triangular matrix L and an upper triangular matrix U, i.e.,

$$A = LU, \tag{3}$$

where L has unit diagonal elements and $U = A^{(n-1)}$. This factorization is computed directly by the **Doolittle method** without calculating the intermediate $A^{(k)}$. The **Crout method** also produces a similar factorization (3) in which U has unit diagonal elements. The **Cholesky method** determines a similar factorization (3) of a †positive definite matrix A, in which $U = L^t$. Once the **triangular factorizations** (3) are formed, the solution of the system (1) is determined by solving the two triangular systems $Ly = b$ and $Ux = y$ successively.

The number of multiplicative operations required for these factorizations are about $n^3/3$ for forward elimination, Dootittle's method, and Crout's method, and about $n^3/6$ for Cholesky's method. The solution of each triangular system requires about $n^2/2$ multiplicative operations. Special properties of A, such as a banded structure, can be exploited to reduce the number of operations and memory requirements considerably [4]. **Gauss-Jordan elimination** is similar to Gaussian elimina-

tion, but the elements above the diagonal are also eliminated to dispense with back substitution. However, it requires about $n^3/2$ multiplications. Modern computers can easily handle problems of size $n=100$ by these direct methods.

C. Iterative Methods

An iterative method is a dynamical process that generates a sequence of approximations $\{x^k\}$ converging to the exact solution. At each step of the iteration, an improved approximation is obtained from the previous ones. The accuracy of the solution depends on the number of iterations performed. Most iterative methods retain the coefficient matrix in its original form throughout the process and hence have the advantage of requiring minimal memory. They are suitable for solving large **sparse** systems arising in finite-dimensional approximations to †partial differential equations, where A is sparse if most of its elements are zero.

Linear stationary iterative processes are the most frequently used iterative methods. A method in this class is written in the form

$$x^k = x^{k-1} + R(b - Ax^{k-1}), \qquad k = 1, 2, \ldots, \tag{4}$$

where R is chosen to approximate the inverse of A. Usually, R and $b - Ax^{k-1}$ are not expressed explicitly in the actual algorithms. If the †spectral radius of the **iteration matrix** $I - RA$ is less than one, the method converges to the solution for an arbitrary initial approximation x^0 [5]. The key matrix R can be chosen quite freely as long as this condition is satisfied. In the **Richardson method**, $R = \alpha A^t$, $0 < \alpha < 2/\|A\|^2$. $R = (A+E)^{-1}$ is the usual choice, in which E is a perturbing matrix that makes $A+E$ easily invertible. In the **Gauss-Seidel method** and the **Jacobi method**, R is chosen to be the inverse of the lower triangular and the diagonal submatrices of A, respectively, i.e., $R = (L+D)^{-1}$ and $R = D^{-1}$, where $L = (a_{ij})$ $(i > j)$ and $D = (a_{ii})$.

Direct methods can be combined with the iterative method (4) to obtain an approximate inverse R. In the course of factorization by a direct method, an artificial perturbation E is introduced to produce an **incomplete factorization**,

$$\tilde{L}\tilde{U} = A + E, \tag{5}$$

so that $\tilde{L}$ and $\tilde{U}$ are of low †complexity of computation. A combined method is then constructed by putting $R = \tilde{U}^{-1}\tilde{L}^{-1}$. Even if E is not added intentionally, each of the direct methods actually produces an incomplete factorization (5), in which E is a matrix of small elements accounting for the effect of roundoff errors [1–4]. In this case, the combined method is called the **iterative improvement** of the direct method. If applied to a first solution $x^0 = \tilde{U}^{-1}\tilde{L}^{-1}b$, it produces more accurate solution in a few steps with only a modest increase in computation time, when the system is not too ill-conditioned [4]. It is essential, however, that the residual $b - Ax^0$ be computed with higher precision.

Various convergence criteria have been established for a variety of methods [5, 6]. For example, the Gauss-Seidel method is convergent for a symmetric positive definite matrix A. The smaller the spectral radius, the faster the convergence of the method. In general, a larger amount of computation is required in each iteration to get faster convergence. If the spectral radius $\rho(I - RA)$ is close to one, the convergence is slow, and an acceleration of the process is needed. **SOR** (successive overrelaxation) is an accelerated version of the Gauss-Seidel method, in which $R_\omega = (L + \omega^{-1}D)^{-1}$ and the optimum **acceleration parameter** ω $(0 < \omega < 2)$ is chosen to minimize $\rho(I - R_\omega A)$ (→ 304 Numerical Solution of Partial Differential Equations; [6]). There is an adaptive acceleration method [7], which, if applied to a scalar sequence, reduces to Aitken's δ^2-method.

D. The Conjugate Gradient Method

The **conjugate gradient (CG) method** is a nonlinear stationary iterative method for solving a system with a symmetric positive definite coefficient matrix. The method generates $\{x^k\}$, $\{r^k\}$, and $\{p^k\}$ by means of the formulas

$$\begin{aligned}
&\alpha_k = (r_k, r_k)/(p^k, Ap^k),\\
&x^{k+1} = x^k + \alpha_k p^k,\\
&r^{k+1} = b - Ax^{k+1} = r^k - \alpha_k Ap^k,\\
&\beta_k = (r^{k+1}, r^{k+1})/(r^k, r^k),\\
&p^{k+1} = r^{k+1} + \beta_k p^k,
\end{aligned} \tag{6}$$

where x^0 is arbitrary and $p^0 = r^0 = b - Ax^0$, $(r, s) = s^t r$.

The CG method shares a feature with the direct method. In theory, $\{x^k\}$ converges to the solution in less than n steps, and the p^i are mutually **conjugate**, i.e., $(p^j)^t Ap^i = 0$ $(i \neq j)$. When Hestenes and Stiefel [9] proposed the method in 1952, they created a great sensation because of the method's theoretical elegance. The CG method turned out, however, to be highly sensitive to roundoff errors. In practice, nice theoretical properties, such as finite termination, do not hold in the presence of error. Recently, the CG method has regained its

popularity as an iterative method for solving large sparse systems. The iteration (6) is usually restarted periodically to accelerate the convergence. It may converge in a small number of steps, as it takes advantage of the distribution of eigenvalues of A adaptively. The CG method is most effective when it is used to accelerate linear stationary iterative methods or when it is applied to the **preconditioned** system

$$\tilde{L}^{-1}A(\tilde{L}^{-1})^t y = L^{-1}b, \quad x = (L^{-1})^t y, \tag{7}$$

where $\tilde{L}$ is computed by the incomplete Cholesky factorization $\tilde{L}\tilde{L}^t = A + E$. The matrix $\tilde{L}^{-1}A(\tilde{L}^{-1})^t$ is not to be formed explicitly in the CG method.

E. Linear Least Squares Problem

Let $Ax = b$ be an overdetermined system of linear equations, where A is an $m \times n$ matrix ($m \geqslant n$). The **linear least squares problem** is to find the x that minimizes the Euclidean norm $\|b - Ax\|$. We assume here that A is of rank n. The most straightforward method for solving the problem is to apply the Cholesky method to the **normal equation**

$$(A^t A)x = A^t b. \tag{8}$$

This may result in ill-conditioning of the system, since $\operatorname{cond} A^t A = (\operatorname{cond} A)^2$. Ill-conditioning can sometimes be avoided by forming the factorization $A = LU$ by Gaussian elimination, where L is an $m \times n$ lower trapezoidal matrix and U is an upper triangular matrix. The least squares solution is then obtained by solving successively the systems $L^t Ly = L^t b$ and $Ux = y$.

Another approach to avoiding ill-conditioning is based on orthogonalization. The modified Gram-Schmidt orthogonalization method produces in an effective manner the factorization

$$A = QU, \tag{9}$$

where Q is an $m \times n$ matrix whose columns are a set of †orthonormal vectors and U is an upper triangular matrix [10, 11]. The least squares solution is obtained by solving $Ux = Q^t b$. A sequence of **Householder transformations**, $P_k = I - 2w_k w_k^t / \|w_k\|^2$, can produce the factorization

$$P_n P_{n-1} \ldots P_2 P_1 A = \begin{pmatrix} U \\ \cdots \\ 0 \end{pmatrix}, \tag{10}$$

where U is an $n \times n$ upper triangular matrix [12]. The least squares solution is obtained by solving $Ux = \bar{b}$, where $\bar{b}$ is formed by the first n elements of the vector $P_n P_{n-1} \ldots P_2 P_1 b$. A similar factorization can also be produced by using a sequence of **Givens transformations**, which have the advantage of exploiting the sparseness of A. The matrices U in (9) and (10) are unique and coincide with the transpose L^t of the lower triangular matrix L of the Cholesky factorization, $A^t A = LL^t$, of $A^t A$. The multiplicative operations required for the factorizations (9) and (10) are about mn^2 and $mn^2 - n^3/3$, respectively.

When the rank of the matrix is unknown, we can determine an "effective rank" p of A, based on the **singular value decomposition (SVD)**:

$$A = UDV^t,$$

where U and V are †orthogonal matrices of dimensions m and n, respectively, and D is an $m \times n$ diagonal matrix whose diagonal elements d_{ii} are the singular values σ_i of A [13]. If singular values smaller than σ_p can be ignored, the approximate least squares solution is given by $\tilde{x}^p = V\tilde{D}_p^+ U^t b$, where the ith element $\tilde{d}_{ii}^p$ of the diagonal matrix $\tilde{D}_p^+$ is σ_i^{-1} if $i \leqslant p$ and 0 otherwise.

These methods certainly solve the system (1) but generally require more computational work than those based on the triangular factorization. The least squares solution is computed also by applying the CG method (6) to the normal equation (8), in which $A^t A$ is not to be formed explicitly [9]. Iterative methods for solving linear systems with singular and rectangular coefficient matrices are characterized in terms of the †range of RA and the †null space of AR [15].

Besides the methods discussed here, numerous other methods have been proposed for solving the system (1) [16–18]. Recently, various sparse techniques have been developed for direct methods to control the growth of nonzero entries (**fill-in**) in the process of matrix factorization [19]. In general, direct methods have an advantage over iterative methods; however, their relative computational efficiency varies according to the scale, sparsity, and type of the coefficient matrix and to the available computational devices.

References

[1] J. H. Wilkinson, Error analysis of direct methods of matrix inversion, J. ACM, 8 (1961), 281–330.

[2] J. H. Wilkinson, Rounding errors in algebraic processes, Prentice-Hall, 1963.

[3] J. H. Wilkinson, The algebraic eigenvalue problem, Oxford Univ. Press, 1965.

[4] G. E. Forsythe and C. B. Moler, Computer solution of linear algebraic systems, Prentice-Hall, 1967.

[5] R. S. Varga, Matrix iterative analysis, Prentice-Hall, 1962.
[6] D. M. Young, Iterative solution of large linear systems, Academic Press, 1971.
[7] K. Tanabe, An adaptive acceleration of general linear iterative process for solving a system of linear equations, Proc. Fifth Hawaii Int. Conf. System Sci., (1972), 116–118.
[8] C. Lanczos, An iteration method for the solution of the eigenvalue problem of linear differential and integral operators, J. Res. Nat. Bur. Standards, sec. B, 45 (1950), 255–282.
[9] M. R. Hestenes and E. Stiefel, Methods of conjugate gradients for solving linear systems, J. Res. Nat. Bur. Standards, sec. B, 49 (1952), 409–436.
[10] F. L. Bauer, Elimination with weighted row combinations for solving linear equations and least squares problems, Numer. Math., 7 (1965), 338–352.
[11] A. Björck, Solving linear least squares problems by Gram-Schmidt orthogonalization, BIT, 7 (1967), 1–21.
[12] G. H. Golub, Numerical methods for solving linear least squares problems, Numer. Math., 7 (1965), 206–216.
[13] G. Golub and W. Kahan, Calculating the singular values and pseudoinverse of a matrix, SIAM J. Numer. Anal., 2 (1965), 205–224.
[14] C. L. Lawson and R. J. Hanson, Solving least squares problems, Prentice-Hall, 1974.
[15] K. Tanabe, Characterization of linear stationary iterative processes for solving a singular system of linear equations, Numer. Math., 22 (1974), 349–359.
[16] D. K. Faddeev and V. N. Faddeeva, Computational methods of linear algebra, Freeman, 1963. (Original in Russian, 1963.)
[17] G. E. Forsythe, Solving linear algebraic equations can be interesting, Bull. Amer. Math. Soc., 59 (1953), 299–329.
[18] J. R. Westlake, A handbook of numerical matrix inversion and solution of linear equations, Wiley, 1968.
[19] J. K. Reid (ed.), Large sparse sets of linear equations, Academic Press, 1971.

303 (XV.8) Numerical Solution of Ordinary Differential Equations

A. General Remarks

Applications of the numerical solution of †ordinary differential equations include †initial value problems, †boundary value problems, and †eigenvalue problems. Although solution by an †analog computer is among the numerical methods in the wide sense, we usually mean by "numerical solution" a solution obtained by approximating a problem of infinite degrees of freedom and a continuous variable by one of finite degrees of freedom. The approximation of an arbitrary function by a linear combination of a given finite system of functions is an example (→ Section I). Successive substitution and power series expansion are also used but have no particular advantages over other methods. Among various methods of approximation, the so-called **difference methods** (or **discrete variable methods**) are the most flexible and have the largest field of application. A difference method reduces a given problem to an approximate problem in which we deal only with a systematically chosen set of discrete values of the variable and with values of the unknown functions at the chosen points. A difference method is usually carried out by a simple iterative computation, so that it is suitable for †digital computers. We confine ourselves almost exclusively to difference methods [1–7].

B. Initial Value Problems

To solve numerically general initial value problems, it suffices to consider the problem of determining numerically the values on an interval $[a,b]$ of m functions $y^i(x)$ $(i=1,\dots,m)$ that satisfy a system of ordinary differential equations of the form $y^{i\prime}(x)=f^i(x,y^1(x),\dots,y^m(x))$ and the initial conditions $y^i(a)=\eta^i$ $(i=1,\dots,m)$, where the f^i are given functions with appropriate smoothness, and a, b, and η^i are given constants. We define the mesh points $x_n=a+nh$ $(n=0,1,\dots)$, call h the **step size**, and determine numerically the values y_n^i approximating $y^i(x_n)$.

Assuming that the y_n^i were computed with infinite accuracy, i.e., without †roundoff error, we call $e_n^i\equiv y_n^i-y^i(x_n)$ the global **truncation error** (or global **discretization error**), while we call $r_n^i\equiv\tilde{y}_n^i-y_n^i$ the global **roundoff error**, where the $\tilde{y}_n^i$ are the values we actually have by means of finite-accuracy computation. On the other hand, if the assumption holds locally that the solutions at the previous steps are exact, we call the e_n^i and r_n^i the local truncation error (or local discretization error) and the local roundoff error, respectively. To avoid confusion, we denote the local truncation error by t_n^i. The numerical solution must have the property that, for every $x\in[a,b]$, $\lim_{h\to0,\,x_n=x}e_n^i=0$, which is called the condition of **convergence**. In particular, if $t_n^i=O(h^{p+1})$ $(x_n=x, h\to 0$, where $O(h^{p+1})$ is Landau's symbol), then p

is called the **order** of the solution process.

We shall show the typical methods of numerical solution, using the abbreviation f_n^i for $f^i(x_n, y_n^j)$.

C. Overall Approximation

If we rewrite the differential equations on the interval $[x_n, x_{n+P}]$ as $y^i(x_{n+q}) - y^i(x_n) = \int_{x_n}^{x_{n+q}} f^i(x, y^j(x))\,dx$, $q = 1, \ldots, P$, and substitute suitable †numerical integration formulas for the integrals on the right-hand side, then we have the **overall approximation formulas**:

$$y_{n+q}^i - y_n^i = h \sum_{r=0}^{P} c_{qr} f_{n+r}^i, \qquad q = 1, \ldots, P,$$

from which we can obtain $y_{n+1}^i, \ldots, y_{n+P}^i$ when the y_n^i are given. Since f_{n+r}^i contains y_{n+r}^i, these equations are nonlinear in y_{n+q}^i, so that we have to use, for example, the iterative procedure of starting from suitable 0th approximations and adopting as the lth approximations of the y_{n+q}^i values computed from the overall approximation formulas by substituting the $(l-1)$th approximations for the y_{n+r}^j in the f_{n+r}^i on the right-hand sides. These formulas are often used to determine a set of starting values for a multistep method. The truncation error of the formulas depends on that of the numerical integration formula substituted for the integral. A few examples: $P=1$, $(c_{10}, c_{11}) = (1/2, 1/2)$ (error $h^3 y^{i(3)}(x_n)/12 + O(h^4)$); $P = 2$, $(c_{10}, c_{11}, c_{12}) = (5/12, 8/12, -1/12)$ (error $-h^4 y^{i(4)}(x_n) + O(h^5)$), $(c_{20}, c_{21}, c_{22}) = (1/3, 4/3, 1/3)$ (error $h^5 y^{i(5)}(x_n) + O(h^6)$); $P = 3$, $(c_{10}, c_{11}, c_{12}, c_{13}) = (9/24, 19/24, -5/24, 1/24)$, $(c_{20}, c_{21}, c_{22}, c_{23}) = (1/3, 4/3, 1/3, 0)$, $(c_{30}, c_{31}, c_{32}, c_{33}) = (3/8, 9/8, 9/8, 3/8)$.

D. Runge-Kutta Methods

By a **general Runge-Kutta method** we mean a method of determining $y_0^i = \eta^i, y_1^i, y_2^i, \ldots$ successively by means of a formula $y_{n+1}^i - y_n^i = h\Phi^i(x_n, y_n^j; h)$, where the functions Φ^i are defined in terms of parameters α_r and β_{rs} as $\Phi^i(x, y^j; h) = \sum_{r=0}^{P} \alpha_r k_r^i$, $k_0^i = f^i(x, y^j)$, $k_r^i = f^i(x + \beta_{r0} h, y^j + h(\beta_{r0} - \sum_{s=1}^{P} \beta_{rs}) k_0^j + h \sum_{s=1}^{P} \beta_{rs} k_s^j)$ $(r = 1, \ldots, P)$. A Runge-Kutta method is called **explicit** if $\beta_{rs} = 0$ for $r \leqslant s$, and **implicit** otherwise. In the latter case, if $\beta_{rs} = 0$ for $r < s$, the method is called **semi-explicit** (or **semi-implicit**). Unless otherwise stated, the Runge-Kutta methods considered below are assumed to be explicit. In order to proceed one step from y_n^i to y_{n+1}^i we have to compute each function f^i $(P+1)$ times. Hence this is called the **$(P+1)$-stage method**. If we denote by $z^i(t)$ the solutions of $z^{i\prime}(t) = f^i(t, z^j(t))$ with the initial conditions $z^i(x) = y^i$ and put $h\Delta^i(x, y^j; h) = z^i(x+h) - z^i(x)$, then we have $\Phi^i(x, y^j; h) - \Delta^i(x, y^j; h) = h^p \varphi^i(x, y^j) + O(h^{p+1})$, where the $\varphi^i(x, y^j)$ are expressible in terms of the $f^i(x, y^j)$ and their (partial) derivatives. Various Runge-Kutta methods have been devised by searching for values of the α_r and β_{rs} that make p as large as possible for a given p (p is called the order of the method). When searching for these values of α_r and β_{rs}, we usually impose at least one of the following conditions: (1) α_r and β_{rs} are simple; (2) the truncation error is small; (3) the region of absolute stability (→ Section G) is large; (4) α_r and β_{rs} give smaller roundoff errors; (5) only a small computer memory is required. The ordered pairs (P, p), where p is the highest order that can be attained by the $(P+1)$-stage method, are $(P \leqslant 3, P+1)$, $(4,4)$, $(5,5)$, $(6,6)$, $(7,6)$, $(8,7)$, $(9\text{–}10, 8)$, $(p+2 \leqslant (P+1) \leqslant (1/2)(p^2 - 2p + 4), p \geqslant 9)$. The following formulas are frequently used since they are accurate and have simple parameters: for $P = 1$ and $p = 2$, the formulas with $\alpha_0 = 0$, $\alpha_1 = 1$, $\beta_{10} = 1/2$ (modified Euler method), with $\alpha_0 = \alpha_1 = 1/2$, $\beta_{10} = 1$ (improved Euler method) and with $\alpha_0 = 1/4$, $\alpha_1 = 3/4$, $\beta_{10} = 2/3$ (Ralston's second-order method); for $P = 2$ and $p = 3$, the formulas with $\alpha_0 = 2/9$, $\alpha_1 = 1/3$, $\alpha_2 = 4/9$, $\beta_{10} = 1/2$, $\beta_{20} = 3/4$, $\beta_{21} = 3/4$ (Ralston's third-order method), and with $\alpha_0 = 1/4$, $\alpha_1 = 0$, $\alpha_2 = 3/4$, $\beta_{10} = 1/3$, $\beta_{20} = 2/3$, $\beta_{21} = 2/3$ (Heun's third-order method). For $P = 3$, $p = 4$, the formula with $\alpha_0 = \alpha_3 = 1/6$, $\alpha_1 = \alpha_2 = 1/3$, $\beta_{10} = \beta_{20} = \beta_{21} = 1/2$, $\beta_{30} = \beta_{32} = 1$, $\beta_{31} = 0$ is well known and is frequently referred to as "the fourth-order Runge-Kutta method" or "the Runge-Kutta method." It has various desirable features [8]. Gill's modification of the classical Runge-Kutta method (sometimes called the **Runge-Kutta-Gill method**) has some advantage in regard to roundoff errors and to the necessary computer memory size, while Ralston's second-, third-, and fourth-order methods have minimum error bounds in Lotkin's sense for the local truncation error [7]. There exist "substantially fifth-order" methods for $P = 4$ [9]. Today we frequently prefer to use higher-order methods, especially fifth-order methods, instead of lower-order methods, because the former require less computation time for solutions of given accuracy [10].

For the estimation of the local truncation error, the following two practical techniques are well known: (i) **one-step–two-half-steps error estimate**. If we let y_{n+1}^{i2} and y_{n+1}^i denote the approximations that are computed with an mth-order method by taking two half-steps and one full step $(= h_n)$, respectively, then an estimate of the local truncation error per half-step associated with y_{n+1}^{i2} is given by the

formula

$$t_{n+1}^i \approx \gamma_m^i (h_n/2)^{m+1} \approx (y_{n+1}^{i2} - y_{n+1}^i)/2(1-2^m),$$

where we assume that γ_m^i does not vary much over the interval $[x_n, x_{n+1}]$. (ii) **The formula of embedding form**. The following formula of embedding form is more effective than the traditional method above. By use of the pth-order method $y_{n+1}^i = y_n^i + \sum_{r=0}^{P} \alpha_r k_r^i$ and the $(p+1)$th-order method $y_{n+1}^{i*} = y_n^i + \sum_{r=0}^{P} \alpha_r^* k_r^i$, we obtain an estimate of the local truncation error in y_{n+1}^i as $t_{n+1}^i \approx y_{n+1}^i - y_{n+1}^{i*}$ [11–13]. At present, formulas of this type are known for $p = 2$–8, where the method for $p=5$ is said to be the most efficient. Recently, a similar method for the estimation of the global truncation error has been investigated [14]. In addition, multistep generalizations of explicit Runge-Kutta methods, which are known as **pseudo-Runge-Kutta methods**, have also been investigated [15].

E. Multistep Methods

A **linear multistep method** approximates given differential equations by difference equations $\rho(E) y_n^i = h\sigma(E) f_n^i$, where E is the operator of increasing n by 1, $\rho(\zeta) = \alpha_k \zeta^k + \alpha_{k-1}\zeta^{k-1} + \ldots + \alpha_1 \zeta + \alpha_0$, $\sigma(\zeta) = \beta_k \zeta^k + \beta_{k-1}\zeta^{k-1} + \ldots + \beta_1 \zeta + \beta_0$, $\alpha_k \neq 0$, and $|\alpha_0| + |\beta_0| \neq 0$. (If ρ and σ are of degree k in ζ, we speak of a **linear k-step method**.) By means of these difference equations we can determine y_{n+k}^i from y_{n+k-1}^i, $y_{n+k-2}^i, \ldots, y_n^i$. Since the y_{n+k}^i are determined explicitly if $\beta_k = 0$ and implicitly if $\beta_k \neq 0$, the difference equations are called **explicit** and **implicit**, respectively. In the implicit case, if $|h| < |(\alpha_k/\beta_k)L|$, where L is a †Lipschitz constant for $f^i(x, y^j)$, then the y_{n+k}^i can be calculated by successive substitutions. In order to obtain the y_n^i successively by means of a k-step method with $k \geqslant 0$, it is necessary, to begin with, to give the $k-1$ sets of m values $y_1^i, \ldots, y_{k-1}^i$ (called the **starting values**) in addition to the initial values $y_0^i = \eta^i$. To determine the starting values, Runge-Kutta methods, overall approximation formulas, etc., are ordinarily used. The number of times we have to compute the f^i to proceed one step from y_n^i to y_{n+1}^i is only 1 for an explicit case, while for an implicit case it is equal to the number of iterations required to secure the convergence of the successive substitutions (ordinarily, the step size h as well as the 0th approximation for y_{n+k}^i are chosen so that the convergence is attained after a few iterations). We have $\lim_{h \to 0, x_n = x} e_n^i = 0$ $(x \in [a, b])$ for any f^i and η^i and any starting values such that $\lim_{h \to 0} y_\mu^i = \eta^i$ $(\mu = 0, 1, \ldots, k-1)$, if and only if the polynomials ρ and σ satisfy the following two conditions: (i) **consistency**: $\rho(1) = 0$, $\rho'(1) = \sigma(1)$; (ii) zero **stability**: Every root of $\rho(\zeta) = 0$ lies on or inside the unit circle, and every root lying on the unit circle is simple. We call $t_n^i = (\rho(E) y^i(x_n) - h\sigma(E) y^{i\prime}(x_n))/\sigma(1)$ the **local truncation error**. For a consistent and stable method determined by (ρ, σ) there exist a real constant C_{p+1} $(\neq 0)$ and an integer p $(\geqslant 1)$ such that $t_n^i = h^{p+1} C_{p+1} y^{i(p+1)}(x_n)/\sigma(1) + O(h^{p+2})$; p is called the **order** of the method and $C \equiv C_{p+1}/\sigma(1)$ the **error constant**. For a given ρ satisfying (ii), σ is chosen to make the order as small as possible and then to make the error constant as small as possible. However, p cannot exceed $k+1$ if k is odd or $k+2$ if k is even; moreover, p can be equal to $k+2$ only if all the roots of $\rho(\zeta) = 0$ lie on the unit circle.

The following are examples of multistep methods.

(1) Explicit Methods.
(a) **Adams-Bashforth methods**.

$k=1$, $\rho(\zeta) = \zeta - 1$, $\sigma(\zeta) = 1$, $p=1$,
$C = 1/2$ **(Euler method)**;

$k=2$, $\rho(\zeta) = \zeta(\zeta - 1)$, $\sigma(\zeta) = (3\zeta - 1)/2$,
$p=2$, $C = 5/12$;

$k=3$, $\rho(\zeta) = \zeta^2(\zeta - 1)$,
$\sigma(\zeta) = (23\zeta^2 - 16\zeta + 5)/12$, $p=3$,
$C = 3/8$;

etc.

(b) **Midpoint rule**.

$k=2$, $\rho(\zeta) = \zeta^2 - 1$, $\sigma(\zeta) = 2\zeta$,
$p=2$, $C = 1/6$.

(c) **Milne's predictor**.

$k=4$, $\rho(\zeta) = \zeta^4 - 1$,
$\sigma(\zeta) = (8\zeta^3 - 4\zeta^2 + 8\zeta)/3$,
$p=4$, $C = 7/90$.

(2) Implicit Methods.
(a) **Adams-Moulton methods**.

$k=1$, $\rho(\zeta) = \zeta - 1$, $\sigma(\zeta) = (\zeta + 1)/2$,
$p=2$, $C = -1/12$ **(trapezoidal rule)**;

$k=2$, $\rho(\zeta) = \zeta(\zeta - 1)$,
$\sigma(\zeta) = (5\zeta^2 + 8\zeta - 1)/12$,
$p=3$, $C = -1/24$;

etc.

(b) **Milne's corrector** (or the **Milne-Simpson formula**).

$k=2$, $\rho(\zeta) = \zeta^2 - 1$, $\sigma(\zeta) = (\zeta^2 + 4\zeta + 1)/3$,
$p=4$, $C = -1/90$.

When an implicit formula (ρ, σ) is used to obtain y_{n+k}^i, the 0th approximation y_{n+k}^{i*} for y_{n+k}^i is usually determined by an explicit formula (ρ^*, σ^*) of the same order as (ρ, σ) (where (ρ^*, σ^*) itself may not necessarily be stable). This kind of combination of implicit and explicit methods is called a **predictor-corrector method** (or **PC method**), where the formula (ρ^*, σ^*) is called the **predictor** and (ρ, σ) the **corrector**. Typical combinations are the midpoint rule and the trapezoidal rule, Milne's predictor and Milne's corrector (this combination is called **Milne's method**), and an Adams-Bashforth method and an Adams-Moulton method of the same order. A predictor-corrector method has the advantage that the local truncation error (of the corrector) can be estimated in the course of calculation without extra computation. In fact, if the orders of the predictor (ρ^*, σ^*) and the corrector (ρ, σ) are equal to p and their error constants to C^* and C, respectively, then the local truncation error t_n^i can be estimated by $t_n^i = KD_n^i + O(h^{p+2})$ $(n = 0, 1, \ldots)$, where $D_n^i \equiv y_{n+k}^{i*} - y_{n+k}^i$ is the difference between the value y_{n+k}^{i*} at x_{n+k} obtained by the predictor and the y_{n+k}^i obtained by the corrector, and $K = \alpha_k^* C/[(C - C^*)\sigma^*(1)]$ [16].

The term **multistep methods** originates from the fact that these use the values of the dependent variables at more than two different mesh points in order to proceed one step. These are also called **multivalue methods** since they use more than one value of the dependent variable. The multivalue method is, however, a more general concept than the multistep method.

Linear multistep methods are not only examples of the PC method; they are also examples of the **variable-step variable-order algorithms (VSVO algorithms)**, where the order of the formula as well as the step size are automatically chosen according to the behavior of the solution. In practical VSVO algorithms, the Adams-Bashforth-Moulton family of PC pairs of order 1 to 13 are usually used, and for solving stiff systems (→ Section G) those correctors with large regions of absolute stability are used. In these algorithms we use the multivalue method, which saves the information at different steps in a form convenient for the change of order and of step size. The multivalue method also facilitates error estimation and stability, and results in an efficient use of memory and reduction of the computational cost. For example, in Gear's algorithm the information required for computing y_{n+1} is saved in the following form: $\mathbf{y}_n = (y_n, hy_n', \ldots, (h^k/k!)y_n^{(k)})$, where $y'(x)$ is the polynomial $P_{k,n}(x)$ interpolating $f_n, f_{n-1}, \ldots, f_{n-k+1}$, where $y_n^{(i)} = d^{i-1}P_{k,n}(x)/dx^{i-1}|_{x=x_n}$ hold. All the local truncation error estimators are based essentially on Milne's device. In the VSVO algorithms heuristics play an important role [2, 17, 18].

F. Extrapolation Methods

Let us assume that the numerical solution with the step size h at some fixed point x has the form

$$y(x,h) = y(x) + \sum_{i=1}^{m} \tau_i h^{ri} + O(h^{r(m+1)}),$$

where r is a positive integer. Then we can approximate $y(x,h)$ by a function $R_m(x,h)$ with $(m+1)$ unknowns determined by the requirement that $R_m(x,h_j) = y(x,h_j)$, $j = 0, 1, 2, \ldots, m$, $h_i > h_j$ $(i < j)$ in order to approximate $y(x)$ by $R_m(x,0)$. If $R_m(x,h)$ is a polynomial of degree mr in h, we call the foregoing method a **polynomial extrapolation method**, where $R_m(x,h) = y(x) + o(h^{r(m+1)})$ as h tends to 0. Defining $R_0^j = y(x,h_j)$ and $R_m(x,0) = R_m^i$ for $j = i, i+1, \ldots, i+m$, we can calculate the R_m^i by the recursion relation $R_m^i = R_{m-1}^{i+1} + (R_{m-1}^{i+1} - R_{m-1}^i)/((h_i/h_{i+m})^r - 1)$, $m \geqslant 1$. It is well known that, for the following method (Gragg's method), there exists an asymptotic expansion with $r = 2$ [19]: $x_n = nh$, $\eta(0,h) = y(0)$, $\eta(x_1,h) = y_0 + hf(y_0, 0)$, $\eta(x_{n+1},h) = \eta(x_{n-1},h) + 2hf(\eta(x_n,h), x_n)$, $n = 1, 2, \ldots, N-1$, where $x_N = x$, $y(x,h) = 1/2[\eta(x_{N-1},h) + \eta(x_N,h) + hf(\eta(x_N,h), x_N)]$. If $R_m(x,h)$ is a rational function

$$R_m(x,h_j) = (p_0 + p_1 h + p_2 h^2 + \ldots + p_\mu h^\mu) / (q_0 + q_1 h + q_2 h^2 + \ldots + q_\nu h^\nu),$$

where $\mu = [m/2]$, $\nu = m - \mu = [(m+1)/2]$, $j = i, i+1, \ldots, i+m$, we call this method a **rational extrapolation method**. For $r = 2$, the following formulas were derived by Bulirsch and Stoer [20]:

$$R_m^i = R_{m-1}^{i+1} + (R_{m-1}^{i+1} - R_{m-1}^i)/((h_i/h_{i+m})^2 \times [1 - (R_{m-1}^{i+1} - R_{m-1}^i)/(R_{m-1}^{i+1} - R_{m-2}^{i+1})] - 1),$$

$$m \geqslant 1, \; R_{-1}^i = 0, \quad R_0^i = y(t, h_i).$$

G. Stability

Consider the application of the general k-step method at the nth point (which is consistent and 0-stable)

$$\sum_{j=0}^{k} \alpha_j y_{n+j} = h \sum_{j=0}^{k} \beta_j f_{n+j}$$

to $y' = f(x,y)$, $y(x_0) = y_0$. Let $\tilde{y}_n$ be the numerical solution at $x = x_n$. Let $\tilde{e}_n = y(x_n) - \tilde{y}_n$ be the global error, and let φ_n be the total error at the nth application. Then we find

$$\sum_{j=0}^{k} \alpha_j \tilde{e}_{n+j} = h \sum_{j=0}^{k} \frac{\partial f}{\partial y}(x_{n+j}, \xi_{n+j}) \beta_j \tilde{e}_{n+j} + \varphi_{n+k},$$

where ξ_{n+j} lies in the open interval whose endpoints are y_{n+j} and $y(x_{n+j})$. If we make two assumptions, $\partial f/\partial y=\lambda$ (const), $\varphi_n=\varphi$ (const), the above equation reduces to $\sum_{j=0}^{k}(\alpha_j-h\lambda\beta_j)\tilde{e}_{n+j}=\varphi$, whose general solution is given by

$$\tilde{e}_n=\sum_{s=0}^{k} d_s r_s^n-\varphi/h\lambda\sum_{j=0}^{k}\beta_j,$$

where the d_s are arbitrary constants and the r_s are the roots, assumed distinct, of the polynomial equation

$$\sum_{j=0}^{k}(\alpha_j-h\lambda\beta_j)r^j=0.$$

We call the linear k-step method **absolutely stable** for a given $h\lambda$ if $|r_s|<1$, $s=1,2,\ldots,k$. On the other hand, we call the linear k-step method **relatively stable** for a given $h\lambda$ if $|r_s|<|r_1|$ (or $|r_s|<e^{\lambda h}$), $s=2,3,\ldots,k$, where r_1 is the root corresponding to the theoretical solution. We call the region $S_a=\{h\lambda\,|\,|r_s|<1, s=1,2,\ldots,k\}$ (or $S_r=\{h\lambda\,|\,|r_s|<e^{\lambda h}$ (or $|r_1|$), $s=2,3,\ldots,k\}$) in the complex plane the **region of absolute stability** (or the **region of relative stability**) of the linear k-step method. Also we call $S_a\cap R$ (or $S_r\cap R$) the **interval of absolute stability** (or the **interval of relative stability**) of the linear multistep method, where R is the real line. The explicit methods and the PC methods have a finite interval of absolute stability. The implicit methods usually have larger intervals of absolute stability than the corresponding explicit methods. The higher-order PC methods have smaller intervals, while Runge-Kutta methods do not. The PECE methods usually have larger intervals of stability than the corresponding PEC or $\mathrm{P(EC)}^2$ methods, where P indicates an application of predictor, C an application of corrector, and E an evaluation of f [21, 22].

Let us consider the linear system $\mathbf{y}'=A\mathbf{y}+\varphi(x)$, where A is an $m\times m$ constant matrix and $\mathbf{y}(x)$, $\mathbf{y}'(x)$, $\varphi(x)\in\mathbf{R}^m$. If A possesses m distinct eigenvalues $\lambda_t=\mu_t+i\nu_t$, $t=1,2,\ldots,m$, the theoretical solution of this system is given by

$$\mathbf{y}(x)=\sum_{t=1}^{m}K_t\exp((\mu_t+i\nu_t)x)C_t+\psi(x),$$

where K_t and C_t, $t=1,2,\ldots,m$, are, respectively, arbitrary constants and the eigenvectors corresponding to λ_t and $\psi(x)$ is a particular solution. We call the linear system **stiff** if (i) $\mu_t<0$, $t=1,2,\ldots,m$, (ii) $s\gg 1$, where $s=\max_{1\leqslant t\leqslant m}|\mu_t|/\min_{1\leqslant t\leqslant m}|\mu_t|$. We call the nonlinear system $\mathbf{y}'=f(x,\mathbf{y})$, $\mathbf{y}(x)$, $\mathbf{y}''(x)$, $f(x,\mathbf{y})\in\mathbf{R}^m$ **stiff in an interval** I **of** x if, for every $x\in I$, the eigenvalues $\lambda_t(x)$ of the †Jacobian matrix of f satisfy (i) and (ii). The ratio s is called the **stiffness ratio**. To solve stiff systems effectively a variety of methods with infinite stability regions have been proposed. Some of them are:

(1) A-**stability** (Dahlquist): $S_a\supseteq\{h\lambda\,|\,\mathrm{Re}(h\lambda)<0\}$
(2) **Stiff-stability** (Gear): $S_a\supseteq S_1\cup S_2$, $S_1=\{h\lambda\,|\,\mathrm{Re}(h\lambda)<-a<0\}$, $S_2=\{h\lambda\,|\,-a\leqslant\mathrm{Re}(h\lambda)\leqslant b,\ -c\leqslant\mathrm{Im}(h\lambda)\leqslant c, b\geqslant 0, c>0\}$
(3) $A(\alpha)$-**stability** (Widlund): $S_a\supseteq\{h\lambda\,|\,-\alpha<\pi-\arg(h\lambda)<\alpha, \alpha\in(0,\pi/2)\}$
(4) A_0-**stability** (Cryer): $S_a\supseteq\{h\lambda\,|\,\mathrm{Re}(h\lambda)<0, \mathrm{Im}(h\lambda)=0\}$.

The order p and the step number k of linear multistep methods are restricted by the following stability requirements:
(i) A-stability: implicit, $p\leqslant 2$; trapezoidal rule is the most accurate method.
(ii) Stiff-stability: implicit, $p\leqslant k$; backward differentiation formulas $\sum_{j=0}^{k}\alpha_j y_{n+j}=hf_{n+k}$ are stiffly stable for $p=k=1,2,\ldots,6$, 0-unstable for $k=6$.
(iii) $A(\alpha)$-stability: implicit; there exist high-order $A(0)$-stable linear multistep methods. (A method is said to be $A(0)$-**stable** if it is $A(\alpha)$-stable for some sufficiently small $\alpha\in(0,\pi/2)$.)

For the Runge-Kutta (P,p) methods, we can write

$$\tilde{e}_{n+1}=\left\{1+h\lambda+(h\lambda)^2/2!+\ldots+(h\lambda)^p/p!+\sum_{q=p+1}^{P+1}\gamma_q(h\lambda)^q\right\}\tilde{e}_n+\varphi_{n+1},$$

where $\partial f/\partial y=\lambda$ (const) and γ_q are functions of the coefficients of the method in use. We call a region $R=\{h\lambda\,|\,|1+h\lambda+(h\lambda)^2/2!+\ldots+(h\lambda)^p/p!+\sum_{q=p+1}^{P+1}\gamma_q(h\lambda)^q|<1\}$ a **region of absolute stability of the Runge-Kutta** (P,p) **method**. The implicit Runge-Kutta methods have a larger region of absolute stability than the corresponding explicit Runge-Kutta methods. Therefore the implicit Runge-Kutta methods are suitable for stiff equations, and the explicit Runge-Kutta methods are suitable for nonstiff or mildly stiff equations.

Recently, Yamaguti and others have pointed out that the instabilities occurring in the numerical solution of ordinary differential equations are closely connected to the phenomena of **chaos** studied by Li, Yorke, and others [23, 24].

H. Boundary Value Problems

A boundary value problem is generally formulated as the problem of obtaining functions $y^i(x)$ that satisfy the differential equations $y^{i\prime}(x)=f^i(x,y^1(x),\ldots,y^m(x))$ and boundary conditions $B_k(y^1(x_{n_1}),\ldots,y^m(x_{n_1}); y^1(x_{n_2}),\ldots,y^m(x_{n_2});\ldots;y^1(x_{n_s}),\ldots,y^m(x_{n_s}))=0$ $(i,k=1,\ldots,m)$, where the f^i and B_k are given functions. If the f^i and B_k are linear in the y^i, we can

first calculate the solutions $y_0^i(x)$ of the differential equations under an appropriate initial condition, as well as a set of m independent solutions $y_l^i(x)$ $(l=1, \ldots, m)$ of the homogeneous differential equations (say, under the initial conditions $y_l^i(a)=\delta_l^i$ at some point $x=a$), and then substitute the expression for the desired solution $y^i(x)=y_0^i(x)+\sum_{l=1}^m \alpha_l y_l^i(x)$ in the boundary conditions to obtain simultaneous linear equations for the unknowns α_l (→ 302 Numerical Solution of Linear Equations). No universally powerful method is known for the case where the f^i or the B_k or both are nonlinear. Ordinarily, we resort to the trial-and-error method of solving the differential equations iteratively under different initial conditions until the solution fully satisfies the boundary conditions, or to approximating the differential operators by suitable difference operators to obtain a set of (generally nonlinear) simultaneous equations approximating at the same time the differential equations and the boundary conditions. For related topics → 298 Numerical Computation of Eigenvalues; 315 Ordinary Differential Equations (Boundary Value Problems); [25–27].

I. Methods Other than Difference Methods

Besides difference methods there are frequently used methods for finding in a given finite-dimensional function space a function that best satisfies the differential equations as well as the initial or boundary conditions. Denote the equations by $L[y(x)]=0$ and assume the conditions to be linear. We choose a function $y_0(x)$ that satisfies the conditions and that is considered to approximate the exact solution, and also functions $y_l(x)$ $(l=1, \ldots, q)$ that are considered to represent the typical deviations of $y_0(x)$ from the exact solution and each of which satisfies the homogeneous conditions. We then determine the α_l in such a way that $y(x)=y_0(x)+\sum_{l=1}^q \alpha_l y_l(x)$ best satisfies the differential equations. Corresponding to different interpretations of the words "best satisfy" there are different methods. The **collocation method** determines the α_l so as to nullify the values of $L[y(x)]$ at some prescribed points x_i; the **method of least squares** minimizes the integral of $|L[y(x)]|^2$ over the considered region; the **Galerkin method** makes $L[y(x)]$ orthogonal to every $y_l(x)$ $(l=1, \ldots, q)$; and the **Ritz method** considers the variational problem $\delta J[y(x)]=0$ whose †Euler equation is $L[y(x)]=0$ (→ 46 Calculus of Variations F), if such a problem exists, and determines the α_l so as to have $J[y(x)]$ take an extremum value with $y(x)=y_0(x)+\sum_{l=1}^q \alpha_l y_l(x)$.

References

[1] J. D. Lambert, Computational methods in ordinary differential equations, Wiley, 1973.
[2] C. W. Gear, Numerical initial value problems in ordinary differential equations, Prentice-Hall, 1971.
[3] P. Henrici, Discrete variable methods in ordinary differential equations, Wiley, 1962.
[4] G. Hall and J. M. Watt (eds.), Modern numerical methods for ordinary differential equations, Clarendon Press, 1976.
[5] H. J. Stetter, Analysis of discretization methods for ordinary differential equations, Springer, 1970.
[6] L. Lapidus and J. H. Seinfeld, Numerical solution of ordinary differential equations, Academic Press, 1971.
[7] A. Ralston and P. Rabinowitz, A first course in numerical analysis, McGraw-Hill, second edition, 1978.
[8] M. Tanaka, T. Arakawa, and S. Yamashita, On the characteristics of Runge-Kutta methods, Information Processing in Japan, 17 (1977), 175–181.
[9] E. B. Shanks, Solutions of differential equations by evaluations of functions, Math. Comp., 20 (1966), 21–38.
[10] K. R. Jackson, W. H. Enright, and T. E. Hull, A theoretical criterion for comparing Runge-Kutta formulas, SIAM J. Numer. Anal., 15 (1978), 618–641.
[11] M. Tanaka, On Kutta-Merson process and its allied processes, Information Processing in Japan, 8 (1968), 44–52.
[12] J. H. Verner, Explicit Runge-Kutta methods with estimates of the local truncation error, SIAM J. Numer. Anal., 15 (1978), 772–790.
[13] H. Shintani, Two step processes by one step methods of order 3 and order 4, J. Sci. Hiroshima Univ., ser. A-I, 30 (1966), 183–195.
[14] P. Merluzzi and C. Brosilow, Runge-Kutta integration algorithms with built-in estimates of the accumulated truncation error, Computing, 20 (1978), 1–16.
[15] G. D. Byrne and R. J. Lambert, Pseudo-Runge-Kutta methods involving two points, J. Assoc. Comput. Math. 13 (1966), 114–123.
[16] H. J. Stetter, Global error estimation in Adams PC codes, ACM Trans. Math. Software, 5 (1979), 415–430.
[17] L. F. Shampine and M. K. Gordon, Computer solution of ordinary differential equations, the initial value problem, Freeman, 1975.
[18] F. T. Krogh, A variable-step variable-order multistep method for the numerical solution of ordinary differential equations, Proc. Information Processing 68, North-Holland, 1969, vol. 1.

[19] W. B. Gragg, On extrapolation algorithms for ordinary initial value problems, SIAM J. Numer. Anal., 2 (1965), 384–403.
[20] R. Bulirsch and J. Stoer, Numerical treatment of ordinary differential equations by extrapolation methods, Numer. Math., 8 (1966).
[21] G. Dahlquist, Stability questions for some numerical methods for ordinary differential equations, Amer. Math. Soc. Proc. Symp. in Applied Math. 15, 1963.
[22] M. Iri, A stabilizing device for unstable numerical solutions of ordinary differential equations—Design principle and applications of a "filter," Information Processing in Japan, 4 (1964), 65–73.
[23] T. Y. Li and J. A. Yorke, Period three implies chaos, Amer. Math. Monthly, 82 (1975), 985–992.
[24] M. Yamaguti and H. Matano, Euler's finite difference scheme and chaos, Proc. Japan Acad., ser. A, 55 (1979), 78–80.
[25] A. K. Aziz (ed.), Numerical solutions of boundary-value problem, Academic Press, 1975.
[26] H. B. Keller, Numerical methods for two-point boundary-value problems, Blaisdell, 1968.
[27] H. B. Keller, Numerical solutions of two-point boundary-value problems, SIAM Regional Conf. Ser. in Appl. Math., 1976.

304 (XV.9) Numerical Solution of Partial Differential Equations

A. General Remarks

The numerical solution of partial differential equations became practical in the 1950s with the advent of automatic digital computers. Nowadays, by means of modern high-performance computers, the numerical solution of partial differential equations is carried out extensively and often on a very large scale for problems in physics, engineering, and other fields of applied analysis, in order to obtain approximate solutions of rigorous equations or to simulate real phenomena by means of numerical experiments. Various numerical-solution schemes have been proposed, applied, and studied, and programming techniques required to implement numerical solutions are being invented regularly; some are of a general nature, and others are devised for particular problems. The variational method, which includes the Ritz-Galerkin method and the finite element method, and the difference method deserve special mention here because of their generality and scope of application.

B. The Ritz-Galerkin Method

Having been applied mainly to boundary value problems for elliptic partial differential equations, the **Ritz method** is a classical method, which has been used since the old era of mechanical calculators, and is mathematically nothing other than the †direct method in the calculus of variations (→ 46 Calculus of Variations F) applied to †variational problems that are equivalent to the original boundary value problems. In order to exemplify the method, let Ω be a bounded domain in $\mathbf{R}^N$ with piecewise smooth boundary S, and consider the Dirichlet boundary value problem comprised of the Poisson equation

$$\Delta u = -f \tag{1}$$

and the boundary condition

$$u|_s = \beta. \tag{2}$$

Here f and β are given functions on Ω and S, respectively. (In the following, given functions are assumed to be sufficiently smooth.) The boundary value problem (1), (2) is equivalent to the variational problem of minimizing the †functional

$$J[u] = \frac{1}{2}a(u,u) - (u,f) \tag{3}$$

within the set $D(J)$ of **admissible functions** u subject to (2). In (3) the following notation is used:

$$a(u,v) = (\nabla u, \nabla v)_{L_2(\Omega)} = \int_\Omega \nabla u \cdot \nabla v \, dx$$

and

$$(u,v) = (u,v)_{L_2(\Omega)} = \int_\Omega uv \, dx,$$

while

$$D(J) = \{u \in L_2(\Omega) \mid \nabla u \in L_2(\Omega), \quad u|_s = \beta\}.$$

This variational problem is further reduced to the condition:

$$u \in D(J),$$

$$a(u,\varphi) = (f,\varphi) \qquad (\forall \varphi \in V), \tag{4}$$

where V stands for the †Sobolev space $H_0^1(\Omega)$ (→ 168 Function Spaces B), that is,

$$V = H_0^1(\Omega) = \{u \in L_2(\Omega) \mid \nabla u \in L_2(\Omega), \ u|_s = 0\}.$$

Note that V is equal to $D(J)$ with $\beta \equiv 0$. The condition (4) is often called a **weak form of the boundary value problem**. In applying the Ritz

method to this problem, we introduce unknown parameters $\alpha_1, \alpha_2, \ldots, \alpha_n$ and set the approximate solution u_n in the form

$$u_n = F(x; \alpha_1, \alpha_2, \ldots, \alpha_n) \in D(J)$$

and minimize $J[u_n]$ as a function of the n variables $\alpha_1, \alpha_2, \ldots, \alpha_n$. A standard way to form u_n is to choose first a $b \in D(J)$ subject to the inhomogeneous boundary condition and $\varphi_1, \varphi_2, \ldots, \varphi_n \in V$ subject to homogeneous boundary condition, and then to set

$$u_n = b + \alpha_1 \varphi_1 + \alpha_2 \varphi_2 + \ldots + \alpha_n \varphi_n. \tag{5}$$

The functions φ_j ($j = 1, 2, \ldots, n$) are called **basis functions** or **coordinate functions** in the Ritz method. Denoting by D_n the set of all possible functions appearing on the right-hand side of (5) and by V_n the linear space generated by $\varphi_1, \varphi_2, \ldots, \varphi_n$, we can state the condition to determine the approximate solution u_n as:

$$u_n \in D_n$$

$$a(u_n, \varphi) = (f, \varphi) \qquad (\forall \varphi \in V_n). \tag{6}$$

Thus the Ritz method is a kind of **projective approximation method** which approximates the weak equation (4) by its projection on a finite-dimensional space V_n. The condition (6) is equivalent to the equations

$$a(u_n, \varphi_j) = (f, \varphi_j) \qquad (j = 1, 2, \ldots, n). \tag{7}$$

Particularly when $\beta \equiv 0$ in (2), D_n coincides with V_n, and the equations (7) that determine the coefficients α_j are reduced to a linear equation

$$K\alpha = \gamma \tag{8}$$

for $\alpha = {}^t(\alpha_1, \alpha_2, \ldots, \alpha_n)$, where the matrix $K = (K_{ij})$ and the vector $\gamma = {}^t(\gamma_1, \gamma_2, \ldots, \gamma_n)$ are given respectively by $K_{ij} = a(\varphi_i, \varphi_j)$ and $\gamma_j = (f, \varphi_j)$. Solving (8) numerically by the †Gauss elimination method or by †iteration, one eventually obtains the approximate solution u_n by the Ritz method. The Ritz method is also applicable, for instance, to the eigenvalue problem $\Delta u = \lambda u$, $u|_s = 0$, since this eigenvalue problem is equivalent to the variational problem of finding a stationary **Rayleigh's quotient** $R[u] = a(u, u)/(u, u)$ with V as the set of admissible functions. In the approximation to restrict the admissible functions to V_n, the approximate eigenvalue $\lambda^{(n)}$ is determined through the matrix eigenvalue problem

$$K\alpha = \lambda^{(n)} M\alpha, \tag{9}$$

where the matrix K is as above and the matrix $M = (M_{ij})$ is given by $M_{ij} = (\varphi_i, \varphi_j)$. Sometimes, optimization of $J[u_n]$ or $R[u_n]$ is carried out more directly, e.g., by the †gradient method, particularly when u_n contains free parameters α_j in some nonlinear way.

Many studies have been made of convergence and error estimation for the Ritz method in general (→ e.g., [1–4]). For instance, it is known that, if $\beta = 0$, $f \in L^2(\Omega)$, and $\{\varphi_1, \varphi_2, \ldots, \varphi_n, \ldots\}$ spans a linear subspace dense in $V = H_0^1(\Omega)$, then the approximate solution u_n obtained through (6) (or through (12) below) converges to the rigorous solution in the H^1-topology (and hence in the L^2-topology). Nevertheless, success in using the method in practical applications can be gained only by a clever choice of trial functions. A typical but systematic example of such a choice is made by the finite element method (→ Section C).

We proceed to the **Galerkin method** [1]. Suppose that we are to solve the equation

$$L[u] = \Delta u + (b \cdot \nabla) u = -f, \tag{10}$$

which is the equation in (1) perturbed by a lower-order term (the convection term). Here $b = (b_1(x), \ldots, b_n(x))$ and $(b \cdot \nabla) u = \sum_j b_j(x)(\partial u / \partial x_j)$. For simplicity, the boundary condition (2) is assumed to be homogeneous (i.e., $\beta \equiv 0$). Although the boundary value problem (10), (2) cannot be reduced to a variational problem since (10) is not symmetric, it is equivalent to its weaker form

$$a(u, \varphi) - ((b \cdot \nabla) u, \varphi) = (f, \varphi) \qquad (\forall \varphi \in V), \tag{11}$$

provided that $u \in V$. Note that (11) is obtained from $(L[u] + f, \varphi)_{L_2} = 0$ by transforming $(L[u], \varphi)$ through integration by parts. Actually, this boundary value problem has a unique solution u for any $f \in L_2(\Omega)$ if $|\operatorname{div} b|$ is sufficiently small. According to the Galerkin method as applied to the present problem, one replaces V by V_n in (11) and determines the approximate solution $u_n = \alpha_1 \varphi_1 + \ldots + \alpha_n \varphi_n \in V_n$ by the condition

$$a(u_n, \varphi) - ((b \cdot \nabla) u_n, \varphi) = (f, \varphi) \qquad (\forall \varphi \in V_n), \tag{12}$$

which is equivalent to the equations

$$a(u_n, \varphi_j) - ((b \cdot \nabla) u_n, \varphi_j) = (f, \varphi_j) \qquad (j = 1, 2, \ldots, n). \tag{13}$$

Sometimes it is convenient to project $(L[u] + f, \varphi) = 0$ onto the finite-dimensional space and to deal directly with $(L[u_n] + f, \varphi_j) = 0$ ($j = 1, 2, \ldots, n$). When the coefficients of the equation and the given function f are periodic in space variables and when sine or cosine functions are chosen as basis functions, the Galerkin method turns out to be the same as the so-called Fourier approximation method and can be efficiently implemented with the aid of a †fast Fourier transform (FFT) to yield the required Fourier coefficients numerically [5].

The Galerkin method can be applied to the numerical solution of evolution equations, namely, in order to approximate time-

dependent solutions. This is exemplified through the following initial-boundary value problem for the diffusion equation (heat equation):

$$\frac{\partial u}{\partial t}=\Delta u \quad (t\geqslant 0,\ x\in\Omega) \tag{14}$$

with the homogeneous boundary condition

$$u|_s=0 \tag{15}$$

and the initial condition

$$u|_{t=0}=a(x) \quad (x\in\Omega). \tag{16}$$

We first set the required approximate solution $u_n=u_n(t,x)$ in terms of the basis functions $\varphi_1,\varphi_2,\ldots,\varphi_n$ of V_n as

$$u_n=\alpha_1(t)\varphi_1+\alpha_2(t)\varphi_2+\ldots+\alpha_n(t)\varphi_n, \tag{17}$$

and then determine the coefficients $\alpha_1(t),\alpha_2(t),\ldots,\alpha_n(t)$ from the conditions

$$\frac{d}{dt}(u_n,\varphi)=-a(u_n,\varphi) \quad (t\geqslant 0,\ \forall\varphi\in V_n) \tag{18}$$

and

$$(u_n(0),\varphi)=(a,\varphi) \quad (\forall\varphi\in V_n). \tag{19}$$

Since (18) is equivalent to the equations

$$\frac{d}{dt}(u_n,\varphi_j)=-a(u_n,\varphi_j) \quad (j=1,2,\ldots,n),$$

the vector function $\alpha(t)={}^t(\alpha_1(t),\alpha_2(t),\ldots,\alpha_n(t))$ is obtained by solving the ordinary differential equation

$$M\frac{d}{dt}\alpha=-K\alpha, \tag{20}$$

where the matrices M and K are those in (8) and (9). The initial value $\alpha(0)$ of α should be given in accordance with (19). Sometimes the procedure above is called a **semidiscrete approximation** based on the Galerkin method. Obviously, this method can be applied to more general evolution equations, for instance, to the **diffusion-convection equation**

$$\frac{\partial u}{\partial t}=L[u]=\Delta u+(b\cdot\nabla)u, \tag{21}$$

which governs time-dependent states corresponding to stationary states subject to (10). As a concrete example, we can refer to the finite element approximation for evolution equations described in Section D. If the geometry of Ω is simple enough, e.g., for 1-dimensional Ω, one can adopt eigenfunctions of $L[\ ]$ as the basis functions; the resulting version of the semidiscrete Galerkin method is called a **spectral method**.

In order to obtain $u_n(t)$ one has to carry out numerical integration for (20), discretizing the time variable. The Galerkin approximation with discretization of the space and time variables is called a **full discrete approximation**. Finally, an advantage of the Ritz-Galerkin method is that if the boundary condition imposed on rigorous solutions is a natural boundary condition, then the admissible functions and particularly the basis functions need not satisfy it [4].

C. Finite Element Methods for Boundary Value Problems

In recent years the method of numerical solution of partial differential equations most extensively employed in structural mechanics and many other fields of engineering has been the **finite element method**. In its standard form, the finite element method can be regarded, at least mathematically, as a type of Ritz-Galerkin method that adopts as its basis functions piecewise polynomials of low degree and with narrow supports. Although the idea leading to the finite element method can be traced back to a paper by R. Courant in 1943 (*Bull. Amer. Math. Soc.*), the method acquired its popularity in the late 1950s when it was rediscovered by engineers on the basis of mechanical considerations [6]. Here we apply the method to the boundary value problem (1), (2), assuming that $\beta=0$ and Ω a polygonal domain in $\mathbf{R}^2$ [7–9]. First, $\bar\Omega=\Omega\cup S$ is divided into small triangles of which the length of any side does not exceed $h>0$. Each triangle T appearing in this decomposition of $\bar\Omega$ is called a **triangular element**, or simply an **element** (following the terminology in structural mechanics), and we denote the set of all triangular elements by $\mathcal{T}_h$, h being equal to $\max_{T\in\mathcal{T}_h}\{$the longest side of $T\}$. The set of all **nodal points**, i.e., the vertices of the triangular elements, is denoted by N, and we put

$$N_i=\{P\in N\mid P\in\Omega\}=\{P_1,P_2,\ldots,P_n\}. \tag{22}$$

Then a standard choice of admissible functions is to adopt as V_n in (6) the following $V_h\subset V=H_0^1(\Omega)$:

$$V_h=\{v_h\in C(\bar\Omega)\mid v_h \text{ is linear on each element and } v_h|_s=0\}, \tag{23}$$

namely, V_h is the set of all elementwise linear continuous functions satisfying the boundary condition. Now the approximate solution $u_h\in V_h$ is determined by the condition

$$a(u_h,\varphi_h)=(f,\varphi_h) \quad (\forall\varphi_h\in V_h). \tag{24}$$

A basis of V_h is formed by pyramidal functions $\varphi_j\ (j=1,2,\ldots,n)\in V_h$ such that

$$\varphi_j(P_j)=1 \quad\text{and}\quad \varphi_j(Q)=0 \quad (Q\in N\setminus\{P_j\}). \tag{25}$$

In order to obtain u_h, one has to solve (8) numerically for α. Since the support of φ_j is confined to elements adjacent to P_j, the matrix K turns out to be a sparse matrix of **multi-diagonal type**, and hence particular elimination procedures can be applied with high efficiency. In the finite element method, the matrix K is often called the **stiffness matrix**. The convergence of u_h to u as $h \to 0$ is guaranteed if $\mathscr{T}_h$ $(0 < h \leqslant h_0)$ satisfies a certain regularity condition. For instance, if any angle of a triangle T is not smaller than $\theta_0 > 0$, then u_h converges in the $H^1(\Omega)$-topology. Moreover, when Ω is convex, it holds that

$$\|u - u_h\|_{H^1} \leqslant Ch\|f\|_{L_2},$$

$$\|u - u_h\|_{L_2} \leqslant Ch^2\|f\|_{L_2} \qquad (26)$$

[7, 9]. Uniform convergence of u_h to u and L_∞-estimation of the error have also been established (e.g., [7, chap. 3]). u_h is subject to the maximum principle, provided that the triangulation $\mathscr{T}_h$ is of acute type. Here, $\mathscr{T}_h$ is said to be of **acute type** (or of **strongly acute type**) if any angle θ of $T \in \mathscr{T}_h$ is in $0 < \theta \leqslant (\pi/2)$ $(0 < \theta \leqslant \theta_1 < (\pi/2)$. To acquire higher accuracy, more sophisticated admissible functions that are piecewise polynomials of higher degree are used. To take account of curved boundaries, several devices, including the **isoparametric method**, have been proposed.

In dealing with biharmonic equations or other equations of higher order, one can conveniently apply a modified version of the finite element method of **mixed type** or of **nonconforming type**; in the latter the approximate solution is sought within a class of functions which are of less regularity at the interface of elements and hence do not belong to the domain of the original variational problem. i.e., which are not admissible in the classical sense [10, 11].

D. Finite Element Methods for Initial Value Problems

The most basic type of finite element method applicable to evolution problems is described here, and concerns the initial boundary value problem (14)–(16), assuming that Ω is the same as in Section B [8, 9]. In the semidiscrete finite element approximation one uses the functions φ_j of the preceding section as the basis for the approximate solution, now denoted by u_h, and then determines u_h through (17)–(19) with V_n replaced by V_h. Then u_h converges to u as $h \to 0$, provided that the triangulation $\mathscr{T}_h$ is regular. Furthermore, under the same condition that makes (26) hold, the error is given by the following estimate [12]:

$$\|u - u_h\|_{L_2} \leqslant Ch^2\|a\|_{L^2}/t \qquad (t > 0), \qquad (27)$$

which reflects the smoothing property of the diffusion equation. In the terminology of the finite element method, the matrix M in (20) is called the **mass matrix**. In order to gain a fully discrete finite element approximation $u_{\tau,h}$ for u, one discretizes the time variable with the †mesh length τ and solves the difference analog of (20), i.e., either

$$M(\alpha(t+\tau) - \alpha(t))/\tau = -K\alpha(t) \qquad (28)$$

or

$$M(\alpha(t+\tau) - \alpha(t))/\tau = -K\alpha(t+\tau), \qquad (29)$$

where t is restricted to $t = k\tau$ $(k = 0, 1, 2, \dots)$. The difference schemes (28) and (29) are of **forward type** and of **backward type**, respectively. From (29) follows $\|\alpha(k\tau)\| \leqslant \|\alpha(0)\|$, and consequently $\|u_{\tau,h}(t)\|_{L_2} \leqslant C\|a\|_{L_2}$. In the generally accepted terminology, the approximation is **stable** if for each $T > 0$ there exists a positive constant C_T depending on T such that

$$\|u_{\tau,h}(t)\| \leqslant C_T\|a\| \qquad (0 \leqslant t = k\tau \leqslant T). \qquad (30)$$

Thus the approximation through (29) is unconditionally stable, while the one through (28) is stable only under certain conditions on the triangulation $\mathscr{T}_h$ and on the time mesh τ. For instance, the forward scheme (28) is stable if $\mathscr{T}_h$ is regular and satisfies the **inverse assumption**

$$\sup_{0 < h \leqslant h_0} \max_{T \in \mathscr{T}_h} \frac{h}{(\text{longest side of } T)} < +\infty,$$

and if the ratio τ/h^2 is sufficiently small [13, 14].

Sometimes, the mass matrix M in (28) and (29) is modified by the following procedure, called **mass lumping**: For each P_j, one joins alternately the center of gravity of the elements with P_j as one of their vertices and the middle point of the sides with P_j as one of their endpoints, thus forming a closed broken line Γ_j surrounding P_j. We denote by $\bar{\varphi}_j$ the characteristic function of the polygonal domain B_j bounded by Γ_j. When mass lumping is applied, the matrix M is replaced by the matrix $\overline{M} = ((\bar{\varphi}_i, \bar{\varphi}_j))$. Generally, mass lumping relaxes the conditions for the stability and the maximum principle to hold, although it may lower the accuracy to some extent (H. Fujii, *Proc. US-Japan Seminar*, Univ. Tokyo Press, 1973). Usually, schemes without mass lumping are called **consistent mass schemes**.

In principle, there is no difficulty in applying the finite element method to the diffusion-convection equation (21). However, when $\|b\|$ is large, the stability condition becomes strin-

gent. To meet these difficulties, M. Tabata, T. Ikeda [15], and others have devised some schemes that enjoy better stability, the maximum principle, and even the law of conservation of mass by improving mass lumping and introducing the idea of an upstream approximation or an artificial viscosity for discretizing the convection term. Moreover, finite element methods are currently being applied to wave equations, the Navier-Stokes equations, and various other evolution equations.

E. Difference Methods for Boundary Value Problems

In difference methods for solving partial differential equations we reduce the original equation to its **difference analog**, replacing †differential quotients of the unknown function by the corresponding difference quotients. The difference analog, usually a system of algebraic equations, is then solved numerically, yielding the desired approximate solution. Since the last part of the method, i.e., the numerical solution of the difference analog, requires a large amount of computation, difference methods were not feasible until the development of automatic computers.

In terms of the variable x, difference approximations for the derivative df/dx are the **forward difference**

$$(D_x^h f)(x)=(f(x+h)-f(x))/h,$$

the **backward difference**

$$(D_{\bar{x}}^h f)(x)=(f(x)-f(x-h))/h,$$

and the **central difference (symmetric difference)**

$$(\delta_x^h f)(x)=(f(x+h)-f(x-h)/2h.$$

In order to exemplify the difference method applied to boundary value problems, we return to equations (1) and (2), supposing that Ω is a 2-dimensional square $Q=\{(x,y)\mid 0<x<1, 0<y<1\}$ [1, 16–18]. Then we cover $Q\cup S$ by a **square net** (lattice, grid) with mesh length $h=1/N$, for N a large integer. The set of net points $P_{m,n}=(mh,nh)$ lying in Q is denoted by Ω_h, and S_h is the set of net points on S. We want to have a net function u_h on $R_h=\Omega_h\cup S_h$ that approximates the solution u of (1) and (2), where u_h is determined by the difference equation

$$\Delta_h u_h(x,y)=-f(x,y)\qquad((x,y)\in\Omega_h),\tag{31}$$

and the boundary condition

$$u_h(x,y)=\beta(x,y)\qquad((x,y)\in S_h).\tag{32}$$

Here Δ_h is a difference operator, the 5-point difference analog of Δ, which is defined by

$$\begin{aligned}\Delta_h\varphi&=D_{\bar{x}}^h D_x^h\varphi+D_{\bar{y}}^h D_y^h\varphi\\&=(\varphi(x+h,y)+\varphi(x-h,y)+\varphi(x,y+h)\\&\quad+\varphi(x,y-h)-4\varphi(x,y))/h^2.\end{aligned}$$

In view of the maximum principle for net functions on R_h, it is seen that (31) and (32) admit only $u_h=0$ as a solution if $f=\beta=0$, which implies that the solution u_h of (31) and (32) exists uniquely for all f and β. When we are concerned with the convergence of u_h to u as $h\to 0$, we have to consider (31) and (32) for all small $h>0$. Such a family of difference equations is called a **difference scheme**. Actually, u_h obtained through (31) and (32) converges to u as $h\to 0$, and we can say that the difference scheme (31) and (32) is convergent. In general, the order of the error $u-u_h$ depends on the smoothness of u. For instance, if $u\in C^4(\bar{\Omega})$, then $|u-u_h|\leqslant Ch^2$ holds uniformly on R_h. When the asymptotic behavior of the error is known to satisfy $u-u_h=wh^2+o(h^2)$ for some $w=w(x,y)$, it is possible to attain a higher accuracy by eliminating the h^2-order term if we compute u_h for two different values of h and form an appropriate linear combination of the u_h thus obtained (Richardson's extrapolation; → e.g., [17]). The convergence of difference schemes applied to boundary value problems is not affected if rectangular nets are used instead of square nets. Many techniques have been devised to treat curved boundaries. There have been attempts recently to map Ω into a domain Q of simpler geometry, say, a rectangular one, and then discretize the arising partial differential equations with variable coefficients in Q (e.g., J. F. Thompson et al., *J. Comput. Phys.*, 1982.) Moreover, when information regarding the location and nature of singularities of solutions is available, one can refine the net near these points.

The difference analog of elliptic equations described above is a system of linear equations with the values of the u_h as its unknowns. Since the coefficient matrix of this system is of multidiagonal type, direct elimination methods can be efficiently applied to solve it. Particularly, some elimination algorithms employing vector computations have been invented [19]. Iteration methods of the Gauss-Seidel type can also be applied, where acceleration of convergence by SOR is effective (→ 302 Numerical Solution of Linear Equations C).

F. Difference Methods for Initial Value Problems

To demonstrate characteristics of the difference method applied to evolution equa-

tions [17, 20, 21], we first consider the initial boundary value problem (14)–(16) for the diffusion equation supposing that $\Omega=(0,1)$, a 1-dimensional interval. Then we cover $Q=[0,\infty)\times\bar{\Omega}$ by a rectangular net $t=n\tau$ ($n=0,1,2,\ldots$) and $x=jh$ ($j=0,1,\ldots$) with mesh lengths $\Delta t=\tau$ and $\Delta x=h$. The value of the approximate function $u_{\tau,h}$ at the net point $(n\tau, jh)$ is denoted by U_j^n. Then the simplest **difference scheme of forward type** for (14) is written as

$$\frac{U_j^{n+1}-U_j^n}{\tau}=\frac{U_{j+1}^n+U_{j-1}^n-2U_j^n}{h^2}. \tag{33}$$

Namely, we adopt the difference operator $L_{\tau,h}[\ ]=D_t-D_{\bar{x}}^h D_x^h$ as the difference analog of $(\partial/\partial t)-(\partial^2/\partial x^2)$. Equation (33) is **consistent** in the sense that $L_{\tau,h}[u]=$ "the residual of u" $\to 0$ ($\tau, h\to 0$) for any smooth solution u of (14). This is the case also with the **difference scheme of backward type**

$$\frac{U_j^{n+1}-U_j^n}{\tau}=\frac{U_{j+1}^{n+1}+U_{j-1}^{n+1}-2U_j^{n+1}}{h^2}. \tag{34}$$

To compute the approximate solution $U^n=\{U_j^n\,|\,0\leqslant j\leqslant N\}$, (33) or (34) must be combined with the approximate initial condition corresponding to (16), say $U_j^o=a(jh)$, and with the boundary condition $U_0^n=U_N^n=0$. Then one can proceed from U^0 to U^1, from U^1 to U^2, and so on. Actually, (33) is rewritten in the form

$$U_j^{n+1}=\lambda U_{j+1}^n+\lambda U_{j-1}^n+(1-2\lambda)U_j^n, \tag{35}$$

with $\lambda=\tau/h^2$, which gives U^{n+1} explicitly in terms of U^n. Hence (33) is an **explicit scheme**. On the other hand, in order to acquire U^{n+1} from U^n according to (34), it is necessary to solve a system of linear equation with U^{n+1} as its unknown. Hence (34) is an **implicit scheme**. Since the coefficient matrix of the system of linear equations to determine U^{n+1} according to (34) is tridiagonal, one can employ a particularly efficient elimination algorithm.

We introduce the maximum norm $\|\varphi_h\|_h=\max_{R_h}|\varphi(jh)|$ for net functions φ_h on $R_h=\Omega_h\cup S_h=\{x=jh\,|\,0\leqslant j\leqslant N\}$. Then, as is obvious from (35), the approximate solution $U^n=u_{\tau,h}(n\tau)$ obtained through (33) satisfies $\|U^n\|_h\leqslant\|U^0\|_h$ ($n=0,1,\ldots$), provided that

$$0<\lambda=\frac{\tau}{h^2}\leqslant\frac{1}{2}. \tag{36}$$

Generally, a difference scheme approximating an initial value problem is said to be **stable** if the approximate solution $u_{\tau,h}$ for the initial value $a_h=u_{\tau,h}(0)$ satisfies, for small τ and h,

$$\|u_{\tau,h}(t)\|_h\leqslant M_T\|a_h\|_h \quad (0\leqslant t=n\tau\leqslant T) \tag{37}$$

for any $T>0$, M_T being a constant depending on T. **Lax's equivalence theorem** asserts that if the original initial value problem is †well-posed and if the approximating difference scheme is consistent, then the stability of the difference scheme is necessary and sufficient for the convergence of the approximate solution $u_{\tau,h}$ to the rigorous solution u as τ, $h\to 0$. For instance the explicit difference scheme (33) for the 1-dimensional diffusion equation is convergent under the mesh condition (36). On the other hand, the implicit scheme (34) is unconditionally stable and hence is convergent. Many difference schemes more sophisticated than (33) and (34), of higher accuracy or with other favorable properties, have been proposed and studied. All these difference schemes can be generalized to the case of many-dimensional Ω. For instance, if $\Omega\subset\mathbf{R}^2$, then by means of the difference operator Δ_h in (31), the forward scheme and the backward scheme are given by $D_\tau U^n=\Delta_h U^n$ and $D_\tau U^n=\Delta_h U^{n+1}$, respectively. The former is stable if $0<\lambda=\tau/h^2\leqslant 1/4$, while the latter is unconditionally stable. Moreover, in a method called the **ADI method** (alternating direction implicit method) one introduces $U^{n+1/2}$ on **fractional steps** $t=n\tau+\tau/2$ and discretizes (14) as

$$\frac{U^{n+1/2}-U^n}{\tau/2}=D_{\bar{x}}^h D_x^h U^n+D_{\bar{y}}^h D_y^h U^{n+1/2},$$

$$\frac{U^{n+1}-U^{n+1/2}}{\tau/2}=D_{\bar{x}}^h D_x^h U^{n+1}+D_{\bar{y}}^h D_y^h U^{n+1/2}. \tag{38}$$

Then, starting from U^0, the computation goes as $U^0\to U^{1/2}\to U^1\to\ldots\to U^n\to U^{n+1/2}\ldots$ by solving at each step alternatively a difference equation implicit with respect to x and y, for which a particular elimination method can be used with good efficiency. Furthermore, the ADI method is unconditionally stable. Various types of fractional-step method that generalize the ADI method have been proposed [12, 17].

We now proceed to hyperbolic equations. First consider the Cauchy problem where the spatial domain is the whole $\mathbf{R}^1$. Noting that the wave equation $v_{tt}=v_{xx}$, for instance, can be rewritten as

$$\frac{\partial u}{\partial t}=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\frac{\partial u}{\partial x}$$

by putting $u={}^t(v_t, v_x)$, we here deal with a hyperbolic system of the form

$$\frac{\partial u}{\partial t}=A\frac{\partial u}{\partial x}, \tag{39}$$

where the unknown u is an m-vector $u={}^t(u_1,u_2,\ldots,u_m)$, and A is a constant $m\times m$ matrix that is supposed to be real and symmetric. For hyperbolic equations, explicit schemes are generally preferred, and as a typ-

ical one we mention the **Friedrichs scheme**. In the Friedrichs scheme applied to (39), the approximate solution $U={}^t(U_1, U_2, \ldots, U_m)$ is determined by

$$\frac{U(t+\tau,x)-\frac{1}{2}(U(t,x+h)+U(t,x-h))}{\tau}$$

$$=A\delta_x^h U(t,x). \tag{40}$$

Here t is restricted to $t=n\tau$ $(n=0,1,\ldots)$, while x is regarded as a variable ranging over $\mathbf{R}^1$. Introducing translation operators T_h and T_{-h} by

$$T_h: \varphi(x)\to\varphi(x+h) \quad \text{and} \quad T_{-h}: \varphi(x)\to\varphi(x-h),$$

we obtain from (40)

$$U(t+\tau,\cdot)=S_h U(t,\cdot), \tag{41}$$

where, with the mesh ratio $\lambda=\tau/h$, S_h is given by

$$S_h=\left(\frac{1}{2}+\frac{\lambda}{2}A\right)T_h+\left(\frac{1}{2}-\frac{\lambda}{2}A\right)T_{-h}. \tag{42}$$

Generally, if a difference scheme is reduced to the form of (41), then the operator S_h, which yields evolution of the solution of the difference equation for one time step, is called the **amplification operator** of the scheme. The stability of the difference scheme is implied by the uniform boundedness of the operator norm $\|S_h^n\|$ of S^n acting in the Hilbert space $(L_2(R^1))^m$ for $0\leqslant n\leqslant T/\tau$. Consequently, the scheme is stable if $\|S_h\|\leqslant 1+C\tau$, C being a constant. The †symbol $\tilde{S}_h(\xi)$ of S_h with respect to the Fourier transform is called the **amplification matrix** of the scheme. The L_2-stability of the scheme mentioned above is equivalent to the uniform boundedness of the matrix norm $\|\tilde{S}_h^n(\xi)\|$ for $0\leqslant n\tau\leqslant T$ and $\xi\in\mathbf{R}^1$. Therefore, denoting by $r_h(\xi)$ the spectral radius of $\tilde{S}_h(\xi)$, one can give a necessary condition for the stability by $|r_h(\xi)|\leqslant 1+C\tau$, which is called the **von Neumann condition**. Concerning general critera for the uniform boundedness of powers of $\tilde{S}_h(\xi)$ when $\tilde{S}_h(\xi)$ is not necessarily normal, a fundamental theorem was given by H. O. Kreiss in 1962 [21].

If the largest-modulus eigenvalue of A is denoted by μ_0, then the Friedrichs scheme is stable and hence is convergent under the condition on the mesh ratio $\lambda=\tau/h$ given by

$$\lambda\leqslant 1/|\mu_0|. \tag{43}$$

In particular, (43) implies that for the stable Friedrichs scheme

propagation speed of the difference scheme

$\geqslant$ propagation speed of the original equation. (44)

Condition (44) is a necessary condition for general difference schemes approximating hyperbolic equations to be stable, and is known as the **CFL condition** (Courant-Friedrichs-Lewy condition) [22]. For a scheme more accurate than the Friedrichs scheme we refer to the Lax-Wendroff scheme, which proceeds by way of the following amplification operator S_h:

$$S_h=I+\tau A\delta_x^h+\frac{\tau^2}{2}A^2 D_{\bar{x}}^h D_x^h$$

$$=I+\frac{\lambda}{2}A(T_h+T_{-h})+\frac{\lambda^2}{2}(T_h-2I+T_{-h}).$$

The Lax-Wendroff scheme is stable under (43).

There are many other difference schemes to approximate (39), most of which are applicable to higher-dimensional spaces. Some of these schemes can be conveniently employed to solve nonlinear hyperbolic equations, for instance, those arising in the gas dynamics of compressible fluids, for which the Friedrichs scheme and the Lax-Wendroff scheme were originally intended [23]. Concerning difference schemes with variable coefficients which approximate a †regularly hyperbolic system with an x-dependent principal part, criteria for the stability of the scheme in terms of the symbol $\tilde{S}_h(x,\xi)$ of S_h have been obtained, making use of the theory of †pseudodifferential operators, by P. D. Lax and L. Nirenberg (*Comm. Pure Appl. Math.*, 1966), M. Yamaguti and T. Nogi (*Publ. Res. Inst. Math. Sci.*, 1967), R. Vaillancourt (*Math. Comput.*, 1970), Z. Koshiba (*J. Math. Anal. Appl.*, 1981), and others.

References

[1] L. V. Kantorovich and V. I. Krylov, Approximate methods of higher analysis, Interscience and Noordhoff, 1958. (Original in Russian, 1950.)
[2] J. Necas, Les méthodes directes en théorie des équations elliptiques, Masson, 1967.
[3] R. Temam, Numerical analysis, Reidel, 1973.
[4] R. Courant and D. Hilbert, Methods of mathematical physics, Interscience, I, 1953; II, 1962.
[5] J. W. Cooley and J. W. Tukey, An algorithm for the machine calculation of complex Fourier Series, Math. Comp., 19 (1964), 297–301.
[6] M. J. Turner, R. W. Clough, H. C. Martin, and L. J. Topp, Stiffness and deflection analysis of complex structures, J. Aero. Sci., 23 (1956), 805–823.
[7] P. G. Ciarlet, The finite element method for elliptic problems, North-Holland, 1978.
[8] A. R. Mitchell and R. Wait, The finite element method in partial differential equations, Wiley, 1977.

[9] G. Strang and G. Fix, An analysis of the finite element method, Prentice-Hall, 1973.
[10] F. Kikuchi and Y. Ando, On the convergence of a mixed finite scheme for plate bending, Nuclear Eng. Design, 24 (1973), 357–373.
[11] T. Miyoshi, A mixed finite element method for the solution of the von Karman equations, Numer. Math., 26 (1976), 255–269.
[12] H. Fujita and A. Mizutani, On the finite element method for parabolic equations I, Approximation of holomorphic semi-groups, J. Math. Soc. Japan, 28 (1976), 749–771.
[13] T. Kato, Perturbation theory for linear operators, Springer, second edition, 1976.
[14] T. Ushijima, Approximation theory for semi-groups of linear operators and its application to approximation of the wave equation, Japan. J. Math., 1 (1975), 185–224.
[15] T. Ikeda, Maximum principle in finite element models for convection-diffusion phenomena, Kinokuniya and North-Holland, 1983.
[16] G. E. Forsythe and W. R. Wasow, Finite-difference methods for partial differential equations, Wiley, 1960.
[17] G. I. Marchuk, Methods of numerical mathematics, Springer, second edition, 1982.
[18] G. D. Smith, Numerical solution of partial differential equations: Finite difference methods, Clarendon Press, second edition, 1978.
[19] D. L. Book (ed.), Finite-difference techniques for vectorized fluid dynamics calculations, Springer, 1981.
[20] F. John, Lectures on advanced numerical analysis, Gordon & Breach, 1967.
[21] R. D. Richtmyer and K. W. Morton, Difference method for initial value problems, Interscience, 1957.
[22] R. Courant, K. Friedrichs, and H. Lewy, Über die partiellen Differenzengleichungen der mathematischen Physik, Math. Ann., 100 (1928), 32–74.
[23] R. Peyret and T. D. Taylor, Computational methods for fluid flow, Springer, 1983.
[24] B. Carnahan, H. A. Luther, and J. O. Wilkes, Applied numerical methods, Wiley, 1969.

O

305 (IX.11) Obstructions

A. History

The theory of **obstructions** aims at measuring the extensibility of mappings by means of algebraic tools. Such classical results as the †Brouwer mapping theorem and Hopf's extension and †classification theorems in homotopy theory might be regarded as the origins of this theory. A systematic study of the theory was initiated by S. Eilenberg [1] in connection with the notions of †homotopy and †cohomology groups, which were introduced at the same time. A. Komatu and P. Olum [2] extended the theory to mappings into spaces not necessarily †n-simple. For mappings of polyhedra into certain special spaces, the †homotopy classification problem, closely related to the theory of obstructions, was solved in the following cases (K^m denotes an m-dimensional polyhedron): $K^{n+1} \rightarrow S^n$ (N. Steenrod [5]), $K^{n+2} \rightarrow S^n$ (J. Adem), $K^{n+k} \rightarrow Y$, where $\pi_i(Y)=0$ for $i<n$ and $n<i<n+k$ (M. Nakaoka). There are similar results by L. S. Pontryagin, M. Postnikov, and S. Eilenberg and S. MacLane. Except for the special cases already noted, it is extremely difficult to discuss higher obstructions in general since they involve many complexities. Nevertheless, it is significant that the idea of obstructions has given rise to various important notions in modern algebraic topology, including cohomology operations (→ 64 Cohomology Operations) and characteristic classes (→ 56 Characteristic Classes).

The notion of obstruction is also very useful in the treatment of cross sections of fiber bundles (→ 147 Fiber Bundles), †diffeomorphisms of differentiable manifolds, etc.

B. General Theory for an n-Simple Space Y

The question of whether two (continuous) mappings of a topological space X into another space Y are †homotopic to each other can be reduced to the extensibility of the given mapping: $(X \times \{0\}) \cup (X \times \{1\}) \rightarrow Y$ to a mapping of the product $X \times I$ of X and the unit interval $I=[0,1]$ into Y. Therefore the problem of classifying mappings can be treated in the same way as that of the extension of mappings.

Let K be a †polyhedron, L a subpolyhedron of K, and $\hat{K}^n = L \cup K^n$ the union of L and the †n-skeleton K^n of K. Let Y be an †arcwise connected n-simple space, and f' be a mapping of L into Y. Denote by $\bar{\Phi}^n(f')$ the set of mappings of $\hat{K}^n$ into Y that are extensions of f', and by $\Phi^n(f')$ the set of †homotopy classes of mappings in $\bar{\Phi}^n(f')$ relative to L. The set $\bar{\Phi}^0(f')$ consists of a single element because of the arcwise connectedness of Y, $\bar{\Phi}^1(f')$ is nonempty, and $\bar{\Phi}^n(f')$ $(n \geq 2)$ may be empty. Let f^n be an element of $\Phi^n(f')$. If we consider the restriction of f^n to the boundary $\dot{\sigma}^{n+1}$ of an oriented $(n+1)$-cell σ^{n+1} of K, then $f^n: \dot{\sigma}^{n+1} \rightarrow Y$ determines an element $c(f^n, \sigma^{n+1})$ of the †homotopy group $\pi_n(Y)$ (→ 202 Homotopy Theory). This element gives a measure of obstruction for extending f^n to the interior of σ^{n+1}. We obtain an $(n+1)$-†cocycle $c^{n+1}(f^n)$ of the †simplicial pair (K, L) with coefficients in $\pi_n(Y)$, called the **obstruction cocycle** of f^n, by assigning $c(f^n, \sigma^{n+1})$ to each $(n+1)$-cell σ^{n+1}. This obstruction cocycle $c^{n+1}(f^n)$ is the measure of obstruction for extending f^n to $\hat{K}^{n+1}$. A necessary and sufficient condition for the extensibility is given by $c^{n+1}(f^n)=0$. Clearly, $c^{n+1}(f^n)$ is uniquely determined for each element f of $\bar{\Phi}^n(f')$. The set of all $c^{n+1}(f^n)$ with $f^n \in \Phi^n(f')$ forms a subset $\mathbf{o}^{n+1}(f')$ of the group of cocycles $Z^{n+1}(K, L; \pi_n(Y))$. $\Phi^{n+1}(f')$ is nonempty if and only if $\mathbf{o}^{n+1}(f')$ contains the zero element 0.

Let $K^{\square} = K \times I$, $L^{\square} = (K \times 0) \cup (L \times I) \cup (K \times 1)$. Given two mappings $f_0, f_1: K \rightarrow Y$ satisfying $f_0 | L = f_1 | L$, we can define a natural mapping $F': L^{\square} \rightarrow Y$ such that an element F^n of $\Phi^n(F')$ corresponds to a †homotopy h^{n-1} relative to L connecting $f_0 | K^{n-1}$ with $f_1 | K^{n-1}$. Given an element $F^n \in \Phi^n(F')$, we have the element $c^{n+1}(F^n)$ of $Z^{n+1}(K^{\square}, L^{\square}; \pi_n(Y))$, which we identify with $Z^n(K, L; \pi_n(Y))$ through the natural isomorphism of chain groups of the pair $(K^{\square}, L^{\square})$ to those of the pair (K, L). Thus we can regard $c^{n+1}(F^n)$ as an element of $Z^n(K, L; \pi_n(Y))$, which is denoted by $d^n(f_0, h^{n-1}, f_1)$, and call it the **separation** (or **difference) cocycle**. If $f_0 | \hat{K}^{n-1} = f_1 | \hat{K}^{n-1}$, we have the canonical mapping $F^n: L^{\square} \cup (K^{\square})^n \rightarrow Y$, and the separation cocycle is denoted simply by $d^n(f_0, f_1)$. The set of separation cocycles corresponding to elements of $\Phi^n(F')$ is considered to be a subset of $Z^n(K, L; \pi_n(Y))$ and is denoted by $\mathbf{o}^n(f_0, f_1)$. A necessary and sufficient condition for h^{n-1} to be extensible to a homotopy on $\hat{K}^n$ is $d^n(f_0, h^{n-1}, f_1)=0$. Therefore a necessary and sufficient condition for $f_0 | \hat{K}^n \simeq f_1 | \hat{K}^n$ (rel L) (i.e., relative to L) is $0 \in \mathbf{o}^n(f_0, f_1)$. Given $f_0^n, f_1^n: \hat{K}^n \rightarrow Y$ with $f_0^n | L = f_1^n | L$, then $d^n(f_0^n, h^{n-1}, f_1^n)$ $(\in \mathbf{o}^n(f_0^n, f_1^n))$ is an element of $Z^n(\hat{K}^n, L; \pi_n(Y))$, which is also considered to be a cochain of the pair (K, L). In this sense, we call $d^n(f_0^n, h^{n-1}, f_1^n)$ the **separation** (or **deformation) cochain** over (K, L). The coboundary of the separation cochain $d^n(f_0^n, h^{n-1}, f_1^n)$ coincides (except possibly for sign) with $c^{n+1}(f_0^n) - c^{n+1}(f_1^n)$.

For a fixed $f_0^n \in \Phi^n(f')$, any n-cochain d^n

of the pair (K, L) with coefficients in $\pi_n(Y)$ is expressible as a separation cochain $d^n = d^n(f_0^n, f_1^n)$ where $f_1^n \in \Phi^n(f')$ is a suitable mapping such that $f_0^n | \hat{K}^{n-1} = f_1^n | \hat{K}^{n-1}$ (existence theorem).

Therefore if we take an element f^{n-1} of $\bar{\Phi}^{n-1}(f')$ whose obstruction cocycle $c^n(f^{n-1})$ is zero, the set of all obstruction cocycles $c^{n+1}(f^n)$ of all such $f^n \in \Phi^n(f')$ that are extensions of f^{n-1} forms a subset of $\mathbf{o}^{n+1}(f')$ and coincides with a coset of $Z^{n+1}(K, L; \pi_n(Y))$ factored by $B^{n+1}(K, L; \pi_n(Y))$. Thus a cohomology class $\bar{c}^{n+1}(f^{n-1}) \in H^{n+1}(K, L; \pi_n(Y))$ corresponds to an $f^{n-1} \in \Phi^{n-1}(f')$ such that $c^n(f^{n-1}) = 0$, and $\bar{c}^{n+1}(f^{n-1}) = 0$ is a necessary and sufficient condition for f^{n-1} to be extensible to $\hat{K}^{n+1}$ **(first extension theorem)**.

For the separation cocycle, $\bar{d}^n(f_0, h^{n-2}, f_1) \in H^n(K, L; \pi_n(Y))$ corresponds to each homotopy h^{n-2} on $\hat{K}^{n-2}$ such that $d^{n-1}(f_0, h^{n-2}, f_1) = 0$, and $\bar{d}^n(f_0, h^{n-2}, f_1) = 0$ is a necessary and sufficient condition for h^{n-2} to be extensible to a homotopy on $\hat{K}^n$ **(first homotopy theorem)**.

The subset of $H^{n+1}(K, L; \pi_n(Y))$ corresponding to $\mathbf{o}^{n+1}(f')$ is denoted by $\mathbf{O}^{n+1}(f')$ and is called the **obstruction to an** $(n+1)$**-dimensional extension** of f'. Similarly, the subset $\mathbf{O}^n(f_0, f_1)$ of $H^n(K, L; \pi_n(Y))$ corresponding to $\mathbf{o}^n(f_0, f_1)$ is called the **obstruction to an** n**-dimensional homotopy** connecting f_0 with f_1. Clearly, a condition for f' to be extensible to $\hat{K}^{n+1}$ is given by $0 \in \mathbf{O}^{n+1}(f')$, and a necessary and sufficient condition for $f_0 | K^n \simeq f_1 | K^n$ (rel L) is given by $0 \in \mathbf{O}^n(f_0, f_1)$.

A continuous mapping $\varphi : (K', L') \to (K, L)$ induces homomorphisms of cohomology groups $\varphi^* : H^{n+1}(K, L; \pi_n(Y)) \to H^{n+1}(K', L'; \pi_n(Y))$, $H^n(K, L; \pi_n(Y)) \to H^n(K', L'; \pi_n(Y))$. Then for $f' : L \leftarrow Y$, $\mathbf{O}^{n+1}(f' \circ \varphi) \supset \varphi^* \mathbf{O}^{n+1}(f')$, and for $f_0, f_1 : K \to Y$ such that $f_0 | L = f_1 | L$, $\mathbf{O}^n(f_0 \circ \varphi, f_1 \circ \varphi) \supset \varphi^* \mathbf{O}^n(f_0, f_1)$. Therefore we also find that the obstruction to an extension and the obstruction to a homotopy are independent of the choice of subdivisions of K, L, and consequently are topological invariants.

Let f_0, f_1, and f_2 be mappings $K \to Y$ such that $f_0 | L = f_1 | L = f_2 | L$. Given homotopies $h_{01}^{n-1} : f_0 | \hat{K}^{n-1} \simeq f_1 | \hat{K}^{n-1}$ (rel L), $h_{12}^{n-1} : f_1 | \hat{K}^{n-1} \simeq f_2 | \hat{K}^{n-1}$ (rel L), then for the composite $h_{02}^{n-1} = h_{12}^{n-1} \circ h_{01}^{n-1}$, we have

$$d^n(f_0, h_{01}^{n-1}, f_1) + d^n(f_1, h_{12}^{n-1}, f_2) = d^n(f_0, h_{02}^{n-1}, f_2),$$

and for the inverse homotopy $h_{10}^{n-1} : f_1 | \hat{K}^{n-1} \simeq f_0 | \hat{K}^{n-1}$ of h_{01}^{n-1}, clearly

$$d^n(f_1, h_{10}^{n-1}, f_0) = -d^n(f_0, h_{01}^{n-1}, f_1).$$

Therefore $\mathbf{O}^n(f_0, f_0)$ forms a subgroup of $H^n(K, L; \pi_n(Y))$ that is determined by the homotopy class of $f_0 | \hat{K}^{n-1}$ relative to L. In general, if $\mathbf{O}^n(f_0, f_1)$ is nonempty, it is a coset of $H^n(K, L; \pi_n(Y))$ factored by the subgroup $\mathbf{O}^n(f_0, f_0)$. Combined with the existence theorem on separation cochains, this can be utilized to show the following theorem.

Assume that $\Phi^n(f')$ is nonempty. The set of all elements $\bar{\Phi}^n(f')$ that are extensions of an element of $\bar{\Phi}^{n-1}(f')$ is put in one-to-one correspondence with the quotient group of $H^n(\hat{K}^n, L; \pi_n(Y))$ modulo $\mathbf{O}^n(f_0^n, f_0^n)$ by pairing the obstruction $\mathbf{O}^n(f_0^n, f^n)$ with each f^n for a fixed f_0^n. Among such elements of $\bar{\Phi}^n(f')$, the set of f^n that are extensible to $\hat{K}^{n+1}$ is in one-to-one correspondence with the quotient group of $H^n(\hat{K}^{n+1}, L; \pi_n(Y)) = H^n(K, L; \pi_n(Y))$ modulo the subgroup $\mathbf{O}^n(f_0^{n+1}, f_0^{n+1})$, assuming that f_0^n is extended to f_0^{n+1} **(first classification theorem)**.

C. Primary Obstructions

Assume that $H^{i+1}(K, L; \pi_i(Y)) = H^i(K, L; \pi_i(Y)) = 0$, where $0 < i < p$ (e.g., $\pi_i(Y) = 0$, $0 < i < p$). In this case, by consecutive use of the first extension theorem and the first homotopy theorem, we can show that each $\bar{\Phi}^i(f')$ $(i < p)$ consists of a single element and $\mathbf{O}^{p+1}(f')$ also consists of a single element $\bar{c}^{p+1}(f') \in H^{p+1}(K, L; \pi_p(Y))$. The element $\bar{c}^{p+1}(f')$, called the **primary obstruction** of f', vanishes if and only if f' can be extended to $\hat{K}^{p+1}$ **(second extension theorem)**. When $H^{i+1}(K, L; \pi_i(Y)) = 0$ for $i > p$ (for example, when $\pi_i(Y) = 0$ for $p < i < \dim(K - L)$), f' is extendable to K if and only if the first obstruction of f' vanishes **(third extension theorem)**.

Correspondingly, if $H^i(K, L; \pi_i(Y)) = H^{i-1}(K, L; \pi_i(Y)) = 0$ $(0 < i < p)$, then for any two mappings $f_0, f_1 : K \to Y$, $f_0 | L = f_1 | L$, $\mathbf{O}^p(f_0, f_1)$ consists of a single element $\bar{d}^p(f_0, f_1) \in H^p(K, L; \pi_p(Y))$, which we call the **primary difference** of f_0 and f_1. This element vanishes if and only if $f_0 | \hat{K}^p \simeq f_1 | \hat{K}^p$ (rel L) **(second homotopy theorem)**. Moreover, when $H^i(K, L; \pi_i(Y)) = 0$ $(i > p)$, the primary difference is zero if and only if $f_0 \simeq f_1$ (rel L) **(third homotopy theorem)**.

Assume that the hypotheses of the second extension theorem and second homotopy theorem are satisfied. If we assign to each element f^p of $\bar{\Phi}^p(f')$ the primary difference of f^p and the fixed element f_0^p, then $\bar{\Phi}^p(f')$ is in one-to-one correspondence with $H^p(\hat{K}^p, L; \pi_p(Y))$ by the first classification theorem **(second classification theorem)**. Similarly, assume that the hypotheses of the third extension theorem and third homotopy theorem are satisfied. If $f_0 : K \to Y$, $f' = f_0 | L$, then homotopy classes relative to L of extensions f of f' are put in one-to-one correspondence with the

elements of $H^p(K, L; \pi_p(Y))$ by pairing $d^p(f, f_0)$ with f (**third classification theorem**).

D. Secondary Obstructions

For simplicity, assume that $\pi_i(Y)=0$ ($i<p$ and $p<i<q$). If the primary obstruction $\bar{c}^{p+1}(f')\in H^{p+1}(K, L; \pi_p(Y))$ of $f': L \rightarrow Y$ vanishes, we can define $\mathbf{O}^{q+1}(f')\subset H^{q+1}(K, L; \pi_q(Y))$, which we call the **secondary obstruction** of f'. When $Y=S^p$, $q=p+1$, $p>2$, the secondary obstruction $\mathbf{O}^{p+2}(f')$ coincides with a coset of $H^{p+2}(K, L; Z_2)$ modulo the subgroup $Sq^2(H^p(K, L; Z))$, where Sq^2 denotes the †Steenrod square operation [5]. In this case, if $L=K^p$, then $\mathbf{O}^{p+2}(f')$ reduces to a cohomology class, $Sq^2(i^*)^{-1}f'^*(\sigma)$ with $i: L \rightarrow K$, where σ is a generator of $H^p(S^p, Z)$ (in this case $(i^*)^{-1}f'^*(\sigma)\neq\varnothing$ is equivalent to $\bar{c}^{p+1}(f')=0$) [5]. Moreover, if $Sq^2 f'^*(\sigma)=0$, then there exists a suitable extension $f^{p+2}: \hat{K}^{p+2}\rightarrow Y=S^p$ of f'. The set of obstruction cocycles of all such f^{p+2} defines the **tertiary obstruction** $\mathbf{O}^{p+3}(f')$, which coincides with a coset of $H^{p+3}(K, L; Z_2)$ modulo the subgroup $Sq^2(H^{p+1}(K, L; Z_2))$. By using the †secondary cohomology operation Φ of J. Adem, it can be expressed as $\Phi((i^*)^{-1}f'^*(\sigma))$ (→ 64 Cohomology Operations).

All the propositions in this article remain true if we take †CW complexes instead of polyhedra K.

References

[1] S. Eilenberg, Cohomology and continuous mappings, Ann. Math., (2) 41 (1940), 231–251.
[2] P. Olum, Obstructions to extensions and homotopies, Ann. Math., (2) 52 (1950), 1–50.
[3] P. Olum, On mappings into spaces in which certain homotopy groups vanish, Ann. Math., (2) 57 (1953), 561–574.
[4] N. E. Steenrod, The topology of fibre bundles, Princeton Univ. Press, 1951.
[5] N. E. Steenrod, Products of cocycles and extensions of mappings, Ann. Math., (2) 48 (1947), 290–320.
[6] E. H. Spanier, Algebraic topology, McGraw-Hill, 1966.

306 (XII.20) Operational Calculus

A. General Remarks

The term "operational calculus" in the usual sense means a method for solving †linear differential equations by reducing the operations of differentiation and integration into algebraic ones in a symbolic manner. The idea was initiated by P. S. Laplace in his *Théorie analytique des probabilités* (1812), but the method has acquired popularity since O. Heaviside used it systematically in the late 19th century to solve electric-circuit problems. The method is therefore also called **Heaviside calculus**, but Heaviside gave only a formal method of calculus without bothering with rigorous arguments. The mathematical foundations were given in later years, first in terms of †Laplace transforms, then by applying the theory of †distributions. One of the motivations behind L. Schwartz's creation of this latter theory in the 1940s was to give a sound foundation for the formal method, but the theory obtained has had a much larger range of applications. Schwartz's theory was based on the newly developed theory of †topological linear spaces. On the other hand, J. Mikusiński gave another foundation, based only on elementary algebraic notions and on Titchmarsh's theorem, whose proof has recently been much simplified.

In this article, we first explain the simple theory established by Mikusiński [2] and later discuss its relation to the classical Laplace transform method.

B. The Operational Calculus of Mikusiński

The set $\mathscr{C}$ of all continuous complex-valued functions $a=\{a(t)\}$ defined on $t\geqslant 0$ is a †linear space with the usual addition and scalar multiplication. $\mathscr{C}$ is a †commutative algebra with multiplication $a\cdot b$ defined by the †convolution $\{\int_0^t a(t-s)b(s)ds\}$. The ring $\mathscr{C}$ has no †zero divisors (**Titchmarsh's theorem**). (There have been several interesting proofs of Titchmarsh's theorem since the first demonstration given by Titchmarsh himself [3]. For example, a simple proof has been published by C. Ryll-Nardzewski (1952).) Hence we can construct the †quotient field $\mathscr{Q}$ of the ring $\mathscr{C}$. An element of $\mathscr{Q}$ is called a **Mikusiński operator**, or simply an **operator**. If we define $a(t)=0$ for $t<0$ for the elements $\{a(t)\}$ in $\mathscr{C}$, then $\mathscr{C}$ is a subalgebra of $\mathscr{U}$, which is the set of all locally integrable (locally L_1) functions in $(-\infty, \infty)$ whose †support is bounded below. Here we identify two functions that coincide almost everywhere. The algebra $\mathscr{U}$ has no zero divisor, and its quotient field is also $\mathscr{Q}$.

The unity element for multiplication in $\mathscr{Q}$, denoted by $\delta=b/b$ ($b\neq 0$), plays the role of the †Dirac δ-function. It is sometimes called the **impulse function**. The operator $l=\{1\}\in\mathscr{C}$ is the

function that takes the values 0 and 1 according as $t<0$ or $t>0$. This operator is **Heaviside's function** and is sometimes called the **unit function**. Usually it is denoted by $\mathbf{1}(t)$ or simply $\mathbf{1}$. The value $\mathbf{1}(0)$ may be arbitrary, but usually it is defined as 1/2, the mean of the limit values from both sides. The operator l is an **integral operator**, because, as an operator carrying a into $l\cdot a$, it yields

$$l\cdot a=\left\{\int_0^t a(s)ds\right\}=\text{the integral of } a \text{ over } [0,t].$$

More generally, the operator $\{t^{\lambda-1}/\Gamma(\lambda)\}$ ($\operatorname{Re}\lambda>0$) gives the λth-order integral. The operator $s=\delta/l$, which is the inverse operator of l, is a **differential operator**. If $a\in\mathscr{C}$ is of †class C^1, then we have

$$s\cdot a=a'+a(+0)\delta=a'+\{a(+0)\}/\{1\}. \quad (1)$$

Similarly, if $a\in\mathscr{C}$ is of class C^n, we have

$$s^n\cdot a=a^{(n)}+a^{(n-1)}(0)\delta+a^{(n-2)}(0)s+\ldots+a(0)s^{n-1}. \quad (2)$$

The operator $a\to s\cdot a$ can be applied to functions a that are not differentiable in the ordinary sense, and considering the application of s to be the operation of differentiation, we can treat the differential operator algebraically in the field $\mathscr{Q}$. In particular, we have $s\cdot\mathbf{1}=\delta$, and this relation is frequently represented by the formula

$$d\mathbf{1}(t)/dt=\delta(t). \quad (3)$$

A rational function of s whose numerator is of lower degree than its denominator is an †elementary function of t. For example, we have the relations

$$1/(s-\alpha)^n=\{t^{n-1}e^{\alpha t}/(n-1)!\},$$

$$1/(s^2+\beta^2)=\{\beta^{-1}\sin\beta t\},$$

$$s/(s^2+\beta^2)=\{\cos\beta t\}. \quad (4)$$

The solution of an ordinary linear differential equation with constant coefficients $\sum_{r=0}^{n}\alpha_r\varphi^{(r)}(t)=f(t)$ under the †initial condition $\varphi^{(i)}(0)=\gamma_i$ $(0\leqslant i\leqslant n-1)$ is thus reduced to an equation in s by using formulas (1) and (2), and is computed by decomposing the following operator into partial fractions:

$$\{\varphi(t)\}=\frac{\beta_{n-1}s^{n-1}+\ldots+\beta_0+f}{\alpha_n s^n+\ldots+\alpha_1 s+\alpha_0}, \quad (5)$$

where $\beta_r=\alpha_{r+1}\gamma_0+\alpha_{r+2}\gamma_1+\ldots+\alpha_n\gamma_{n-r-1}$, $0\leqslant r\leqslant n-1$. The general solution is represented by (5) if we consider the constants $\gamma_0,\ldots,\gamma_{n-1}$ or $\beta_0,\ldots,\beta_{n-1}$ as arbitrary parameters. If the rational function in the right-hand side of (5) is $M(s)/D(s)$ and the degree of the numerator is less than that of the denominator, then the right-hand side of (5) is explicitly represented by

$$\sum_{r=0}^{l-1}\frac{t^{l-r-1}}{(l-r-1)!r!}\left[\frac{d^r}{d\lambda^r}\left(\frac{M(\lambda)}{D(\lambda)/(\lambda-\lambda_0)^l}\right)\right]_{\lambda=\lambda_0}e^{\lambda_0 t}+\sum_{i=1}^{m}\frac{M(\lambda_i)}{D(\lambda_i)}e^{\lambda_i t},$$

where we assume that $\lambda_0,\lambda_1,\ldots,\lambda_m$ exhaust the roots of the equation $D(\lambda)=0$, λ_0 is a multiple root of degree l, and all other roots are simple $(m=n-l)$. The above formula is called the **expansion theorem**.

C. Limits of Operators

A sequence a_n of operators is said to **converge** to the **limit** $a=b/q$ if there exists an operator $q(\neq 0)$ such that $q\cdot a_n\in\mathscr{C}$ and the sequence of functions $q\cdot a_n$ converges †uniformly to b on compact sets. The limit a is determined uniquely without depending on the operator q. Based on this notion of limits of operators, we can construct the theory of series of operators and differential and integral calculus of functions of an independent variable λ whose values are operators. They are completely parallel to the usual theories of elementary calculus (→ 106 Differential Calculus; 216 Integral Calculus; 379 Series). A linear partial differential equation in the function $\varphi(x,t)$ of two variables, in particular its initial value problem, reduces to a linear ordinary differential equation of an operational-valued function of an independent variable x.

For a given operator w, the solution (if it exists) of the differential equation $\varphi'(\lambda)=w\cdot\varphi(\lambda)$ with the initial condition $\varphi(0)=\delta$ is unique, is called the **exponential function of an operator** w, and is denoted by $\varphi(\lambda)=e^{\lambda w}$. If the power series

$$\sum_{n=0}^{\infty}\lambda^n w^n/n! \quad (6)$$

converges, the limit is identical to the exponential function $e^{\lambda w}$. However, there are several cases in which $e^{\lambda w}$ exists even when the series (6) of operators does not converge.

For example, for $w=-\sqrt{s}$, we have

$$e^{-\lambda\sqrt{s}}=\{(\lambda/2\sqrt{\pi t^3})\exp(-\lambda^2/4t)\}, \quad (7)$$

and for $w=-s$, we have

$$e^{-\lambda s}=h^\lambda=s\cdot H_\lambda(t), \quad (8)$$

where the function $H_\lambda(t)$ takes the values 0 and 1 according as $t<\lambda$ or $t>\lambda$. $H_\lambda(t)$ belongs to the ring $\mathscr{U}$ and is called the **jump function** at λ. For $f(t)\in\mathscr{U}$, we have

$$h^\lambda\cdot\{f(t)\}=\{f(t-\lambda)\},$$

and hence we call (8) the **translation operator**

(or **shift operator**). For $w=-s$, the series (6) does not converge, but if we apply the formal relation $e^{-\lambda s}=\sum_{n=0}^{\infty}(-\lambda s)^n/n!$ to $f(t)$, we have a formal Taylor expansion

$$f(t-\lambda)=\sum_{n=0}^{\infty} f^{(n)}(t)(-\lambda)^n/n!.$$

The solution of linear †difference equations are represented by rational functions of h^λ. The power series $\sum \alpha_n h^n$ of operators always converges. This fact gives an explicit example of a representation by formal power series. Note that the operators $e^{-\lambda s}$ and $e^{-\lambda\sqrt{s}}$ play an essential role in the solution of the †wave equation

$$\frac{\partial^2\varphi}{\partial x^2}=\frac{\alpha^2\partial^2\varphi}{\partial t^2}$$

and the †heat equation

$$\frac{\partial^2\varphi}{\partial x^2}=\frac{\alpha^2\partial\varphi}{\partial t}.$$

The operator (7) converges to δ for $\lambda\to 0$, and this gives a †regularization of the Dirac δ-function.

D. Laplace Transform

For every function $\{f(t)\}\in\mathscr{C}$, the limit

$$\lim_{\beta\to+\infty}\int_0^\beta e^{-\lambda s}f(\lambda)d\lambda=\int_0^\infty e^{-\lambda s}f(\lambda)d\lambda \qquad (9)$$

always exists (in the sense of the limit of operators), and as an operator coincides with the original function $\{f(t)\}$. Therefore, if the usual Laplace transform (→ 240 Laplace Transform) of the function $f(\lambda)$ exists and (9) is a function $g(s)$, then as a function of the differential operator s, $g(s)$ is the operator that is given by the inverse Laplace transform $f(t)$ of $g(s)$. Formulas (4) and (7) are indeed typical examples of this relation, where the left-hand side is the usual Laplace transform of the right-hand side. In the practical computation of (5), we can compute the Laplace inverse transform of the right-hand side. However, if we took the Laplace transform as the foundation of the theory, it would not only be complicated but also be subject to the artificial restriction caused by the convergence condition on Laplace transforms.

In the theory of operational calculus, the transform

$$g(p)=p\int_0^\infty e^{-pt}f(t)dt \qquad (10)$$

is sometimes used instead of the Laplace transform itself. But the difference is not essential; we obtain the latter transform merely by changing the variable from s to p and multiplying the former transform by p.

E. Relation to Distributions

For $f\in\mathscr{C}$, an operator of the form $h^\lambda\cdot s^k\cdot f$ is identified with a distribution of L. Schwartz with support bounded from below. We can identify with a Schwartz distribution the limit of a sequence f_n (or a suitable equivalence class of sequences) of operators of the form $h^\lambda\cdot s^k\cdot f$ such that $f_n, f_{n+1}, \ldots$ are identical in the interval $(-n,n)$. The notions of Schwartz distributions and of Mikusiński operators do not include each other, but both are generalizations of the notions of functions and their derivatives. For formulas and examples → Appendix A, Table 12.

References

[1] O. Heaviside, On operators in physical mathematics I, II, Proc. Roy. Soc. London, ser. A, 52 (1893), 504–529; 54 (1893), 105–442.
[2] J. Mikusiński, Operational calculus, Pergamon, 1959. (Original in Polish, 1953.)
[3] E. C. Titchmarsh, The zeros of certain integral functions, Proc. London Math. Soc., 25 (1926), 283–302.
[4] K. Yosida, Operational calculus: A theory of hyperfunctions, Springer, 1984.

307 (XIX.13) Operations Research

A. General Remarks

Operations research in the most general sense can be characterized as the application of scientific methods, techniques, and tools to the operations of systems so as to provide those in control with optimum solutions to problems. This definition is due to Churchman, Ackoff, and Arnoff [1]. Operational research began in a military context in the United Kingdom during World War II, and it was quickly taken up under the name operations research (OR) in the United States. After the war it evolved in connection with industrial organization, and its many techniques have found expanding areas of application in the United States, the United Kingdom, and in other industrial countries. Nowadays OR is used widely in industry for solving practical problems, such as planning, scheduling, inventory, transportation, and marketing. It also has various important applications in the fields of agricul-

ture, commerce, economics, education, public service, etc., and some techniques developed in OR have influenced other fields of science and technology.

B. Applications

Applications of OR to practical problems are often carried out by project teams because knowledge of disparate aspects of the problems are required, and interdisciplinary cooperation is indispensable. The following are the major phases of an OR project: (i) formulating the problem, (ii) constructing a mathematical model to represent the system under study, (iii) deriving a solution from the model, (iv) testing the model and the solution derived from it, (v) establishing controls over the solution and putting it to work (implementation).

When the mathematical model that has been constructed in phase (ii) is complicated and/or the amount of data to be handled is large, a digital †computer is often utilized in phases (iii) and (iv).

C. Mathematical Models [2]

Typical mathematical models and tools that appear frequently in OR are:

(1) **Optimization model** (→ 264 Mathematical Programming). This model is characterized by one or more real-valued functions, which are called **objective functions**, to be minimized (or maximized) under some constraints. According to the number of objective functions, the types of objective functions, and the types of constraints, this model is classified roughly as follows: (i) **Single-objective model**, which includes linear, quadratic, convex, nonlinear and integer programming models (→ 215 Integer Programming, 255 Linear Programming, 292 Nonlinear Programming, 349 Quadratic Programming); (ii) **multi-objective model**; (iii) **stochastic programming model** (→ 408 Stochastic Programming); (iv) **dynamic programming model** (→ 127 Dynamic Programming); (v) **network flow model** (→ 281 Network Flow Problems).

(2) **Game-theoretic model** (→ 173 Game Theory). Game theory is a powerful tool for deriving a solution to practical problems in which more than one person is involved, with each player having different objectives.

(3) **Inventory model** (→ 227 Inventory Control). It is necessary for most firms to control stocks of resources, products, etc., in order to carry out their activities smoothly; various inventory models have been developed for such problems. Mathematically, optimization techniques (→ 264 Mathematical Programming), Markovian decision processes (→ 127 Dynamic Programming, 261 Markov Processes), and basic †probability theory are used to construct models for these problems.

(4) **Queuing model** (→ 260 Markov Chains H). In a telephone system, calls made when all the lines of the system are busy are lost. The problem of computing the **probability of loss** involved was first solved in the pioneering article on †queuing theory by A. K. Erlang in 1917. For systems in which calls can be put on hold when all lines are busy, one deals with the †waiting time distribution instead of the probability of loss. In the 1930s, F. Pollaczeck and A. Ya. Khinchin gave explicit formulas for the characteristic function of the waiting time distribution. In many situations, such as in telephone systems, air and surface traffic, production lines, and computer systems, various congestion phenomena are often observed, and many kinds of queuing models are utilized to analyze the congestion. Mathematically, almost all such models are formulated by using Markov processes. For practical uses, approximation and computational methods are important as well as theoretical results.

(5) **Scheduling model** (→ 376 Scheduling and Production Planning). **Network scheduling** is used to schedule complicated projects (for example, construction of buildings) that consist of a large number of jobs related to each other in some natural order. **PERT** (program evaluation and review technique) and **CPM** (critical path method) are popular computational methods for this model (→ 281 Network Flow Problems). **Job shop scheduling** is used when we have m jobs and n machines and each job requires some of the machines in a given order. The model allows us to find an optimal order (in some certain sense) of jobs to be implemented on each machine.

(6) **Replacement model**. There are two typical cases. One is the **preventive maintenance model**, which is suitable when replacements are done under a routine policy because a replacement or a repair before a failure is more effective than after a failure. Probabilistic treatments are mainly used, and this model resembles those for queues and †Markov processes. The other is a model for deciding whether to replace a piece of equipment in use. In this case, we need to compare costs of both used and new equipment, and evaluations of various types of present cost are important.

(7) †**Simulation**. This is a numerical experiment in a simulated model of a phenomenon which we want to analyze. Simulation is one of the most popular techniques in OR.

(8) **Other models**. Besides the models listed above, many problems are formulated by way of various other models in OR. In modeling,

†probability theory, and mathematical †statistics, especially, †Markov chain, †multivariate analysis, †design of experiments, †regression analysis, †time series analysis, etc. often play important roles.

References

[1] C. W. Churchman, R. L. Ackoff, and E. L. Arnoff, Introduction to operations research, Wiley, 1957.
[2] H. M. Wagner, Principles of operations research, Prentice-Hall, 1975.

308 (XII.19) Operator Algebras

A. Preliminaries

Let $\mathfrak{H}$ be a †Hilbert space. The set of †bounded linear operators on $\mathfrak{H}$ is denoted by $\mathscr{B}(\mathfrak{H})=\mathscr{B}$. It contains the identity operator I. The notions of operator sum $A+B$, operator product AB, and †adjoint A^* are defined on it. A subalgebra of $\mathscr{B}(\mathfrak{H})$ is called an **operator algebra**. In this article we consider mainly von Neumann algebras. For C^*-algebras → 36 Banach Algebras G–K.

Any †Hermitian operator A (i.e., an operator such that $A=A^*$) has the property that (Ax,x) is always real for any $x\in\mathfrak{H}$. If $(Ax,x)\geqslant 0$ for any x, A is called **positive**, and we write $A\geqslant 0$. When Hermitian operators A and B satisfy $A-B\geqslant 0$, we write $A\geqslant B$. Thus we introduce an ordering $A\geqslant B$ between Hermitian operators. A set $\{A_\alpha\}$ of positive Hermitian operators is said to be an **increasing directed set** if any two of them A_α, A_β always have a common majorant A_γ, that is, $A_\alpha\leqslant A_\gamma$ and $A_\beta\leqslant A_\gamma$. If a Hermitian operator A satisfies $(Ax,x)=\sup(A_\alpha x,x)$ for such a set, it is called the **supremum** and is denoted by $\sup A_\alpha$. The supremum $\sup A_\alpha$ exists if and only if the $\sup(A_\alpha x,x)$ is finite for any x, and then A_α converges to A with respect to the weak and strong operator topologies.

B. Topologies in $\mathscr{B}$

Various topologies are introduced in $\mathscr{B}=\mathscr{B}(\mathfrak{H})$: the †uniform operator topology, the †strong operator topology, and the †weak operator topology (→ 251 Linear Operators). These topologies are listed above in order of decreasing fineness. The operation in $\mathscr{B}$ of taking the adjoint, $A\to A^*$, is continuous with respect to the uniform operator topology and weak operator topology, but not with respect to the strong operator topology. The operation from $\mathscr{B}\times\mathscr{B}$ to $\mathscr{B}$ of taking the product, $(A,B)\to AB$, is continuous with respect to the uniform operator topology, is continuous with respect to the strong operator topology when the first factor is restricted to a set bounded in the operator norm, but is not continuous on $\mathscr{B}\times\mathscr{B}$. It is continuous with respect to the weak operator topology when one of the factors is fixed (i.e., it is separately continuous). The set $\mathscr{B}$ is a †Banach space with respect to the operator norm, or, more precisely, a †C^*-algebra. It is a †locally convex topological linear space with respect to the strong or weak operator topology.

The Banach space $\mathscr{B}$ is the †dual of the Banach space $\mathscr{B}_*$ of all †nuclear operators in $\mathfrak{H}$ (→ 68 Compact and Nuclear Operators I). The weak* topology in $\mathscr{B}$ as the dual of $\mathscr{B}_*$ is called the **σ-weak topology**.

C. Von Neumann Algebras

A subset $\mathscr{M}$ of $\mathscr{B}$ is called a **$*$-subalgebra** if it is a subalgebra (i.e., A, $B\in\mathscr{M}$ implies $\lambda A+\mu B$, $AB\in\mathscr{M}$) and contains the adjoint A^* of any $A\in\mathscr{M}$. The **commutant** $\mathscr{A}'$ of a subset $\mathscr{A}$ of $\mathscr{B}$ is the set of operators that commute with both A and A^* for $A\in\mathscr{A}$. The commutant is a $*$-subalgebra, and $\mathscr{A}'=\mathscr{A}'''$.

A **von Neumann algebra** $\mathscr{M}$ is a $*$-subalgebra of $\mathscr{B}$ that is defined by one of the following four equivalent conditions: (i) $\mathscr{M}$ is a $*$-subalgebra of $\mathscr{B}$ containing I, closed under the weak operator topology; (ii) $\mathscr{M}$ is a $*$-subalgebra of $\mathscr{B}$ containing I, closed under the strong operator topology; (iii) $\mathscr{M}$ is the commutant of a subset of $\mathscr{B}$ (or, equivalently, $\mathscr{M}=\mathscr{M}''$); (iv) $\mathscr{M}$ is a $*$-subalgebra of $\mathscr{B}$ containing I, closed under the uniform operator topology, and, as a Banach space, coinciding with the conjugate space of some Banach space. Note that a $*$-subalgebra of $\mathscr{B}$, closed under the uniform operator topology, is a C^*-algebra. Von Neumann algebras are also called **rings of operators** or **W^*-algebras**. The latter term is usually used for a C^*-algebra $*$-isomorphic to a von Neumann algebra in contrast to a concrete von Neumann algebra.

The study of these algebras was started by J. von Neumann in 1929. He showed the equivalence of conditions (i)–(iii) (**von Neumann's density theorem**), and established a foundation for the theory named after him [1]. The notion of von Neumann algebras can be regarded as a natural extension of the notion of matrix algebras in a finite-dimensional space, and therein lies the importance of the theory. The

fourth condition of the definition was given by S. Sakai (*Pacific J. Math.*, 1956).

The following theorem is of use in the theory of von Neumann algebras: Given a $*$-subalgebra $\mathcal{A}$ of $\mathcal{B}$ containing I, its closure $\mathcal{M}$ with respect to the weak or strong operator topology is von Neumann algebra; and when we denote the set of elements of operator norm $\leqslant 1$ in $\mathcal{A}$ (resp. $\mathcal{M}$) by $\mathcal{A}_1$ (resp. $\mathcal{M}_1$), $\mathcal{M}_1$ is likewise the closure of $\mathcal{A}_1$ with respect to the weak or strong operator topology (**Kaplansky's density theorem** (*Ann. Math.*, 1951).

If E is a projection operator in a von Neumann algebra $\mathcal{M}$, then $E\mathcal{M}E=\{EAE|A\in\mathcal{M}\}$ is a $*$-subalgebra of $\mathcal{B}$ closed with respect to the weak operator topology. It is not a von Neumann algebra because it does not contain I, but since its elements operate exclusively in the closed subspace $E\mathfrak{H}$, we can regard it as an algebra of operators on $E\mathfrak{H}$. In this sense, $E\mathcal{M}E$ can be regarded as a von Neumann algebra on $E\mathfrak{H}$, which we call the **reduced von Neumann algebra** of $\mathcal{M}$ on $E\mathfrak{H}$ and write $\mathcal{M}_E$. If E is a projection operator in $\mathcal{M}'$, $E\mathcal{M}E=E\mathcal{M}$ restricted to the subspace $E\mathfrak{H}$ is called the **induced von Neumann algebra** of $\mathcal{M}$ on $E\mathfrak{H}$ and is denoted also by $\mathcal{M}_E$. In the latter case, the mapping $A\in\mathcal{M}\rightarrow EA\in\mathcal{M}_E$ is a $*$-homomorphism and is called the **induction** of $\mathcal{M}$ onto $\mathcal{M}_E$.

The **tensor product** $\mathfrak{H}_1\otimes\mathfrak{H}_2$ of two Hilbert spaces $\mathfrak{H}_i$ $(i=1,2)$ is the †completion of their †tensor product as a complex linear space equipped with the unique †inner product satisfying $(f_1\otimes f_2, g_1\otimes g_2)=(f_1,g_1)_1(f_2,g_2)_2$ for all $f_i, g_i\in\mathfrak{H}_i$. For any $A_i\in\mathcal{B}(\mathfrak{H}_i)$, there exists a unique operator in $\mathcal{B}(\mathfrak{H}_1\otimes\mathfrak{H}_2)$ denoted by $A_1\otimes A_2$ satisfying $(A_1\otimes A_2)(f_1\otimes f_2)=A_1f_1\otimes A_2f_2$ for all $f_i\in\mathfrak{H}_i$. For von Neumann algebras $\mathcal{M}_i\subset\mathcal{B}(\mathfrak{H}_i)$, the von Neumann algebra generated by $A_1\otimes A_2$ with $A_i\in\mathcal{M}_i$ is denoted by $\mathcal{M}_1\otimes\mathcal{M}_2$ and is called the **tensor product** of $\mathcal{M}_1$ and $\mathcal{M}_2$. The $*$-isomorphism $A\in\mathcal{M}\rightarrow A\otimes I\in\mathcal{M}\otimes\mathbf{I}$, where $\mathbf{I}$ is the trivial von Neumann algebra consisting solely of complex multiples of the identity operator, is called an **amplification**.

For two von Neumann algebras $\mathcal{M}_i\subset\mathcal{B}(\mathfrak{H}_i)$ $(i=1,2)$, a $*$-isomorphism π from $\mathcal{M}_1$ into $\mathcal{M}_2$ is called **spatial** if there exists a unitary (i.e., a bijective isometric linear) mapping U from $\mathfrak{H}_i$ to $\mathfrak{H}_2$ such that $UAU^*=\pi A$ for all $A\in\mathcal{M}_1$, and a $*$-homomorphism π is called **normal** if $\sup_\alpha\pi A_\alpha=\pi(\sup_\alpha A_\alpha)$ whenever A_α is a bounded increasing net in $\mathcal{M}$. Any normal $*$-homomorphism is continuous in the strong and weak operator topologies and is a product of an amplification, a spatial $*$-isomorphism, and an induction. Its kernel is of the form $E\mathcal{M}$, where E is a projection operator belonging to the **center** $\mathcal{Z}=\mathcal{M}\cap\mathcal{M}'$ of $\mathcal{M}$. This gives a complete description of all possible **normal representations** (i.e., a normal $*$-homomorphism into some $\mathcal{B}$) of a von Neumann algebra.

D. States, Weights, and Traces

A **state** φ of a C^*-algebra $\mathcal{A}$ is a complex-valued function on $\mathcal{A}$ that is (1) complex linear: $\varphi(A+B)=\varphi(A)+\varphi(B)$, $\varphi(cA)=c\varphi(A)$ for A, $B\in\mathcal{A}$, $c\in\mathbf{C}$, (2) **positive**: $\varphi(A^*A)\geqslant 0$ for $A\in\mathcal{A}$, and (3) normalized: $\|\varphi\|=1$ (equivalent to $\varphi(I)=1$ if $I\in\mathcal{A}$). For any positive linear functional φ on $\mathcal{A}$, there exists a triplet $(\mathfrak{H}_\varphi,\pi_\varphi,\xi_\varphi)$ (unique up to the unitary equivalence) of a Hilbert space $\mathfrak{H}_\varphi$, a representation π_φ (i.e., a $*$-homomorphism into $\mathcal{B}(\mathfrak{H}_\varphi)$) of $\mathcal{A}$, and a vector ξ_φ in $\mathfrak{H}_\varphi$ such that $\varphi(A)=(\pi_\varphi(A)\xi_\varphi,\xi_\varphi)$ and $\mathfrak{H}_\varphi$ is the closure of $\pi_\varphi(\mathcal{A})\xi_\varphi$. The space $\mathfrak{H}_\varphi$ is constructed by defining the inner product $(\eta(A),\eta(B))=\varphi(B^*A)$ in the quotient of $\mathcal{A}$ by its left ideal $\{A\,|\,\varphi(A^*A)=0\}$, where η is the quotient mapping, and by completion. Then π_φ is defined by $\pi_\varphi(A)\eta(B)=\eta(AB)$. This is called the **GNS construction** after its originators I. M. Gel'fand, M. A. Naĭmark, and I. E. Siegel.

A **weight** φ on a von Neumann algebra $\mathcal{M}$ is a function defined on the positive elements of $\mathcal{M}$, with positive real or infinite values, which is additive and homogeneous ($\varphi(A+B)=\varphi(A)+\varphi(B)$ and $\varphi(\lambda A)=\lambda\varphi(A)$ for all A, $B\in\mathcal{M}$ and $\lambda\geqslant 0$ with the convention $0\cdot\infty=0$ and $\infty+a=\infty$). It is said to be **faithful** if it does not vanish except for $\varphi(0)=0$, **normal** if $\varphi(A)=\sup\varphi(A_\alpha)$ whenever A_α is an increasing net of positive elements of $\mathcal{M}$ and $A=\sup A_\alpha$, and **semifinite** if the left ideal $\mathfrak{N}_\varphi$, consisting of all elements $A\in\mathcal{M}$ for which $\varphi(A^*A)$ is finite, has the property that the linear span $\mathfrak{M}_\varphi$ of $\mathfrak{N}_\varphi^*\mathfrak{N}_\varphi$ is dense in $\mathcal{M}$. The restriction of φ to positive elements of $\mathfrak{M}_\varphi$ has a unique extension to a linear functional on $\mathfrak{M}_\varphi$, which we denote by the same letter φ. Canonically associated with a normal semifinite weight φ, there exists a Hilbert space $\mathfrak{H}_\varphi$, a normal $*$-representation π_φ of $\mathcal{M}$, and a complex linear mapping η from $\mathfrak{N}_\varphi$ into a dense subset of $\mathfrak{H}_\varphi$ such that $(\eta(B'),\eta(B))=\varphi(B^*B)$, and $\pi_\varphi(A)\eta(B)=\eta(AB)$, where B, $B'\in\mathfrak{N}_\varphi$ and $A\in\mathcal{M}$. If φ is **finite** (i.e., $\mathfrak{N}_\varphi=\mathcal{M}$), then its extension to $\mathcal{M}$ is a positive linear functional for which the triplet $(\mathfrak{H}_\varphi,\pi_\varphi,\eta(1))$ is given by the GNS construction.

The linear span of all normal states of a von Neumann algebra $\mathcal{M}$ is a norm-closed subspace of its dual $\mathcal{M}^*$, called its **predual** and denoted by $\mathcal{M}_*$, because $\mathcal{M}$ turns out to be the dual of $\mathcal{M}_*$.

A **trace** t on a von Neumann algebra $\mathcal{M}$ is a weight satisfying $t(UAU^*)=t(A)$ for U unitary

in $\mathcal{M}$ and for all positive A in $\mathcal{M}$ (equivalently, $t(A^*A)=t(AA^*)$ for all $A\in\mathcal{M}$).

E. The von Neumann Classification

A von Neumann algebra for which a semifinite normal trace does not exist is called a **purely infinite** von Neumann algebra or von Neumann algebra of **type III**. In contrast to this, a von Neumann algebra $\mathcal{M}$ is called **semifinite** (resp. **finite**) if for each positive Hermitian operator A ($\neq 0$) in $\mathcal{M}$ there is a semifinite (resp. finite) normal trace t such that $t(A)\neq 0$. Every Abelian von Neumann algebra is finite. If there are no central projection operators $E\neq 0$ such that $\mathcal{M}_E$ is finite, $\mathcal{M}$ is called **properly infinite**. Purely infinite $\mathcal{M}$ and $\mathcal{B}(\mathfrak{H})$ for infinite $\mathfrak{H}$, for example, are properly infinite. A nonzero projection operator E in $\mathcal{M}$ is called **Abelian** when $\mathcal{M}_E$ is Abelian. We call $\mathcal{M}$ a von Neumann algebra of **type I** (or **discrete**) when it contains an Abelian projection E for which I is the only central projection P covering E (i.e., $E\leqslant P$). A von Neumann algebra is of **type II** if it is semifinite and contains no Abelian projection. A von Neumann algebra of type II is called of **type II$_1$** if it is finite and of **type II$_\infty$** if it is properly infinite. A finite von Neumann algebra is also characterized as a von Neumann algebra in which the operation of taking the adjoint is continuous with respect to the strong operator topology on bounded spheres (Sakai, *Proc. Japan Acad.*, 1957). A properly infinite von Neumann algebra $\mathcal{M}$ is characterized by the property that $\mathcal{M}$ and $\mathcal{M}\otimes\mathcal{B}(\mathfrak{H})$ for any separable $\mathfrak{H}$ is $*$-isomorphic.

Given a von Neumann algebra $\mathcal{M}$, there exist mutually orthogonal projections E_{I}, E_{II}, E_{III} in the center $\mathcal{Z}$ of $\mathcal{M}$ such that $E_{\mathrm{I}}+E_{\mathrm{II}}+E_{\mathrm{III}}=I$, and $\mathcal{M}_{E_{\mathrm{I}}}$, $\mathcal{M}_{E_{\mathrm{II}}}$, $\mathcal{M}_{E_{\mathrm{III}}}$ are von Neumann algebras of type I, type II, type III, respectively. This decomposition is unique. There also exist unique central projections E_f and E_i such that $\mathcal{M}_{E_f}$ is finite, $\mathcal{M}_{E_i}$ is properly infinite, and $E_f+E_i=1$. The two decompositions can be combined. (If some of the projections E are 0, the condition on the corresponding $\mathcal{M}_E$ is to be waived.)

F. Factors

A von Neumann algebra whose center consists exclusively of scalar multiples of the identity operator is called a **factor**. Von Neumann's reduction theory (→ Section G) reduces the study of arbitrary von Neumann algebras on a separable Hilbert space more or less to the study of factors. Factors are classified into **types I$_n$** ($n=\infty,1,2,\dots$), **II$_1$** (i.e., of type II and finite), **II$_\infty$** (i.e., of type II and not finite), and III. A factor of type I$_n$ is isomorphic to the algebra $\mathcal{B}(\mathfrak{H})$ of all bounded operators on an n-dimensional Hilbert space $\mathfrak{H}$. Since the discovery of two nonisomorphic examples of factors of type II$_1$ by F. J. Murray and von Neumann (1943), classification of factors has been a central problem in the theory of von Neumann algebras. After 1967, great progress was made in the investigation of isomorphism classes of factors, and we have uncountably many nonisomorphic examples of factors of types II$_1$, II$_\infty$, and III. After the discovery of the third to ninth nonisomorphic examples of factors of type II$_1$ by J. Schwartz (1963) (the third example, 1963), W. Ching (the fourth), Sakai (the fifth), J. Dixmier and E. C. Lance (the sixth and seventh), and G. Zeller-Meier (the eighth and ninth), D. McDuff showed that there exist countably many nonisomorphic examples of factors of type II$_1$, and finally McDuff (*Ann. Math.*, 1969) and Sakai (*J. Functional Anal.*, 1970) showed the existence of uncountably many nonisomorphic examples of factors of type II$_1$. For type III factors → Section I.

G. The Integral Direct Sum and Decomposition into Factors

The Hilbert spaces considered in this section are all †separable. Let $(\mathfrak{M},\mathcal{E},\mu)$ be a †measure space; with each $\zeta\in\mathfrak{M}$ we associate a Hilbert space $\mathfrak{H}(\zeta)$. We consider functions $x(\zeta)$ on $\mathfrak{M}$ whose values are in $\mathfrak{H}(\zeta)$ for each ζ. Let $\mathbf{K}$ be a set of these functions having the following properties: (i) $\|x(\zeta)\|$ is measurable for $x(\zeta)\in\mathbf{K}$; (ii) if for a function $y(\zeta)$, the numerical function $(x(\zeta),y(\zeta))$ is measurable any $x(\zeta)\in\mathbf{K}$, then $y(\zeta)\in\mathbf{K}$; (iii) there is a countable family $\{x_1(\zeta),x_2(\zeta),\dots\}$ of functions in $\mathbf{K}$ such that for each fixed $\zeta\in\mathfrak{M}$, the set $\{x_1(\zeta),x_2(\zeta),\dots\}$ is dense in $\mathfrak{H}(\zeta)$. Then $\mathbf{K}$ is a linear space. We call each function in $\mathbf{K}$ a **measurable vector function**. We introduce in the set of measurable vector functions $x(\zeta)$ with

$$\int\|x(\zeta)\|^2\,d\mu(\zeta)<\infty$$

an equivalence relation by defining $x(\zeta)$ and $y(\zeta)$ as equivalent when

$$\int\|x(\zeta)-y(\zeta)\|^2\,d\mu(\zeta)=0.$$

Thus we obtain a space of equivalence classes which we denote by $\mathfrak{H}$. $\mathfrak{H}$ is a Hilbert space with the inner product

$$(x,y)=\int(x(\zeta),y(\zeta))\,d\mu(\zeta),$$

which is called the **integral direct sum** (or **direct integral**) of $\mathfrak{H}(\zeta)$. An operator function $A(\zeta)$ that associates with each $\zeta \in \mathfrak{M}$ a bounded linear operator $A(\zeta)$ on $\mathfrak{H}(\zeta)$ is called **measurable** if for any measurable $x(\zeta)$, $A(\zeta)x(\zeta)$ is also measurable. If, moreover, $\|A(\zeta)\|$ is bounded, $A(\zeta)$ transforms a function in $\mathfrak{H}$ to a function in $\mathfrak{H}$ and thus defines a bounded linear operator on $\mathfrak{H}$. This operator is called the **integral direct sum** (or **direct integral**) of $A(\zeta)$, and an operator on $\mathfrak{H}$ that can be reduced to this form is called **decomposable**.

Generally, let $\mathfrak{H}$ be a Hilbert space, and consider an Abelian von Neumann algebra $\mathscr{A}$ on $\mathfrak{H}$. Then we construct a measure space $(\mathfrak{M}, \mathscr{E}, \mu)$ and represent $\mathscr{A}$ as the set of bounded measurable functions on $\mathfrak{M}$. (This is possible in different ways. The Gel'fand representation is an example.) Then a Hilbert space $\mathfrak{H}(\zeta)$ can be constructed so that $\mathfrak{H}$ is represented as the integral direct sum of $\mathfrak{H}(\zeta)$. Operators in $\mathscr{A}$ are all decomposable and are called **diagonalizable**. A von Neumann algebra $\mathscr{M}$ on $\mathfrak{H}$ whose elements are all decomposable is characterized by $\mathscr{M} \subset \mathscr{A}'$. The $A(\zeta)$ yielded by the decomposition of operators A in $\mathscr{M}$ generate a von Neumann algebra $\mathscr{M}(\zeta)$ on $\mathfrak{H}(\zeta)$. If we take as $\mathscr{A}$ the center $\mathscr{Z}$ of $\mathscr{M}$, then almost all the $\mathscr{M}(\zeta)$ are factors (**von Neumann's reduction theory**), and if we take as $\mathscr{A}$ a maximal Abelian von Neumann algebra contained in $\mathscr{M}'$, then almost all the $\mathscr{M}(\zeta)$ are type I factors (F. I. Mautner, *Ann. Math.*, 1950).

H. Tomita-Takesaki Theory

Motivated by the problem of proving the commutant theorem for tensor products (i.e., $(\mathscr{M}_1 \otimes \mathscr{M}_2)' = \mathscr{M}_1' \otimes \mathscr{M}_2'$), which remained unsolved for algebras of type III up until that time, Tomita succeeded in 1967, after years of effort, in generalizing the theory of Hilbert algebras, previously developed only for semifinite von Neumann algebras. The most important ingredient of this theory lies in certain one-parameter groups of $*$-automorphisms of a von Neumann algebra, called modular automorphisms (see below), each one-parameter group of modular automorphisms being intrinsically associated with a faithful semifinite normal weight of the algebra. Tomita's theory was perfected by Takesaki [13], who also showed that modular automorphisms satisfy (and are characterized by) a condition originally introduced in statistical mechanics by the physicists R. Kubo, P. C. Martin, and J. Schwinger and accordingly known as the KMS condition. In the mathematical foundations of statistical mechanics, this condition characterizes equilibrium states of a physical system for a given one-parameter group of automorphisms (of a C^*-algebra) describing the time-development of the system. It was a fortunate coincidence that this condition was formulated in a so-called C^*-algebra approach to statistical mechanics by R. Haag, N. M. Hugenholtz, and M. Winnink [14] at about the same time that Tomita's work appeared in 1967. The original proofs of the Tomita-Takesaki theory have been simplified considerably by the work of M. Rieffel [16] and A. Van Daele [15]. Deeper insight into the significance of modular automorphisms is also provided by the work of A. Connes [19], showing that the group of modular automorphisms (up to inner automorphisms) is intrinsic to the von Neumann algebra (i.e., independent of the weight) and belongs to the center of Out $\mathscr{M}$ (the group Aut $\mathscr{M}$ of all $*$-automorphisms of the von Neumann algebra $\mathscr{M}$ modulo the subgroup Int $\mathscr{M}$ of all inner $*$-automorphisms).

Some of the basic definitions and results of the **Tomita-Takesaki theory** are as follows. If φ is a normal semifinite faithful weight on $\mathscr{M}$, the antilinear operator S_φ, defined on a dense subset $\eta(\mathfrak{N}_\varphi \cap \mathfrak{N}_\varphi^*)$ of $\mathfrak{H}_\varphi$ (→ Section D) by the relation $S_\varphi \eta(A) = \eta(A^*)$, is †closable and the polar decomposition $\overline{S_\varphi} = J_\varphi \Delta_\varphi^{1/2}$ of its closure defines a positive self-adjoint operator Δ_φ, called a **modular operator**, and an antiunitary involution J_φ. The principal results of the Tomita-Takesaki theory are (1) if $x \in \mathscr{M}$, then $\sigma_t^\varphi(A) \equiv \Delta_\varphi^{it} A \Delta_\varphi^{-it} \in \mathscr{M}$ for all real t, and this defines a continuous one-parameter group of $*$-automorphisms σ_t^φ of $\mathscr{M}$, called **modular automorphisms**, and (2) if $A \in \mathscr{M}$, then $j_\varphi(A) \equiv J_\varphi A J_\varphi \in \mathscr{M}'$, and j_φ is a conjugate-linear isomorphism of $\mathscr{M}$ onto $\mathscr{M}'$. A weight φ on $\mathscr{M}$ is said to satisfy the **KMS condition** at β (a real number) relative to a one-parameter group of $*$-automorphisms σ_t of $\mathscr{M}$ if, for every pair $A, B \in \mathfrak{N}_\varphi \cap \mathfrak{N}_\varphi^*$, there exists a bounded continuous function $F(z)$ (depending on A, B), on $0 \leqslant \operatorname{Im} z \leqslant \beta$ holomorphic in $0 < \operatorname{Im} z < \beta$ and such that $F(t) = \varphi(A\sigma_t(B))$ and $F(t + i\beta) = \varphi(\sigma_t(B)A)$. A given one-parameter group σ_t coincides with a group of modular automorphisms σ_t^φ if and only if φ satisfies the KMS condition at $\beta = -1$ relative to σ_t. (In statistical mechanics, $\beta = (kT)^{-1}$, where k is the Boltzmann constant and T the absolute temperature of the system.)

I. Structure and Classification of Factors of Type III

At the Baton Rouge Conference in 1967, R. T. Powers reported his results [17] on non-isomorphism of the one-parameter family

of factors of type III (now called **Powers's factors**), which had been constructed by von Neumann in 1938 in terms of an infinite tensor product of factors of type I (abbreviated as **ITPFI**). Prior to Powers's work only three different kinds of factors of type III, along with the same number of factors of type II_1, had been distinguished. A systematic classification of ITPFIs was subsequently given by H. Araki and E. J. Woods [18] in terms of two invariants, i.e., the **asymptotic ratio set** $r_\infty(\mathscr{M})$ and the **ρ-set** $\rho(\mathscr{M})$ of the von Neumann algebra $\mathscr{M}$. Using the Tomita-Takesaki theory, Connes [19] introduced two invariants, namely the **S-set** $S(\mathscr{M})$ (the intersection of the spectra of all modular operators) and the **T-set** $T(\mathscr{M})$ (the set of all real t for which the modular automorphism σ_t^φ is inner), and, when $\mathscr{M}$ is an ITPFI, proved the equality $S(\mathscr{M})=r_\infty(\mathscr{M})$ and the relation $T(\mathscr{M})=2\pi|\log\rho(\mathscr{M})|^{-1}$. The S-set $S(\mathscr{M})$ of a factor of type III on a separable Hilbert space is either the set of all nonnegative reals (**type III_1**), the set of all integral powers of λ (where $0<\lambda<1$) and 0 (**type III_λ**), or the set $\{0,1\}$ (**type III_0**). The work of Araki and Woods shows that there exists only one ITPFI of type III_λ for each $\lambda\in(0,1)$ (the examples of Powers) as well as for $\lambda=1$, while there exist continuously many ITPFIs of type III_0. Woods [20] has shown that the classification of ITPFIs of type III_0 is not smooth.

A structural analysis of factors of type III, given independently by Connes [19], Takesaki (*Acta Math.*, 1973), and Araki (*Publ. Res. Inst. Math. Sci.*, 1973) expressed independently a certain class of factors of type III as a kind of crossed product of semifinite von Neumann algebras with their injective endomorphisms (automorphism in the case of Connes). These analyses led Takesaki [21] to the discovery of a duality theorem for crossed products of von Neumann algebras with locally compact groups of their $*$-automorphisms (→ Section J) and its application to the following **structure theorem for von Neumann algebras of type III**. The crossed product of a von Neumann algebra $\mathscr{M}$ with the group of modular automorphisms σ_t^φ is a von Neumann algebra $\mathscr{N}$ of type II_∞, with a canonical action θ_t of the dual group as a one-parameter group of $*$-automorphisms which is trace-scaling, i.e., $\tau\circ\theta_t=e^{-t}\tau$ for some faithful normal trace τ. If $\mathscr{M}$ is properly infinite, the crossed product of $\mathscr{N}$ with θ_t is isomorphic to the original von Neumann algebra $\mathscr{M}$. In particular, any von Neumann algebra $\mathscr{M}$ of type III can be written as the crossed product of a von Neumann algebra $\mathscr{N}$ of type II_∞ with a one-parameter group of trace-scaling $*$-automorphisms θ_t. The isomorphism class of $\mathscr{M}$ is determined by the isomorphism class of $\mathscr{N}$ together with the conjugacy class of θ_t modulo inner automorphisms. The restriction $\tilde{\theta}_t$ of θ_t to the center $\mathscr{Z}$ of $\mathscr{N}$ is of special importance. $\mathscr{M}$ is a factor if and only if $\tilde{\theta}_t$ is ergodic. In that case, $\mathscr{M}$ is of type III_1 if $\mathscr{N}$ is a factor, $\mathscr{M}$ is of type III_λ, $0<\lambda<1$, if $\tilde{\theta}_t$ is periodic with period $-\log\lambda$, and $\mathscr{M}$ is of type III_0 if $\tilde{\theta}_t$ is aperiodic and not isomorphic to the one-parameter group of $*$-automorphisms of $L_\infty(\mathbf{R})$ induced by the translations of the real line $\mathbf{R}$. (The excluded case does not occur for $\mathscr{M}$ of type III.)

A von Neumann algebra on a separable Hilbert space is called **approximately finite-dimensional** (or **approximately finite** or **hyperfinite**) if it is generated by an increasing sequence of finite-dimensional $*$-subalgebras. This class of von Neumann algebras includes many important examples, such as ITPFI and the von Neumann algebra generated by any representation of canonical commutation (or anticommutation) relations on a separable Hilbert space. The classification of approximately finite-dimensional factors is almost complete. In fact the uniqueness of an approximately finite-dimensional factor of type II_1 has been known since the work of von Neumann. It is called the **hyperfinite factor**. The uniqueness of approximately finite-dimensional factors of type II_∞ (which is then the tensor product of the hyperfinite factor with $\mathscr{B}(\mathfrak{H})$) and of type III_λ, $0<\lambda<1$, (which are then Powers's factors) has been demonstrated by Connes [22]. Approximately finite-dimensional factors of type III_0 are classified exactly by the isomorphism classes of the ergodic groups $\tilde{\theta}_t$ of $*$-automorphisms of commutative von Neumann algebras $\mathscr{Z}$. Any such factor is a **Krieger's factor**, i.e., a crossed product of a commutative von Neumann algebra with a single $*$-automorphism. Examples of such factors have been extensively studied by Krieger, who has also shown [23] that isomorphism of a Krieger's factor is equivalent to weak equivalence of the associated nonsingular transformation of the standard measure space.

A von Neumann algebra on a separable Hilbert space is approximately finite-dimensional if and only if it is injective (→ 36 Banach Algebras H).

J. Crossed Products

The **crossed product** $\mathscr{M}\otimes_\alpha G$ of a von Neumann algebra $\mathscr{M}$ (acting on a Hilbert space $\mathfrak{H}$) and a locally compact Abelian group G relative to a continuous action α of G on $\mathscr{M}$ (by $*$-automorphisms α_g, $g\in G$) is the von Neumann algebra $\mathscr{N}$ generated by the operators $\pi(A)$, $A\in\mathscr{M}$ and $\lambda(h)$, $h\in G$, defined on the Hilbert space $L_2(G,\mathfrak{H})$ of all $\mathfrak{H}$-valued L_2-

functions on G (relative to the Haar measure) by

$$[\pi(A)\xi](g)=\alpha_g^{-1}(A)\xi(g),$$

$$[\lambda(h)\xi](g)=\xi(h^{-1}g),$$

where $\xi\in L_2(G,\mathfrak{H})$. The canonical action $\hat{\alpha}$ of the dual $\hat{G}$ on $\mathcal{N}$ is defined by $\hat{\alpha}_p(B)=\mu(p)B\mu(p)^*$ for $B\in\mathcal{N}$ and $p\in\hat{G}$, where $\mu(p)$ is defined by $[\mu(p)\xi](g)=\overline{\langle g,p\rangle}\xi(g)$. The **duality theorem of Takesaki** [21] asserts that $[\mathcal{M}\otimes_\alpha G]\otimes_{\hat{\alpha}}\hat{G}$ is isomorphic to $\mathcal{M}\otimes\mathcal{B}(L_2(G))$, where the second factor $\mathcal{B}(L_2(G))$ is the algebra of all bounded linear operators on $L_2(G)$.

K. Natural Positive Cone

The closure V^α of the set of vectors $\Delta_\varphi^\alpha\eta(A)$ for all positive A in $\mathfrak{N}_\varphi\cap\mathfrak{N}_\varphi^*$ reflects certain properties of the von Neumann algebra $\mathcal{M}$ for $0\leqslant\alpha\leqslant 1/2$ [24, 25]. In particular, $V^{1/4}$ is called the **natural positive cone**. It is a self-dual closed convex cone, and is intrinsic to the von Neumann algebra $\mathcal{M}$ (i.e., independent of the weight φ). Every normal positive linear functional ψ on $\mathcal{M}$ has a unique representative $\xi(\psi)$ in this cone (i.e., $\psi(A)=(\pi_\varphi(A)\xi(\psi),\xi(\psi))$), and the mapping ξ is a concave, monotone, bijective homeomorphism, homogeneous of degree 1/2. The group of all $*$-automorphisms of $\mathcal{M}$ has a natural unitary representation $U(g)$, $g\in\operatorname{Aut}\mathcal{M}$, satisfying the relations $U(g)AU(g)^*=g(A)$, $U(g)\xi(\varphi)=\xi(\varphi\circ g^{-1})$.

L. C^*-Algebras and von Neumann Algebras

Let a C^*-algebra $\mathcal{A}$ be given. A $*$-representation $x\rightarrow T_x$ gives rise to a von Neumann algebra $\mathcal{M}$, generated by T_x, $x\in\mathcal{A}$. The type of this representation is defined according to the type of $\mathcal{M}$. A C^*-algebra is called a **C^*-algebra of type I** if its $*$-representations are always of type I. It is known that this class is exactly the class of GCR algebras (→ 36 Banach Algebras H). It is also known that a separable non–type I C^*-algebra has a representation of type II and a general non–type I C^*-algebra has a representation of type III (J. Glimm, *Ann. Math.*, 1961; Sakai [8]).

For a C^*-algebra $\mathcal{A}$, all its representations generate †injective von Neumann algebras if and only if $\mathcal{A}$ is †nuclear [26, 27].

M. Topological Groups and von Neumann Algebras

Consider a unitary representation $g\rightarrow U_g$ of a locally compact Hausdorff group G (→ 423 Topological Groups). If this representation is a †factor representation, the type of this representation is defined according to the type of the von Neumann algebra $\mathcal{M}$ generated by U_g, $g\in G$. A †group of type I is a group whose factor representations are all of type I. For example, connected semisimple Lie groups and connected nilpotent Lie groups are of type I (Harish-Chandra, *Trans. Amer. Math. Soc.*, 1953; J. M. G. Fell, *Proc. Amer. Math. Soc.*, 1962). Examples of groups that are not of type I are known (→ 437 Unitary Representation E).

References

[1] J. von Neumann, Collected works II, III, Pergamon, 1961.
[2] W. Arveson, An invitation to C^*-algebras, Springer, 1976.
[3] O. Bratteli and D. W. Robinson, Operator algebras and quantum statistical mechanics, Springer, 1979.
[4] J. Dixmier, Von Neumann algebras, North-Holland, 1981.
[5] J. Dixmier, Les C^*-algèbres et leurs représentations, Gauthier-Villars, 1964.
[6] I. Kaplansky, Rings of operators, Benjamin, 1968.
[7] G. K. Pedersen, C^*-algebras and their automorphism groups, Academic Press, 1979.
[8] S. Sakai, C^*-algebras and W^*-algebras, Springer, 1971.
[9] J. T. Schwartz, W^*-algebras, Gordon & Breach, 1967.
[10] S. Stratila and L. Zsido, Lectures on von Neumann algebras, Editura Academiei/Abacus Press, 1975.
[11] M. Takesaki, Theory of operator algebras I, Springer, 1979.
[12] Y. Nakagami and M. Takesaki, Duality for crossed products of von Neumann algebras, Lecture notes in math. 731, Springer, 1979.
[13] M. Takesaki, Tomita's theory of modular Hilbert algebras and its applications, Lecture notes in math. 128, Springer, 1970.
[14] R. Haag, N. M. Hugenholtz, and M. Winnink, On the equilibrium states in quantum statistical mechanics, Comm. Math. Phys., 5 (1967), 215–236.
[15] A. Van Daele, A new approach to the Tomita-Takesaki theory of generalized Hilbert algebras, J. Functional Anal., 15 (1974), 378–393.
[16] M. Rieffel and A. van Daele, A bounded approach to Tomita-Takesaki theory, Pacific J. Math., 69 (1977), 187–221.
[17] R. T. Powers, Representations of uniformly hyperfinite algebras and their asso-

ciated von Neumann rings, Ann. Math., (2) 86 (1967), 138–171.
[18] H. Araki and E. J. Woods, A classification of factors, Publ. Res. Inst. Math. Sci., (A) 4 (1968), 51–130.
[19] A. Connes, Une classification des facteurs de type III, Ann. Sci. Ecole Norm. Sup., (4) 6 (1973), 133–252.
[20] E. J. Woods, The classification of factors is not smooth, Canad. J. Math., 25 (1973), 96–102.
[21] M. Takesaki, Duality for crossed products and the structure of von Neumann algebras of type III, Acta Math., 131 (1973), 249–310.
[22] A. Connes, Classification of injective factors, Ann. Math., 104 (1976), 73–115.
[23] W. Krieger, On ergodic flows and the isomorphism of factors, Math. Ann., 223 (1976), 19–70.
[24] H. Araki, Some properties of modular conjugation operators of von Neumann algebras and a non-commutative Radon-Nikodym theorem with a chain rule. Pacific J. Math., 50 (1974), 309–354.
[25] A. Connes, Caractérisation des espaces vectoriels ordonnés sous-jacents aux algèbres de von Neumann, Ann. Inst. Fourier, 24, Fasc. 4 (1974), 127–156.
[26] M. Choi and E. Effros, Separable nuclear C^*-algebras and injectivity, Duke Math. J., 43 (1976), 309–322.
[27] M. Choi and E. Effros, Nuclear C^*-algebras and injectivity: The general case, Indiana Univ. Math. J., 26 (1977), 443–446.
[28] R. V. Kadison and J. R. Ringrose, Fundamentals of the theory of operators I, Academic Press, 1983.
[29] S. Stratila, Modular theory in operator algebras, Editura Academiei/Abacus Press, 1981.

309 (XX.7) Orbit Determination

A. General Remarks

The purposes of the theory of **orbit determination** are (1) to study properties of orbits of celestial bodies, (2) to determine orbital elements from observed positions of the celestial bodies, and (3) to compute their predicted positions utilizing the orbital elements. Celestial bodies to which the theory is applied are mainly planets, asteroids, comets, satellites, and artificial satellites in the solar system, although orbits of meteors and visual, photometric, and eclipsing binaries can be determined by similar methods.

B. Kepler's Orbital Elements

Consider, for example, an asteroid moving on an ellipse with one focus at the sun. The elliptic orbit is fixed by the initial conditions of the motion or the integration constants of the †Hamilton-Jacobi equation (→ 55 Celestial Mechanics) and is described by **Kepler's orbital elements** a, e, ω, i, Ω, and t_0 (Fig. 1).

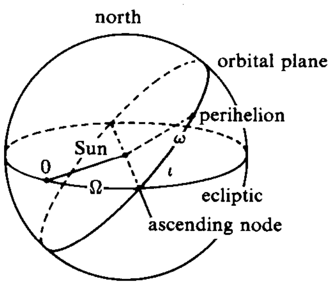


Fig. 1
Orbital elements

The size and shape of the ellipse are determined by the semimajor axis (half the †major axis) a and the †eccentricity e, while the argument ω of perihelion, measured from the ascending node to the perihelion, shows the direction of the major axis. (Sometimes, we adopt as one of the main parameters the **perihelion distance** $q=a(1-e)$ instead of the semimajor axis a.) The position of the orbital plane is determined by the inclination angle i to the ecliptic and the longitude Ω of the ascending node, and then the position of the asteroid on the orbit is determined by the time t_0 of the perihelion passage. The period T of one revolution, or **mean motion** $n=2\pi/T$, which is the mean angular velocity, is computed by Kepler's third law $n^2a^3=\mu$, with μ a constant depending on the mass of the asteroid. The mean motion is a fundamental frequency in the solution of the Hamilton-Jacobi equation and is obtained by differentiating the energy constant $-\mu/2a$ with respect to an action variable $\sqrt{\mu a}$.

To express the position of the asteroid on the ellipse as a function of time, we use the **true anomaly** v, which is the angular distance of the asteroid from the perihelion, the **eccentric anomaly** E, and the **mean anomaly** $M=n(t-t_0)$. Of these three anomalies the mean anomaly can be derived directly from Kepler's elements, although it must be transformed to the true anomaly or to the eccentric anomaly when we compute the coordinates of the aster-

oid. **Kepler's equation**

$$E - e\sin E = M \tag{1}$$

holds between E and M. Solving this equation, we obtain an expression for E as a function of M:

$$E = M + \sum_{n=1}^{\infty} \frac{2}{n} J_n(ne)\sin nM, \tag{2}$$

where J_n is the †Bessel function of order n. However, in practical computations, we often solve equation (1) directly by numerical methods or by using tables.

C. Orbit Determination

An astrometric observation of a celestial body usually consists of measurements of two coordinates (right ascension and declination) on the celestial sphere. Therefore, to derive six orbital elements, three sets of observations should be made at three moments separated by appropriate time intervals. If the topocentric distance of the celestial body is known, the orbital elements can be computed directly from observations. However, since the distance is not usually known, special methods have had to be developed. A method for orbit determination was worked out by C. F. †Gauss at the beginning of the 19th century to find the orbit of Ceres, the first asteroid to be discovered. Although the topocentric distances are not known, we know that orbits of asteroids are planar, and Kepler's second law, the law of conservation of areal velocity, holds approximately. Therefore we can assume that the area of the triangle made by the sun and the two positions of the asteroid observed at different moments is proportional to the corresponding time interval. Using this property of the orbit we can derive the topocentric distance and then the orbital elements. This method is called the indirect method, and similar methods can be developed for parabolic and hyperbolic orbits.

D. Osculating Elements and Orbit Improvement

For the †two-body problem the orbit is a fixed and invariable ellipse, and therefore Kepler's orbital elements are constants. On the other hand, when gravitational interactions from other bodies cannot be disregarded, the orbital elements are found to be variable by computing the †perturbations by the †method of variation of constants. The perturbations are expressed as sums of periodic, †secular, and long-periodic terms.

Because of the perturbations, the orbit deviates from the fixed ellipse, although at every moment the instantaneous velocity and position of the asteroid determines an ellipse. The orbital elements of the ellipse thus defined at each moment, called **osculating elements**, are variable with time. To compute perturbations that cause this change of osculating elements, it is necessary to observe the initial conditions of motion, i.e., the osculating elements at the initial moment. During a time interval shorter than the period of one revolution, the variations of the osculating elements are usually very small. Therefore, by three sets of observations made at three moments at short intervals, it is possible to determine the orbital elements that can be identified with the osculating elements observed at the mean moment. However, if the intervals are very short, errors in the determined values often are very large, and it becomes necessary to carry out observations at distant moments also. When such additional observations are made, those data are compared with the respective values that follow from the initial observations, and the perturbations computed from them; then the method of least squares is used to improve the estimation of the orbital elements.

E. Artificial Satellites

Since the periods of revolution of asteroids are of the order of a few years, the osculating elements change very little in a few weeks. On the other hand, for artificial satellites moving around the earth, the periodic as well as secular perturbations become very large after a few hours because the period of revolution may be as short as two hours. Therefore, to determine orbital elements for artificial satellites, observed positions should be corrected by subtracting the effects of periodic perturbations computed from approximate orbital elements already known. By using the observations thus corrected, mean orbital elements are derived by the method of orbit improvement. The approximate orbital elements can be computed if the launching conditions of the satellites are known. In this manner, mean orbital elements can be derived every day, and variations of the mean orbital elements, or amounts of secular perturbations, for a certain period (say, for 100 days) are found. From them, information on atmospheric density and the gravitational potential of the earth are derived. It should be remarked that for artificial satellites distance measurements have been made by radar, and velocity determinations have been made by measuring the Doppler effect.

For satellites of other planets, measurements of two coordinates with respect to the centers of the planets are made. Masses of planets can be computed by Kepler's third law when the orbital elements of satellites are known, and gravitational potentials of the planets can be determined from their secular perturbations.

F. Binaries

In the study of visual binaries, methods similar to those for satellites can be applied, although the exact estimation of the distances to binaries is often impossible. For photometric binaries radial components of velocities are derived by measuring the Doppler effect; and for eclipsing binaries important information, such as their masses, densities, and sizes, as well as data regarding their internal constitutions, can be derived from the observed orbital elements.

References

[1] G. Stracke, Bahnbestimmung der Planeten und Kometen, Springer, 1929.
[2] A. D. Dubyago, The determination of orbits, Macmillan, 1961.
[3] C. F. Gauss, Theoria motus, Dover, 1963.
[4] P. R. Escobal, Methods of orbit determination, Wiley, 1965.

310 (XII.4) Ordered Linear Spaces

A. History

Many spaces used in functional analysis, such as †Hilbert spaces, †Banach spaces, and †topological linear spaces, are generalizations of Euclidean spaces, where the leading idea has been to generalize the distance in Euclidean spaces in various ways. On the other hand, generalizing the order concept for real numbers has led to spaces of another kind: ordered linear spaces and vector lattices. The theory of vector lattices was presented in a lecture by F. Riesz at the International Congress of Mathematicians in 1928 [1] and has been developed by many authors. Among them we cite H. Freudenthal, L. V. Kantorovitch (*Mat. Sb.*, 2 (44) (1937)), Riesz (*Ann. Math.*, (2) 41 (1940)), S. Kakutani, F. Bohnenblust, G. Birkhoff [2], H. Nakano [3], B. C. Vulikh [5], and H. H. Schaefer [8]. Vector lattices have been used in lattice-theoretic treatments of integration (→ Section I), †spectral resolution, and †ergodic theory. The central notion is Banach lattices (→ Section F), but the theory has been extended to the case where E is a †locally convex topological linear space with the structure of a vector lattice [6–8].

B. Definitions

A real †linear space E is said to be an **ordered linear space** if E is supplied with an †order relation $\geqslant$ with the following two properties: (i) $x \geqslant y \Rightarrow x+z \geqslant y+z$; (ii) $x \geqslant y$ and $\lambda \geqslant 0$ (λ is a real number) $\Rightarrow \lambda x \geqslant \lambda y$.

If, in addition, E forms a †lattice under this order $\geqslant$, we call E a **vector lattice** (**Riesz space** or **lattice ordered linear space**).

For Sections B through E, we assume E to be a vector lattice. For any $x, y \in E$, the †join and †meet of x, y are denoted by $x \vee y$ and $x \wedge y$ respectively. The following relations are obvious:

$$(x+z)\vee(y+z)=(x\vee y)+z,$$

$$(x+z)\wedge(y+z)=(x\wedge y)+z,$$

$$\lambda x\vee\lambda y=\lambda(x\vee y),\quad \lambda x\wedge\lambda y=\lambda(x\wedge y)\quad (\lambda\geqslant 0),$$

$$\lambda x\vee\lambda y=\lambda(x\wedge y),\quad \lambda x\wedge\lambda y=\lambda(x\vee y)\quad (\lambda\leqslant 0)$$

and

$$(x\vee y)\wedge z=(x\wedge z)\vee(y\wedge z),$$

$$(x\wedge y)\vee z=(x\vee z)\wedge(y\vee z).$$

The last relation means that E is a †distributive lattice.

For $x \in E$, the elements $x \vee 0$, $(-x)\vee 0$, and $x \vee(-x)$ are called respectively the **positive part**, **negative part**, and **absolute value** of the element x, and are denoted respectively by x^+, x^-, and $|x|$. The following identities hold: $x = x^+ - x^-$ (**Jordan decomposition**), $|x| = x^+ + x^-$, $x^+ \wedge x^- = 0$, $x \vee y + x \wedge y = x + y$, $|\lambda x| = |\lambda||x|$, and $|x-y| = x \vee y - x \wedge y$.

For $a, b \in E$ with $a \leqslant b$, the set $\{x \mid a \leqslant x \leqslant b\}$ is called an **interval** and is denoted by $[a, b]$. A subset of E is called **(order) bounded** if it is included in an interval. An element e of E is said to be a **unit** or an **Archimedean unit** if for any $x \in E$ there exists a natural number n such that $x \leqslant ne$. A linear subspace I of E is called an **ideal** (or **order ideal**) of E if $x \in I$ and $|y| \leqslant |x|$ imply $y \in I$.

C. Order Limits

Given a subset $\{x_\alpha\}$ of E, if an element x of E is an upper bound of $\{x_\alpha\}$ and any upper bound y of $\{x_\alpha\}$ satisfies the relation $y \geqslant x$, then it is called the **least upper bound** (or **su-**

premum) of $\{x_\alpha\}$ and is denoted by $\sup_\alpha x_\alpha$ or $\bigvee_\alpha x_\alpha$. The **greatest lower bound** (or **infimum**) of $\{x_\alpha\}$, denoted by $\inf_\alpha x_\alpha$ or $\bigwedge_\alpha x_\alpha$, is defined dually.

A sequence $\{x_n\}$ $(x_n \in E)$ is said to be **order convergent** to x if there exists a nonincreasing sequence $\{u_n\}$ $(u_n \in E)$ such that $\bigwedge_n u_n = 0$ and $|x_n - x| \leqslant u_n$. In this case x is called the **order limit** of $\{x_n\}$ and is denoted by $x = \text{o-lim}\, x_n$. For order convergent sequences $\{x_n\}$ and $\{y_n\}$, we can show the following relations: $\text{o-lim}(\lambda x_n + \mu y_n) = \lambda(\text{o-lim}\, x_n) + \mu(\text{o-lim}\, y_n)$ and $\text{o-lim}(x_n \curlyvee y_n) = (\text{o-lim}\, x_n) \curlyvee (\text{o-lim}\, y_n)$. We say that E is **Archimedean** if the relations $0 \leqslant nx \leqslant y$ $(n = 1, 2, 3, \ldots)$ imply $x = 0$. If E is Archimedean, then the relations $x = \text{o-lim}\, x_n$ and $\lambda = \lim \lambda_n$ imply $\lambda x = \text{o-lim}\, \lambda_n x_n$. We say that E is **complete** (**σ-complete**) if any (countable) subset of E that is bounded above has a least upper bound. A σ-complete vector lattice is always Archimedean. If E is σ-complete, for any sequence $\{x_n\}$ bounded above, $\text{o-lim}\sup x_n$ is defined to be $\bigwedge_n \bigvee_{m \geqslant n} x_m$; we define $\text{o-lim}\inf x_n$ similarly. With these definitions, $x = \text{o-lim}\, x_n$ is equivalent to $x = \text{o-lim}\sup x_n = \text{o-lim}\inf x_n$. Any Archimedean vector lattice can be extended to a complete vector lattice in the same way as the real numbers are constructed from the rational numbers by Dedekind cuts (→ 294 Numbers).

D. Examples of Vector Lattices

†Sequence spaces, such as c, m, and l_p, and †function spaces, such as C, M, and L_p, form vector lattices under pointwise ordering (→ 168 Function Spaces). Among these spaces c and C are not σ-complete, but the others are complete. We give two examples. First, let Σ be a σ-algebra of subsets of a space Ω, and let $A(\Omega, \Sigma)$ be the set of all finite σ-additive †set functions defined on Σ. Then $A(\Omega, \Sigma)$ is a complete vector lattice if we define $\mu_1 \geqslant \mu_2$ to mean $\mu_1(S) \geqslant \mu_2(S)$ for any $S \in \Sigma$. The second example is an ordered space consisting of all †bounded symmetric operators T on a Hilbert space H, where we define $T_1 \geqslant T_2$ to mean $(T_1 x, x) \geqslant (T_2 x, x)$ for any $x \in H$. In general, this space is not a vector lattice. However, if A is a commutative †W^*-algebra of operators on H and S is the set of †symmetric operators belonging to A, then S is a complete vector lattice under the ordering just defined. We can replace the conditions of finiteness in $A(\Omega, \Sigma)$ and boundedness in S with weaker ones and still obtain the same situation. The †Radon-Nikodym theorem in $A(\Omega, \Sigma)$ and the †spectral resolution theorem of symmetric operators in S can be extended to theorems of †spectral representations in general vector lattices.

Let E_1 be a linear space of functions defined on a set Ω ordered pointwise. If there exists a bijective mapping defined on a vector lattice E onto E_1 that is linear and order isomorphic, we call E_1 a **representation** of E. If E has an Archimedean unit and is simple (which means that E and $\{0\}$ are the only ideals of E), then E can be represented as the set of real numbers such that the Archimedean unit of E is represented by the number 1 (H. Freudenthal, *Proc. Akad. Amsterdam*, 39 (1936)).

E. Dual Spaces

Let $\mathfrak{L}(E, F)$ be the set of **order bounded** linear mappings of a vector lattice E into a vector lattice F, where order boundedness means that any bounded (in the sense of the order) subset of E is mapped into a bounded set of F. For any $\varphi_1, \varphi_2 \in \mathfrak{L}(E, F)$, define $\varphi_1 \geqslant \varphi_2$ to mean $\varphi_1(x) \geqslant \varphi_2(x) (x \geqslant 0, x \in E)$. If F is complete, then $\mathfrak{L}(E, F)$ is a complete vector lattice. An element $\varphi \in \mathfrak{L}(E, F)$ is called a **positive operator** if $\varphi \geqslant 0$. If F is the set of real numbers $\mathbf{R}$, then $\mathfrak{L}(E, F)$ is the set of all (order) bounded †linear functionals on E. This space, called the **dual lattice** of the vector lattice E and denoted by E^b, is a complete vector lattice. For $f \in E^b$ and $x \geqslant 0$, $x \in E$, we have

$$f^+(x) = \sup_{0 \leqslant y \leqslant x} f(y), \quad f^-(x) = -\inf_{0 \leqslant y \leqslant x} f(y),$$

$$|f|(x) = \sup_{|y| \leqslant x} f(y).$$

F. Banach Lattices

A linear space E is called a **normed vector lattice** if E is a vector lattice having the structure of a †normed space satisfying $|x| \leqslant |y| \Rightarrow \|x\| \leqslant \|y\|$. Furthermore if a normed vector lattice E is complete relative to the norm, we call E a **Banach lattice**. The examples in Section D are Banach lattices (for $\mu \in A(\Omega, \Sigma)$ we define $\|\mu\| = |\mu|(\Omega)$).

In Banach lattices, $\|x_n - x\| \to 0$ and $\|y_n - y\| \to 0$ imply $\|x_n \curlyvee y_n - x \curlyvee y\| \to 0$. Among relations between order convergence and norm convergence in Banach lattices, the following is one of the most fundamental: In a Banach lattice E, norm convergence of a sequence $\{x_n\}$ to x is equivalent to **relative uniform star convergence** of $\{x_n\}$ to x, i.e., for any subsequence $\{x_{n(m)}\}$ of $\{x_n\}$, there exists a subsequence $\{x_{n(m(l))}\}$ of $\{x_{n(m)}\}$ and an element y of E satisfying the relations $|x_{n(m(l))} - x| \leqslant y/l$ $(l = 1, 2, \ldots)$.

Any set bounded relative to the order is bounded relative to the norm, but the converse does not hold in general. For a linear functional, however, these two concepts of boundedness coincide, and the order dual of

E is the same as the norm dual of E. Moreover, the dual (in any sense) of a Banach lattice is also a Banach lattice.

G. Abstract M Spaces and Abstract L Spaces

For a Banach lattice E, we consider the following three conditions: (M) $x, y \geqslant 0 \Rightarrow \|x \vee y\| = \max(\|x\|, \|y\|)$. (L) $x, y \geqslant 0 \Rightarrow \|x+y\| = \|x\| + \|y\|$. ($L_p$) $x \wedge y = 0 \Rightarrow \|x+y\|^p = \|x\|^p + \|y\|^p$ $(1 < p < \infty)$. If E satisfies one of the conditions (M), (L), or (L_p), we say that E is an **abstract M space**, and **abstract L space**, or an **abstract L_p space**, written AM, AL, and AL_p, respectively. If the unit ball of an AM space has a †greatest element, it is called the **Kakutani unit** of E. The duals of AM spaces, AL spaces, and AL_p spaces are AL spaces, AM spaces with the Kakutani units, and AL_q spaces $(1/p+1/q=1)$, respectively. An AM space with a Kakutani unit is represented by $C(\Omega)$, i.e., the set of all real-valued continuous functions defined on a compact Hausdorff space Ω. The AL spaces and AL_p spaces are represented by L_1 and L_p, respectively, on a †measure space. Here the representation of a Banach lattice means a representation of a vector lattice preserving the norm (Kakutani, *Ann. Math.*, (2) 42 (1941); Bohnenblust, *Duke Math. J.*, 6 (1940)).

H. Spectral Properties of Positive Operators

The n-dimensional real vector space E_n is a vector lattice under pointwise order (→ Section D). An element x in E_n is called **strictly positive** if $x_i > 0$ for all i. A square matrix $A = (a_{i,j})$ of order n is called **positive** if $a_{i,j} \geqslant 0$ for all i and j. It corresponds to a positive operator in E_n (→ Section E). A is called **irreducible** if there exists no permutation matrix P such that $P^{-1}AP = \begin{pmatrix} A_1 & A_2 \\ 0 & A_3 \end{pmatrix}$, where A_1 and A_3 are square matrices of order n_i $(1 \leqslant n_i < n)$. We denote by $\sigma(A)$ the †spectrum of A and by $r(A)$ the †spectral radius of A, i.e., $\sup\{|\alpha| \mid \alpha \in \sigma(A)\}$. The spectral circle of A is the circle of radius $r(A)$ having the origin as its center. O. Perron (*Math. Ann.*, 64 (1907)) and G. Frobenius (*S. B. Preuss. Akad. Wiss.*, 1908 and 1912) established the following remarkable result on the spectral properties of positive matrices.

Theorem (Perron-Frobenius). Let A be a positive square matrix. Then its spectral radius $r(A)$ belongs to $\sigma(A)$, and for this spectrum $r(A)$ there exists an eigenvector $x \geqslant 0$. Assume further that A is irreducible and the order of A is greater than 1. Then $r(A) > 0$ and the eigenspace of A for $r(A)$ is a 1-dimentional subspace spanned by a strictly positive element. In this case the eigenvalues of A on the spectral circle are the kth roots of unity for some k multiplied by $r(A)$, each of which is a simple root of the eigenequation of A.

Since a positive matrix of order n corresponds to a positive operator in E_n, extensions of this theorem to positive operators in ordered linear spaces have been studied by many mathematicians. For these extensions, see the following articles: M. G. Kreĭn and M. A. Rutman (*Amer. Math. Soc. Transl.*, 26 (1950); original in Russian, 1948), F. F. Bonsall (*J. London Math. Soc.*, 30 (1955)), S. Karlin (*J. Math. Mech.*, 8 (1959)), T. Ando (*J. Fac. Sci. Hokkaido Univ.*, ser. 1, 13 (1957)), H. H. Schaefer [8], H. P. Lotz (*Math. Z.*, 108 (1968)), F. Niiro and I. Sawashima (*Sci. Pap. Coll. Gen. Educ., Univ. Tokyo*, 1966), I. Sawashima and F. Niiro (*Nat. Rep. Ochanomizu Univ.*, 30 (1979)), and S. Miyajima (*J. Fac. Sci. Univ. Tokyo*, 27 (1980)).

I. Integrals Based on Ordering

As applications of ordered linear space theory, we state the integrals of Daniel-Stone and of Banach. Let us begin with a set $\mathfrak{E}$ of real-valued functions defined on an abstract space S and assume that $\mathfrak{E}$ is a vector lattice with respect to the usual order relation, addition, and scalar multiplication. Assume further that a functional $E(f)$ defined on $\mathfrak{E}$ satisfies the following conditions: (i) additivity, i.e., $E(f+g) = E(f)+E(g)$; (ii) positivity, i.e., $f \geqslant 0$ implies $E(f) \geqslant 0$; (iii) $f, f_n \in \mathfrak{E}$ $(n=1,2,\ldots)$ and $|f| \leqslant \sum_{n=1}^{\infty} |f_n|$ imply $E(|f|) \leqslant \sum_{n=1}^{\infty} E(|f_n|)$, where $|f|$ means $f \vee (-f)$. A functional on $\mathfrak{E}$ satisfying both conditions (i) and (ii) is called a positive linear functional. A positive linear functional on $\mathfrak{E}$ satisfies M. H. Stone's condition (iii) if and only if it satisfies P. J. Daniell's condition (iii)': $f_1 \geqslant f_2 \geqslant \ldots$ and $\lim_{n\to\infty} f_n = 0$ imply $\lim_{n\to\infty} E(f_n) = 0$. Next, we define, for every function φ on S admitting $\pm\infty$ as values, a functional $N(\varphi)$ as follows:

$$N(\varphi) = \inf\left\{ \sum_{n=1}^{\infty} E(|f_n|) \,\middle|\, f_n \in \mathfrak{E}, \quad |\varphi| \leqslant \sum_{n=1}^{\infty} |f_n| \right\}.$$

Here we put $N(\varphi) = +\infty$ when for a function φ there are no functions $\{f_n\}$ such that $|\varphi| \leqslant \sum_{n=1}^{\infty} |f_n|$. A function φ is, by definition, a **null function** if $N(\varphi) = 0$ holds, and a set A is a **null set** if its †characteristic function is a null function. Since each function of $\mathfrak{F}_0 = \{\varphi \mid N(\varphi) < +\infty\}$ takes finite values except on a null set, we can define addition and the scalar multiplication for such functions except on a null set. Let $\mathfrak{F}$ be the set of equivalence classes of $\mathfrak{F}_0$

with respect to the relation $\varphi \sim \psi$ defined as $N(\varphi-\psi)=0$. Then $\mathfrak{F}$ is a Banach lattice with the norm N, and $\mathfrak{E}$ is included in $\mathfrak{F}$ (by identifying f and g of $\mathfrak{E}$ when $E(|f-g|)=0$). Let us denote now by $\mathfrak{L}$ the closure of $\mathfrak{E}$ in $\mathfrak{F}$. Then any function φ belonging to $\mathfrak{L}$ is said to be **Daniell-Stone integrable**, and $L(\varphi)=N(\varphi^+)-N(\varphi^-)(\varphi^+=\frac{1}{2}(|\varphi|+\varphi), \varphi^-=\frac{1}{2}(|\varphi|-\varphi))$ is called the **Daniell-Stone integral** of φ. The integral L thus defined is, as a functional on $\mathfrak{L}$, an extension of the functional E on $\mathfrak{E}$. For this integral, Lebesgue's convergence theorem is easily proved, and a result corresponding to Fubini's theorem has been obtained [9]. Furthermore, the concepts of measurable functions, measurable sets, and measure can be defined by using L and $\mathfrak{L}$. Also, the relation between L and the Lebesgue integral with respect to this measure is known [9]. Since $\mathfrak{L}$ is an †abstract L-space, the Daniell-Stone integral $L(\varphi)$ is represented by the Lebesgue integral of φ on a certain measure space. The Daniell-Stone integral introduced above is due to M. H. Stone [9]. Daniell (*Ann. Math.*, (2) 19 (1917–1918)) originally defined the upper integral $\bar{I}(\varphi)$ by using $E(f)$ on $\mathfrak{E}$ satisfying conditions (i), (ii), and (iii'). Also, he defined the set $\mathfrak{L}$ of Daniell-Stone integrable functions by $L=\{\varphi \mid \bar{I}(\varphi)=-\bar{I}(-\varphi)\}$. S. Banach defined an integral by using methods similar to Daniell's, replacing condition (iii') for a positive linear functional $E(f)$ on $\mathfrak{E}$ by condition (iii''): $\lim_{n\to\infty} f_n=0$, $|f_n|\leqslant g$, and $f_n, g\in\mathfrak{E}$ imply $\lim_{n\to\infty} E(f_n)=0$ [10]. Furthermore, N. Bourbaki [11] and E. J. McShane (*Proc. Nat. Acad. Sci. US*, 1946) have defined a more general integral than the Daniell-Stone with a condition analogous to (iii'), replacing the sequence in it by a †directed family of functions $\{f_\alpha\}$.

Specifically if, in the Daniell-Stone integral, S is a locally compact Hausdorff space and $\mathfrak{E}$ is the set of continuous functions with compact supports, then a functional $E(f)$ on $\mathfrak{E}$ satisfying conditions (i) and (ii) is proved to satisfy the condition (iii'), and the Daniell-Stone integral $L(\varphi)$ can be constructed from $E(f)$ [11].

Banach also defined another integral for all real-valued bounded functions on $[0,1)$ by using the †Hahn-Banach extension theorem [12]. His definition is as follows: Let $\mathfrak{F}$ be the set of all real-valued bounded functions on $[0,1)$ and $\mathfrak{A}$ be the family of all finite sets of real numbers $\alpha=(\alpha_1,\alpha_2,\ldots,\alpha_n)$. Furthermore, we define, for $x(s)\in\mathfrak{F}$ and $\alpha\in\mathfrak{A}$,

$$M(x,\alpha)=\sup_{-\infty<s<+\infty}\frac{1}{n}\sum_{k=1}^{n}x(s+\alpha_k),$$

where $x(s)$ is considered as the periodic extension to $(-\infty,+\infty)$ of the function defined originally on $[0,1)$ and

$$p(x)=\inf_{\alpha\in\mathfrak{A}}M(x,\alpha).$$

Then, by the Hahn-Banach extension theorem, there exists a linear functional F on $\mathfrak{F}$ satisfying $F(x)\leqslant p(x)$. If we write $\int x(s)ds$ for $F(x)$, then we can prove immediately that $\int x(s)ds$ has the following properties: (1) $\int\{ax(s)+by(x)\}ds=a\int x(s)ds+b\int y(s)ds$, where a and b are real constants. (2) $x(s)\geqslant 0$ implies $\int x(s)ds\geqslant 0$. (3) $\int x(s+s_0)ds=\int x(s)ds$, where s_0 is an arbitrary real number. (4) $\int 1\,ds=1$. If necessary, we can add the property (5) $\int x(1-s)ds=\int x(s)ds$ by defining

$$\int x(s)ds=\frac{1}{2}\{F(x(s))+F(x(1-s))\}.$$

Then

$$F(x)=\int x(s)ds$$

or

$$\frac{1}{2}\{F(x(s))+F(x(1-s))\}=\int x(s)ds$$

is called the **Banach integral** of $x(s)$.

The construction of the Daniell-Stone integral and the Banach integral opened avenues to several other abstract integrals based on the order relation, such as an integral for more general functions with values in a vector lattice, or an integral considered as a mapping from a vector lattice into another vector lattice (or from an ordered set into another ordered set). Indeed, if the function takes values in a complete vector lattice, then almost all results in this section (e.g., the Hahn-Banach extension theorem) hold trivially. For discussions of these and other abstract integrals → [13–15].

References

[1] F. Riesz, Sur la décomposition des opérations fonctionnelles linéaires, Atti del Congresso Internazionale dei Matematici, 1928, Bologna, 3, 143–148.
[2] G. Birkhoff, Lattice theory, Amer. Math. Soc. Colloq. Publ., revised edition, 1948.
[3] H. Nakano, Modulared semi-ordered linear spaces, Maruzen, 1950.
[4] M. M. Day, Normed linear spaces, Erg. Math., Springer, 1958.
[5] B. C. Vulikh, Introduction to the theory of partially ordered vector spaces, Gronigen, 1967.
[6] I. Namioka, Partially ordered linear topological spaces, Mem. Amer. Math. Soc., 1957.
[7] A. L. Peressini, Ordered topological vector spaces, Harper & Row, 1967.

[8] H. H. Schaefer, Banach lattices and positive operators, Springer, 1974.
[9] M. H. Stone, Notes on integration, Proc. Nat. Acad. Sci. US I–III, 34 (1948), 336–342, 447–455, 483–490; IV, 35 (1949), 50–58.
[10] S. Banach, The Lebesgue integral in abstract spaces, Theory of the Integral, S. Saks (ed.), Stechert, 1937, 320–330 (Dover, 1964).
[11] N. Bourbaki, Eléments de mathématique, Integration, ch. 1–9, Actualités Sci. Ind., Hermann: ch. 1–4, 117a, second edition, 1965; ch. 5, 1244b, second edition, 1967; ch. 6, 1281a, 1959; ch. 7, 8, 1306, 1963; ch. 9, 1343, 1969.
[12] S. Banach, Théorie des opérations linéaires, Warsaw, 1932, 30–32. (Chelsea, 1963).
[13] S. Izumi, N. Matuyama, M. Nakamura, M. Orihara, and G. Sunouchi, An abstract integral, Proc. Imp. Acad. Tokyo, I–III, 16 (1940), 21–25, 87–89, 518–523; IV, 17 (1941), 1–4; V–X, 18 (1942), 45–49, 50–52, 53–56, 535–538, 539–542, 543–547.
[14] E. J. McShane, Order-preserving maps and integration processes, Ann. Math. Studies, Princeton Univ. Press, 1953.
[15] T. H. Hildebrandt, Integration in abstract spaces, Bull. Amer. Math. Soc., 59 (1953), 111–139.
[16] N. Dunford and J. T. Schwartz, Linear operators, Wiley, I (1958), III (1971).
[17] K. Yosida, Functional analysis, Springer, 1965.

311 (II.12) Ordering

A. Ordering

The concept of ordering is abstracted from various relations, such as the inequality relation between real numbers and the inclusion relation between sets. Suppose that we are given a set $X=\{x, y, z, \ldots\}$; the relation between the elements of X, denoted by $\leqslant$ or other symbols, is called an **ordering (partial ordering, semiordering, order relation**, or simply **order**) if the following three laws hold: (i) the **reflexive law**, $x \leqslant x$; (ii) the **antisymmetric law**, $x \leqslant y$ and $y \leqslant x$ imply $x=y$; and (iii) the **transitive law**, $x \leqslant y$ and $y \leqslant z$ imply $x \leqslant z$.

A set X with an ordering between its elements is called an **ordered set (partially ordered set** or **semiordered set**). A subset of an ordered set X is also an ordered set with respect to the same ordering as in X. If for an arbitrary pair of elements x, y of an ordered set A either $x \leqslant y$ or $y \leqslant x$ must hold, then the ordering $\leqslant$ is called a **total** (or **linear**) **ordering**, and A is called a **totally ordered** (or **linearly ordered**) **set**.

We sometimes write $x \leqslant y$ as $y \geqslant x$. The binary relation $\geqslant$ is called the **dual ordering** of $\leqslant$; it is also an ordering. More generally, the duals of concepts, conditions, and propositions concerning an ordering are defined by replacing the ordering with its dual. For example, $x<y$ means that $x \leqslant y$ and $x \neq y$, while $x>y$ means that $x \geqslant y$ and $x \neq y$; and $>$ is the dual of $<$. If a universal proposition concerning the ordering is true, then its dual is also true; this principle is called the **duality principle for ordering**. Incidentally, $x \leqslant y$ is equivalent to the statement $x<y$ or $x=y$, according to the definition of $<$.

B. Definitions

A subset of an ordered set X of the form $\{x \mid a<x<b\}$ is denoted by (a, b), and a set of the form (a, b), $\{x \mid x<a\}$, or $\{x \mid x>a\}$ is called an **interval**. In particular, $S(c)=\{x \mid x<c\}$ is called the **segment** of X determined by c. A pair of elements a, b satisfying $a \leqslant b$ is called a **quotient** of X and is denoted by b/a.

When $a<c<b$ or $b<c<a$, c is said to lie **between** a and b. A totally ordered set A is said to be **dense** if for any pair of distinct elements a and b in A there exists a third element c lying between a and b. When $a<b$ and there is no element lying between a and b, then a is called a **predecessor** of b, and b a **successor** of a. In this manner, most of the terminology associated with the inequality of numbers is carried over to general ordering.

In an ordered set A, an element a is called an **upper bound** of a subset X if $x \leqslant a$ for every element x of X. When an upper bound exists, X is said to be **bounded from above** (or **bounded above**). The dual concept of an upper bound is a **lower bound** of the subset; and if the subset has a lower bound, it is said to be **bounded from below** (or **bounded below**). A set bounded both from above and from below is simply said to be **bounded**. When a is an upper bound of X and $a \in X$, then a is called the **greatest element** (or **maximum element**) of X. Such an element a (if it exists) is unique and is denoted by $\max X$; its dual is the **least element** (or **minimum element**) and is denoted by $\min X$. If there is a least element in the set of upper bounds of X, it is called the **least upper bound** (or **supremum**) of X and is denoted by **l.u.b.** X or **sup** X. Its dual is the **greatest lower bound** (or **infimum**) and is denoted by **g.l.b.** X or **inf** X.

If the ordered set X is the image $\varphi(\Lambda)$ of a set Λ under a †mapping φ, where Λ is of the form $\{\lambda \mid C(\lambda)\}$, then $\sup X$ is also written

$\sup_{C(\lambda)} \varphi(\lambda)$ and is called the supremum of $\varphi(\lambda)$ for all λ that satisfy $C(\lambda)$. When there is no danger of misunderstanding, it may be written as $\sup_\lambda \varphi(\lambda)$ or $\sup \varphi(\lambda)$ and called simply the supremum of $\varphi(\lambda)$; similar conventions hold for inf, max, min, etc.

An element a of a set X is called a **maximal element** if $a<x$ never holds for any element x of X; its dual is a **minimal element**. If the greatest (least) element exists, it is the only maximal (minimal) element. But in general, a maximal (minimal) element is not necessarily unique.

C. Chain Conditions

An ordered set X is said to satisfy the **minimal condition** if every nonempty subset of X has a minimal element. The dual condition is called the **maximal condition**. An infinite sequence $\{a_1, a_2, \ldots, a_n, \ldots\}$ of elements of an ordered set X such that $a_1 < a_2 < \ldots < a_n < \ldots$ is called an **ascending chain**, and the condition that X has no ascending chain is called the **ascending chain condition**. The notions dual to those of ascending chain and ascending chain condition are **descending chain** and **descending chain condition**, respectively. By the **chain condition**, we mean either the ascending or the descending chain condition. Under the †axiom of choice, the maximal condition is equivalent to the ascending chain condition, and the minimal condition to the descending chain condition.

If a totally ordered set X satisfies the minimal condition or, equivalently, if every nonempty subset of X has a least element, then the set X is called a **well-ordered set**, and its ordering is called a **well-ordering**.

The following theorem is called the principle of **transfinite induction**: Let $P(x)$ be a proposition concerning an element x of a well-ordered set X such that (i) $P(x_0)$ is true for the least element x_0 of X, and (ii) $P(x)$ is true if $P(y)$ is true for all y satisfying $y<x$. Then $P(x)$ is true for all x in X. †Mathematical induction is a special case of this principle, where X is the set of all natural numbers. To define a mapping F from a well-ordered set X into a set Y, we may use the following principle: Suppose that $F(x_0)$ is defined for the least element x_0 of X, and for each element x of X there is given a method to associate an element $G(f)$ of Y uniquely with each mapping $f: S(x) \to Y$ with domain $S(x)$, where $S(x)$ is the segment of X determined by x. Then there exists a unique mapping $F: X \to Y$ satisfying $F(x) = G(F \mid S(x))$ for all x. The definition of the mapping F by this principle is called a **definition by transfinite induction**. The principles of induction are often used for proving propositions or giving definitions concerning ordinal numbers (→ 312 Ordinal Numbers).

D. Directed Sets

An ordered set (or in general a preordered set (→ Section H)) in which every finite subset is bounded from above is called a **directed set**. Let B be a subset of a directed set A. If $\{b \mid b \geqslant a\} \cap B \neq \varnothing$ for every element a of A, then B is said to be **cofinal** in A; such a subset B is itself a directed set. If $\{b \mid b \geqslant a\} \subset B$ for some element a of A, then B is said to be **residual** in A; such a subset B is also cofinal in A. The condition that B is cofinal in A is equivalent to the condition that $A - B$ is not residual in A.

E. Order-Preserving Mappings

A mapping $\varphi: A \to A'$ of an ordered set A into an ordered set A' is called an **order-preserving mapping (monotone mapping** or **order homomorphism)** if $a \leqslant b$ always implies $\varphi(a) \leqslant \varphi(b)$. Moreover, if φ is bijective and φ^{-1} is also an order-preserving mapping from A' onto A, then φ is called an **order isomorphism**. A' is said to be **order homomorphic (order isomorphic)** to A when there exists an order homomorphism (order isomorphism) φ such that $A' = \varphi(A)$. If a mapping $\varphi: A \to A'$ gives an order isomorphism of A to the dual of A', φ is called a **dual isomorphism** (or **anti-isomorphism**).

F. Direct Sum and Direct Product

Let S be a set that is the †disjoint union of a family $\{A_\lambda\}_{\lambda \in \Lambda}$ of its subsets, and suppose that each A_λ is an ordered set. For $a, b \in S$, define $a \leqslant b$ to mean that $a, b \in A_\lambda$ for some $\lambda \in \Lambda$ and $a \leqslant b$ with respect to the ordering in A_λ. The ordered set S obtained in this way is called the **direct sum** (or **cardinal sum**) of the family $\{A_\lambda\}_{\lambda \in \Lambda}$ of ordered sets. When $(a_\lambda)_{\lambda \in \Lambda}$ and $(b_\lambda)_{\lambda \in \Lambda}$ are elements of the †Cartesian product $P = \prod_{\lambda \in \Lambda} A_\lambda$ of a family $\{A_\lambda\}_{\lambda \in \Lambda}$ of ordered sets, we define $(a_\lambda)_{\lambda \in \Lambda} \leqslant (b_\lambda)_{\lambda \in \Lambda}$ to mean that $a_\lambda \leqslant b_\lambda$ holds for all $\lambda \in \Lambda$. The ordered set P obtained in this way is called the **direct product** (or **cardinal product**) of the family $\{A_\lambda\}_{\lambda \in \Lambda}$ of ordered sets.

G. Ordinal Sum and Ordinal Product

Suppose that $\mathfrak{A} = \{A, B, \ldots\}$ is a family of mutually disjoint ordered sets and is itself an

ordered set. Then an ordering $\leqslant$ can be defined in the disjoint union $S=\bigcup X$ $(X\in\mathfrak{A})$ as follows: $x\leqslant y$ in S means that either (i) there exists an A satisfying $x, y\in A\in\mathfrak{A}$ and $x\leqslant y$ holds with respect to the ordering in A; or (ii) for A and B satisfying $x\in A\in\mathfrak{A}$, $y\in B\in\mathfrak{A}$, we have $A<B$. The ordered set S obtained in this way is called the **ordinal sum** obtained from $\mathfrak{A}$ and is denoted by $\sum_{\mathfrak{A}}X$. In particular, if $\mathfrak{A}=\{A,B\}$ and $A<B$, the ordinal sum is denoted by $A+B$.

Suppose that X is a subset of the Cartesian product $\prod_\lambda X_\lambda$ of a family of ordered sets indexed by an ordered set Λ, and the subset $\{\lambda\mid x_\lambda\neq y_\lambda\}$ of Λ has a least element whenever $x=(x_\lambda)_{\lambda\in\Lambda}$ and $y=(y_\lambda)_{\lambda\in\Lambda}$ are two distinct elements of X. The ordering in X defined by setting $x<y$ when $x_\mu<y_\mu$ for the least element μ of $\{\lambda\mid x_\lambda\neq y_\lambda\}$ is called the **lexicographic ordering** in X. It can be applied to $X=\prod_\lambda X_\lambda$ if Λ is well-ordered; X is then called the **ordinal product**. When $A, B, \ldots$ are ordered sets, $AB\ldots$ is often used to denote the ordinal product obtained from $X_1=A$, $X_2=B, \ldots$ with the ordering $1<2<\ldots$ of indices; the ordering in this ordinal product is called the **lexicographic ordering** in the Cartesian product $A\times B\times\ldots$.

H. Preordering

A relation R between elements of a set X is called a **preordering** (or **pseudoordering**) if it satisfies the reflexive law and the transitive law, but not necessarily the antisymmetric law. By defining $(x_1,y_1)R\,(x_2,y_2)\Leftrightarrow x_1\leqslant x_2$, for example, a preordering of pairs (x,y) of real numbers is obtained. From a preordering R an equivalence relation $\sim$ can be defined in X by $x\sim y\Leftrightarrow(xRy$ and $yRx)$. Let $[X]=X/\sim$ be the †quotient set of set X by this equivalence relation, and let $[x]$, $[y]$ be the equivalence classes determined by $x, y\in X$; then an ordering $\leqslant$ can be defined in $[X]$ by $[x]\leqslant[y]\Leftrightarrow xRy$. (For further topics → 52 Categories and Functors; 409 Structures.)

References

[1] N. Bourbaki, Eléments de mathématique, I. Théorie des ensembles, ch. 3, Actualités Sci. Ind., 1243b, Hermann, second edition, 1967; English translation, Theory of sets, Addison-Wesley, 1968.
[2] G. Birkhoff, Lattice theory, Amer. Math. Soc. Colloq. Publ., 1940, revised edition, 1948.
Also → references to 381 Sets.

312 (II.13) Ordinal Numbers

A. General Remarks

Let $A\cong B$ mean that two †ordered sets A, B are †order isomorphic; then the relation $\cong$ is an †equivalence relation. An equivalence class under this relation is called an **order type**, and the class to which an ordered set A belongs is called the **order type of** A. Historically, an ordinal number was first defined as the order type of a †well-ordered set (Cantor [2]). However, it was found that a contradiction occurs if order type defined in this way are considered to form a set. Hence, another definition was given by J. von Neumann [3], which is stated in Section B. A similar situation was found concerning the definition of †cardinal numbers, which led to a new definition of cardinal numbers using ordinal numbers, which is given in Section D.

B. Definitions

A set α is called an **ordinal number** if it satisfies the following two conditions: (i) α is a well-ordered set with the †binary relation $\in$ as its ordering; and (ii) $\beta\in\alpha$ implies $\beta\subset\alpha$. According to this definition, the empty set is an ordinal number, which is denoted by 0. Also, $1=\{0\}$, $2=\{0,1\}$, $3=\{0,1,2\},\ldots$ are ordinal numbers. These ordinal numbers, which are finite sets, are called **finite ordinal numbers**. The finite ordinal numbers are identified with the natural numbers (including 0). The set $\omega=\{0,1,2,\ldots\}$ of all natural numbers is also an ordinal number. An ordinal number that is an infinite set, like ω, is called a **transfinite ordinal number**.

For every well-ordered set A, there exists one and only one ordinal number order isomorphic to A. This ordinal number is called the **ordinal number** of A. (Throughout this article, lower-case Greek letters denote ordinal numbers.) We also write $\alpha\in\beta$ as $\alpha<\beta$, which defines an †ordering of the ordinal numbers. The least ordinal number is 0, and the ordering of the finite ordinal numbers coincides with the usual ordering of the natural numbers. The least transfinite ordinal number is ω. The ordering $\leqslant$, introduced by defining $\alpha\leqslant\beta$ to mean either $\alpha<\beta$ or $\alpha=\beta$, is a †linear ordering and, in fact, a †well-ordering of the ordinal numbers. Therefore †transfinite induction can be applied to ordinal numbers.

For any ordinal number α, the set $\alpha'=\{\xi\mid\xi\leqslant\alpha\}$ is also an ordinal number, and is the †successor of α. There exists at most one

ordinal number that is the †predecessor of α. A transfinite ordinal number without a predecessor is called a **limit ordinal number**, and all the other ordinal numbers are called **isolated ordinal numbers**. The first limit ordinal number is ω. For any set A of ordinal numbers, $\{\xi \mid \exists\eta\,(\xi<\eta\in A)\}$ is an ordinal number and is $\sup A$, the †supremum of A.

C. Sum, Product, and Power

The **sum** $\alpha+\beta$, the **product** $\alpha\cdot\beta$ (or $\alpha\beta$), and the **power** α^β of ordinal numbers α, β are defined by transfinite induction on β and have the following properties:

$$\alpha+0=\alpha, \quad \alpha+\beta'=(\alpha+\beta)',$$

$$\alpha+\gamma=\sup\{\alpha+\xi\mid\xi<\gamma\};$$

$$\alpha\cdot 0=0, \quad \alpha\cdot\beta'=\alpha\cdot\beta+\alpha,$$

$$\alpha\cdot\gamma=\sup\{\alpha\cdot\xi\mid\xi<\gamma\};$$

$$\alpha^0=1, \quad \alpha^{\beta'}=\alpha^\beta\cdot\alpha, \quad \alpha^\gamma=\sup\{\alpha^\xi\mid\xi<\gamma\}.$$

Here γ is a limit ordinal number, and for the power we assume that $\alpha>0$. The sum and product thus defined satisfy the associative laws $(\alpha+\beta)+\gamma=\alpha+(\beta+\gamma)$, $(\alpha\cdot\beta)\cdot\gamma=\alpha\cdot(\beta\cdot\gamma)$ and the left distributive law $\alpha\cdot(\beta+\gamma)=\alpha\cdot\beta+\alpha\cdot\gamma$; the power satisfies the laws $\alpha^{\beta+\gamma}=\alpha^\beta\cdot\alpha^\gamma$, $\alpha^{\beta\gamma}=(\alpha^\beta)^\gamma$. If α and β are the ordinal numbers of the well-ordered sets A and B, respectively, then $\alpha+\beta$ is the ordinal number of the †ordinal sum $A+B$, and $\alpha\cdot\beta$ is the ordinal number of the †ordinal product BA.

When $\pi>1$, any ordinal number α can be written uniquely in the form

$$\alpha=\pi^{\beta_1}\cdot\gamma_1+\pi^{\beta_2}\cdot\gamma_2+\ldots+\pi^{\beta_n}\cdot\gamma_n;$$

$$\beta_1>\beta_2>\ldots\beta_n>0, \quad 0<\gamma_i<\pi, \quad 1\leqslant i\leqslant n,$$

which is called the **π-adic normal form** for α; when $\pi=\omega$, it is called **Cantor's normal form**.

Let f be an ordinal number-valued function of ordinal numbers. We say that f is **strictly monotone** when $\alpha<\beta$ implies $f(\alpha)<f(\beta)$. If f is strictly monotone, then $\alpha\leqslant f(\alpha)$. We say that f is **continuous** when $f(\gamma)=\sup\{f(\xi)\mid\xi<\gamma\}$ for each limit ordinal number γ. A strictly monotone continuous function is called a **normal function.** If f is a normal function, then for any α there exists a β that satisfies $f(\beta)=\beta>\alpha$. In fact, it suffices to define $\beta_n (n<\omega)$ by $\beta_0=f(\alpha+1)$, $\beta_{n+1}=f(\beta_n)$ and put $\beta=\sup\{\beta_n\mid n<\omega\}$. Since $f(\alpha)=\omega^\alpha$ is a normal function, there exists an ε that satisfies $\omega^\varepsilon=\varepsilon$. Such an ordinal number ε is called an **ε-number**. We say that β is **cofinal** to α when there exists a monotone function f that satisfies $\alpha=\sup\{f(\xi)'\mid\xi<\beta\}$. The first ordinal number that is cofinal to α is called the **cofinality** of α and is denoted by $\mathrm{cf}(\alpha)$.

D. Cardinal Numbers

Let $M\sim N$ mean that a one-to-one correspondence exists between the two sets M and N. An ordinal number α with the property that $\alpha\sim\xi$ implies $\alpha\leqslant\xi$ is called an **initial number** or a **cardinal number**.

With the †axiom of choice, it can be shown that for each set M there exists one and only one cardinal number α satisfying $M\sim\alpha$. This unique α is called the **cardinality** (or **cardinal number**) **of the set** M and is denoted by $\overline{\overline{M}}$.

All finite ordinal numbers are cardinal numbers, and ω is the least transfinite cardinal number. There exists one and only one monotone function that maps the class of ordinal numbers onto the class of transfinite cardinal numbers, and it is a normal function. The value of this function corresponding to α is denoted by $\aleph_\alpha$ **(aleph alpha)** or ω_α. In particular, $\aleph_0=\omega$, and $\aleph_1$ is both the smallest uncountable cardinal number and the smallest uncountable ordinal number. A finite ordinal number is called an **ordinal number of the first number class**, and an ordinal number α satisfying $\aleph_0\leqslant\alpha<\aleph_1$ is called an **ordinal number of the second number class**. The concept of **ordinal number of the third** (or **higher**) **number class** is defined similarly.

E. Inaccessible Ordinal Numbers

The cofinality $\mathrm{cf}(\alpha)$ of α always satisfies $\mathrm{cf}(\alpha)\leqslant\alpha$. An ordinal number is said to be **regular** when $\mathrm{cf}(\alpha)=\alpha$ and singular when $\mathrm{cf}(\alpha)<\alpha$. For any ordinal number α, $\mathrm{cf}(\alpha)$ is a regular cardinal number; therefore any regular ordinal number is a cardinal number. When $\alpha=\omega_\beta$ is regular and β is a limit ordinal number, α is said to be **weakly inaccessible**. Let R be the set-valued function of ordinal numbers, defined by $R(0)=\varnothing$ and $R(\alpha)=\bigcup\{\mathfrak{P}(R(\xi))\mid\xi<\alpha\}$ (by †transfinite induction), where $\mathfrak{P}(M)$ denotes the †power set of M. A regular ordinal number α is said to be **strongly inaccessible** when $\alpha>\omega$ and the following condition is satisfied: If x, y are a pair of sets such that $x\in R(\alpha)$, $y\subset R(\alpha)$, and there exists a mapping of x onto y, then $y\in R(\alpha)$. If a regular ordinal number α is strongly inaccessible, it is weakly inaccessible. A strongly inaccessible ordinal number is usually defined as a regular number $\alpha>\omega$ such that $\beta<\alpha$ implies $\overline{\overline{\mathfrak{P}(\beta)}}<\alpha$. Under the axiom of choice, this definition is equivalent to the one given here. Moreover, under the †generalized continuum hypothesis, strong

inaccessibility and weak inaccessibility are equivalent.

References

[1] N. Bourbaki, Eléments de mathématique, I. Théorie des ensembles, ch. 3, Actualités Sci. Ind., 1243b, second edition, Hermann, 1967; English translation, Theory of sets, Addison-Wesley, 1968.
[2] G. Cantor, Beiträge zur Begründung der transfiniten Mengenlehre II, Math. Ann., 49 (1897), 207–246. (Gesammelte Abhandlungen, Springer, 1932; English translation, Contributions to the founding of the theory of transfinite numbers, Open Court, 1915.)
[3] J. von Neumann, Zur Einführung der transfiniten Zahlen, Acts Sci. Math. Szeged., 1 (1923), 199–208. (Collected works I, Pergamon, 1961.)
Also → references to 381 Sets.

313 (XIII.2) Ordinary Differential Equations

A. General Remarks

Let x be a real (complex) variable and y a real (complex) function of x. Assume that $y=F(x)$ is a differentiable function of class C^n if x, y are real, and a holomorphic function if x, y are complex. We write $y', y'', \ldots, y^{(n)}$ for the first n derivatives of y. A relation among $x, y, y', \ldots, y^{(n)}$,

$$f(x, y, y', \ldots, y^{(n)})=0 \tag{1}$$

(which holds identically with respect to x), is called an **ordinary differential equation** for the function $y=F(x)$. Here we assume that the function f in the left-hand side of (1) is a real (complex) function of the $n+2$ variables $x, y, y', \ldots, y^{(n)}$ and is defined in a given domain of $\mathbf{R}^{n+2}$ ($\mathbf{C}^{n+2}$). Usually we assume further that f has a certain regularity, such as being of class C^r ($r=0,1,\ldots,\infty$), †real analytic, or †complex analytic. A function $y=F(x)$ that satisfies (1) is called a **solution** of (1). To find a solution of (1) is to **solve** or **integrate** it. Ordinary differential equations may be contrasted to **partial differential equations**, which are equations similar to (1) but in which y is a function of two or more variables $x_1, x_2, \ldots$ and which contain the partial derivatives $\partial y/\partial x_1, \partial y/\partial x_2, \ldots$ (— 320 Partial Differential Equations). Ordinarily, the term **differential equation** refers to an ordinary or partial differential equation. In this article, since we are concerned only with ordinary differential equations, we omit the word "ordinary." If the left-hand side f of (1) contains $y^{(n)}$ explicitly or $\partial f/\partial y^{(n)} \neq 0$, then we say that the **order** of (1) is n, and if further f is a polynomial in $y, y', \ldots, y^{(n)}$ that is of degree m with respect to $y^{(n)}$, we say that the **degree** of (1) is m. In particular, if f is a linear form in $y, y', \ldots, y^{(n)}$, then (1) is said to be **linear**. A differential equation that is not linear is said to be **nonlinear** (→ 252 Linear Ordinary Differential Equations; 291 Nonlinear Problems).

Let $\varphi(x, y, c_1, \ldots, c_n)$ be a function of the $n+2$ variables $x, y, c_1, \ldots, c_n$ of class C^r in a domain D, and let $(x_0, y_0, c_1^0, \ldots, c_n^0) \in D$, $\varphi(x_0, y_0, c_1^0, \ldots, c_n^0)=0$, and $\varphi_y(x_0, y_0, c_1^0, \ldots, c_n^0) \neq 0$. Then the equation $\varphi(x, y, c_1^0, \ldots, c_n^0)=0$ defines an †implicit function $y(x)$ of class C^r satisfying the condition $y(x_0)=y_0$. Consider $c_1, \ldots, c_n$ to be constants in $\varphi(x, y, c_1, \ldots, c_n)=0$ and differentiate φ n times with respect to x. Then we obtain a system of n equations in the variables $x, y, y', \ldots, y^{(n)}, c_1, \ldots, c_n$. If we can eliminate $c_1, \ldots, c_n$ from these n equations and $\varphi=0$, then we obtain an nth-order differential equation of the form (1). Conversely, a solution of an nth-order differential equation can usually be written in the form

$$\varphi(x, y, c_1, \ldots, c_n)=0, \tag{2}$$

which contains n **arbitrary constants** $c_1, \ldots, c_n$ (sometimes called **integration constants**). A solution containing n arbitrary constants of the form (2) of an nth-order differential equation is called a **general solution**, and a solution $\varphi(x, y, c_1^0, \ldots, c_n^0)=0$ obtained from a general solution $\varphi=0$ by giving particular values $c_1^0, \ldots, c_n^0$ to the arbitrary constants is called a **particular solution**. Some equations admit solutions that are not particular solutions. They are called **singular solutions** (for example, †Clairaut differential equations; → Appendix A, Table 14.I).

B. Systems of Differential Equations

A set of n differential equations containing n unknown functions $y_1, \ldots, y_n$ of a variable x is called a **system of ordinary differential equations**. Here each equation of the system has a form similar to (1), but each left-hand side contains $y_1, \ldots, y_n$ and their derivatives. A set of n functions $y_1, \ldots, y_n$ of x is called a **solution** if the functions satisfy the given system of differential equations. The highest order of derivatives in the left-hand sides is called the **order** of the system of differential equations.

We consider most frequently a first-order system of the form

$$y_i'=f_i(x, y_1, \ldots, y_n), \qquad i=1,2,\ldots,n. \tag{3}$$

If we put $y=y_1, y'=y_2, \ldots, y^{(n-1)}=y_n$ and solve (1) with respect to $y^{(n)}$ to get $y^{(n)}=f_n(x, y_1, \ldots, y_n)$, then (1) is equivalent to a system of equations of the form (3), where $f_1=y_2, f_2=y_3, \ldots, f_n=f$. In an analogous way, a general system of equations can be transformed to a system of the form (3). Therefore (3) is called the **normal form** of differential equations.

C. The Geometric Interpretation

When $x, y_1, \ldots, y_n$ are real, (3) can be interpreted as follows: Let $I=(a,b)$ be an open interval and D a domain of $\mathbf{R}^n$. Let

$$y_i=\varphi_i(x, c_1, \ldots, c_n), \qquad i=1,2,\ldots,n, \tag{4}$$

be functions of class C^1 defined for $(x, c_1, \ldots, c_n) \in I \times D$, and let $\mathfrak{D}(x_0)$ be the image in the $y_1, \ldots, y_n$-space of D under the mapping $y_i=\varphi_i(x_0, c_1, \ldots, c_n)$ $(i=1,2,\ldots,n)$ for a fixed $x_0 \in I$. We assume that for each fixed $x_0 \in I$ we have $\partial(\varphi_1, \ldots, \varphi_n)/\partial(c_1, \ldots, c_n) \neq 0$ in D. Then for every $x=x_0 \in I$, $c_1, \ldots, c_n$ are considered to be functions of $(y_1, \ldots, y_n)$ defined in a neighborhood of every point $(y_1^0, \ldots, y_n^0)$ of $\mathfrak{D}(x_0)$, and we have $y_i'=\varphi_i'(x, c_1, \ldots, c_n)= f_i(x, y_1, \ldots, y_n)$ $(i=1,2,\ldots,n)$, i.e., $y_1, \ldots, y_n$ satisfy a system of differential equations of the form (3). On the other hand, (4) represents a family of curves of class C^1 in the $x, y_1, \ldots, y_n$-space $\mathbf{R}^{n+1}$ containing n parameters $c_1, \ldots, c_n$, for which $(y_1', \ldots, y_n')$ is the tangent vector (in the terminology of physics, $(y_i', \ldots, y_n')$ gives the speed and the direction of a stationary flow in $\mathbf{R}^{n+1}$ at each point). By solving (3) we find the family of curves of class C^1 in $\mathbf{R}^{n+1}$ (in the terminology of physics, we find a stationary flow of which the speed and the direction are given at each point). A solution containing n parameters analogous to (4) is called a **general solution** of (3), and a solution obtained from a general solution by giving particular values to the n parameters is called a **particular solution**.

As may be imagined by the interpretation in this section, there exists in general one and only one particular solution passing through the point $(x_0, y_1^0, \ldots, y_n^0)$ for $x_0 \in I$, $(y_1^0, \ldots, y_n^0) \in \mathfrak{D}(x_0)$. The problem of finding this solution, i.e., the solution of (3) for which $y_i(x_0)=y_i^0$ for $x=x_0$, is called the **initial value problem** (→ 316 Ordinary Differential Equations (Initial Value Problems)).

D. Methods of Integration

We have different methods of solving differential equations. To solve differential equations by a finite number of integrations is called the **method of quadrature**. This method is useful for some special types of differential equations (→ Appendix A, Table 14.I). S. Lie gave theoretical foundations for this method by using Lie transformation groups (→ 431 Transformation Groups; Appendix A, Table 14.III). There are many other methods, for example, power series methods (assuming that the solution can be expanded in a power series $\sum a_\nu (x-a)^\nu$, substituting the series for y in (1), and finding its coefficients); methods of successive approximation; methods using †Laplace transforms or †Fourier transforms; †perturbation methods; numerical methods; etc. (→ 303 Numerical Solution of Ordinary Differential Equations).

Historically, finding explicit solutions of various kinds of differential equations has been the main object of the theory. Recently, however, the importance of qualitative studies, in particular theorems on the existence and uniqueness of solutions, has been recognized. For example, if a solution with a property A is given, and if the uniqueness of the solution having the property A and the existence of solutions having the properties A and B can be shown, then the given solution necessarily has the property B. In this way, topological and analytic studies of differential equations are applied to find their solutions (→ 314 Ordinary Differential Equations (Asymptotic Behavior of Solutions); 315 (Boundary Value Problems); 316 (Initial Value Problems); 126 Dynamical Systems).

References

[1] W. Kaplan, Ordinary differential equations, Addison-Wesley, 1958.
[2] P. Hartman, Ordinary differential equations, Wiley, 1964.
[3] E. Hille, Lectures on ordinary differential equations, Addition-Wesley, 1968.
Also → references to 316 Ordinary Differential Equations (Initial Value Problems).

314 (XIII.5) Ordinary Differential Equations (Asymptotic Behavior of Solutions)

A. Linear Differential Equations

A system of linear ordinary differential equations can be written as

$$\mathbf{x}'=A(t)\mathbf{x}, \tag{1}$$

where t is a real independent variable, $\mathbf{x} = (x_1, \ldots, x_n)$ is an n-dimensional complex vector function of t, and $A(t)$ is an $n \times n$ matrix whose elements are complex-valued functions of t. If $A(t)$ is a continuous function of t defined on an open interval I, any solution of (1) is continuously differentiable for $t \in I$. The question naturally arises as to how the solutions behave as t approaches either one of the endpoints of I; that is, the question of the **asymptotic properties** of the solutions. The interval I can always be taken to be $0 \leqslant t < \infty$, by applying a suitable transformation of the independent variable if necessary.

The study of the †asymptotic expansions of solutions when the coefficient $A(t)$ is an analytic function of t was initiated by H. Poincaré in 1880. This work was continued by J. Horn, J. C. C. A. Kneser, and others in the direction of removing assumptions on the structure of $A(t)$ and extending the domain where the expansions are valid. The theory has been almost completed by W. J. Trjitzinsky, J. Malmquist, and M. Hukuhara (→ 254 Linear Ordinary Differential Equations (Local Theory)). On the other hand, O. Perron initiated a new direction of research by weakening the regularity conditions on the coefficients. His work was continued by F. Lettenmeyer, R. A. Späth, Hukuhara, and others. The methods used in these two lines of investigation were originally distinct, but Hukuhara established a unified method of treating the problems arising in these two different types of investigations. Furthermore, he succeeded in sharpening those results previously obtained.

Here we assume that $A(t)$ need not be analytic. The following asymptotic properties of a solution $\mathbf{x}(t)$ as $t \to \infty$ are considered: (i) boundedness of $\limsup t^{-1} \log|\mathbf{x}(t)|$; (ii) boundedness of solution: $\limsup|\mathbf{x}(t)| < \infty$; (iii) convergence of solution: $\lim \mathbf{x}(t)$; (iv) integrability: $\int^\infty |\mathbf{x}(s)|^p ds < \infty$, etc. We call $\chi(\mathbf{x}) = \limsup t^{-1} \log|\mathbf{x}(t)|$ the **type number** (or **Lyapunov characteristic number**) of the solution $\mathbf{x}(t)$. The fact that all solutions of (1) are bounded is equivalent to the †stability of the solution $\mathbf{x} = 0$, and the fact that all solutions of (1) tend to zero as $t \to \infty$ is equivalent to the †asymptotic stability of the solution $\mathbf{x} = 0$.

B. Constant Coefficients and Periodic Coefficients

We begin with the particular case of (1), where $A(t)$ is a constant matrix:

$$\mathbf{x}' = A\mathbf{x}. \tag{2}$$

To study the asymptotic properties of the solutions of (2), it suffices to transform the matrix A into a †Jordan canonical form, since the structure of the solution space of (2) is completely determined by the Jordan canonical form of A. Thus all solutions of (2) are bounded if and only if every eigenvalue of A has a real part not greater than zero, and those with zero real parts are of simple type, that is, the corresponding blocks in the Jordan canonical form are all 1×1 matrices; all solutions of (2) tend to zero as $t \to \infty$ if and only if every eigenvalue of A has negative real part.

Consider the linear system

$$\mathbf{x}' = A_0(t)\mathbf{x}, \tag{3}$$

where $A_0(t)$ is a periodic matrix function of period ω. According to †Floquet's theorem, (3) is transformed into a system with constant coefficients by means of a suitable transformation $\mathbf{x} = P(t)\mathbf{y}$, where $P(t)$ is a nonsingular periodic matrix of period ω. Thus, at least theoretically, the information on the asymptotic behavior of the solutions of the periodic system (3) can be derived from the corresponding theory for the system with constant coefficients (2).

C. Asymptotic Integration

Suppose that $A(t)$ is bounded. Then the type number $\chi(\mathbf{x})$ is finite for any nontrivial solution $\mathbf{x}(t)$ of (1), and the number of distinct type numbers does not exceed n.

Consider the linear system

$$\mathbf{x}' = [A + B(t)]\mathbf{x}, \tag{4}$$

where A is a constant matrix and $B(t)$ is a matrix function such that $\int_t^{t+1} \|B(s)\| \, ds \to 0$ as $t \to \infty$. For any nontrivial solution $\mathbf{x}(t)$ of (4), the limit $\mu = \lim t^{-1} \log|\mathbf{x}(t)|$ exists and is equal to the real part of one of the eigenvalues of A. Conversely, if at least one eigenvalue of A has real part μ, then there exists a nontrivial solution $\mathbf{x}(t)$ of (4) satisfying $\lim t^{-1} \log|\mathbf{x}(t)| = \mu$. Suppose in addition that $B(t) \to 0$ as $t \to \infty$. Let $\mu_1 \leqslant \mu_2 \leqslant \ldots \leqslant \mu_n$ be the real parts of the eigenvalues of A. Then there exists a †fundamental system of solutions of (4), $\{\mathbf{x}_1(t), \ldots, \mathbf{x}_n(t)\}$, such that for any c_i, $c_k \neq 0$,

$$\log|c_1\mathbf{x}_1(t) + \ldots + c_k\mathbf{x}_k(t)| = \mu_k t + o(t).$$

A sharp estimate of the term $o(t)$ was given by Hukuhara.

Next consider the linear system

$$\mathbf{x}' = [A(t) + B(t)]\mathbf{x}, \tag{5}$$

where the matrices $A(t)$ and $B(t)$ satisfy $\int^\infty \|A'(s)\| \, ds < \infty$ and $\int^\infty \|B(s)\| \, ds < \infty$.

Let $\lambda_1(t), \ldots, \lambda_n(t)$ and $\lambda_1, \ldots, \lambda_n$, $\lambda_k = \lim \lambda_k(t)$,

be the eigenvalues of $A(t)$ and $A=\lim A(t)$, respectively. N. Levinson proved the following theorem: Assume that $\lambda_1, \ldots, \lambda_n$ are mutually distinct and $M_{jk}(t)=\mathrm{Re}\int_0^t[\lambda_j(s)-\lambda_k(s)]\,ds$ satisfy either $M_{jk}(t)\to\infty$ as $t\to\infty$ and for each pair (j,k), $M_{jk}(t_2)-M_{jk}(t_1)\geqslant -K$ for $t_1<t_2$, or $M_{jk}(t)\to-\infty$ as $t\to\infty$ and $M_{jk}(t_2)-M_{jk}(t_1)\leqslant K$ for $t_1<t_2$, or $|M_{jk}(t_2)-M_{jk}(t_1)|\leqslant K$ for all t_1, t_2, where K is a positive constant. Then (5) has a fundamental system of solutions $\{\mathbf{x}_1(t), \ldots, \mathbf{x}_n(t)\}$ such that

$$\mathbf{x}_j(t)=\exp\left(\int_0^t \lambda_j(s)\,ds\right)[\xi_j+o(1)], \qquad j=1,\ldots,n,$$

where ξ_j is an eigenvector of A corresponding to λ_j.

D. Boundedness and Convergence of Solutions

Consider again the linear system (5) satisfying $\int^\infty\|A'(s)\|\,ds<\infty$ and $\int^\infty\|B(s)\|\,ds<\infty$. Suppose that all eigenvalues of $A(t)$ have nonpositive real parts and that the eigenvalues of $A=\lim A(t)$ whose real parts vanish are simple. Then all solutions of (5) are bounded. This result is a generalization, due to L. Cesari, of the so-called **Dini-Hukuhara theorem**.

In the case of general $A(t)$, it is known that not all solutions of (5) are bounded even if all solutions of (1) tend to zero as $t\to\infty$ and if the matrix $B(t)$ is such that $\int^\infty\|B(s)\|\,ds<\infty$ and $B(t)\to 0$ as $t\to\infty$. However, if $A(t)$ is periodic or satisfies $\liminf \mathrm{Re}\int^t \mathrm{tr}\,A(s)\,ds>-\infty$, then under the assumption that $\int^\infty\|B(s)\|\,ds<\infty$, the boundedness of all solutions of (1) implies the boundedness of all solutions of (5).

The following inequalities often provide useful information about the asymptotic behavior of solutions of (1):

$$|\mathbf{x}(0)|\exp\left(-\int_0^t \mu[-A(s)]\,ds\right)\leqslant|\mathbf{x}(t)|$$

$$\leqslant|\mathbf{x}(0)|\exp\left(\int_0^t \mu[A(s)]\,ds\right), \qquad t\geqslant 0,$$

where $\mu[A(t)]=\lim_{h\to+0}[\|I+hA(t)\|-1]/h$. ($\mu[A(t)]$ was introduced by Lozinskiĭ.) If $\limsup\int^t\mu[A(s)]\,ds<\infty$, then all solutions of (1) are bounded; if $\lim\int^t\mu[A(s)]\,ds$ exists, then for every solution $\mathbf{x}(t)$ of (1), $|\mathbf{x}(t)|$ tends to a finite limit as $t\to\infty$; and if $\lim\int^t\mu[A(s)]\,ds=-\infty$, then all solutions of (1) tend to zero as $t\to\infty$. It can be shown that every solution of (1) tends to a finite limit as $t\to\infty$, provided that $\int^\infty\|A(s)\|\,ds<\infty$.

If all solutions of (1) are bounded, then $\limsup\mathrm{Re}\int^t\mathrm{tr}\,A(s)\,ds<\infty$. If $\liminf\mathrm{Re}\int^t\mathrm{tr}\,A(s)\,ds>-\infty$, then (1) has a solution $\mathbf{x}(t)$ with the property that $\limsup|\mathbf{x}(t)|>0$. When $\lim|\mathbf{x}(t)|$ exists for every solution $x(t)$ of (1), if there exists a nontrivial solution $\mathbf{x}(t)$ of (1) such that $\lim x(t)=0$, then $\lim\mathrm{Re}\int^t\mathrm{tr}\,A(s)\,ds=-\infty$, but if there is no such solution, then $\mathrm{Re}\int^t\mathrm{tr}\,A(s)\,ds$ is bounded.

E. Nonlinear Differential Equations

Consider a system of nonlinear differential equations of the form

$$\mathbf{x}'=A\mathbf{x}+\mathbf{f}(t,\mathbf{x}), \tag{6}$$

where A is an $n\times n$ constant matrix and $\mathbf{f}(t,\mathbf{x})$ is an n-vector function that is continuous for $t\geqslant 0$, $|\mathbf{x}|<\Delta$, and that satisfies $\mathbf{f}(t,\mathbf{0})=\mathbf{0}$. Suppose that $\mathbf{f}(t,\mathbf{x})/|\mathbf{x}|\to\mathbf{0}$ as $|\mathbf{x}|\to 0$ and $t\to\infty$. Then for every eventually nontrivial solution $\mathbf{x}(t)$ of (6) that tends to zero as $t\to\infty$, $\mu=\lim t^{-1}\log|\mathbf{x}(t)|$ exists and equals the real part of one of the eigenvalues of A. Conversely, if at least one eigenvalue of A has real part $\mu<0$, then there exists a solution $\mathbf{x}(t)$ of (6) such that $\lim t^{-1}\log|\mathbf{x}(t)|=\mu$. Suppose that $\mathbf{f}(t,\mathbf{x})/|\mathbf{x}|\to 0$ as $|\mathbf{x}|\to 0$ uniformly with respect to t. Then if all eigenvalues of A have negative real parts, the zero solution $\mathbf{x}(t)=\mathbf{0}$ of (6) is asymptotically stable, and if A has an eigenvalue whose real part is positive, then the zero solution of (6) is †unstable. Suppose that $\mathbf{f}_{\mathbf{x}}(t,\mathbf{x})=(\partial f_j(t,x)/\partial x_k)\to 0$ as $|\mathbf{x}|\to 0$ uniformly with respect to t. In this case, if A is a matrix such that its k eigenvalues have negative real parts and the other $n-k$ eigenvalues have positive real parts, then there exists a k-dimensional manifold S containing the origin with the following property: For t_0 sufficiently large, any solution $\mathbf{x}(t)$ of (6) tends to zero as $t\to\infty$, provided that $\mathbf{x}(t_0)\in S$, and if $\mathbf{x}(t_0)\notin S$, $\mathbf{x}(t)$ cannot remain in the vicinity of the origin no matter how close $\mathbf{x}(t_0)$ is to the origin.

In the nonlinear system

$$\mathbf{x}'=\mathbf{F}(t,\mathbf{x}), \tag{7}$$

suppose that $\mathbf{F}(t,\mathbf{x})$ is of period ω with respect to t and has continuous partial derivatives with respect to $\mathbf{x}$. Suppose, moreover, that (7) has a solution $\mathbf{p}(t)$ of period ω. If all the †characteristic exponents of the †variational system of (7) with respect to $\mathbf{p}(t)$, $\mathbf{y}'=\mathbf{F}_{\mathbf{x}}(t,\mathbf{p}(t))\mathbf{y}$ with $\mathbf{F}_{\mathbf{x}}(t,\mathbf{x})=(\partial F_j(t,x)/\partial x_k)$, have negative real parts, then the periodic solution $\mathbf{p}(t)$ is asymptotically stable. If an autonomous system $\mathbf{x}'=F(\mathbf{x})$ has a periodic solution $\mathbf{p}(t)$ and the corresponding variational system $\mathbf{y}'=\mathbf{F}_{\mathbf{x}}(\mathbf{p}(t))\mathbf{y}$ has $n-1$ characteristic exponents with negative eigenvalues, then there exists an $\varepsilon>0$ such that for any solution $\mathbf{x}(t)$ satisfying $|\mathbf{x}(t_1)-\mathbf{p}(t_0)|<\varepsilon$ for some t_0 and t_1, we have $|\mathbf{x}(t)-\mathbf{p}(t+c)|\to 0$ as $t\to\infty$ for a suitable choice of c (asymptotic phase).

F. Scalar Differential Equations

The aforementioned results can be specialized to the case of higher-order scalar (or single) ordinary differential equations. Much sharper results can often be derived through direct analysis of scalar equations themselves. In particular, detailed and deep results have been obtained for second-order linear differential equations of the form

$$x''+q(t)x=0, \tag{8}$$

e.g., †Mathieu's equation.

If $\int^{\infty} s|q(s)|\,ds<\infty$, then (8) has a fundamental system of solutions $\{x_1(t), x_2(t)\}$ satisfying, for $t\to\infty$,

$$x_1(t)=1+o(1), \quad x_2(t)=t[1+o(1)],$$

$$x_1'(t)=t^{-1}o(1), \quad x_2'(t)=1+o(1);$$

if $\int^{\infty}|q(s)+1|\,ds<\infty$, then (8) has a fundamental system of solutions satisfying

$$x_1(t)=e^{t}[1+o(1)], \quad x_2(t)=e^{-t}[1+o(1)],$$

$$x_1'(t)=e^{t}[1+o(1)], \quad x_2'(t)=-e^{-t}[1+o(1)];$$

and if $\int^{\infty}|q(s)-1|\,ds<\infty$, then (8) has a fundamental system of solutions

$$x_1(t)=e^{it}[1+o(1)], \quad x_2(t)=e^{-it}[1+o(1)],$$

$$x_1'(t)=ie^{it}[1+o(1)], \quad x_2'(t)=-ie^{-it}[1+o(1)].$$

Suppose that $q(t)\to c>0$ as $t\to\infty$ and $\int^{\infty}|q'(s)|\,ds<\infty$. Then $x(t)$ and $x'(t)$ are bounded for every solution $x(t)$ of (8). The same is true if $q(t)$ is a positive periodic function of period ω such that $\omega\int_0^{\omega}q(s)\,ds\leqslant 4$. If $q(t)$ is negative, then (8) always has both bounded and unbounded monotone solutions.

The number of linearly independent solutions $x(t)$ of (8) satisfying $\int^{\infty}|x(s)|^2\,ds<\infty$ plays an important role in †eigenvalue problems. It is known that the ordinary differential operator $l[x]=x''+q(t)x$ is of †limit point type at infinity if there exist a positive function $M(t)$ and positive constants k_1, k_2 such that $q(t)\leqslant k_1M(t)$, $|M'(t)M^{-3/2}(t)|\leqslant k_2$, and $\int^{\infty}M^{-1/2}(s)\,ds=\infty$, and that $l[x]$ is of †limit circle type at infinity if $q(t)>0$, $\int^{\infty}q^{-1/2}(s)\,ds=\infty$, and $\int^{\infty}|[q^{-3/2}(s)q'(s)]'+(1/4)q^{-5/2}(s)q'^2(s)|\,ds<\infty$.

Finally consider the nonlinear equation

$$x''+q(t)|x|^{\gamma}\operatorname{sgn}x=0, \tag{9}$$

where γ is a positive constant and $q(t)$ is a positive function. If $q'(t)\geqslant 0$, then all solutions of (9) are bounded; if either $q'(t)\geqslant 0$ and $\lim q(t)<\infty$ or $q'(t)\leqslant 0$ and $\lim q(t)>0$, then all solutions $x(t)$ of (9) are bounded together with their derivatives $x'(t)$; and if $q'(t)\geqslant 0$, $\lim q(t)=\infty$ and either $q''(t)\geqslant 0$ or $q''(t)\leqslant 0$, then all solutions of (9) converge to zero as $t\to\infty$.

Equation (9) is said to be **oscillatory** if every solution of (9) that is continuable to $t=\infty$ has arbitrarily large zeros. If (9) is oscillatory and if $q_1(t)\geqslant q(t)$, then the equation $x''+q_1(t)|x|^{\gamma}\operatorname{sgn}x=0$ is also oscillatory. When $\gamma=1$, (9) is oscillatory if $q(t)\geqslant(1+\varepsilon)/4t^2$ for some $\varepsilon>0$, and is not oscillatory if $q(t)\leqslant 1/4t^2$. A necessary and sufficient condition for equation (9) with $\gamma\neq 1$ to be oscillatory is as follows: $\int^{\infty}sq(s)\,ds=\infty$ if $\gamma>1$; $\int^{\infty}s^{\gamma}q(s)\,ds=\infty$ if $0<\gamma<1$.

References

[1] M. Hukuhara, Sur les points singuliers des équations différentielles linéaires; domaine réel, J. Fac. Sci. Hokkaido Univ., (I) 2 (1934), 13–88.

[2] R. Bellman, Stability theory of differential equations, McGraw-Hill, 1953.

[3] L. Cesari, Asymptotic behavior and stability problems in ordinary differential equations, Springer, third edition, 1970.

[4] W. A. Coppel, Stability and asymptotic behavior of differential equations, Heath, 1965.

[5] P. Hartman, Ordinary differential equations, Wiley, 1964.

[6] E. A. Coddington and N. Levinson, Theory of ordinary differential equations, McGraw-Hill, 1955.

[7] J. S. W. Wong, On the generalized Emden-Fowler equation, SIAM Rev., 17 (1975), 339–360.

315 (XIII.4) Ordinary Differential Equations (Boundary Value Problems)

A. General Remarks

Consider the differential equation in the real variable x

$$f(x,y,y',\ldots,y^{(n)})=0. \tag{1}$$

Let $a_1,\ldots,a_k$ be points in an interval $I\subset\mathbf{R}$ and consider several relations between nk values $y(a_i), y'(a_i),\ldots,y^{(n-1)}(a_i)$, $i=1,\ldots,k$. The problem of finding solutions of (1) satisfying these relations is called a **boundary value problem** of (1), and the relations considered are called **boundary conditions**. When $k=2$ and a_1, a_2 are the endpoints of I, the problem, called a **two-point boundary value problem**, has been a main subject of study. We can consider boundary value problems in the same way for systems of differential equations.

B. Linear Differential Equations

Consider a linear ordinary differential operator L defined by

$$L[y]=p_0(x)y^{(n)}+p_1(x)y^{(n-1)}+\ldots+p_n(x)y,$$

where $p_k(x)$ is a complex-valued function of class C^{n-k} defined on a compact interval $a\leqslant x\leqslant b$ and $p_0(x)\neq 0$ for any $x\in[a,b]$. We define a system of **linear boundary operators** $U_1,\ldots,U_m$ by

$$U_i[y]=\sum_{j=1}^{n}M_{ij}y^{(j-1)}(a)+\sum_{j=1}^{n}N_{ij}y^{(j-1)}(b).$$

Given a function $f(x)$ and complex constants $\gamma_1,\ldots,\gamma_m$, the linear boundary value problem defined by

$$L[y]=f(x),\qquad U_i[y]=\gamma_i,\qquad i=1,\ldots,n,\qquad (2)$$

is a two-point boundary value problem. When $f(x)\equiv 0$, $\gamma_i=0$, $i=1,\ldots,n$, the problem is called **homogeneous**; otherwise it is called **inhomogeneous**. Let $L^*[y]$ be a formally †adjoint differential operator of $L[y]$. A set of m^* linear boundary conditions $U_i^*[y]=0$, $i=1,\ldots,m^*$, is said to be an **adjoint boundary condition** of $U_i[y]=0$, $i=1,\ldots,m$, if for any function y of class C^n satisfying $U_i[y]=0$, $i=1,\ldots,m$, and any function y^* of class C^n satisfying $U_i^*[y^*]=0$, $i=1,\ldots,m^*$, we have $\int_a^b L[y]\overline{y^*}\,dx=\int_b^a y\overline{L^*[y^*]}\,dx$. The boundary value problem

$$L^*[y]=0,\qquad U_i^*[y]=0,\qquad i=1,\ldots,m^*,\qquad (3)$$

is said to be an **adjoint boundary value problem** of

$$L[y]=0,\qquad U_i[y]=0,\qquad i=1,\ldots,m.\qquad (4)$$

We say that the problem (4) is **self-adjoint** if $L[y]=L^*[y]$ and the conditions $U_i[y]=0$, $i=1,\ldots,m$, are equivalent to the conditions $U_i^*[y]=0$, $i=1,\ldots,m^*$.

The boundary value problem containing a parameter λ

$$L[y]=\lambda y,\qquad U_i[y]=0,\qquad i=1,\ldots,n,\qquad (5)$$

admits nontrivial solutions only for special values of λ. Such values of λ are called the **eigenvalues** (or **proper values**) of (5), and the corresponding solutions $\not\equiv 0$ are called the **eigenfunctions** (or **proper functions**) of (5). For any value of λ that is not an eigenvalue, there exists a unique function $G(x,\xi,\lambda)$ such that the conditions $L[y]=\lambda y+f$, $U_i[y]=0$ are equivalent to $y=\int_a^b G(x,\xi,\lambda)f(\xi)\,d\xi$. The function $G(x,\xi,\lambda)$ is called the **Green's function** of (5). If $\lambda=0$ is not an eigenvalue, then (5) is equivalent to

$$y(x)=\lambda\int_a^b G(x,\xi)y(\xi)\,d\xi,$$

where $G(x,\xi)=G(x,\xi,0)$. For
function $G(x,\xi,\lambda)$ of (5) and th
tion $G^*(x,\xi,\lambda)$ of $L^*[y]=\lambda y$,
have the relation $G(x,\xi,\lambda)=\bar{G}$
the assumption that (5) is self-
the following four proposition
has only real eigenvalues whic
or countably infinite discrete s
functions corresponding to tw
values are orthogonal to each
is an †orthonormal set of eiger
that no eigenfunction is linear
of $\{\varphi_n\}$, then the system $\{\varphi_n\}$ i
orthonormal set in the Hilbert
consisting of functions that ar
grable on (a,b), and hence for
expansion $f=a_1\varphi_1+a_2\varphi_2+\ldots$
$L_2(a,b)$ the †Parseval equality
if f is a function of class C^n sat
0, then the Fourier expansion
uniformly to f on $[a,b]$.

The boundary value proble
order equation

$$(p(x)y')'+(q(x)+\lambda r(x))y=0,$$

$$\alpha y(a)+\beta y'(a)=0,\qquad \gamma y(b)+\delta y'$$

is called a **Sturm-Liouville prol**
r are real-valued functions defi
and α, β, γ, are real constants.
q, r are continuous and $p(x)>$
$[a,b]$. Then (i) the eigenvalues
ing sequence tending to $+\infty$; (
tion $\varphi_n(x)$ associated with λ_n h
zeros in $a<x<b$, and there ex
adjacent zeros of $\varphi_n(x)$ a zero
(iii) the set of eigenfunctions is
set on $[a,b]$ with weight functi

$$\int_a^b r(x)\varphi_m(x)\varphi_n(x)\,dx=0,\qquad m$$

When the coefficients $p_0,\ldots,$
defined in an open interval $-$
$b\leqslant\infty$ and p_k is of class C^{n-k}, L
natural way operators in the H
consisting of functions that are
grable in $a<x<b$, and the gen
based on operator theory in H
(→ 390 Spectral Analysis of O

C. Nonlinear Differential Equ

Boundary value problems for
ferential equations are very dif
sults are obtained only for equ
form.

Consider, for example, the s
equation

$$y''=f(x,y,y')$$

and boundary conditions $y(a)=A$, $y(b)=B$. The following theorem has been proved: Suppose that $f(x,y,y')$ is continuous for $a\leqslant x\leqslant b$, $\underline{\omega}(x)\leqslant y\leqslant\bar{\omega}(x)$, $-\infty<y'<+\infty$, and $|f(x,y,y')|\leqslant M(1+y'^2)$; $\underline{\omega}''(x)>f(x,\underline{\omega}(x),\underline{\omega}'(x))$ and $\bar{\omega}''<f(x,\bar{\omega}(x),\bar{\omega}'(x))$ for $a\leqslant x\leqslant b$; and $\underline{\omega}(a)\leqslant A\leqslant\bar{\omega}(a)$ and $\underline{\omega}(b)\leqslant B\leqslant\bar{\omega}(b)$. Then (6) admits a solution $y(x)$ such that $y(a)=A$, $y(b)=B$, and $\underline{\omega}(x)\leqslant y(x)\leqslant\bar{\omega}(x)$ for $a\leqslant x\leqslant b$. If in addition f is an increasing function with respect to y, the solution is unique. Moreover, under suitable conditions, the solution is obtainable by the method of successive approximations.

The boundary value problem

$$y'''+2yy''+2\lambda(k^2-y'^2)=0,$$

$$y(0)=y'(0)=0, \qquad y'(x)\to k\ (x\to\infty),$$

where λ and k are constants, appears in the theory of fluid dynamics. It is known that if $\lambda\geqslant 0$, the problem has a solution, and that if $0\leqslant\lambda\leqslant 1$, the solution is unique.

Consider the system of differential equations

$$y_j'=f_j(x,y_1,\ldots,y_n), \qquad j=1,\ldots,n.$$

The problem of finding a solution such that $y_j(a_j)=b_j$, $j=1,\ldots,n$, called **Hukuhara's problem**, reduces to the initial value problem when the a_j coincide. The problem of solving

$$y^{(n)}=f(x,y,y',\ldots,y^{(n-1)}),$$

$$y(a_j)=b_j, \qquad j=1,\ldots,n,$$

is reduced to Hukuhara's problem by a suitable change of variables. The following result is a generalization of †Perron's theorem: Let $\underline{\omega}_j(x)$, $\bar{\omega}_j(x)$, $j=1,\ldots,n$, be continuous and right and left differentiable functions and $\underline{\omega}_j(x)\leqslant\bar{\omega}_j(x)$ for $a\leqslant x\leqslant b$. Suppose that the $f_j(x,y_1,\ldots,y_n)$ are continuous for $\alpha\leqslant x\leqslant\beta$ and $\underline{\omega}_k(x)\leqslant y_k\leqslant\bar{\omega}_k(x)$, $k=1,\ldots,n$; satisfy $(x-a_j)(D^{\pm}\bar{\omega}_j(x)-f_j(x,y_1,\ldots,y_n))\geqslant 0$ for $y_j=\bar{\omega}_j(x)$ and $\underline{\omega}_k(x)\leqslant y_k\leqslant\bar{\omega}_k(x)$, $k\neq j$; satisfy $(x-a_j)(D^{\pm}\underline{\omega}_j(x)-f_j(x,y_1,\ldots,y_n))\leqslant 0$ for $y_j=\underline{\omega}_j(x)$ and $\underline{\omega}_k(x)\leqslant y_k\leqslant\bar{\omega}_k(x)$, $k\neq j$; and satisfy $\underline{\omega}_j(a_j)\leqslant b_j\leqslant\bar{\omega}_j(a_j)$. Then there exists a solution $y(x)$ such that $y_j(a_j)=b_j$ and $\underline{\omega}_j(x)\leqslant y(x)\leqslant\bar{\omega}_j(x)$. This theory was applied by M. Hukuhara to the study of singular points of ordinary differential equations.

References

[1] E. A. Coddington and N. Levinson, Theory of ordinary differential equations, McGraw-Hill, 1955.
[2] P. Hartman, Ordinary differential equations, Wiley, 1964.
[3] E. Hille, Lectures on ordinary differential equations, Addison-Wesley, 1969.
[4] M. A. Naĭmark (Neumark), Linear differential operators I, II, Ungar, 1967, 1968. (Original in Russian, 1954.)
[5] N. Dunford and J. T. Schwartz, Linear operators II, Interscience, 1963.
[6] B. M. Levitan and I. S. Sargsyan (Sargsjan), Introduction to spectral theory: Self-adjoint ordinary differential operators, Amer. Math. Soc. Transl. of Math. Monographs, 1975. (Original in Russian, 1970.)

316 (XIII.3) Ordinary Differential Equations (Initial Value Problems)

A. General Remarks

Consider a system of ordinary differential equations

$$dy_i/dx=f_i(x,y_1,\ldots,y_n), \qquad i=1,\ldots,n. \tag{1}$$

A. L. Cauchy first gave a rigorous proof for the existence and uniqueness of solutions: If f_i, $i=1,\ldots,n$, and their derivatives $\partial f_i/\partial y_k$ are continuous in a neighborhood of a point $(a,b_1,\ldots,b_n)$, then there exists a unique solution of (1) satisfying the conditions $y_i(a)=b_i$, $i=1,\ldots,n$. These conditions are called **initial conditions**, and the values $a,b_1,\ldots,b_n$ **initial values**. The problem of finding solutions that satisfy initial conditions is called an **initial value problem** (or **Cauchy problem**). If we consider $(x,y_1,\ldots,y_n)$ as the coordinates of a point in the $(n+1)$-dimensional space $\mathbf{R}^{n+1}$, then a solution of (1) represents a curve in this space called a **solution curve** (or **integral curve**). The statement that a solution satisfies initial conditions $y_i(a)=b_i$, $i=1,\ldots,n$, means that the integral curve represented by it passes through the point $(a,b_1,\ldots,b_n)$.

Since, in general, we can transform a differential equation of higher order into a system of differential equations of the form (1) by introducing new dependent variables, all definitions and theorems concerning the system (1) can be interpreted as applying to a higher-order equation. For example, for the equation $y^{(n)}=f(x,y,y',\ldots,y^{(n-1)})$ the conditions $y(a)=b$, $y'(a)=b',\ldots,y^{(n-1)}(a)=b^{(n-1)}$ constitute initial conditions, and the values $a,b,b',\ldots,b^{(n-1)}$ are initial values. If f and its derivatives $\partial f/\partial y^{(i)}$ are continuous, then there exists a unique solution satisfying given initial conditions.

Suppose that the f_i are continuous. Then a system of functions $(y_1(x),\ldots,y_n(x))$ is a solu-

tion of (1) if and only if

$$y_i(x)=b_i+\int_a^x f_i(x,y_1(x),\dots,y_n(x))\,dx,\qquad i=1,\dots,n.$$

When the f_i are not continuous, we define $(y_1(x),\dots,y_n(x))$ to be a solution of (1) for the initial value problem $y_i(a)=b_i$ if $(y_1(x),\dots,y_n(x))$ satisfied the integral equation just given.

We use the vectorial notation: $\mathbf{y}=(y_1,\dots y_n)$, $\mathbf{f}=(f_1,\dots,f_n)$ together with $\|\mathbf{y}\|^2=y_1^2+\dots+y_n^2$. The equations (1) are then written as the single equation

$$\mathbf{y}'=\mathbf{f}(x,\mathbf{y}).$$

B. Equations in the Real and Complex Domains

We state main theorems for differential equations in the real domain in Sections C–F and in the complex domain in Section G.

C. Existence Theorems

Suppose that $\mathbf{f}(x,\mathbf{y})$ is continuous for $|x-a|\leqslant r$ and $\|\mathbf{y}-\mathbf{b}\|\leqslant\rho$, and that $\|\mathbf{f}(x,\mathbf{y})\|\leqslant M$ there. Then equation (1) admits a solution satisfying $\mathbf{y}(a)=\mathbf{b}$ and defined in an interval $|x-a|\leqslant\min(r,\rho/M)$ **(existence theorem)**. There are two methods of proving this theorem, one using **Cauchy polygons** and one using †fixed-point theorems for function spaces. From this theorem we deduce that if $\mathbf{f}(x,\mathbf{y})$ is continuous in a domain D of $\mathbf{R}^{n+1}$, then there exists a solution curve passing through any point of D. Let $\mathbf{y}=\boldsymbol{\varphi}_1(x)$ and $\mathbf{y}=\boldsymbol{\varphi}_2(x)$ be solutions of (1) defined in the intervals I_1 and I_2, respectively. If $I_1\subset I_2$ and $\boldsymbol{\varphi}_1(x)=\boldsymbol{\varphi}_2(x)$ for $x\in I_1$, we say that $\boldsymbol{\varphi}_2$ is a **prolongation** or **extension of** $\boldsymbol{\varphi}_1$. Given a solution of (1), there exists a nonextendable solution that is an extension of the solution. The solution curve of a nonextendable solution tends to the boundary of D as x tends to any one of the ends of its interval of definition.

D. Uniqueness Theorems

Continuity does not imply uniqueness of the solution. If (1) admits at most one solution satisfying a given condition, we call this condition a **uniqueness condition**. Various kinds of **uniqueness theorems**, which state uniqueness conditions, are known.

The **Lipschitz condition**:

$$\|\mathbf{f}(x,\mathbf{y})-\mathbf{f}(x,\mathbf{z})\|\leqslant L\|\mathbf{y}-\mathbf{z}\|,\quad L>0 \text{ a constant,}$$

is one of the simplest. When $\mathbf{f}$ is continuous and satisfies the Lipschitz condition, the **method of successive approximation**, initiated by C. E. Picard, is often used to prove the existence of solutions. This method is as follows: We choose a suitable function, for example $\mathbf{y}_0(x)\equiv\mathbf{b}$, and then define $\mathbf{y}_k(x)$, $k=1,2,\dots$, recursively by $\mathbf{y}_k(x)=b+\int_a^x\mathbf{f}(x,\mathbf{y}_{k-1}(x))dx$. Then $\{\mathbf{y}_k(x)\}$ is uniformly convergent, and its limit is a solution of (1) satisfying $\mathbf{y}(a)=\mathbf{b}$.

Assuming the continuity of $\mathbf{f}$, H. Okamura gave a necessary and sufficient condition for uniqueness: Suppose that $\mathbf{f}$ is continuous in D. Then a necessary and sufficient condition for there to exist a unique solution curve of (1) going from any point of D to the right is that there exist a C^1-function $\varphi(x,\mathbf{y},\mathbf{z})$ defined for $(x,\mathbf{y},\mathbf{z})$ such that $(x,\mathbf{y})$ and $(x,\mathbf{z})\in D$ and satisfying the conditions $\varphi(x,\mathbf{y},\mathbf{z})=0$ for $\mathbf{y}=\mathbf{z}$, $\varphi(x,\mathbf{y},\mathbf{z})>0$ for $\mathbf{y}\neq\mathbf{z}$, and

$$\frac{\partial\varphi}{\partial x}+\sum_i\frac{\partial\varphi}{\partial y_i}f_i(x,\mathbf{y})+\sum_i\frac{\partial\varphi}{\partial z_i}f_i(x,\mathbf{z})\leqslant 0.$$

E. Perron's Theorem

Consider the scalar equation $y'=f(x,y)$. We have **Perron's theorem**: Let $\underline{\omega}(x)$ and $\bar{\omega}(x)$ be continuous functions that are right differentiable in $\alpha\leqslant x<\beta$ and satisfy $\underline{\omega}(x)\leqslant\bar{\omega}(x)$, and let f be a continuous function defined on $D:\alpha\leqslant x<\beta$, $\underline{\omega}(x)\leqslant y\leqslant\bar{\omega}(x)$. Suppose that $D^+\underline{\omega}(x)\leqslant f(x,\underline{\omega}(x))$ and $D^+\bar{\omega}(x)\geqslant f(x,\bar{\omega}(x))$. ($D^+\omega$ denotes the †right derivative of ω.) Then for any $(a,b)\in D$ there exists a solution defined on $a\leqslant x<\beta$ and satisfying $y(a)=b$. The fact that the interval of definition is $a\leqslant x<\beta$ can be expressed by saying that if we denote the set $\infty<x<\beta$, $|y|<\infty$ by Ω, then D is closed in Ω and there exists, among solution curves going from a point in D to the right, a curve that reaches the boundary of Ω.

Perron's theorem was generalized by M. Hukuhara and M. Nagumo. Let Ω be an open set in $\mathbf{R}^{n+1}$, D a closed set in Ω, and $\mathbf{f}$ a continuous function in D. A necessary and sufficient condition for (1) to admit a solution curve going from any point $(a,\mathbf{b})$ in D to the right is that there exist a sequence of points in D, $\{(a_k,\mathbf{b}_k)\}$, such that $a_k\downarrow a$ and $(\mathbf{b}_k-\mathbf{b})/(a_k-a)\to\mathbf{f}(a,\mathbf{b})$. Moreover, every solution curve is prolonged to the right to the boundary of Ω. Let $S(\mathbf{y})$ be a continuous †subadditive and positively homogeneous function and $\omega(x)$ a function continuous and right differentiable on $\alpha\leqslant x<\beta$. A sufficient condition for $D:\alpha\leqslant x<\beta$, $S(\mathbf{y})\leqslant\omega(x)$ to possess the property in the statement of Perron's theorem is given by $D^+\omega(x)\geqslant S(\mathbf{f}(x,\mathbf{y}))$ for $\|\mathbf{y}\|=\omega(x)$.

A continuous function $\omega(x)$ is said to be a

right majorizing function of (1) with respect to $S(\mathbf{y})$ if for any solution $\varphi(x)$, $S(\varphi(a)) \leqslant \omega(a)$ implies $S(\varphi(x)) \leqslant \omega(x)$ for $x \geqslant a$ if both $S(\varphi(x))$ and $\omega(x)$ are defined. In order for $\omega(x)$ to be a right majorizing function, it suffices that $D^+\omega(x) > S(\mathbf{f}(x,\mathbf{y}))$ for $\|\mathbf{y}\| = \omega(x)$. A function satisfying this inequality is called a **right superior function** of (1) with respect to $S(\mathbf{y})$. If $F(x, S(\mathbf{y})) > S(\mathbf{f}(x,\mathbf{y}))$, then any solution of $y' = F(x,y)$ is a right superior function of (1). Theorems stating such facts are called **comparison theorems**.

If (1) has a unique solution, the condition $D^+\omega(x) \geqslant S(\mathbf{f}(x,\mathbf{y}))$ for $S(\mathbf{y}) = \omega(x)$ implies that $\omega(x)$ is a right majorizing function of (1). Conversely, we can derive from comparison theorems general uniqueness theorems, one of which we state. Suppose that $G(x,y)$ is continuous for $\alpha < x < \beta$ and $0 \leqslant y < r(x)$; $G(x,0) \equiv 0$; a solution of $y' = G(x,y)$ such that $y = o(r(x))$ as $x \to \alpha + 0$ vanishes identically; and finally that $S(\mathbf{f}(x,\mathbf{y}_1) - \mathbf{f}(x,\mathbf{y}_2)) \leqslant G(x, S(\mathbf{y}_1 - \mathbf{y}_2))$. Then for two solutions φ_1, φ_2 of (1) such that $S(\varphi_1 - \varphi_2) = o(r(x))$ as $x \to \alpha + 0$, we have $\varphi_1 \equiv \varphi_2$. Assuming that $\mathbf{f}$ is continuous at $(a, \mathbf{b})$ and taking $y/(x-a)$ as G, we obtain Nagumo's condition $(x-a)S(\mathbf{f}(x,\mathbf{y}_1) - \mathbf{f}(x,\mathbf{y}_2)) \leqslant S(\mathbf{y}_1 - \mathbf{y}_2)$.

G. Peano proved the following theorem: With the same notation and assumption as in Perron's theorem, there exist a **maximum solution** $\bar{\varphi}$ and a **minimum solution** $\underline{\varphi}$ of $y' = f(x,y)$ such that $y(\alpha) = b$ for $\underline{\omega}(\alpha) \leqslant b \leqslant \bar{\omega}(\alpha)$, and such that there exists a solution curve passing through any point in $\alpha \leqslant x < \beta$, $\underline{\varphi}(x) \leqslant y \leqslant \bar{\varphi}(x)$. This theorem was extended by Hukuhara as follows. Suppose that $\mathbf{f}(x,\mathbf{y})$ is continuous and bounded in $D: \alpha \leqslant x \leqslant \beta$, $\|\mathbf{y}\| < \infty$. Let C be a †continuum in D, and let $\mathfrak{F}(C)$ denote the set of solutions intersecting C. Then $\mathfrak{F}(C)$ is a continuum of the †function space $C([\alpha,\beta])$. From this theorem we can deduce the **Kneser-Nagumo theorem**, which says that the intersection of the set of points belonging to the members of $\mathfrak{F}(C)$ and a hyperplane $x = \xi$ is a continuum. It was proved by Hukuhara that if C is in the hyperplane $x = \alpha$, then (1) admits a solution connecting the two hyperplanes $x = \alpha$ and $x = \beta$ and passing through the boundary of the set of points belonging to the members of $\mathfrak{F}(C)$.

F. Equations Containing Parameters

We assume uniqueness of the solution of

$$\mathbf{y}' = \mathbf{f}(x, \mathbf{y}, \boldsymbol{\lambda}), \qquad \boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m), \tag{2}$$

where $\mathbf{f}$ is a continuous function of $(x, \mathbf{y}, \boldsymbol{\lambda})$. Let $\varphi(x, a, \mathbf{b}, \boldsymbol{\lambda})$ denote the solution of (2) satisfying $\mathbf{y}(a) = \mathbf{b}$. Then $\varphi(x, a, \mathbf{b}, \boldsymbol{\lambda})$ is continuous with respect to $(x, a, \mathbf{b}, \boldsymbol{\lambda})$ in its region of definition. If the derivatives $\partial\mathbf{f}/\partial y_k$ are also continuous, then $\varphi(x, a, \mathbf{b}, \boldsymbol{\lambda})$ is a continuously differentiable function of $(x, \mathbf{b})$; $z_{jk} = \partial\varphi_j/\partial b_k$, $j = 1, \ldots, n$, satisfy the system of linear ordinary differential equations and the initial condition

$$\frac{d}{dx} z_{jk} = \sum_{l=1}^{n} \left(\frac{\partial f_j}{\partial y_l}\right) z_{lk}, \qquad z_{jk}(a) = \delta_{jk};$$

and $z_j = \partial\varphi_j/\partial a$, $j = 1, \ldots, n$, satisfy the same system with the initial condition $z_j(a) = -f_j(a, \mathbf{b}, \boldsymbol{\lambda})$, where $(\partial f_j/\partial y_l)$ means $(\partial f_j/\partial y_l)(x, \varphi(x, a, \mathbf{b}, \boldsymbol{\lambda}), \boldsymbol{\lambda})$. If $\mathbf{f}$ further admits continuous derivatives $\partial\mathbf{f}/\partial\lambda_l$, then $\varphi(x, a, \mathbf{b}, \boldsymbol{\lambda})$ is continuously differentiable with respect to λ_l, and moreover, $w_{jl} = \partial\varphi_j/\partial\lambda_l$, $j = 1, \ldots, n$, satisfy the system

$$\frac{d}{dx} w_{jl} = \sum_{k=1}^{n} \left(\frac{\partial f_j}{\partial y_k}\right) w_{kl} + \left(\frac{\partial f_j}{\partial \lambda_l}\right), \qquad w_{jl}(a) = 0.$$

These differential systems are called the **variational equations** of (1).

C. Carathéodory proved the existence of solutions of (1) under the less restrictive assumption that $\mathbf{f}$ is continuous with respect to $\mathbf{y}$ for any fixed x and measurable with respect to x for any fixed $\mathbf{y}$.

Suppose that $\mathbf{f}$ is continuous and satisfies a Lipschitz condition. Let $\mathbf{z}(x)$ be a function such that $\mathbf{z}(a) = \mathbf{b}$ and $\|z_i'(x) - f_i(x, \mathbf{z}(x))\| \leqslant \varepsilon(x)$, and let $\mathbf{y}(x)$ be a solution of (1) such that $\mathbf{y}(a) = \mathbf{b}$. Then we obtain

$$\|z_i(x) - y_i(x)\| \leqslant e^{L|x-a|} \left| \int_a^x \varepsilon(x) e^{-L|x-a|} dx \right|,$$

which gives approximate solutions of (1).

G. Equations in the Complex Domain

We assume that the variables $x, y_1, \ldots, y_n$ all have complex values. We have the following theorem: If $\mathbf{f}$ is holomorphic at $(a, \mathbf{b})$, then (1) has a unique solution that is holomorphic at $x = a$ and takes the value $\mathbf{b}$ at $x = a$. This theorem can be proved by utilizing the method of successive approximations and fixed-point theorems. Cauchy proved the theorem by using **majorant series**. This method, called the **method of majorants**, proceeds for the scalar equations as follows: Let $f(x,y) = \sum a_{jk}(x-a)^j(y-b)^k$ and $y = \sum c_n(x-a)^n$. Substituting the latter series into both members, we can successively determine the coefficients c_n by the method of undetermined coefficients. Assuming that $|f| < M$ for $|x-a| < r$ and $|y-b| < \rho$, consider the solution $Y = \sum C_n(x-a)^n$ of

$$\frac{dY}{dx} = \frac{M}{(1-(x-a)/r)(1-(Y-b)/\rho)}$$

satisfying $Y(a) = b$. We have $C_n \geqslant |c_n|$ for any n,

which shows that $\sum C_n(x-a)^n$ is a majorant series of $\sum c_n(x-a)^n$.

We have the following **uniqueness theorem**: Suppose that $\mathbf{f}(x,\mathbf{y})$ is holomorphic at $(a,\mathbf{b})$. Let C be a curve having the point a as one of its ends and $\boldsymbol{\varphi}(x)$ be a solution with the following properties: $\boldsymbol{\varphi}$ is holomorphic on C except possibly at $x=a$, and there exists a sequence of points on C, $\{a_k\}$, such that $a_k \to a$ and $\boldsymbol{\varphi}(a_k)\to \mathbf{b}$. Then $\boldsymbol{\varphi}(x)$ is holomorphic at a. By a †theorem of identity, the analytic continuation of a solution continues to be a solution if it does not encounter any singularity of $\mathbf{f}$. If a solution $\boldsymbol{\varphi}(x)$ is holomorphic on a smooth curve $x=\chi(t)$, with $0 \leqslant t \leqslant t_0$ and $\chi(0)=a$, and $\boldsymbol{\varphi}(a)=\mathbf{b}$, then $\mathbf{y}=\boldsymbol{\varphi}(\chi(t))$ is a solution for $0 \leqslant t \leqslant t_0$ of

$$\mathbf{y}'=\chi'(x)\mathbf{f}(\chi(t),\mathbf{y}) \tag{3}$$

satisfying $\mathbf{y}(0)=\mathbf{b}$. Conversely, if (3) has a solution $\mathbf{y}=\boldsymbol{\psi}(t)$ defined in $0 \leqslant t \leqslant t_0$ and satisfying $\mathbf{y}(0)=\mathbf{b}$, and if $\mathbf{f}(x,\mathbf{y})$ is holomorphic at $(\chi(t),\boldsymbol{\psi}(t))$ for $0 \leqslant t \leqslant t_0$, then (1) has a solution $\boldsymbol{\varphi}(x)$ holomorphic on C and $\boldsymbol{\psi}(t)=\boldsymbol{\varphi}(\chi(t))$ for $0 \leqslant t \leqslant t_0$.

Suppose that $f=f_1/f_2$, where f_1 and f_2 are holomorphic at (a,b). If $f_1(a,b)\neq 0$, $f_2(a,y)\not\equiv 0$, and $f_2(a,b)=0$, then the equation $y'=f(x,y)$ admits a unique solution such that $y\to b$ as $x\to a$, and this solution can be expanded into a †Puiseux series:

$$y=\sum_{n=0}^{\infty} c_n(x-a)^{n/r}.$$

If $\mathbf{f}$ is holomorphic at $(a,\mathbf{b})$, then the solution $\mathbf{y}=\boldsymbol{\varphi}(x,x_0,\mathbf{y}_0)$ of (1) satisfying $\mathbf{y}(x_0)=\mathbf{y}_0$ is holomorphic with respect to $(x,x_0,\mathbf{y}_0)$ at $(a,a,\mathbf{b})$. If $\mathbf{f}(x,\mathbf{y},\lambda)$ is holomorphic at $(a,\mathbf{b},\lambda_0)$, then the solution of (2), $\mathbf{y}=\boldsymbol{\varphi}(x,x_0,\mathbf{y}_0,\lambda)$, satisfying $\mathbf{y}(x_0)=\mathbf{y}_0$ is holomorphic at $(a,a,\mathbf{b},\lambda_0)$. Suppose that x is a real variable and $\mathbf{y}$ is a complex vector. If $\mathbf{f}$ is continuous with respect to $(x,\mathbf{y})$ and holomorphic with respect to $\mathbf{y}$, then the solution $\boldsymbol{\varphi}(x,x_0,\mathbf{y}_0)$ is holomorphic with respect to $\mathbf{y}_0$. If $\mathbf{f}(x,\mathbf{y},\lambda)$ is continuous with respect to $(x,\mathbf{y},\lambda)$ and holomorphic with respect to $(\mathbf{y},\lambda)$, then $\boldsymbol{\varphi}(x,x_0,\mathbf{y}_0,\lambda)$ is holomorphic with respect to $(\mathbf{y}_0,\lambda)$.

References

[1] E. Kamke, Differentialgleichungen reeller Funktionen, Akademische Verlag, 1930.
[2] E. A. Coddington and N. Levinson, Theory of ordinary differential equations, McGraw-Hill, 1955.
[3] P. Hartman, Ordinary differential equations, Wiley, 1964.
[4] H. Okamura, Condition nécessaire et suffisante remplie par les équations différentielles ordinaires sans points de Peano, Mem. Coll. Sci. Univ. Kyoto, (A) 24 (1941).
[5] O. Perron, Ein neuer Existenz Beweis für die Integrale der Differentialgleichung $y'=f(x,y)$, Math. Ann., 15 (1945).
[6] M. Hukuhara, Sur la théorie des équations différentielles ordinaires, J. Fac. Sci. Univ. Tokyo, sec. I, vol. VII, pt. 5, 1958.
[7] M. Nagumo, Über die Lage der Integralkurven gewöhnlicher Differentialgleichungen, Proc. Phys.-Math. Soc. Japan, 24 (1943).
[8] E. Hille, Ordinary differential equations in the complex domain, Wiley, 1976.
[9] R. Abraham and J. E. Marsden, Foundations of mechanics, Benjamin-Cummings, second edition, 1978.

317 (X.21) Orthogonal Functions

A. Orthogonal Systems

Let (X,μ) be a †measure space. For complex-valued functions f, g on X belonging to the †function space $L_2(X)$, we define the inner product $(f,g)=\int_X f(x)\overline{g(x)}\,d\mu(x)$ and the norm $\|f\|=(f,f)^{1/2}$. If $(f,g)=0$, then we say that f and g are **orthogonal** on X with respect to the measure μ. If X is a subset of a Euclidean space and μ is the †Lebesgue measure m, then we simply say that they are orthogonal. If the measure has a †density function $\varphi(x)$ with respect to the Lebesgue measure and $(f,g)=\int_X f(x)\overline{g(x)}\varphi(x)\,dm(x)=0$, we say that they are orthogonal with respect to the **weight function** $\varphi(x)$. If $\|f\|^2=1$, then f is said to be **normalized**. A set of functions $\{f_n(x)\}$ $(n=1,2,\dots)$ is said to be an **orthogonal system** (or **orthogonal set**), and we write $\{f_n\}\in O(X)$, if any pair of functions in the set are orthogonal. The orthogonal set $\{f_n(x)\}$ is said to be **orthonormal**, and we write $\{f_n\}\in ON(X)$ if each f_n is normalized.

Let $\{f_n\}$ be a set of linearly independent functions in $L_2(X)$, and let R be a subset of $L_2(X)$. If we can approximate any function $f\in R$ arbitrarily closely by a finite linear combination of the $f_n(x)$ with respect to the norm in $L_2(X)$, we say that $\{f_n\}$ is **total** in R. Let $\{f_n\}\in O(X)$. If $(\varphi,f_n)=0$ for all n implies $\varphi(x)=0$ almost everywhere (a.e.), then $\{f_n\}$ is said to be **complete** in $L_2(X)$. An orthogonal system $\{f_n\}$ is complete in $L_2(X)$ if and only if the system is total in $L_2(X)$.

If $\{f_n\}\in O(X)$, then the series $\sum_{n=1}^{\infty} c_n f_n(x)$ is called an **orthogonal series**. If the series converges to $\varphi(x)$ †in the mean of order 2, then $c_n=(\varphi,f_n)/\|f_n\|^2$. We call the c_n $(n=1,2,\dots)$ the

expansion coefficients or **Fourier coefficients** of $\varphi(x)$ with respect to $\{f_n\}$.

If $\{g_n\} \subset L_2(X)$ are linearly independent, we can construct an orthonormal system $\{f_n\}$ by forming suitable linear combinations of the g_n; $\{f_n\}$ spans the same subspace as $\{g_n\}$. For this purpose we set $f_1(x)=g_1(x)/\|g_1\|$, $f_n(x)=\varphi_n(x)/\|\varphi_n\|$, where $\varphi_n(x)=g_n(x)-\sum_{\nu=1}^{n-1}(g_n, f_\nu)f_\nu(x)$, $n \geqslant 2$. This procedure is called **Schmidt orthogonalization** or **Gram-Schmidt orthogonalization.**

If the c_n are Fourier coefficients of $\varphi(x) \in L_2(X)$ with respect to $\{f_n(x)\} \in ON(X)$, then we have the †Bessel inequality $\sum_{n=1}^{\infty}|c_n|^2 \leqslant \|\varphi\|^2$. Equality in the Bessel inequality for all $\varphi \in L_2(X)$ (the †Parseval identity) is equivalent to completeness of $\{f_n\}$ in $L_2(X)$. In this case $\sum_{n=1}^{\infty} c_n f_n(x)$ is called the **orthogonal expansion** of φ with respect to $\{f_n\}$, and conversely, for any sequence $\{c_n\}$ such that $\sum |c_n|^2 < \infty$, there is a function $\varphi \in L_2(X)$ that has the c_n as its Fourier coefficients, and

$$\sum_{n=1}^{\infty} |c_n|^2 = \|\varphi\|^2,$$

$$\varphi(x) = \underset{n\to\infty}{\text{l.i.m.}}\left(\sum_{m=1}^{n} c_m f_m(x)\right).$$

This is called the †**Riesz-Fischer theorem.**

B. Orthogonal Systems on the Real Line

We assume that X is a finite interval (a, b) and that functions on X are real-valued. We write $O(a,b)$ or $ON(a,b)$ instead of $O(X)$, $ON(X)$.
(1) If $\{f_n\} \in ON(a,b)$, $|f_n(x)| \leqslant M$ (const.), and $\sum c_n f_n(x)$ converges †a.e., then $c_n \to 0$ as $n \to \infty$.
(2) We can construct a complete orthonormal system $\{f_n(x)\}$ and a function $\varphi(x) \in L_1(a,b)$ such that its orthogonal expansion $\sum c_n f_n(x)$ diverges everywhere. (3) If $\{f_n(x)\} \in ON(a,b)$ and $\sum c_n^2 \log^2 n < \infty$, then $\sum c_n f_n(x)$ converges a.e. The factor $\log^2 n$ cannot be replaced by any other monotone increasing factor $\omega(n)$ satisfying $0 \leqslant \omega(n) = o(\log^2 n)$ (**Rademacher-Men'shov theorem**). K. Tandori proved that if $c_n \downarrow 0$ and $\sum c_n \varphi_n$ converges a.e. for any orthonormal system $\{\varphi_n\}$, then $\sum |c_n|^2 \log^2 n < \infty$.
(4) If the orthogonal expansion of a function $\varphi \in L_2(a,b)$ is †summable by Abel's method on a set E, then it is †$(C,1)$-summable a.e. on E. $(C,1)$-summability a.e. of the orthogonal expansion of a function of $\varphi \in L_2(a,b)$ is equivalent to convergence a.e. of the partial sums $s_{2^n}(x)$ $(n=1,2,\ldots)$ of its expansion. (5) Suppose that $\{f_n(x)\} \in ON(a,b)$, $|f_n(x)| \leqslant M$. Then: (i) If the a_n are the Fourier coefficients of $\varphi(x)$ with respect to $\{f_n(x)\}$, then

$$\left(\sum_{n=1}^{\infty} |a_n|^{p'}\right)^{1/p'} \leqslant M^{(2-p)/p}\left(\int_a^b |\varphi(x)|^p dx\right)^{1/p},$$

where $1 < p \leqslant 2$, $1/p + 1/p' = 1$. Conversely, if $(\sum |a_n|^p)^{1/p} < \infty$ $(1 < p \leqslant 2)$, there exists a function $\varphi(x)$ which has the a_n as its Fourier coefficients and such that

$$\left(\int_a^b |\varphi(x)|^{p'} dx\right)^{1/p'} \leqslant M^{(2-p)/p}\left(\sum_{n=1}^{\infty} |a_n|^p\right)^{1/p}$$

(**F. Riesz's theorem**). When the orthonormal system is the trigonometric system, this is called the **Hausdorff-Young theorem.** (ii) Let $\{a_n^*\}$ be the decreasing rearrangement of $\{|a_n|\}$; then

$$\sum_{n=1}^{\infty} a_n^{*p} n^{p-2} \leqslant A_p \int_a^b |\varphi(x)|^p dx \qquad (1 < p \leqslant 2).$$

If $q \geqslant 2$ and $\sum a_n^{*q} n^{q-2} < \infty$, then there exists a function $\varphi(x)$ which has the a_n as its Fourier coefficients and such that

$$\int_a^b |\varphi(x)|^q dx \leqslant A_q \sum_{n=1}^{\infty} a_n^{*q} n^{q-2}$$

(**Paley's theorem**). When the system is trigonometric, this is called the **Hardy-Littlewood theorem.** (6) If for some positive ε we have $\sum |c_n|^{2-\varepsilon} < \infty$, then $\sum c_n f_n(x)$ converges a.e.
(7) If we set $s^*(x) = \sup_n |\sum_{\nu=1}^{n} c_\nu f_\nu(x)|$, then $\|s^*\|_q \leqslant A_q (\sum c_\nu^{*q} \nu^{q-2})^{1/q}$ $(q > 2)$, where $\{c_n^*\}$ is the decreasing rearrangement of $\{|c_n|\}$.

C. Examples of Orthogonal Systems

(1) $\{\cos nx\} \in O(0,\pi)$, $\{\sin nx\} \in O(0,\pi)$. (2) $\{1, \cos nx, \sin nx\} \in O(0, 2\pi)$ (→ 159 Fourier Series). (3) Suppose that $A(x)$ is positive and continuous, and let $y_n(x)$ be solutions of $y''(x) + \lambda_n A(x) y(x) = 0$ satisfying the condition $y_n(a) = y_n(b) = 0$, where λ_n is any †eigenvalue. Then $\{\sqrt{A(x)}\, y_n(x)\} \in O(a,b)$ (for orthogonality of eigenfunctions → 315 Ordinary Differential Equations (Boundary Value Problems) B). (4) Set $r_n(x) = -1$ or 1 according as the nth digit of the binary expansion of x $(0 < x \leqslant 1)$ is 1 or 0, and $r_n(x) = 0$ if x is expandable in two ways. Then $\{r_n(x)\} \in O(0,1)$. This is called **Rademacher's system of orthogonal functions.** The system is not complete, but it is interpreted as a †sample space of coin tossing. Rademacher's system is useful for constructing various counter-examples. (5) Let the binary expansion of n be $n = 2^{\nu_1} + 2^{\nu_1} + \ldots + 2^{\nu_p}$ $(\nu_1 < \nu_2 < \ldots < \nu_p)$, and set $w_n(x) = r_{\nu_1+1}(x) r_{\nu_2+1}(x) \ldots r_{\nu_p+1}(x)$. Then $\{w_n(x)\}$ is a complete orthonormal system called **Walsh's system of orthogonal functions.** This system is interpreted as a system of characters of the group of binary numbers, and there are many theorems for this system analogous to those for the trigono-

metric system. (6) In the interval [0, 1], set

$$\chi_m^k(x)=\sqrt{2^m},\quad x\in((k-1)/2^m,\ (k-1/2)/2^m)$$
$$=-\sqrt{2^m},\quad x\in((k-1/2)/2^m,\ k/2^m)$$
$$=0,\quad x\in((l-1)/2^m,\ l/2^m),\quad l\neq k,\ 1\leqslant l\leqslant 2^m.$$

The orthonormal system $\chi_m^k(x)$ ($1\leqslant k\leqslant 2^m$, $1\leqslant m$) is called **Haar's system of orthogonal functions**. The Haar expansion of the continuous function $f(x)$ converges to $f(x)$ uniformly.

D. Orthogonal Polynomials (→ Appendix A, Table 20)

Suppose that we are given a weight function $\varphi(x)\geqslant 0$ ($\varphi(x)>0$ a.e.) defined on (a,b) and that the inner product of functions f, g on (a,b) is defined by $(f,g)=\int_a^b f(x)g(x)\varphi(x)dx$. We orthogonalize $\{x^n\}$ by Schmidt orthogonalization and obtain polynomials $p_n(x)$ of degree n. Here the sign of $p_n(x)$ can be determined so that the sign of the coefficient of the highest power of x is positive. We call $\{p_n(x)\}$ **the system of orthogonal polynomials** belonging to the weight function $\varphi(x)$. This system is complete in $L_2^{(\varphi)}(a,b)$, which is defined to be the space of functions f such that $\int_a^b|f(x)|^2\varphi(x)dx<\infty$. In other words, the system $\{\sqrt{\varphi(x)}p_n(x)\}$ is a complete orthonormal system in the ordinary $L_2(a,b)$ space. Concerning the convergence problem of the orthogonal expansion by $\{p_n(x)\}$, the **Christoffel-Darboux formula**

$$\sum_{k=0}^{n}p_k(t)p_k(x)=\frac{\alpha_n}{\alpha_{n+1}}\cdot\frac{p_n(x)p_{n+1}(t)-p_n(t)p_{n+1}(x)}{t-x}$$

plays an important role.

Several important special functions in classical mathematical physics are given by orthogonal polynomials:

(1) Setting $\varphi(x)=(1-x)^\alpha(1+x)^\beta$ ($\alpha>-1,\beta>-1$) in $[-1,1]$, we get the **Jacobi polynomials**, although they are sometimes defined in [0, 1] with respect to $\varphi(x)=x^\alpha(1-x)^\beta$ (→ Appendix A, Table 20.V). If we set $\alpha=\beta$ in the Jacobi polynomials, we get the **ultraspherical polynomials** (or **Gegenbauer polynomials**) (→ Appendix A, Table 20.I). Furthermore, if $\alpha=\beta=0$, then we get the †Legendre polynomials, and if $\alpha=\beta=-1/2$, we get the **Chebyshev polynomials** $T_n(x)=\cos(n\arccos x)$. The $T_n(x)$ also appear in the best approximation problem (→ Appendix A, Table 20.II; 336 Polynomial Approximation).

(2) If we set $\varphi(x)=x^\alpha e^{-x}$ in $(0,\infty)$, we get the **Sonine polynomials** (or **associated Laguerre polynomials**) with appropriate constant factors. If α is a positive integer m, we get

$$S_n^{(m)}(x)=x^{-m}e^x\left(\frac{d^n}{dx^n}(x^{n+m}e^{-x})\right).$$

The particular case $m=0$ gives the **Laguerre polynomials**. In this case, however, it is customary to normalize them as $L_n(x)=(e^x/n!)(d^n/dx^n)(x^ne^{-x})$ (→ Appendix A, Table 20.VI). Laguerre polynomials are used in †numerical integrations of a Gaussian type in $(0,\infty)$. Furthermore, associated Laguerre polynomials appear in the solutions of the Schrödinger equation for the behavior of hydrogen atoms. This system of orthogonal polynomials is useful in the expansions of approximate eigenfunctions of atoms analogous to hydrogen, velocity distribution functions of molecules in gas theory, and so on.

(3) Setting $\varphi(x)=e^{-x^2}$ (or $e^{-x^2/2}$) in $(-\infty,\infty)$, we get **Hermite polynomials** $H_n(x)=(-1)^ne^{x^2}(d^ne^{-x^2}/dx^n)$, modulo constant factors (→ Appendix A, Table 20.VI). Hermite polynomials are special cases of parabolic cylinder functions (→ 167 Functions of Confluent Type). These polynomials appear as eigenfunctions of the Schrödinger equation for harmonic oscillators. They are also connected with probability integrals and are used in mathematical statistics.

(4) Replacing the integral by a finite sum $\sum_{m=0}^{n}f(m)\overline{g(m)}$ in the definition of inner product, we get so-called **orthogonality for a finite sum**. (Regarding orthogonal polynomials with respect to a finite sum (→ Appendix A, Table 20.VII) and their application to the mean square approximation → 19 Analog Computation F.) Since orthogonal polynomials with respect to a finite sum are often called simply orthogonal polynomials by engineers, one must be careful not to confuse these with the ordinary ones.

References

[1] S. Kaczmarz and H. Steinhaus, Theorie der Orthogonalreihen, Warsaw, 1935 (Chelsea, 1951).
[2] G. Szegö, Orthogonal polynomials, Amer. Math. Soc. Colloq. Publ., 1939.
[3] R. Courant and D. Hilbert, Methods of mathematical physics I, Interscience, 1953.
[4] F. G. Tricomi, Vorlesungen über Orthogonalreihen, Springer, 1955.
[5] G. Alexits, Convergence problems of orthogonal series, Pergamon, 1961, revised translation of the German edition of 1960.
[6] G. Sansone, Orthogonal functions, Interscience, revised English edition, 1959.

318 (XX.14) Oscillations

A. General Remarks

A **vibration** or **oscillation** is a phenomenon that repeats periodically, either exactly or approximately. Exactly periodic oscillations are studied in the theory of †periodic solutions of differential equations. The †period of a solution $f(t)$ is called the **period** of the oscillation, and its reciprocal the **frequency**. The difference between the greatest and least values of $f(t)$ (globally or in an interval) is the **amplitude**. The theory of vibrations has its origin in the study of mechanical vibrations, but its nomenclature has been used also for electric circuits. As examples of practical applications of the theory of oscillations, we mention, in engineering, the prevention of vibrations and the generation of stable sustained oscillations, and in geophysics, investigations concerning the free oscillation of the earth, the existence of which has recently been confirmed.

B. Linear Oscillation

Periodic solutions of †linear differential equations have been studied in detail for a long time. Perhaps the simplest case of such an oscillation is represented by the differential equation

$$d^2x/dt^2+n^2x=0, \qquad (1)$$

where the **restitutive force** is proportional to the displacement from the equilibrium position. Typical examples are the free vibration of a simple pendulum with small amplitude and an electric circuit composed of a self-inductance and capacity (without resistance). The solution is given by $x=A\cos(nt+\alpha)$. This is called **harmonic oscillation** or **simple harmonic motion**. Here the amplitude is A, the period is $2\pi/n$, n is the **circular frequency**, and α is the **initial phase**.

A system of m †degrees of freedom $(x_1, \ldots, x_m)$ is said to be in free harmonic oscillation if the coordinates can be expressed as

$$x_i=\sum_{k=1}^{m} A_{ik}\cos(n_kt+\alpha_k), \qquad i=1,2,\ldots,m.$$

Each of these simple harmonic oscillations is called a **normal vibration**. As a limiting case, where the number of degrees of freedom is infinite, we have the vibration of a string:

$$\frac{\partial^2u(x,t)}{\partial t^2}=n^2\frac{\partial^2u(x,t)}{\partial x^2},$$

$$u(0,t)=u(l,t)=0.$$

The solution is given by a series

$$\sum_k A_k\sin(k\pi x/l)\cos(k\pi nt/l),$$

which is just the superposition of the fundamental vibration (corresponding to $k=1$) and simply harmonic motions of frequencies equal to multiples of the fundamental frequency.

If a resisting force proportional to the velocity is acting, the equation becomes

$$d^2x/dt^2+2\varepsilon\,dx/dt+n^2x=0, \qquad n>\varepsilon, \qquad (2)$$

whose solution

$$x=Ae^{-\varepsilon t}\cos(\sigma t+\alpha), \qquad \sigma=\sqrt{n^2-\varepsilon^2} \qquad (3)$$

is not periodic. However, x becomes zero at a fixed interval π/σ, and the extremal values in the intervals decrease to zero in a geometrical progression with the common ratio $v=\exp(-\pi\varepsilon/\sigma)$. This phenomenon is called **damped oscillation** with **damping ratio** v and **logarithmic decrement** $\log v=-\pi\varepsilon/\sigma$. In this case, too, $2\pi/\sigma$ is called the period.

When a driving force term $\varphi(t)$ is present in the right-hand side of (2), the solution takes on the additional term

$$\frac{e^{-\varepsilon t}}{\sigma}\left(\sin\sigma t\int\varphi(t)e^{\varepsilon t}\cos\sigma t\,dt\right.$$
$$\left.-\cos\sigma t\int\varphi(t)e^{\varepsilon t}\sin\sigma t\,dt\right),$$

which represents the **forced oscillation** due to $\varphi(t)$.

If $\varepsilon<0$ in (2) (**negative resistance**), the solution (3) increases in amplitude, so that a small disturbance is amplified, resulting in an automatic generation of oscillation. This phenomenon is called **self-excited vibration**. Besides being caused by some special kinds of circuit elements (e.g., tunnel diodes), such a situation often occurs when the vibrating system has time delay characteristics (→ 163 Functional-Differential Equations).

Among sustained vibrations, other than forced oscillations and self-excited vibrations, are the **parametrically sustained vibrations** caused by periodic variation of a parameter of the vibrating system. Electric wires and pantographs for use in high-speed electric railways must be designed to prevent unwanted parametrically sustained vibrations. On the other hand, a parametron is an electric element utilizing parametrically sustained vibration.

C. Nonlinear Oscillation

Actual vibrating systems contain more or less nonlinear elements, which give rise to various kinds of oscillations different from those described by the linear theory (→ 290 Non-

linear Oscillation). For example, $d^2x/dt^2 - \varepsilon(1-x^2)dx/dt + x = 0$ ($\varepsilon > 0$) represents a stable sustained oscillation such that for large values of ε, two nearly stationary states occur alternately, the transition from one to the other taking place abruptly. This is called **relaxation oscillation**.

References

[1] Lord Rayleigh, The theory of sound, Macmillan, second revised edition, I, 1894; II, 1896 (Dover, 1945).
[2] A. A. Andronov and C. E. Chaĭkin, Theory of oscillations, Princeton Univ. Press, 1949. (Original in Russian, 1937.)
[3] L. S. Pontryagin, Ordinary differential equations, Addison-Wesley, 1962. (Original in Russian, 1961.)
Also → references to 287 Nonlinear Lattice Dynamics, 290 Nonlinear Oscillation, 291 Nonlinear Problems.

P

319 (I.3) Paradoxes

A. General Remarks

A statement that is apparently absurd but not easily disproved is called a **paradox**. A contradiction between a proposition and its negation is called an **antinomy** if both statements can be supported by logically equivalent reasoning. In practical use, however, "paradox" and "antinomy" often mean the same thing.

B. Paradoxes in Set Theory

(1) The Russell Paradox (1903). We classify sets into two kinds as follows: Any set that does not contain itself as an element is called a **set of the first kind**, and any set that contains itself as an element is called a **set of the second kind**. Every set is either a set of the first kind or of the second kind. Denote the set of all sets of the first kind by M. If M is a set of the first kind, M cannot be an element of M. But if M is of the first kind, then M must be an element of M, by definition. This is contradictory. On the other hand, if M is a set of the second kind, M must be an element of M; but since M is an element of M, M is a set of the first kind, so M cannot be an element of M, by definition. This is contradictory.

Since the kind of reasoning employed in this paradox is very simple and is often utilized in mathematics, it became popular in set theory. To remove this paradox from set theory, Russell suggested †ramified type theory. If we adopt this theory, however, it becomes very hard to develop even an ordinary theory of real numbers (→ 156 Foundations of Mathematics). On the other hand, this paradox, together with the Burali-Forti paradox, indicates that the definition of a set should be restrictive. This realization led to the development of †axiomatic set theory.

(2) The Burali-Forti Paradox (1897). Let $W=\{0,1,2,\ldots,\omega,\ldots\}$ be the †well-ordered set (→ 312 Ordinal Numbers A) of all †ordinal numbers. Let Ω be the ordinal number of W. Then every ordinal number, being an element of W, is less than Ω. But Ω is an ordinal number. Hence, $\Omega<\Omega$. This is contradictory.

(3) The Richard Paradox (1905). The expressions in the English language can be enumerated by the device that is applied to the usual enumeration of the algebraic equations with integral coefficients. From the specified enumeration of all expressions in the English language, by striking out those which do not define a real number in the interval $(0,1]$, we obtain an enumeration of those which do. Consider the following expression: "The greatest real number represented by a proper nonterminating decimal fraction whose nth digit, for any natural number n, is not equal to the nth digit of the nonterminating decimal fraction representing the real number defined by the nth expression in the last-described enumeration." Then we have before us a definition of a real number in the interval $(0,1]$ by means of an expression in the English language. This real number, by its definition, must differ from every real number definable by an expression in the English language. This is contradictory.

The following paradox was given by Berry (1906): "The least natural number not nameable in fewer than twenty-two syllables" is actually named by this expression, which has twenty-one syllables. The Epimenides paradox is a traditional ancient Greek paradox of this kind: Epimenides (a Cretan) said, "Cretans are always liars...."

C. Paradoxes of the Continuum

The problem of the continuum is important in both mathematics and philosophy. There are several **paradoxes of Zeno** concerning the continuum, among which the following two are best known:

(1) Assume that Achilles and a tortoise start simultaneously from the points A and B, respectively, Achilles running after the tortoise. When Achilles reaches the point B, the tortoise advances to a point B_1. When Achilles reaches the point B_1, the tortoise advances further to a point B_2. Thus Achilles can never overtake the tortoise.

(2) A flying arrow occupies a certain point at each moment. In other words, at each moment the arrow stands still. Therefore the arrow can never move.

References

[1] A. N. Whitehead and B. Russell, Principia mathematica, second edition, Cambridge Univ. Press, 1925.
[2] S. C. Kleene, Introduction to metamathematics, Van Nostrand, 1952.
[3] E. Mendelson, Introduction to mathematical logic, Van Nostrand, 1966.

320 (XIII.19) Partial Differential Equations

A. General Remarks

A partial differential equation is a functional equation

$$F\left(x_1,\ldots,x_n,z,\frac{\partial z}{\partial x_1},\ldots,\frac{\partial^2 z}{\partial x_1^2},\frac{\partial^2 z}{\partial x_1\partial x_2},\ldots\right)=0$$

that involves a function z of independent variables $x_1, x_2, \ldots, x_n$, its †partial derivatives, and the independent variables $x_1, \ldots, x_n$. The definition of a system of partial differential equations is similar to that of a system of †ordinary differential equations. (The partial differential equation becomes an ordinary differential equation if the number of independent variables is one.)

The †order of the highest derivative appearing in a partial differential equation is called the **order** of the partial differential equation.

Usually we write p_i for $\partial z/\partial x_i$, x for x_1, and y for x_2 when $n=2$, and $p=\partial z/\partial x$, $q=\partial z/\partial y$, $r=\partial^2 z/\partial x^2$, $s=\partial^2 z/\partial x\partial y$, $t=\partial^2 z/\partial y^2$ when z is a function of x and y.

A partial differential equation is called **linear** if it is a linear relation with respect to z and its partial derivatives. For example, the equation

$$A(x,y)r+B(x,y)s+C(x,y)t+D(x,y)p$$
$$+E(x,y)q+F(x,y)z=G(x,y)$$

is a linear partial differential equation. A partial differential equation is **quasilinear** if it is a linear relation with respect to the highest-order partial derivatives. A partial differential equation is called **nonlinear** if it is not linear (→ 291 Nonlinear Problems). A function $z=\varphi(x_1,x_2,\ldots,x_n)$ that satisfies the given partial differential equation is called a **solution** of the partial differential equation. Obtaining such a solution for a given partial differential equation is called **solving** this equation, and by analogy to the case $n=2$, the **integral hypersurface** of the equation is

$$z-\varphi(x_1,x_2,\ldots,x_n)=0.$$

For a system of partial differential equations, we define solutions in the same manner.

Example 1. $X_1, X_2, \ldots, X_n$ are functions of n independent variables $x_1, x_2, \ldots, x_n$. Then solving the partial differential equation

$$X_1\frac{\partial z}{\partial x_1}+X_2\frac{\partial z}{\partial x_2}+\ldots+X_n\frac{\partial z}{\partial x_n}=0 \qquad (1)$$

is equivalent to solving the system of ordinary differential equations

$$\frac{dx_1}{X_1}=\frac{dx_2}{X_2}=\ldots=\frac{dx_n}{X_n}. \qquad (2)$$

In other words, if $f_1, f_2, \ldots, f_{n-1}$ are $n-1$ independent †integrals of (2), then for an arbitrary function Φ, $z=\Phi(f_1,\ldots,f_{n-1})$ is a general solution of (1) (→ Section C).

Example 2. If $P_1, P_2, \ldots, P_n, R$ are functions of independent variables $x_1, \ldots, x_n$ and the dependent variable z, and if the quasilinear partial differential equation **(Lagrange's differential equation)**

$$P_1\frac{\partial z}{\partial x_1}+P_2\frac{\partial z}{\partial x_2}+\ldots+P_n\frac{\partial z}{\partial x_n}=R \qquad (3)$$

has an integral hypersurface $V(z,x_1,\ldots,x_n)=0$, then we have

$$P_1\frac{\partial V}{\partial x_1}+P_2\frac{\partial V}{\partial x_2}+\ldots+P_n\frac{\partial V}{\partial x_n}+R\frac{\partial V}{\partial z}=0, \qquad (4)$$

which is an equation of type (1). From this we can obtain a general solution by the method of example 1. The same procedure is applicable to solving other systems of partial differential equations.

B. Characteristic Manifolds

We consider a partial differential equation of the nth order of two independent variables x, y:

$$F(x,y,z,p_{10},p_{01},\ldots,P_{n0},\ldots,p_{0n})=0,$$
$$\text{where } p_{jk}=\frac{\partial^{j+k}z}{\partial x^j\partial y^k}. \qquad (5)$$

With this equation, we associate a manifold defined by a real parameter λ,

$$x=x(\lambda),\quad y=y(\lambda),\quad p_{jk}=p_{jk}(\lambda),$$
$$j,k=0,1,\ldots,n-1;\quad j+k\leqslant n-1, \qquad (6)$$

and consider the following problem: Find a solution $\varphi(x,y)$ of (5) that satisfies

$$\frac{\partial^{j+k}\varphi(x,y)}{\partial x^j\partial y^k}=p_{jk}(\lambda),$$
$$j,k=0,1,2,\ldots,n-1;\quad j+k\leqslant n-1$$

on the curve $x=x(\lambda)$, $y=y(\lambda)$. We call this problem the **Cauchy problem** for equation (5).

If F vanishes for a system of values x^0, y^0, p^0_{jk} $(j,k=0,1,2,\ldots,n; p^0_{00}=z^0; j+k\leqslant n)$ and is †holomorphic in a neighborhood of this system of values, if x, y, p_{jk} are holomorphic functions of λ in a neighborhood of $\lambda=0$ and take the respective values x^0, y^0, p^0_{jk} at $\lambda=0$, and if

$$P_{n0}\,dy^n-P_{n-1,1}\,dx\,dy^{n-1}+\ldots$$
$$+(-1)^nP_{0n}\,dx^n\neq 0 \qquad (7)$$

for (x^0, y^0, p^0_{jk}), then we have a unique holomorphic solution of this Cauchy problem in a neighborhood of (x^0, y^0, p^0_{jk}). (Here we use the notation $P_{jk} = \partial F / \partial p_{jk}$.) This is **Cauchy's existence theorem**. We cannot apply this theorem when the left-hand side of (7) vanishes at (x^0, y^0, p^0_{jk}). At such a point, uniqueness of the solution fails, and there may be several solutions through the point (x^0, y^0, p^0_{jk}). There are n curves on which the left-hand side of (7) vanishes on the integral surface $z = \varphi(x, y)$. These curves are called **characteristic curves** of (5). We associate the values

$$p_{jk}(\lambda) = \left.\frac{\partial^{j+k}\varphi(x,y)}{\partial x^j \partial y^k}\right|_{x=x(\lambda),\, y=y(\lambda)},$$

$$j, k = 0, 1, \ldots, n; \quad j+k \leqslant n,$$

with each point (x, y) on these curves. The manifold $\{x(\lambda), y(\lambda), p(\lambda)\}$ $(p(\lambda) = \{p_{jk}(\lambda)\})$ of the parameter λ is called a **characteristic manifold** of equation (5). Cauchy's existence theorem cannot be applied on the characteristic manifold.

The foregoing considerations can be extended, to some extent, to the space of higher dimensions $\mathbf{R}^n$ or $\mathbf{C}^n$. Let P be a linear partial differential operator of order m:

$$P(x, D) = \sum_{|\alpha| \leqslant m} a_\alpha(x) D^\alpha,$$

$$D^\alpha = \left(\frac{\partial}{\partial x_i}\right)^{\alpha_1} \cdots \left(\frac{\partial}{\partial x_n}\right)^{\alpha_n}, \qquad |\alpha| = \alpha_1 + \ldots + \alpha_n.$$

The coefficients are assumed smooth and real in $\mathbf{R}^n$ or holomorphic in $\mathbf{C}^n$. Its homogeneous part of order m, denoted by $P_m(x, D)$, is called the **principal part** of P. Let S be a hypersurface defined by $\varphi(x) = 0$ with $\operatorname{grad}\varphi(x) \neq 0$. S is called a **characteristic surface** if $P_m(x, \operatorname{grad}\varphi(x)) = 0$ holds for x in a neighborhood of S or merely for $x \in S$, and $\varphi(x)$ is called a phase function. A real vector $\xi^0 (\neq 0)$, or a complex vector $\xi^0 \neq 0$ is called a characteristic direction at the point x_0 if $P_m(x_0, \xi^0) = 0$. The zeros (x, ξ) $(\xi \neq 0)$ of $P_m(x, \xi)$ are called the **characteristic set**. Furthermore (x_0, ξ^0) is called simple if $\operatorname{grad}_\xi P_m(x_0, \xi^0) \neq 0$. Suppose that (x_0, ξ^0) is simple. The integral curve $(x(t), \xi(t))$ that satisfies

$$\dot{x} = \operatorname{grad}_\xi P_m(x, \xi),$$

$$\dot{\xi} = -\operatorname{grad}_x P_m(x, \xi)$$

is called the **bicharacteristic strip** of P_m. Evidently $P_m(x, \xi)$ is constant along the bicharacteristic strip. In particular, if this constant is zero, it is said to be **null-bicharacteristic**. The phase function $\varphi(x)$ can be obtained at least locally by using the bicharacteristic strip.

C. Classification of Solutions

First, we consider a partial differential equation of the first order of two independent variables:

$$F(x, y, z, p, q) = 0. \tag{8}$$

A solution of (8) that contains two arbitrary constants is called a **complete solution**. If we get one complete solution of (8), then we can obtain all the solutions by differentiations and eliminations. Let (9) be a complete solution:

$$V(x, y, z, a, b) = 0. \tag{9}$$

Differentiating this, we get

$$\frac{\partial V}{\partial x} + p\frac{\partial V}{\partial z} = 0, \qquad \frac{\partial V}{\partial y} + q\frac{\partial V}{\partial z} = 0. \tag{10}$$

Eliminating a and b from (9) and (10), we get the original equation (8). Furthermore, solving equation (8) is equivalent to getting three functions a, b, z of x, y from (8), (9), and (10). If we regard a, b as functions of x, y, in (9), we get

$$\frac{\partial V}{\partial a}\frac{\partial a}{\partial x} + \frac{\partial V}{\partial b}\frac{\partial b}{\partial x} = 0, \qquad \frac{\partial V}{\partial a}\frac{\partial a}{\partial y} + \frac{\partial V}{\partial b}\frac{\partial b}{\partial y} = 0. \tag{11}$$

Therefore we can replace (9) and (10) by (9) and (11). Hence we have the following three cases: (i) When a, b are constants, we get a complete solution. (ii) When $V = 0$, $\partial V / \partial a = 0$, and $\partial V / \partial b = 0$, we get a solution that does not contain arbitrary constants, because z, a, b are all functions of x and y. We call this solution a **singular solution** of (8). (iii) When $\partial V / \partial a$, $\partial V / \partial b$ do not vanish simultaneously, the Jacobian $D(a, b)/D(x, y)$ vanishes because of (11). This means that there exists a functional relation between a and b. If there are two such relations, a and b are constants and the solution z becomes a complete solution. Therefore we assume that there is only one such relation between a and b, whose form is assumed to be $b = \varphi(a)$. Then we get

$$V(x, y, z, a, \varphi(a)) = 0, \qquad \frac{\partial V}{\partial a} + \frac{\partial V}{\partial b}\varphi'(a) = 0. \tag{12}$$

If we solve (12) for the unknowns a and z, we get a solution z of (8) that contains an arbitrary function φ instead of arbitrary constants. Such a solution is called a **general solution** of the partial differential equation (8) of the first order. By specializing this function φ, we obtain a **particular solution** of (8). Thus (i)–(iii) exhaust all the cases, and by obtaining a complete solution of (8) we can get all the solutions of (8). The number of complete solutions may be more than 1, or it may be infinite. These complete solutions can be transformed into each other by †contact transformations.

Moreover, they are contained in the general solutions.

Now we consider the case where the number of independent variables is n. Take an equation that contains $n-r+1$ arbitrary constants $a_1, a_2, \ldots, a_{n-r+1}$:

$$V(x_1, x_2, \ldots, x_n : a_1, a_2, \ldots, a_{n-r+1}) = 0. \qquad (13)$$

Differentiate this equation assuming that $a_1, \ldots, a_{n-r+1}$ take fixed values. Then we get

$$\frac{\partial V}{\partial x_i} + p_i \frac{\partial V}{\partial z} = 0, \quad p_i = \frac{\partial z}{\partial x_i}, \quad i = 1, 2, \ldots, n. \qquad (14)$$

If we eliminate $a_1, a_2, \ldots, a_{n-r+1}$ from (13) and (14), we obtain r partial differential equations of the first order:

$$F_j(x_1, x_2, \ldots, x_n, z, p_1, p_2, \ldots, p_n) = 0, \quad j = 1, 2, \ldots, r. \qquad (15)$$

We call (13) a **complete solution** of (15). In this case, as in the case when $n=2$, we get all the solutions of (15) from a complete solution (13) of (15). We have the same classification as in the case $n=2$: (i) When we regard $a_1, \ldots, a_{n-r+1}$ as constants, then we have a complete solution of (15). (ii) When we can eliminate the constants $a_1, \ldots, a_{n-r+1}$ from equations

$$V = 0, \quad \frac{\partial V}{\partial a_1} = 0, \ldots, \frac{\partial V}{\partial a_{n-r+1}} = 0,$$

we get a solution that does not contain an arbitrary constant. Such a solution is called a **singular solution** of (15). (iii) If not all of the $\partial V / \partial a_i$ vanish, there exists at least one functional relation among $a_1, a_2, \ldots, a_{n-r+1}$. We assume that there exist exactly k ($\leqslant n-r$) relations among $a_1, a_2, \ldots, a_{n-r+1}$:

$$f_j(a_1, a_2, \ldots, a_{n-r+1}) = 0, \quad j = 1, \ldots, k. \qquad (16)$$

Then there exist numbers $\lambda_1, \lambda_2, \ldots, \lambda_k$ such that

$$\frac{\partial V}{\partial a_l} = \lambda_1 \frac{\partial f_1}{\partial a_l} + \ldots + \lambda_k \frac{\partial f_k}{\partial a_l}, \quad l = 1, \ldots, n-r+1. \qquad (17)$$

Hence, by eliminating $a_1, a_2, \ldots, a_{n-r+1}, \lambda_1, \lambda_2, \ldots, \lambda_k$ from (13), (16), and (17), we generally obtain exactly one relation between $x_1, \ldots, x_n$ and z. This is a solution of (15) that contains exactly k arbitrary functions $f_1, f_2, \ldots, f_k$. Such a solution is called a **general solution** of (15). In particular, if $k = n-r+1$, then it is a complete solution. We might think that there are $n-r$ general solutions corresponding to $k = 1, \ldots, n-r$. But these general solutions are not essentially different. For the partial equation of the second order $F(x, y, z, p, q, r, s, t) = 0$, this definition of a general solution is not applicable since we cannot successfully define a general solution by using the number of arbitrary functions contained in a solution.

Instead, we now use the following definition, due to J. G. Darboux: A solution of a general partial differential equation is called a **general solution** if by specializing its arbitrary functions and constants appropriately we obtain a solution whose existence is established by Cauchy's existence theorem. A solution $z = \varphi(x, y)$ of a general partial differential equation is called a **singular solution** if Cauchy's existence theorem cannot be applied on any curves on the manifold formed by $z = \varphi(x, y)$, $p = \partial\varphi/\partial x$, $q = \partial\varphi/\partial y$.

D. Cauchy's Method

We can regard equation (8) as a relation between the point (x, y, z) on the integral surface S and the direction cosines of a tangent plane at that point. Therefore the tangent planes at all points of the surface form a one-parameter family. They envelop a cone (T) whose vertex is (x, y, z) on S. The tangent plane at a point M on the integral surface S is tangent to this cone (T) along one generating line G of (T).

A curve on S whose tangents are all generating lines of (T) is a characteristic curve. If we write

$$X = \frac{\partial F}{\partial x}, \quad Y = \frac{\partial F}{\partial y}, \quad Z = \frac{\partial F}{\partial z}, \quad P = \frac{\partial F}{\partial p}, \quad Q = \frac{\partial F}{\partial q},$$

then the characteristic curve is given by the system of ordinary differential equations:

$$\frac{dx}{P} = \frac{dy}{Q} = \frac{dz}{Pp + Qq} = \frac{-dp}{X + pZ} = \frac{-dp}{Y + qZ}. \qquad (18)$$

We call this system the **characteristic differential equation** or **Charpit subsidiary (auxiliary) equation** of the partial differential equation (8) of the first order. System (18) determines not only x, y, z but also p and q. The set of these †surface elements (x, y, z, p, q) is the characteristic manifold. This characteristic manifold is considered as a part of the integral surface with infinitesimal width, and in this case we call it the **characteristic strip**. The characteristic strip is represented by the equations $x = x(\lambda)$, $y = y(\lambda)$, $z = z(\lambda)$, $p = p(\lambda)$, $q = q(\lambda)$ containing a parameter. On the integral surface $z = z(x, y)$, we have

$$\frac{dz}{d\lambda} = \frac{\partial z}{\partial x}\frac{dx}{d\lambda} + \frac{\partial z}{\partial y}\frac{dy}{d\lambda}$$

and

$$dz = p\,dx + q\,dy. \qquad (19)$$

Equation (19) is called the **strip condition**. The equations (18) evidently satisfy this condition.

The solution $u(t,x)=u(t,x_1,\dots,x_n)$ of

$$u_t+f(t,x_1,\dots,x_n,u_{x_1},\dots,u_{x_n})=0,$$

$$u(0,x)=\varphi(x)$$

is obtained (at least locally) as follows. Let the solution of the differential equations

$$\frac{dx_i}{dt}=f_{\xi_i}(t,x_1,\dots,x_n,\xi_1,\dots,\xi_n),$$

$$\frac{d\xi_i}{dt}=-f_{x_i}(t,x_1,\dots,x_n,\xi_1,\dots,\xi_n),\quad 1\leqslant i\leqslant n,$$

issuing from (x_0,ξ^0) at $t=0$ be $(x(t;x_0,\xi^0)$, $\xi(t;x_0,\xi^0))$. Then specializing $\xi_j^0=\varphi_{x_j}(x_0)$ $(1\leqslant j\leqslant n)$, the solution $u(t,x)$ is obtained by quadrature along these curves (characteristic strips) from the relation

$$\frac{du}{dt}=-f(t,x,\xi)+\sum_{j=1}^{n}\xi_j f_{\xi_j}(t,x,\xi).$$

In particular, when $f(t,x,\xi)$ is homogeneous of degree 1 in ξ, by Euler's identity the right-hand side is identically zero. This means that u is constant along the characteristic strip.

E. Homogeneous Partial Differential Equations

Assume that $f(\xi_1,\xi_2,\dots,\xi_m)$ is a †homogeneous polynomial of m independent variables $\xi_1,\xi_2,\dots,\xi_m$. We denote the differential operator $\partial/\partial x_s$ by D_s. Then consider a homogeneous partial differential equation

$$f(D_1,D_2,\dots,D_m)u=0. \tag{20}$$

We can obtain a homogeneous equation from an inhomogeneous partial differential equation by transformation of the dependent variable. For example, $D_1^2w=D_2w$ becomes the homogeneous partial differential equation $(D_1^2-D_2D_3)u=0$ by the transformation of the dependent variable $u=e^{x_3}w$. The equation $(D_1^2-D_2D_3)u=0$ corresponds to the homogeneous polynomial $f(\xi_1,\xi_2,\xi_3)=\xi_1^2-\xi_2\xi_3$. Generally, for equation (20), we consider the solution

$$u=F(\theta_1,\theta_2,\dots,\theta_s), \tag{21}$$

where $\theta_1,\theta_2,\dots,\theta_s$ are functions of $x_1,\dots,x_m$ and F is an arbitrary function of θ_i. Such a solution is called a **primary solution** of (20). For example, for the equation $(D_1^2-D_2^2)u=0$, there are two primary solutions, $u=F(x_1+x_2)$ and $u=F(x_1-x_2)$. For †Laplace's equation

$$\Delta u=\frac{\partial^2u}{\partial x^2}+\frac{\partial^2u}{\partial y^2}+\frac{\partial^2u}{\partial z^2}=0,$$

we have a primary solution $u=F(z+ix\cos\alpha+iy\sin\alpha)$, where α is a parameter.

Second, for the †wave equation

$$\frac{\partial^2u}{\partial x^2}+\frac{\partial^2u}{\partial y^2}+\frac{\partial^2u}{\partial z^2}-\frac{1}{c^2}\frac{\partial^2u}{\partial t^2}=0, \tag{22}$$

we have a solution

$$u=\gamma f(\alpha,\beta),\quad \gamma=\frac{1}{r},\quad \alpha=t-\frac{r}{c},\quad \beta=\frac{z-r}{x+iy}, \tag{23}$$

where $r^2=x^2+y^2+z^2$, f is an arbitrary function, and γ is a particular solution of (22). Furthermore, $v=f(\alpha,\beta)$ satisfies the equation of a characteristic curve of (22):

$$\left(\frac{\partial v}{\partial x}\right)^2+\left(\frac{\partial v}{\partial y}\right)^2+\left(\frac{\partial v}{\partial z}\right)^2=\frac{1}{c^2}\left(\frac{\partial v}{\partial t}\right)^2.$$

Such a solution, which is the product of particular solutions and some function that contains an arbitrary function, is called a **primitive solution** of the original equation. Equation (22) has another primitive solution of the type

$$u=\frac{1}{r}g\left(t+\frac{r}{c},\ \frac{z-r}{x+iy}\right),$$

where g is an arbitrary function.

Laplace's equation has a primitive solution

$$u=\frac{1}{r}f\left(\frac{z-r}{x+iy}\right).$$

A **basic equation** is an equation, such as Laplace's equation, that has a primary solution and a primitive solution. A solution of an equation that has the same characteristic curves as a basic equation can be obtained from a particular solution of the basic equation by integrations and additions. For example, if we choose a particular solution $u=r^{-1}$ of Laplace's equation, which is a specialization of the primitive solution $u=r^{-1}f((z-r)/(x+iy))$, then

$$u=\iiint[(x-\xi)^2+(y-\eta)^2+(z-\zeta)^2]^{-1/2}\times F(\xi,\eta,\zeta)\,d\xi\,d\eta\,d\zeta$$

is a solution of

$$\Delta u+4\pi F(x,y,z)=0$$

in the interior of the domain of integration.

F. Determined Systems

The general form of a system of partial differential equations in two independent variables is

$$F_i(x,y,u^{(1)},u^{(2)},\dots,u^{(m)},u_x^{(1)},u_y^{(1)},\dots,u_x^{(m)},u_y^{(m)},u_{xx}^{(1)},\dots)=0,\qquad i=1,2,\dots,h, \tag{24}$$

i.e., a system of h equations for m functions

$u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ of the independent variables x and y. The system is called a **determined system** if $h=m$, an **overdetermined system** if $h>m$, and an **underdetermined system** if $h<m$.

An example of a determined system is the †Cauchy-Riemann equation

$$u_x - v_y = 0, \quad u_y + v_x = 0,$$

for $u(x,y)$, $v(x,y)$, which can be further reduced to two determined equations $\Delta u=0$ and $\Delta v=0$.

A simple example of an overdetermined system is

$$u_x = f(x,y), \quad u_y = f(x,y).$$

A necessary and sufficient condition for the existence of a solution of this system is $f_y = f_x$. The **Cauchy-Riemann differential equations** for a holomorphic function $f(z_1, z_2) = u + iv$ of two complex variables $z_1 = x_1 + iy_1$, $z_2 = x_2 + iy_2$ are

$$u_{x_1} = v_{y_1}, \quad u_{x_2} = v_{y_2}, \quad u_{y_1} = -v_{x_1},$$

$$u_{y_2} = -v_{x_2},$$

which can be reduced to

$$u_{x_1x_1} + u_{y_1y_1} = 0, \quad u_{x_1x_2} + u_{y_1y_2} = 0,$$

$$u_{x_2x_2} + u_{y_2y_2} = 0, \quad u_{x_1y_2} - u_{x_2y_1} = 0,$$

which is also an overdetermined system.

An example of an underdetermined system is

$$\frac{\partial(u,v)}{\partial(x,y)} = u_x v_y - u_y v_x = 0.$$

If this equation holds, there exists a functional relation $w(u,v)=0$ that can be regarded as a solution of this underdetermined system.

G. General Theory of Differential Operators

In recent developments of the theory of partial differential equations, there is a trend to construct a general theory for †differential operators regardless of the classical types of differential equations (→ 112 Differential Operators). For example, we take a property that is satisfied by some equation of classical type (e.g., an †elliptic differential equation) and proceed to characterize all equations that have this property. (For example, †hypoellipticity is a property of classical †parabolic and elliptic equations.) There are several basic problems in this general theory: the existence of a fundamental solution, the existence of a local solution, unique continuation of solutions, the differentiability and analyticity of solutions, and the propagation of smoothness. We explain here only two of them: the fundamental solution and the local existence of solutions.

H. Fundamental Solutions

L is assumed to be a linear partial differential operator with constant coefficients. If a †distribution E satisfies the equation

$$LE(x) = \delta(x),$$

where $\delta(x)$ is the †Dirac δ-function, then we call $E(x)$ a **fundamental solution** (or **elementary solution**) of L. Also, if L is a linear differential operator and E satisfies the equation

$$L_x E(x,y) = \delta(x-y),$$

then we call this distribution $E(x,y)$ a **fundamental kernel** (or **elementary kernel**) of L.

Let L be a differential operator with constant coefficients and $E(x)$ be a fundamental solution of L. Then $E(x,y)$ is a fundamental kernel of L. Sometimes $E(x,y)$ itself is called a fundamental solution. L. Ehrenpreis and B. Malgrange proved that any linear differential operator with constant coefficients has a fundamental solution [4].

If we take a fundamental solution (fundamental kernel) E and add to it an arbitrary solution of the equation $Lu=0$, then we get another fundamental solution (fundamental kernel) of L. This freedom of the fundamental solution (fundamental kernel) can be used to construct †Green's functions of the boundary value problem of elliptic equations and of the mixed initial-boundary value problem for parabolic equations. A Green's function is a fundamental solution (fundamental kernel) that satisfies given boundary conditions (→ 188 Green's Functions; 189 Green's Operator). The fundamental solutions (fundamental kernels) relative to the Cauchy problem are also defined as in this section. For example, consider a fundamental solution of the Cauchy problem concerning the future behavior of a differential operator $L = \partial/\partial t - P(\partial/\partial x)$, namely, a distribution $E(t,x)$ that satisfies $LE = 0$ $(t>0)$ and $E(t,x)|_{t=0} = \delta(x)$. If we put $\tilde{E}(t,x) = E(t,x)$ $(t \geqslant 0)$ and $\tilde{E}(t,x) = 0$ $(t<0)$, then $\tilde{E}(t,x)$ is a fundamental solution (or kernel) of the differential operator L, that is, $L\tilde{E} = \delta(t,x)$. Sometimes a fundamental solution of the Cauchy problem for a parabolic equation is called a Green's function. On the other hand, a fundamental solution (or kernel) of the Cauchy problem for a hyperbolic equation is called a †Riemann function. A Riemann function actually is not always a function; in general it is a distribution.

Example 1. A fundamental solution of the 3-dimensional Laplacian

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2}$$

is $E(x) = -1/4\pi r$, where $r = \sqrt{x_1^2 + x_2^2 + x_3^2}$.

Example 2. A fundamental solution of the Cauchy problem for the future of the 3-dimensional wave operator

$$L=\frac{\partial^2}{\partial t^2}-\sum_{i=1}^{3}\frac{\partial^2}{\partial x_i^2},$$

i.e., a distribution $E(t,x)$ $(t\geqslant 0)$ that satisfies $LE=0$ $(t>0)$, $E(0,x)=0$, and $(\partial/\partial t)E(0,x)=\delta(x)$, is given by $E(t,x)=(1/4\pi t)\delta(r-t)$ $(t>0)$, $r=\sqrt{x_1^2+x_2^2+x_3^2}$. A fundamental solution for L is given by $\tilde{E}(t,x)=E(t,x)$ $(t\geqslant 0)$; $=0$ $(t<0)$ (→ Appendix A, Table 15.V).

I. Existence of Local Solutions

Given a linear differential operator L and the equation $Lu=f$, we have the problem of determining whether this equation always has a solution in some neighborhood of a given point. If the coefficients of L and f are holomorphic in a neighborhood of this point and if the homogeneous part of highest order does not vanish, then there exists a solution that is holomorphic in a neighborhood of the given point (†Cauchy-Kovalevskaya existence theorem).

If L is a linear differential operator with constant coefficients, E is a fundamental solution of L, and f is a function (or distribution) that is zero outside of a compact set, then we have a solution u that is the †convolution of E and f: $u=E*f$. On the other hand, H. Lewy proposed the following example [3]:

$$Lu\equiv\frac{\partial u}{\partial x_1}+i\frac{\partial u}{\partial x_2}+2i(x_1+ix_2)\frac{\partial u}{\partial x_3}=f(x_3),$$

where f is a real function of x_3. He showed that if this equation has a solution that is of class C^1, then f must be real analytic. Therefore, if f is of class C^∞ but not real analytic, then this equation has no C^1-solution. Actually, no solution exists even in the distribution sense. (Note that, since the coefficients of L are now complex-valued, the results mentioned at the beginning of this section are no longer applicable.)

For linear differential operators L, a study by L. Hörmander gives some necessary conditions and also some sufficient conditions for the local existence of a solution of the equation $Lu=f$ for sufficiently general f [4]. This result has been developed and completed by L. Nirenberg and F. Treves [18] and by R. Beals and C. Fefferman [19]. The operator considered by S. Mizohata (*J. Math. Kyoto Univ.*, 1 (1962)),

$$L\equiv\frac{\partial}{\partial x_i}+ix_1^k\frac{\partial}{\partial x_2},$$

serves as a standard model in this problem. In the neighborhood of the origin, if k is even, L is locally solvable, and if k is odd, it is not locally solvable. However, for linear partial differential operators with multiple characteristics, the problem of local solvability becomes extremely difficult.

References

[1] E. Goursat, Cours d'analyse mathématique, Gauthier-Villars, II, second edition, 1911; III, fourth edition, 1927.
[2] R. Courant and D. Hilbert, Methods of mathematical physics II, Interscience, 1962.
[3] H. Lewy, An example of a smooth linear partial differential equation without solution, Ann. Math., (2) 66 (1957), 155–158.
[4] L. Hörmander, Linear partial differential operators, Springer, 1963.
[5] S. Mizohata, Theory of partial differential equations, Cambridge Univ. Press, 1973. (Original in Japanese, 1965.)
[6] C. Carathéodory, Calculus of variations and partial differential equations of the first order, Holden-Day, 1967.
[7] I. G. Petrovskiĭ, Über das Cauchysche Problem für ein System linearer partieller Differentialgleichungen im Gebiete der nichtanalytischen Funktionen, Bull. Univ. Moscou, ser. internat., sec. A, vol. 1, fasc. 7 (1938), 1–74.
[8] J. Hadamard, Le problème de Cauchy et les équations aux dérivées partielles linéaires hyperboliques, Hermann, 1932.
[9] I. G. Petrovskiĭ, Lectures on partial differential equations, Interscience, 1954. (Original in Russian, 1953.)
[10] S. L. Sobolev, Partial differential equations of mathematical physics, Pergamon, 1964. (Original in Russian, 1954.)
[11] J. L. Lions, Equations différentielles opérationelles et problèmes aux limites, Springer, 1961.
[12] J. L. Lions and E. Magenes, Problèmes aux limites non homogènes et applications, Dunod, I, 1968; II, 1968; III, 1969.
[13] A. Friedman, Partial differential equations, Holt, Rinehart and Winston, 1969.
[14] F. Treves, Locally convex spaces and linear partial differential equations, Springer, 1967.
[15] F. Treves, Linear partial differential equations with constant coefficients, Gordon & Breach, 1966.
[16] F. John, Partial differential equations, Springer, 1971.
[17] L. Bers, F. John, and M. Schechter, Partial differential equations, Interscience, 1964.
[18] L. Nirenberg and F. Treves, On local solvability of linear partial differential equations I, II, Comm. Pure Appl. Math., 23 (1970), 1–38, 459–510.

[19] R. Beals and C. Fefferman, On local solvability of linear partial differential equations, Ann. Math., 97 (1973), 482–498.
[20] F. Cardoso and F. Treves, A necessary condition of local solvability for pseudodifferential equations with double characteristics, Ann. Inst. Fourier, 24 (1974), 225–292.

321 (XIII.21) Partial Differential Equations (Initial Value Problems)

A. General Remarks

First, we give two examples of initial value problems for †partial differential equations.

(I) Consider the partial differential equation $u_x - u_y = 0$ of two independent variables x and y. If the function $\varphi(y)$ is of class C^1, then $u = \varphi(x+y)$ is a solution of this equation that satisfies $u(0,y)=\varphi(y)$.

(II) We denote a point of $\mathbf{R}^{n+1}$ (or $\mathbf{C}^{n+1}$) by (t,x), $x=(x_1,\ldots,x_n)$. Let L be a linear partial differential operator of order m:

$$L=\frac{\partial^m}{\partial t^m}+\sum_{|\nu|\leqslant m} a_{\nu_0\nu_1\ldots\nu_n}(t,x)\frac{\partial^{|\nu|}}{\partial t^{\nu_0}\partial x_1^{\nu_1}\ldots\partial x_n^{\nu_n}},$$

$$|\nu|=\nu_0+\nu_1+\ldots+\nu_n,\quad \nu_0<m,$$

where the coefficients $a_{\nu_0\nu_1\ldots\nu_n}(t,x)$ are functions of class C^ω (i.e., †real analytic functions or holomorphic functions) in a neighborhood of $(t,x)=(0,0)$. If the functions $v(t,x)$ and $w_k(x)$ $(0\leqslant k\leqslant m-1)$ are of class C^ω in a neighborhood of $(t,x)=(0,0)$, then there exists a unique solution $u(t,x)$ of class C^ω in a neighborhood of $(t,x)=(0,0)$ that satisfies

$$L[u]=v(t,x),$$

$$\frac{\partial^k u}{\partial t^k}(0,x)=w_k(x),\qquad 0\leqslant k\leqslant m-1. \qquad (1)$$

This is called the **Cauchy-Kovalevskaya existence theorem** (for linear partial differential equations).

As in (II) we choose one of the independent variables as the principal variable and regard the others as parameters. When we assign a value a to the principal variable t, the values of the dependent variables (unknown functions) and their derivatives are called **initial values, initial data**, or **Cauchy data**. Conditions to determine initial values are called **initial conditions**. The problem of finding a solution of (1) under given initial conditions is called an **initial value problem** or **Cauchy problem**. We may consider initial value problems not only for initial conditions on a hyperplane $t=a$, but also for initial conditions on a hypersurface, called an **initial surface** (→ Section C).

Let $a(x,D)$ be a linear partial differential operator of order m:

$$a(x,D)=\sum_{|\alpha|\leqslant m} a_\alpha(x)D^\alpha,\qquad D^\alpha=\frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1}\ldots\partial x_n^{\alpha_n}},$$

$$|\alpha|=\alpha_1+\ldots+\alpha_n,$$

where the coefficients $a_\alpha(x)$ are of class C^ω in a neighborhood of $x=0$. Its **characteristic polynomial** is

$$h(x,\xi)=\sum_{|\alpha|=m} a_\alpha(x)\xi^\alpha,\qquad \xi^\alpha=\xi_1^{\alpha_1}\ldots\xi_n^{\alpha_n}.$$

Let $S: s(x)=0$ be a regular surface (i.e., $s_x=(\partial s/\partial x_1,\ldots,\partial s/\partial x_n)\neq(0)$) of codimension 1. We suppose that S is a †noncharacteristic surface, that is, $h(x,s_x)\neq 0$ on S. Let $v(x)$ and $w_k(x)$ $(0\leqslant k\leqslant m-1)$ be the functions of class C^ω in a neighborhood of $x=0$ and on S, respectively. We consider the Cauchy problem

$$a(x,D)u(x)=v(x),\qquad \frac{\partial^k u}{\partial n^k(x)}=w_k(x)\quad \text{on } S,$$

$$0\leqslant k\leqslant m-1,\qquad (1)$$

where n is the outward normal direction of S. S is thus the initial surface. Then there exists a unique solution $u(x)$ of class C^ω in a neighborhood of $x=0$. In fact, by the change of variables $X_1=s(x)$, $X_i=x_i$ $(2\leqslant i\leqslant n)$ if $\partial s/\partial x_1\neq 0$ on S, this problem can be reduced to the Cauchy problem (1) by taking account of the fact that $h(x,s_x)\neq 0$ on S.

The Cauchy-Kovalevskaya theorem asserts the local existence of solution when the initial values are of class C^ω. Indeed, J. Hadamard noted that if the initial values are not of class C^ω, the initial value problem does not always have a solution. For example, consider the initial value problem

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}=0$$

with the initial values

$$u(0,y,z)=w(y,z),\qquad \frac{\partial u}{\partial x(0,y,z)}=0.$$

If the function $w(y,z)$ is not of class C^ω in any neighborhood of $y=z=0$, the solution of this problem can never exist in (or even on one side $x\geqslant 0$ of) any neighborhood of $x=y=z=0$.

B. The Cauchy-Kovalevskaya Existence Theorem for a System of Partial Differential Equations in the Normal Form

The Cauchy-Kovalevskaya existence theorem (1) is extended for more general systems of

partial differential equations in the normal form studied by Kovalevskaya. Consider

$$\frac{\partial^{p_i} u_i}{\partial t^{p_i}} = F_i\left(t, x, u_1, \ldots, u_m, \ldots, \frac{\partial^{|\nu|} u_j}{\partial t^{\nu_0} \partial x_1^{\nu_1} \ldots \partial x_n^{\nu_n}}, \ldots\right),$$

where $1 \leqslant i, j \leqslant m$, $|\nu| = \nu_0 + \nu_1 + \ldots + \nu_n \leqslant p_j$, $\nu_0 < p_j$, and $x = (x_1, \ldots, x_n)$. We assume that F_i $(1 \leqslant i \leqslant m)$ are functions of class C^ω with respect to arguments t, x, $u_1, \ldots, u_m, \ldots, \partial^{|\nu|} u_j / \partial t^{\nu_0} \partial x_1^{\nu_1} \ldots \partial x_n^{\nu_n}, \ldots$, in a neighborhood of $(0, 0, \ldots, 0)$. If the functions $w_{ik}(x)$ $(1 \leqslant i \leqslant m, 0 \leqslant k \leqslant p_i - 1)$ are of class C^ω in a neighborhood of $x = 0$, then there exists a unique solution $u_1(t, x), \ldots, u_m(t, x)$ of class C^ω in a neighborhood of $(t, x) = (0, 0)$ that satisfies the equations and the initial values

$$\frac{\partial^k u_i}{\partial t^k}(0, x) = w_{ik}(x), \quad 1 \leqslant i \leqslant m, \quad 0 \leqslant k \leqslant p_i - 1$$

[2, 4].

C. Single Equations of the First Order

For a single partial differential equation given in the normal form

$$\frac{\partial u}{\partial x} = F\left(x, y_1, \ldots, y_k, u, \frac{\partial u}{\partial y_1}, \ldots, \frac{\partial u}{\partial y_k}\right), \tag{2}$$

we assume that $F(x, y_1, \ldots, y_k, u, q_1, \ldots, q_k)$ is a real-valued function of class C^2 in a neighborhood of $x = a$, $y_i = b_i$, $u = c$, $q_i = d_i$, and that $\varphi(y_1, \ldots, y_k)$ is also a function of class C^2 such that $\varphi = c$, $\partial\varphi/\partial y_i = d_i$ at $y_i = b_i$. Then there is a solution u of (2) in a neighborhood of $x = a$ and $y_i = b_i$ that satisfies $u(a, y_1, \ldots, y_k) = \varphi(y_1, \ldots, y_k)$ and is of class C^2.

For the uniqueness of solutions of this Cauchy problem, A. Haar (*Atti del Congresso Internazionale dei Matematici*, 1928, Bologna, vol. 3) showed that if F satisfies the †Lipschitz condition

$$|F(x, y, u', q') - F(x, y, u, q)| \leqslant A \sum_{i=1}^{k} |q_i' - q_i| + B|u' - u|,$$

then the solution of the initial value problem for (2) is unique. To obtain this result he studied the partial differential inequality

$$\left|\frac{\partial u}{\partial x}\right| \leqslant A \sum_{i=1}^{k} \left|\frac{\partial u}{\partial y_i}\right| + B|u| + C. \tag{3}$$

Next, we consider more general equations of the first order

$$F\left(x_1, \ldots, x_k, u, \frac{\partial u}{\partial x_1}, \ldots, \frac{\partial u}{\partial x_k}\right) = 0. \tag{4}$$

Suppose that $F(x_1, \ldots, x_k, u, p_1, \ldots, p_k)$ is a real-valued function of class C^2 in a neighborhood of $x_i = a_i$, $u = b$, $p_i = c_i$; $\varphi(x_1, \ldots, x_k)$, $S(x_1, \ldots, x_k)$ are functions of class C^2 in a neighborhood of $x_i = a_i$ that satisfy $b = \varphi(a_1, \ldots, a_k)$, $c_i = (\partial\varphi/\partial x_i)_{x=a}$, $S(a_1, \ldots, a_k) = 0$; and

$$\left[\sum_{\nu=1}^{k} \frac{\partial F}{\partial p_\nu} \frac{\partial S}{\partial x_\nu}\right]_{x=a, u=b, p=c} \neq 0. \tag{5}$$

Then there exists a solution u of (4) of class C^2 in a neighborhood of $x = a$ that satisfies $u = \varphi(x)$ on the hypersurface $S(x) = 0$.

Furthermore, if F, φ, S are of class C^1 and satisfy (5), then there is at most one solution u of (4) of class C^1 in a neighborhood of a that satisfies $u = \varphi(x)$ on the hypersurface $S(x) = 0$. These facts can be proved in the following way. By choosing $S_1(x), \ldots, S_{k-1}(x)$ and then $S(x)$ so that the †Jacobian $\partial(S, S_1, \ldots, S_{k-1})/\partial(x_1, \ldots, x_k)$ does not vanish and by changing variables from x to $S, S_1, \ldots, S_{k-1}$, we obtain a normal form solved for $\partial u/\partial S$ by condition (5) (this condition means that the hypersurface $S(x) = 0$ is not tangent to the †characteristic curves).

D. Quasilinear Equations of the Second Order

Consider the equation

$$\sum_{i,j=1}^{k} a_{ij}(x, u, p) \frac{\partial^2 u}{\partial x_i \partial x_j} + b(x, u, p) = 0,$$

where $x = (x_1, \ldots, x_k)$, $p = (p_1, \ldots, p_k)$, and $p_i = \partial u/\partial x_i$. We assume that the initial conditions are $u = \varphi(x)$ on $S(x) = 0$ and

$$\sum_{i=1}^{k} \frac{\partial S}{\partial x_i} \frac{\partial u}{\partial x_i} = \psi(x)$$

on the same hypersurface. Taking the other functions $S_1(x), \ldots, S_{k-1}(x)$ and then $S(x)$ so that the Jacobian $\partial(S, S_1, \ldots, S_{k-1})/\partial(x_1, \ldots, x_k)$ does not vanish, we change the variables x to $s_0, s_1, \ldots, s_{k-1}$ $(s_0 = S, s_i = S_i)$. Then we get

$$\sum_{\mu,\nu=0}^{k-1} Q(S_\mu, S_\nu) \frac{\partial^2 u}{\partial s_\mu \partial s_\nu} + b^* = 0, \tag{6}$$

where

$$Q(S_\mu, S_\nu) = \sum_{i,j=1}^{k} a_{ij} \frac{\partial S_\mu}{\partial x_i} \frac{\partial S_\nu}{\partial x_j},$$

$$p_i = \sum_{\mu=0}^{k-1} \frac{\partial S_\mu}{\partial x_i} \frac{\partial u}{\partial s_\mu},$$

$$b^* = b + \sum_{\lambda=0}^{k-1} \sum_{i,j=1}^{k} a_{ij} \frac{\partial^2 S_\lambda}{\partial x_i \partial x_j} \frac{\partial u}{\partial s_\lambda}.$$

When $s_0 = 0$, the initial conditions are trans-

formed to

$$u=\varphi^*(s_1,\dots,s_{k-1})=\varphi(x(0,s_1,\dots,s_{k-1})),$$

$$\frac{\partial u}{\partial s_0}=\psi^*(s_1,\dots,s_{k-1})$$

$$=\left[\sum_{i=1}^{k-1}\left(\frac{\partial S}{\partial x_i}\right)^2\right]^{-1}\left[\psi(x(0,s_1,\dots,s_{k-1}))\right.$$

$$\left.-\sum_{\lambda=1}^{k-1}\frac{\partial\psi^*}{\partial s_\lambda}\sum_{i=1}^{k}\frac{\partial S_\lambda}{\partial x_i}\frac{\partial S}{\partial x_i}\right].$$

Suppose that $Q(S,S)\neq 0$, and set $\partial u/\partial s_i=q_i$. Then equation (6) added to the initial conditions $u=\varphi^*(s_1,\dots,s_{k-1})$, $q_0=\psi^*(s_1,\dots,s_{k-1})$ at $s_0=0$ is equivalent to the system in the normal form of partial differential equations of the first order

$$\frac{\partial u}{\partial s_0}=q_0,\quad \frac{\partial q_\mu}{\partial s_0}=\frac{\partial q_0}{\partial s_\mu},\quad \mu=1,\dots,k-1,$$

$$\frac{\partial q_0}{\partial s_0}=-\frac{1}{Q(S,S)}\left[\sum_{\mu=1}^{k-1}\left(\sum_{\nu=1}^{k-1}Q(S_\mu,S_\nu)\frac{\partial q_\mu}{\partial s_\nu}\right.\right.$$

$$\left.\left.-2Q(S_\mu,S)\frac{\partial q_0}{\partial s_\mu}\right)-b^*\right]$$

with initial conditions

$$u=\varphi^*(s_1,\dots,s_{k-1}),\qquad q_0=\psi^*(s_1,\dots,s_{k-1}),$$

$$q_u=\partial\varphi^*/\partial s_\mu$$

when $s_0=0$. Thus, if the coefficients are of class C^ω, the preceding theory applies to this equation.

E. Continuous Dependence of Solutions on the Initial Values

First, we consider the following simple example of a linear equation. The wave equation

$$\frac{\partial^2 v}{\partial t^2}=c^2\frac{\partial^2 v}{\partial x^2}$$

for $v(x,t)$ is the simplest †hyperbolic equation. Its solution satisfying the initial conditions $v(x,0)=f(x)$, $(\partial v/\partial t)(x,0)=g(x)$ is given by

$$v(x,t)=\frac{1}{2}(f(x+ct)+f(x-ct))+\frac{1}{2}\int_{x-ct}^{x+ct}g(\lambda)\,d\lambda.$$

It is obvious from this expression that if we regard $f(x)$ and $g(x)$ as elements of the †function space $C(\mathbf{R})$ of continuous functions of $x\in\mathbf{R}$ with the topology of †uniform convergence on compact sets, then $v(x,t)$ is determined as the value of a linear operator from $C(\mathbf{R}^2)$ to $C(\mathbf{R}^2)$ on xt-space. If such continuous dependence on the initial values is established, or more precisely, if there is a unique solution for sufficiently smooth initial values that depends on the initial values continuously in a suitable sense, we say that the initial value problem is **well posed (properly posed** or **correctly posed)**.

Systematic research on the well-posedness of Cauchy problems was initiated by I. G. Petrovskiĭ, who considered the following system of partial differential equations, which is more general than differential equations of the normal form. (The coefficients are all assumed to be functions of t only.)

$$\frac{\partial^{n_j}u_j}{\partial t^{n_j}}=\sum_{k=1}^{N}\sum_{|\nu|\leqslant m}a_{jk\nu_0\nu_1\dots\nu_n}(t)\frac{\partial^{\nu_0+\nu_1+\dots+\nu_n}u_k(x,t)}{\partial t^{\nu_0}\partial x_1^{\nu_1}\dots\partial x_n^{\nu_n}}$$

$$+B_j(x_1,\dots,x_n,t),\quad j=1,\dots,N,$$

$$|\nu|=\nu_0+\nu_1+\dots+\nu_n,\qquad \nu_0<n_k. \tag{7}$$

This has a normal form if $n_k=m$ $(k=1,\dots,N)$. Now, taking the derivatives

$$\left(\frac{\partial}{\partial t}\right)^{n_j-1}u_j(x,t),\quad \left(\frac{\partial}{\partial t}\right)^{n_j-2}u_j(x,t),\dots,$$

$$\left(\frac{\partial}{\partial t}\right)u_j(x,t)$$

as new unknowns, we get another system:

$$\frac{\partial u_j}{\partial t}=\sum_{k=1}^{N'}\sum_{|\nu|\leqslant m}a_{jk\nu_1\dots\nu_n}(t)\frac{\partial^{\nu_1+\dots+\nu_n}}{\partial x_1^{\nu_1}\dots\partial x_n^{\nu_n}}u_k(x,t)$$

$$+C_j(x,t),\quad j=1,\dots,N'. \tag{8}$$

We take as the space of initial values the †topological linear space composed of all functions whose derivatives up to a sufficiently large order are bounded on the whole space $\mathbf{R}^n$ and equipped with the topology determined by the †seminorms that are the maximums of derivatives up to a given order on the whole space, and we take as the range space a similar space on the xt-space, where $x\in\mathbf{R}^n$ and $0\leqslant t\leqslant T$. Then we can formulate a necessary and sufficient condition for the well-posedness of the initial value problem for the future (the problem is regarded as specifying a mapping that assigns to the initial values on $t=0$ the values $u(x,t)$ for $t>0$). To give such a condition we consider the following system of ordinary differential equations, which are given by a †Fourier transformation on the x-space of system (8):

$$\frac{d\hat{u}_j}{dt}=\sum_{k=1}^{N'}\sum_{|\nu|\leqslant m}a_{jk\nu_1\dots\nu_n}(t)$$

$$\times(2\pi i\xi_1)^{\nu_1}\dots(2\pi i\xi_n)^{\nu_n}\hat{u}_k(\xi,t)+C_j(\xi,t),$$

$$j=1,\dots,N'. \tag{9}$$

If the †fundamental system of solutions of (9) is $v_i^{(j)}(\xi,t)$ $(i=1,\dots,N',j=1,\dots,N')$, then the condition is that these functions satisfy the inequalities

$$|v_i^{(j)}(\xi,t)|\leqslant C(1+|\xi|)^L,\qquad 0\leqslant t\leqslant T, \tag{10}$$

where C and L are constants independent of ξ. This is the condition obtained by Petrovskiĭ.

If system (7) is of normal form and is well posed for the future, it is also well posed for the past. In this case, equation (7) is said to be of **hyperbolic type** (→ 325 Partial Differential Equations of Hyperbolic Type). An example that is not of a normal form and is well posed is a parabolic equation (→ 327 Partial Differential Equations of Parabolic Type). In Petrovskiĭ's theory it is sufficient to assume that the coefficients in (7) are continuous and bounded, and we can take T arbitrarily large provided that (10) is satisfied. Hence Petrovskiĭ's theory guarantees global existence of solutions.

F. Uniqueness of Solutions

If a linear partial differential equation of the first order of normal form has coefficients of class C^ω, then the solution of class C^1 satisfying the prescribed initial conditions is unique (**Holmgren's uniqueness theorem**, 1901). Every system of partial differential equations of higher order of normal form can be reduced to a system of the first order of normal form. Therefore, if the coefficients are of class C^ω, there is only one solution for the original initial value problem with continuous partial derivatives up to the order of the equation. Moreover, if an analytic manifold S of dimension $n-1$ (n is the number of independent variables) is not †characteristic for the given equation of order m, there is at most one solution whose derivatives of order up to $m-1$ coincide with given functions on the manifold S. The proof of this fact relies on the Cauchy-Kovalevskaya existence theorem.

The uniqueness problem for the initial value problem is in general very difficult even for linear equations when the coefficients are not of class C^ω.

In particular, if the number of independent variables is 2 and the coefficients are all real, then we have a result of T. Carleman (1939) about the system:

$$\frac{\partial u_\mu}{\partial x}=\sum_{\nu=1}^{m} a_{\mu\nu}(x,y)\frac{\partial u_\nu}{\partial y}+\sum_{\nu=1}^{m} b_{\mu\nu}(x,y)u_\nu, \qquad (11)$$

where the $a_{\mu\nu}$ are of class C^2 and the $b_{\mu\nu}$ are continuous. He proved that if the eigenvalues of the matrix $(a_{\mu\nu})$ are all distinct, then even when some of the eigenvalues are complex, there is at most one solution of class C^1 for the initial value problem. If we omit the assumption about the eigenvalues, however, the theorem does not hold in general, because we have a counterexample due to A. Pliś (1954) where all coefficients are of class C^∞, $m=2$, $b_{\mu\nu}\equiv 0$.

A. P. Calderón showed that Carleman's result can be extended to the case $n>2$ (*Amer. J. Math.*, 80 (1958)). Consider the following linear partial differential equation of the kth order:

$$\frac{\partial^k u}{\partial x^k}+\sum_{j=1}^{k} P_j\left(x,y,\frac{\partial}{\partial y}\right)\left[\frac{\partial^{k-j}u}{\partial x^{k-j}}\right]+B[u]=0, \qquad (12)$$

where $P_j(x,y,\xi)$ is a homogeneous polynomial of degree j of $\xi=(\xi_1,\dots,\xi_n)$ with real coefficients and B is a differential operation in (x,y) of order at most $k-1$. We assume that the coefficients of $P_j(x,y,\xi)$ are functions of (x,y) of class C^1, their derivatives are Hölder continuous, and the coefficients of B are bounded and continuous. If the characteristic equation of (12),

$$\lambda^k+\sum_{j=1}^{k} P_j(x,y,\xi)\lambda^{k-j}=0,$$

has only distinct roots for $\xi\neq 0$, then the C^k-solution of the Cauchy problem is unique in a neighborhood of $x=a$. Calderón proved this except for the cases $k\geqslant 4$, $n=2$, where a certain topological difficulty arises. S. Mizohata (*J. Math. Soc. Japan*, 11 (1959)) succeeded in obtaining the proof for the exceptional cases.

This result can be extended to systems of equations under similar assumptions. See L. Hörmander [4] for an extension to the complex coefficient case. S. Mizohata, T. Shirota, and H. Kumanogo discuss the uniqueness theorem for equations of double characteristics or of parabolic type.

For nonlinear equations there are, in general, very few results about the global existence of solutions. For example, if the function F in equation (2) in the normal form satisfies the Lipschitz condition with constants A and B that are independent of x, then we get a global existence theorem. The method of proof of this theorem is as follows: First, we prove the existence for $|x|\leqslant\varepsilon_1$ and sufficiently small ε_1 by Picard's †successive iteration method. Then we regard $x=\varepsilon_1$ as the hyperplane on which the initial values are assigned and proved the existence of a solution on $\varepsilon_1\leqslant|x_1|\leqslant\varepsilon_2$, and so on. The same method can be applied to nonlinear systems of the first order if they are special types of quasilinear systems.

G. Construction of Solutions by Asymptotic Expansion

Let $a(x,D)=\sum_{|\alpha|\leqslant m} a_\alpha(x)D^\alpha$ be a linear partial differential operator of order m, with coefficients of class C^∞. We write $a(x,\xi)=\sum_{|\alpha|\leqslant m} a_\alpha(x)\xi^\alpha=h(x,\xi)+h'(x,\xi)+\dots$ with h and h' homogeneous in ξ of degree m and $m-1$, respectively. Let $K:\varphi(x)=0$ be a regular

surface of codimension 1. We assume that K is a simple characteristic, i.e., $h(x, \varphi_x)=0$ and $(\partial h/\partial \xi_i(x, \varphi_x)) \neq (0)$ on K. Let $f_j(t)$ $(j=0,1,\ldots)$ be a sequence of functions satisfying $df_j/dt(t)=f_{j-1}(t)$, $j=1,2,\ldots$. Then the equation $a(x,D)u(x)=0$ has a formal solution in the form

$$u(x)=\sum_{j=0}^{\infty} f_j(\varphi(x))u_j(x).$$

The coefficients $u_j(x)$ are obtained by solving successively the equations

$$L_1[u_j]=\sum_{k=2}^{m} L_k[u_{j-k+1}], \qquad (13)$$

where

$$L_1=\sum_{i=1}^{n}\frac{\partial h}{\partial \xi_i}(x,\varphi_x)D_i+\frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial^2 h}{\partial \xi_i \partial \xi_j}(x,\varphi_x)\varphi_{x_i x_j}$$
$$+h'(x,\varphi_x)$$

and L_k $(2 \leqslant k \leqslant m)$ are differential operators of order k depending only on a and φ. By this method of asymptotic expansion of solution, fundamental solutions and singularities of solutions for hyperbolic equations have been studied (Hadamard [5, 6]; R. Courant and P. Lax, *Proc. Nat. Acad. Sci. US*, 42 (1956); Lax, *Duke Math. J.*, 24 (1957); D. Ludwig, *Comm. Pure Appl. Math.*, 22 (1960); S. Mizohata, *J. Math. Kyoto Univ.*, 1 (1962); → 325 Partial Differential Equation of Hyperbolic Type).

If $a(x,D)$ is an operator with analytic coefficients, this formal solution is convergent. By using this fact, Mizohata constructed †null solutions. In fact, put $f_j(t)=t^{p+m+j}/\Gamma(p+m+j)$ $(p \geqslant 0)$, $j=0,1,\ldots$. Define $u(x)$ by $u(x)=0$ for $\varphi(x) \leqslant 0$ and $u(x)=\sum_{j=0}^{\infty} f_j(\varphi(x))u_j(x)$ for $\varphi(x) \geqslant 0$ obtained by the preceding process. Then $u(x)$ is a null solution of $a(x,D)u(x)=0$ (Mizohata, *J. Math. Kyoto Univ.*, 1 (1962)).

The Cauchy problem in the case when the initial surface has characteristic points had been studied by J. Leray and by L. Gårding, T. Kotake, and Leray. Let $a(x,D)$ be a differential operator with holomorphic coefficients in a complex domain. Let $S: s(x)=0$ be a regular surface and T be a subvariety of codimension 1 of S. Suppose that S is noncharacteristic on $S-T$, but characteristic on T. Consider the Cauchy problem (1) of Section A. Then the solution $u(x)$ is ramified around a characteristic surface K that is tangent to S on T, and it can be uniformized (Leray, *Bull. Soc. Math. France*, 85 (1957); Gårding, Kotake, and Leray, *Bull. Soc. Math. France*, 92 (1964)).

Next we consider the Cauchy problem (1) when the initial surface S $(x_1=0)$ is noncharacteristic, but the initial values $w_k(x)$ $(0 \leqslant k \leqslant m-1)$ have singularities on a regular subvariety T $(x_1=x_2=0)$ of S. We assume that $h(0,\xi_1,1,0,\ldots,0)=0$ has m distinct roots. Then there exist m characteristic surfaces $K_i: \varphi_i(x)=0$ $(1 \leqslant i \leqslant m)$ originating from T. $\varphi_i(x)$ is obtained by solving the †Hamilton-Jacobi equation $h(x,\varphi_x)=0$, $\varphi(0,x_2,\ldots,x_n)=x_2$. Now, if $w_k(x)$ $(0 \leqslant k \leqslant m-1)$ has a pole along T, the Cauchy problem (1) has a unique solution in the form

$$u(x)=\sum_{i=1}^{m}\left\{\frac{F_i(x)}{[\varphi_i(x)]^{p_i}}+G_i(x)\log\varphi_i(x)\right\}+H(x),$$

where $F_i(x)$, $G_i(x)$ $(1 \leqslant i \leqslant m)$ and $H(x)$ are holomorphic functions in a neighborhood of $x=0$ and p_i $(1 \leqslant i \leqslant m)$ are integers >0. In order to obtain this solution, we set $u(x)$ in the form $u(x)=\sum_{i=1}^{m}\sum_{j=0}^{\infty} f_j(\varphi_i(x))u_{i,j}(x)$, where $f_0(t)$ is a function with a pole or a logarithmic singularity at $t=0$ chosen so that $u(x)$ satisfies the initial conditions. Thus we can determine the coefficients $u_{i,j}(x)$ by solving successively the equations (13) on each K_i $(1 \leqslant i \leqslant m)$ (Y. Hamada, *Publ. Res. Inst. Math. Sci.*, 5 (1969); C. Wagschal, *J. Math. Pures Appl.*, 51 (1972)). When the multiplicity of characteristic roots is more than 1, the situation is not the same. For example, consider the Cauchy problem

$$D_1^2u-D_2u=0, \qquad u(0,x_2)=\frac{1}{x^2},$$

$$D_1u(0,x_2)=0.$$

The solution is

$$u(x)=\sum_{k=0}^{\infty}(-1)^n n!\,x_1^{2n}/(2n)!\,x_2^{n+1}$$

with essential singularities along $x_2=0$. This situation occurs quite generally. We factor $h(x,\xi)=h_1(x,\xi)^{r_1}\ldots h_s(x,\xi)^{r_s}$, where h_i $(1 \leqslant i \leqslant s)$ are irreducible polynomials of degree m_i in ξ with holomorphic coefficients. We assume that the equation $\prod_{i=1}^{s} h_i(0,\xi_1,1,0,\ldots,0)=0$ has p distinct roots $(p=m_1+\ldots+m_s)$. Then there exist p characteristic surfaces K_i $(1 \leqslant i \leqslant p)$ originating from T. We suppose more generally that the initial values $w_k(x)$ are multivalued functions ramified around T. Then the Cauchy problem (1) has a unique holomorphic solution on the universal covering space of $V-\bigcup_{i=1}^{p} K_i$, where V is a neighborhood of $x=0$. In fact, this is solved by transforming this problem into †integrodifferential equations. Such a method of solution is closely related to the method discussed in this section. See Hamada, Leray, and Wagschal (*J. Math. Pures Appl.*, 55 (1976)). In this case, even if the initial values have only poles, the solution in general may have essential singularities along $\bigcup_{i=1}^{p} K_i$, but if $a(x,D)$ satisfies Levi's condition ($a(x,D)$ is well decomposable), the solution does not yield essential singularities along $\bigcup_{i=1}^{p} K_i$. See J. De Paris (*J. Math. Pures Appl.*, 51 (1972)).

References

[1] R. Courant and D. Hilbert, Methods of mathematical physics II, Interscience, 1962.
[2] I. G. Petrovskiĭ, Lectures on partial differential equations, Interscience, 1954. (Original in Russian, second edition, 1953.)
[3] L. Hörmander, Linear partial differential operators, Springer, 1963.
[4] E. Goursat, Cours d'analyse mathématique, Gauthier-Villars, II, second edition, 1911; III, fourth edition, 1927.
[5] J. Hadamard, Leçons sur la propagation des ones et les équations de l'hydrodynamique, Hermann, 1903 (Chelsea, 1949).
[6] J. Hadamard, Le problème de Cauchy et les équations aux dérivées partielles linéaires hyperboliques, Hermann, 1932.
[7] C. Carathéodory, Calculus of variations and partial differential equations of first order, Holden-Day, 1965. (Original in German, 1935.)
[8] S. Mizohata, Theory of partial differential equations, Cambridge University Press, 1973. (Original in Japanese, 1965.)
[9] F. Treves, Linear partial differential equations with constant coefficients, Gordon & Breach, 1966.
[10] F. John, Partial differential equations, Springer, 1971.
[11] J. Dieudonné, Eléments d'analyse, 4, 7, 8, Gauthier-Villars, 1971–1978.

322 (XIII.20) Partial Differential Equations (Methods of Integration)

A. General Remarks

The methods of integrating partial differential equations are not as simple as those for †ordinary differential equations. For ordinary differential equations, we often obtain the desired solution by first finding the general solution containing several arbitrary constants and then specializing those constants to satisfy prescribed additional conditions. The situation is more complicated, however, for partial differential equations. In the formal general solution of a partial differential equation we have arbitrary functions instead of arbitrary constants, and there are cases where it is impossible, or very difficult, to find a suitable specialization of these functions so that the given additional conditions are fulfilled. For this reason, methods of solution are rather specific and are classified according to the types of additional conditions. In many cases, we assume that a problem is mathematically well posed, guess the possible solutions, and verify it directly.

A problem is said to be mathematically **well posed (properly posed** or **correctly posed)** if, under assigned additional conditions, the solution (i) exists, (ii) is uniquely determined, and (iii) depends continuously on the assigned data. By carefully examining problems in physics and engineering, we usually obtain many well-posed and important problems. In these problems, usually the data are sufficiently smooth functions. In these cases, part (iii) of the above definition (the continuous dependence of the solutions on the data) follows often from assumptions (i) and (ii). For example, †elliptic equations like †Laplace's equation $u_{xx}+u_{yy}=0$ for $u(x,y)$ describe laws of static or stationary phenomena such as the field of universal gravitation, the electrostatic field, the magnetostatic field, the steady flow of incompressible fluids without vortices, and the steady flow of electricity or heat. For this equation, †boundary value problems are well posed, but †initial value problems are not (→ 323 Partial Differential Equations of Elliptic Type). By contrast, for †hyperbolic equations like $u_{tt}-u_{xx}=0$ for $u(x,t)$ and †parabolic equations like $u_t-u_{xx}=0$ for $u(x,t)$ which control the change (in reference to time) of the various stationary phenomena, initial value problems or mixed problems with both boundary conditions and initial conditions are well posed (→ 325 Partial Differential Equations of Parabolic Type).

For the rest of this article we explain fundamental and typical methods of integration (→ Appendix A, Table 15).

B. The Lagrange-Charpit Method

For the partial differential equation of the first order

$$F(x,y,u,p,q)=0, \qquad p=\frac{\partial u}{\partial x}, \qquad q=\frac{\partial u}{\partial y}, \tag{1}$$

we consider a system of ordinary differential equations called †characteristic differential equations:

$$\frac{dx}{F_p}=\frac{dy}{F_q}=\frac{du}{pF_p+qF_q}=\frac{-dp}{F_x+pF_u}=\frac{-dq}{F_y+qF_u}. \tag{2}$$

If we obtain at least one †integral of this system containing p, q, and an arbitrary constant a in the form

$$G(x,y,u,p,q)=a, \tag{3}$$

and if we find p and q from (1) and (3), then $du=p\,dx+q\,dy$ is an †exact differential form, and by integrating it we get a solution $\Phi(x,y,u,a,b)$

$=0$ of (1). A solution of (1) containing two arbitrary constants is called a †complete solution. Here, setting $b=g(a)$ and eliminating a from the two equations $\Phi(x,y,u,a,g(a))=0$ and $\Phi_a+\Phi_b g'(a)=0$, we obtain a solution involving an arbitrary function. Such a solution is called a †general solution of (1). For example, consider $pq=1$. Since the characteristic differential equations are

$$\frac{dx}{q}=\frac{dy}{p}=\frac{du}{2pq}=\frac{-dp}{0}=\frac{-dq}{0},$$

we have $p=a$ (constant), and hence from the original equation we get $q=a^{-1}$. Then $du=a\,dx+a^{-1}\,dy$ is an exact differential form, and by integrating it a complete solution $u=ax+a^{-1}y+b$ is obtained. A general solution is found by eliminating a from $u=ax+a^{-1}y+g(a)$ and $0=x-a^{-2}y+g'(a)$.

Likewise, when we have n independent variables $x_1,\dots,x_n$ in the equation

$$F(x_1,\dots,x_n,u,p_1,\dots,p_n)=0, \qquad p_i=\frac{\partial u}{\partial x_i}, \tag{1'}$$

we can use the †characteristic differential equations to find a complete solution and a general solution (the **Lagrange-Charpit method**; → 82 Contact Transformations).

C. Separation of Variables and the Principle of Superposition

The simplest and most useful method is the **separation of variables.** Concerning the equation $u_x^2+u_y^2=1$ in $u(x,y)$, for example, by setting $u=\varphi(x)+\psi(y)$ we obtain $\varphi'(x)^2+\psi'(y)^2=1$ or $\varphi'(x)^2=1-\psi'(y)^2$. Since the right-hand side is independent of x and the left-hand side is independent of y, both sides must be equal to the same constant α^2. From this we get a (complete) solution involving two arbitrary constants α, β:

$$u=\alpha x+\sqrt{1-\alpha^2}\;y+\beta.$$

For †linear equations, it is often effective to write the solution as a product $u=\varphi(x)\psi(y)$. From the †heat equation $u_y=u_{xx}$, we obtain by this process a relation $\psi'(y)/\psi(y)=\varphi''(x)/\varphi(x)$, from which we get a particular solution $u=ae^{-v^2y}\sin v(x-\alpha)$ containing a parameter v.

Next, when the equation is linear and homogeneous, by forming a linear combination of particular solutions that correspond to various values of a parameter v, we obtain a new solution (the **principle of superposition**). For example, by integrating the solution $e^{-v^2y}\cos vx$ with respect to the parameter v between the limits $-\infty$ and ∞ (namely, by a superposition consisting of a linear combination and a limiting process), we obtain a new solution

$$u=\int_{-\infty}^{\infty}e^{-v^2y}\cos vx\,dv=\sqrt{\frac{\pi}{y}}\exp\left(-\frac{x^2}{4y}\right) \qquad (y>0), \tag{4}$$

which is the †fundamental solution of the heat equation. This name refers to the fact that the function (4) can be used to obtain solutions of the heat equation under some initial conditions. More exactly, a solution of $u_y-u_{xx}=0$ that is a function of class C^2 in $y>0$, is continuous in $y\geqslant 0$, and coincides with a bounded continuous function $\varphi(x)$ on $y=0$ can be obtained by a superposition of the solution (4) such as

$$u(x,y)=\frac{1}{2\sqrt{\pi y}}\int_{-\infty}^{\infty}\varphi(\xi)\exp\left(-\frac{(x-\xi)^2}{4y}\right)d\xi.$$

The method of separation of variables applied after a suitable transformation of variables is often successful. In particular, by using orthogonal coordinates, †polar coordinates, or †cylindrical coordinates according to the form of boundary, we often obtain satisfactory results. For example, concerning the boundary value problem for Laplace's equation $\Delta u=u_{xx}+u_{yy}=0$, which is smooth in the circle $r^2=x^2+y^2<1$ and takes the value of a given continuous function $g(\theta)$ on the circumference $r=1$, we can use polar coordinates to rewrite the equation in the form

$$\Delta u=u_{rr}+\frac{u_r}{r}+\frac{1}{r^2}u_{\theta\theta}=0$$

and apply the method of separation of variables to obtain particular solutions $r^n\cos n\theta$, $r^n\sin n\theta$. Hence it is reasonable to suspect that by a superposition of these particular solutions we can obtain the desired solution:

$$u(x,y)=\frac{a_0}{2}+\sum_{n=1}^{\infty}(a_n\cos n\theta+b_n\sin n\theta)r^n.$$

In fact, this series is a desired solution if the coefficients a_n, b_n can be chosen so that the series converges uniformly for $0\leqslant r\leqslant 1$ and can be differentiated twice term by term for $0\leqslant r<1$, and if we have

$$g(\theta)=\frac{a_0}{2}+\sum_{n=1}^{\infty}(a_n\cos n\theta+b_n\sin n\theta).$$

By virtue of the uniqueness of the solution of an elliptic equation, this is the unique desired solution. The boundary value problem in the preceeding paragraph is well posed.

D. Mixed Problems

For linear homogeneous equations of hyperbolic or parabolic type, **mixed problems** fre-

quently appear, i.e., problems in which both initial and boundary conditions are assigned. These problems are, furthermore, classified into two types.

For the first type, homogeneous boundary conditions are assigned. For example, in vibration problems of a nonhomogeneous string between $0 \leqslant x \leqslant l$, we must find the solution of the equation

$$(T(x)u_x)_x = \rho u_{tt}$$

satisfying an initial condition $u = f(x)$, $u_t = g(x)$ for $t = 0$, under homogeneous boundary conditions: (i) $u = 0$ for $x = 0$ and $x = l$ in the case of two fixed ends; (ii) $u_x = 0$ for $x = 0$ and $x = l$ in the case of two free ends; (iii) $-u_x + \sigma_0 u = 0$ for $x = 0$, and $u_x + \sigma_1 u = 0$ for $x = l$ ($\sigma_0 > 0$, $\sigma_1 > 0$), where the two ends are tied elastically to the fixed points; (iv) $T(0)u(0) = T(l)u(l)$, $T(0)u'(0) = T(l)u'(l)$ (periodicity condition); (v) u, u' are finite at $x = 0$ and $x = l$ (regularity condition).

The method of separation of variables is applicable to problems of this type also. For example, suppose that we have two fixed ends. Disregarding the initial condition for a while, we find a particular solution fulfilling only the boundary condition by setting $u(x,t) = y(x)\exp ivt$. Here, $y(x)$ must satisfy

$$(T(x)y')' + v^2\rho(x)y = 0, \quad y(0) = y(l) = 0. \tag{5}$$

Except when the solution is trivial (i.e., $u(x,t) = 0$), $y(x) \not\equiv 0$ must hold. But in the special case $T(x) \equiv 1$, $\rho(x) \equiv 1$, $l = \pi$, the values of v for which functions of this kind exist are only $1, 2, 3, \ldots$ (the corresponding $y(x)$ is $\sin vx$).

Also, in more general cases, nontrivial solutions $y(x)$ exist only for some discrete values of v. These v are called †eigenvalues of (5), and the corresponding solutions $y(x)$ are called †eigenfunctions for the eigenvalues v. That is, the desired particular solution can be obtained by solving the †eigenvalue problem (5). Again, when $T(x) \equiv \rho(x) \equiv 1$, $l = \pi$, particular solutions are $\sin nx \exp int$, $\sin nx \exp(-int)$ ($n = 1, 2, \ldots$). We consider

$$u(x,t) = \sum_{n=1}^{\infty} (a_n \cos nt + b_n \sin nt) \sin nx, \tag{6}$$

which is obtained by a superposition of these particular solutions. If we can determine the coefficients a_n, b_n so that this series converges uniformly and is twice differentiable term by term, and

$$f(x) = \sum_{n=1}^{\infty} a_n \sin nx, \quad g(x) = \sum_{n=1}^{\infty} n b_n \sin nx,$$

then the series (6) is a solution of the mixed problem in question. If the uniqueness of the solution of the mixed problem is proved, it is not necessary to look for any solution other than the one obtained by combining the method of separation of variables and the principle of superposition.

Furthermore, the method described in this section is applicable to solving the following nonhomogeneous equation, which characterizes the motion of a string under the influence of an external force $f(x,t)$:

$$u_{tt} - u_{xx} = f(x,t) \tag{7}$$

under the boundary condition for the first type of mixed problem and the initial condition $u = u_t = 0$ for $t = 0$. In this case, we expand the unknown u and the function $f(x,t)$ in terms of the system of eigenfunctions $\{\sin nx\}$, and by substituting

$$u = \sum_{n=1}^{\infty} a_n(t) \sin nx, \qquad f(x,t) = \sum_{n=1}^{\infty} A_n(t) \sin nx$$

into (7), we reduce the problem to determining $a_n(t)$ ($n = 1, 2, \ldots$). When the external force varies with a harmonic oscillation over time as in $f(x,t) = -\varphi(x)\exp(-i\omega t)$, a similar method can be applied by setting $u = v(x)\exp(-i\omega t)$.

For mixed problems of the second type, the homogeneous initial condition $u = 0$, $u_t = 0$ for $t = 0$ is assigned, but the boundary condition is nonhomogeneous. For example, when an oscillating string is at rest until $t = 0$, and for $t > 0$ its right end is fixed and its left end moves subject to an assigned rule, the behavior of the string is described by the solution of $u_{tt} - u_{xx} = 0$ under the boundary condition $u(0,t) = f(t)$, $u(l,t) = 0$ ($t > 0$). This is called a **transient problem**. If we now choose an arbitrary function $B(x,t)$ that fulfills all the boundary and initial conditions, and if we set $u - B(x,t) = v(x,t)$ and $B_{xx} - B_{tt} = f(x,t)$, then v satisfies

$$v_{tt} - v_{xx} = f(x,t) \quad \text{for } t > 0$$

with $v = v_t = 0$ for $t = 0$; $v = 0$ for $x = 0$ and $x = l$. Then $v(x,t)$ describes the oscillation of a string that is at rest until $t = 0$ and moves under the effect of an external force represented by $f(x,t)$ for $t > 0$. Problems of this kind often appear in electrical engineering.

Such problems can be reduced to problems of the first type in the manner described in the previous paragraph, but there are some direct methods that are more effective, the first being **Duhamel's method**. Consider the case where $f(t)$ is the unit impulse function:

$$f(t) = \begin{cases} 1, & t > 0, \\ 0, & t < 0, \end{cases}$$

and let $U(x,t)$ be a solution corresponding to this case and vanishing for $t \leqslant 0$, $x > 0$. Then a solution corresponding to the general case is given by

$$u = \int_0^t \frac{\partial}{\partial t} U(x, t-\tau) f(\tau)\, d\tau.$$

The second method is based on application of the †Laplace transformation. Denoting the Laplace transform of a solution $u(x,t)$ by

$$\int_0^\infty ue^{-pt}\,dt=\frac{v(x,p)}{p},$$

we multiply both sides of $u_{tt}-u_{xx}=0$ by e^{-pt} and integrate the result with respect to t from 0 to ∞. Then, taking account of the initial condition, we have

$$v_{xx}-p^2v=0$$

with the boundary condition

$$v(0,p)=p\int_0^\infty f(t)e^{-pt}\,dt, \qquad v(l,p)=0.$$

If for $p=\alpha+i\beta$ $(\alpha>\alpha_0)$ we can find a solution $v(x,p)$ of this boundary value problem for the ordinary differential equation, then the desired solution is given by

$$u(x,t)=\frac{1}{2\pi i}\int_L \frac{v(x,p)}{p}e^{pt}\,dp,$$

where L is a path in $\alpha>\alpha_0$ parallel to the imaginary axis. This is called the **Bromwich integral**.

E. Green's Formula and Application of Fundamental Solutions

Given a linear partial differential operator

$$L[u]\equiv\sum_p a_p(x)D^pu, \quad p=(p_1,\ldots,p_n),$$

$$D^p=\partial^{p_1+\ldots+p_n}/\partial x_1^{p_1}\ldots dx_n^{p_n},$$

we call the operator

$$L^*[v]\equiv\sum_p(-1)^{p_1+\ldots+p_n}D^p(\bar{a}_p(x)v)$$

the **adjoint operator** of L, and the operator

$${}^tL[v]\equiv\sum_p(-1)^{p_1+\ldots+p_n}D^p(a_p(x)v)$$

the **transposed operator** of L. Sometimes ${}^tL$ is also called the adjoint operator of L. In the complex Hilbert space, the adjoint operators are more appropriate than the transposed operators. In this case, the operator L^* defined above is usually called the **formal adjoint operator** to distinguish it from the one defined by

$$(L[u],v)=(u,L^*[v]) \text{ for all } u\in D(L),$$

$D(L)$ being the domain of the definition of L (→ 251 Linear Operators). These transposed or adjoint operators are often used to represent (at least locally) the solutions u of $L[u]=f$.

We explain in more detail the specific case where the number of independent variables is 2. For the linear partial differential operator

$$L(u)=A(x,y)\frac{\partial^2u}{\partial x^2}+2B(x,y)\frac{\partial^2u}{\partial x\,\partial y}+C(x,y)\frac{\partial^2u}{\partial y^2}$$
$$+D(x,y)\frac{\partial u}{\partial x}+E(x,y)\frac{\partial u}{\partial y}+F(x,y)u$$

and its adjoint partial differential operator

$$M(v)=\frac{\partial^2(Av)}{\partial x^2}+2\frac{\partial^2(Bv)}{\partial x\,\partial y}+\frac{\partial^2(Cv)}{\partial y^2}$$
$$-\frac{\partial(Dv)}{\partial x}-\frac{\partial(Ev)}{\partial y}+Fv,$$

we have †Green's formula:

$$\iint_D(vL(u)-uM(v))\,dx\,dy$$
$$=-\int_{\partial D}\left(P\frac{\partial x}{\partial n}+Q\frac{\partial y}{\partial n}\right)ds, \tag{8}$$

where

$$P=v\{Au_x+Bu_y\}-u\{(Av)_x+(Bv)_y\}+Duv,$$
$$Q=v\{Bu_x+Cu_y\}-u\{(Bv)_x+(Cv)_y\}+Euv, \tag{9}$$

and ∂D denotes the boundary curve of the domain D, n the internal normal of ∂D at a point of ∂D, and s the arc length.

We can apply this formula for solving a nonhomogeneous equation $L(u)=f$ as follows. Assume that there exists a solution of $L(u)=f$ satisfying the assigned additional condition, and choose a †fundamental solution of $M(v)=0$ having an adequate singularity at a point (x_0,y_0) of D and fulfilling a suitable boundary condition. Then, if we substitute these solutions u and v into (8), we obtain an explicit representation of $u(x_0,y_0)$. If we can verify that the function $u(x,y)$ thus obtained is a solution of $L(u)=f$ fulfilling the assigned additional conditions, then we see, under the assumption of uniqueness of solutions, that this and only this function $u(x,y)$ is the desired solution. For example, consider a boundary problem for $L(u)=\frac{\partial^2u}{\partial x^2}+\frac{\partial^2u}{\partial y^2}$, for which $M(v)=\frac{\partial^2v}{\partial x^2}+\frac{\partial^2v}{\partial y^2}$. The problem is, for a circle of radius r with center at the origin, to find a function $u(x,y)$ that is continuous in the interior and on the circumference C_r of the circle, and that satisfies $L(u)=f$ in the interior of the circle (f is bounded and continuous in the interior of the circle) and is equal to a given continuous function g on the circumference C_r. In this case, we consider the circumference K_ε of sufficiently small radius ε with center (x_0,y_0) contained in the interior of the first circle, and set

$$v(x,y)=(1/2\pi)\log 1/\rho+h(x,y),$$
$$\rho=((x-x_0)^2+(y-y_0)^2)^{1/2}. \tag{10}$$

We apply formula (8) to the domain D_ε enclosed by the circumferences K_ε and C_r. If, in (10), h satisfies $M(v)=0$ in the interior of the circle enclosed by C_r and vanishes on C_r, then we get

$$\iint_{D_\varepsilon} vf\,dx\,dy = -\int_{K_\varepsilon}(vu_n - uv_n)\,ds + \int_{C_r} gv\,ds.$$

By letting ε tend to zero, the first term on the right-hand side yields

$$-u(x_0, y_0)\int_0^{2\pi}\frac{1}{2\pi}\frac{\partial\log\rho}{\partial\rho}\rho\,d\theta = -u(x_0, y_0),$$

by the logarithmic singularity of v at the point (x_0, y_0). Therefore we have an explicit representation of u,

$$u(x_0, y_0) = \int_{C_r} gv_n\,ds - \iint_{x^2+y^2\leqslant r^2} vf\,ds\,dy, \qquad (11)$$

and it is easily verified that this is the desired solution.

As stated in the previous paragraph, to apply Green's formula it is necessary to find a solution v of $M(v)=0$ possessing a fundamental singularity as the logarithmic singularity, i.e., a so-called fundamental solution. As fundamental singularities for

$$M(v)=\frac{\partial v}{\partial y}-\frac{\partial^2 v}{\partial x^2} \text{ and } M(v)=\frac{\partial^2 v}{\partial y^2}-\frac{\partial^2 v}{\partial x^2}$$

of parabolic type and of hyperbolic type, we must take those given respectively by

$$v(x,y) = \begin{cases} \dfrac{1}{2\sqrt{\pi(y-y_0)}}\exp\left(-\dfrac{(x-x_0)^2}{4(y-y_0)}\right), & y>y_0, \\ 0, \quad y\leqslant y_0, \end{cases}$$

and

$$v(x,y)=\begin{cases} 1/2, & |x-x_0|<y-y_0, \\ 0, & |x-x_0|\geqslant y-y_0 \end{cases}$$

(→ 323 Partial Differential Equations of Elliptic Type; 325 Partial Differential Equations of Hyperbolic Type; 327 Partial Differential Equations of Parabolic Type).

References

[1] R. Courant and D. Hilbert, Methods of mathematical physics, Interscience, I, 1953; II, 1962.
[2] E. Goursat, Cours d'analyse mathématique, Gauthier-Villars, III, 1927.
[3] A. G. Webster and G. Szegö, Partielle Differentialgleichungen der mathematischen Physik, Teubner, 1930.
[4] L. Schwartz, Mathematics for the physical sciences, Hermann, revised edition, 1966.
[5] J. L. Lions, Equations différentielles opérationelles et problèmes aux limites, Springer, 1961.
[6] A. Friedman, Partial differential equations, Holt, Rinehart and Winston, 1969.

323 (XIII.24) Partial Differential Equations of Elliptic Type

A. General Remarks

Suppose that we are given a †linear second-order partial differential equation

$$L[u]\equiv\sum_{i,j=1}^n a_{ij}(x)\frac{\partial^2 u}{\partial x_i\partial x_j}+\sum_{i=1}^n b_i(x)\frac{\partial u}{\partial x_i}+c(x)u = f(x), \qquad (1)$$

where $x=(x_1,\dots,x_n)$, $a_{ji}=a_{ij}$. If the quadratic form $\sum a_{ij}\xi_i\xi_j$ in ξ is †positive definite at every point x of a domain G, this equation (or the operator L) is said to be **elliptic** (or of **elliptic type**) in G. For $n=2$, $a_{11}(x)a_{22}(x)-(a_{12}(x))^2>0$ is the condition of ellipticity. In this case, by a change of independent variables, the equation is transformed locally into the canonical form

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+b_1(x,y)\frac{\partial u}{\partial x}+b_2(x,y)\frac{\partial u}{\partial y}+c(x,y)u = f(x,y).$$

The operator $\sum_{i=1}^n \partial^2/\partial x_i^2$, denoted by Δ, is called **Laplace's operator** (or the **Laplacian**). The simplest examples of elliptic equations are $\Delta u=0$ (**Laplace's differential equation**) and $\Delta u=f(x)$ (**Poisson's differential equation**) (→ Appendix A, Table 15).

B. Fundamental Solutions

Let K be the n-dimensional ball with radius R, center x_0, boundary Ω (an $(n-1)$-dimensional sphere), and area S_n, and let r be the distance from x_0 to x. Then for any function $u(x)$ of class C^2 we have

$$u(x_0)=\frac{1}{S_n}\int_\Omega u\,dS-\frac{R^{n-1}}{(n-2)S_n}\times\int_K (r^{2-n}-R^{2-n})\Delta u\,dx \qquad \text{for } n>2,$$

$$u(x_0)=\frac{1}{2\pi}\int_\Omega u\,dS-\frac{1}{2\pi}\int_K \log\frac{R}{r}\Delta u\,dx \qquad \text{for } n=2.$$

Thus, if u is a solution of Poisson's equation $\Delta u=f(x)$, we have a representation of $u(x_0)$ by replacing Δu by f in the integrals just given. Next, concerning the solutions of the equation $\Delta u+cu=0$ ($c>0$, constant), the following relation holds:

$$\frac{1}{S_n}\int_\Omega u\,dS=\frac{\Gamma(n/2)J_{n'}(R\sqrt{c})}{(R\sqrt{c}/2)^{n'}}u(x_0),$$

where Γ is the †gamma function, $n'=(n-2)/2$, and J_ν is the †Bessel function of order ν.

Now, if we put

$$V(x,\xi)=\begin{cases}\left(\sum_{i=1}^{n}(x_i-\xi_i)^2\right)^{1-n/2}=r^{2-n}, & n\geqslant 3,\\ \log\left(\sum_{i=1}^{2}(x_i-\xi_i)^2\right)^{-1/2}=\log\frac{1}{r}, & n=2,\end{cases}$$

then the function $u(x)$ defined by

$$u(x)=-\frac{1}{\omega_n}\int_G V(x,\xi)f(\xi)\,d\xi,$$

$$\omega_n=2\pi^{n/2}\frac{n-2}{\Gamma(n/2)}$$

represents a †particular solution of $\Delta u=f(x)$, where $V(x,\xi)$ as a function of the variables x satisfies $\Delta V(x,\xi)=0$ except at $x=\xi$.

Consider now the more general case (1). A function $E(x,\xi)$ is called a **fundamental solution** (or **elementary solution**) of (1) or of L if

$$u(x)=\int_G E(x,\xi)f(\xi)\,d\xi$$

provides a solution of (1) for any $f(x)\in C^\infty$ with compact support. A fundamental solution $E(x,\xi)$ is a solution of the equation $L[E]=0$ having a singularity at $x=\xi$ of the form $-\omega_n^{-1}\sqrt{a(\xi)}\,r^{2-n}$ ($n\geqslant 3$) or $\omega_2^{-1}\sqrt{a(\xi)}\log r$ ($n=2$), where $a(\xi)=\det(a^{ij}(\xi))$ and $r=(\sum a^{ij}(\xi)(x_i-\xi_i)(x_j-\xi_j))^{1/2}$ and (a^{ij}) is the inverse matrix of (a_{ij}).

Roughly, there are three different methods of constructing fundamental solutions. The first and most general is to use pseudodifferential operators (→ 345 Pseudodifferential Operators). The second is that of J. Hadamard [1], which uses the geodesic distance between two points x and ξ with respect to the Riemannian metric $\sum a^{ij}(x)dx_i dx_j$. This is important in applications of elliptic equations to geometry. The third method is due to E. E. Levi [2] and is as follows: Let

$$V(x,\xi)=\begin{cases}r^{2-n}, & n\geqslant 3,\\ -\log r, & n=2.\end{cases}$$

To obtain a solution of $L[u]=f(x)$ we set

$$u=\int_G V(x,\xi)\varphi(\xi)\sqrt{a(\xi)}\,d\xi.$$

Writing $L[V(x,\xi)]=\chi(x,\xi)$, we obtain the following †integral equation of Fredholm type in $\varphi(x)$:

$$\varphi(x)-\frac{1}{\omega_n}\int_G \chi(x,\xi)\varphi(\xi)\sqrt{a(\xi)}\,d\xi=\frac{1}{\omega_n}f(x).$$

If we denote the †resolvent of this equation by $\bar\chi(x,\xi)$ and put

$$\gamma(x,\xi)=V(x,\xi)-\omega_n\int_G V(x,\xi')\bar\chi(\xi',\xi)\sqrt{a(\xi')}\,d\xi',$$

then $E(x,\xi)=-\gamma(x,\xi)\sqrt{a(\xi)}/\omega_n$ is seen to be a fundamental solution of L. Thus Levi's method enables us to construct a fundamental solution locally by successive approximation, because the integral equation in φ as above has a unique solution expressible by Neumann series if the domain G is small enough (→ 189 Green's Operator).

C. The Dirichlet Problem

Let G be a bounded domain with boundary Γ. We call the problem of finding a solution u of the given elliptic equation in G that is continuous on $G\cup\Gamma$ and takes the assigned continuous boundary values on T the **first boundary value problem** (or **Dirichlet problem**). In particular, the Dirichlet problem for $\Delta u=0$ has been studied in detail (→ 120 Dirichlet Problem).

If $c(x)<0$ and $f(x)\leqslant 0$ or if $c(x)\leqslant 0$ and $f(x)<0$ in (1), a solution u does not attain its local negative minimum in G. This is called the **strong maximum principle** (→ [3, 4] for **Hopf's maximum principle** and **Giraud's theorem**). The maximum principle is one of the most powerful tools available for the treatment of elliptic equations of the second order with real coefficients. From this it follows that the solution of the Dirichlet problem for (1) is unique if $c(x)<0$. Furthermore, concerning the uniqueness of the solution, we have the following criterion: If there exists a function $\omega(x)>0$ of class C^2 in G and continuous on $G\cup\Gamma$ such that $L[\omega(x)]<0$, then the solution of the Dirichlet problem is unique.

Let G be a †regular domain in the plane. If we denote †Green's function of Δ relative to the Dirichlet problem in G by $K(x,y;\xi,\eta)$, then the solution u of $\Delta u=f(x,y)$ vanishing on Γ is given by

$$u(x,y)=-\frac{1}{2\pi}\iint_G K(x,y;\xi,\eta)f(\xi,\eta)\,d\xi\,d\eta,$$

where $f(x,y)$ satisfies a †Hölder condition.

The Dirichlet problem for

$$L[u]\equiv\Delta u+a(x,y)\frac{\partial u}{\partial x}+b(x,y)\frac{\partial u}{\partial y}+c(x,y)u=f(x,y)$$

with prescribed boundary values reduces to the problem in the previous paragraph. In fact, let $h(x, y)$ be a function that coincides with the prescribed boundary value on Γ. Then if we put $u = h(x, y) + v$, the problem is reduced to finding a solution v satisfying an equation similar to the one for $L[u]$ (replacing f by $f - L[h]$) and vanishing on Γ. Now suppose that Γ consists of a finite number of †Jordan curves whose †curvatures vary continuously, a and b are continuously differentiable, and c and f satisfy Hölder conditions. Let $K(x, y; \xi, \eta)$ be Green's function for the Dirichlet problem relative to Δ. To find a solution of $L[u] = f$ vanishing on Γ we set

$$u = -\frac{1}{2\pi}\iint_G K(x, y; \xi, \eta)\rho(\xi, \eta)\,d\xi\,d\eta.$$

Then ρ is a solution of the integral equation

$$\rho(x, y) + \iint_G H(x, y, \xi, \eta)\rho(\xi, \eta)\,d\xi\,d\eta = f(x, y),$$

where $H(x, y, \xi, \eta) = (-1/2\pi)\{a(x, y)K_x(x, y; \xi, \eta) + b(x, y)K_y(x, y; \xi, \eta) + c(x, y)K(x, y; \xi, \eta)\}$. Therefore we have the following alternatives: Either $L[u] = f$ has a unique solution for any f and any given boundary values on Γ, or $L[u] = 0$ has nontrivial solutions vanishing on Γ (in this case the number of linearly independent solutions is finite).

More general equations of type (1) have been studied by J. Schauder and others. Schauder proved, first, that when the $b_i(x)$ and $c(x)$ are zero, $a_{ij}(x)$ and $f(x)$ satisfy Hölder conditions, and the boundary Γ of G is of class C^2, then there exists a unique solution $u(x)$ vanishing on Γ. Next, when the $b_i(x)$ and $c(x)$ are continuous and a_{ij} and f satisfy Hölder conditions, he showed the following alternatives: Either $L[u] = f$ admits a unique solution vanishing on Γ for every f, or $L[u] = 0$ has nontrivial solutions vanishing on Γ (in this case, the number of linearly independent solutions is finite).

In (1), suppose that a_{ij}, b_i, and c are Hölder continuous of exponent α $(0 < \alpha < 1)$ uniformly on $\bar{G}$ and that Γ is of class $C^{2+\alpha}$. Then we have the following inequality **(Schauder's estimate)** for any $u \in C^{2+\alpha}(\bar{G})$:

$$\|u\|_{2+\alpha, \bar{G}} \leqslant K_1(\|L[u]\|_{\alpha, \bar{G}} + \|u\|_{2+\alpha, \Gamma}) + K_2\|u\|_{0, \bar{G}}, \qquad (2)$$

where $\|\cdot\|_{k,S}$ stands for the norm in the function space $C^k(S)$ (→ 168 Function Spaces). K_1 and K_2 are positive constants depending on L, G, and α but independent of u. More precisely, K_1 depends only on the ellipticity constant λ of L defined as the smallest number $\geqslant 1$ such that

$$\lambda^{-1}|\xi|^2 \leqslant \sum a_{ij}(x)\xi_i\xi_j \leqslant \lambda|\xi|^2 \qquad (3)$$

for any $(x, \xi) \in \bar{G} \times \mathbf{R}^n$. The inequality (2) is one of the most important **a priori estimates** in the theory of elliptic equations [5, 6] (→ inequality (9) in Section H).

D. Quasilinear Partial Differential Equations

Consider the second-order partial differential equation in $u(x_1, \ldots, x_n)$:

$$F(x_1, \ldots, x_n, u, p_1, \ldots, p_n, p_{11}, \ldots, p_{ij}, \ldots, p_{nn}) = 0,$$

where $p_i = \partial u/\partial x_i$, $p_{ij} = \partial^2 u/\partial x_i \partial x_j$. If, for a solution $u(x)$, $\sum_{i,j=1}^n (\partial F/\partial p_{ij})\xi_i\xi_j$ is a positive definite form, we say that the equation is **elliptic** at $u(x)$. Moreover, if for any values of u, p_i, p_{ij} this quadratic form is positive definite, we say simply that the equation is of **elliptic type**. The equation is called **quasilinear** if F is linear in p_{ij}. For example, the equation of †minimal surfaces (→ 334 Plateau's Problem)

$$(1 + p_2^2)p_{11} - 2p_1p_2p_{12} + (1 + p_1^2)p_{22} = 0$$

is a quasilinear elliptic equation. Furthermore,

$$\Delta u = f(x, y, u, u_x, u_y)$$

is also elliptic. E. Picard solved this equation by the **method of successive approximation** [7]. Specifically, let $h(x, y)$ be the †harmonic function taking the assigned boundary values on Γ. Starting from $u_0(x, y) = h(x, y)$, we define the functions $u_n(x, y)$ successively as the solutions of

$$\Delta u_n = f(x, y, u_{n-1}, \partial u_{n-1}/\partial x, \partial u_{n-1}/\partial y),$$

coinciding with h on Γ. Let $K(x, y; \xi, \eta)$ be Green's function in G and in the region defined by $|u - h(x, y)| \leqslant A$, $|p - h_x| \leqslant B$, $|q - h_y| \leqslant B$, $(x, y) \in G$, and let the supremum of $|f(x, y, u, p, q)|$ be C. Assume now that

$$\frac{C}{2\pi}\iint_G K\,d\xi\,d\eta \leqslant A, \qquad \frac{C}{2\pi}\iint_G |K_x|\,d\xi\,d\eta \leqslant B,$$

$$\frac{C}{2\pi}\iint_G |K_y|\,d\xi\,d\eta \leqslant B,$$

and that f satisfies a Hölder condition in (x, y). Moreover, let

$$|f(x, y, u', p', q') - f(x, y, u, p, q)| \leqslant L|u - u'| + L'(|p' - p| + |q' - q|).$$

Assume finally that

$$\iint_G (LK + L'(|K_x| + |K_y|))\,d\xi\,d\eta \leqslant \gamma < 1.$$

Under these assumptions, $\{u_n(x, y)\}$ is uniformly convergent, and the limit $u(x, y)$ coinciding with $h(x, y)$ on Γ satisfies

$$\Delta u = f(x, y, u, \partial u/\partial x, \partial u/\partial y).$$

Furthermore, it is known that the solution is unique within the region mentioned above. This method can be applied when G is small and the values of h_x and h_y are limited.

When f does not contain p and q, the following method is known. Let $\underline{\omega}(x, y)$ and $\bar{\omega}(x, y)$ both be continuous on $G \cup \Gamma$ and of class C^2 in G. Suppose that $f(x, y, u)$ satisfies a Hölder condition in $\underline{\omega}(x, y) \leqslant u \leqslant \bar{\omega}(x, y)$, and that

$$\Delta\underline{\omega} \geqslant f(x, y, \underline{\omega}(x, y)), \qquad \Delta\bar{\omega} \leqslant f(x, y, \bar{\omega}(x, y)).$$

Then, given a continuous function φ on Γ such that $\underline{\omega} \leqslant \varphi \leqslant \bar{\omega}$, there exists a unique solution u of $\Delta u = f(x, y, u)$ such that

$$\underline{\omega}(x, y) \leqslant u(x, y) \leqslant \bar{\omega}(x, y) \text{ in } G,$$

$$u(x, y) = \varphi \text{ on } \Gamma.$$

Finally, consider the equation

$$A\frac{\partial^2 u}{\partial x^2} + 2B\frac{\partial^2 u}{\partial x \partial y} + C\frac{\partial^2 u}{\partial y^2} = 0,$$

where A, B, and C are functions of x, y, u, p, q and $AC - B^2 > 0$. Under the following conditions, there exists a solution of the Dirichlet problem: A, B, C are of class C^2, and their derivatives of order 2 always satisfy Hölder conditions; G is †convex; and the boundary value φ along Γ considered as a curve in xyu-space represented by the parameter of arc length is of class $C^{3+\alpha}$ $(\alpha > 0)$. Moreover, any plane having 3 common points with this curve has slope less than a fixed number Λ. The proof of this theorem is carried out in the following way: For any function u satisfying

$$|u(x, y)| \leqslant \max|\varphi|, \qquad |u_x| \leqslant \Lambda, \qquad |u_y| \leqslant \Lambda,$$

replacing u, p, q by $u(x, y)$, $u_x(x, y)$, $u_y(x, y)$, respectively, in A, B, and C, we have the linear equation in v:

$$A[u]\frac{\partial^2 v}{\partial x^2} + 2B[u]\frac{\partial^2 v}{\partial x \partial y} + C[u]\frac{\partial^2 v}{\partial y^2} = 0.$$

We can obtain the solution v taking the boundary value φ on Γ. Thus we have a mapping $u \to v$. Applying the †fixed-point theorem in function space to this mapping, we have the desired solution $v(x, y) = u(x, y)$.

Concerning the Dirichlet problem for the second-order **semilinear** elliptic partial differential equation

$$\sum_{i,j=1}^{n} a_{ij}(x)\frac{\partial^2 u}{\partial x_i \partial x_j} = f\left(x, u, \frac{\partial u}{\partial x_1}, \ldots, \frac{\partial u}{\partial x_n}\right),$$

a work by M. Nagumo (*Osaka Math. J.*, 6 (1954)) establishes a general existence theorem.

For the general nonlinear equation

$$F(x_1, \ldots, x_n, u, p_1, \ldots, p_n, p_{11}, \ldots, p_{ij}, \ldots, p_{nn}) = 0,$$

if $\sum(\partial F/\partial p_{ij})\xi_i\xi_j > 0$, $F_u \leqslant 0$, the solution of the Dirichlet problem for this equation is unique. Even when $F_u \leqslant 0$ does not hold, the conclusion remains the same if we can reduce the equation to this case by a suitable change of variables.

Quasilinear equations in **divergence form**

$$\sum_{i=1}^{n} \frac{\partial}{\partial x_i} a_i\left(x, u, \frac{\partial u}{\partial x_1}, \ldots, \frac{\partial u}{\partial x_n}\right) = f\left(x, u, \frac{\partial u}{\partial x_1}, \ldots, \frac{\partial u}{\partial x_n}\right), \tag{4}$$

or more generally, any quasilinear elliptic equation

$$\sum_{i,j=1}^{n} a_{ij}\left(x, u, \frac{\partial u}{\partial x_1}, \ldots, \frac{\partial u}{\partial x_n}\right)\frac{\partial^2 u}{\partial x_i \partial x_j} = f\left(x, u, \frac{\partial u}{\partial x_1}, \ldots, \frac{\partial u}{\partial x_n}\right), \tag{5}$$

and even quasilinear elliptic systems have been treated in detail in several recent works [8, 9]. J. Serrin [10] treated the Dirichlet problem and established the existence and the uniqueness of solutions for some classes of equations of type (5) containing the minimal surface equation. His method is to estimate the maximum norms of u and of its first derivatives, to apply a result of O. A. Ladyzhenskaya and N. N. Ural'tseva [9] and the Schauder estimate (2), and finally to use the **Leray-Schauder fixed-point theorem** [11].

E. Relation to the Calculus of Variations

Consider the bilinear form

$$J = \int_G \left(\sum a_{ij}(x)\frac{\partial u}{\partial x_i}\frac{\partial u}{\partial x_j} + 2\sum b_i(x)\frac{\partial u}{\partial x_i}u + c(x)u^2 + 2f(x)u\right)dx,$$

where we assume $\sum a_{ij}(x)\xi_i\xi_j > 0$. Under the boundary conditions imposed on u, if there exists a function u that makes J minimum, then assuming some differentiability condition on $a_{ij}(x)$, $b_i(x)$, $c(x)$, we have the †Euler-Lagrange equation

$$\sum_j \frac{\partial}{\partial x_j}\left(\sum_i a_{ij}(x)\frac{\partial u}{\partial x_i}\right) - \left(c(x) - \sum_i \frac{\partial}{\partial x_i}b_i(x)\right)u - f(x) = 0,$$

which is a linear second-order self-adjoint elliptic equation.

B. Riemann treated the simplest case, where $a_{ij}(x) = \delta_i^j$, $b_i(x) = c(x) = 0$, i.e., the case $\Delta u = 0$. He proved, assuming the existence of the minimum of J, the existence of the solution of $\Delta u = 0$ with assigned boundary values. This result,

called **Dirichlet's principle**, was used by D. Hilbert, R. Courant, H. Weyl, O. Nikodym, and others to show the existence of solutions for linear self-adjoint elliptic equations.

If $F(x_1, \ldots, x_n, u, p_1, \ldots, p_n)$ satisfies $\Sigma(\partial^2 F/\partial p_i \partial p_j)\xi_i \xi_j > 0$ (and F has some regularity), then the function that minimizes the integral

$$J = \int_G F\left(x_1, \ldots, x_n, u, \frac{\partial u}{\partial x_1}, \ldots, \frac{\partial u}{\partial x_n}\right) dx_1, \ldots, dx_n$$

with the given boundary condition satisfies the Euler-Lagrange equation (of type (4))

$$\sum_{i,j=1}^n F_{p_i p_j} \frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{i=1}^n F_{u p_i} \frac{\partial u}{\partial x_i} + \sum_{i=1}^n F_{x_i p_i} - F_u = 0$$

($p_i = \partial u/\partial x_i$) and the boundary condition as well. This is also an elliptic partial differential equation in u. The case where F is a function of p alone (in particular, the case of †minimal surfaces) has been studied by A. Haar, T. Radó, J. Serrin, and others, particularly for $F = (1 + p_1^2 + \ldots + p_n^2)^{1/2}$ in the case of the minimal surface equation.

F. The Second and Third Boundary Value Problems

Let G be a domain in $\mathbf{R}^n$ with a smooth boundary consisting of a finite number of hypersurfaces. Also, let $B[u]$ be the boundary operator defined by

$$B[u] = a\frac{\partial u}{\partial v} + \beta u, \tag{6}$$

where $a = (\sum_{i=1}^n (\sum_{j=1}^n a_{ij} \cos(v_0 x_j))^2)^{1/2}$, v_0 is the outer normal of unit length at the point $x \in S$, and v is the **conormal** defined by

$$\cos(v x_i) = \sum_{j=1}^n a_{ij} \cos(v_0 x_j)/a, \qquad i = 1, \ldots, n.$$

The problem of finding the solution $u(x)$ of the equation $L[u] = f$ continuous on the closed domain $\bar{G}$ and satisfying $B[u] = \varphi$ on the boundary S of G is called the **second boundary value problem** (or **Neumann problem**) when $\beta \equiv 0$, and the **third boundary value problem** (or **Robin problem**) when $\beta \not\equiv 0$. In general, in boundary value problems, the condition that the solutions must satisfy at the boundary is called the **boundary condition**. We assume that the boundary S of G is expressed locally by a function with †Hölder continuous first derivatives (G is then called a **domain of class** $C^{1,\lambda}$). Assume that G is such a domain, $c \leqslant 0$, $\beta \geqslant 0$, and at least one of c and β is not identically 0. Then the second and third boundary value problems admit one and only one solution. When $c \equiv 0$ and $\beta \equiv 0$, the solutions of the second boundary value problem are determined uniquely up to additive constants. Furthermore, let M be the †adjoint operator of L, and let

$$B'[v] = a\frac{\partial v}{\partial v} + (\beta - b)v,$$

where

$$b = \sum_{i=1}^n \cos(v_0 x_i)\left[b_i - \sum_{j=1}^n \frac{\partial a_{ij}}{\partial x_j}\right].$$

Then if the boundary S of G is of class C^1 and f, φ are continuous, in order that there exist at least one solution u of the second or third boundary value problem relative to $L[u] = f$, it is necessary and sufficient that

$$\int_G fv\, dx - \int_S \varphi v\, dS = 0,$$

where v is any solution of $M[v] = 0$ with the boundary condition $B'[v] = 0$. Here the necessity is easily derived from Green's formula. G. Giraud used the notion of fundamental solution to reduce the second and third boundary value problems relative to $L[u] = f$ to a problem of integral equations, under the assumptions that G is a domain of class C^1, the coefficients of L and f satisfy Hölder conditions, and φ and β are continuous [3; also 12].

G. Method of Orthogonal Projection

The theory of †Hilbert spaces is applicable to the boundary value problems in Section F. In general, let $H^m(G)$ be the space of functions in $L_2(G)$ whose partial derivatives in the sense of †distributions up to order m belong to $L_2(G)$. For elements f and g in $H^m(G)$, we define the following inner product:

$$(f, g)_m = \sum_{|\alpha| \leqslant m} \int_G D^\alpha f(x) \overline{D^\alpha g(x)}\, dx,$$

where

$$D^\alpha = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \ldots \partial x_n^{\alpha_n}}, \qquad |\alpha| = \alpha_1 + \ldots + \alpha_n$$

(→ 168 Function Spaces). With respect to this inner product, $H^m(G)$ is a Hilbert space. When u satisfies $L[u] = f$ ($f \in L_2(G)$) and $\varphi(x)$ is an arbitrary element of $H^1(G)$, Green's formula yields

$$-\sum_{i,j=1}^n \left(a_{ij}\frac{\partial u}{\partial x_i}, \frac{\partial \varphi}{\partial x_j}\right) + \sum_{i=1}^n \left(b_i'(x)\frac{\partial u}{\partial x_i}, \varphi\right)$$

$$+ (c(x)u, \varphi) + \int_S a\frac{\partial u}{\partial v} \cdot \bar{\varphi}\, dS = (f, \varphi),$$

where $b_i'(x) = b_i(x) = -\sum_j (\partial a_{ij}/\partial x_j)$. Taking account of the boundary condition on u:

$a\,\partial u/\partial\nu+\beta u=0$, we get

$$\sum_{i,j}\left(a_{ij}\frac{\partial u}{\partial x_i},\frac{\partial\varphi}{\partial x_j}\right)-\sum_i\left(b_i'\frac{\partial u}{\partial x_i},\varphi\right)-(cu,\varphi)$$
$$+\int_S\beta u\bar{\varphi}\,dS=-(f,\varphi).$$

Thus the problem is reduced to finding $u(x)\in H^1(G)$ satisfying this equation for all $\varphi\in H^1(G)$. This equation can be regarded as an equation in $H^1(G)$. If necessary, by replacing $c(x)$ by $c(x)-t$ for a large t, we can show that for any $f(x)\in L_2(G)$, there exists a unique solution $u(x)\in H^1(G)$ [12, 13]. Now, for a solution of this functional equation, if we take $\varphi(x)\in\mathcal{D}(G)$, we have $(u,L^*[\varphi])=(f,\varphi)$, where L^* is the †adjoint operator of L. This means that $u(x)$ is a solution of $L[u]=f$ in the sense of distributions, and we call such a $u(x)$ a **weak solution**. Such a treatment may be called the **method of orthogonal projection**, following Weyl. In this case, it can be shown that if we assume smoothness of the coefficients, the boundary S, and β, then the solution $u(x)\in H^1(G)$ belongs to $H^{s+2}(G)$ when $f(x)\in H^s(G)$ $(s=0,1,\ldots)$. Thus, if we apply Green's formula, we can see that u satisfies the boundary condition $a\,\partial u/\partial\nu+\beta u=0$. In particular, when $s>n/2$, we see that $u(x)\in C^2(\bar{G})$ by †Sobolev's theorem [12]. In other words, $u(x)$ is a **genuine solution**.

Next, we introduce the complex parameter λ and consider the boundary value problem $(L+\lambda I)[u]=f$, $f\in L_2(G)$, $a\,\partial u/\partial\nu+\beta u=0$. If t is large, $(L-tI)$ is a one-to-one mapping from the domain $\mathcal{D}(L)=\{u\in H^2(G)\mid a\,\partial u/\partial\nu+\beta u=0\}$ onto $L_2(G)$. Thus, denoting its inverse, which acts on the equation from the left, by G_t, we have

$$(I+(\lambda+t)G_t)[u]=G_tf.$$

Conversely, since the solution $u(x)$ contained in $L_2(G)$ (hence also contained in $\mathcal{D}(L)$) satisfies the equation and the boundary condition, the problem is reduced to the displayed equation in $L_2(G)$ considered above. Now, since G is bounded and G_t is a continuous mapping from $L_2(G)$ into $H^2(G)$, **Rellich's theorem** yields that G_t is a †compact operator when it is regarded as an operator in $L_2(G)$. So we can apply the †Riesz-Schauder theorem (→ 189 Green's Operator).

H. Elliptic Equations of Higher Order

The differential operator of order m:

$$L=\sum_{|\alpha|\leqslant m}a_\alpha(x)D^\alpha,\qquad D^\alpha=\frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1}\ldots\partial x_n^{\alpha_n}},$$
$$|\alpha|=\alpha_1+\ldots+\alpha_n,$$

is called an **elliptic operator** if $\sum_{|\alpha|=m}a_\alpha(x)\xi^\alpha\neq0$ $(\xi\neq0)$. In particular, if

$$\operatorname{Re}\sum_{|\alpha|=m}a_\alpha(x)\xi^\alpha\geqslant c|\xi|^m,\qquad c>0,$$

L is called a **strongly elliptic operator**. In this case, m is even. L. Gårding studied the Dirichlet problem for strongly elliptic operators [14]. If we put $m=2b$, the boundary value condition is stated as $\partial^ju/\partial\nu^j=f_j(x)$ $(j=0,1,\ldots,b-1)$, where ν is the normal of unit length at the boundary. Using the notion of function space, this boundary condition means that the solutions belong to the closure $\mathring{H}^b(G)$ of $\mathcal{D}(G)$ in $H^b(G)$ (→ 168 Function Spaces). In this treatment, **Gårding's inequality**

$$(-1)^b\operatorname{Re}(L[u],u)\geqslant\delta\|u\|_b^2-c\|u\|^2,\qquad u\in\mathring{H}^b(G),\quad(7)$$

where δ and c are positive constants, plays an important role.

In general, for an elliptic operator L defined in an open set G, if $u(x)$ satisfies $L[u]=f(x)$ and $f(x)$ belongs to H^s on any compact set in G, then $u(x)$ belongs to H^{m+s} on every compact set in G (**Friedrich's theorem** [15]).

General boundary value problems for elliptic equations of higher order have been considered by S. Agmon, A. Douglis, and L. Nirenberg [16], M. Schechter [17], and others. These problems are formulated as follows:

$$L[u]=f(x),\qquad B_j(x,D)u(x)=\varphi_j(x),$$
$$x\in S,\qquad j=1,2,\ldots,b(=m/2),\quad(8)$$

where the $B_j(x,D)$ are differential operators at the boundary and f and $\{\varphi_j\}$ are given functions. Under certain algebraic conditions (**Shapiro-Lopatinskiĭ conditions**) on $(L,\{B_j\})$, the problems are treated also in $H^m(G)$; hence the L^2 **a priori estimates** play a fundamental role: If $u\in H^m(G)$,

$$\|u\|_m\leqslant K\Big(\|Lu\|+\sum_{j=1}^b\|B_ju\|_{m-m_j-(1/2),S}+\|u\|\Big),\quad(9)$$

where K is a constant determined by (L,B_j,G), $\|\cdot\|_{k,S}$ is the norm in $H^k(S)$, and m_j are the orders of B_j (compare (9) with (2)). Under these estimates the boundary value problem is said to be **coercive**. In applications, the theory of interpolation of function spaces are also used [18] (→ 168 Function Spaces). Variational general boundary value problems have been treated by D. Fujiwara and N. Shimakura (*J. Math. Pures Appl.*, 49 (1970)) and others. For systems of such equations, there are works by F. E. Browder (*Ann. math. studies* 33, Princeton Univ. Press, 1954, 15–51) and others.

I. Analyticity of Solutions

In a linear elliptic equation $Lu=f$, suppose that all the coefficients and f are of class C^∞ (resp. real analytic) in an open set G and that u is a distribution solution in G. Then u is also of class C^∞ (resp. real analytic) in G [19]. Hence the linear elliptic operators are **hypoelliptic** (resp. **analytically hypoelliptic**) (→ 112 Differential Operators). In particular, †harmonic functions (i.e., solutions of $\Delta u=0$) are (real) analytic in the domain of existence, whatever the boundary values may be.

Hilbert conjectured that when $F(x,y,u,p,q,r,s,t)$ $(p=p_1, q=p_2, r=p_{11}, s=p_{12}, t=p_{22})$ is analytic in the arguments, then any solution u of the elliptic equation $F=0$ is analytic on the domain of existence (1900, Hilbert's 19th problem; → 196 Hilbert). This conjecture was proved by S. Bernshteĭn, Radó, and others, and then H. Lewy proposed a method of extending this equation to a complex domain so that it can be regarded as a †hyperbolic equation (*Math. Ann.*, 101 (1929)). This result was further extended by I. G. Petrovskiĭ to a general system of nonlinear differential equations of elliptic type (*Mat. Sb.*, 5 (47), (1939)).

J. The Unique Continuation Theorem

Since all the solutions of Laplace's equation $\Delta u=0$ are analytic, it follows that if $u(x)$ vanishes on an open set in a domain, then $u(x)$ vanishes identically in this domain. This **unique continuation theorem** can be extended to linear elliptic partial differential equations with analytic coefficients in view of the analyticity of solutions. This fact is also proved by applying †Holmgren's uniqueness theorem (→ 321 Partial Differential Equations (Initial Value Problems)). The unique continuation theorem, first established by T. Carleman for second-order elliptic partial differential equations $L[u]=0$ with C^1-coefficients in the case of two independent variables, was extended to second-order linear elliptic equations with C^2-coefficients in the case of any number of independent variables by C. Müller, E. Heinz, and finally by N. Aronszajn [20]. This research was extended by A. P. Calderón [21] and others in the direction of establishing the uniqueness of the Cauchy problem. However, it is to be remarked that even if we assume that the coefficients are of class C^∞, we cannot affirm the unique continuation property for general elliptic equations. A counterexample was given by A. Pliś (*Comm. Pure Appl. Math.*, 14 (1961)). See also the work of K. Watanabe (*Tohoku Math J.*, 23 (1971)).

K. Elliptic Pseudodifferential Operators and the Index

A pseudodifferential operator $P(x,D)$ with symbol $p(x,\xi)\in S^m_{0,1}$ (→ 345 Pseudodifferential Operators) is said to be **elliptic**, provided there exists a positive constant c such that $|p(x,\xi)|\geqslant c(1+|\xi|)^m$ for all $x\in\mathbf{R}^n$ and $|\xi|\geqslant c^{-1}$. The notion of ellipticity can be extended to operators on a manifold. The theory of elliptic pseudodifferential operators has been widely applied to the study of elliptic differential equations, and is particularly useful in the calculation of the †index of elliptic operators. B. R. Vaĭnberg and V. V. Grushin [22] calculated the index i of the †coercive boundary value problem for an elliptic operator by showing that i is equal to the index of some elliptic pseudodifferential operator on the boundary.

Example [23]: Given a real vector field (v_1, v_2) on the unit circle $x_1^2+x_2^2=1$, suppose the vector $(v_1(x), v_2(x))$ rotates l times around the origin as the point $x=(x_1,x_2)$ moves once around the unit circle in the positive direction. Then the index of the boundary value problem

$$\left(\frac{\partial^2}{\partial x_1^2}+\frac{\partial^2}{\partial x_2^2}\right)u(x)=f(x), \qquad x_1^2+x_2^2<1,$$

$$v_1(x)\frac{\partial u}{\partial x_1}+v_2(x)\frac{\partial u}{\partial x_2}=g(x), \qquad x_1^2+x_2^2=1,$$

is equal to $2-2l$.

M. F. Atiyah and I. M. Singer determined the index of a general elliptic operator on a manifold in terms of certain topological invariants of the manifold (→ 237 K-Theory H). The index of noncoercive boundary value problems has also been studied by Vaĭnberg and Grushin, R. Seeley (*Topics in pseudo-differential operators*, C.I.M.E. 1968, 335–375), and others.

L. The Giorgi-Nash-Moser Result

Let us state the following result (J. Moser [24]): Let L be of the form

$$Lu=\sum_{i,j=1}^{n}\frac{\partial}{\partial x_j}\left\{a_{ij}(x)\frac{\partial u}{\partial x_i}\right\},$$

where $a_{ij}=a_{ji}$ are real-valued, of class $L^\infty(G)$, and such that the ellipticity condition (3) holds at almost everywhere in G with some $\lambda\geqslant 1$. Also, let G' be any subdomain of G whose distance from ∂G is not smaller than $\delta>0$. Then, for any weak solution $u\in H^1(G)$ of the equation $Lu=0$ and for any two points x and y in G', we have the inequality

$$|u(x)-u(y)|\leqslant A|x-y|^\alpha\|u\|_{L^2(G)},$$

where A and α ($A>0, 0<\alpha\leqslant 1$) depend only on (n, λ, δ) and are independent of the particular choice of (L, G, G', u). Moser proved that the above inequality is a corollary to a Harnack-type inequality (→ 327 Partial Differential Equations of Parabolic Type G).

M. Asymptotic Distribution of Eigenvalues

Let L be an elliptic operator on $\bar{G}$ of order m with smooth coefficients realized as a self-adjoint operator in $L^2(G)$ under a nice boundary condition, where G is a bounded domain in $\mathbf{R}^n$ with smooth boundary. Let $N(T)(T>0)$ be the number of eigenvalues of L smaller than T. Then, it holds that

$N(T)=CT^{n/m}+$an error term, as $T\to+\infty$,

$$C=(2\pi)^{-n}\int_G dx\int_{S^{n-1}} a(x,\xi)^{-n/m}\,dS_\xi, \qquad (10)$$

where $a(x,\xi)$ is the †principal symbol of L and S^{n-1} is the unit sphere on $\mathbf{R}^n$. C is independent of the boundary condition and, if L is of constant coefficients, of the shape of G.

Formula (10) was at first established by H. Weyl [25] for the case of $L=$ Laplacian, and hence it is often called **Weyl's formula**. Weyl's method is based on the minimax principle [26]. T. Carleman (*Ber. Math.-Phys. Klasse der Sächs. Akad. Wiss. Leipzig*, 88 (1936)) studied the behavior of the trace of the Green's function of zI-L as $|z|\to\infty$ in the complex plane (→ [27]). S. Minakshisundaram (*Canad. J. Math.*, 1 (1947)) discussed this formula in connection with the heat equation; see also S. Mizohata and R. Arima (*J. Math. Kyoto Univ.*, 4 (1964)). L. Hörmander (*Acta Math.*, 121 (1968)) treated the case of compact manifolds without boundary and obtained the best possible error estimate. H. P. McKean and I. M. Singer (*J. Differential Geometry*, 1 (1967)) treated the case of manifolds and discussed the geometric meaning of this formula. In general, $N(T)$ is no more than $O(T^{n/m})$ if L is of degenerate elliptic type (C. Nordin, *Ark. Mat.*, 10 (1972)).

N. Equations of Degenerate Elliptic Type

An operator L of the form (1) is said to be degenerate at $x^0\in\bar{G}$ in the direction $\xi\in\mathbf{R}^n$ if ξ is a null vector of the matrix $(a_{ij}(x^0))$. L is said to be of degenerate elliptic type if $(a_{ij}(x))$ is nonnegative definite at any $x\in\bar{G}$ and if L is degenerate at some point of $\bar{G}$ in some direction.

Suppose that the coefficients of L and the boundary Γ of G are smooth enough. At $x\in\Gamma$, denote by $\nu(x)=(\nu_1(x),\ldots,\nu_n(x))$ the unit outer normal vector to G. Let Σ_3 be the set of $x\in\Gamma$ at which L is not degenerate in the normal direction. Also, let Σ_1, Σ_2, and Σ_0 be the sets of $x\in\Gamma\setminus\Sigma_3$ at which $b(x)>0$, <0, and $=0$, respectively, where $b(x)$ is defined by

$$b(x)=-\sum_{i=1}^n \nu_i(x)\left\{b_i(x)-\sum_{j=1}^n\frac{\partial a_{ij}(x)}{\partial x_j}\right\}. \qquad (11)$$

Then the Dirichlet problem for equation (1) is to find a function $u(x)$ defined on $G\cup\Sigma_2\cup\Sigma_3$ satisfying

$$L[u]=f \text{ in } G, \qquad (1)$$

$$u=g \text{ on } \Sigma_2\cup\Sigma_3, \qquad (12)$$

where f and g are given functions.

Let $1<p<\infty$. We set $q=p/(p-1)$. We also put

$$c^*(x)=c(x)-\sum_{i=1}^n\frac{\partial b_i(x)}{\partial x_i}+\sum_{i,j=1}^n\frac{\partial^2 a_{ij}(x)}{\partial x_i\partial x_j}. \qquad (13)$$

We have the following existence theorem [28]: If (i) either $c<0$ on $\bar{G}$ or $c^*<0$ on $\bar{G}$ and if (ii) $pc+qc^*<0$ on $\bar{G}$, the Dirichlet problem (1) and (12) (with $g=0$) has a weak solution $u\in L^p(G)$ for any $f\in L^p(G)$. The regularity of solutions is also discussed in [28]. The value of $b(x)$ is closely related to the regularity near the point $x\in\Gamma$ if L is degenerate at x in the normal direction (→ also M. S. Baouendi and C. Goulaouic, *Arch. Rational Mech. Anal.*, 34 (1969)).

Degenerate elliptic equations of type (1) have also been investigated from the probabilistic viewpoint (→ 115 Diffusion Processes). The general boundary value problems for degenerate elliptic equations of higher order have been treated by M. I. Vishik and V. V. Grushin [29], N. Shimakura (*J. Math. Kyoto Univ.*, 9 (1969)), P. Bolley and J. Camus (*Ann. Scuola Norm. Sup. Pisa*, IV-1 (1974)), and others.

References

[1] J. Hadamard, Lectures on Cauchy's problem in linear partial differential equations, Yale Univ. Press and Oxford Univ. Press, 1923.
[2] E. E. Levi, Sulle equazioni lineari totalmente elliptichee alle derivate parziali, Rend. Circ. Mat. Palermo, 24 (1907), 275–317.
[3] C. Miranda, Partial differential equations of elliptic type, Springer, 1970.
[4] M. H. Protter and H. F. Weinberger, Maximum principles in differential equations, Prentice-Hall, 1967.
[5] J. Schauder, Über lineare elliptische Differentialgleichungen zweiter Ordnung, Math. Z., 38 (1934), 257–282.
[6] J. Schauder, Numerische Abschätzungen in

elliptischen linearen Differentialgleichungen, Studia Math., 5 (1934), 34–42.
[7] E. Picard, Leçon sur quelques problèmes aux limites de la théorie des équations différentielles, Gauthier-Villars, 1930.
[8] C. B. Morrey, Multiple integrals in the calculus of variations, Springer, 1966.
[9] O. A. Ladyzhenskaya and N. N. Ural'tseva, Linear and quasilinear elliptic equations, Academic Press, 1968. (Original in Russian, 1964.)
[10] J. Serrin, The problem of Dirichlet for quasilinear elliptic equations, Philos. Trans. Roy. Soc. London, (A) 264 (1969), 413–493.
[11] J. Leray and J. Schauder, Topologie et équations fonctionnelles, Ann. Ecole Norm. Sup., 51 (1934), 45–78.
[12] S. Mizohata, The theory of partial differential equations, Cambridge Univ. Press, 1973. (Original in Japanese, 1965.)
[13] L. Nirenberg, Remarks on strongly elliptic partial differential equations, Comm. Pure Appl. Math., 8 (1955), 649–675.
[14] L. Gårding, Dirichlet's problem for linear elliptic partial differential equations, Math. Scand., 1 (1953), 55–72.
[15] K. O. Friedrichs, On the differentiability of solutions of linear elliptic equations, Comm. Pure Appl. Math., 6 (1953), 299–325.
[16] S. Agmon, A. Douglis, and L. Nirenberg, Estimates near the boundary for solutions of elliptic partial differential equations satisfying general boundary conditions I, II, Comm. Pure Appl. Math., 12 (1959), 623–727; 17 (1964), 35–92.
[17] M. Schechter, General boundary value problems for elliptic equations, Comm. Pure Appl. Math., 12 (1959), 457–486.
[18] J. L. Lions and E. Magenes, Problèmes aux limites non homogènes et applications I–III, Dunod, 1968.
[19] F. John, Plane waves and spherical means applied to partial differential equations, Interscience, 1955.
[20] N. Aronszajn, A unique continuation theorem for solutions of elliptic equations or inequalities of second order, J. Math. Pure Appl., 36 (1957), 235–249.
[21] A. P. Calderón, Uniqueness in the Cauchy problem for partial differential equations, Amer. J. Math., 80 (1958), 16–36.
[22] B. R. Vaĭnberg and V. V. Grushin, Uniformly nonelliptic problems I, II, Math. USSR-Sb., 1 (1967), 543–568; 2 (1967), 111–133. (Original in Russian, 1967.)
[23] I. N. Vekua, System von Differentialgleichungen erster Ordnung vom elliptischen Typus und Randwertaufgaben, Berlin, 1956.
[24] J. Moser, On Harnack's theorem for elliptic differential equations, Comm. Pure Appl. Math., 14 (1961), 577–591.
[25] H. Weyl, Das asymptotische Verteilungsgesetze der Eigenverte linearer partieller Differentialgleichungen, Math. Ann., 71 (1912), 441–479.
[26] R. Courant and D. Hilbert, Methods of mathematical physics I, II, Interscience, 1962.
[27] S. Agmon, Lectures on elliptic boundary value problems, Van Nostrand, 1965.
[28] O. A. Oleĭnik and E. V. Radkevich, Second-order equations with non-negative characteristic form, Plenum, 1973. (Original in Russian, 1971.)
[29] M. I. Vishik and V. V. Grushin, Boundary value problems for elliptic equations degenerate on the boundary of a domain, Math. USSR-Sb., 9 (1969), 423–454. (Original in Russian, 1969.)

324 (XIII.22) Partial Differential Equations of First Order

A. Quasilinear Partial Differential Equations and Their Characteristic Curves

Suppose that we are given a †quasilinear partial differential equation

$$\sum_{i=1}^{n} P_i(x,u)\frac{\partial u}{\partial x_i}=Q(x,u), \qquad x=(x_1,\ldots,x_n). \qquad (1)$$

A curve defined by a solution $x_i=x_i(t)$, $u=u(t)$ of the system of ordinary differential equations

$$\frac{dx_i}{dt}=P_i(x,u), \qquad i=1,\ldots,n; \qquad \frac{du}{dt}=Q(x,u)$$

is called a **characteristic curve** of (1) (→ 320 Partial Differential Equations; 322 Partial Differential Equations (Methods of Integration)). A necessary and sufficient condition for $u=u(x)$ to be a solution of (1) is that the characteristic curve passing through any point on the hypersurface $u=u(x)$ (in the $(n+1)$-dimensional xu-space) always be contained in this hypersurface. For example, the characteristic curve of $\sum_{i=1}^{n} x_i\partial u/\partial x_i=ku$ is $x_i=x_i^0e^t$, $u=u^0e^{kt}$ (a solution of $x_i'=x_i, u'=ku$). Therefore the solution $u=u(x)$ is a function such that $u(\lambda x_1,\ldots,\lambda x_n)=\lambda^k u(x_1,\ldots,x_n)$ $(\lambda=e^t>0)$, i.e., a homogeneous function of degree k.

B. Nonlinear Partial Differential Equations and Their Characteristic Strips

We denote the value of $\partial u/\partial x_i$ by p_i and define the **surface element** (or **hypersurface element**)

by the $(2n+1)$-dimensional vector $(x,u,p)=(x_1,\ldots,x_n,u,p_1,\ldots,p_n)$. Consider the partial differential equation

$$F(x_1,\ldots,x_n,u,p_1,\ldots,p_n)=0, \qquad p_i=\partial u/\partial x_i. \tag{2}$$

A set $(x(t),u(t),p(t))$ of surface elements depending on a parameter t and satisfying the system of ordinary differential equations

$$\frac{dx_i}{dt}=F_{p_i}, \qquad \frac{du}{dt}=\sum_{i=1}^{n}p_iF_{p_i}, \qquad \frac{dp_i}{dt}=-(F_{x_i}+p_iF_u)$$

is called a **characteristic strip** of equation (2), and the curve $x=x(t)$, $u=u(t)$ is called a **characteristic curve** of (2). For quasilinear equations, this definition of characteristic curve coincides with the one mentioned in Section A. Furthermore, an r-dimensional †differentiable manifold consisting of surface elements satisfying the relation

$$du-\sum_{i=1}^{n}p_i\,dx_i=0$$

is called an r-dimensional **union of surface elements**. A solution of the partial differential equation (2) is, in general, formed by the set of all characteristic strips possessing, as initial values, surface elements belonging to an $(n-1)$-dimensional union of surface elements satisfying $F(x,u,p)=0$.

An example of a nonlinear partial differential equation of first order is $pq-z=0$ (where $x_1=x$, $x_2=y$, $u=z$, $p_1=p$, $p_2=q$). The equations of the characteristic strip are $x'=q$, $y'=p$, $z'=2pq$, $p'=p$, $q'=q$, and therefore the characteristic strip is given by $y=y_0+(p_0/q_0)x$, $z=z_0+2p_0x+(p_0/q_0)x^2$, $p=p_0+(p_0/q_0)x$, $q=q_0+x$ (if we take x as an independent variable and impose an initial condition $y=y_0$, $z=z_0$, $p=p_0$, $q=q_0$ at $x=0$). Putting $z_0=W(y_0)$ (an arbitrary function) for $x=0$, we have, furthermore, $y=y_0+W(y_0)/(W'(y_0))^2$, $z=W(y_0)+2W(y_0)/W'(y_0)+W(y_0)x^2/(W'(y_0))^2$. The elimination of y_0 from these expressions yields a general solution $z=z(x,y)$. In this case, a †complete solution is $4az=(x+ay+b)^2$ (where a, b are constants), and a †singular solution is $z=0$.

C. Complete Systems of Linear Partial Differential Equations

For functions $P_i(x)$ $(i=1,\ldots,x=(x_1,\ldots,x_n))$ of class C^∞, define a differential operator X by

$$X=\sum_{\nu=1}^{n}P_\nu(x)\frac{\partial}{\partial x_\nu}.$$

We call k differential operators $X_i=\sum_{\nu=1}^{n}P_\nu^i(x)\partial/\partial x_\nu$ $(i=1,\ldots,k)$ mutually **independent** when the rank of the matrix (P_ν^i) is equal to k. If a system of k independent linear partial differential equations involving one unknown function $f(x)$,

$$X_1f=0,\ldots,X_kf=0, \tag{3}$$

has the maximum number $n-k$ of independent †integrals, then the system (3) is called a **complete system**. A necessary and sufficient condition for the system (3) to be a complete system is that there exist k^3 functions $\lambda_{ij}^\nu(x)$ of class C^∞ such that

$$[X_j,X_i]\equiv X_jX_i-X_iX_j=\sum_{\nu=1}^{k}\lambda_{ij}^\nu(x)X_\nu,$$

that is,

$$\sum_{h=1}^{n}\left(P_h^j\frac{\partial P_\mu^i}{\partial x_h}-P_h^i\frac{\partial P_\mu^j}{\partial x_h}\right)=\sum_{\nu=1}^{k}\lambda_{ij}^\nu P_\mu^\nu.$$

Here, $[X_j,X_i]$ is a differential operator of first order, called the **commutator** of the differential operators X_i, X_j or the **Poisson bracket**.

D. Involutory Systems

For two functions $F(x,u,p)$, $G(x,u,p)$ of x, u, p of class C^∞, we define the **Lagrange bracket** $[F,G]$ by

$$[F,G]=\sum_{\nu=1}^{n}\left(\frac{\partial F}{\partial p_\nu}\left(\frac{\partial G}{\partial x_\nu}+p_\nu\frac{\partial G}{\partial u}\right)-\frac{\partial G}{\partial p_\nu}\left(\frac{\partial F}{\partial x_\nu}+p_\nu\frac{\partial F}{\partial u}\right)\right).$$

If F, G do not contain u and are homogeneous linear forms with respect to p, then F and G are differential operators $F=X_1u$ and $G=X_2u$ with respect to u (for $p_\nu=\partial u/\partial x_\nu$), and we see that $[F,G]=[X_1,X_2]_u$. This bracket has the following properties:

$$[F,G]=-[G,F],$$

$$[F,\varphi(G_1,\ldots,G_k)]=\sum_{i=1}^{k}\frac{\partial\varphi}{\partial G_i}[F,G_i],$$

$$[[F,G],H]+[[G,H],F]+[[H,F],G]$$
$$=\frac{\partial F}{\partial u}[G,H]+\frac{\partial G}{\partial u}[H,F]+\frac{\partial H}{\partial u}[F,G].$$

When F, G are functions of x and p only, we usually use the notation (F,G) and call it also the **Poisson bracket**. In this case, the right-hand side of the third relation vanishes.

Consider k partial differential equations involving one unknown function $u(x_1,\ldots,x_n)$,

$$F_i(x,u,p)=0, \quad i=1,\ldots,k; \qquad p_\nu=\frac{\partial u}{\partial x_\nu}. \tag{4}$$

If a common solution $u(x)$ of these equations exists, it is also a solution of $[F_i,F_j]=0$ $(i,j=$

$1, \ldots, k$). Therefore, from the equations thus obtained, we take independent equations and add them to the original system. If we then have more than $n+1$ equations, the original system has no solution. Otherwise, we obtain a system $F_j=0$ ($j=1, \ldots, r$) for which the F_j are independent (i.e., the rank of $(\partial F_i/\partial p_\nu)$ is equal to r), and all $[F_i, F_j]=0$ can be derived from $F_j=0$. A system (4) such that $[F_i, F_j]\equiv 0$ for all i, j is called an **involutive** (or **involutory**) **system.** We always treat a system by extending it to an involutory system. When $k=1$, we regard the equation itself as an involutory system.

When the equations (4) are mutually independent, a necessary and sufficient condition for them to have in common a solution with $n-k$ degrees of freedom (a solution that coincides with an arbitrary function on an adequate manifold of dimension $n-k$) is that the system (4) be a system of equations involving unknowns p and equivalent to an involutory system.

An involutory system (4) can be extended to an involutory system consisting of n independent equations by adding $n-k$ suitable equations

$$F_{k+1}(x, u, p)=a_{k+1}, \ldots, F_n(x, u, p)=a_n. \qquad (4')$$

That is, we can find successively $f=F_l$ ($l=k+1, \ldots, n$) such that F_l satisfies the system of equations

$$[F_i, f]=0, \qquad i=1, \ldots, l-1,$$

which is a complete system of linear partial differential equations for f, and the F_i ($i=1, \ldots, n$) are mutually independent. Then, if we find that p_ν as functions of (x, u, a) ($a=(a_{k+1}, \ldots, a_n)$) from (4) and (4'), the system of †total differential equations

$$\frac{\partial u}{\partial x_\nu}=p_\nu(x, u, a), \quad \nu=1, \ldots, n,$$

is †completely integrable, and we can find u as a solution containing essentially $n-k+1$ parameters $c, a_{k+1}, \ldots, a_n$, that is, a complete solution of (4). Moreover, if we have an involutory system of $n+1$ independent equations $F_1=0, \ldots, F_k=0$, $F_{k+1}=a_{k+1}, \ldots, F_{n+1}=a_{n+1}$, then we find a complete solution by eliminating $p_1, \ldots, p_n$ between the equations. This method of integrating an involutory system is called **Jacobi's second method of integration**.

E. Relation to the Calculus of Variations

Consider a partial differential equation of first order $F(x_1, \ldots, x_n, u, p_1, \ldots, p_n)=0$. If a solution $u(x)$ of this equation is given as an implicit function by $\varphi(x_1, \ldots, x_n, u)=0$, we have

$$F\left(x_1, \ldots, x_n, u, -\frac{\partial\varphi/\partial x_1}{\partial\varphi/\partial u}, \ldots, -\frac{\partial\varphi/\partial x_n}{\partial\varphi/\partial u}\right)=0.$$

This gives formally a partial differential equation with independent variables $u, x_1, \ldots, x_n$ and a dependent variable φ. It does not contain φ explicitly. That is, this equation has the form

$$F(x_1, \ldots, x_{n+1}, \partial\varphi/\partial x_1, \ldots, \partial\varphi/\partial x_{n+1})=0.$$

Then, by finding a partial derivative, say $\partial\varphi/\partial x_{n+1}$, as a function of the remaining ones from the displayed equation, we get a partial differential equation of the form

$$\partial\varphi/\partial t+H(t, x, \partial\varphi/\partial x)=0,$$

which is called the **normal form** of the partial differential equation of first order. Setting $p_i=\partial\varphi/\partial x_i$, the equations of the characteristic curve of this equation are

$$\frac{dx_i}{dt}=H_{p_i}(t, x, p), \qquad \frac{dp_i}{dt}=-H_{x_i}(t, x, p),$$
$$i=1, \ldots, n,$$

which are called **Hamilton's differential equations.**

Now, consider the †Euler-Lagrange differential equations

$$dF_{x_i'}/dt-F_{x_i}=0, \quad i=1, \ldots, n,$$

for the integral

$$J=\int_{t_0}^{t_1} F(t, x, x')\,dt, \qquad x'=\frac{dx}{dt}.$$

Under the assumption that $\det(F_{x_i' x_j'})\neq 0$, we put $F_{x_i'}=p_i$, and solve these relations with respect to x_i' in the form, say, $x_i'=\varphi_i(t, x, p)$. Furthermore, if we put

$$\left(\sum_{\nu=1}^{n} x_\nu' F_{x_\nu'}-F\right)_{x'=\varphi(t,x,p)}=H(t, x, p),$$

then the Euler-Lagrange equations are equivalent to Hamilton's differential equations

$$dx_i/dt=H_{p_i}, \qquad dp_i/dt=-H_{x_i}, \qquad i=1, \ldots, n,$$

since $F(t, x, x')=\sum_{i=1}^{n} p_i H_{p_i}-H$.

A curve represented by a solution of the Euler-Lagrange equations is called a **stationary curve.** Now consider a family of stationary curves in a domain G of the $(n+1)$-dimensional tx-space such that passing through every point of G there is one and only one curve in this family, and suppose that the family is †transversal to an r-dimensional manifold $\mathfrak{A}$ ($r\leqslant n$) (that is, $F\delta t-\sum F_{x_i'}\delta x_i=0$ for the differentials $\delta t, \delta x_i$ along $\mathfrak{A}$; in particular, if $\mathfrak{A}$ consists of only one point ($r=0$), a stationary curve passing through this point is transversal to $\mathfrak{A}$). In

this case, if we denote by $V(t,x)$ the value of the integral J along the stationary curve from $\mathfrak{A}$ to any point (t,x) of G, then the equation

$$\partial V/\partial t+H(t,x,\partial V/\partial x)=0$$

holds. This equation is called the **Hamilton-Jacobi differential equation** or the **canonical** or **eikonal equation**. Conversely, a solution of this equation is equal to the value of the integral J for a family of stationary curves transversal to an adequate $\mathfrak{A}$.

F. The Monge Differential Equation

Consider the partial differential equation (2). By eliminating p and t between

$$\frac{dx_i}{dt}=F_{p_i}, \quad \frac{du}{dt}=\sum_{i=1}^{n} p_i F_{p_i}, \quad \text{and} \quad F(x,u,p)=0$$

(for example, by eliminating p between $dx_i/dx_1 = F_{p_i}/F_{p_1}$ and $F=0$ when $F_{p_1}\neq 0$), we obtain

$$M(x_1,\ldots,x_n,u,\partial x_2/\partial x_1,\ldots,\partial x_n/\partial x_1)=0.$$

This equation is called the **Monge differential equation**, and the curve represented by its solution is called an **integral curve** of the equation. In the $(n+1)$-dimensional tx-space, a curve that is an envelope of a 1-parameter family of characteristic curves of the partial differential equation (2) is a solution of the Monge equation. A characteristic curve is also an integral curve. When $n=2$, an integral curve that is not a characteristic curve is a †line of regression of the surface generated by the family of characteristic curves tangent to the integral curve under consideration (which is an integral surface of $F(x,u,p)=0$). If F is linear in p_i, i.e., quasilinear, all integral curves coincide with characteristic curves.

References

[1] R. Courant and D. Hilbert, Methods of mathematical physics II, Interscience, 1962.
[2] I. G. Petrovskiĭ, Lectures on partial differential equations, Interscience, 1954. (Original in Russian, 1953.)
[3] E. Goursat, Cours d'analyse mathématique, Gauthier-Villars, II, second edition, 1911; III, fourth edition, 1927.
[4] F. John, Partial differential equations, Springer, 1971.
[5] C. Carathéodory, Variationsrechnung und Partielle Differentialgleichungen Erste Ordnung, Teubner, 1935; English translation, Calculus of variations and partial differential equations of the first order, Holden-Day, 1967.

325 (XIII.25) Partial Differential Equations of Hyperbolic Type

A. Second-Order Linear Hyperbolic Equations

A †linear partial differential equation in $n+1$ variables t, $x=(x_1,\ldots,x_n)$ of the second order,

$$L[u]=\frac{\partial^2 u}{\partial t^2}-\sum_{i=1}^{n} a_{0i}\frac{\partial^2 u}{\partial t\partial x_i}-\sum_{i,j=1}^{n} a_{ij}\frac{\partial^2 u}{\partial x_i\partial x_j} -a_0\frac{\partial u}{\partial t}-\sum_{i=1}^{n} a_i\frac{\partial u}{\partial x_i}-au=0, \qquad (1)$$

with coefficients $a_{0i},\ldots,a$ that are functions in (t,x) is said to be **hyperbolic** (or of **hyperbolic type**) (with respect to the t-direction) in tx-space if the **characteristic equation** of equation (1) considered at each point of tx-space,

$$H(t,x;\lambda,\xi)=\lambda^2-\sum_{i=1}^{n} a_{0i}\xi_i\lambda-\sum_{i,j=1}^{n} a_{ij}\xi_i\xi_j=0, \qquad (2)$$

has two distinct real roots $\lambda=\lambda_1(t,x,\xi)$, $\lambda_2(t,x,\xi)$ for any n-tuple of real numbers $\xi=(\xi_1,\ldots,\xi_n)\neq(0,\ldots,0)$. In particular, (1) is called **regularly hyperbolic** if these two roots are separated uniformly, that is,

$$\lim_{(t,x),|\xi|=1} |\lambda_t(t,x,\xi)-\lambda_2(t,x,\xi)|=c>0.$$

A typical example of hyperbolic equations is the **wave equation**

$$\Box u=\frac{\partial^2 u}{\partial t^2}-\frac{\partial^2 u}{\partial x_1^2}-\ldots-\frac{\partial^2 u}{\partial x_n^2}=0. \qquad (3)$$

Equation (3) is also called the **equation of a vibrating string**, the **equation of a vibrating membrane**, or the **equation of sound propagation** according as $n=1$, 2, or 3. Another example is

$$\frac{\partial^2 u}{\partial t^2}-c^2\frac{\partial^2 u}{\partial x^2}-2\frac{\partial u}{\partial t}=0,$$

which describes the propagation of electric current in a conducting wire with leakage and is called the **telegraph equation** (→ Appendix A, Table 15).

A hyperplane $\lambda(t-t^0)+\ldots+\xi_n(x_n-x_n^0)=0$ passing through a point $p^0=(t^0,x^0)$ in tx-space and having normal direction (λ,ξ) is called a **characteristic hyperplane** of (1) at p^0 if the direction (λ,ξ) satisfies the characteristic equation at p^0: $H(t^0,x^0;\lambda,\xi)=0$. A hypersurface S: $s(t,x)=0$ in tx-space is a **characteristic hypersurface** of (1) if at each point of S the tangent hyperplane of S is a characteristic hyperplane of (1), that is, $H(t,x;s_t,s_x)=0$ everywhere on S. According to the theory of first-order par-

tial differential equations, a characteristic hypersurface of S is generated by so-called **bicharacteristic curves**, i.e., solution curves $t=t(\tau)$, $x(\tau)$ of a system of ordinary differential equations

$$\frac{dt}{d\tau}=H_\lambda, \quad \ldots, \quad \frac{dx_n}{d\tau}=H_{\xi_n},$$

$$\frac{d\lambda}{d\tau}=-H_t, \quad \ldots, \quad \frac{d\xi_n}{d\tau}=-H_{x_n},$$

$$H(t,x;\lambda,\xi)=0.$$

Now if (1) is hyperbolic, the set of all characteristic hyperplanes at $p^0:\{\lambda(t-t^0)+\ldots+\xi_n(x_n-x_n^0)=0 \mid H(t^0,x^0;\lambda,\xi)=0\}$ has as its envelope a cone $C(p^0)$ with the vertex p^0. Moreover, since the intersection of any hyperplane $t=$ constant and the cone $C(p^0)$ is an $(n-1)$-dimensional ellipsoid or two points for $n=1$, a conical body $D_+(p^0)$ $(D_-(p^0))$ is determined, whose boundary consists of the part of $C(p^0)$ with $t\geqslant t_0$ $(t\leqslant t_0)$ and the interior of the ellipsoid on the hyperplane $t=$constant. A †smooth curve γ in tx-space is called **timelike** if the tangent vector of γ at each point p on γ belongs to $D_+(p)$ or $D_-(p)$. Consider the set of points that can be connected with the point p^0 by a timelike curve. We call its closure an **emission**, and a subset $\mathscr{D}_+(p^0)$ $(\mathscr{D}_-(p^0))$ of the closure for which $t\geqslant t_0$ $(t\leqslant t_0)$ a **forward (backward) emission**. An emission is a conical body surrounded by characteristic hyperplanes in some neighborhood of the vertex p^0. If the coefficients of (1) are bounded functions, the emissions $\mathscr{D}_\pm(p^0)$ are contained in a conical body

$$K=\left\{(t,x) \mid \lambda_{\max}^2(t-t^0)^2\geqslant\sum_{i=1}^n(x_i-x_i^0)^2\right\}$$

independently of the situation of p^0, where $\lambda_{\max}=\max_{(t,x),|\xi|=1}(|\lambda_1(t,x,\xi)|,|\lambda_2(t,x,\xi)|)$.

B. The Cauchy Problem

Important for the hyperbolic equation (1) is the †Cauchy problem, i.e., the problem of finding a function $u=u(t,x)$ that satisfies (1) in $t>0$ and the initial conditions

$$u(0,x)=u_0(x), \qquad \partial u/\partial t(0,x)=u_1(x), \tag{4}$$

where the functions $u_0(x)$ and $u_1(x)$ are given on the initial hyperplane $t=0$.

Suppose that (1) is regularly hyperbolic and the coefficients are bounded and sufficiently smooth (i.e., of class C^ν with ν sufficiently large). Then for the Cauchy problem the following theorem holds. Theorem (C): There exists a positive integer l $(=[n/2]+3)$, depending on the dimension $n+1$ of the tx-space, such that if the functions $u_0(x)$ and $u_1(x)$ in (4) are of class C^l, then there exists a unique solution $u=u(t,x)$ of class C^2 in the domain $0\leqslant t<\infty$, $-\infty<x_i<\infty$ $(1\leqslant i\leqslant n)$. Moreover, this correspondence $\{u_0(x),u_1(x)\}\rightarrow u(t,x)$ is continuous in the following sense: If a sequence of initial functions $\{u_{0k}(x),u_{1k}(x)\}$ $(k=1,2,\ldots)$ and their derivatives up to the lth order tend to 0 uniformly on every compact set in the hyperplane $t=0$, then the sequence of corresponding solutions $u_k(t,x)$ also tends to 0 uniformly on every compact set in each hyperplane $t=$constant. In other words, the Cauchy problem for regularly hyperbolic equations is †well posed in the sense of Hadamard [2].

For dependence of the solution on initial data, the following proposition is valid: The values of the solution u at a point $p^0=(t^0,x^0)$ depend only on the initial data on a domain G_0 (**domain of dependence**) of the initial hyperplane, which is determined as the intersection of the backward emission $\mathscr{D}_-(p^0)$ and the initial hyperplane. We have the following dual proposition: A change in the initial conditions in a neighborhood of a point Q_0 of the initial hyperplane induces a change of values of the solution only in some neighborhood of the forward emission $\mathscr{D}_+(Q_0)$ (**domain of influence**). If the coefficients of the equation are bounded, the intersection of emissions $\mathscr{D}_\pm(p^0)$ and the hyperplane $t=$constant is always compact. It follows that the domain of dependence and the domain of influence are bounded. In some special cases, there exists a proper subdomain of G_0 such that the solution depends only on the initial data on the subdomain. For example, for the wave equation (3) with $n=3$, the solution for the Cauchy problem at a point $p^0=(t^0,x^0)$ (→ Section D) is determined, as can be seen from the solution formula (12), by the initial data in a neighborhood of the cone with vertex $p^0:(t-t^0)^2=\sum_{i=1}^n(x_i-x_i^0)^2$, namely, in a neighborhood of the intersection of bicharacteristic curves (lines, in this case) passing through p^0 and the initial hyperplane. If the solution of the Cauchy problem has such a property, it is said that **Huygens's principle** is valid, or that diffusion of waves does not occur. For the wave equation (3), Huygens's principle is valid only for odd $n>1$.

C. The Energy Inequality

The energy conservation law for the wave equation (3),

$$E(t) =\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left(\left(\frac{\partial u}{\partial t}\right)^2+\sum_{i=1}^n\left(\frac{\partial u}{\partial x_i}\right)^2\right)dx_1\ldots dx_n =\text{constant},$$

is generalized to the so-called energy inequality for hyperbolic equations, which plays an essential role in deducing the well-posedness of the Cauchy problem and the properties of the domain of dependence of the solution. Let the coefficients of a hyperbolic equation (1) be bounded functions, and let $G(\tau)$ be the intersection of the conical body $K=\{(t,x)\mid\lambda_{\max}^2(t-t^1)^2\geqslant\sum_{i=1}^n(x_i-x_i^1)^2\}$ with the hyperplane $t=\tau$ $(\tau<t^1)$. The $k(\geqslant 1)$th-order energy integral of the solution $u(t,x)$ of (1) on $G(\tau)$ is defined by

$$E_\tau^{(k)}(u,G(\tau))$$
$$=\int_{G(\tau)}\sum_{\alpha_0+\ldots+\alpha_n\leqslant k}\left|\frac{\partial^{|\alpha|}u}{\partial t^{\alpha_0}\ldots\partial x_n^{\alpha_n}}(\tau,x)\right|^2 dx.$$

Then the following inequality holds:

$$E_\tau^{(k)}(u,G(\tau))\leqslant CE_{t^0}^{(k)}(u,G(t^0)),\qquad t^0\leqslant\tau<t^1,\qquad(5)$$

where the constant C is independent of u. We call (5) the **energy inequality** (J. Schauder [5]). For the wave equation (3), the hypothesis $l=[n/2]+3$ in theorem (C) can be replaced by a weaker condition $l=[n/2]+2$, but if we take $l=[n/2]+1$, there is an example for which no global solution of class C^1 exists. In general, even though the initial functions are of class C^1, the solution in the Cauchy problem for hyperbolic equations may not be of class C^1, while if the energy of the initial functions is bounded, the energy of the solution is also bounded.

D. Representation Formulas for Solutions of the Cauchy Problem

We consider solution formulas that represent solutions of the Cauchy problem explicitly as functionals of the initial functions. The problem of solving (1) under the condition (4), or more generally, the problem of solving an equation $L[u]=f(t,x)$ under (4), can be reduced, by transforming the unknown function u and applying †Duhamel's method, to the Cauchy problem with initial conditions on the hyperplane $t=\tau$:

$$u(\tau,x)=0,\qquad \partial u/\partial t(\tau,x)=\varphi(x)\qquad(4')$$

for arbitrary τ. We define a function $\chi_q(s)$ of a real variable s by

$$\chi_q(s)=|s|^q/4(2\pi i)^{n-1}q!\qquad(n\text{ odd}),$$
$$=-s^q\log|s|/(2\pi i)^n q!\qquad(n\text{ even}),\qquad(6)$$

where n is the dimension of the x-space and q is a positive integer such that $q+n$ is even. Then for a function $\varphi(x)$ of class C^ν (with ν sufficiently large) with compact support, the following equality holds [7]:

$$\varphi(x)=\int_{-\infty}^{\infty}\Delta_y^{(n+q)/2}\varphi(y)\,dy\int_{|\omega|=1}\chi_q((x-y)\omega)\,d\omega,\qquad(7)$$

where Δ_y is the †Laplacian with respect to the variables $y=(y_1,\ldots,y_n)$ and $d\omega$ is the surface element of the unit sphere $|\omega|=1$ in x-space. Now, since the †principle of superposition is valid because (1) is linear, we can infer from formula (7) that the Cauchy problem (1) with initial condition (4′) can be reduced to the one for initial conditions with parameters y, ω:

$$u(\tau,x)=0,\qquad \partial u/\partial t(\tau,x)=\chi_q((x-y)\omega).\qquad(4'')$$

In fact, since $\chi_q(s)$ is $(q-1)$-times differentiable by definition, the Cauchy problem (1) with initial condition (4″) has a unique solution $R_q(t,x;\tau,y;\omega)$ for q chosen large enough so that theorem (C) can be applied to (1) with initial condition (4″). Moreover, $R_q(t,x;\tau,y;\omega)$ is a function of (t,x,τ,y,ω) of class C^ν, and ν increases with q. Now, let $\varphi(x)$ be a function of class C^ν with sufficiently large ν and with compact support. Then, by (7) and the definition of R_q, the integral

$$\int_{-\infty}^{\infty}\Delta_y^{(n+q)/2}\varphi(y)\,dy\int_{|\omega|=1}R_q(t,x;\tau,y;\omega)\,d\omega\qquad(8)$$

is a solution of the Cauchy problem (1) with initial condition (4′). Therefore, when R_q is found explicitly, (8) yields a solution formula of the Cauchy problem (1) with initial condition (4′) as a functional of the initial functions. Since in (8), the integral $\int_{|\omega|=1}R_q\,d\omega$ is not necessarily of class C^{n+q} as a function of $(t,x;\tau,y)$, $\Delta_y^{(n+q)/2}\int_{|\omega|=1}R_q\,d\omega$ is in general not a function in the ordinary sense. But we denote it by $R(t,x;\tau,y)$ formally, and understand that a linear operator

$$u(t,x)=\int R(t,x;\tau,y)\varphi(y)\,dy\qquad(9)$$

is defined by (8). The kernel $R(t,x;\tau,y)$ in this sense is called a **fundamental solution** or **Riemann function** of the Cauchy problem. If we extend the function $\int_{|\omega|=1}R_q(t,x;\tau,y;\omega)\,d\omega$ defined for $t\geqslant\tau$ to $t<\tau$, assigning it the value 0 there, then $R(t,x;\tau,y)=\Delta_y^{(n+q)/2}\int_{|\omega|=1}R_q(t,x;\tau,y;\omega)\,d\omega$ can be considered a †distribution on $(t,x;\tau,y)$-space, and the equality $L(t,x,\partial/\partial t,\partial/\partial x)R(t,x;\tau,y)=L^*(\tau,y,\partial/\partial\tau,\partial/\partial y)R(t,x;\tau,y)=\delta(t-\tau)\delta(x-y)$ is valid, where L^* is the †adjoint operator of L and δ is †Dirac's δ-function. In other words, $R(t,x;\tau,y)$ is a fundamental solution of L in the sense of distribution theory.

The fundamental solution $R(t,x;\tau,y)$ can be analyzed using the asymptotic expansion with

respect to the sequence of functions $\{\chi_q(s)\}$, and we have the following important result: If the coefficients of (1) are of class C^∞ (resp. real analytic), then the fundamental solution $R(t,x;\tau,y)$ is of class C^∞ (real analytic) in (t,x) except for points that are on bicharacteristic curves of (1) passing through the point (τ,y). In the language of the Cauchy problem, the smoothness of the solution u at a point $p=(t,x)$ depends only on the smoothness of the initial conditions on a neighborhood of the intersection of the initial hyperplane and all bicharacteristic curves passing through p. This fact is called **Huygens's principle in the wider sense**. Behavior of the fundamental solution $R(t,x;\tau,y)$ near discontinuous points has also been investigated [2].

For the wave equation (3), the fundamental solution can be constructed, and therefore we can write the solution formula explicitly. The solution formula for (3) with initial condition (4') for $n\geqslant 3$ is

$$u(t,x)=\frac{1}{(n-2)!}\frac{\partial^{n-2}}{\partial t^{n-2}}\int_0^t (t^2-\tau^2)^{(n-3)/2}\tau Q(x,\tau)\,d\tau,$$

$$Q(x,\tau)=\frac{1}{\omega_n}\int_{|\omega|=1}\varphi(x+\tau\omega)\,d\omega,$$

where $\omega_n=2\sqrt{\pi^n}/\Gamma(n/2)$ is the surface area of the unit sphere of n-dimensional space. Solutions of the Cauchy problem (3) with initial condition (4) for $n=1$, 2, and 3 are, respectively,

$$u(t,x)=\frac{u_0(x+t)+u_0(x-t)}{2}+\frac{1}{2}\int_{x-t}^{x+t}u_1(\xi)\,d\xi \tag{10}$$

(d'Alembert's solution),

$$u(t,x_1,x_2)=\frac{1}{2\pi}\frac{\partial}{\partial t}\int_{C_t}\frac{u_0(\xi_1,\xi_2)\,d\xi_1\,d\xi_2}{\sqrt{t^2-(x_1-\xi_1)^2-(x_2-\xi_2)^2}}+\frac{1}{2\pi}\int_{C_t}\frac{u_1(\xi_1,\xi_2)\,d\xi_1\,d\xi_2}{\sqrt{t^2-(x_1-\xi_1)^2-(x_2-\xi_2)^2}} \tag{11}$$

(Poisson's solution), and

$$u(t,x_1,x_2,x_3)=\frac{1}{4\pi}\frac{\partial}{\partial t}\int_{D_t}\frac{u_0(\xi_1,\xi_2,\xi_3)}{t}\,d\omega_t+\frac{1}{4\pi}\int_{D_t}\frac{u_1(\xi_1,\xi_2,\xi_3)}{t}\,d\omega_t \tag{12}$$

(Kirchhoff's solution), where C_t is a disk in the $\xi_1\xi_2$-plane with center (x_1,x_2) and radius t, D_t is a sphere in the $\xi_1\xi_2\xi_3$-space with center (x_1,x_2,x_3) and radius t, and $d\omega_t$ is the surface element of D_t.

E. Second-Order Nonlinear Hyperbolic Equations

A second-order nonlinear differential equation

$$\frac{\partial^2 u}{\partial t^2}=A\left(t,x,u,\frac{\partial u}{\partial t},\frac{\partial u}{\partial x_i},\frac{\partial^2 u}{\partial t\partial x_i},\frac{\partial^2 u}{\partial x_i\partial x_j}\right),\qquad 1\leqslant i,\quad j\leqslant n, \tag{13}$$

is called **hyperbolic** in a neighborhood of a function $U(t,x)$ if the linear equation of the form (1) obtained from (13) is hyperbolic, where $a_{0i}(t,x)$ and $a_{ij}(t,x)$ are determined by substituting u by $U(t,x)$ in the partial derivatives of A with respect to $\partial^2u/\partial t\partial x_i$ and $\partial^2 u/\partial x_i\partial x_j$, respectively. If the functions A in (13) and $U=x_0(x)+tu_1(x)$ determined by (4) are sufficiently smooth with respect to $t,x,u,\dots,\partial^2u/\partial x_i\partial x_j$, the Cauchy problem for (13) with initial condition (4) has a unique solution in some neighborhood of the initial hyperplane under the condition that equation (13) is hyperbolic in a neighborhood of U. In general, initial value problems for nonlinear equations have only local solutions.

F. Higher-Order Hyperbolic Equations

An Nth-order linear differential equation in $n+1$ variables t, $x=(x_1,\dots,x_n)$ with constant coefficients

$$L\left(\frac{\partial}{\partial t},\frac{\partial}{\partial x}\right)u=\sum_{\alpha_0+|\alpha|\leqslant N}a_{\alpha_0\alpha}\left(\frac{\partial}{\partial t}\right)^{\alpha_0}\left(\frac{\partial}{\partial x}\right)^{\alpha}u=0, \tag{14}$$

where $\alpha=(\alpha_1,\dots,\alpha_n)$, $|\alpha|=\alpha_1+\dots+\alpha_n$, and

$$\left(\frac{\partial}{\partial x}\right)^{\alpha}=\frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1}\dots\partial x_n^{\alpha_n}},$$

is **hyperbolic in the sense of Gårding** if the following two conditions are satisfied: (i) the partial derivative $\partial^N/\partial t^N$ appears in L; (ii) the real parts of the roots $\lambda=\lambda_1(\xi),\dots,\lambda_N(\xi)$ of the **characteristic equation** $L(\lambda,i\xi)=0$ are bounded functions of real variables $\xi=(\xi_1,\dots,\xi_n)$. When L is a homogeneous equation of the Nth order, condition (ii) is equivalent to the following condition: (ii') $\lambda_1(\xi),\dots,\lambda_N(\xi)$ are purely imaginary for all real $\xi=(\xi_1,\dots,\xi_n)\neq(0,\dots,0)$. The principal part (consisting of the highest-order terms) of a hyperbolic equation is also hyperbolic. If (14) is hyperbolic, a theorem analogous to theorem (C) holds for the Cauchy problem for (14) with initial conditions

$$\partial^k u/\partial t^k(0,x)=u_k(x),\qquad 0\leqslant k\leqslant n-1; \tag{15}$$

that is, the Cauchy problem for (14) with initial condition (15) is well posed in the sense of Hadamard. Conversely, if the Cauchy problem

for (14) with initial condition (15) is well posed, then (14) must satisfy the hyperbolicity conditions (i) and (ii) in the sense of Gårding. In other words, well-posedness of the Cauchy problem is equivalent to hyperbolicity in the sense of L. Gårding [14]. Gårding's conditions for hyperbolicity cannot be generalized to the case of variable coefficients, since the influence of the lower-order terms in the equation is taken into account in the definition of hyperbolicity. However, in the case of constant coefficients, an Nth-order homogeneous equation remains hyperbolic for any addition of lower-order terms if and only if the characteristic equation has N distinct purely imaginary roots for any real $\xi=(\xi_1,\ldots,\xi_n)\neq(0,\ldots,0)$. In this case the equation is called **hyperbolic in the strict sense**. Thus a linear equation with variable coefficients

$$L[u]=\frac{\partial^N u}{\partial t^N}+\sum_{\substack{\alpha_0+|\alpha|\leqslant N\\ \alpha_0<N}} a_{\alpha_0\alpha}(t,x)\left(\frac{\partial}{\partial t}\right)^{\alpha_0}\left(\frac{\partial}{\partial x}\right)^{\alpha}u$$

$$=0 \qquad (16)$$

is called **hyperbolic in the sense of Petrovskiĭ** if the characteristic equation

$$\lambda^N+\sum_{\alpha_0+|\alpha|=N} a_{\alpha_0\alpha}(t,x)\lambda^{\alpha_0}(i\xi)^{\alpha}=0$$

has N distinct purely imaginary roots (called **characteristic roots**) $\lambda_1(t,x,\xi),\ldots,\lambda_N(t,x,\xi)$ for each point $p=(t,x)$ and each $\xi\neq 0$. Moreover, if the characteristic roots $\lambda_1,\ldots,\lambda_N$ are separated uniformly, i.e., the inequality

$$\lim_{(t,x),|\xi|=1,j\neq k}|\lambda_j(t,x,\xi)-\lambda_k(t,x,\xi)|=c>0$$

holds, (16) is said to be **regularly hyperbolic**. In the second-order case, this definition is equivalent to the previous one. Theorem (C) holds for the Cauchy problem for a regularly hyperbolic equation (16) with initial conditions (15). For the domain of dependence of the solution, a result analogous to the case of the second-order equation can be obtained using an energy inequality [9, 10]. If the coefficients are of class C^∞, Huygens's principle in the wider sense is valid, that is, discontinuity of the solution is carried over only along bicharacteristic curves.

G. Systems of Hyperbolic Equations

For systems of equations

$$\sum_{j=1}^{l} L_{ij}[u_j]=0, \qquad 1\leqslant i\leqslant l,$$

where the L_{ij} are higher-order linear differential operators of the form (16), several types of hyperbolicity are formulated in connection with the well-posedness of the Cauchy problem. We take up two important types, hyperbolicity in the sense of Petrovskiĭ and symmetric hyperbolicity due to Friedrichs.

We call a system of linear differential equations

$$\sum_{j=1}^{l}\sum_{\alpha_0+|\alpha|\leqslant n_j} a^{ij}_{\alpha_0\alpha}(t,x)\left(\frac{\partial}{\partial t}\right)^{\alpha_0}\left(\frac{\partial}{\partial x}\right)^{\alpha}u_j=0,$$

$$1\leqslant i\leqslant l, \quad (17)$$

a **system of hyperbolic differential equations** (in the sense of Petrovskiĭ) if the determinant

$$\det\left(\sum_{\alpha_0+|\alpha|\leqslant n_j} a^{ij}_{\alpha_0\alpha}(t,x)\left(\frac{\partial}{\partial t}\right)^{\alpha_0}\left(\frac{\partial}{\partial x}\right)^{\alpha}\right), \qquad (18)$$

calculated formally using the matrix of differential operators in the system, is hyperbolic in the sense of Petrovskiĭ as a single equation of $N(=\sum_{j=1}^{l} n_j)$th order. Petrovskiĭ showed that the Cauchy problem for a system that is hyperbolic in this sense is well posed [10]. There were some imperfections in his argument, which have been corrected by others (→ S. Mizohata [12]). In the case of constant coefficients, the Cauchy problem for (17) is well posed if and only if (18) is hyperbolic in the sense of Gårding.

K. O. Friedrichs, observing that the energy inequality played an essential role in Petrovskiĭ's research, studied symmetric hyperbolic systems of equations, since for them the energy inequalities are valid most naturally. A system of first-order linear differential equations

$$A_0(t,x)\frac{\partial u}{\partial t}=\sum_{i=1}^{n} A_i(t,x)\frac{\partial u}{\partial x_i}+Bu \qquad (19)$$

is called **symmetric hyperbolic** (in the sense of Friedrichs) if the matrices $A_i(t,x)$ $(0\leqslant i\leqslant n)$ are symmetric and $A_0(t,x)$ is positive definite. A typical example is provided by Maxwell's equations. For this system it has been shown that the Cauchy problem is well posed and the domain of dependence of the solution is bounded [13].

H. Weakly Hyperbolic Operators

We adopt the following definition of hyperbolicity: a linear differential operator of Nth order

$$L=\frac{\partial^N}{\partial t^N}+\sum_{\substack{\alpha_0+|\alpha|\leqslant N\\ \alpha_0\leqslant N-1}} a_{\alpha_0\alpha}(t,x)\left(\frac{\partial}{\partial t}\right)^{\alpha_0}\left(\frac{\partial}{\partial x}\right)^{\alpha}$$

is called **hyperbolic** if the Cauchy problem for $L[u]=0$ with initial condition (15) is well posed in Hadamard's sense. A hyperbolic operator L is called **strongly hyperbolic** if L remains hyperbolic for any addition of lower-order terms, and **weakly hyperbolic** otherwise.

A necessary condition for hyperbolicity is that all characteristic roots of $L_N(t,x,\lambda,\xi)=0$ be real for any (t,x,ξ) (P. D. Lax [15], Mizohata [16]). In the case of constant coefficients, strongly hyperbolic operators have been characterized by K. Kasahara and M. Yamaguti (*Mem. Coll. Sci. Kyoto Univ.* (1960)).

As for operators with variable coefficients, it is known that not only regularly hyperbolic operators but also some special classes of not regularly hyperbolic operators are strongly hyperbolic (V. Ya. Ivriĭ, *Moscow Math. Soc.*, 1976). Let us consider the hyperbolicity of an operator L which is not regularly hyperbolic under the assumption that the multiplicities of the characteristic roots are constant and at most 2; namely, the characteristic polynomial

$$L_N(t,x,\lambda,\xi)=\lambda^N+\sum_{\substack{\alpha_0+|\alpha|=N\\ \alpha_0\leqslant N-1}} a_{\alpha_0\alpha}(t,x)\lambda^{\alpha_0}\xi^\alpha$$

is decomposed in the following way:

$$L_N(t,x,\lambda,\xi)$$
$$=\prod_{j=1}^{s}(\lambda-\lambda_j(t,x,\xi))^2\prod_{j=s+1}^{N-s}(\lambda-\lambda_j(t,x,\xi)),$$

$$\lim_{\substack{(t,x)\\ |\xi|=1,\, j\neq k}}|\lambda_j(t,x,\xi)-\lambda_k(t,x,\xi)|=c>0.$$

Assume that the $\lambda_j(t,x,\xi), j=1,2,\ldots,N-s$, are real. Then L is hyperbolic if and only if it satisfies **E. E. Levi's condition**, i.e.,

$$l_j(t,x,\xi)=\left[L_{N-1}+\frac{1}{2}\left(\frac{\partial^2 L_N}{\partial\lambda^2}\frac{\partial\lambda_j}{\partial t}+\sum_{\alpha=1}^{n}\frac{\partial^2 L_N}{\partial\lambda\partial\xi_\alpha}\frac{\partial\lambda_j}{\partial x_\alpha}\right)\right]_{\lambda=\lambda_j}=0$$
$$\text{for all } (t,x,\xi),\quad j=1,2,\ldots,s,$$

where L_{N-1} denotes the homogeneous part of $(N-1)$th order among lower-order terms of L (Mizohata and Y. Ohya [17]). Thus, for a **weakly hyperbolic** operator with variable coefficients, even the principal part is not necessarily hyperbolic. J. Chazarain [18] has studied weakly hyperbolic operators with characteristic roots of arbitrary constant multiplicity. O. A. Oleĭnik [19] studied the Cauchy problem with nonconstant multiplicities for second-order equations. For higher-order equations, if the multiplicity of characteristic roots at $(\hat t,\hat x)$ is at most p, or more precisely, if there exist positive rational numbers q and r $(q\geqslant r)$ such that

$$\left(\frac{\partial}{\partial\lambda}\right)^p L_N(\hat t,\hat x,0,\xi)\neq 0$$

and

$$\left(\frac{\partial}{\partial\lambda}\right)^{\alpha_0}\left(\frac{\partial}{\partial\xi}\right)^{\alpha}\left(\frac{\partial}{\partial t}\right)^{\beta_0}\left(\frac{\partial}{\partial x}\right)^{\beta}L_N(\hat t,\hat x,0,\xi)=0 \qquad ({}^\forall\xi\in\mathbf{R}^n\setminus 0)$$

for $\alpha_0+|\alpha|+q\beta_0+r|\beta|<p$, then it is necessary for the well posedness of the Cauchy problem that

$$\left(\frac{\partial}{\partial\lambda}\right)^{\alpha_0}\left(\frac{\partial}{\partial\xi}\right)^{\alpha}\left(\frac{\partial}{\partial t}\right)^{\beta_0}\left(\frac{\partial}{\partial x}\right)^{\beta}L_{N-s}(\hat t,\hat x,0,\xi)=0 \qquad ({}^\forall\xi\in\mathbf{R}^n\setminus 0)$$

for $\alpha_0+|\alpha|+q\beta_0+r|\beta|+s(1+q)<p$ are satisfied, where L_{N-s} $(1\leqslant s\leqslant N)$ are the homogeneous lower-order terms of order $N-s$ of L (Ivriĭ and V. M. Petkov [20]).

On the other hand, the sufficient condition in the case of multiplicity 2 is given by some conditions related to the subprincipal symbol

$$L_{N-1}-\frac{1}{2}\left(\frac{\partial^2 L_N}{\partial\lambda\partial t}+\sum_{\alpha=1}^{n}\frac{\partial^2 L_N}{\partial\xi_\alpha\partial x_\alpha}\right),$$

which corresponds to Levi's condition in the case of constant multiplicity, and to the †Poisson brackets (A. Menikoff, *Amer. J. Math.*, 1976; Ohya, *Ann. Scuola Norm. Sup. Pisa*, 1977; L. Hörmander, *J. Anal. Math.*, 1977).

The Cauchy problem for a weakly hyperbolic system of equations is more complicated, because of the essential difficulty that the matrix structure

$$\left(\sum_{\alpha_0+|\alpha|=n_j} a^{ij}_{\alpha_0\alpha}(t,x)\lambda^{\alpha_0}\xi^\alpha\right)$$

associated with (17) is not clear in general (→ references in [20]).

I. Gevrey Classes

Classically, the functions of †class s $(s\geqslant 1)$ of Gevrey (→ 58 C^∞-Functions and Quasi-Analytic Functions G; 168 Function Spaces B (14) were introduced into the studies of the fundamental solution for the heat equation:

$\gamma^{(s)}_{\mathrm{loc}}(\mathbf{R}^n)=\{\varphi(x)\in C^\infty(\mathbf{R}^n)\,|$ for any compact subset K of $\mathbf{R}^n$ and any multi-indices α, there exist constants C_K and A such that $\sup_K|(\partial/\partial x)^\alpha\varphi(x)|\leqslant C_K A^{|\alpha|}|\alpha|!^s\}$.

This class of functions was used efficiently in the studies of the Cauchy problem for †weakly hyperbolic partial differential equations:

$$L\left(t,x,\frac{\partial}{\partial t},\frac{\partial}{\partial x}\right)u(t,x)=f(t,x) \qquad \text{in}\quad [0,T]\times\mathbf{R}^n=\Omega,$$

$$\left(\frac{\partial}{\partial t}\right)^j u(0,x)=u_j(x),\qquad j=0,1,\ldots,N-1.$$

We assume that the multiplicities of the characteristic roots are constant, i.e.,

$$L_N(t,x,\lambda,\xi)=\prod_{i=1}^{k}(\lambda-\lambda_i(t,x,\xi))^{\nu_i},$$

where ν_i is constant for any $(t,x,\xi)\in\Omega\times\mathbf{R}^n$, $\lambda_i(t,x,\xi)$ is real and distinct, and $\sum_{i=1}^k \nu_i=N$. Let $\max_{1\leqslant i\leqslant k}\nu_i=p$. If we suppose that $L(t,x,\partial/\partial t,\partial/\partial x)-\tilde{L}(t,x,\partial/\partial t,\partial/\partial x)$ is a (†pseudo) differential operator of order at most q, where

$$\tilde{L}=\prod_{i=1}^{p} a_i\left(t,x,\frac{\partial}{\partial t},\frac{\partial}{\partial x}\right)$$

is a (pseudo)differential operator with $a_i(t,x,\partial/\partial t,\partial/\partial x)$ being strictly hyperbolic (pseudo) differential operators associated with L_N, then for any s such that $1\leqslant s<p/q$, the Cauchy problem is well posed in $\gamma^{(s)}_{\mathrm{loc}}(\Omega)$, provided that all $a_{\alpha_0\alpha}(t,x)$ of L and $f(t,x)$ belong to $\gamma^{(s)}_{\mathrm{loc}}(\Omega)$, and that $\{u_j(x)\}_{0\leqslant j\leqslant N-1}$ are given in $\gamma^{(s)}_{\mathrm{loc}}(\mathbf{R}^n)$ (Ohya [21], J. Leray and Ohya [22]). This result was proved even for the case of arbitrary nonconstant multiplicity of characteristic roots by M. D. Bronshteĭn [23].

J. Lacunas for Hyperbolic Operators

The theory of **lacunas** of fundamental solutions of hyperbolic operators, initiated by Petrovskiĭ [24], has been developed further in a paper [25] by M. F. Atiyah, R. Bott, and L. Gårding.

Let $L(\xi)=L_N(\xi)+M(\xi)$ be a hyperbolic polynomial with respect to the vector $\theta\in\mathbf{R}^n-0$, where $L_N(\xi)$ is the principal part of L; this means that $L_N(\theta)\neq 0$ and $L(\xi+t\theta)\neq 0$ when $|\mathrm{Im}\, t|$ is sufficiently large. Then L has a fundamental solution $E=E(L,\theta,x)$ in the form

$$E(L,\theta,x)=(2\pi)^{-n}\int L(\xi-ic\theta)^{-1}e^{ix(\xi-ic\theta)}\,d\xi,$$

where c is sufficiently large and the integral is taken in the sense of †distribution. The convex hull of the support of E, denoted by $K=K(L,\theta)=K(A,\theta)$, is a cone depending only on the real part $\mathrm{Re}\,A$ of the complex hypersurface $A:L(\xi)=0$, and contained in the union of the origin and the half-space $x\cdot\theta>0$. Let A_ξ be the †tangent conoid of A at ξ, transported to the origin, and define the wavefront surface $W(A,\theta)$ by the union of all $K(A_\xi,\theta)$ for $\xi\neq 0$. Then it can be shown that the singular supports of $E(L,\theta)$ and all the $E(L_N^k,\theta)$ are contained in $W(A,\theta)$ and, moreover, that they are locally holomorphic outside W. In [25], the Herglotz-Petrovskiĭ-Leray formulas are generalized to any nonstrict $L(\xi)$. Thus we have

$$D^\beta E(L_N^k,\theta,x)=\mathrm{const}\int_{\alpha^*}(x\cdot\xi)^q\xi^\beta L_N^k(\xi)^{-k}\omega(\xi),\quad q>0, \tag{20}$$

$$D^\beta E(L_N^k,\theta,x)=\mathrm{const}\int_{t_x\cdot\partial\alpha^*}(x\cdot\xi)^q\xi^\beta L_N^k(\xi)^{-k}\omega(\xi),\quad q\leqslant 0 \tag{20'}$$

when $x\in K(A,\theta)\setminus W(A,\theta)$. Here, $q=mk-|\beta|-n$ is the degree of homogeneity of the left-hand side, and $\omega=\sum_{j=1}^n(-1)^{j-1}\xi_j d\xi_1\wedge\ldots\wedge\widehat{d\xi_j}\wedge\ldots\wedge d\xi_n$. The integrands are closed †rational $(n-1)$-forms on $(n-1)$-dimensional complex †projective space and are integrated over certain †homology classes $\alpha^*=\alpha(A,\theta,x)^*$ and $t_x\cdot\partial\alpha^*$. These formulas provide means of obtaining topological criteria for lacunas. Let $\mathscr{L}\subset\mathring{K}$ be a maximal connected open set, where $E(L,\theta,x)$ is holomorphic. $\mathscr{L}$ is said to be a **weak (strong) lacuna** of L if $E(L,\theta,x)$ is the restriction of an entire function to $\mathscr{L}$($E(L,\theta,x)=0$ in $\mathscr{L}$). In [25] it is shown that x belongs to a weak lacuna for all $E(L_N^k,\theta,\cdot)$ if and only if $\partial\alpha^*=0$. The sufficiency directly follows from (20′) and the necessity follows from a theorem of A. Grothendieck (*Publ. Math. Inst. HES*, 1966) which implies that the rational forms which appear in (20′) span all the †cohomology classes in question.

K. Mixed Initial-Boundary Value Problems

Let Ω be a domain in $\mathbf{R}^n$ with a sufficiently smooth boundary Γ, let $L(t,x,\partial/\partial t,\partial/\partial x)$ be a linear hyperbolic operator of Nth order defined in $[0,\infty)\times\bar{\Omega}=\{(t,x)\,|\,t\in[0,\infty),x\in\bar{\Omega}\}$, and let $B_j(t,x,\partial/\partial t,\partial/\partial x)$, $j=1,2,\ldots,b$, be linear differential operators of N_jth order defined in a neighborhood of $[0,\infty)\times\Gamma$.

The problem of finding a function $u(t,x)$ satisfying the conditions

$$L[u]=0\quad\text{in}\quad(0,\infty)\times\Omega,$$
$$B_j[u]=0\quad\text{on}\quad(0,\infty)\times\Gamma,\quad j=1,2,\ldots,b,$$
$$\partial^k u/\partial t^k(0,x)=u_k(x),\quad 0\leqslant k\leqslant N-1, \tag{21}$$

is called a **mixed initial-boundary value problem**. A typical example of such a mixed problem is provided by the case $L=\square$ $(n=2)$ and $B[u]=u(t,x)$, which describes the vibration of membranes with a fixed boundary.

The mixed problem (21) is said to be well posed if for any initial data $u_k(x)\in C^\infty(\bar{\Omega})$, $0\leqslant k\leqslant N-1$, which are compatible with the boundary conditions, there exists a unique solution $u(t,x)\in C^\infty([0,\infty)\times\bar{\Omega})$.

Mixed problems for second-order hyperbolic equations are considered in [6]. In regard to mixed problems for hyperbolic equations of higher order, we make the following four assumptions: (i) $\Omega=\mathbf{R}^n_+=\{(x',x_n)\,|\,x'\in\mathbf{R}^{n-1},x_n>0\}$; (ii) $\Gamma=\{x\,|\,x_n=0\}$ is not characteristic for L or B_j; (iii) L is regularly hyperbolic; (iv) $N_j\leqslant N-1$ and $N_j\neq N_k$ if $j\neq k$.

We denote the †principal parts of L and B_j by $L_N(t,x',x_n,\partial/\partial t,\partial/\partial x',\partial/\partial x_n)$ and $B_{j0}(t,x',\partial/\partial t,\partial/\partial x',\partial/\partial x_n)$, respectively. By the hyper-

bolicity of L, an equation in κ

$$L_N(t,x',0,\lambda,\xi',\kappa)=0 \quad \text{for} \quad \operatorname{Im}\lambda<0, \quad \xi'\in\mathbf{R}^{n-1}$$

has μ roots κ_j^+ with $\operatorname{Im}\kappa_j^+>0$, and $N-\mu$ roots κ_j^- with $\operatorname{Im}\kappa_j^-<0$, and the number μ is independent of (t,x') and (λ,ξ'). A necessary condition for the well-posedness of the mixed problem (21) is that the number of boundary conditions coincide with this integer μ. The function R defined by

$$R(t,x',\lambda,\xi') = \det\left[\oint_C \frac{B_{j0}(t,x',\lambda,\xi',\kappa)\kappa^{l-1}}{L^+(t,x',\lambda,\xi',\kappa)}d\kappa\right]_{j,l=1,2,\ldots,\mu},$$

where $L^+=\prod_{j=1}^{\mu}(\kappa-\kappa_j^+(t,x',\lambda,\xi'))$ and C is a contour enclosing all κ_j^+, is called a **Lopatinski determinant**.

We say that L and B_j satisfy the uniform Lopatinski condition if

$$\inf_{\substack{(t,x'),\operatorname{Im}\lambda<0\\|\xi'|+|\lambda|=1}} |R(t,x',\lambda,\xi')|=c>0.$$

When the uniform Lopatinski condition is satisfied, the mixed problem (21) is well posed, and (21) represents a phenomenon with a finite propagation speed, which is the same as that of the Cauchy problem for $L[u]=0$ (T. Balaban [26], H. O. Kreiss [27], and R. Sakamoto [28]). An analogous result holds in the case of a domain Ω with a compact boundary Γ, provided that L and B_j satisfy the uniform Lopatinski condition at every point of Γ.

In the treatment of mixed problems for L and B_j not satisfying the uniform Lopatinski condition, the well-posed problems have been characterized for operators with constant coefficients when $\Omega=\mathbf{R}^n_+$ (Sakamoto [29]). For general domains, however, the well-posedness of mixed problems depends not only on the properties of the Lopatinski determinant but also on the shape of the domain (M. Ikawa [30]).

L. Asymptotic Solutions

In order to explain some properties of phenomenon governed by hyperbolic equations, asymptotic solutions play an important role. Consider, for example, the **acoustic problem**

$$\Box u=0 \quad \text{in } (0,\infty)\times\Omega,$$

$$u=0 \quad \text{on } (0,\infty)\times\Gamma.$$

Let $w(t,x;k)$ be a function defined in $(0,\infty)\times\Omega$ with parameter $k\geqslant 1$ of the form

$$w(t,x;k)=e^{ik(\varphi(x)-t)}\sum_{j=0}^{N} v_j(t,x)k^{-j}, \tag{22}$$

where φ is a smooth function satisfying the **eikonal equation** $|\nabla\varphi|^2=1$. If v_j, $j=0,1,\ldots,N$, satisfy the **transport equations**

$$2\frac{\partial v_j}{\partial t}+2\nabla\varphi\cdot\nabla v_j+\Delta\varphi v_j=-i\Box v_{j-1}, \quad v_{-1}=0,$$

we have

$$\Box w=O(k^{-N}).$$

Then w of (22) is an approximate solution of $\Box u=0$ for large k, and it represents a wave propagating in the direction $\nabla\varphi$.

When $\operatorname{supp} w\cap(0,\infty)\times\Gamma\neq\varphi$, if w hits the boundary Γ transversally, we can construct

$$w^+(t,x;k)=e^{ik(\varphi^+(x)-t)}\sum_{j=0}^{N} v_j^+(t,x)k^{-j}$$

such that

$$|\nabla\varphi^+|^2=1 \ \text{in } \Omega, \quad \varphi^+=\varphi, \quad \text{and} \quad \frac{\partial\varphi}{\partial\nu}>0 \ \text{on } \Gamma,$$

and v_j^+ satisfy the transport equations and $v_j^+=-v_j$ on $(0,\infty)\times\Gamma$, where ν is the unit inner normal of Γ. Then $w+w^+$ is an approximate solution, and w^+ represents a **reflected wave** propagating in the direction

$$\nabla\varphi^+=\nabla\varphi-2(\nabla\varphi\cdot\nu)\nu.$$

These asymptotic solutions show that the high-frequency waves propagate approximately according to the laws of †geometric optics.

If asymptotic solution (22) has a **caustic**, i.e., $\{\{x+l\nabla\varphi(x)\mid l\in R\}\mid x\in\operatorname{supp} w\}$ has an envelope, w of the form (22) cannot be an asymptotic solution near the caustic. The asymptotic behavior of high-frequency solutions near the caustic was first considered by G. B. Airy (*Trans. Cambridge Philos. Soc.*, 1838). Under the condition that the principal curvatures of the caustic are positive, w in the form (22) can be prolonged to a domain containing the caustic satisfying the asymptotic solution

$$w(t,x;k)=e^{ik(\theta(x)-t)}\{\operatorname{Ai}(-k^{2/3}\rho(x))g_0(t,x;k) + ik^{-1/3}\operatorname{Ai}'(-k^{2/3}\rho(x))g_1(t,x;k)\}, \tag{23}$$

where Ai is the Airy function

$$\operatorname{Ai}(z)=\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i(zt+t^3/3)}dt$$

and

$$g_l(t,x;k)=\sum_j g_{lj}(t,x)k^{1/6-j}$$

(D. Ludwig [31]).

Concerning the reflection of **grazing rays** by

strictly convex obstacles, the reflected wave can be constructed by the superposition of asymptotic solutions of the type (23).

The methods of construction of asymptotic solutions of the forms (22) and (23) are also applicable to Maxwell equations or more general hyperbolic systems (R. K. Luneberg [32]; Ludwig and Granoff, *J. Math. Anal. Appl.*, 1968; Guillemin and Sternberg, *Amer. Math. Soc. Math. Surveys*, 14 (1977)).

M. Propagation of Singularities

Let L be a hyperbolic operator with C^∞ coefficients and consider the Cauchy problem $L[u]=0$ with initial condition (15). When the initial data have singularities, the solution also has singularities for $t>0$, which is a property of hyperbolic equations quite different from the properties of parabolic ones. It should be noted that the †propagation of singularities cannot be derived from the Huygens principle in the wider sense, i.e., even for regularly hyperbolic operators of second order we cannot determine the location and the type of singularities of the solutions for initial data with singularities directly from the singularities of the fundamental solution $R(t,x;\tau,y)$.

Suppose that the multiplicities of characteristic roots of L are constant. Assume that u_k, $k=0,1,\dots,N-1$, have, on either side of a sufficiently smooth $(n-1)$-dimensional manifold Γ, continuous derivatives of sufficiently high order to suffer jump discontinuities across Γ. Then the solution u has continuous partial derivatives of sufficiently high order everywhere except on the characteristic surfaces of L issuing from Γ, and across these the partial derivatives of u have jump discontinuities (Courant and Hilbert [1]).

For more general singularities of initial data, it is known that the †wavefront propagates along the †bicharacteristic strips satisfying $\varphi_t-\lambda_l(t,x,\nabla\varphi)=0$, $l=1,2,\dots,N-s$, that is, $\mathrm{WF}u(\cdot,t)$ is contained in $\{(x(t),\xi(t))\in T^*(\mathbf{R}^n)\mid (dx_j/dt)(s)=(\partial\lambda_l/\partial\xi_j)(s,x,\xi),\ (d\xi_j/dt)(s)=-(\partial\lambda_l/\partial x_j)(s,x,\xi),\ (x(0),\xi(0))\in\bigcup_k \mathrm{WF}(u_k)\}$ (J. Chazarain [18]).

The propagation of singularities is more complicated in mixed problems because of the reflections of singularities at the boundary. For the †acoustic problem, R. B. Melrose [33] showed the following: Suppose $\mathcal{O}=C\Omega\subset\{x\mid |x|<R\}$ for some $R>0$, and all the broken rays according to the geometric optics starting from $\Omega_R=\Omega\cap\{|x|<R\}$ go out of Ω_R in a fixed time. Then for initial data with singularities in Ω_R, the solution becomes smooth in Ω_R for sufficiently large t.

References

[1] R. Courant and D. Hilbert, Methods of mathematical physics II, Interscience, 1962.
[2] J. Hadamard, Le problème de Cauchy et les équations aux dérivées partielles linéaires hyperboliques, Hermann, 1932.
[3] J. Hadamard, Leçons sur la propagation des ondes et les équations de l'hydrodynamique, Hermann, 1903.
[4] R. Courant and K. O. Friedrichs, Supersonic flow and shock waves, Interscience, 1948.
[5] J. Schauder, Das Anfangswertproblem einer quasilinearen hyperbolischen Differentialgleichung zweiter Ordnung, Fund. Math., 24 (1935), 213–246.
[6] M. Krzyżański and J. Schauder, Quasilineare Differentialgleichungen zweiter Ordnung vom hyperbolischen Typus, gemischte Randwertaufgaben, Studia Math., 6 (1936), 162–189.
[7] F. John, Plane waves and spherical means applied to partial differential equations, Interscience, 1955.
[8] S. L. Sobolev, Applications of functional analysis in mathematical physics, Amer. Math. Soc., 1963. (Original in Russian, 1950.)
[9] J. Leray, Hyperbolic differential equations, Lecture notes, Institute for Advanced Study, Princeton, 1952.
[10] I. G. Petrovskiĭ, Über das Cauchysche Problem für Systeme von partiellen Differentialgleichungen, Mat. Sb. (N.S.), 2 (44) (1937), 815–870.
[11] I. G. Petrovskiĭ, Lectures on partial differential equations, Interscience, 1954. (Original in Russian, 1953.)
[12] S. Mizohata, The theory of partial differential equations, Cambridge, 1973. (Original in Japanese, 1965.)
[13] K. O. Friedrichs, Symmetric hyperbolic linear differential equations, Comm. Pure Appl. Math., 7 (1954), 345–392.
[14] L. Gårding, Linear hyperbolic equations with constant coefficients, Acta Math., 85 (1951), 1–62.
[15] P. D. Lax, Asymptotic solutions of oscillatory initial value problems, Duke Math. J., 24 (1957), 627–646.
[16] S. Mizohata, Some remarks on the Cauchy problem, J. Math. Kyoto Univ., 1 (1961), 109–127.
[17] S. Mizohata and Y. Ohya, Sur la condition de E. E. Levi concernant des équations hyperboliques, Publ. Res. Inst. Math. Sci., 4 (1968), 511–526; Sur la condition d'hyper-

bolicité pour les équations à caractéristiques multiples II, Japan. J. Math., 40 (1971), 63–104.
[18] J. Chazarain, Opérateurs hyperboliques à caractéristiques de multiplicité constante, Ann. Inst. Fourier, 24 (1974), 173–202; Propagation des singularités pour une classe d'opérateurs à caractéristiques multiples et résolubilité locale, Ann. Inst. Fourier, 24 (1974), 203–223.
[19] O. A. Oleĭnik, On the Cauchy problem for weakly hyperbolic equations, Comm. Pure Appl. Math., 23 (1970), 569–586.
[20] V. Ya. Ivriĭ and V. M. Petkov, Necessary conditions for the correctness of the Cauchy problem for nonstrictly hyperbolic equations, Russian Math. Surveys, 29 (1974), 1–70. (Original in Russian, 1974.)
[21] Y. Ohya, Le problème de Cauchy pour les équations hyperboliques à caractéristique multiple, J. Math. Soc. Japan, 16 (1964), 268–286.
[22] J. Leray and Y. Ohya, Systèmes linéaires, hyperboliques non stricts, 2ième Colloque sur l'Analyse Fonctionnelle, CBRM (1964), 105–144.
[23] M. D. Bronshteĭn, The parametrix of the Cauchy problem for hyperbolic operators with characteristics of variable multiplicity, Functional Anal. Appl., 10 (1976) 4, 83–84. (Original in Russian, 1976.)
[24] I. G. Petrovskiĭ, On the diffusion of waves and the lacunas for hyperbolic equations, Mat. Sb. (N.S.), 17 (59) (1945), 289–370.
[25] M. F. Atiyah, R. Bott, and L. Gårding, Lacunas for hyperbolic differential operators with constant coefficients I, II, Acta Math., 124 (1970), 109–189; 134 (1973), 145–206.
[26] T. Balaban, On the mixed problem for a hyperbolic equation, Mem. Amer. Math. Soc., 112 (1971).
[27] H. O. Kreiss, Initial boundary value problems for hyperbolic systems, Comm. Pure Appl. Math., 23 (1970), 277–298.
[28] R. Sakamoto, Mixed problems for hyperbolic equations I, II, J. Math. Kyoto Univ., 10 (1970), 349–373, 403–417.
[29] R. Sakamoto, $\mathscr{E}$-well posedness for hyperbolic mixed problems with constant coefficients, J. Math. Kyoto Univ., 14 (1974), 93–118.
[30] M. Ikawa, On the mixed problems for the wave equation in an interior domain II, Osaka J. Math., 17 (1980), 253–279.
[31] D. Ludwig, Uniform asymptotic expansions at a caustic, Comm. Pure Appl. Math., 19 (1966), 215–250.
[32] R. K. Luneberg, Mathematical theory of optics, Brown Univ. Press, 1944.
[33] R. B. Melrose, Singularities and energy decay in acoustical scattering, Duke Math. J., 45 (1979), 43–59.

326 (XIII.27) Partial Differential Equations of Mixed Type

A. General Remarks

Let $A[u(x)]=0$ be a †quasilinear second-order partial differential equation. The type (†elliptic, †hyperbolic, or †parabolic) of the equation depends on the location of the point x. If the type varies as the point x moves, the equation is said to be of **mixed type**. An example is the equation

$$\left(1-\frac{u^2}{c^2}\right)\frac{\partial^2\varphi}{\partial x^2}-\frac{2uv}{c^2}\frac{\partial^2\varphi}{\partial x\partial y}+\left(1-\frac{v^2}{c^2}\right)\frac{\partial^2\varphi}{\partial y^2}=0 \qquad (1)$$

of 2-dimensional stationary flow without rotation of a compressible fluid without viscosity, where φ is the velocity potential, $u=\partial\varphi/\partial x$ and $v=\partial\varphi/\partial y$ are the velocity components, and c is the local speed of sound, which is a known function of the speed $q=(u^2+v^2)^{1/2}$ of the flow. Equation (1) is of elliptic type if $q<c$, i.e., the flow is **subsonic**, and of hyperbolic type if $q>c$, i.e., the flow is **supersonic**. If there exist points where the flow is subsonic as well as points where it is supersonic, (1) is of mixed type. The study of equations of mixed type has become important with the development of high-speed jet planes.

B. Chaplygin's Differential Equation

It is difficult to solve equation (1) directly since it is nonlinear. However, we can linearize it by taking q and $\theta=\arctan(v/u)$ as independent variables (the hodograph transformation). The linearized equation takes the form

$$\frac{\partial^2 z}{\partial x^2}-K(x)\frac{\partial^2 z}{\partial y^2}=0, \qquad xK(x)\geqslant 0, \qquad (2)$$

which is called **Chaplygin's differential equation**. Equation (2) is hyperbolic for $x>0$ and elliptic for $x<0$. The study of general equations of mixed type, even when they are linear, is much more difficult and less developed than the study of equations of nonmixed type. Almost all research so far has been on equation (2) or slight modifications of it.

C. Tricomi's Differential Equation

The simplest equation of the form (2) is

$$\frac{\partial^2 z}{\partial x^2}-x\frac{\partial^2 z}{\partial y^2}=0, \qquad (3)$$

which is called **Tricomi's differential equation**. F. G. Tricomi considered the following boundary value problem for (3): In Fig. 1, AC and BC are two †characteristic curves of (3) and σ is a Jordan curve connecting A and B. We seek a solution of (3) in the domain D bounded by AC, BC, and σ that takes given values on σ and on one of the two characteristic curves, say on AC. This boundary value problem is called the **Tricomi problem**. Tricomi proved the existence and uniqueness of the solution of his problem under some conditions on the shape of σ and the smoothness of the boundary values. After Tricomi, much research has been done on his and similar problems for equations of form (2) [2]. We can also consider problems such as finding a solution of (3) (or of (2)) satisfying the initial conditions

$$z(0,y)=z_1(y), \quad (\partial z/\partial x)(0,y)=z_2(y)$$

on the common boundary $x=0$ of the elliptic domain and the hyperbolic domain of the equation. This is called the **singular initial value problem**. S. Bergman [3] obtained an integral formula for the solution under the condition that $z_1(y)$ and $z_2(y)$ are real analytic.

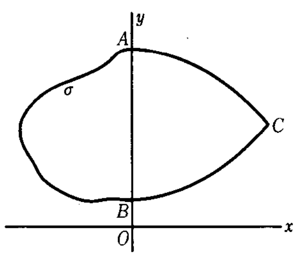


Fig. 1

D. Friedrichs's Theory

For the study of equations of mixed type it would of course be most convenient if there existed a general theory of boundary value problems independent of the type of the equation. However, constructing such a general theory is considered very difficult, because the †well-posedness of boundary conditions as well as the analytic properties of solutions are quite different according to the type. The first contributor to the solution of this difficult problem was K. O. Friedrichs [4], who noticed that although the methods of solving the †Cauchy problem and the †Dirichlet problem are quite different, both methods utilize energy integrals in the proof of the uniqueness of solutions. Using this observation, he succeeded in constructing a unified theory that enables us to treat various types (including the mixed type) of linear equations in a single scheme—an admissible boundary value problem for a **symmetric positive system** of first-order linear partial differential equations. There is, however, a difficulty in Friedrichs's theory since it does not give a unified procedure for reducing a given boundary value problem for a given equation to an admissible boundary value problem for a symmetric positive system of partial differential equations. The study of equations of mixed type that are of more general form than (2) by means of Friedrichs's theory is an open problem.

E. Further Studies

Work on equations of more general type than (2) or (3) has appeared (not all depending on Friedrichs's theory). For example, the following equations are treated in [5, 6, 7], respectively:

$$\sum_{i=1}^{n}\frac{\partial^2 z}{\partial x_i^2}+\frac{t}{|t|}\frac{\partial z}{\partial t}=0,$$

$$\left(\frac{\partial^2}{\partial x^2}+\frac{y}{|y|}\frac{\partial^2}{\partial y^2}\right)^n z=0,$$

$$G(y)\frac{\partial^2 z}{\partial x^2}+\frac{\partial^2 z}{\partial y^2}-K(y)z=h,$$

where $z=(z_1,\ldots,z_n)$ and $G(y)$ and $K(y)$ are symmetric matrices.

References

[1] F. Tricomi, Sulle equazioni lineari alle derivate parziali di 2° ordine di typo misto, Atti Acad. Naz. Lincei, (5) 14 (1923), 133–247.
[2] L. Bers, Mathematical aspects of subsonic and transonic gas dynamics, Wiley, 1958.
[3] S. Bergman, An initial value problem for a class of equations of mixed type, Bull. Amer. Math. Soc., 55 (1949), 165–174.
[4] K. O. Friedrichs, Symmetric positive linear differential equations, Comm. Pure Appl. Math., 11 (1958), 333–418.
[5] A. V. Bitsadze, On the problem of Equations of mixed type in multi-dimensional domains (in Russian), Dokl. Akad. Nauk SSSR, 110 (1956), 901–902.
[6] V. I. Zhegalov, Boundary value problem for higher-order equations of mixed type (in Russian), Dokl. Akad. Nauk SSSR, 136 (1961), 274–276.
[7] V. P. Didenko, Some systems of differential equations of mixed type (in Russian), Dokl. Akad. Nauk SSSR, 144 (1962), 709–712.
[8] A. V. Bitsadze, Equations of the mixed type, Pergamon, 1964. (Original in Russian, 1959.)
[9] M. M. Smirnov, Equations of mixed type, Amer. Math. Soc. Transl., 1978. (Original in Russian, 1970.)

327 (XIII.26) Partial Differential Equations of Parabolic Type

A. General Remarks

Consider a second-order linear partial differential equation

$$\sum_{i,j=1}^{n}\left[a_{ij}(x,t)\frac{\partial^2 u}{\partial x_i \partial x_j}+\sum_{i=1}^{n} b_i(x,t)\frac{\partial u}{\partial x_i}\right]+c(x,t)u-\frac{\partial u}{\partial t}=f \quad (1)$$

for an unknown function u of $(n+1)$ independent variables $(x,t)=(x_1,\ldots,x_n,t)$, where $a_{ij}=a_{ji}$. This equation is said to be **parabolic** (or of **parabolic type**) if and only if the quadratic form $\sum a_{ij}\xi_i\xi_j$ in ξ is positive definite at each point (x,t) of the region under consideration. x and t are sometimes called the variables of space and of time, respectively.

The most widely studied of the parabolic equations is the **equation of heat conduction** (or the **heat equation**):

$$L[u]\equiv\Delta u-\frac{\partial u}{\partial t}=0, \quad (2)$$

where $\Delta=\sum_{i=1}^{n}\partial^2/\partial x_i^2$ is the Laplacian taken over the space variables.

B. The Equation of Heat Conduction

The 1-dimensional case of the heat equation is

$$L[u]\equiv\frac{\partial^2 u}{\partial x^2}-\frac{\partial u}{\partial t}=0 \quad (3)$$

for the temperature $u(x,t)$ in a rod, considered as a function of the distance x measured along the rod and the time t. Equation (3) was one of the first treated in the theory of partial differential equations. Consider a finite rod with constant temperature 0 at its ends $x=a$ and $x=b$. Thermodynamics suggests that the initial temperature $\varphi(x)$ $(\varphi(a)=\varphi(b)=0)$ prescribed at $t=0$ is sufficient to determine the distribution of heat $u(x,t)$ in the rod at all later times $t>0$. On such physical grounds, we can expect that a solution to the following problem exists: Find a continuous function $u(x,t)$ that satisfies equation (3) for $a\leqslant x\leqslant b$, $t>0$, and the boundary conditions

$$u(a,t)=u(b,t)=0,$$

$$\lim_{t\downarrow 0}u(x,t)=\varphi(x), \qquad a\leqslant x\leqslant b. \quad (4)$$

According to J. Fourier the answer to this problem is expressed in a series $\sum_{n=1}^{\infty}c_n u_n$ constructed by superposition of the particular solutions $u_n=\sin\sqrt{\lambda_n}\,(x-a)\exp(-\lambda_n t)$. Here the λ_n are the roots of $\sin\sqrt{\lambda_n}\,(b-a)=0$, and the c_n are chosen so that $\sum_{n=1}^{\infty}c_n u_n(x,0)=\varphi(x)$. In fact, if $\varphi(x)$ is continuously differentiable, then $\sum_{n=1}^{\infty}c_n u_n(x,t)$ is the required solution. Thus we are led to the problem of expanding a given function $\varphi(x)$ in a †Fourier series.

The temperature distribution in an infinite rod is given by a continuous function $u(x,t)$ that satisfies equation (3) for $t>0$ and that, for $t=0$, takes values given by $\varphi(x)$, where

$$\lim_{t\downarrow 0}u(x,t)=\varphi(x). \quad (5)$$

If $\varphi(x)$ is bounded, then it can also be represented by superposition of particular solutions $e^{-(\alpha-x)^2/4t}$ of (3) as

$$u(x,t)=\begin{cases}\dfrac{1}{2\sqrt{\pi t}}\displaystyle\int_{-\infty}^{\infty}\varphi(\alpha)e^{-(\alpha-x)^2/4t}\,d\alpha, & t>0,\\ \varphi(x), & t=0.\end{cases} \quad (6)$$

Partial differential equations of parabolic type are important because of their connection with various phenomena in the physical world; they include not only equations that govern the flow of heat but also those that describe diffusion processes (→ 115 Diffusion Processes).

C. Partial Differential Equations of Parabolic Type in Two Variables

We are concerned mainly with the partial differential equation of parabolic type in two variables:

$$a(x,t)\frac{\partial^2 u}{\partial x^2}+2b(x,t)\frac{\partial^2 u}{\partial x\,\partial t}+c(x,t)\frac{\partial^2 u}{\partial t^2}+d(x,t)\frac{\partial u}{\partial x}+e(x,t)\frac{\partial u}{\partial t}+f(x,t)u=g, \quad (7)$$

with $ac=b^2$. In the region where $|a|+|c|>0$, equation (7) can be reduced to the form

$$\frac{\partial^2\omega}{\partial\xi^2}+d'(\xi,\tau)\frac{\partial\omega}{\partial\xi}+e'(\xi,\tau)\frac{\partial\omega}{\partial\tau}+f'(\xi,\tau)\omega=g' \quad (7')$$

by an appropriate change of variables $\xi=U(x,t)$, $\tau=V(x,t)$. If $e'<0$ in this region, we can assume without loss of generality that our equation takes the canonical form

$$a(x,t)\frac{\partial^2 u}{\partial x^2}+b(x,t)\frac{\partial u}{\partial x}+c(x,t)u-\frac{\partial u}{\partial t}=g, \quad (8)$$

with $a>0$, from the outset. It has the single family of †characteristics

$$t=\text{constant}. \quad (9)$$

There are four typical problems to be posed with regard to equation (8).

The first consists of determining, in some neighborhood of a given curve C nowhere tangent to a characteristic, a solution u that possesses prescribed values of u and $\partial u/\partial n$, or of a linear combination of u and $\partial u/\partial n$, along C. For instance, the problem of finding a solution $u(x,t)$ such that $u(x_0,t)=g(t)$ and $u_x(x_0,t)=h(t)$ for given functions $g(t)$ and $h(t)$ is a problem of this type. Consider the equation

$$\frac{\partial^2 u}{\partial x^2}-\frac{\partial u}{\partial t}=0 \tag{10}$$

in the region $a<t<b$, $x_0\leqslant x$. According to E. Holmgren, a solution $u(x,t)$ satisfying the conditions

$$\lim_{x\downarrow x_0} u(x,t)=g(t), \qquad \lim_{x\downarrow x_0} u_x(x,t)=h(t), \tag{11}$$

where $g'(t)$ is bounded and continuous, exists if and only if

$$h(t)+\frac{1}{\sqrt{\pi}}\int_a^t \frac{g'(\tau)}{\sqrt{t-\tau}}d\tau=k(t) \tag{12}$$

is of class C^∞ and satisfies

$$|k^{(n)}(t)|\leqslant M(n!)^2/r^n \tag{13}$$

for positive constants M and r.

In the second type of problem we are required to find in a region of the form

$$\varphi_1(t)\leqslant x\leqslant\varphi_2(t), \qquad t_1<t<t_2, \tag{14}$$

a solution of (8) that takes prescribed values on part of the boundary of that region. Here we impose the hypothesis on the curves $x=\varphi_1(t)$ and $x=\varphi_2(t)$ that they are nowhere tangent to a characteristic (9). M. Gevrey [3] showed that if such a solution does exist, the functions $\varphi_1(t)$ and $\varphi_2(t)$ must satisfy the †Hölder condition with exponent $\alpha>1/2$:

$$|\varphi_i(t+h)-\varphi_i(t)|\leqslant c|h|^\alpha, \qquad c=\text{constant}, \tag{15}$$

for sufficiently small h. The problem of heat conduction mentioned in Section B corresponds to the particular case $\varphi_1(t)=\text{constant}$, $\varphi_2(t)=\text{constant}$, for which condition (15) is automatically fulfilled.

The third type of problem is to find in a region of the form

$$a\leqslant x\leqslant b, \qquad t>0, \tag{14'}$$

a solution of (8) that satisfies the conditions

$$\lim_{t\downarrow 0} u(x,t)=\varphi(x);$$

$$\frac{\partial u}{\partial x}-hu=0 \quad \text{for} \quad x=a, \quad h=\text{constant}>0;$$

$$\frac{\partial u}{\partial x}+Hu=0 \quad \text{for} \quad x=b, \quad H=\text{constant}>0. \tag{16}$$

This problem, posed for equation (10), is also a mathematical formulation of the problem of heat conduction in a rod [4]. If $\varphi(x)$ is of class C^1, then the solution to this problem can be expanded as

$$u(x,t)=\sum_{n=1}^{\infty} c_n e^{-\lambda_n t}\varphi_n(x),$$

$$c_n=\int_a^b \varphi(x)\varphi_n(x)\,dx, \tag{17}$$

where $\varphi_n(x)$ is a normalized function ($\int_a^b\varphi_n^2(x)dx=1$) that satisfies the boundary conditions

$$\varphi_n'(a)-h\varphi_n(a)=0, \qquad \varphi_n'(b)+H\varphi_n(b)=0 \tag{18}$$

and the equation

$$\frac{d^2\varphi_n(x)}{dx^2}=-\lambda_n\varphi_n(x). \tag{19}$$

The fourth type of problem is to find a function $u(x,t)$ that satisfies (8) for $t>0$ and the initial condition $\lim_{t\downarrow 0}u(x,t)=\varphi(x)$. It corresponds to the problem of heat conduction in an infinite rod.

D. Green's Formula

The †adjoint of the differential operator $L[u]$ in (3) is given by

$$M[v]\equiv\frac{\partial^2 v}{\partial x^2}+\frac{\partial v}{\partial t}. \tag{20}$$

Integration by parts yields the identity

$$\iint_G (\psi L[\varphi]-\varphi M[\psi])\,dx\,dt$$

$$=\int_C \varphi\psi\,dx+\int_C\left(\psi\frac{\partial\varphi}{\partial x}-\varphi\frac{\partial\psi}{\partial x}\right)dt, \tag{21}$$

where G is the region bounded by the closed curve C, and the line integral on the right is evaluated in the counterclockwise direction over C. We call (21) **Green's formula** for the partial differential equation of parabolic type (3). As in the case of partial differential equations of elliptic type (→ 323 Partial Differential Equations of Elliptic Type), this formula is used to establish the uniqueness of the solution of (3) and to derive an integral representation for it.

For example, the uniqueness of the solution is established in the following way: Let the curve $\widehat{AD}$ and $\widehat{BE}$ in Fig. 1 be such that no characteristic meets either of them in more than one point. If $u(x,t)$ is continuous in the closed region $(ABED)$, vanishes on $\widehat{AD}$, $\widehat{BE}$, and the segment $\overline{AB}$, and satisfies equation (3) in $(ABED)$ except on $\overline{AB}$, then it vanishes identically. Green's formula (21), applied to the region $G=(ABQP)$ and the functions $\psi\equiv 1$,

$\varphi\equiv u^2$, yields

$$2\iint_G\left(\frac{\partial u}{\partial x}\right)^2 dx\,dt+\int_{\overline{PQ}}u^2\,dx=0.$$

In a similar way, by extending Green's formula suitably, we are able to prove uniqueness theorems for the four problems stated in Section C for more general linear parabolic equations.

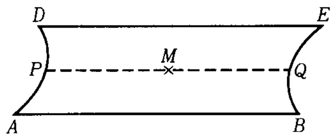


Fig. 1

To obtain a representation for solutions we proceed as follows: Let $u(x,t)$ be a solution of (3) and Let

$$U(\alpha,\beta,x,t)=\frac{1}{\sqrt{4\pi(\beta-t)}}e^{-(\alpha-x)^2/4(\beta-t)} \tag{22}$$

be a particular solution. Applying Green's formula to the region $(PABQMP)$ and the functions $\varphi=u(x,t)$, $\psi=U(x_0,t_0+h,x,t)$, where h is a positive number and M is a point with coordinates (x_0,t_0), we obtain

$$\int_{\overline{PQ}}u(x,t_0)e^{-(x_0-x)^2/4h}\frac{dx}{\sqrt{4\pi h}}$$

$$=\int_{\overline{PABQ}}u(x,t)U(x_0,t_0+h,x,t)\,dx$$

$$+\left(U\frac{\partial u}{\partial x}-u\frac{\partial U}{\partial x}\right)dt.$$

Since the integral in the left-hand side of this equality approaches $u(x_0,t_0)$ as $h\downarrow 0$, we can establish the basic representation formula

$$u(x_0,t_0)=\int_{\overline{PABQ}}u(x,t)U(x_0,t_0,x,t)\,dx$$

$$+\left(U\frac{\partial u}{\partial x}-u\frac{\partial U}{\partial x}\right)dt \tag{23}$$

for solutions u of (3). Formula (23) shows that $u(x_0,t_0)$ is determined in terms of the particular solution (22) if we know the values of u and $\partial u/\partial x$ on the part $\overline{PABQ}$ of the boundary of the region $(ABED)$. The function (22) is called the **fundamental solution** of (3) because it plays the same role as the fundamental solution $\log r$ $(r=((\alpha-x)^2+(\beta-y)^2)^{1/2})$ of Laplace's equation $\partial^2u/\partial x^2+\partial^2u/\partial y^2=0$.

Similarly, the following function E (called the **Gauss kernel**) is a fundamental solution of equation (2):

$$E(\alpha,\beta;x,t)=(4\pi(\beta-t))^{-n/2}\exp\left\{-\frac{|\alpha-x|^2}{4(\beta-t)}\right\}, \tag{24}$$

where $(x,\alpha)\in\mathbf{R}^n\times\mathbf{R}^n$ and $t<\beta$. For equation (3), the following **maximum principle** holds: In the region $(ABED)$ of Fig. 1, suppose that $L[u]\geqslant 0$ and that u takes its maximum value K at an interior point M. Then u is identically equal to K on the segment $\overline{QP}$ and in the region $(ABQP)$. More generally, various versions of the maximum principle are known for equation (1) [5].

E. The Laplace Transform Method

Let $u(x,t)$ be a solution of (3) for $t>0$ and

$$v(x,\lambda)=\int_0^\infty e^{-\lambda t}u(x,t)\,dt,\qquad \lambda>0, \tag{25}$$

be its †Laplace transform with respect to t. Utilizing integration by parts, we have

$$\int_0^\infty e^{-\lambda t}u_t(x,t)\,dt=[e^{-\lambda t}u(x,t)]_{t=0}^{t=\infty}$$

$$+\lambda\int_0^\infty e^{-\lambda t}u(x,t)\,dt$$

$$=-\varphi(x)+\lambda v(x,\lambda), \tag{26}$$

provided that $\lim_{t\to\infty}e^{-\lambda t}u(x,t)=0$ and $\lim_{t\downarrow 0}u(x,t)=\varphi(x)$. We find in view of (3):

$$\frac{\partial^2 v}{\partial x^2}=\lambda v(x,\lambda)-\varphi(x). \tag{26'}$$

Once the solution of (26′) has been found, the desired solution of (3) can be derived by inverting the Laplace transform (25). This idea can also be applied to the solution of parabolic equations with constant coefficients in $(n+1)$ variables, such as (1).

F. General Second-Order Equations of Parabolic Type

Consider the equation (1) with $f=0$, which can be written as

$$\frac{\partial u}{\partial t}=A(t)u, \tag{27}$$

where $A(t)$ is a second-order †elliptic operator with parameter t. Let D be a region (bounded or unbounded) of points x whose boundary is a smooth hypersurface S. We pose the following initial boundary value problem for (27): Find a function $u(x,t)$ that satisfies in $D\times(0,\infty)$ equation (27) together with the conditions

$$\lim_{t\downarrow 0}u(x,t)=\varphi(x),\quad x\in D,$$

$$\partial u(x,t)/\partial n+h(x,t)u(x,t)=f(x,t),\quad x\in S, \tag{16'}$$

where $\partial/\partial n$ is the directional derivative in the

outward †conormal direction at (x, t) $(x \in S)$, and $h(x, t) \geqslant 0$.

The Laplace transform is not suitable for solving problems (27) and (16′). Instead, the theory of 1-parameter semigroups of linear operators (→ 378 Semigroups of Operators and Evolution Equations) can be applied to establish similar fundamental results. Let m be a large positive integer. For $t>0$, put $t_k = k\delta$ for $k = 0, 1, \ldots, m-1$ with $\delta = t/m$. By the Laplace transform method as described above, we can associate with ψ a unique solution v of $A(t_k)v = \lambda v - \psi$ with $\lambda = 1/\delta$. We put $R_k\psi = \lambda v$. By iterating this procedure m times, we have a function $u_m(x, t) = R_{m-1}R_{m-2}\ldots R_0\varphi$ starting from the initial value φ at $t=0$. Then we obtain a solution $u(x, t)$ of (27) and (16′) as the limit of $u_m(x, t)$ as $m \to \infty$. The following results are known [6, 7]: (i) There exists a $u = U(\xi, \tau, x, t)$ $(x, \xi \in D, t > \tau \geqslant 0)$ that, as a function of x and t, satisfies equation (27) and the homogeneous boundary conditions (16′) with $\varphi = 0$, $f = 0$. (ii) The function $u(x, t)$ defined by

$$u(x,t) = \int_D \varphi(\xi) U(\xi, 0, x, t) d\xi + \int_0^t d\tau \int_S f(\xi, \tau)\left(U(\xi, \tau, x, t) - \frac{\partial U(\xi, \tau, x, t)}{\partial n_\xi}\right) dS_\xi \quad (23')$$

is a solution of (27) and satisfies (16′). Thus $U(\xi, \tau, x, t)$ is a generalization of the function (24), called the **fundamental solution** of the linear parabolic equation (27) with boundary conditions (16′). Besides the properties (i) and (ii), the fundamental solution satisfies $U(\xi, \tau, x, t) \geqslant 0$, $\int_D U(\xi, \tau, z, w) U(z, w, x, t) dz = U(\xi, \tau, x, t)$ $(\tau < w < t)$, and further $\int_D U(\xi, \tau, x, t) dx = 1$ under some additional assumptions. Therefore this theory is of considerable significance from the point of view of the theory of probability (→ 115 Diffusion Processes).

It can be shown that a weak solution of the parabolic equation (27) is a genuine solution. That is, if $u(x, t)$ is locally summable and

$$\int_0^\infty \int_D u(x,t)\left(\frac{\partial \varphi(x,t)}{\partial t} + A(t)^* \varphi(x,t)\right) dx\, dt = 0$$

for any function $\varphi(x, t)$ of class C^2 in $D \times (0, \infty)$ and vanishing outside a compact subset of D ($A(t)^*$ is the adjoint of the partial differential operator $A(t)$ and $dx = dx_1 \ldots dx_{n-1}$), then $u(x, t)$ satisfies (27) in $D \times (0, \infty)$ in the usual sense. In particular, when the coefficients of $A(t)$ are infinitely differentiable, any solution $u(x, t)$ of (27) in the distribution sense is a genuine solution (→ 125 Distributions and Hyperfunctions).

If the function h in the boundary conditions (16′) and the coefficients of A are independent of t, then the fundamental solution $U(\xi, \tau, x, t)$ depends only on ξ, x, and $t - \tau$ and is written as $U(\xi, x; t-\tau)$ $(t > \tau)$. Furthermore, if A is †self-adjoint, then there exist a sequence of †eigenfunctions $\{\psi_p(x;\lambda) \mid p = 1, 2, \ldots\}$ $(A\psi_p + \lambda\psi_p = 0)$ and a sequence $\{\rho_p(\Lambda)\}$ of measures on the real line for which the following hold: (1) The fundamental solution $U(\xi, x; t)$ is expressed in the form

$$U(\xi, x; t) = \sum_{p=1}^\infty \int_{-\infty}^\infty e^{-\lambda t} \psi_p(\xi;\lambda)\psi_p(x;\lambda)\, d\rho_p(\lambda).$$

(2) The solution $u(x, t)$ of (27) satisfying (16′) with $f(x, t) \equiv 0$ is expanded as

$$u(x,t) = \sum_{p=1}^\infty \int_{-\infty}^\infty e^{-\lambda t} \psi_p(x;\lambda)\varphi_p(\lambda)\, d\rho_p(\lambda),$$

where

$$\varphi_p(\lambda) = \int_D \psi_p(x;\lambda)\varphi(x)\, dx$$

($\varphi(x)$ is the function given in (16′).

G. Nash's Results

Let us consider a parabolic equation

$$\frac{\partial u}{\partial t} = \sum_{i,j=1}^n \frac{\partial}{\partial x_j}\left\{ a_{ij}(x,t) \frac{\partial u}{\partial x_i}\right\}, \tag{28}$$

where $a_{ij} = a_{ji}$ are real-valued functions of class C^∞ and equal to constants outside a fixed compact set of $\mathbf{R}^n$ for all $t \geqslant t_0$ (this regularity assumption can be relaxed). Suppose that there exists a constant $\lambda \geqslant 1$ such that $\lambda^{-1}|\xi|^2 \leqslant \sum a_{ij}(x, t)\xi_i\xi_j \leqslant \lambda|\xi|^2$ for all $(t, x, \xi) \in (t_0, \infty) \times \mathbf{R}^n \times \mathbf{R}^n$. Then, for any bounded solution $u(x, t)$ of (28) in $(t_0, \infty) \times \mathbf{R}^n$ and for any (x, y, t, s) such that $x \in \mathbf{R}^n$, $y \in \mathbf{R}^n$, and $t_0 < t \leqslant s$, the inequality

$$|u(x,t) - u(y,s)| \leqslant AB\left\{\left(\frac{x-y}{\sqrt{t-t_0}}\right)^\alpha + \left(\frac{s-t}{t-t_0}\right)^\beta\right\}, \tag{29}$$

holds, where $B = \sup\{|u(x, t)| \mid t \geqslant t_0, x \in \mathbf{R}^n\}$ and $\beta = \alpha/(2\alpha + 2)$. In this inequality, the constants α and A are positive, depending on (n, λ) but independent of the particular choice of (a_{ij}), t_0 and of u [9].

As a corollary to this theorem, J. Nash proved that, if the a_{ij} do not contain t and if $v(x)$ is a bounded solution in $\mathbf{R}^n$ of the elliptic equation obtained by replacing $\partial u/\partial t$ by 0 in (28), the inequality

$$|v(x) - v(y)| \leqslant A'B'|x-y|^\sigma \tag{30}$$

holds for any $(x, y) \in \mathbf{R}^n \times \mathbf{R}^n$, where $\sigma = \alpha/(\alpha+1)$ with α in (29) and $B' = \sup|v(x)|$. The constant

A' depends only on (n, λ) (→ 323 Partial Differential Equations of Elliptic Type L).

H. Partial Differential Equations of p-Parabolic Type

Let p and m be given positive integers. Let us consider an equation for an unknown function u of $(n+1)$ independent variables (x, t) of the type

$$\frac{\partial^m u}{\partial t^m}+\sum_{\alpha,j} a_{\alpha,j}(x,t)\frac{\partial^\alpha}{\partial x^\alpha}\frac{\partial^j u}{\partial t^j}=f, \qquad (31)$$

where $\alpha=(\alpha_1,\ldots,\alpha_n)$ and $\partial^\alpha/\partial x^\alpha=(\partial/\partial x_1)^{\alpha_1}\ldots(\partial/\partial x_n)^{\alpha_n}$. We write also $|\alpha|=\alpha_1+\ldots+\alpha_n$. In (31), $\Sigma_{\alpha,j}$ is the summation taken over the (α,j) such that $pj+|\alpha|\leqslant pm$ and $0\leqslant j<m$. Let us denote by $\{\lambda_k(x,t,\xi)\}_{k=1}^m$ the roots λ of the equation

$$\lambda^m+\sum_{\alpha,j}{}' a_{\alpha,j}(x,t)(i\xi)^\alpha\lambda^j=0, \qquad (32)$$

where $\Sigma'_{\alpha,j}$ is the summation over the (α,j)'s such that $pj+|\alpha|=pm$ and that $0\leqslant j<m$. We say that the equation (31) is **p-parabolic** (or of **p-parabolic type**) in the sense of I. G. Petrovskiĭ if and only if there exists a positive number δ such that

$$\operatorname{Re}\lambda_k(x,t,\xi)\leqslant-\delta|\xi|^p, \quad 1\leqslant k\leqslant m, \qquad (33)$$

for any (x,t) in the region under consideration and for any $\xi\in\mathbf{R}^n$. The integer p is then seen to be even. Equation (27) is p-parabolic if $-A(t)$ is strongly elliptic of order p. The heat equation (2) is 2-parabolic in this sense. Similarly, we can define the p-parabolic systems of equations [10].

p-parabolic equations are known to be †hypoelliptic if the coefficients are of class C^∞ [11]. S. D. Eĭdel'man obtained precise estimates of the fundamental solutions and of their derivatives for p-parabolic equations [10]. The mixed initial boundary value problems are investigated in detail also by Eĭdel'man [10] and by R. Arima (*J. Math. Kyoto Univ.*, 4 (1964)).

References

[1] L. Bers, S. Bochner, and F. John, Contributions to the theory of partial differential equations, Ann. Math. Studies, Princeton Univ. Press, 1954.
[2] A. Friedman, Partial differential equations of parabolic type, Prentice-Hall, 1964.
[3] M. Gevrey, Les équations aux dérivées partielles du type paraboliques, J. Math. Pures Appl., 9 (1913), 305–471; 10 (1914), 105–147.
[4] R. Courant and D. Hilbert, Methods of mathematical physics, Interscience, I, 1953, II, 1962.
[5] M. H. Protter and H. F. Weinberger, Maximum principles in differential equations, Prentice-Hall, 1967.
[6] S. Ito, Fundamental solutions of parabolic differential equations and boundary value problems, Japan, J. Math., 27 (1957), 55–102.
[7] A. M. Il'in, A. S. Kalashnikov, and O. A. Oleĭnik, Linear equations of the second order of parabolic type, Russian Math. Surveys, 17-3 (1962), 1–143. (Original in Russian, 1962.)
[8] J. L. Lions, Equations différentielles opérationnelles et problèmes aux limites, Springer, 1961.
[9] J. Nash, Continuity of solutions of parabolic and elliptic equations, Amer. J. Math., 80 (1958), 931–954.
[10] S. D. Eĭdel'man, Parabolic systems, North-Holland, 1969. (Original in Russian, 1963.)
[11] S. Mizohata, Hypoellipticité des équations paraboliques, Bull. Soc. Math. France, 85 (1957), 15–50.

328 (V.6) Partitions of Numbers

A **partition** of a positive integer n is an expression of n as the sum of positive integers. The number of partitions of n, where the order of the summands is ignored and repetition is permitted, is denoted by $p(n)$ and is called the **number of partitions** of n. For example $p(5)=7$ since $5=4+1=3+2=3+1+1=2+2+1=2+1+1+1=1+1+1+1+1$. Therefore, $p(n)$ equals the number of †conjugate classes of the †symmetric group of order n and is closely related to the †representation theory of this group.

The †generating function of $p(n)$ is

$$f(x)=1+\sum_{n=1}^{\infty}p(n)x^n=\left(\sum_{n=1}^{\infty}(1-x^n)\right)^{-1}.$$

The unit circle $|x|=1$ is the †natural boundary of $f(x)$, which is holomorphic in $|x|<1$. The **Dedekind eta function**, which is closely related to $f(x)$, is defined by the following formula for the complex variable τ taking values in the upper half-plane:

$$\eta(\tau)=\exp(\pi i\tau/12)\prod_{n=1}^{\infty}(1-\exp(2\pi i n\tau)).$$

Hence $\eta(\tau+1)=\exp(\pi i/12)\eta(\tau)$. L. Euler (1748) obtained the following formula (called the **pentagonal number theorem** because $n(3n-1)/2$

is a †pentagonal number):

$$\prod_{n=1}^{\infty}(1-x^n)$$

$$=1+\sum_{n=1}^{\infty}(-1)^n(x^{n(3n-1)/2}+x^{n(3n+1)/2}).$$

This follows easily from the **Jacobi-Biehler equality**

$$\prod_{n=1}^{\infty}\{(1-q^{2n})(1+q^{2n-1}z)(1+q^{2n-1}z^{-1})\}$$

$$=1+\sum_{n=1}^{\infty}q^{n^2}(z^n+z^{-n}) \qquad (|q|<1, \quad z\neq 0).$$

By using the †transformation formula for ϑ-functions, we can infer from the pentagonal number theorem that $\eta(-1/\tau)=\sqrt{\tau/i}\,\eta(\tau)$. Hence, if a, b, c, d are integers satisfying $ad-bc=1$ and $c>0$, then

$$\eta\left(\frac{a\tau+b}{c\tau+d}\right)=\varepsilon\sqrt{\frac{c\tau+d}{i}}\eta(\tau),$$

where ε is a 24th root of unity. It is known that $\eta(\tau)$ is a †cusp form of weight $-1/2$ with respect to the full †modular group $\Gamma(1)$ [9, 11]. C. L. Siegel (1955) gave a simple proof of the formula $\eta(-1/\tau)=\sqrt{2/i}\,\eta(\tau)$. Later S. Iseki (1957) gave another proof by using a new method, known as the $\alpha-\beta$ formula [12].

The size of $p(n)$ increases rapidly with n; for instance $p(10)=42$ and $p(100)=190{,}569{,}292$. By making use of a remarkable identity, G. H. Hardy and S. Ramanujan (1918) proved the following inequalities, where A and B are suitable constants:

$$(A/n)e^{2\sqrt{n}}<p(n)<(B/n)e^{2\sqrt{2}\sqrt{n}}.$$

Subsequently they obtained

$$p(n)\sim(1/4\sqrt{3}\,n)\exp(\pi\sqrt{2n/3}).$$

After, P. Erdös (1942) and D. J. Newman (1951), A. G. Postnikov [8] succeeded in proving

$$p(n)=\frac{\exp(\pi\sqrt{2\pi/3})}{4\sqrt{3}\,n}\left(1+O\left(\frac{\log n}{n^{1/4}}\right)\right)$$

by means of an elementary function-theoretic method. Multiplying both sides of Euler's formula by the generating function of $p(n)$ and comparing the coefficients, we obtain

$$\sum_{0\leqslant\omega_k\leqslant n}(-1)^{k-1}p(n-\omega_k)=0,$$

where $\omega_k=k(3k-1)/2$ $(k=0,\pm1,\pm2,\dots)$ is a pentagonal number. From this formula we can calculate $p(n)$ successively; in fact P. A. MacMahon obtained in this way the values of $p(n)$ for n up to 200.

Hardy and Ramanujan proved the following **transformation formula** for the generating function $f(x)$ of $p(n)$:

$$f\left(\exp\left(\frac{2\pi ih}{k}-\frac{2\pi z}{k}\right)\right)$$

$$=W_{h,k}\sqrt{z}\exp\left(\frac{\pi}{12kz}-\frac{\pi z}{12k}\right)$$

$$\times f\left(\exp\left(\frac{2\pi ih'}{k}-\frac{2\pi}{kz}\right)\right),$$

$$(h,k)=1, \quad hh'\equiv-1(\mathrm{mod}\,k),$$

where $W_{h,k}$ is defined by

$$W_{h,k}=\exp(\pi is(h,k))$$

and the value of $s(h,k)$ was given by Rademacher in the form

$$s(h,k)=\sum_{m=1}^{k-1}\frac{m}{k}\left(\left(\frac{hm}{k}\right)\right).$$

Here the symbol $((t))$ in the sum denotes the function that is 0 for integral t and $t-[t]-1/2$ otherwise ([ ] is the †Gauss symbol). With regard to the **Dedekind sum** $s(h,k)$, we have the **reciprocity law for Dedekind sums**:

$$s(h,k)+s(k,h)=-\frac{1}{4}+\frac{1}{12}\left(\frac{h}{k}+\frac{k}{h}+\frac{1}{hk}\right).$$

If we make substitutions $a=h'$, $b=(hh'+1)/k$, $c=k$, and $d=-h$ in the Hardy-Ramanujan transformation formula, then the ε appearing in the transformation formula of $\eta(\tau)$ is seen to be equal to $\exp(-\pi is(c,d)+\pi i(a+d)/12c)$. A direct proof of the transformation formula and the reciprocity law was given by K. Iseki (1952).

According to Cauchy's integral formula, $p(n)$ can be represented as an integral:

$$\frac{1}{2\pi i}\int_{\Gamma}\frac{f(x)}{x^{n+1}}dx,$$

where the contour Γ is taken inside the unit circle around the origin. The generating function $f(x)$ varies greatly: namely, letting $r\to1-0$ in $x=r\exp(2\pi ip/q)$, where p and q are fixed integers, it follows from the transformation formula that $f(x)\sim\exp(\pi^2/6q^2(1-r))$. Nevertheless, we can deal with the integral by the †circle method, introduced by Hardy and Ramanujan, which threw light on recent additive number theory. Hardy and Ramanujan thus obtained

$$p(n)=\frac{1}{2\pi\sqrt{2}}\frac{d}{dn}\left(\frac{\exp(C\lambda_n)}{\lambda_n}\right)+O(\exp(D\sqrt{n})),$$

where

$$\lambda_n=\sqrt{n-1/24}, \quad C=\pi\sqrt{2/3}, \quad D>C/2.$$

The theory was improved by Rademacher (1937, 1943), who expanded $p(n)$ into the

series

$$p(n)=\frac{1}{\pi\sqrt{2}}\sum_{k=1}^{\infty}A_k(n)k^{1/2}\frac{d}{dn}\left(\frac{\sinh(C\lambda_n/k)}{\lambda_n}\right),$$

where

$$A_k(n)=\sum_{h(\mathrm{mod}\,k),(h,k)=1}W_{h,k}\exp(-2\pi ihn/k).$$

Rademacher (1954) proved further that

$$p(n)=2\pi\left(\frac{\pi}{12}\right)^{3/2}$$
$$\times\sum_{k=1}^{\infty}A_k(n)k^{-5/2}L_{3/2}\left(\left(\frac{\pi}{12k}\right)^2(24n-1)\right),$$

where

$$L_\nu(z)=\sum_{n=0}^{\infty}\frac{z^n}{n!\Gamma(\nu+n+1)}.$$

Rademacher (1943) had developed an ingenious proof by taking "Ford's circle" as the contour Γ.

Ramanujan observed that $p(5m+4)\equiv 0$ (mod 5), $p(7m+5)\equiv 0$ (mod 7), and $p(11m+6)\equiv 0$ (mod 11). Rademacher (1942) and Newman studied these cases by using $\eta(\tau)$. More generally, A. O. L. Atkin proved that if $24n\equiv 1$ (mod $5^a7^b11^c$) then $p(n)\equiv 0$ (mod $5^a7^{1+[b/2]}11^c$) (*Glasgow Math. J.*, 8 (1967)). At present, this is the best result.

Let $n=n_1+n_2+\ldots+n_s$ be a partition of n. Many problems arise when we put additional conditions on the n_j. For instance, we may require that the n_j satisfy certain congruence relations (L. K. Hua (1942), S. Iseki (1959)) or are powers of integers (E. M. J. Wright (1934), L. Schoenfeld (1944), S. Iseki (1959)) or are powers of primes (T. Mitsui (1957)). The partition problem can also be extended to the case of an algebraic number field of finite degree (Rademacher, G. Meinardus (1953), Mitsui (1978)).

References

[1] R. G. Ayoub, An introduction to the analytic theory of numbers, Amer. Math. Soc. Math. Surveys, 1963.
[2] G. H. Hardy and S. Ramanujan, Asymptotic formulae in combinatory analysis, Proc. London Math. Soc., (2) 17 (1918), 75–115.
[3] G. H. Hardy and E. M. Wright, An introduction to the theory of numbers, Clarendon Press, fourth edition, 1965.
[4] K. Iseki, A proof of a transformation formula in the theory of partitions, J. Math. Soc. Japan, 4 (1952), 14–26.
[5] H. A. Rademacher, Lectures on analytic number theory, Tata Inst. Fund. Res., 1954–1955.
[6] S. Ramanujan, Collected papers of Srinivasa Ramanujan, Cambridge Univ. Press, 1927 (Chelsea, 1962).
[7] H. Petersson, Über Modulfunktionen und Partitionenprobleme, Abh. Deutsch. Akad. Wiss., Berlin Kl. Math. Allg. Nat., 1954, no. 2.
[8] A. G. Postnikov, Introduction to the analytical theory of numbers (in Russian), Moscow, 1971.
[9] M. I. Knopp, Modular functions in analytic number theory, Markham, 1970.
[10] T. Mitsui, On the partition problem in an algebraic number field, Tokyo J. Math., 1 (1978).
[11] H. A. Rademacher, Topics in analytic number theory, Springer, 1973.
[12] S. Iseki, The transformation formula for the Dedekind modular function and related functional equations, Duke Math. J., 24 (1957), 653–662.

329 (XXI.38)
Pascal, Blaise

Blaise Pascal (June 19, 1623–August 19, 1662) was born in Clermont-Ferrand in southern France. He lost his mother when still an infant and was brought up by his father, Etienne Pascal (discoverer of the curve called Pascal's †limaçon). As a youth, he demonstrated a remarkable ability for mathematics. In 1640, under the influence of Desargues, he discovered †Pascal's theorem on conic sections, and in 1642 invented an adding machine. After hearing of Toricelli's experiments in 1646, he became interested in the theory of fluids and on his own began to conduct experiments; this research put to rest the prevailing opinion that nature abhors a vacuum and that, therefore, a vacuum cannot exist. Pascal formulated the principle stating that pressure, when applied at any point within a contained liquid, is transmitted throughout the fluid. By means of this principle, he explained various phenomena concerning fluids such as the atmosphere and laid the foundations for hydrostatics.

Between 1652 and 1654, Pascal was preoccupied with social affairs, but subsequently he began to devote himself to religion. He entered the Abbey Port-Royal of the Jansenist sect, where he remained until his death. Immediately before his entry, however, he and Fermat exchanged correspondence about games of chance, and these letters proved to be the beginning of the theory of †probability. Concerning games of chance, Pascal had conducted research on †Pascal's triangle, and in this study he formulated and used †mathemat-

ical induction. He also indicated a way to obtain the sum of the mth powers of the consecutive terms of an arithmetic progression, and with an intuitive idea of limits obtained the formula $\int_0^a x^m dx = a^{m+1}/(m+1)$. While in Port-Royal, he published *Lettres provinciales* (1657), in which he carried on a dispute with the Jesuits. His book *Pensées* shows his deep involvement with religion; however, he did not abandon mathematics. In 1658, he determined the area enclosed by a †cycloid and its base, the barycenter and area of the figure enclosed by a cycloid and straight lines parallel to its base, and the volume of the figure obtained by rotating it around these lines. The study of the methods used by Pascal to obtain these results, which were forerunners of differential and integral calculus, led †Leibniz to discover the fundamental theorem of calculus. Pascal also formulated clear ideas about axioms.

References

[1] L. Brunschvicg (ed.), B. Pascal, Oeuvres I–XIV, Hachette, 1904–1914.
[2] J. Chevalier (ed.), B. Pascal, Oeuvres complètes, Bibliothèque de Pléiades, Gallimard, 1954.
[3] Kôkiti Hara, L'Oeuvre mathématique de Pascal, Mém. de la Faculté des Lettres de l'Univ. d'Osaka, no. 21, 1981.

330 (II.8) Permutations and Combinations

Let there be given a set Ω of n elements. If we choose k distinct elements of Ω and arrange them in a row, we have a **k-array** or **k-permutation** of elements of Ω. The number of such arrays is $(n)_k = n(n-1)\ldots(n-k+1)$. The polynomial $(x)_k = x(x-1)\ldots(x-k+1)$ in x of degree k is called the **Jordan factorial** of degree k. In particular, $(n)_n = n!$, **n factorial**, is the number of permutations of Ω. A subset of Ω is called a **k-subset** if it contains exactly k elements. The number of k-subsets (or **k-combinations**) of Ω is $\binom{n}{k} = (n)_k/k!$. The **binomial coefficients** $\binom{x}{n}$ are defined by the generating function $(1+z)^x = \sum_{n=0}^{\infty}\binom{x}{n} z^n$. For any complex number x, the series is convergent for $|z|<1$, and it is verified that $\binom{x}{n} = (x)_n/n!$ in terms of the Jordan factorial $(x)_n$. The same results hold in a †complete field with †valuation, in particular in a †p-adic number field. In any case, we have the recursive relation

$$\binom{x}{n} = \binom{x-1}{n} + \binom{x-1}{n-1} \qquad (n \geqslant 1),$$

$$\binom{x}{0} = 1,$$

and in general

$$\sum_{k=0}^{n}\binom{x}{k}\binom{y}{n-k} = \binom{x+y}{n},$$

which leads to many identities involving binomial coefficients. The recursive relation allows us to compute the values of $\binom{n}{k}$ easily for small integers n, k, as was noticed by Pascal. The arrangement of these values in a triangular form:

```
        1
      1   1
    1   2   1
  1   3   3   1
1   4   6   4   1
```

is called **Pascal's triangle**. For integral values x, $(1+z)^x$ are polynomials, and we have $(a+b)^n = \sum_{k=0}^{n}\binom{n}{k} a^{n-k} b^k$ (**binomial theorem**). As a generalization, we have

$$(a_1 + \ldots + a_m)^n = \sum \frac{n!}{p_1! \ldots p_m!} a_1^{p_1} \ldots a_m^{p_m}$$

(**multinomial theorem**), where the sum is extended over all nonnegative p_i with $\sum p_i = n$.

The number of ways of choosing k elements, allowing repetition, from a set of n elements is $(-1)^k\binom{-n}{k} = \binom{n+k-1}{\mathrm{k}}$. This is also the number of nonnegative integral solutions of $\sum_{i=1}^{n} x_i = k$. As an example of binomial coefficients with noninteger arguments, we have

$$\binom{-1/2}{n} = (-1)^n 2^{-2n}\binom{2n}{n}.$$

References

[1] E. Netto, Lehrbuch der Combinatorik, Teubner, second edition, 1927 (Chelsea, 1958).
[2] P. A. MacMahon, Combinatory analysis I, II, Cambridge Univ. Press, 1915–1916 (Chelsea, 1960).
[3] J. Riordan, Combinatorial identities, Wiley, 1968.

331 (XII.13) Perturbation of Linear Operators

A. General Remarks

Historically, the perturbation method was developed as an approximation device in classical and quantum mechanics. In the perturbation theory of eigenvalues and eigenfunctions, created by L. Rayleigh and E. Schrödinger, the main concern was to find solutions as power series in a parameter κ that could be regarded as small. In the perturbation theory for linear operators, however, we are concerned more generally with the behavior of spectral properties of linear operators when the operators undergo small change. The foundation of the mathematical theory, including a complete convergence proof of perturbation series, was laid down by F. Rellich [1] and T. Kato [2, 3]. Another major topic in the perturbation theory for linear operators is the perturbation of continuous spectra, which was initiated by K. O. Friedrichs (*Math. Ann.*, 115 (1938); [4]). It is closely related to scattering theory and is discussed more fully in 375 Scattering Theory. A standard reference in this field is [5] (also → [6, 7]). Most of the material presented in this article is taken from [5].

For problems in Hilbert spaces there are two general frameworks in which to formulate perturbation situations: the operator formulation and the form formulation. In the former we deal with a family of operators $T(\kappa)$ directly, while in the latter, we deal with associated semibounded Hermitian (or, more generally, sectorial) forms $\mathbf{t}(\kappa)$. The latter is applicable only when there is semiboundedness (or a sectorial property) inherent in the problem, but is usually more general than the former in such problems, since the latter (resp. the former) requires roughly the constancy of the domain of the "square root" of $T(\kappa)$ (resp. the domain of $T(\kappa)$). In this article we discuss problems in the operator formulation. For the form formulation → [5] and [7].

In this article X, Y, ..., are complex Banach spaces and T, A, ... are linear operators unless other specifications are made. The following notations defined in 251 Linear Operators are used without further explanation: $D(T)$, $R(T)$, $\mathbf{B}(X, Y)$, $\mathbf{B}(X)$, $\Gamma(T)$ (the graph of T), $\sigma(T)$ (the spectrum of T), $\rho(T)$ (the resolvent set of T), and $R(\zeta; T)$ (the resolvent of T). We also use $C(X, Y)$ (resp. $A(X, Y)$) to denote the set of all †closed linear operators (resp. all †linear operators) from X to Y and $C(X)=C(X,X)$.

B. Stability of Basic Properties

(1) Let $T\in C(X, Y)$. Important notions for characterizing the smallness of $A\in A(X,Z)$ relative to T are the following. (i) A is said to be **relatively bounded** with respect to T (or simply **T-bounded**) if $D(A)\supset D(T)$ and there exist a, $b\geqslant 0$ such that

(*) $\quad \|Au\|_Z\leqslant a\|u\|_X+b\|Tu\|_Y$ for all $u\in D(T)$.

The infimum, denoted by $\|A\|_T$, of b for which (*) holds with some a is called the T-bound of A. (ii) A is said to be **relatively compact** with respect to T (or **T-compact**) if $D(A)\supset D(T)$ and A is compact from $D(T)$ with the graph norm of T to Z (→ 68 Compact and Nuclear operators F). T-compactness of A implies T-boundedness (and in Hilbert spaces $\|A\|_T=0$).

(2) Let $T\in C(X, Y)$, and let $A\in A(X, Y)$ be T-bounded. (i) If $\|A\|_T<1$ (or if A is T-compact), then $T+A\in C(X, Y)$. (ii) If, in addition, $X=Y$ is a Hilbert space, T is †self-adjoint, and A is †symmetric, then $T+A$ is self-adjoint (**Rellich-Kato theorem**) [1, 8]. (iii) Suppose that T is a †Fredholm operator. If either A is T-compact or the inequality (*) holds with constants a, b satisfying $b\rho+a<\rho$ for a certain positive number ρ determined by T, then $T+A$ is a Fredholm operator and $\operatorname{ind}(T+A)=\operatorname{ind}T$ (for ind T, nul T, and def T → 251 Linear Operators). In the latter case where $b\rho+a<\rho$, we also have $\operatorname{nul}(T+A)\leqslant\operatorname{nul}T$ and $\operatorname{def}(T+A)\leqslant\operatorname{def}T$.

C. Continuity and Analyticity of Families of Closed Operators

In order to handle unbounded operators, which are important in applications, it is necessary to introduce generalized notions of convergence and analyticity of families of closed operators.

(1) $C(X, Y)$ becomes a †metric space by a distance function $\hat{d}(S, T)$ having the property that $\hat{\delta}(\Gamma(S), \Gamma(T))\leqslant\hat{d}(S, T)\leqslant 2\hat{\delta}(\Gamma(S), \Gamma(T))$, where for closed subspaces M and N we put

$$\hat{\delta}(M,N)=\max[\delta(M,N),\delta(N,M)],$$

$$\delta(M,N)=\sup_{u\in M,\|u\|=1}\operatorname{dist}(u,N);\ \delta(0,N)=0.$$

$\hat{\delta}(M,N)$ is called the **gap** between M and N [5]. When $T_n\rightarrow T$ in this metric, T_n is said to converge to T in the generalized sense. This **generalized convergence** coincides with the norm convergence if T_n, $T\in\mathbf{B}(X, Y)$. If $X=Y$ and $\rho(T)\neq\varnothing$, then $T_n\rightarrow T$ in the generalized sense if and only if for some (or equivalently all) $\zeta\in\rho(T)$ we have $\zeta\in\rho(T_n)$ for sufficiently large n and $\|R(\zeta;T_n)-R(\zeta;T)\|\rightarrow 0$, $n\rightarrow\infty$. This is called **norm resolvent convergence**.

(2) When $X=Y$, there is also the notion of strong convergence in the generalized sense [5], which is roughly the strong convergence of resolvents. In particular, when T_n and T are self-adjoint operators in a Hilbert space, $T_n \to T$ strongly in the generalized sense if $R(\zeta; T_n) \to R(\zeta; T)$ strongly for some (or equivalently all) ζ with $\operatorname{Im}\zeta \neq 0$. This is called **strong resolvent convergence**.

(3) Let $D \subset \mathbf{C}$ be a domain. The notion of analyticity (holomorphic property) of a family $T(\kappa) \in \mathbf{B}(X, Y)$, $\kappa \in D$, of bounded operators is well known (→ 37 Banach Spaces K). This notion is generalized to a family $T(\kappa) \in C(X, Y)$, $\kappa \in D$, of closed operators [1, 5]. Namely, $T(\kappa)$ is said to be **holomorphic** in D if at each $\kappa_0 \in D$ there exist a Banach space Z and $U(\kappa) \in \mathbf{B}(Z, X)$, $V(\kappa) \in \mathbf{B}(Z, Y)$, defined near κ_0, such that (i) $U(\kappa)$ and $V(\kappa)$ are holomorphic at κ_0 as families of bounded operators; (ii) $U(\kappa)$ is one to one and onto $D(T(\kappa))$; (iii) $T(\kappa)U(\kappa) = V(\kappa)$. Let us mention several special cases. (I) if $X=Y$ and if $\zeta \in \rho(T(\kappa))$ for all $\kappa \in D$, then $T(\kappa)$ is holomorphic in D if and only if $R(\zeta; T(\kappa))$ is holomorphic in D. (II) If $D(T(\kappa))$ is independent of κ and if $T(\kappa)u$ is holomorphic in D for every $u \in D(T(\kappa))$ then $T(\kappa)$ is holomorphic in D (holomorphic family of type (A) [5]). (III) Let $T \in C(X, Y)$, and let $T^{(n)} \in A(X, Y)$ such that $D(T^{(n)}) \supset D(T)$ and $\|T^{(n)}u\| \leqslant c^{n-1}(a\|u\| + b\|Tu\|)$, $u \in D(T)$, where a, b, $c \geqslant 0$. Then $T(\kappa)u = Tu + \kappa T^{(1)}u + \ldots + \kappa^n T^{(n)} + \ldots$, $u \in D(T)$ defines a holomorphic family of type (A) in $D = \{\kappa \mid |\kappa| < (b+c)^{-1}\}$. (IV) If $X=Y$ is a Hilbert space and if $T(\kappa)$ is self-adjoint for real κ, $T(\kappa)$ is said to be a self-adjoint family. In particular, the family discussed in (III) is a self-adjoint holomorphic family if T is self-adjoint and $T^{(n)}$ is symmetric.

D. Perturbation of Isolated Eigenvalues

(1) Separation of the spectrum. Let $T \in C(X)$. Suppose that a bounded subset Δ of $\sigma(T)$ is separated from the rest of $\sigma(T)$ by a simple closed contour Γ (i.e., $\Gamma \subset \rho(T)$ and $\Delta(\sigma(T) \setminus \Delta)$ lies inside (outside) of Γ). Then the operator

$$P = \frac{1}{2\pi i}\int_\Gamma R(\zeta; T)\,d\zeta,$$

which is independent of Γ, is a projection (i.e., $P \in \mathbf{B}(X)$ and $P^2 = P$). The closed subspaces $X_1 \equiv PX$ and $X_2 \equiv (I-P)X$ †reduce T and give rise to the decomposition $T = T|_{X_1} \oplus T|_{X_2} \equiv T_1 \oplus T_2$. In particular, $\sigma(T_1) = \sigma(T) \cap \{\text{inside of } \Gamma\}$ and $\sigma(T_2) = \sigma(T) \cap \{\text{outside of } \Gamma\}$.

(2) Let $T(\kappa)$ be holomorphic in D. We assume that $0 \in D$ and regard $T^{(0)} = T(0)$ as the unperturbed operator. Suppose that Δ and Γ are as in (1) with T replaced by $T^{(0)}$. Then there exists $\delta > 0$ such that $|\kappa| < \delta$ implies $\Gamma \subset \rho(T(\kappa))$. This follows from the upper continuity of compact components of the spectrum with respect to the metric $\hat{d}$ of $C(X)$ [5]. Thus the separation of the spectrum discussed in (1) is applicable to $T(\kappa)$. In particular, corresponding to the projection

$$P(\kappa) = \frac{1}{2\pi i}\int_\Gamma R(\zeta; T(\kappa))\,d\zeta, \qquad |\kappa| < \delta,$$

$T(\kappa)$ is decomposed as $T(\kappa) = T_1(\kappa) \oplus T_2(\kappa)$; and the problem of determining the spectrum of $T(\kappa)$ inside Γ is reduced to the problem of determining the spectrum of $T_1(\kappa)$ ($|\kappa| < \delta$). Suppose now that $\Delta = \{\lambda_0\}$ is an isolated eigenvalue of $T^{(0)}$ and that $m \equiv \dim P(0)X < \infty$. Then $\dim P(\kappa)X = m$, $|\kappa| < \delta$. Moreover, a base $\{\varphi_1(\kappa), \ldots, \varphi_m(\kappa)\}$ of $P(\kappa)X$ can be constructed in such a way that the $\varphi_j(\kappa)$ are holomorphic in $\{|\kappa| < \delta' \leqslant \delta\}$ [3, 5]. Thus the problem for $T_1(\kappa)$ in this case is just the finite-dimensional eigenvalue problem $\det\{\lambda\delta_{jk} - (T(\kappa)\varphi_j(\kappa), \varphi_k(\kappa))\} = 0$. The totality $\{\lambda_j(\kappa)\}$ of solutions of this equation, i.e., the totality of eigenvalues of $T(\kappa)$ near λ_0, is expressed by one or several power series of $\kappa^{1/p}$ with a suitable integer $p > 0$. If $T(\kappa)$ is a self-adjoint family, we can take $p = 1$ so that the eigenvalues are holomorphic near λ_0. If $H(\kappa) = H^{(0)} + \kappa H^{(1)} + \ldots$ is a self-adjoint holomorphic family described in example (IV) in (3) of Section C and if $m = 1$, the power series $\lambda(\kappa) = \sum \lambda_j \kappa^j$ can be explicitly computed as $\lambda_1 = (H^{(1)}u_0, u_0)$, $\lambda_2 = (H^{(2)}u_0, u_0) + (SH^{(1)}u_0, H^{(1)}u_0)$, ..., where $H^{(0)}u_0 = \lambda_0 u_0$ with $\|u_0\| = 1$ and where $S = \lim_{\zeta \to \lambda_0} R(\zeta; H^{(0)})(I - P(0))$ is the reduced resolvent. This series is known as the **Rayleigh-Schrödinger series**. The power series for the associated eigenvectors $u(\kappa) = \sum \kappa^j u_j$ can also be computed. For details, including the case of a degenerate λ_0 ($m > 1$), in which the situation becomes more complicated due to the splitting of eigenvalues, → [5]. The perturbation theory discussed in this subsection is called **analytic** (or **regular**) **perturbation theory**.

(3) Even when a problem cannot be handled by means of analytic perturbation, it may happen that the coefficients λ_j and u_j of formal power series can be computed up to a certain j. In many such cases it can be shown under general assumptions that an asymptotic expansion such as $\lambda(\kappa) = \lambda_0 + \lambda_1\kappa + o(\kappa)$ is valid as long as the coefficients involved can be computed legitimately [2, 5]. Estimates for $o(\kappa)$ can also be given. This provides a rigorous foundation for the perturbation method in many important practical problems. The case of degenerate λ_0 can be treated similarly. The strong convergence in the generalized sense mentioned in (2) of Section C is used here. This theory is called **asymptotic perturbation theory**.

E. Perturbation of Continuous Spectra

For continuous spectra, studying the mode of change under perturbation is not usually a tractable problem. Rather, certain parts of continuous spectra tend to be stable under perturbation; and the study of this stability has been a major topic in perturbation theory (also → 375 Scattering Theory). In this section we discuss only self-adjoint operators and let $H=\int\lambda dE(\lambda)$, $H_0, \ldots,$ be self-adjoint operators in a Hilbert space X. For $\mathbf{B}_p$ and $\| \ \|_p$ to be used below → 68 Compact and Nuclear Operators.

(1) The essential spectrum (→ 390 Spectral Analysis of Operators E) is stable under compact perturbation. Namely, if $H=H_0+K$ with compact K, then $\sigma_e(H)=\sigma_e(H_0)$ (H. Weyl, *Rend. Circ. Mat. Palermo*, 27 (1909)). More generally, it suffices to assume that $R(\zeta;H)-R(\zeta;H_0)$ is compact for some (or equivalently all) $\zeta\in\rho(H)\cap\rho(H_0)$. Conversely, if X is separable and if $\sigma_e(H)=\sigma_e(H_0)$, then there exist a unitary operator U and a compact operator K such that $H=UH_0U^{-1}+K$ (J. von Neumann, *Actualités Sci. Ind.*, 229 (1935)). Moreover, any self-adjoint operator H in a separable Hilbert space can be changed into $H+K$ with a pure point spectrum by adding a $K\in\mathbf{B}_p$ with $\|K\|_p<\varepsilon$ for any $p>1$ and $\varepsilon>0$ (S. T. Kuroda, *Proc. Japan Acad.*, 34 (1958)). I. D. Berg (1971), W. Sikonia (1971), J. Voigt (1977), and D. Voiculescu (1979) have extended these results to normal operators and m-tuples of commutative self-adjoint operators. Also → 390 Spectral Analysis of Operators I, J.

(2) The absolutely continuous spectrum (→ 390 Spectral Analysis of Operators E) is stable under perturbation by the †trace class. Namely, if $H=H_0+K$, with $K\in\mathbf{B}_1$, then the absolutely continuous parts of H_0 and H are †unitarily equivalent, and in particular $\sigma_{ac}(H)=\sigma_{ac}(H_0)$ (M. Rosenblum, *Pacific J. Math.*, 7 (1957); T. Kato, *Proc. Japan Acad.*, 33 (1957)). Among generalizations we mention the following two. (i) If $R(\zeta;H)-R(\zeta;H_0)\in\mathbf{B}_1$ for some $\zeta\in\rho(H)\cap\rho(H_0)$, then the absolutely continuous parts of $\varphi(H_0)$ and $\varphi(H)$ are unitarily equivalent for any smooth strictly increasing real function φ (M. Sh. Birman, *Izv. Akad. Nauk SSSR*, ser. mat., 27, (1963); T. Kato, *Pacific J. Math.*, 15 (1965)). (ii) If H_0 and H act in different Hilbert spaces X_0 and X, respectively, and if there exists $J\in\mathbf{B}(X_0,X)$ such that $JD(H_0)\subset D(H)$ and such that the closure of $HJ-JH_0$ belongs to $\mathbf{B}_1(X_0,X)$, then the same conclusion as in (i) holds (D. Pearson, *J. Functional Anal.*, 28 (1978)). (i) can be derived from (ii). Perturbation theory for absolutely continuous spectra is closely related to the study of generalized wave operators in scattering theory. In fact, the existence and the completeness of the latter implies the unitary equivalence of absolutely continuous parts. All the results mentioned above are proved by scattering-theoretic methods, either by the wave operator approach or by the abstract stationary approach (→ 375 Scattering Theory, esp. B, C).

F. Some Other Topics

(1) For the perturbation theory for semigroups of operators and evolution equations, not discussed in this article, → [5, 7, 9].

(2) The detailed structure of continuous spectra is hard to analyze. An eigenvalue λ_0 of H_0 which is embedded in the continuous spectrum may diffuse into the continuous spectrum in the presence of a perturbation. In such a case, $H(\kappa)$, $\kappa\neq 0$, has no eigenvalues near λ_0 but may have a continuous spectrum highly concentrated around λ_0. This phenomenon of **spectral concentration** is studied, especially for some concrete problems, in relation to **resonance poles** (or poles of the holomorphic continuation of the resolvent or the scattering matrix). In some problems, it is proved that the first few terms of the perturbation series for $\lambda(\kappa)$ that are still computable are related to the real part of the resonance. Some problems of resonance can be treated by the technique of dilation analyticity, a technique which is also effective in other problems of spectral analysis (J. Aguilar and J. M. Combes, *Comm. Math. Phys.*, 22 (1971)).

(3) A vast quantity of results in the spectral theory of the Schrödinger operators appearing in the †Schrödinger equation in quantum mechanics can be obtained by perturbation methods.

For the topics mentioned in (2) and (3) → [7].

References

[1] F. Rellich, Störungstheorie der Spektralzerlegung I–V, Math. Ann., 113 (1937), 600–619; 113 (1937), 677–685; 116 (1939), 555–570; 117 (1940), 356–382; 118 (1942), 462–484.
[2] T. Kato, On the convergence of the perturbation method, J. Fac. Sci. Univ. Tokyo, sec. I, 6 (1951), 145–226.
[3] T. Kato, On the perturbation theory of closed linear operators, J. Math. Soc. Japan, 4 (1952), 323–337.
[4] K. O. Friedrichs, On the perturbation of continuous spectra, Comm. Pure Appl. Math., 1 (1948), 361–406.
[5] T. Kato, Perturbation theory for linear operators, Springer, 1966.

[6] N. Dunford and J. T. Schwartz, Linear operators I–III, Wiley-Interscience, 1958, 1963, 1971.
[7] M. Reed and B. Simon, Methods of modern mathematical physics I–IV, Academic Press, 1972, 1975, 1979, 1978.
[8] T. Kato, Fundamental properties of Hamiltonian operators of Schrödinger type, Trans. Amer. Math. Soc., 70 (1951), 195–211.
[9] H. Tanabe, Equation of evolution, Pitman, 1979. (Original in Japanese, 1975.)

332 (VI.7) Pi (π)

The **ratio of the circumference of a circle to its diameter** in a Euclidean plane is denoted by π, the initial letter of $\pi\varepsilon\rho\iota\mu\varepsilon\tau\rho o\varsigma$ (perimeter). Thus π can be defined as

$$2\int_0^1 dx/\sqrt{1-x^2}.$$

The symbol π has been used since W. Jones (1675–1749) and L. Euler. The fact that this ratio is a constant is stated in Euclid's *Elements*; however, Euclid gave no statement about the numerical value of π. As an approximate value of π, 3 has been used from antiquity. According to the Rhind Papyrus, $(4/3)^4$ was used in ancient Egypt. Let $L_n(l_n)$ be the perimeter of a regular n-gon circumscribed about (inscribed in) a circle of radius 1. Then the relations

$$L_n>\pi>l_n, \qquad \frac{2}{L_{2n}}=\frac{1}{L_n}+\frac{1}{l_n},$$

$$l_{2n}=\sqrt{l_nL_{2n}}$$

hold. Archimedes obtained $3\frac{10}{71}<\pi<3\frac{1}{7}$ by calculating L_{96} and l_{96}. In 3rd-century China Liu Hui used $\pi\fallingdotseq 3.14$. In 5th-century China, Tsu Chung-Chih mentioned 22/7 as an inaccurate approximate value and 355/113 as an accurate approximate value of π. These values were obtained by methods similar to those of Archimedes. In 5th-century India, Aryabhatta obtained $\pi\fallingdotseq 3.1416$, and in 16th-century Europe, Adriaen van Roomen obtained $\pi\fallingdotseq 355/113$.

F. Viète represented $2/\pi$ in the following infinite product:

$$\prod_{n=2}^{\infty}\cos\frac{\pi}{2^n}=\sqrt{\frac{1}{2}}\sqrt{\frac{1}{2}+\frac{1}{2}\sqrt{\frac{1}{2}}}\times\sqrt{\frac{1}{2}+\frac{1}{2}\sqrt{\frac{1}{2}+\frac{1}{2}\sqrt{\frac{1}{2}}}}\cdots.$$

Using this formula, L. van Ceulen (1540–1610) calculated π to 35 decimals. In the 17th and 18th centuries, the Japanese mathematicians T. Seki, K. Takebe, and Y. Matunaga computed π to 50 decimals. Since the 17th century, many formulas that represent π as a sum of infinite series or as a limit have been used to obtain more accurate approximate values. The following are representations of π known in those days:

$$\frac{\pi}{2}=\frac{2\cdot 2\cdot 4\cdot 4\cdot 6\cdot 6\ldots}{1\cdot 3\cdot 3\cdot 5\cdot 5\cdot 7\ldots} \qquad \text{(J. Wallis)}$$

$$\frac{\pi}{4}=\frac{1}{1+}\,\frac{1^2}{2+}\,\frac{3^2}{2+}\,\frac{5^2}{2+}\ldots$$ (W. Brouncker; for the notation → 83 Continued Fractions)

$$=1-1/3+1/5-1/7+\ldots$$ (J. Gregory, G. W. F. Leibniz)

$$=4\,\mathrm{Arctan}\,1/5-\mathrm{Arctan}\,1/239$$ (J. Machin).

A formula combining Machin's representation of π and the power series $\mathrm{Arctan}\,x=x-(1/3)x^3+(1/5)x^5-\ldots$ is called **Machin's formula** and was often used for calculating an approximate value of π. By utilizing this formula, in 1873 W. Shanks obtained an approximate value of π up to 707 decimals. No improvement of his approximation was obtained until 1946 when D. F. Ferguson calculated 710 digits of π and found that Shanks's value was correct only up to the 527th digit. The computation of an accurate approximate value of π has been made easier by the recent development of computing machines, and an approximate value up to 1,000,000 decimals has been obtained. P. Beckmann [2] gives a detailed and humorous historical account of the calculation of π from ancient times up to the present computer age. Various numerical results obtained by electronic computers are not formally published, some being deposited in the UMT repository of the editorial office of the journal *Mathematics of computation*. Choong et al. [3], using information in [1], obtained the first 21,230 partial denominators of the regular continued fraction representation of π and described how their numerical evidence tallies with theoretical results, obtained by the metrical theory of continued fractions, which is valid for almost all irrational numbers (e.g. → [4]).

In 1761, J. H. Lambert used Brouncker's expression of π in a continued fraction to prove that π is irrational. In 1882, C. L. F. Lindemann proved that π is a †transcendental number using Euler's formula $e^{\pi i}=-1$. The approximate value of π up to 50 decimals is 3.14159265358979323846264338327950288419716939937510... (→ Appendix B, Table 6).

References

[1] D. Shanks and J. W. Wrench, Calculation of π to 100,000 decimals, Math. Comp., 16 (1962), 76–99.
[2] P. Beckmann, A history of pi, second edition, Golem Press, 1971.
[3] K. Y. Choong, D. E. Daykin, and C. R. Rathbone, Rational approximations to π, Math. Comp., 25 (1971), 387–392.
[4] A. Khinchin, Continued fractions, Noordhoff, 1963.

333 (II.18) Plane Domains

A. Domains in the Complex Plane

A †domain (i.e., a †connected open set) in the †complex plane or on the †complex sphere is called a **plane domain**. The †closure of such a domain is called a **closed plane domain**. In this article, we consider only subsets of the complex plane (or sphere), and a plain domain is called simply a domain. The †interior of a †Jordan curve J in the complex plane is a domain called a **Jordan domain**. In a domain D, a Jordan arc whose two endpoints lie on the boundary of D is called a **cross cut** of D.

For a domain D, each of the following three conditions is equivalent to the condition that D is †simply connected: (1) For every cross cut Q of D, $D-Q$ has exactly two †connected components. (2) Every Jordan curve in D is †homotopic to one point, that is, it can always be continuously deformed to a point. (3) The †monodromy theorem holds in D.

If D is a domain on a complex sphere, each of the following three conditions is equivalent to the condition that D is simply connected: (4) The boundary of D consists of a single †continuum or a single point. (5) For every Jordan curve C in D, either the interior or the exterior of C is contained in D. (6) The complement of D with respect to the complex sphere is a connected (not necessarily arcwise connected) closed set. Jordan domains are simply connected.

Let $n \geqslant 2$ be an integer and D a plane domain. The †homology group $H_1(D, \mathbf{Z})$ is identical to $\mathbf{Z}^{n-1}$ if and only if the complement of D in the complex sphere has n connected components. Then D is said to be **n-ply connected** or **multiply connected** without specifying n. If D is an n-ply connected domain, there exist $n-1$ suitable mutually disjoint cross cuts $Q_1, \ldots, Q_{n-1}$ such that $D-(Q_1 \cup \ldots \cup Q_{n-1})$ is simply connected.

Some typical examples of domains are as follows: (1) **Circular domain**: $|z-c|<r$. (2) **Half-plane**: $\operatorname{Re} z>0$, or $\operatorname{Im} z>0$. (3) **Angular domain**: $\alpha<\arg(z-c)<\beta$. (4) **Annular domain**: $r<|z-c|<R$. (5) **Slit domain**: a domain obtained by excluding a Jordan arc Γ from a domain D, where all points on Γ (except an endpoint lying on the boundary of D) are contained in D. In this case, the Jordan arc Γ is called the **slit** of the domain.

B. Boundary Elements

A boundary point P of a domain D is called **accessible** if there exists a sequence of points P_ν tending to P such that the line segments $P_1 P_2, \ldots$ lie completely in D. For example, for the domain obtained by removing $x=1/(n+1)$, $0<y\leqslant 1/2$ $(n=1,2,\ldots)$ from the square $0<x<1$, $0<y<1$, the boundary points with $x=0$, $0\leqslant y<1/2$ are all inaccessible (Fig. 1).

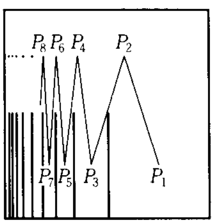


Fig. 1

Let the domain D be bounded by a smooth Jordan curve, and let P be a boundary point of D. Take an angular domain D' with vertex at the point P and the initial parts of the two sides of D' lying in D. A curve in D converging to the point P from the interior of the angular domain D' is called a **Stolz's path** or a **nontangential path** ending at the point P.

Let D be a simply connected domain. A sequence $\{q_\nu\}$ of cross cuts mutually disjoint except for their endpoints is called a **fundamental sequence of cross cuts** if it satisfies the following two conditions (Fig. 2): (1) Every q_ν separates $q_{\nu-1}$ and $q_{\nu+1}$ on D. (2) For $\nu\to\infty$, the sequence q_ν tends to a point on the boundary. Let $\{q_\nu\}$ be a fundamental sequence of cross cuts, and denote by D_ν the subdomain of D separated by q_ν that contains $q_{\nu+1}$. The intersection $\bigcap \bar{D}_\nu$ consists only of the boundary points of D. Two fundamental sequences $\{q_\nu\}$, $\{q'_\nu\}$ of cross cuts are equivalent if every D_μ contains all q'_ν except for a finite number of ν, and every D'_μ contains all q_ν except for a finite number of ν. Here D_μ, D'_μ are the subdomains constructed from q_μ and q'_μ as above. This condition determines an equivalence relation, under which the equivalence class of fundamental sequences of cross cuts is called a **boundary element**. This notion is due to C. Carathéodory [2]. The boundary element of a

multiply connected domain is defined similarly for each isolated component of the boundary. For example, each point of a slit domain, except for the endpoint of Γ lying on the boundary of the domain, determines two distinct boundary elements on each side. A closed domain is usually considered to be the union of a domain and the set of all its boundary elements. Various notions of †ideal boundary come from considering suitable boundary elements for various purposes (→ 207 Ideal Boundaries).

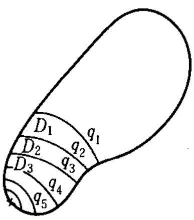


Fig. 2

C. Domain Kernels

Let $\{G_\nu\}$ be a sequence of domains containing the origin 0. If a suitable neighborhood of the origin is contained in G_ν for all ν, there exists a domain G such that every closed domain containing the origin and contained in G is contained in G_ν except for a finite number of ν. The union K of such domains G is called the **domain kernel** of the sequence $\{G_\nu\}$ (Carathéodory). If there is no neighborhood of the origin contained in G_ν for all ν, we put $K=\{0\}$.

If every infinite subsequence of $\{G_\nu\}$ has the same domain kernel K, then we say that the sequence $\{G_\nu\}$ converges to K. The notion of domain kernel is important in considering the limits of a sequence of conformal mappings (→ 77 Conformal Mappings).

References

[1] M. H. A. Newman, Elements of the topology of plane sets of points, Cambridge Univ. Press, second edition, 1951.

[2] C. Carathéodory, Über die Begrenzung einfach zusammenhängender Gebiete, Math. Ann., 73 (1913), 323–370.

334 (X.33) Plateau's Problem

A. Origin

Because of surface tension, a soap membrane bounded by a given closed space curve takes the shape of a minimal surface, i.e., a surface of the least area. This experiment was performed by the Belgian scientist J. A. Plateau (1873) to realize minimal surfaces; hence **Plateau's problem** is that of determining the minimal surfaces bounded by given closed space curves. It is a problem of the †calculus of variations.

B. Formulation

Let Γ be a †simple closed curve in xyz-space such that its projection C on the xy-plane is also a simple closed curve. Let D be the finite domain bounded by C. We consider surfaces $z=z(x,y)$ having common boundary Γ. Then under suitable assumptions on the smoothness of $z(x,y)$, the problem is to minimize the †functional

$$J[z]=\iint_D \sqrt{1+p^2+q^2}\,dx\,dy;$$

$$p=\frac{\partial z}{\partial x},\qquad q=\frac{\partial z}{\partial y},$$

with the condition that $z=z(x,y)$ has Γ as its boundary. The †Euler-Lagrange differential equation for the functional $J[z]$ is

$$\frac{\partial}{\partial x}\frac{p}{\sqrt{1+p^2+q^2}}+\frac{\partial}{\partial y}\frac{q}{\sqrt{1+p^2+q^2}}=0,$$

or $(1+q^2)r-2pqs+(1+p^2)t=0$, $r=\partial^2 z/\partial x^2$, $s=\partial^2 z/\partial x\partial y$, $t=\partial^2 z/\partial y^2$, which is a second-order †quasilinear partial differential equation of elliptic type and whose geometric interpretation had already been given by M. C. Meusnier (1776).

To formulate the problem more generally, let a surface be expressed in vector form $\mathfrak{x}=\mathfrak{x}(u,v)$ by means of parameters u, v. Let its †first fundamental form be $d\mathfrak{x}^2=E\,du^2+2F\,du\,dv+G\,dv^2$ and its †second fundamental form be $-d\mathfrak{x}\,d\mathfrak{n}=L\,du^2+2M\,du\,dv+N\,dv^2$, with $\mathfrak{n}=\mathfrak{n}(u,v)$ the unit normal vector. By equating to zero the †first variation of the **areal functional**

$$\iint \sqrt{EG-F^2}\,du\,dv$$

based on infinitesimal displacement in the normal direction, we obtain the Euler-Lagrange equation in the form

$$2H\equiv(NE-2MF+LG)/(EG-F^2)=0,$$

where $H=(R_1^{-1}+R_2^{-1})/2$ is the †mean curvature of the surface and R_1 and R_2 are the †radii of principal curvature. Since †Beltrami's second differential form satisfies $\Delta_2\mathfrak{x}=H\mathfrak{n}$, the condition for a minimal surface becomes $\Delta\mathfrak{x}=0$ (with Δ the †Laplace operator) provided that **isothermal parameters** u, v satisfying $E=$

$G, F=0$ are chosen, i.e., the vector $\mathfrak{x}(u, v)$ representing a minimal surface is harmonic (the components of this vector are †harmonic functions of u, v). Let $\mathfrak{y}(u, v)$ be a harmonic vector conjugate to $\mathfrak{x}(u, v)$. Then isothermality is expressed by the condition that the analytic vector $\mathfrak{F}(w)=\mathfrak{x}(u, v)+i\mathfrak{y}(u, v)$ $(w=u+iv, i=\sqrt{-1})$ satisfies $\mathfrak{F}'(w)^2=0$ (Weierstrass). In general, a **minimal surface** is defined as a surface with everywhere vanishing mean curvature, and Plateau's problem is to determine the minimal surface with a preassigned boundary. In this formulation, the problem can be easily generalized to an n-dimensional Euclidean space $\mathbf{R}^n$ (→ 275 Minimal Submanifolds).

C. Existence of a Solution

The existence of a solution of Plateau's problem was discussed by S. N. Bernshteĭn (1910) from the viewpoint of a †boundary value problem of the first kind for the elliptic partial differential equation in the previous section. A. Haar (1927) dealt with the minimal problem for the functional $J[z]$ by a †direct method in the calculus of variations. Previously, Riemann, Weierstrass, Schwarz, and others discussed the case where the given space curve Γ is a polygon, in connection with the †monodromy group concerning a second-order linear ordinary differential equation. Subsequently, R. Garnier (1928) investigated the existence of a solution by the limit process when Γ is a simple closed curve with bounded curvature. However, when Γ is assumed merely to be †rectifiable, the existence of a solution was first shown by the limit process by T. Radó (1930). He further discussed the general case where Γ can bound a surface with finite area. On the other hand, by introducing a new functional depending on boundary values instead of the areal functional, J. Douglas (1931) succeeded in giving a satisfactory result for the existence problem. R. Courant (1937) gave another existence proof by reducing Plateau's problem to the †Dirichlet principle [3].

At present, the methods of discussing the existence of solutions of Plateau's problem can be classified into the following three sorts (represented, respectively, by Radó, Courant, and Douglas):

(1) The first method is to minimize directly the areal functional $\iint\sqrt{EG-F^2}\,du\,dv$. The variational equation of the areal functional becomes $H=0$.

(2) **Dirichlet's functional** for a scalar function $f(u,v)$ is defined by $D[f]=\iint(f_u^2+f_v^2)\,du\,dv$, and for an n-dimensional vector function $\mathfrak{f}(u,v)$ with components $f_j(u,v)$ $(j=1,\ldots,n)$ by $D[\mathfrak{f}]=\sum_{j=1}^n D[f_j]$. The existence of a solution of Plateau's problem can be discussed by starting from the variational problem of minimizing $D[\mathfrak{f}]$. The variational equation of $D[\mathfrak{f}]$ is $\Delta\mathfrak{f}=0$.

(3) An analytic vector $\mathfrak{F}(w)$ is representable in terms of the boundary values of its real part. For instance, if the domain of w is the unit disk $|w|<1$, then Poisson's integral formula

$$\mathfrak{F}(w)=\frac{1}{2\pi}\int_0^{2\pi}\mathfrak{h}(\theta)\frac{e^{i\theta}+w}{e^{i\theta}-w}d\theta+i\,\mathrm{Im}\,\mathfrak{F}(w)$$

with the boundary function $\mathfrak{h}(\theta)=\mathrm{Re}\,\mathfrak{F}(e^{i\theta})$ can be used. On the other hand, the vector function that minimizes the Dirichlet integral among functions with fixed boundary values is harmonic. Based on these facts, Douglas transformed Dirichlet's functional with harmonic argument functions into a functional whose arguments are boundary functions. Specifically, by starting from the problem of minimizing **Douglas's functional**

$$A[\mathfrak{h}]=\frac{1}{4\pi}\int_0^{2\pi}\int_0^{2\pi}\frac{(\mathfrak{h}(\theta)-\mathfrak{h}(\varphi))^2}{4\sin^2(\theta-\varphi)/2}d\theta\,d\varphi,$$

we can prove the existence of solution of Plateau's problem satisfactorily.

D. The Generalized Case

Up to now we have been concerned with Plateau's problem in the case of a single simple closed curve. Douglas, Courant, and others treated the generalized case of a finite number of boundary curves, where †genus and orientability are assigned as the topological structure of the surface to be found. The existence of a solution has been shown in this case also. The problem is further generalized from the case of fixed boundary to the case where the boundary is merely restricted to lie on a given manifold [3]. On the other hand, C. B. Morrey (1948) generalized the problem by replacing the ambient space $\mathbf{R}^n$ by an n-dimensional †Riemannian manifold and gave the existence proof in considerable generality [6].

E. Relation to Conformal Mappings

There is a notable relation between Plateau's problem and conformal mapping when the dimension of the space is 2. Namely, the existence proof of the solution of the former for a Jordan domain implies †Riemann's mapping theorem together with W. F. Osgood and C. Carathéodory's result on boundary correspondence (→ 77 Conformal Mappings).

F. New Developments

Among recent contributions to the study of Plateau's problem, the following remarkable

results have emerged. One of them is connected with the final result of Douglas (1939) on the existence of solution surfaces. The mapping of a 2-dimensional manifold with boundary into $\mathbf{R}^n$ defining Douglas's solution of the Plateau problem for a finite number of simple closed curves is a †minimal immersion with the possible exception of isolated points where it fails to be an immersion. These points are called branch points. It was then proved by R. Osserman (1970) and R. D. Gulliver (1973) that for $n=3$ the mapping of Douglas's theorem, which is a surface of least area, is free of branch points, i.e., is an immersion. Osserman also gave examples of generalized minimal surfaces in $\mathbf{R}^n$ $(n>3)$ with true branch points. In this connection, Gulliver also dealt with an analogous problem for surfaces of prescribed mean curvature.

Next, we mention the question of boundary regularity. H. Lewy (1951) proved that if the boundary of a minimal surface is analytic, then the surface is analytic up to the boundary. Subsequently, S. Hildebrandt (1969) and others proved that if the boundary is of class $C^{m,\alpha}$, $m \geqslant 1$, the surface is also of class $C^{m,\alpha}$ up to the boundary. There are also some recent results on the number of solutions of Plateau's problem. For instance, J. C. C. Nitsche (1973) proved the uniqueness of solutions for analytic boundaries of †total curvature at most 4π.

Further developments in connection with Plateau's problem have emerged in the work of E. R. Reifenberg (1960) and others, who sought to minimize the †Hausdorff measure among general classes of geometric objects, not as parametrized manifolds, but as subsets of $\mathbf{R}^n$. The existence and regularity of solutions of Plateau's problem from this point of view have been discussed by H. Federer (1969), W. H. Fleming, F. J. Almgren, and others [10].

References

[1] T. Radó, On the problem of Plateau, Erg. Math., Springer, 1933.
[2] J. Douglas, Solution of the problem of Plateau, Trans. Amer. Math. Soc., 33 (1931), 263–321.
[3] R. Courant, Dirichlet's principle, conformal mapping and minimal surfaces, Interscience, 1950.
[4] J. C. C. Nitsche, Vorlesungen über Minimalflächen, Springer, 1975.
[5] R. Osserman, A survey of minimal surfaces, Van Nostrand, 1969.
[6] C. B. Morrey, Multiple integrals in the calculus of variations, Springer, 1966.
[7] R. Osserman, A proof of the regularity everywhere of the classical solution to Plateau's problem, Ann. Math., (2) 91 (1970), 550–569.
[8] S. Hildebrandt, Boundary behavior of minimal surfaces, Arch. Rational Mech. Anal., 35 (1969), 47–82.
[9] J. C. C. Nitsche, A new uniqueness theorem for minimal surfaces, Arch. Rational Mech. Anal., 52 (1973), 319–329.
[10] H. Federer, Geometric measure theory, Springer, 1969.
[11] F. J. Almgren, Plateau's problem. An invitation to varifold geometry, Benjamin, 1966.

335 (XXI.39) Poincaré, Henri

Henri Poincaré (April 29, 1854–July 17, 1912) was born in Nancy, France. After graduating from the Ecole Polytechnique, he taught at the University of Caen in 1879, then at the University of Paris in 1881. He was made a member of the Académie des Sciences in 1887 and of the Académie Française in 1908. He died in Paris.

His achievements center on analysis and applications to theoretical physics and astronomy. However, his work covered many fields of mathematics, including arithmetic, algebraic geometry, spectral theory, and topology. His †uniformization of analytic functions by means of the theory of †automorphic functions in 1880 is especially notable. His paper on the †three-body problem won the prize offered by the king of Sweden in 1889.

The methods he developed in his three-volume *Mécanique céleste* (1892–1899) began a new epoch in celestial mechanics. In addition, Poincaré opened the road to †algebraic topology and made suggestive contributions to the †theory of relativity and †quantum theory. He asserted that science is for science's own sake [4], and his popular philosophical works concerning the foundations of natural science and mathematics exhibit a lucid style.

References

[1] H. Poincaré, Oeuvres I–XI, Gauthier-Villars, 1916–1956.
[2] H. Poincaré, Les méthodes nouvelles de la mécanique céleste 1–III, Gauthier-Villars, 1892–1899.
[3] H. Poincaré, Science et hypothèse, Flammarion, 1903.
[4] H. Poincaré, La valeur de la science, Flammarion, 1914.

[5] H. Poincaré, Science et méthode, Flammarion, 1908.

336 (X.20) Polynomial Approximation

A. General Remarks

On the existence of polynomial approximations, we have **Weierstrass's approximation theorem**, which is formulated in the following two forms: (i) If $f(x)$ is a function that is continuous in the finite interval $[a,b]$, then for every $\varepsilon>0$ there exists a polynomial $P_n(x)$ of degree $n=n(\varepsilon)$ such that the inequality $|f(x)-P_n(x)|\leqslant\varepsilon$ holds throughout the interval $[a,b]$. (ii) If $f(\theta)$ is a continuous function of period 2π, then corresponding to every positive number ε there exists a trigonometric polynomial of degree $n=n(\varepsilon)$,

$$P_n(\theta)=a_0+\sum_{k=1}^{n}(a_k\cos k\theta+b_k\sin k\theta), \tag{1}$$

such that the inequality $|f(\theta)-P_n(\theta)|\leqslant\varepsilon$ holds for all values of θ. The second form of Weierstrass's theorem follows from the first, and conversely. M. H. Stone obtained a theorem that generalizes Weierstrass's theorem to the case of functions of several variables. Of the many direct proofs now available for Weierstrass's theorem, we mention two simple ones. To prove version (i) of the theorem, we can assume that the given function $f(x)$ is defined in the segment $[0,1]$. Consider the **Bernshteĭn polynomial**

$$B_n(x)=\sum_{k=0}^{n}f\left(\frac{k}{n}\right)\binom{n}{k}x^k(1-x)^{n-k}.$$

Then $B_n(x)$ converges to $f(x)$ †uniformly. To prove (ii), we can apply †Fejér's theorem on †Fourier series. We have the following generalization of (i): Let $p_1, p_2, \ldots$ be a sequence of positive numbers such that $\lim p_n=\infty$. Then linear combinations of $x^0=1, x^{p_1}, x^{p_2}, \ldots$ can uniformly approximate each continuous function on $[0,1]$ with arbitrary precision if and only if $\Sigma p_n^{-1}=-\infty$ (**Müntz's theorem**).

B. Best Approximations

Let $\varphi_0(x), \varphi_1(x), \ldots$ be a sequence of linearly independent continuous functions on a bounded closed domain A in $\mathbf{R}^n$. For any given continuous function $f(x)$, a function $P_n(x)=\sum_{k=0}^{n}c_k\varphi_k(x)$ attains

$$\inf_{c_0,\ldots,c_n}\max_{x\in A}|f(x)-P_n(x)|$$

is called the **best approximation** of $f(x)$ by a linear combination of $\{\varphi_k(x)\}$. For any given n there is a best approximation of $f(x)$ by a linear combination of $\varphi_0(x),\ldots,\varphi_n(x)$, but such an approximation is not always unique. For such an approximation to be unique it is necessary and sufficient that the determinant of the matrix $(\varphi_k(x_i))$ $(k,i=0,1,2,\ldots,n)$ is not zero, where $x_0, x_1, \ldots, x_n$ are $n+1$ arbitrary distinct points of A (**Haar's condition**) (*Math. Ann.*, 78 (1918)). If $\{\varphi_k(x)\}$ satisfies this condition, the system of functions $\{\varphi_k\}_{k=0,\ldots,n}$ is called a **Chebyshev system** (or **unisolvent system**). The sets $\{1,x,x^2,\ldots,x^n\}$ on $[a,b]$, $\{1,\cos x,\ldots,\cos nx\}$ on $[0,\pi]$ and $\{\sin x,\ldots,\sin nx\}$ on $[0,\pi]$ are Chebyshev systems. For a Chebyshev system $\{\varphi_k(x)\}$ on $[a,b]$, let $P_n(x)$ be a linear combination of $\varphi_0(x),\ldots,\varphi_n(x)$ that is not identical to the function $f\in C[a,b]$. Then $P_n(x)$ is the best approximation for $f(x)$ if and only if there are at least $n+2$ distinct points $x_1<\ldots<x_{n+2}$ of $[a,b]$, where $|f(x)-P_n(x)|$ attains its maximum (these points are called **deviation points**) and $(f(x_i)-P_n(x_i))(f(x_{i+1})-P_n(x_{i+1}))<0$ $(i=1,\ldots,n+1)$ (**Chebyshev's theorem**).

For example, consider the polynomial $P_n(x)=a_{n-1}x^{n-1}+\ldots+a_1x+a_0$ with real coefficients such that

$$\max_{-1\leqslant x\leqslant 1}|x^n-a_{n-1}x^{n-1}-\ldots-a_0|$$

takes its smallest value. Then $x^n-P_n(x)=2^{-(n-1)}T_n(x)$, where $T_n(x)=\cos(n\arccos x)$ is the Chebyshev polynomial of degree n.

Since the best approximation is desired for numerical computation, several methods have been developed to find it (→ 300 Numerical Methods). However, when the set $A\subset\mathbf{R}^n$ $(n\geqslant 2)$ contains three nonintersecting arcs emanating from a common point, A admits no Chebyshev system. Thus we do not always have a unique best-approximation polynomial [16].

C. Degrees of Approximation and Moduli of Continuity

For a continuous function $f(x)$ defined on $[a,b]$, the **modulus of continuity of kth order** is defined by

$$\omega_k(f;t)=\sup_{\substack{|h|\leqslant t\\ a\leqslant x\leqslant b\\ a\leqslant x+kh\leqslant b}}\left|\sum_{v=0}^{k}(-1)^{k-v}\binom{k}{v}f(x+vh)\right|$$

for $t\leqslant(b-a)/k$. In particular, ω_1 is the ordinary modulus of continuity. Put $E_n^*(f)=\inf_{a_0,\ldots,a_n,b_1,\ldots,b_n}\max_{a\leqslant x\leqslant b}|f(x)-P_n(x)|$, where f is a continuous periodic function of period 2π and $P_n(x)$ is a trigonometric polynomial of the form (1). Then $E_n^*(f)\leqslant c_k\omega_k(f;1/(n+1))$ (**Jackson's theorem** [1]), where c_k is indepen-

dent of f. The best possible coefficient c_k has been determined by J. Favard [2]. Further investigations on the relation between $E_n^*(f)$ and $\omega_k(f;t)$ have been carried out by S. N. Bernshteĭn [3] and A. Zygmund [4]. S. B. Stechkin obtained the following results:

$$\omega_k\left(f;\frac{1}{n}\right)\leqslant\frac{c_k}{n^k}\sum_{v=0}^{n}(v+1)^{k-1}E_v^*(f)$$

[5, 6]. For the approximation of $f\in C^r([-1,1])$ by polynomials, there exists a polynomial $P_n(x)$ of degree at most n such that for any $x\in[-1,1]$,

$$|f(x)-P_n(x)|\leqslant M_r(t(x))^r\omega_1(f^{(r)};t(x)),$$

where M_r is a constant not depending on f, x, and n, $f^{(r)}(x)$ is the rth derivative of $f(x)$, and $t(x)=(1/n)(\sqrt{1-x^2}+(|x|/n))$. We also have theorems evaluating $\omega_k(f^{(r)};t)$ in terms of $|f(x)-P_n(x)|$. For the proof of these theorems, estimation of the magnitude of the derivative of the polynomial of degree n plays an essential role. For example, we have the **Bernshteĭn inequality** $\max_x|T_n'(x)|\leqslant n\max_x|T_n(x)|$ for any trigonometric polynomial $T_n(x)$ of degree n and the **Markov inequality**

$$|P_n'(x)|\leqslant n\min[1/\sqrt{1-x^2},n]\max_{-1\leqslant x\leqslant 1}|P_n(x)|$$

for $x\in[-1,1]$ and any polynomial $P_n(x)$ of degree n.

D. Approximation by Fourier Expansions

If $\{\varphi_n(x)\}$ is an †orthonormal system of functions in $L_2(a,b)$ and f is any function in $L_2(a,b)$, then among all linear combinations of $\varphi_0(x),\ldots,\varphi_n(x)$ the one that gives the best mean square approximation to f (i.e., the one for which the integral

$$\int_a^b\left(f(x)-\sum_{k=0}^{n}c_k\varphi_k(x)\right)^2dx$$

attains its minimum) is the Fourier polynomial $\sum_{k=0}^{n}a_k\varphi_k(x)$, where $a_k=\int_a^b f(x)\varphi_k(x)\,dx$. Consequently, the **least square approximation** (or best approximation with respect to the L_2-norm) by trigonometric polynomials is given by the partial sum $s_n(x)$ of the Fourier series of $f(x)$. For L_p $(1<p<\infty)$, $s_n(x)$ also gives the best approximation up to a constant factor, but in the case of uniform approximation we have $|f(x)-s_n(x)|\leqslant A(\log n)\omega_k(f;n^{-1})$, and this result cannot be improved in general. There is no linear operation that gives the best trigonometric approximation. In approximation with a linear combination of $\varphi_0(x),\ldots,\varphi_n(x)$, the saturation phenomenon of approximation often appears. For example, observe the arithmetic means of $s_n(x)$ (i.e., †Fejér means $\sigma_n(x)$). If $f\in\operatorname{Lip}\alpha$ $(0<\alpha<1)$ (→ 84 Continuous Functions A), then $|f(x)-\sigma_n(x)|=O(n^{-\alpha})$. However, $|f(x)-\sigma_n(x)|=O(n^{-1})$ if and only if the †conjugate function $\tilde{f}(x)\in\operatorname{Lip}1$; $|f(x)-\sigma_n(x)|=o(n^{-1})$ if and only if $f(x)$ is constant (M. Zamansky [7], G. Sunouchi and C. Watari [8]).

E. Trigonometric Interpolation Polynomials

Since the trigonometric system is a Chebyshev system, given $2n+1$ distinct points $x_0,x_1,\ldots,x_{2n}$ and arbitrary numbers $c_0,c_1,\ldots,c_{2n}$, there is always a unique trigonometric polynomial of degree n with prescribed values c_k at the points x_k. Given any continuous function $f(x)$ with period 2π, the trigonometric polynomial that coincides with $f(x)$ at the points x_k is called the **trigonometric interpolation polynomial** with nodes at x_k. If $x_k=2\pi k/(2n+1)$ $(k=0,1,\ldots,2n)$, then the interpolating trigonometric polynomial is given by

$$U_n(f,x)=\frac{1}{2n+1}\sum_{j=0}^{2n}f(x_j)\frac{\sin((n+1/2)(x-x_j))}{\sin((x-x_j)/2)}$$

$$=\frac{1}{2\pi}\int_0^{2\pi}f(t)\frac{\sin((n+1/2)(x-t))}{\sin((x-t)/2)}d\varphi_n(t),$$

where $\varphi_n(t)$ is a step function that has the value $2\pi j/(2n+1)$ in $[2\pi j/(2n+1),2\pi(j+1)/(2n+1)]$. $U_n(f,x)$ resembles the partial sum $s_n(x)$ of the Fourier series of $f(x)$. If $f(x)$ is continuous and of †bounded variation, then $U_n(f,x)$ converges uniformly to $f(x)$ (D. Jackson [1]). Although the partial sum $s_n(x)$ of the Fourier series of a continuous function $f(x)$ converges almost everywhere to $f(x)$, there is a continuous function for which $U_n(f,x)$ diverges everywhere (J. Marcinkiewicz [9]). Moreover, there exists a continuous function for which $(1/n)(\sum_{k=1}^{n}U_k(f,x))$ diverges everywhere (P. Erdős [10], G. Grünwald [11]). Restating these facts for the algebraic polynomial case, we can conclude that there is a continuous function defined in $[-1,1]$ for which the †Lagrange interpolation polynomial and its arithmetic mean are both divergent everywhere if we take as nodes the roots of the Chebyshev polynomial of degree n.

F. The Case of a Complex Domain

If a given function $f(z)$ is holomorphic in a bounded †simply connected domain E in the complex plane and continuous in $\bar{E}$, then $f(z)$ is approximated uniformly by polynomials on any compact set in E (**Runge's theorem**). This theorem was first studied by C. Runge, and his results were developed by J. L. Walsh and M. V. Keldysh (e.g., [12]). When E contains no interior point, the polynomial approxima-

tion of a continuous function defined in E was given by M. A. Lavrent'ev. Unifying these two results, S. N. Mergelyan obtained the following theorem [13]: A necessary and sufficient condition for an arbitrary function continuous on a compact set E and holomorphic inside E to be approximated on E uniformly by polynomials is that the set E does not divide the complex plane.

On the degree of approximation of polynomials to $f(z)$ on a simply connected domain, there are the following results: Let D be a closed bounded set whose complement K is connected and regular in the sense that K possesses a †Green's function $G(x, y)$ with a pole at infinity. Let D_R be the locus $G(x, y) = \log R > 0$. When $f(z)$ is holomorphic on D, there exists the largest number ρ with the following property: $f(z)$ is single-valued and holomorphic at every interior point of D_ρ. If $R < \rho$, there exist polynomials $P_n(z)$ of degree n $(n = 1, 2, \ldots)$ such that $|f(z) - P_n(z)| \leqslant M/R^n$ for $z \in D$, where M is a constant independent of n and z. On the other hand, there exist no such polynomials $P_n(z)$ on D for $R > \rho$ (Bernshteĭn and Walsh [12]).

G. Lagrange's Interpolation Formula

For each n $(n = 0, 1, \ldots)$, let $z_1^{(n)}, z_2^{(n)}, \ldots, z_{n+1}^{(n)}$ be a given set of real or complex numbers, and let $f(z)$ be an arbitrary function. Then there is a unique polynomial of degree n that coincides with $f(z)$ at each point $z_k^{(n)}$ $(k = 1, \ldots, n+1)$. This is called **Lagrange's interpolation polynomial** and is given by

$$P_n(z) = \sum_{k=1}^{n+1} f(z_k^{(n)}) \frac{\omega(z)}{(z - z_k^{(n)})\omega'(z_k^{(n)})},$$

$$\omega(z) = (z - z_1^{(n)}) \ldots (z - z_{n+1}^{(n)}).$$

The sequence $P_n(z)$ does not always converge to $f(z)$. For example, if we take $f(z) = 1/z$ and the $(n+1)$st roots of 1 as $z_k^{(n)}$, then $P_n(z) = z^n$ and $P_n(z)$ converges to $f(z)$ only at the point 1. For real variables also, there are examples of divergent $P_n(z)$. However, if $f(z)$ is holomorphic in $|z| < \rho$ $(\rho > 1)$, then $P_n(z)$ with the $(n+1)$st roots of 1 as nodes converges to $f(z)$ uniformly in $|z| \leqslant 1$.

When $z_k^{(n)}$ is independent of the choice of n, $P_n(z)$ coincides with the sum of the first n terms of †Newton's interpolation formula. In this case, $P_n(z)$ is called **Newton's interpolation polynomial** and is given by

$$P_n(z) = a_0 + a_1(z - z_1) + a_2(z - z_1)(z - z_2) + \ldots$$
$$+ a_n(z - z_1) \ldots (z - z_n),$$

where $a_0 = f(z_1)$; $a_1 = (f(z_2) - f(z_1))/(z_2 - z_1)$ $(z_2 \neq z_1)$, $a_1 = f'(z_1)$ $(z_2 = z_1)$; and so on. Successive coefficients of the polynomial $P_n(z)$ can be calculated by †finite differences. Convergence of Newton's interpolation polynomial is closely connected to convergence of †Dirichlet series.

H. Chebyshev Approximation

Let D be a bounded closed subset of the complex plane, and $f(z)$ a continuous function on D. Then there exists a polynomial $\pi_n(z)$ of degree n such that $\max_{z \in D} |f(z) - \pi_n(z)|$ attains the infimum $E_n(f)$. The polynomial $\pi_n(z)$ is unique and is called the **best approximation polynomial (in the sense of Chebyshev)**. If D is simply connected and $f(z)$ is single-valued and holomorphic on D, then $\pi_n(z)$ converges to $f(z)$ uniformly on D. Moreover, in this case there exist a number M that does not depend on n and a number $R > 1$ such that $|f(z) - \pi_n(z)| \leqslant M/R^n$. Assuming that $f(z)$ satisfies certain additional conditions, W. E. Sewell [14] proved the existence of a constant r such that $|f(z) - \pi_n(z)| \leqslant M/n^r R^n$. Furthermore, by approximating $f(z) = z^n$ by polynomials of degree $n-1$, we can show that there exists a polynomial $T_n(z)$ of degree n such that

$$\min \left\{ \max_{z \in D} |z^n + a_1 z^{n-1} + \ldots + a_n| \right\} =: |T_n(z)|.$$

$T_n(z)$ is called a **Chebyshev polynomial** of degree n with respect to the domain D. Similar statements are valid for functions of a real variable. In particular, when $D = [-1, 1]$, we have

$$T_n(x) = \cos(n \arccos x)/2^{n-1},$$

which is the ordinary (real) Chebyshev polynomial. Generally, the limit

$$\lim_{n \to \infty} \left(\max_D |T_n(z)| \right)^{1/n} = \rho(D)$$

exists, and $\rho(D)$ coincides with the †capacity and †transfinite diameter of D [15]. For new results and applications of Chebyshev polynomials → [17].

If the method of evaluating the degree of approximation using the absolute value $|f(z) - \pi_n(z)|$ is replaced by methods using a †curvilinear integral or †surface integral, as explained below, we still obtain similar results. Let D be a closed domain in the complex plane with a boundary C that is a rectifiable Jordan curve. If $f(z)$ is single-valued and holomorphic on D, then there exists a polynomial $\pi_n(z)$ of degree n that minimizes the integral $\int_C u(z)|f(z) - \pi_n(z)|^p |dz|$ $(p > 0)$, where $u(z)$ is a given positive continuous function on C. Moreover, $|f(z) - \pi_n(z)| \leqslant M/R^n$ for some $R > 1$ (actually $\{\pi_n(z)\}$ is †overconvergent). If D is a closed

Jordan domain and if $f(z)$ is single-valued and holomorphic in D, then there exists a polynomial $\pi_n(z)$ of degree n that minimizes the integral $\int\int_D u(z)|f(z)-\pi_n(z)|^p dS$, where $u(z)$ is a given positive continuous function on D. Moreover, $|f(z)-\pi_n(z)| \leqslant M/R^n$ for some $R>1$ on D.

I. Approximation by Orthogonal Polynomials on a Curve

Let C be a rectifiable Jordan curve in the complex plane, and let $p_k(z) \in {}^\dagger L_2(C)$. If $\int_C p_k(z)\overline{p_n(z)}|dz| = \delta_{kn}$, then $\{p_k(z)\}$ is called an orthonormal system on C. Given a holomorphic function $f(z)$ on D, we set $a_k = \int_C f(z)p_k(z)|dz|$ and consider the formal series $\sum_{k=0}^{\infty} a_k p_k(z)$. If we denote the nth partial sum of this series by $s_n(z)$, then $s_n(z)$ is the least square approximation by a linear combination of $p_0(z), \ldots, p_n(z)$. This and other results, such as †Bessel's inequality, the †Riesz-Fischer theorem, etc., are all valid here as in the theory of general †orthogonal systems. In particular, if we take $|z|=R$ as C, then $\{1, z, z^2, \ldots\}$ is an orthogonal system. Since in this case

$$a_k = \frac{1}{2\pi R^{2k+1}} \int_C f(z)\bar{z}^k |dz| = \frac{1}{2\pi i} \int_C \frac{f(z)}{z^{k+1}} |dz|$$

and $s_n(z) = a_0 + a_1 z + \ldots + a_n z^n$, the †Taylor expansion of $f(z)$ coincides with the orthogonal expansion of $f(z)$.

Given a compact domain D and a holomorphic function on D, if there exist orthogonal polynomials $p_n(z)$ such that the orthogonal expansion of $f(z)$ with respect to $p_n(z)$ converges to $f(z)$ uniformly on D, we say that $\{p_n(z)\}$ belongs to the domain D. The problem of existence and determination of such polynomials for any given domain was proposed and first solved by G. Faber. Generalizations were given by G. Szegö, T. Carleman, and Walsh. Roughly speaking, $p_n(z)$ is given by the orthogonalization of the system $\{1, z, z^2, \ldots\}$ with respect to the curvilinear integral on $C = \partial D$ or the surface integral on D.

J. Numerical Approximation of Functions

The accuracy of the approximation of a given function $f(z)$ by the partial sums of its †Taylor expansion $\sum a_n(x-x_0)^n$ decreases rapidly as the distance $|x-x_0|$ increases. The accuracy of the approximation of $f(x)$ defined on a compact interval $[A, B]$ by a (polynomial) function $\varphi(x)$ can be evaluated by means of the least square approximation, the best approximation with respect to the uniform norm, and so on. The second method is best suited to numerical calculation of functions. To get the best approximating polynomial $\varphi(x) = P_n(x) = \sum_{k=0}^{n} c_k \varphi_k(x)$ of $f(x)$ (→ Section B), we must determine coefficients c_k that satisfy the conditions of Chebyshev's theorem. The first step in this process is the orthogonal development of $f(x)$ by Chebyshev polynomials $\{T_n(x)\}$: $\varphi_n(x) = \sum_{k=0}^{n} a_k T_k(u)$, $u = (x-(A+B)/2)/((B-A)/2)$. The error $|f(x)-\varphi_n(x)|$ is estimated by a constant multiple of $T_{n+1}(u)$: $|f(x)-\varphi_n(x)| \leqslant K|T_{n+1}(u)|$. This **Chebyshev interpolation** is actually given by $a_0 = N^{-1}\sum_{i=1}^{N} f(x_i)$, $a_k = 2N^{-1}\sum_{i=1}^{N} f(x_i)T_k(u_i)$ $(k=1, \ldots, n)$, where $N = n+1$ and the $u_i = (x_i-(A+B)/2)/((B-A)/2)$ $(i=1, \ldots, N)$ are the roots of $T_N(u)$. Let M be the extremum of the error $|f(x)-\varphi_n(x)|$ of such an approximation, and set $f(x_i) - \varphi_n(x_i) = \pm M_i$ $(i=1, 2, \ldots, N)$. Consider a function $\bar{\varphi}_n(x) = \sum \bar{a}_k T_k(u)$ satisfying $f(x_i) - \bar{\varphi}(x_i) = \pm M$. Then solve the linear equation $\bar{\varphi}_n(x_i) - \varphi_n(x_i) = \pm(M-M_i)$ with respect to $\Delta a_k = \bar{a}_k - a_k$ and M. Repeat this process until Δa_k becomes sufficiently small.

A computer can perform the division very quickly, and the rational approximation of a function, for example by its †continued fraction expansion, is often useful.

References

[1] D. Jackson, The theory of approximation, Amer. Math. Soc. Colloq. Publ., 1930.
[2] J. Favard, Sur les meilleurs procédés d'approximation des certains classes des fonctions par des polynomes trigonometrique, Bull. Sci. Math., 61 (1937), 209–224, 243–265.
[3] S. N. Bernshteĭn (Bernstein), Sur l'ordre de meilleure approximation des fonctions continues par des polynomes de degré donné, Mém. Acad. Roy. Belgique, ser. 2, 4 (1912), 1–104.
[4] A. Zygmund, Smooth functions, Duke Math. J., 12 (1945), 47–75.
[5] N. I. Achieser, Theory of approximation, Unger, 1956. (Original in Russian, 1947.)
[6] A. Zygmund, Trigonometric series, Cambridge, 1959.
[7] M. Zamansky, Class de saturation de certains procédés d'approximation des séries de Fourier des fonctions continues et applications à quelques problèmes d'approximation, Ann. Sci. Ecole Norm. Sup., sér. 3, 66 (1949), 19–93.
[8] G. Sunouchi and C. Watari, On the determination of the class of saturation in the theory of approximation of functions II, Tôhoku Math. J., 11 (1959), 480–488.
[9] J. Marcinkiewicz, Sur la divergence des polynomes d'interpolation, Acta Sci. Math. Szeged., 8 (1937), 131–135.

[10] P. Erdös, Some theorems and remarks on interpolations, Acta Sci. Math. Szeged., 12 (1950), 11–17.
[11] G. Grünwald, Über Divergenzerscheinungen der Lagrangeschen Interpolationspolynome der stetigen Functionen, Ann. Math., (2) 37 (1936), 908–918.
[12] J. L. Walsh, Interpolation and approximation by rational functions in the complex domain, Amer. Math. Soc. Colloq. Publ. 1935; fifth edition, 1981.
[13] S. N. Mergelyan (Mergeljan), On the representation of functions by series of polynomials on closed sets, Amer. Math. Soc. Transl., ser. I, 3 (1962), 287–293. (Original in Russian, 1951.)
[14] W. E. Sewell, Degree of approximation by polynomials in the complex domain, Ann. math. studies 9, Princeton Univ. Press, 1942.
[15] E. Hille, Analytic function theory II, Ginn, 1962.
[16] J. R. Rice, The approximation of functions, Addison-Wesley, I, 1964; II, 1969.
[17] S. J. Karlin and W. J. Sneddon, Tchebysheff systems with applications in analysis and statistics, Interscience, 1966.

337 (III.4) Polynomials

A. Polynomials in One Variable

Let R be a commutative †ring and $a_0, a_1, \ldots, a_n$ elements of R. An expression $f(X)$ of the form

$$f(X) = a_0 + a_1 X + \ldots + a_n X^n \tag{1}$$

is called a **polynomial** in a variable X over R; if $a_n \neq 0$, the number n is called the **degree** of the polynomial $f(X)$ and is denoted by $\deg f$. If $a_n = 1$, the polynomial (1) is called a **monic polynomial**. The totality of polynomials in X over R forms a commutative ring with respect to ordinary addition and multiplication (whose definition will be given later). It is called the **ring of polynomials** (or the **polynomial ring**) of X over R and is denoted by $R[X]$. We say that we **adjoin** X to R to obtain $R[X]$.

B. Polynomials in Several Variables

Let $R[X, Y]$ denote the ring $R[X][Y]$, namely, the ring obtained by adjoining Y to $R[X]$. An element of $R[X, Y]$ can then be expressed as $\sum a_{\mu\nu} X^\mu Y^\nu$. This expression is called a polynomial in X and Y over R. Generally, $R[X_1, \ldots, X_m] = R[X_1, \ldots, X_{m-1}][X_m]$ is called the **polynomial ring in** m **variables** (on m **indeterminates**) $X_1, \ldots, X_m$ over R, and its element

$$F(X_1, X_2, \ldots, X_m) = \sum a_{\nu_1 \nu_2 \ldots \nu_m} X_1^{\nu_1} X_2^{\nu_2} \ldots X_m^{\nu_m} \tag{2}$$

($\sum$ denotes a finite sum for nonnegative integral ν_i beginning with $\nu_1 = \nu_2 = \ldots = \nu_m = 0$) is called a **polynomial in** m **variables** $X_1, \ldots, X_m$ over R. We call each summand a **term** of the polynomial, $a_{\nu_1 \nu_2 \ldots \nu_m}$ the **coefficient** and $\nu_1 + \nu_2 + \ldots + \nu_m$ the **degree** of this term. The greatest degree of terms is called the **degree** of the polynomial F. The term $a_{0 \ldots 0}$ of degree 0 is called the **constant term** of F. If a polynomial F in $X_1, \ldots, X_m$ is composed of terms of the same degree n, then F is called a **homogeneous polynomial** (or **form**) of degree n; a polynomial consisting of a single term, such as $aX_1^{\nu_1} X_2^{\nu_2} \ldots X_m^{\nu_m}$, is called a **monomial**. Now let $\alpha_1, \alpha_2, \ldots, \alpha_m$ be elements of R (or a commutative ring S containing R), and let $F(\alpha_1, \alpha_2, \ldots, \alpha_m)$ denote the element of R (or S) obtained by substitution of $\alpha_1, \alpha_2, \ldots, \alpha_m$ for $X_1, X_2, \ldots, X_m$ in $F(X_1, X_2, \ldots, X_m)$. It is also called a polynomial in $\alpha_1, \alpha_2, \ldots, \alpha_m$. If $F(\alpha_1, \alpha_2, \ldots, \alpha_m) = 0$, then $(\alpha_1, \alpha_2, \ldots, \alpha_m)$ is called a **zero point** (in S) of the polynomial $F(X_1, X_2, \ldots, X_m)$ (or a solution of the algebraic equation $F(X_1, \ldots, X_m) = 0$). In the case of one variable, it is called a **root** of $F(X_1)$ (or of $F(X_1) = 0$).

C. Polynomial Rings

Addition and multiplication in $R[X]$ are defined by

$(\sum a_i X^i) + (\sum b_j X^j) = \sum (a_i + b_i) X^i$,

$(\sum a_i X^i)(\sum b_j X^j) = \sum_k (\sum_{i+j=k} a_i b_j) X^k$.

A polynomial $f(X) \in R[X]$ can be regarded as a function of a commutative ring R' containing R into itself such that $c \mapsto f(c)$. In this sense, $f(X) + g(X)$ and $f(X)g(X)$ are the functions such that $c \mapsto f(c) + g(c)$ and $c \mapsto f(c)g(c)$, respectively.

It holds that $\deg(f(X) + g(X)) \leqslant \max\{\deg f(X), \deg g(X)\}$, $\deg f(X)g(X) \leqslant \deg f(X) + \deg g(X)$. If R is an †integral domain, then the latter inequality is an equality, and therefore $R[X]$ is an integral domain. For these inequalities and for convenience elsewhere, we define the degree of 0 to be indefinite.

Assume that R is a field. For given $f, g \in R[X]$ ($\deg g \geqslant 1$), we can find unique $q, r \in R[X]$ such that $f = gq + r$ and $\deg r < \deg g$ or $r = 0$ (**division algorithm**). This q is called the **integral quotient** of f by g, and r is called the **remainder** of f divided by g. The same fact remains true in the general $R[X]$ if $g(X)$ is monic. (→ 369 Rings of Polynomials).

D. Factorization into Primes

Let k be an integral domain. Since $k[X]$ and hence $k[X_1, \ldots, X_m]$ are integral domains, we can define the concepts concerning divisibility (such as a divisor, a multiple, etc.) (→ 67 Commutative Rings). If k is a †unique factorization domain, then so are $k[X]$ and $k[X_1, \ldots, X_m]$. A polynomial over k is said to be **primitive** if the greatest common divisor of all the coefficients is equal to 1. Every polynomial over k can be uniquely expressed as a product of some primitive polynomials and an element of k; a product of primitive polynomials is primitive (**Gauss's theorem**).

If k is a field, then $k[X]$ and $k[X_1, \ldots, X_m]$ are unique factorization domains. Furthermore, to find the greatest common divisor (f, g) of $f, g \in k[X]$, we can use the **Euclidean algorithm**, that is, apply the division algorithm repeatedly to obtain

$$f = gq_1 + r_1, \quad g = r_1 q_2 + r_2, \quad r_1 = r_2 q_3 + r_3, \ldots,$$

$$\deg g > \deg r_1 > \deg r_2 > \ldots,$$

so that after a finite number of steps we attain $r_{v-1} = r_v q_{v+1} (r_{v+1} = 0)$. Then $r_v = (f, g)$. Accordingly, $k[X]$ is a †principal ideal domain. This algorithm is applied to Euclid rings (→ 67 Commutative Rings L).

E. The Remainder Theorem

Let k be an integral domain, $f(X) \in k[X]$, and let $g(X) = X - \alpha (\alpha \in k)$. Then using the division algorithm, we get

$$f(X) = (X - \alpha) q(X) + r,$$

$$q(X) \in k[X], \quad r \in k.$$

Therefore, $f(\alpha) = r$; that is, the remainder of $f(X)$ divided by $X - \alpha$ is equal to $f(\alpha)$. This is called the **remainder theorem**. If $f(\alpha) = 0$, then $f(X)$ is divisible by $X - \alpha$ in $k[X]$.

F. Irreducible Polynomials

Let k be a field. A polynomial $f(X) \in k[X]$ of degree n is said to be **reducible** over k if f is divisible by a polynomial of degree $v < n$ in $k[X]$ $(v \neq 0)$; otherwise, it is said to be **irreducible** over k. Any polynomial of degree 1 is irreducible. A polynomial f is a †prime element of $k[X]$ if and only if f is irreducible over k. Let I be a unique factorization domain. If $f(X)$ is a polynomial (1) in $I[X]$ such that for a prime element p in I, $a_n \not\equiv 0 \pmod p$, $a_{n-1} \equiv a_{n-2} \equiv \ldots \equiv a_0 \equiv 0 \pmod p$ but $a_0 \not\equiv 0 \pmod{p^2}$, then $f(X)$ is irreducible over the field of quotients of I (**Eisenstein's theorem**). If a polynomial (2) in m variables over an †algebraic number field k is irreducible, we can obtain an irreducible polynomial in $X_1, \ldots, X_\mu$ $(0 < \mu < m)$ from the polynomial $F(X_1, \ldots, X_m)$ by assigning appropriate values in k to $X_{\mu+1}, \ldots, X_m$ (**Hilbert's irreducibility theorem**, *J. Reine Angew. Math.*, 110 (1892)). These two theorems have been generalized in many ways and given precise formulations. In Hilbert's irreducibility theorem, the algebraic number field may be replaced, for example, by any infinite field that is †finitely generated over its †prime field (K. Dörge, W. Franz, E. Inaba).

G. Derivatives

Given a polynomial

$$f(X_1, \ldots, X_m) = \sum a_{v_1 v_2 \ldots v_m} X_1^{v_1} X_2^{v_2} \ldots X_m^{v_m}$$

over a field k, we define the (formal) **derivative** of f with respect to X_i as $f_i(X_1, \ldots, X_m) = \sum v_i a_{v_1 v_2 \ldots v_m} X_1^{v_1} \ldots X_i^{v_i - 1} \ldots X_m^{v_m}$ and denote it by $\partial f / \partial X_i$. The map $f \mapsto \partial f / \partial X_i$ is called the (formal) **derivative** with respect to X_i. In particular, if $m = 1$, then $\partial f / \partial X$ is denoted by df/dX. The usual rules of †derivatives also hold for the formal derivative. If $df/dX = 0$ for an irreducible polynomial $f(X)$ in $k[X]$, then $f(X)$ is said to be **inseparable**; otherwise, $f(X)$ is **separable**. If the †characteristic of the field k is 0, then every irreducible polynomial $f(X)$ $(\neq 0)$ is separable. When k is of characteristic $p \neq 0$, an irreducible polynomial $f(X)$ is inseparable if and only if we can write $f(X) = g(X^p)$.

H. Rational Expressions

The †field of quotients of the polynomial ring $k[X_1, \ldots, X_n]$ over a field k is denoted by $k(X_1, \ldots, X_n)$ and is called the **field of rational expressions** (or **field of rational functions**) in variables $X_1, \ldots, X_n$ over k. Its element is called a **rational expression** in $X_1, \ldots, X_n$. It can be written as a quotient of one polynomial $f(X_1, \ldots, X_n)$ by another polynomial $g(X_1, \ldots, X_n) \neq 0$. Also, an expression $f(\alpha_1, \ldots, \alpha_n)/g(\alpha_1, \ldots, \alpha_n)$ obtained by replacing $X_1, \ldots, X_n$ with elements $\alpha_1, \ldots, \alpha_n$ of k in the above expression is called a rational expression in $\alpha_1, \ldots, \alpha_n$ (provided that $g(\alpha_1, \ldots, \alpha_n) \neq 0$).

I. Symmetric Polynomials and Alternating Polynomials

Let $f(X_1, \ldots, X_n)$ be a polynomial in variables $X_1, \ldots, X_n$ over an integral domain I. If

$f(X_1, \ldots, X_n)$ is invariant under every permutation of $X_1, \ldots, X_n$, it is called a **symmetric polynomial** (or **symmetric function**) of $X_1, \ldots, X_n$. If $f(X_1, \ldots, X_n)$ is transformed into $-f(X_1, \ldots, X_n)$ by every †odd permutation of $X_1, \ldots, X_n$, it is called an **alternating polynomial** (or **alternating function**). Also, an expression $f(\alpha_1, \ldots, \alpha_n)$ obtained from $f(X_1, \ldots, X_n)$ by replacing $X_1, \ldots, X_n$ with elements $\alpha_1, \ldots, \alpha_n$ of I is called a symmetric (alternating) function of $\alpha_1, \ldots, \alpha_n$ if $f(X_1, \ldots, X_n)$ is symmetric (alternating).

Let the coefficient of X^{n-k} in the expansion of $(X-X_1)\ldots(X-X_n)$ be denoted by $(-1)^k\sigma_k$. Then we have $\sigma_1 = \sum X_i = X_1 + \ldots + X_n, \sigma_2 = \sum X_i X_j = X_1X_2 + X_1X_3 + \ldots + X_{n-1}X_n, \ldots, \sigma_n = X_1X_2, \ldots, X_n$. Obviously, these are symmetric polynomials of $X_1, \ldots, X_n$. Moreover, for every element φ of the polynomial ring $I[Y_1, Y_2, \ldots, Y_n]$, $\varphi(\sigma_1, \sigma_2, \ldots, \sigma_n)$ is a symmetric polynomial of $X_1, \ldots, X_n$. Conversely, every symmetric polynomial of $X_1, \ldots, X_n$ can be uniquely expressed as a polynomial $\varphi(\sigma_1, \sigma_2, \ldots, \sigma_n)$. Thus the totality of symmetric polynomials of $X_1, \ldots, X_n$ is identical with the ring $I[\sigma_1, \sigma_2, \ldots, \sigma_n]$. This is called the **fundamental theorem on symmetric polynomials**, and $\sigma_1, \sigma_2, \ldots, \sigma_n$ are called **elementary symmetric polynomials** (or **elementary symmetric functions**). For example, for $s_\nu = \sum X_i^\nu$ $(\nu = 1, 2, \ldots)$, we have $s_1 = \sigma_1, s_2 = \sigma_1^2 - 2\sigma_2, s_3 = \sigma_1^3 - 3\sigma_1\sigma_2 + 3\sigma_3, s_4 = \sigma_1^4 - 4\sigma_1^2\sigma_2 + 2\sigma_2^2 + 4\sigma_1\sigma_3 - 4\sigma_4$. Concerning the elementary symmetric polynomials and the s_ν, we have the relations $s_\nu - \sigma_1 s_{\nu-1} + \sigma_2 s_{\nu-2} - \ldots + (-1)^{\nu-1}\sigma_{\nu-1}s_1 + (-1)^\nu \nu\sigma_\nu = 0$ $(\nu = 1, 2, \ldots)$, and $s_\mu - \sigma_1 s_{\mu-1} + \ldots + (-1)^n \sigma_n s_{\mu-n} = 0$ $(\mu = n+1, n+2, \ldots)$ **(Newton's formulas)**.

Let $p(X_1, \ldots, X_n) = (X_2 - X_1)(X_3 - X_1)\ldots(X_n - X_1)(X_3 - X_2)\ldots(X_n - X_{n-1})$ be the product of $n(n-1)/2$ differences between $X_1, \ldots, X_n$. Then the polynomial p is invariant under even permutations of $X_1, \ldots, X_n$, and p becomes $-p$ under odd permutations. Hence p is an alternating polynomial of $X_1, \ldots, X_n$. It is called the **simplest alternating polynomial** of these variables. Because of its particular expression, p is also called the **difference product** of $X_1, \ldots, X_n$. If the characteristic of I is different from 2, an alternating polynomial f is divisible by the simplest alternating polynomial p; it can be written as $f = ps$, where s stands for a symmetric polynomial.

J. Discriminants

The square $D(X_1, \ldots, X_n) = p^2(X_1, \ldots, X_n)$ of the simplest alternating polynomial p is a symmetric polynomial, and it is therefore a polynomial in $\sigma_1, \ldots, \sigma_n$. $D(\alpha_1, \alpha_2, \ldots, \alpha_n) = 0$ gives a criterion for the condition that some of $\alpha_1, \ldots, \alpha_n$ are equal. If $\alpha_1, \ldots, \alpha_n$ are the roots of an †algebraic equation $a_0X^n + a_1X^{n-1} + \ldots + a_n = 0$ of degree n, then $D(\alpha_1, \ldots, \alpha_n)$ is called the **discriminant of the equation**. It can be expressed in terms of coefficients $a_0, a_1, \ldots, a_n$ of the equation. For instance, if $n = 2$, we have $a_0^2 D = a_1^2 - 4a_0a_2$; if $n = 3$, we have $a_0^4 D = a_1^2a_2^2 + 18a_0a_1a_2a_3 - 4a_0a_2^3 - 4a_1^3a_3 - 27a_0^2a_3^2$.

References

[1] B. L. van der Waerden, Algebra I, Springer, seventh edition, 1966.
[2] N. Bourbaki, Eléments de mathématique, Algèbre, ch. 4, Actualités Sci. Ind., 1102b, Hermann, second edition, 1959.
[3] A. G. Kurosh, Lectures on general algebra, Chelsea, 1970. (Original in Russian, 1955.)
[4] R. Godement, Cours d'algèbre, Hermann, 1963.
For Hilbert's irreducibility theorem,
[5] S. Lang, Diophantine geometry, Interscience, 1962, ch. 8.

338 (X.28) Potential Theory

A. Newtonian Potential

In dynamics, a **potential** means a function u of n variables $x_1, \ldots, x_n$ such that $-\operatorname{grad} u = -(\partial u/\partial x_1, \ldots, \partial u/\partial x_n)$ gives a field of force in the n-dimensional $(n \geqslant 2)$ Euclidean space $\mathbf{R}^n$. Given a point P in $\mathbf{R}^n$ and a measure μ, the functions $u(P)$ given by the integrals

$$u(P) = -\int \log \overline{PQ}\, d\mu(Q), \qquad n = 2,$$

$$u(P) = \int \overline{PQ}^{2-n}\, d\mu(Q), \qquad n \geqslant 3,$$

are typical examples of potential functions. They are called the **logarithmic potential** and **Newtonian potential**, respectively. However, some authors mean by Newtonian potential the function $u(P) = \int \overline{PQ}^{-1}\, d\mu(Q)$ in $\mathbf{R}^3$. Usually, the measure μ is taken to be a nonnegative †Radon measure with compact †support. These potentials are †superharmonic in $\mathbf{R}^n$ and harmonic outside the support of μ. Conversely, any harmonic function defined on a domain in $\mathbf{R}^n$ can be expressed as the sum of a potential of a single layer and a potential of a double layer (defined in the next paragraph). Because of this close relation between potentials and harmonic functions, sometimes potential

theory means the study of harmonic functions (→ 193 Harmonic Functions and Subharmonic Functions). (For the representation by potentials of superharmonic functions → 193 Harmonic Functions and Subharmonic Functions S.)

Suppose that the measure μ of $\mathbf{R}^3$ satisfies $d\mu = \rho\, d\tau$ with sufficiently smooth density ρ and volume element $d\tau$. Then the Newtonian potential u of μ satisfies **Poisson's equation** $\Delta u = -4\pi\rho$. If the support of μ is contained in a surface S and $d\mu = \rho\, d\sigma$ with density ρ and surface element $d\sigma$, then the potential u of μ is called the potential of a **single layer** (or **simple distribution**). If ρ is continuous on S, then u is continuous in the whole space, and the †directional derivative of u at a point P in the direction of the normal line to S at P_0 tends to

$$-2\pi\rho(P_0) + \int_S \rho \frac{\partial}{\partial n_P} \frac{1}{\overline{PQ}} \bigg|_{P=P_0} d\sigma(Q)$$

as P approaches P_0 along the normal line. Therefore, as P moves on the line, the directional derivative jumps by $-4\pi\rho(P_0)$ at P_0. If ρ satisfies the †Hölder condition at $P_0 \in S$, then the derivative at P in the direction of any fixed tangent line at P_0 has a finite limit as P tends to P_0 along the normal line. The integral

$$u(P) = \int_S \rho \frac{\partial}{\partial n_Q} \frac{1}{\overline{PQ}} d\sigma(Q)$$

is called the potential of a **double layer** (or **double distribution**). If ρ is continuous on S, then the limits at P_0 of u from the two directions along the normal line at P_0 exist and are $2\pi\rho(P_0) + u(P_0)$ and $-2\pi\rho(P_0) + u(P_0)$. If, further, ρ is of class C^2 on S, then each partial derivative of u has a finite limit as P tends to a point of S.

B. Generalized Potential

The classical notion of potentials is generalized as follows: Let Ω be a space supplied with a measure μ ($\geqslant 0$) and $\Phi(P, Q)$ a real-valued function on the product space $\Omega \times \Omega$. When the integral $\int \Phi(P, Q)\, d\mu(Q)$ is well defined at each point $P \in \Omega$, it is called the **potential** of μ with **kernel** Φ and is denoted by $\Phi(P, \mu)$ or $\Phi\mu(P)$. The function $\check{\Phi}(P, Q) = \Phi(Q, P)$ is called the **adjoint kernel** of Φ. When

$$(\mu, \nu) = \iint \Phi(P, Q)\, d\mu(Q)\, d\nu(P) = \int \Phi(P, \mu)\, d\nu(P)$$

exists for measures $\mu, \nu \geqslant 0$, the value is called the **mutual energy** of μ and ν. In particular, (μ, μ) is called the **energy** of μ. The definition of potential given above may be too general, and some restrictions are called for. We assume that Ω is a †locally compact Hausdorff space; Φ is a †lower semicontinuous function on $\Omega \times \Omega$ satisfying $-\infty < \Phi \leqslant \infty$; and μ, ν, and λ are nonnegative Radon measures with compact support in Ω. In particular, when $\Omega = \mathbf{R}^n$, the potential with the kernel $\Phi(P, Q) = \overline{PQ}^{-\alpha}$ $(0 < \alpha < n)$ is called a **potential of order** α (sometimes of order $n - \alpha$) or a **Riesz potential**.

C. The Maximum Principle and the Continuity Principle

Let $\Omega = \mathbf{R}^n$, and let $\Phi(P, \mu)$ be the kernel of the Newtonian potential. Then $\Phi(P, \mu)$ satisfies the following principles: (1) **Frostman's maximum principle (first maximum principle)**: $\sup_{P \in \Omega} \Phi(P, \mu) \leqslant \sup_{P \in S_\mu} \Phi(P, \mu)$ for any μ, where S_μ is the support of μ. (2) **Ugaheri's maximum principle (dilated maximum principle)**: There is a constant $c \geqslant 0$ such that $\sup_{P \in \Omega} \Phi(P, \mu) \leqslant c \sup_{P \in S_\mu} \Phi(P, \mu)$ for any μ. (3) A variation of Ugaheri's maximum principle: Given any compact set $K \subset \Omega$, there exists a constant c which may depend on K such that $\sup_{P \in K} \Phi(P, \mu) \leqslant c \sup_{P \in S_\mu} \Phi(P, \mu)$ for any μ with $S_\mu \subset K$. (4) **Upper boundedness principle**: If $\Phi(P, \mu)$ is bounded from above on S_μ, then it is bounded on Ω also. (5) For any compact set $K \subset \Omega$ and any μ with $S_\mu \subset K$, if $\Phi(P, \mu)$ is bounded above on S_μ, then it is bounded on K also. (6) **Continuity principle**: If $\Phi(P, \mu)$ is continuous as a function on S_μ, then it is also continuous in Ω. Generally, the relations shown in Fig. 1 hold among principles (1)–(6), where (a)→(b) means that (a) implies (b), and (c)↛(d) means the negation of (c)→(d).

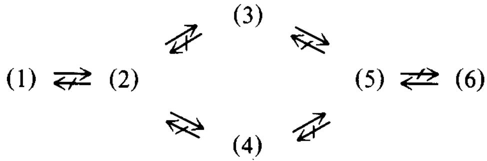


Fig. 1

If the continuity principle holds for a general kernel $\Phi(p, \mu)$ of a potential and if for any μ there is a sequence $\{P_k\}$ of points in $\Omega - S_\mu$ that has an accumulation point on S_μ and along which $\Phi(P_k, \mu) \to \sup_{\Omega - S_\mu} \Phi(P, \mu)$, then (1) holds also. The second condition is valid, for instance, when $\Omega = \mathbf{R}^n$, $\Phi(P, \mu)$ is †subharmonic in $\mathbf{R}^n - S_\mu$, and $\limsup \Phi(P, \mu) < \sup_{Q \in \mathbf{R}^n} \Phi(Q, \mu)$ as P tends to the point at infinity. T. Ugaheri, G. Choquet, and N. Ninomiya studied (2) and (6). Ugaheri showed that for any nonnegative decreasing function $\varphi(r)$ defined in $[0, \infty)$ and satisfying $\varphi(0) = \infty$, the kernel $\Phi(P, Q) = \varphi(\overline{PQ})$ satisfies (2) in $\mathbf{R}^n$. M. Ohtsuka established

that (6)→(5) and (5)↛(6) in general and proved (5)→(6) in the special case where Φ is continuous on $\Omega\times\Omega$ in the wider sense (i.e., Φ may have ∞ as its value) and Φ is finite outside the diagonal set of $\Omega\times\Omega$. The examples in Sections F, I, and J show that the potentials with kernels satisfying a weak condition such as (6) possess a number of important properties. We note that (6) does not necessarily hold in general. (For literature on (1)–(6) and other related principles → [18].)

D. The Energy Principle

Denote by E the class of all measures of finite energy. A symmetric kernel is called **positive definite** (or of **positive type**) if $(\mu-\nu,\mu-\nu)=(\mu,\mu)+(\nu,\nu)-2(\mu,\nu)\geqslant 0$ for any μ, $\nu\in E$. If the equality $(\mu-\nu,\mu-\nu)=0$ always implies $\mu=\nu$, then the kernel is said to satisfy the **energy principle**. Some characterizations for a kernel to be of positive type or to satisfy the energy principle were given by Ninomiya. Using them, he showed that a symmetric kernel that satisfies Frostman's maximum principle or the domination principle (→ Section L) and a certain additional condition is of positive type. Choquet and Ohtsuka generalized these definitions and results [18].

E. Topologies on Classes of Measures

Let C_0 be the space of continuous functions with compact support in Ω and M_0^+ be the class of measures in Ω. The †seminorms $\mu-\nu\to|\int f\,d\mu-\int f\,d\nu|$ $(f\in C_0)$ define the **vague topology** on M_0^+. The class of unit distributions can be topologized by the vague topology, which induces a topology on Ω itself. This topology coincides with the original topology in Ω. A subclass M of M_0^+ is relatively compact with respect to the vague topology if on every compact set in Ω, the values of the measures of M are bounded. Denote by L the class of measures λ such that (λ,μ) is finite for all $\mu\in M_0^+$, and define the **fine topology** on M_0^+ by the seminorms $\mu-\nu\to|(\lambda,\mu)-(\lambda,\nu)|$ $(\lambda\in L)$. This topology was introduced by H. Cartan [4] for the Newtonian kernel. It is the weakest topology that makes each $\check{\Phi}(P,\lambda)$, $\lambda\in L$, continuous. Further, when the kernel is of positive type, the **weak topology** is defined on E by the seminorms $\mu-\nu\to|(\lambda,\mu)-(\lambda,\nu)|$ $(\lambda\in E)$. The **strong topology** is defined on E by the seminorm $\sqrt{(\mu-\nu,\mu-\nu)}$.

For the Newtonian kernel, Cartan [5] showed that vague$\prec$fine$\prec$weak$\prec$strong on E, where vague$\prec$fine, for instance, means that the fine topology is stronger than the vague topology. He proved that the fine, weak, and strong convergences are equivalent for any sequence $\{\mu_n\}$ of measures with bounded energies. B. Fuglede [16] called a kernel of positive type **consistent** when any †Cauchy net with respect to the strong topology that converges vaguely to a measure converges strongly to the same measure. This notion is used to give conditions for E to be †complete with respect to the strong topology [11, 18]. Moreover, Fuglede called a consistent kernel satisfying the energy principle **perfect**, and studied the cases where †convolution kernels on a locally compact topological group are consistent or perfect [11]. For instance, $\overline{PQ}^{-\alpha}$ $(0<\alpha<n)$ in $\mathbf{R}^n$ and the Bessel kernel, which was studied by N. Aronszajn and K. T. Smith, are perfect.

F. Convergence of Sequences of Potentials

We are concerned with determining when a family of potentials corresponding to a class of measures $\{\mu_\omega\}$ with indices ω in a directed set converges. If all S_{μ_ω} are contained in a fixed compact set and μ_ω converges vaguely to μ_0, then $\liminf\Phi(P,\mu_\omega)\geqslant\Phi(P,\mu_0)$ in Ω. If, moreover, Φ is continuous in the wider sense and both Φ and $\check{\Phi}$ satisfy the continuity principle, then equality holds quasi-everywhere (q.e.) in Ω in this inequality [18]; we now define the notion q.e. First, for a nonempty compact set K in Ω, define $W(K)$ to be $\inf(\mu,\mu)$, where $S_\mu\subset K$ and $\mu(K)=1$, and $W(\varnothing)$ to be ∞ for the empty set $\varnothing$. Next, set $W_i(X)=\inf_{K\subset X}W(K)$ for an arbitrary set X in Ω and $W_e(X)=\sup_{X\subset G}W_i(G)$, where G is an open set in Ω. When a property holds except on a set X such that $W_e(X)=\infty$ (resp. $W_i(X)=\infty$), we say that the property holds **quasi-everywhere** (q.e.) (resp. **nearly everywhere** (n.e.)). The terms q.e. and n.e. are also used in the theory of †capacity, although their meaning is not the same as here. (For results on the convergence of sequences of potentials → [18].)

G. Thin Sets

A set $X\subset\Omega$ is called **thin** at P_0 when either P_0 is an isolated point of the set $X\cup\{P_0\}$ with respect to the original topology of Ω or there exists a measure μ such that $\liminf\Phi(P,\mu)>\Phi(P_0,\mu)$ as $P\in X-\{P_0\}$ tends to P_0. If P_0 is an isolated point of $X\cup\{P_0\}$ with respect to the topology weakest among those stronger than both the original topology and the fine topology, then X is thin with respect to the adjoint kernel $\check{\Phi}$. The converse is true in a special case [5]. The notion of thinness was

introduced in 1940 by M. Brelot and investigated in detail in [3]. Let $\varphi(r)$ be a positive decreasing function. Suppose there exist positive numbers r_0, δ, a such that $\varphi(r) \leqslant a\varphi((1+\delta)r)$ in $0<r<r_0$. Assume that Ω is a metric space with distance ρ, and take $\varphi(\rho(P,Q))$ as the kernel $\Phi(P,Q)$. Then a necessary and sufficient condition for X to be thin at P_0 is $\sum_{j=1}^{\infty} s^j/W_e(X_j) < \infty$ $(s>1)$ [19], where $X_j = \{P \in X \mid s^j \leqslant \varphi(\rho(P,P_0)) \leqslant s^{j+1}\}$. This criterion was obtained by N. Wiener in 1924 and utilized to give a condition for a boundary point P_0 of a domain D in $\mathbf{R}^3$ to be †regular with respect to the †Dirichlet problem for D. In this situation, P_0 is regular if and only if the complement of D is thin at P_0. When every compact subset of X is thin at P_0, X is called **internally thin** at P_0. A necessary and sufficient condition for X to be internally thin at P_0 is $\sum s^j/W_i(X_j) < \infty$.

H. Polar Sets

Brelot (1941) called a set A **polar** when there is a measure μ for which $\Phi(P,\mu)=\infty$ on A. Consider $\check{U}(\Omega,K)$ $(=V(K,\Omega))$ as defined in 48 Capacity C, and define $\check{U}_e(\Omega,X)$ by $\sup_{G \supset X} \inf_{K \subset G} \check{U}(\Omega,K)$ for an arbitrary set X, where G is an open set. Then for any μ, $X = \{P \mid \Phi(P,\mu)=\infty\}$ is a G_δ set for which the value of $\check{U}_e(\Omega,X)$ is infinite. Conversely, given a G_δ set A of †Newtonian outer capacity zero in $\mathbf{R}^n$ $(n \geqslant 3)$, there is a measure μ such that the set of points where the Newtonian potential of μ is equal to ∞ coincides with A and $\mu(\mathbf{R}^n - A)=0$ (Choquet [7]). This result is called **Evans's theorem** (or the **Evans-Selberg theorem**) (→ 48 Capacity) in the special case where A is compact.

I. Quasicontinuity

A function f in Ω is called **quasicontinuous** if there is an open subset G in Ω of arbitrarily small capacity such that the restriction of f to $\Omega - G$ is continuous. Naturally, quasicontinuity depends on the definition of capacity. Suppose that whenever the potential of a measure μ with kernel Φ is continuous as a function on S_μ it is quasicontinuous in Ω; Φ is then said to satisfy the **quasicontinuity principle**. Assuming in addition that Φ is positive symmetric and taking $1/U(\Omega,G)$ as the capacity of G, we find that every potential is quasicontinuous in Ω (M. Kishi). A similar result is valid for a nonsymmetric kernel if the continuity principle is assumed (Choquet [6]).

J. The Gauss Variational Problem

Given a compact set K and a function f on K, the problem of minimizing the **Gauss integral** $(\mu,\mu) - 2\int f\,d\mu$ for a measure μ such that $S_\mu \subset K$ is called the **Gauss variational problem**. When an additional condition is imposed on μ, the problem is called conditional. Among many results obtained for this problem [18], the following is typical: If Φ is symmetric, K supports a nonzero measure of finite energy, and f is finite upper semicontinuous on K, then there exists a μ such that $f(P) \leqslant \Phi(P,\mu)$ n.e. on K and $f(P) \geqslant \Phi(P,\mu)$ on S_μ. When Φ is not symmetric, the same relations hold for some μ if Φ is positive and $\check{\Phi}$ satisfies the continuity principle (Kishi [15]), although the method using Gauss variation is not applicable. When Φ is symmetric and of positive type, the Gauss integral with $f(P)=\Phi(P,\nu)$ $(\nu \in E)$ is equal to $\|\mu-\nu\|^2 - \|\nu\|^2$, and the minimizing problem is equivalent to finding the projection of ν to $\{\mu \in E \mid S_\mu \subset K\}$. In some cases, this projection is equal to the measure obtained by the balayage of ν to K (→ Section L).

K. Equilibrium Mass Distributions

A unit measure μ supported by a compact set K is called an **equilibrium mass distribution** on K if $\Phi(P,\mu)$ is equal to a constant a n.e. on K and $\Phi(P,\mu) \leqslant a$ in Ω. The kernel is said to satisfy the **equilibrium principle** if there exists an equilibrium mass distribution on every compact set. If $a>0$, $1/a$ can be regarded as a kind of capacity, and μ/a is called a **capacitary mass distribution**. Corresponding to †inner and outer capacities, inner and outer capacitary mass distributions and their coincidence can be discussed [11]. When Φ is symmetric, Frostman's maximum principle is equivalent to the equilibrium principle.

L. The Sweeping-Out Principle

A kernel is said to satisfy the **sweeping-out principle** (or **balayage principle**) if for any compact set K and measure μ there exists a measure ν supported by K such that $\Phi(P,\nu) = \Phi(P,\mu)$ n.e. on K and $\Phi(P,\nu) \leqslant \Phi(P,\mu)$ in Ω. When we find such a ν, we say that we **sweep out** μ to K, and finding ν is called a **sweeping-out process** (or **balayage**). This terminology originated in the classical process for the Newtonian potential in which the exterior of K is covered by a countable number of balls and the masses inside the balls are repeatedly swept out onto the spherical surfaces. For any general kernel, the balayage principle implies

the **domination principle** (also called **Cartan's maximum principle**), which asserts that if the inequality $\Phi(P,\mu)\leqslant\Phi(P,\nu)$ is valid on S_μ for $\mu\in E$ and any ν, then the same inequality holds in Ω. The converse is true if Φ is positive, symmetric, continuous in the wider sense, and finite outside the diagonal set. In contrast to the domination principle, Φ is said to satisfy the **inverse domination principle** if the inequality $\Phi(P,\mu)\leqslant\Phi(P,\nu)$ on S_ν for $\mu\in E$ and any ν implies the same inequality in Ω. In a special case, the domination principle implies Frostman's maximum principle [17]. For the equilibrium and domination principles for non-symmetric kernels → [14].

Corresponding to inner and outer capacitary mass distributions, we can examine inner and outer balayage mass distributions and inner and outer Gauss variational problems and their coincidences [5, 18]. With respect to the Newtonian potential, a point P is called an **internally (externally) irregular point** of X if the inner (outer) balayage mass distribution to X of the unit measure ε_P at P is different from ε_P. X is thin (internally thin) at P if and only if P is an externally (internally) irregular point of X (Cartan [5]).

M. Other Principles

A kernel Φ is said to satisfy the **uniqueness principle** if $\mu\equiv\nu$ follows from the equality $\Phi(P,\mu)=\Phi(P,\nu)$, which is valid n.e. in Ω. Ninomiya and Kishi studied this principle. A kernel Φ is said to satisfy the **lower envelope principle** if given μ and ν, there is a λ such that $\Phi(P,\lambda)=\min(\Phi(P,\mu),\Phi(P,\nu))$. If Φ satisfies the domination principle and $\check{\Phi}$ satisfies the continuity principle, then Φ satisfies the lower envelope principle on every compact set considered as a space. Conversely, if Φ satisfies the lower envelope principle on every compact set considered as a space, then Φ satisfies the domination principle or the inverse domination principle under some additional conditions (Kishi). A kernel is said to satisfy the **complete maximum principle** if the inequality $\Phi(P,\mu)\leqslant\Phi(P,\nu)+a$ on S_μ implies the same inequality in Ω for $\mu\in E$, any ν, and $a\geqslant0$ (Cartan and J. Deny, 1950). This principle implies both Frostman's maximum principle and the domination principle. Potentials of order α $(n-2\leqslant\alpha<n)$ in $\mathbf{R}^n$ and the **Yukawa potential** with kernel $a\overline{PQ}^{-1}\exp(-\lambda\overline{PQ})$ in $\mathbf{R}^3$ satisfy the complete maximum principle. Relations between this principle and some other principles were studied by Kishi [14].

While all principles discussed so far are global, Choquet and Ohtsuka made a local study.

N. Diffusion Kernels

By means of the bilinear form $\int f\,d\mu$ ($f\in C$, $\mu\in M_0$), we now introduce weak topologies in the space C of continuous functions in Ω and in the class M_0 of Radon measures of general sign with compact support. Similarly, we introduce weak topologies in C_0 and in the class M of measures of general sign with not necessarily compact support. A positive linear mapping G of M_0 into M that is continuous with respect to these two weak topologies is called a **diffusion kernel**. The linear mapping G^* of C_0 into C that is determined by $\int f\,dG\mu=\int G^*f\,d\mu$ is called a **transposed mapping**. We can define the balayage principle for G and the domination principle for G^* as in the case where kernels are functions. Then G satisfies the balayage principle if and only if G^* satisfies the domination principle (Choquet and Deny). The complete maximum principle has been defined and studied for G^* (Deny). G. A. Hunt obtained a relation between this principle and the representation of G^*f in the form $\int_0^\infty P_t f\,dt$ with a †semigroup P_t. His result is important in the theory of †stochastic processes (→ 261 Markov Processes).

O. Convolution Kernels

A diffusion kernel G on a locally compact Abelian group induces a convolution kernel κ if G is translation-invariant. It is called a **Hunt kernel** when there exists a vaguely continuous semigroup $\{\alpha_t\}_{t\geqslant0}$ such that $\kappa=\int_0^\infty\alpha_t\,dt$ and $\alpha_0=\varepsilon_0$, the Dirac measure at the origin. A Hunt kernel satisfies the domination principle and the balayage principle to all open sets, and it satisfies the complete maximum principle if and only if $\{\alpha_t\}_{t\geqslant0}$ is sub-Markovian, $\int d\alpha_t\leqslant1$. The Fourier transform of such a semigroup has a closed connection with a negative-definite function [1].

For a convolution kernel κ satisfying the domination principle, or, equivalently, the balayage principle, the inequality

$$\kappa(\omega_1)\kappa(\omega_2)\leqslant\kappa(\omega_1-\omega_1)\kappa(\omega_1+\omega_2)$$

is valid for all relatively compact open sets ω_1 and ω_2 [8]. It has a unique decomposition $\kappa=\varphi\cdot(\kappa_0+\kappa_1)$, where φ is a continuous exponential function, κ_0 is equal to 0 or a Hunt kernel satisfying the complete maximum principle, and κ_1 is a singular kernel satisfying the domination principle such that $\kappa_1*\varepsilon_x=\kappa_1$ for every $x\in S_{\kappa_0}$ [12].

P. Potentials with Distribution Kernels

A function f in $\mathbf{R}^n$ is called **slowly increasing in the sense of Deny** if there exists a positive

integer q such that $\int f(P)(1+\overline{OP}^2)^{-q}d\tau(P)<\infty$. A †distribution K is called a **distribution kernel** if the †Fourier transform FK of K is a function $k\geqslant 0$ and both k and $1/k$ are slowly increasing in the sense of Deny. Given such a distribution K, the family W of distributions T for which FT are functions and $\|T\|^2=\int k(FT)^2\,d\tau<\infty$ (this is called the energy of T) is a †Hilbert space with inner product $(T_1, T_2)=\int kFT_1FT_2\,d\tau$. However, the family of Newtonian potentials of measures with finite energy is not a Hilbert space (Cartan). The family of functions of class C^∞ with compact support is †dense in W. For every $T\in W$ the function $FK\times FT$ is slowly increasing in the sense of Deny. There exists a distribution $U=U^T$ that satisfies $FU=FK\times FT$, called the K-potential of T. Since W is complete, the method of projection is applicable, and problems of equilibrium, balayage, and capacity can be examined. For instance, if †Dirac's distribution δ is taken as a distribution kernel, then the corresponding capacity is the Lebesgue measure. In the case of the Newtonian kernel, $\partial U^T/\partial x_j=f_j$ is defined a.e. in $\mathbf{R}^n$ for any $T\in W$. These f_j are square integrable, $T=-c_n\sum_{j=1}^n \partial f_j/\partial x_j$, and $\|T\|^2=c_n\sum_{j=1}^n\int|f_j|^2\,d\tau$, where $1/c_n=2(n-2)\pi^{n/2}/\Gamma(n/2)$. Furthermore,

$$U^T=(n-2)\int\sum_{j=1}^n(x_j-y_j)f_j(Q)\overline{PQ}^{-n}\,d\tau(Q),$$

where x_j, y_j are components of P, Q, respectively. Every ordinary potential of a double layer is a special case of U^T. Conversely, let f be a function in $\mathbf{R}^n$ that is absolutely continuous along almost every line parallel to each coordinate axis and whose partial derivatives are square integrable. Then f is equal to the potential of some $T\in W$ with Newtonian kernel up to an additive constant. These results are due to Deny [9].

Q. Dirichlet Spaces

In this section, functions are assumed to be complex-valued. Let Ω be a locally compact Hausdorff space, $\xi\geqslant 0$ a Radon measure in Ω, and C_0 the space of continuous functions with compact support. A Hilbert space D consisting of locally ξ-integrable functions is called a **Dirichlet space** if $C_0\cap D$ is dense in both C_0 and D, the relations $|v(P)-v(Q)|\leqslant|u(P)-u(Q)|$ and $|v(P)|\leqslant|u(P)|$ for $u\in D$ and a function v always imply $v\in D$ and $\|v\|\leqslant\|u\|$, and for any compact set $K\subset\Omega$, there exists a constant $A(K)$ such that $\int_K|u|\,d\xi\leqslant A(K)\|u\|$ for every $u\in D$. The notion of Dirichlet space was introduced by A. Beurling. A function $u\in D$ is called a potential if there exists a Radon measure μ such that $(u,\varphi)=\int\overline{\varphi}\,d\mu$ holds for every $\varphi\in C_0\cap D$. If in addition $\mu\geqslant 0$, then u is called a pure potential. For any pure potential, the lower envelope principle, the equilibrium principle, the balayage principle, and the complete maximum principle hold [2]. Suppose that Ω is a locally compact Abelian group and D is a Dirichlet space such that $U_yu(x)=u(x-y)\in D$ and $\|U_yu\|=\|u\|$ for every $u\in D$ and $y\in\Omega$. Then we call D special and characterize it in terms of a real-valued continuous function on Ω [2]. (For axiomatic potential theory → 193 Harmonic Functions and Subharmonic Functions.)

References

[1] C. Berg and G. Forst, Potential theory on locally compact Abelian groups, Springer, 1975.
[2] A. Beurling and J. Deny, Dirichlet spaces, Proc. Nat. Acad. Sci. US, 45 (1959), 208–215.
[3] M. Brelot, Sur les ensembles effilés, Bull. Sci. Math., (2) 68 (1944), 12–36.
[4] M. Brelot, Eléments de la théorie classique du potentiel, Centre de Documentation Universitaire, third edition, 1965.
[5] H. Cartan, Théorie généale du balayage en potentiel newtonien, Ann. Univ. Grenoble (N.S.), 22 (1946), 221–280.
[6] G. Choquet, Sur les fondements de la théorie fine du potentiel, C. R. Acad. Sci. Paris, 244 (1957), 1606–1609.
[7] G. Choquet, Potentiels sur un ensemble de capacité nulle, Suites de potentiels, C. R. Acad. Sci. Paris, 244 (1957), 1707–1710.
[8] G. Choquet and J. Deny, Noyaux de convolution et balayage sur tout ouvert, Lecture notes in math. 404, Springer, 1974, 60–112.
[9] J. Deny, Les potentiels d'énergie finie, Acta Math., 82 (1950), 107–183.
[10] O. Frostman, Potentiel d'équilibre et capacité des ensembles avec quelques applications à la théorie des fonctions, Medd. Lunds Univ. Mat. Sem., 3 (1935).
[11] B. Fuglede, On the theory of potentials in locally compact spaces, Acta Math., 103 (1960), 139–215.
[12] M. Itô, Caractérisation du principe de domination pour les noyaux de convolution non-bornés, Nagoya Math. J., 57 (1975), 167–197.
[13] O. D. Kellogg, Foundations of potential theory, Springer, 1929.
[14] M. Kishi, Maximum principles in the potential theory, Nagoya Math. J., 23 (1963), 165–187.
[15] M. Kishi, An existence theorem in potential theory, Nagoya Math. J., 27 (1966), 133–137.

[16] N. S. Landkof, Foundations of modern potential theory, Springer, 1972.
[17] N. Ninomiya, Etude sur la théorie du potentiel pris par rapport au noyau symétrique, J. Inst. Polytech. Osaka City Univ., (A) 8 (1957), 147–179.
[18] M. Ohtsuka, On potentials in locally compact spaces, J. Sci. Hiroshima Univ., 25 (1961), 135–352.
[19] M. Ohtsuka, On thin sets in potential theory, Sem. Anal. Funtions, Inst. for Adv. Study, Princeton, 1957, vol. 1, p. 302–313.

339 (XI.2) Power Series

A. General Remarks

Let a and $c_0, c_1, c_2, \ldots$ be elements of a †field K and z be a variable. A series of the form $P=\sum_{n=0}^{\infty} c_n(z-a)^n$ is called a **power series** (in one variable). We assume that K is the field of complex numbers. For a given power series P, we can determine a unique real number R $(0 \leqslant R \leqslant \infty)$ such that P converges if $|z-a|<R$ and diverges if $R<|z-a|$. We call R the **radius of convergence** and the circle $|z-a|=R$ (sometimes $|z-a|<R$) the **circle of convergence** of P. The value of R is given by $R=1/\limsup \sqrt[n]{|c_n|}$ (**Cauchy-Hadamard formula**) with the conventions $0=1/\infty$, $\infty=1/0$. Also $R=\lim|c_n/c_{n+1}|$, provided that the limit on the right-hand side exists.

A power series †converges absolutely and uniformly on every compact subset inside its circle of convergence and defines there a single-valued complex function. Since the series is †termwise differentiable, the function is actually a holomorphic function of a complex variable. Conversely, any function $f(z)$ holomorphic in a domain can be represented by a power series in a neighborhood of each point a of the domain. Such a representation is called the **Taylor expansion** of $f(z)$ at a (or in the neighborhood of a). A power series that represents a holomorphic function is called a holomorphic **function element**. K. Weierstrass defined an analytic function as the set of all elements that can be obtained by †analytic continuations starting from a given function element (→ 198 Holomorphic Functions).

Besides the series $\sum_{n=0}^{\infty} c_n(z-a)^n$, a series of the form $Q=\sum_{n=0}^{\infty} c_n z^{-n}$ is called a **power series with center at the point at infinity**, and its value at ∞ is defined to be c_0. By putting $z-a=t$ when its center a is a finite point and $z^{-1}=t$ when its center is ∞, every power series can be written in the form $\sum_{n=0}^{\infty} c_n t^n$, and such a t is called a **local canonical parameter**.

When t is a local canonical parameter, a series of the form $\sum_{n=-\infty}^{\infty} c_n t^n$ is called a **Laurent series** and a series $\sum_{n=-\infty}^{\infty} c_n t^{n/k}$ (k a fixed natural number) is called a **Puiseux series**, after the French mathematicians A. Laurent and V. A. Puiseux. Power series are sometimes called **Taylor series**.

If we perform †analytic continuations of a power series from its center along radii of its circle of convergence, we encounter a †singularity on the circumference for at least one radius. For a power series with the circle of convergence $|z|=R$, the argument α of the singularity on $|z|=R$ nearest $z=R$ is given in the following way. Suppose, for simplicity, that the radius of convergence R of $\sum c_n z^n$ equals 1, and put

$$\rho_n(h)=\sum_{\nu=0}^{n}\binom{n}{\nu} c_{n-\nu} h^{\nu},$$

$$P(h)=\limsup_{n\to\infty} \sqrt[n]{|\rho_n(h)|}.$$

Then α is obtained from

$$\cos\alpha=P'_+(0)=\lim_{h\to+0}(P(h)-1)/h$$

(S. Mandelbrojt, 1937). In particular, if all the c_n are real and nonnegative, $z=R$ is a singularity (Vivanti's theorem).

B. Abel's Continuity Theorem

As a property of the power series on the circle of convergence, we have **Abel's continuity theorem**: If the radius of convergence of $f(z)=\sum_{n=0}^{\infty} a_n z^n$ is equal to 1 and $\sum_{n=0}^{\infty} a_n$ coverges (or is †(C,k)-summable $(k>-1)$) to A, then $f(z)\to A$ when z approaches 1 in any sector $\{z \,|\, |z|<1, |\arg(1-z)|<(\pi/2)-\delta\}$, $\delta>0$ (†Stolz's path).

The converse of this theorem is not always true. The existence of $\lim_{z\to 1} f(z)$ does not necessarily lead to the convergence of $\sum_{n=0}^{\infty} a_n$. Even Cesàro summability of $\sum_{n=0}^{\infty} a_n$ does not always follow from the existence of $\lim_{z\to 1} f(z)$. If $a_n=o(1/n)$ and $f(z)\to A$ when z approaches 1 along a curve ending at $z=1$, then $\sum a_n$ converges to A (**Tauber's theorem**, 1897). The theorems concerning additional sufficient conditions for the validity of the converse of Abel's theorem are called **theorems of Tauberian type** (or **Tauberian theorems**). In Tauber's theorem, the hypothesis on the a_n may be replaced by $a_n=O(1/n)$ or $n\operatorname{Re}a_n$, $n\operatorname{Im}a_n$ may be bounded from above but not necessarily from below (G. H. Hardy and J. E. Littlewood). Here, the condition $a_n=O(1/n)$ cannot be weakened (Littlewood). Sufficient condi-

tions for $\sum a_n$ to be †summable for various summation methods are also known.

C. Lambert Series

A series of the form

$$\sum_{n=1}^{\infty} a_n \frac{z^n}{1-z^n}, \quad |z| \neq 1, \tag{1}$$

is called a **Lambert series.** If $\sum a_n$ is convergent, (1) converges for any z with $|z| \neq 1$, and moreover, it converges uniformly on any compact set contained in $|z|<1$ or $|z|>1$. If $\sum a_n$ is divergent, (1) and the power series $\sum a_n z^n$ converge or diverge simultaneously for $|z| \neq 1$.

There is a detailed study of Lambert series by K. Knopp (1913). If R is the radius of convergence of $\sum a_n z^n$ and $\sum_{d|n} a_d = A_n$ is the sum extending over all divisors of n, then we have the reciprocity relation

$$\sum_{n=1}^{\infty} a_n \frac{z^n}{1-z^n} = \sum_{n=1}^{\infty} A_n z^n,$$

which holds for $|z|<\min(R, 1)$. As special cases of this relation, we have

$$z = \sum_{n=1}^{\infty} \mu(n) \frac{z^n}{1-z^n},$$

$$\frac{z}{(1-z)^2} = \sum_{n=1}^{\infty} \varphi(n) \frac{z^n}{1-z^n},$$

where μ and φ are the †Möbius function and †Euler function, respectively.

If the na_n are real and bounded from below,

$$\lim_{z \to 1-0} \sum_{n=1}^{\infty} (1-z) n a_n \frac{z^n}{1-z^n} = s$$

implies $\sum a_n = s$. Hardy and Littlewood (1921) showed that this theorem of Tauberian type is equivalent to the †prime number theorem.

D. Singularities of Power Series

Given a power series $P = \sum a_n z^n$, if the †branch in $|z|<R^*$ of the analytic function $f(z)$ determined by P is single-valued meromorphic but the branch in $|z|<R'$ with $R'>R^*$ has singularities other than poles, then R^* is called the **radius of meromorphy** of P and $|z|=R^*$ (sometimes $|z|<R^*$) is called the **circle of meromorphy** of P. R^* can be computed in the following way. Put

$$l_p = \limsup_{n \to \infty} |\sqrt[n]{D_n^{(p)}}|,$$

$$D_n^{(p)} = \begin{vmatrix} a_n & a_{n+1} & \cdots & a_{n+p} \\ a_{n+1} & a_{n+2} & \cdots & a_{n+p+1} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n+p} & a_{n+p+1} & \cdots & a_{n+2p} \end{vmatrix}.$$

Then the sequence of numbers l_{p-1}/l_p is increasing, and $\lim l_{p-1}/l_p = R^*$. If $R_1, R_2, \ldots$ are different valued of l_{p-1}/l_p, then $f(z)$ is holomorphic at points not lying on $|z|=R_n$ (**Hadamard's theorem**, 1892).

When a point a and a set A of points in the complex plane are given, the set of points that can be joined with a by a segment disjoint from A is called the **star region** determined by a and A. Take any half-line starting at a; the point of A lying on it and nearest to a is called a **vertex**. For a power series $\sum c_n z^n$, the set of centers of the function elements obtained by analytic continuations along half-lines starting at the origin is called the star region of $\sum c_n z^n$ with respect to the origin. Let $\{\alpha\}$ and $\{\beta\}$ be the set of vertices of the star regions with respect to the origin of $\sum a_n z^n$ and $\sum b_n z^n$, respectively. Then the star region determined by the origin and the set $\{\alpha\beta\}$ is contained in the star region of $\sum a_n b_n z^n$ (**Hadamard's multiplication theorem**, 1892).

The following are some results concerning conditions for the coincidence of the circle of convergence of a power series and its †natural boundary. Let the a_n be positive numbers and b a natural number greater than 1. If the radius of convergence of $\sum_{n=0}^{\infty} a_n z^{b^n}$ is equal to 1, then $|z|=1$ is its natural boundary (Weierstrass, E. I. Fredholm). If the radius of convergence of $\sum_{n=0}^{\infty} a_n z^{\lambda_n}$ (with $\{\lambda_n\}$ an increasing sequence of natural numbers) is 1 and $\liminf_{n \to \infty} (\lambda_{n+1}-\lambda_n)/\lambda_n > 0$, then $|z|=1$ is the natural boundary (**Hadamard's gap theorem**). The latter condition was weakened to $\liminf_{n \to \infty} (\lambda_{n+1}-\lambda_n)/\sqrt{\lambda_n} > 0$ by E. Borel (1896). E. Fabry (1896) showed that with the radius of convergence of $\sum_{n=0}^{\infty} a_n z^n$ being 1, if there exist a suitable sequence of natural numbers $m_1 < m_2 < \ldots$ and a number θ $(0<\theta<1)$ such that $\lim s_i/m_i = 0$, where the s_i are the number of nonzero a_n contained in the interval $(m_i(1-\theta), m_i(1+\theta))$, then $|z|=1$ is the natural boundary of $\sum a_n z^n$. By applying this theorem to $\sum a_n z^{\lambda_n}$ with radius of convergence 1, it can be shown that if $\lim_{n \to \infty} \lambda_n/n = \infty$, then $|z|=1$ is its natural boundary. These theorems are called **gap theorems** because they concern power series with gaps in their exponents. A generalization of Fabry's theorem was obtained by G. Pólya. It is known that Fabry's last condition above is in a sense the best possible (Pólya, 1942).

Regarding the natural boundary of a power series, we also have the following result: When the radius of convergence of $\sum a_n z^n$ is 1, by a suitable choice of the sequence $\{\varepsilon_n\}$ $(\varepsilon_n = \pm 1)$, the series $\sum \varepsilon_n a_n z^n$ has $|z|=1$ as its natural boundary (A. Hurwitz, P. Fatou, Pólya).

E. Overconvergence

If the radius of convergence of $f(z)=\sum a_n z^n$ is 1, the sequence of partial sums $S(1,z)$, $S(2,z)$, ..., where $S(n,z)=\sum_{v=0}^{n} a_v z^v$, is naturally divergent for $|z|>1$, but a suitable subsequence $S(n_k,z)$ $(k=1,2,\dots)$ may still be convergent. A. Ostrowski [7, 8] called this phenomenon **overconvergence** and proved the following result: By definition, $f(z)=\sum_{n=0}^{\infty} a_n z^{\lambda_n}$ has a **lacunary structure** if the sequence $\{\lambda_n\}$ has a subsequence $\{\lambda_{n_k}\}$ such that $\lambda_{n_k+1}>\lambda_{n_k}(1+\theta)$ $(\theta>0)$. If this situation occurs, then in a sufficiently small neighborhood of a point on $|z|=1$ where $f(z)$ is holomorphic, $S(\lambda_{n_k},z)$ $(k=1,2,\dots)$ converges uniformly. This result includes Hadamard's gap theorem as a special case. Conversely, any power series for which overconvergence takes place can be represented as the sum of a power series having a lacunary structure and a power series whose radius of convergence is greater than 1. G. Bourion [9] gave a unified theory of these results using †superharmonic functions.

R. Jentsch [10] showed that all singularities of a power series on its circle of convergence are accumulation points of the zeros of the partial sums. On the other hand, if the zeros of a subsequence $S(n_k,z)$ $(k=1,2,\dots)$ has no accumulation point on $|z|=1$, then the power series has a lacunary structure and overconvergence takes place for $S(n_k,z)$ $(k=1,2,\dots)$. If $\log n_{k+1}=O(n_k)$ and $S(n_k,z)$ $(k=1,2,\dots)$ is overconvergent, then all boundary points of the domain of overconvergence are accumulation points of the zeros of $S(n_k,z)$ (Ostrowski [8]).

A power series is completely determined by its coefficients, but little is known about the relations between the arithmetical properties of its coefficients and the function-theoretic properties of the function represented by the series. A known result is that if the power series $\sum c_n z^n$ with rational coefficients represents a branch of an †algebraic function, then we can find an integer γ such that the $c_n\gamma^n$ $(n\geqslant 1)$ are all integers (Eisenstein's theorem, 1852).

For power series of several variables → 21 Analytic Functions of Several Complex Variables; for formal power series → 370 Rings of Power Series. For power series expansions → Appendix A, Table 10.IV.

References

For the general theory of power series,
[1] E. G. H. Landau, Darstellung und Begründung einiger neuerer Ergebnisse der Funktionentheorie, Springer, second edition, 1929.
[2] P. Dienes, The Taylor series, Clarendon Press, 1931.
[3] L. Bieberbach, Analytische Fortsetzung, Springer, 1955.
Also → references to 198 Holomorphic Functions.
For singularities,
[4] S. Mandelbrojt, Les singularités des fonctions analytiques représentées par une série de Taylor, Gauthier-Villars, 1932.
For theorems of Tauberian type,
[5] N. Wiener, Tauberian theorems, Ann. Math., (2) 33 (1932), 1–100.
[6] H. R. Pitt, Tauberian theorems, Oxford Univ. Press, 1958.
For overconvergence,
[7] A. Ostrowski, Über eine Eigenschaft gewisser Potenzreihen mit unendlichvielen verschwinden der Koeffizienten, S.-B. Preuss. Akad. Wiss., (1921), 557–565.
[8] A. Ostrowski, Über vollständige Gebiete gleichmässiger Konvergenz von Folgen analytischer Funktionen, Abh. Math. Sem. Univ. Hamburg, 1 (1922), 327–350.
[9] G. Bourion, L'ultraconvergence dans les séries de Taylor, Actualités Sci. Ind., Hermann, 1937.
[10] R. Jentsch, Fortgesetzte Untersuchungen über die Abschnitte von Potenzreihen, Acta Math., 41 (1918), 253–270.

340 (XVII.17) Probabilistic Methods in Statistical Mechanics

A. Introduction

Probabilistic methods are often very useful in the rigorous treatment of the mathematical foundations of statistical mechanics and also in some other problems related to statistical mechanics. As examples of such methods, we explain here (1) Ising models, (2) Markov statistical mechanics, (3) percolation processes, (4) random Schrödinger equations, and (5) the †Boltzmann equation.

B. Ising Models

The **Ising model** was proposed by E. Ising [1] to explain the **phenomena of phase transitions** of a ferromagnet, in which either a + or − spin is put on each site of a crystal lattice, and interaction between nearest neighboring sites is taken into consideration.

Let V be a cube in the d-dimensional integer lattice space $\mathbf{Z}^d$ and $X_V=\{+1,-1\}^V$ be the totality of spin configurations in V. Each ele-

ment of X_V is denoted by $\sigma=\{\sigma_i\}_{i\in V}$ ($\sigma_i=+1$ or -1). We suppose that a spin configuration σ has a potential of the type

$$U_V(\sigma)=-\sum_{\langle i,j\rangle\subset V}\sigma_i\sigma_j+h\sum_{i\in V}\sigma_i,$$

where $\langle i,j\rangle$ means that (i,j) is a nearest neighboring pair of sites and h stands for the parameter of an external field. A †probability measure on X_V is called a **state** on V. For each state μ, the **free energy** is defined by

$$F_V(\mu)=\sum_{\sigma\in X_V}\mu(\sigma)U_V(\sigma)+\frac{1}{\beta}\sum_{\sigma\in X_V}\mu(\sigma)\log\mu(\sigma),$$

where $\beta=1/kT$ (k is the †Boltzmann constant, T is the †absolute temperature). Then there exists a unique state

$$g_V^{\beta,h}(\sigma)=\frac{\exp(-\beta U_V(\sigma))}{\sum_{\gamma\in X_V}\exp(-\beta U_V(\gamma))},\qquad \sigma\in X_V,$$

on V which minimizes the free energy F_V (**variational principle**). $g_V^{\beta,h}$ is called a **Gibbs state** on V of the Ising model with parameter (β,h). Physically, a Gibbs state is an **equilibrium state**, in which various physical quantities are calculated.

Since for each $\mathbf{Z}^d$-homogeneous (i.e., translationally invariant with respect to $\mathbf{Z}^d$) state μ the **mean free energy**

$$f(\mu)=\lim_{V\to\mathbf{Z}^d}\frac{1}{|V|}F_V(\mu)$$

is well defined, a (limiting) Gibbs state can also be defined for the infinite domain $V=\mathbf{Z}^d$ by the above mentioned variational principle. However, at present the probabilistic definition of Gibbs states given by Dobrushin [2] and Lanford and Ruelle [3] is prevalent.

It is known that if an external field is present (i.e., $h\neq 0$) there is only one Gibbs state for any β. On the other hand when an external field is absent ($h=0$) and $d\geqslant 2$, there are at least two Gibbs states, i.e., a phase transition occurs, for a sufficiently low temperature.

Finally, we mention some known facts in this field. In the following we assume $h=0$. (1) In the 1-dimensional case, the phase transition never occurs. (2) For $d\geqslant 2$, there exists a **critical value** $\beta_c(d)$ such that the phase transition does not occur for any $\beta<\beta_c(d)$ but it occurs for every $\beta>\beta_c(d)$. The calculation of $\beta_c(2)$ has been carried out by Onsager [4]. (3) In the 2-dimensional case, every Gibbs state is $\mathbf{Z}^2$-homogeneous [5, 6]. (4) For $d\geqslant 3$, there is a $\mathbf{Z}^d$-inhomogeneous Gibbs state for sufficiently large β [7].

C. Markov Statistical Mechanics

Stochastic Ising models, infinite interacting particle systems, and many models occurring in physics, biology, and sociology are formulated as a class of infinite-dimensional †Markov processes. The field of investigation of stationary states and statistical or ergodic properties of these processes is called **Markov statistical mechanics**, which has made rapid progress during the last decade. We explain this field by looking at a typical class of processes.

Let $\mathbf{Z}^d$ be the d-dimensional integer lattice space. Putting + or − spin on each site on $\mathbf{Z}^d$, let us consider a random motion of spins which evolves while interacting with neighboring spins. Let $X=\{+1,-1\}^{\mathbf{Z}^d}$ be the totality of spin configurations and an element of X be denoted by $\eta=\{\eta(i)\}_{i\in\mathbf{Z}^d}$ ($\eta(i)=+1$ or -1). The process is described in terms of a collection of nonnegative functions $c_i(\eta)$ defined for $i\in\mathbf{Z}^d$ and $\eta\in X$. For the configuration η_t at time t, $\eta_t(i)$ changes to $-\eta_t(i)$ in the time interval $[t,t+\Delta t]$ with probability $c_i(\eta_t)\Delta t+o(\Delta t)$. This process on X is called a spin-flip model. For an initial distribution μ we denote by μ_t the distribution at time t. If $\mu_t=\mu$ for all $t\geqslant 0$, μ is called a **stationary state**.

Example 1. Stochastic Ising models. A **stochastic Ising model** was proposed by Glauber [8] to describe the random motion in a ferromagnet upon contact with a heat bath. Then the flip rate $\{c_i(\eta)\}$ is defined by the potential of the Ising model. It is known that any Gibbs state of the Ising model is a reversible stationary state of the stochastic Ising model, and the converse is also valid. Free energy plays an important role in the study of the ergodic properties of these models. In particular, the mean free energy is a nondecreasing functional along the distributions μ_t, $0\leqslant t<\infty$.

Example 2. Contact processes. A **contact process** was introduced by Harris [11] to investigate the spread of infection. The flip rate of the contact process is given by

$$c_i(\eta)=\begin{cases}1 & \text{if}\quad \eta(i)=+1,\\ k\lambda & \text{if}\quad \eta(i)=-1\quad\text{and}\\ & \qquad\#\{j\mid |j-i|=1,\quad \eta(j)=+1\}=k.\end{cases}$$

Here +1 denotes an infected individual and −1 denotes a healthy one. Denote by $-\mathbf{1}$ the configuration at which all sites are healthy and by $\delta_{-\mathbf{1}}$ the unit point mass at $-\mathbf{1}$; then $\delta_{-\mathbf{1}}$ is a stationary state. The most important result is the following: There exists a **critical value** λ_d ($0<\lambda_d<\infty$) such that if $\lambda<\lambda_d$ $\delta_{-\mathbf{1}}$ is a unique stationary state, and if $\lambda>\lambda_d$ there is another stationary state μ satisfying

$$\mu[\eta\in X;\ \eta(i)=+1 \text{ for infinitely many } i]=1.$$

D. Percolation Processes

A **percolation process** is a mathematical model which describes the random spread of a fluid

through a medium. It can be used to describe phenomena such as the penetration through a porous solid by a liquid or the spread of an infectious disease [14]. Usually, the process is identified as a **site percolation process** or a **bond percolation process**. Here we describe the latter only.

Let $L=\{S,B\}$ be a countable connected †graph with a set of sites (†vertices) S and a set of bonds (†edges) B. Each bond b is open with probability p and closed with probability $1-p$ independently of all other bonds. Suppose that a fixed point o is the source of a fluid which flows from o along open bonds only.

Let $\theta(p)$ be the probability that the fluid spreads infinitely far, and define the **critical percolation probability** $p_H=\inf\{p \mid \theta(p)>0\}$. Then it is known that (1) $p_H=1/2$ for the square lattice [16], (2) $p_H=2\sin(\pi/18)$ for the triangular lattice, and (3) $p_H=1-2\sin(\pi/18)$ for the honeycomb lattice.

E. Random Schrödinger Equations

Random Schrödinger equations are †Schrödinger equations in $\mathbf{R}^n$ with random potentials $U(x,\omega)$; therefore the corresponding operators are of the form

$$A(\omega)=-\Delta+U(\cdot,\omega),$$

where ω denotes a random parameter in a probability space $(\Omega,\mathfrak{F},P)$ and Δ denotes the †Laplacian in $\mathbf{R}^n$. It is assumed that this system of potentials forms a spatially homogeneous random field with the †ergodic property. This system of equations is considered to be a model describing the motion of quantum-mechanical particles in a random medium. Mathematically, the problem is to investigate various spectral properties of the self-adjoint operators $A(\omega)$. Since $A(\omega)$ and its shifted operator $A(\omega(\cdot+x))$ are †unitarily equivalent, the above assumption on the potentials $U(\cdot,\omega)$ implies that the spectral structures are independent of each sample ω a.s. if their structures of $A(\omega)$ are measurable with respect to $(\Omega,\mathfrak{F})$.

Several rigorous results have been obtained. In the 1-dimensional case, if the potentials $U(\cdot,\omega)$ are functionals of a strongly ergodic †Markov process, then it is proved that $A(\omega)$ has only a pure point spectrum [17], and each eigenfunction decays exponentially fast [18]. In multidimensional cases, asymptotic behavior at the left edge of the mean of the resolutions of the identity for a certain $A(\omega)$ have been investigated [19]. It is assumed that the random potential for this $A(\omega)$ takes the form

$$U(x,\omega)=\int_{\mathbf{R}^n}\varphi(x-y)\pi(dy,\omega),$$

where $\{\pi(dy,\omega)\}$ is a †Poisson random measure with mean measure dy and $\varphi(x)$ is a nonnegative measurable function satisfying $\varphi(x)=o(|x|^{-n-2})$ as $|x|\to\infty$. Let $E_\lambda(x,y,\omega)$ be the continuous kernel for the resolution of the identity for $A(\omega)$, and denote by $N(\lambda)$ the mean of $E_\lambda(0,0,\omega)$. Then $N(\lambda)$ is a nondecreasing function vanishing on $(-\infty,0)$ and has the following asymptotic form at $\lambda=0$:

$$\lambda^{n/2}\log N(\lambda)\to-\gamma_1^{n/2}\quad\text{as}\quad\lambda\to 0,$$

where γ_1 is the first eigenvalue of $-\Delta$ with a Dirichlet boundary condition on the ball in $\mathbf{R}^n$ with unit volume. The quantity $N(\lambda)$ can be identified with a limiting state density of $A(\omega)$, namely, the limit function of

$$N_V(\lambda,\omega)=\#\{j \mid \lambda_j^V(\omega)\leqslant\lambda\}/|V|$$

as V tends to $\mathbf{R}^n$ regularly, where $\{\lambda_j^V(\omega)\}$ is the set of eigenvalues of $A(\omega)$ in a smooth bounded domain V in $\mathbf{R}^n$ with a Dirichlet boundary condition and $|V|$ is the volume of V. To obtain the above asymptotic behavior of $N(\lambda)$, the theory of †large deviation for Markov processes plays a crucial role [20].

F. Boltzmann Equation

In the kinetic theory of gases the †Boltzmann equation is derived from the †Liouville equation by considering the BBGKY hierarchy of particle distribution functions for N particles and then by taking the limit $N\to\infty$ under certain conditions (→ 402 Statistical Mechanics). Mathematically rigorous discussions were given by O. E. Lanford [21] for a gas of hard spheres; he showed that solutions of the BBGKY hierarchy converge to those of the Boltzmann hierarchy for small time under the Boltzmann-Grad limit ($N\to\infty$, $Nd^2\to 1$, $d=$ the diameter of the hard spheres).

An approach to the Boltzmann equation in the spatially homogeneous case can also be based on a †master equation. M. Kac [22] considered a Poisson-like process describing the random time evolution of the n-tuple of the velocities of n particles. For a gas of hard spheres this is determined by the master equation

$$\frac{\partial}{\partial t}u(t,x_1,\dots,x_n)$$
$$=\frac{1}{n}\sum_{1\leqslant i<j\leqslant n}\int_{S^2}\{u(t,x_1,\dots,x_i',\dots,x_j',\dots,x_n)-u(t,x_1,\dots,x_n)\}|(x_i-x_j,l)|\,dl,$$
$$t>0,\quad x_1,\dots,x_n\in\mathbf{R}^3,\qquad(1)$$

where S^2 is the 2-dimensional unit sphere, dl is

the uniform distribution on S^2 and

$$x_1'=x_i+(x_j-x_i,l)l, \quad x_j'=x_j-(x_j-x_i,l)l.$$

Let σ be a positive constant and $S(\sqrt{n\sigma})$ denote the $(3n-1)$-dimensional sphere with center 0 and radius $\sqrt{n\sigma}$. Given a symmetric probability density u_n on $S(\sqrt{n\sigma})$ for each $n\geqslant 1$, a sequence $\{u_n\}$ is said to have **Boltzmann's property** or to be **chaotic** (or u-chaotic to stress u), if there exists a probability density u on $\mathbf{R}^3$ such that

$$\lim_{n\to\infty}\int_{S(\sqrt{n\sigma})}\varphi_1(x_1)\ldots\varphi_m(x_m)u_n(x_1,\ldots,x_n)dx_1\ldots dx_n=\prod_{k=1}^m\int_{\mathbf{R}^3}\varphi_k(x)u(x)dx$$

for each $m\geqslant 1$ and $\varphi_k\in C_0(\mathbf{R}^3)$, $1\leqslant k\leqslant m$. Kac's assertion is that the Boltzmann equation is to be derived from the master equation via the **propagation of chaos**; more precisely, if $\{u_n\}$ is a u-chaotic sequence, then $\{u_n(t)\}$ is also $u(t)$-chaotic, where $u_n(t)$ is the solution of the master equation (1) with $u_n(0)=u_n$ and $u(t)$ is the solution of the following Boltzmann equation with $u(0)=u$:

$$\frac{\partial}{\partial t}u(t,x)=\int_{S^2\times\mathbf{R}^3}\{u(t,x')u(t,y')-u(t,x)u(t,y)\}\times|(x-y,l)|\,dl\,dy,$$

$$x'=x+(y-x,l)l, \quad y'=y-(y-x,l)l.$$

The propagation of chaos was verified by Kac [22], H. P. McKean [24], and others for a considerably wide class of nonlinear equations of Boltzmann type (with cutoff). The propagation of chaos is the stage corresponding to the †law of large numbers. The next stage is the †central limit theorem or fluctuation theory, which was also discussed by Kac [23] and McKean [25]. Moreover, based on Kac's work [22], McKean [26] introduced a class of †Markov processes associated with certain nonlinear evolution equations including the Boltzmann equation; a process of this type describes the time evolution of the velocity of a particle interacting with other similar particles. In the case of the spatially homogeneous Boltzmann equation of Maxwellian molecules without cutoff, such a Markov process was constructed by solving a certain †stochastic differential equation [27]. This implies the existence of probability measure-valued solutions of the equation.

References

[1] E. Ising, Beitrag zur Theorie des Ferromagnetismus, Z. Phys., 31 (1925), 253–264.

[2] R. L. Dobrushin, The description of a random field by means of conditional probabilities and conditions of its regularity, Theory Prob. Appl., 13 (1968), 197–224. (Original in Russian, 1968.)

[3] O. E. Lanford III and D. Ruelle, Observables at infinity and states with short range correlations in statistical mechanics, Comm. Math. Phys., 13 (1969), 194–215.

[4] L. Onsager, Crystal statistics I: A two-dimensional model with an order-disorder transition, Phys. Rev., 65 (1944), 117–149.

[5] M. Aizenman, Translation invariance and instability of phase coexistence in the two dimensional Ising system, Comm. Math. Phys., 73 (1980), 83–94.

[6] Y. Higuchi, On the absence of non-translationally invariant Gibbs states for the two-dimensional Ising model, Proc. Conf. on Random Fields, Esztergom, North-Holland, 1981.

[7] R. L. Dobrushin, Gibbs state describing coexistence of phases for a three-dimensional Ising model, Theory Prob. Appl., 17 (1972), 582–600. (Original in Russian, 1972.)

[8] R. J. Glauber, Time-dependent statistics of the Ising model, J. Math. Phys., 4 (1963), 294–307.

[9] R. L. Dobrushin, Markov processes with a large number of locally interacting components: Existence of a limit process and its ergodicity, Problems of Information Transmission, 7 (1971), 149–164; Markov processes with many locally interacting components: The reversible case and some generalizations, ibid., 235–241. (Originals in Russian, 1971.)

[10] F. Spitzer, Interaction of Markov processes, Advances in Math., 5 (1970), 246–290.

[11] T. E. Harris, Contact interactions on a lattice, Ann. Probability, 2 (1974), 969–988.

[12] R. Holley and T. M. Liggett, The survival of contact processes, Ann. Probability, 6 (1978), 198–206.

[13] T. M. Liggett, The stochastic evolution of infinite systems of interacting particles, Lecture notes in math. 598, Springer, 1977, 187–248.

[14] S. R. Broadbent and J. M. Hammersley, Percolation processes I: Crystals and mazes, Proc. Cambridge Philos. Soc., 53 (1957), 629–641.

[15] R. T. Smythe and J. C. Wierman, First-passage percolation on the square lattice, Lecture notes in math. 671, Springer, 1978.

[16] H. Kesten, The critical probability of bond percolation on the square lattice equals 1/2, Comm. Math. Phys., 74 (1980), 41–59.

[17] I. J. Goldseid, S. A. Molchanov, and L. A. Pastur, One-dimensional random Schrödinger operator with purely point spectrum, Functional Anal. Appl., 11 (1977), 1–10. (Original in Russian, 1977.)

[18] S. A. Molchanov, The structure of eigenfunctions of one-dimensional unordered structures, Math. USSR-Izv., 12 (1978), 69–101. (Original in Russian, 1978.)
[19] S. Nakao, On the spectral distribution of the Schrödinger operator with random potential, Japan. J. Math., 3 (1977), 111–139.
[20] M. D. Donsker and S. R. S. Varadhan, Asymptotics for the Wiener sausage, Comm. Pure Appl. Math., 28 (1975), 525–565.
[21] O. E. Lanford III, Time evolution of large classical systems, Lecture notes in physics 38, Springer, 1975.
[22] M. Kac, Probability and related topics in physical sciences, Interscience, 1959.
[23] M. Kac, Some probabilistic aspects of the Boltzmann-equation, Acta Phys. Austrica Supp. 10, Springer, 1973, 379–400.
[24] H. P. McKean, Speed of approach to equilibrium for Kac's caricature of a Maxwellian gas, Arch. Rational Mech. Anal., 21 (1966), 347–367.
[25] H. P. McKean, Fluctuations in the kinetic theory of gases, Comm. Pure Appl. Math., 28 (1975), 435–455.
[26] H. P. McKean, A class of Markov processes associated with nonlinear parabolic equations, Proc. Nat. Acad. Sci. US, 56 (1966), 1907–1911.
[27] H. Tanaka, Probabilistic treatment of the Boltzmann equation of Maxwellian molecules, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 46 (1978), 67–105.

341 (XVII.2) Probability Measures

A. General Remarks

A **probability measure** Φ on a †measurable space $(S, \mathfrak{S})$ is defined to be a †measure on $(S, \mathfrak{S})$ with $\Phi(S)=1$ (→ 270 Measure Theory). In probability theory, probability measure appears usually as the †probability distribution of a †random variable (→ 342 Probability Theory). Unless stated otherwise, we regard a †topological space T as a measurable space endowed with the topological σ-algebra $\mathfrak{B}(T)$ on T, i.e., the †σ-algebra generated by the †open subsets of T. Hence the distribution of an $\mathbf{R}^n$-valued random variable is a probability measure on $(\mathbf{R}^n, \mathfrak{B}^n=\mathfrak{B}(\mathbf{R}^n))$. From this probabilistic background we often call a probability measure on $(\mathbf{R}^n, \mathfrak{B}^n)$ an n-dimensional (probability) distribution. For probability measures on topological spaces → 270 Measure Theory.

B. Quantities Characterizing Probability Distributions

Several different quantities characterize the properties of probability distributions in one dimension: the **mean** (or **mathematical expectation**) $m=\int_{-\infty}^{\infty} x d\Phi(x)$, the **variance** $\sigma^2=\int_{-\infty}^{\infty}|x-m|^2 d\Phi(x)$, the **standard deviation** σ, the kth **moment** $\alpha_k=\int_{-\infty}^{\infty} x^k d\Phi(x)$, the kth **absolute moment** $\beta_k=\int_{-\infty}^{\infty}|x|^k d\Phi(x)$, the kth **moment about the mean** $\mu_k=\int_{-\infty}^{\infty}(x-m)^k d\Phi(x)$, etc.

A one-to-one correspondence exists between a 1-dimensional distribution Φ and its **(cumulative) distribution function** F defined by $F(x)=\Phi((-\infty, x])$. A distribution function is characterized by the following properties: (1) It is monotone nondecreasing; (2) it is right continuous; (3) $\lim_{x\to-\infty}F(x)=0$ and $\lim_{x\to+\infty}F(x)=1$. Similar statements hold for the multidimensional case.

Let $X(\omega)$ be a real random variable on a †probability space $(\Omega, \mathfrak{F}, P)$. Then the distribution of X is given by $\Phi(E)=P(\{\omega \mid X(\omega)\in E\})$, $E\in\mathfrak{B}^1$, and the characteristic quantities of Φ defined above are given in terms of $X(\omega)$ as follows: $m=E(X)$, $\sigma^2=E(X-m)^2$, $F(x)=P(\{\omega \mid X(\omega)\leqslant x\})$, etc. The moments and the moments about the mean are connected by the relation $\mu_r=\sum_{k=0}^{r}\binom{r}{k}\alpha_{r-k}(-m)^k$ $(r=1,2,\ldots)$. When Φ is an n-dimensional distribution, the following quantities are frequently used: the **mean vector**, which is an n-dimensional vector whose ith component is given by $m_i=\int x_i d\Phi(x)$; the **covariance matrix**, which is an $n\times n$ matrix whose (i,j)-element is $\sigma_{ij}=\int(x_i-m_i)(x_j-m_j)d\Phi(x)$; the **moment matrix**, which is an $n\times n$ matrix whose (i,j)-element is $m_{ij}=\int x_i x_j d\Phi(x)$. (The covariance matrix is also called the **variance matrix** or the **variance-covariance matrix**.) The covariance matrix and the moment matrix are †positive definite and symmetric. The quantities listed above are defined only under some integrability conditions.

C. Characteristic Functions

Consider a probability measure Φ defined on a measurable space $(\mathbf{R}^n, \mathfrak{B}^n)$, where $\mathfrak{B}^n$ is the σ-algebra of all †Borel sets in $\mathbf{R}^n$. The **characteristic function** of Φ is the †Fourier transform φ defined by

$$\varphi(z)=\int_{\mathbf{R}^n} e^{i(z,x)}d\Phi(x), \qquad z\in\mathbf{R}^n, \tag{1}$$

where (z,x) denotes the †scalar product of z and x $(z,x\in\mathbf{R}^n)$. Let X be an n-dimensional random variable with probability distribution Φ defined on a †probability space $(\Omega, \mathfrak{B}, P)$.

Then the Fourier transform of Φ is also called the **characteristic function** of X, which can also be written as $E(e^{i(z,X)})$ (→ 342 Probability Theory).

The following properties play a fundamental role in the study of the relationship between probability distributions and characteristic functions: (i) the correspondence defined by (1) between the n-dimensional probability distribution Φ and its characteristic function φ is one-to-one. (ii) For any $a_p, b_p \in \mathbf{R}$, $a_p < b_p$ ($p = 1, 2, \ldots, n$), we have

$$\int_{\mathbf{R}^n} \prod_{p=1}^{n} f(x_p; a_p, b_p) d\Phi(x)$$
$$= \lim_{c\to\infty} \left(\frac{1}{2\pi}\right)^n \int_{-c}^{c} \cdots \int_{-c}^{c} \prod_{p=1}^{n} \frac{e^{-ib_pz_p} - e^{-ia_pz_p}}{-iz_p}$$
$$\times \varphi(z_1, \ldots, z_n) dz_1, \ldots, dz_n, \quad (2)$$

where $f(t; a, b)$ denotes the **modified indicator function** of $[a, b]$ defined by

$$f(t; a, b) = \begin{cases} 1, & t \in (a, b), \\ 1/2, & t = a \text{ or } b, \\ 0, & t \notin [a, b], \end{cases}$$

and $x = (x_1, \ldots, x_n) \in \mathbf{R}^n$. If an n-dimensional interval $I = [a_1, \ldots, a_n; b_1, \ldots, b_n]$ defined by $a_i \leqslant x_i \leqslant b_i$ ($i = 1, 2, \ldots, n$) is an **interval of continuity** for the probability distribution Φ, i.e., $\Phi(\partial I) = 0$, where ∂I denotes the boundary of I, then the left-hand side of (2) is equal to $\Phi(I)$. Equation (2) is called the **inversion formula** for the characteristic function φ.

The characteristic function φ of an n-dimensional probability distribution has the following properties: (i) For any points $z^{(1)}, \ldots, z^{(p)}$ of the n-dimensional space $\mathbf{R}^n$ and any complex numbers $a_1, \ldots, a_p$, we have

$$\sum_{j,k=1}^{p} \varphi(z^{(j)} - z^{(k)}) a_j \bar{a}_k \geqslant 0.$$

(ii) $\varphi(z^{(k)})$ converges to $\varphi(0)$ as $z^{(k)} \to 0$. (iii) $\varphi(0) = 1$. A complex-valued function φ of $z \in \mathbf{R}^n$ is called †positive definite if it satisfies the inequality in (i). Any continuous positive definite function φ on $\mathbf{R}^n$ such that $\varphi(0) = 1$ is the characteristic function of an n-dimensional probability distribution (†Bochner's theorem) (→ 192 Harmonic Analysis). A counterpart to Bochner's theorem holds for any positive definite sequence as well (†Herglotz's theorem).

The characteristic function is often useful for giving probability distributions explicitly. (For characteristic functions of typical probability distributions → Appendix A, Table 22. For general information about criteria that can be used to decide whether a given function is a characteristic function → [8].)

The **moment generating function** defined by $f(z) = \int \exp(-(z, x)) d\Phi(x)$ ($z \in \mathbf{R}^n$) does not necessarily exist for all n-dimensional distributions but does exist for a number of useful probability distributions Φ, and then $f(z)$ uniquely determines Φ.

Given a 1-dimensional distribution Φ with $\beta_k < +\infty$, we denote by γ_k the coefficient of $(iz)^k/k!$ in the †Maclaurin expansion of $\log \varphi(z)$. We call γ_k the (kth order) **semi-invariant** of Φ. The moments and semi-invariants are connected by the relations $\gamma_1 = \alpha_1$, $\gamma_2 = \alpha_2 - \alpha_1^2 = \sigma^2$, $\gamma_3 = \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3$, $\gamma_4 = \alpha_4 - 3\alpha_2^2 - 4\alpha_1\alpha_3 + 12\alpha_1^2\alpha_2 - 6\alpha_1^4, \ldots$.

D. Specific Distributions

Given an n-dimensional distribution Φ, a point a with $\Phi(\{a\}) > 0$ is called a discontinuity point of Φ. The set D of all discontinuity points of Φ is at most countable. When $\Phi(D) = 1$, Φ is called a **purely discontinuous distribution**. In particular, if D is a lattice, Φ is called a **lattice distribution**. If the distribution function of Φ is a continuous function, Φ is called a **continuous distribution**. By virtue of the †Lebesgue decomposition theorem, every probability distribution can be expressed in the form

$$\Phi = a_1\Phi_1 + a_2\Phi_2 + a_3\Phi_3,$$

$$a_1, a_2, a_3 \geqslant 0, \quad a_1 + a_2 + a_3 = 1,$$

where Φ_1 is purely discontinuous, Φ_2 is †absolutely continuous with respect to †Lebesgue measure, and Φ_3 is continuous and †singular. Let Φ be an absolutely continuous distribution. Then there exists a unique (up to Lebesgue measure zero) measurable nonnegative function $f(x)$ ($x \in \mathbf{R}^n$) such that $\Phi(E) = \int_E f(x) dx$. This function $f(x)$ is called the **probability density** of Φ.

We now list some frequently used 1-dimensional lattice distributions (for explicit data → Appendix A, Table 22): the **unit distribution** with $\Phi(\{0\}) = 1$; the **binomial distribution** $Bin(n, p)$ with parameters n and p; the **Poisson distribution** $P(\lambda)$ with parameter λ; the **geometric distribution** $G(p)$ with parameter p; the **hypergeometric distribution** $H(N, n, p)$ with parameters N, n, and p; and the **negative binomial distribution** $NB(m, q)$ with parameters m and q. The following k-dimensional lattice distributions are used frequently: the **multinomial distribution** $M(n, p)$ with parameters n and p; the **multiple hypergeometric distribution**; the **negative multinomial distribution**; etc.

The following 1-dimensional distributions are absolutely continuous: the **normal distribution** (or **Gaussian distribution**) $N(\mu, \sigma^2)$ with mean μ and variance σ^2 (sometimes $N(0, 1)$ is called the **standard normal distribution**); the

Cauchy distribution $C(\mu, \sigma)$ with parameters μ (†median) and σ; the **uniform distribution** $U(\alpha, \beta)$ on an interval $[\alpha, \beta]$; the **exponential distribution** $e(\sigma)$ with parameter σ; the **gamma distribution** $\Gamma(p, \sigma)$; the †χ^2 distribution $\chi^2(n)$; the **beta distribution** $B(p, q)$; the **F-distribution** $F(m, n)$; the **Z-distribution** $Z(m, n)$; the **t-distribution** $t(n)$; etc. Furthermore, there are several k-dimensional absolutely continuous distributions, such as the **k-dimensional normal distribution** $N(\mu, \Sigma)$ with mean vector $\mu = (\mu_1, \mu_2, \dots, \mu_k)$ and covariance matrix $\Sigma = (\sigma_{ij})$, the **Dirichlet distribution**, etc.

E. Convolution

Given any two n-dimensional distributions Φ_1, Φ_2, the n-dimensional distribution $\Phi(E) = \int_{\mathbf{R}^{2n}} \chi_E(x+y)\,d\Phi_1(x)\,d\Phi_2(y)$ is called the **composition** (or **convolution**) of Φ_1 and Φ_2 and is denoted by $\Phi_1 * \Phi_2$, where χ_E is the indicator function of the set E. Let X_1 and X_2 be †independent random variables with distributions Φ_1 and Φ_2. Then the distribution of $X_1 + X_2$ is $\Phi_1 * \Phi_2$. When F_i is the distribution function of Φ_i $(i = 1, 2)$, the distribution function $F_1 * F_2$ of $\Phi_1 * \Phi_2$ is expressed in the form $F_1 * F_2(x) = \int_{\mathbf{R}^n} F_1(x-y)\,dF_2(y)$. If Φ_1 has a density $f_1(x)$, then $\Phi_1 * \Phi_2$ has a density $f(x) = \int_{\mathbf{R}^n} f_1(x-y)\,dF_2(y)$. If $\varphi(z)$ is the characteristic function of the convolution of two probability distributions Φ_1 and Φ_2 with characteristic functions φ_1 and φ_2, then φ is the product of φ_1 and φ_2: $\varphi(z) = \varphi_1(z) \cdot \varphi_2(z)$. Therefore, for every k, the kth order semi-invariant of the convolution of two distributions is equal to the sum of their kth order semi-invariants. Suppose that we are given a family of distributions $\Phi = \{\Phi(\alpha, \beta, \dots)\}$ indexed with parameters $\alpha, \beta, \dots$. If for $(\alpha_1, \beta_1, \dots)$ and $(\alpha_2, \beta_2, \dots)$ there exists $(\alpha_3, \beta_3, \dots)$ such that $\Phi(\alpha_1, \beta_1, \dots) * \Phi(\alpha_2, \beta_2, \dots) = \Phi(\alpha_3, \beta_3, \dots)$, then we say that Φ has a **reproducing property**. Some of the distributions listed above have the reproducing property: $P(\lambda_1) * P(\lambda_2) = P(\lambda_1 + \lambda_2)$, $Bin(n_1, p) * Bin(n_2, p) = Bin(n_1 + n_2, p)$, $NB(m_1, q) * NB(m_2, q) = NB(m_1 + m_2, q)$, $N(\mu_1, \sigma_1^2) * N(\mu_2, \sigma_2^2) = N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$, $\Gamma(p_1, \sigma) * \Gamma(p_2, \sigma) = \Gamma(p_1 + p_2, \sigma)$, $C(\mu_1, \sigma_1) * C(\mu_2, \sigma_2) = C(\mu_1 + \mu_2, \sigma_1 + \sigma_2)$, etc.

Given a 1-dimensional distribution function $F(x)$,

$$Q_F(l) = \max_{-\infty < x < \infty} (F(x+l) - F(x-l)), \quad l > 0,$$

is called the **maximal concentration function** of F (P. Lévy [6]). Since it satisfies the relation $Q_{F_1 * F_2}(l) \leqslant Q_{F_i}(l)$ $(i = 1, 2)$, we can use it to study the properties of sums of independent random variables. The **mean concentration function** defined by

$$C_F(l) = \frac{1}{2l} \int_{-\infty}^{\infty} (F(x+l) - F(x-l))^2\,dx$$

is also useful for similar purposes.

Let $N(m, v)$ be the 1-dimensional normal distribution with mean m and variance v, and let $P(\lambda, a)$ be the distribution obtained through translation by a of the Poisson distribution with parameter λ. If 1-dimensional distributions Φ_k, Ψ_k $(k = 1, 2)$ exist such that $N(m, v) = \Phi_1 * \Phi_2$, $P(\lambda, a) = \Psi_1 * \Psi_2$, we have $\Phi_k = N(m_k, v_k)$, $\Psi_k = P(\lambda_k, a_k)$ $(k = 1, 2)$ for some m_k, v_k, λ_k, a_k $(k = 1, 2)$. These are known, respectively, as Cramér's theorem and Raikov's theorem. Yu. V. Linnik proved a similar fact (the decomposition theorem) for a more general family with reproducing property by using the theory of analytic functions [9].

F. Convergence of Probability Distributions

The concept of convergence of distributions plays an important role in limit theorems and other fields of probability theory. When Ω is a topological space, we consider **convergence** of probability measures on Ω with respect to the †weak topology introduced in the space of measures on Ω (→ 37 Banach Spaces). Such convergence is called **weak convergence** in probability theory. For a sequence of n-dimensional distributions Φ_k $(k = 1, 2, \dots)$ to converge to Φ weakly, each of the following conditions is necessary and sufficient. (1) For every continuous function with compact support, $\lim_{k \to \infty} \int_{\mathbf{R}^n} f(x)\,d\Phi_k(x) = \int_{\mathbf{R}^n} f(x)\,d\Phi(x)$. (2) At every continuity point of the distribution function $F(x_1, \dots, x_n)$ of Φ, $\lim_{k \to \infty} F_k(x_1, \dots, x_n) = F(x_1, \dots, x_n)$ (F_k is the distribution function of Φ_k). (3) For every continuity set E of Φ (namely, a set such that $\Phi(\bar{E} - E^0) = 0$), $\lim_{k \to \infty} \Phi_k(E) = \Phi(E)$. (4) For all open $G \subset \mathbf{R}^n$, $\liminf_{k \to \infty} \Phi_k(G) \geqslant \Phi(G)$. (5) For all closed $F \subset \mathbf{R}^n$, $\limsup_{k \to \infty} \Phi_k(F) \leqslant \Phi(F)$. (6) $\lim_{k \to \infty} \rho(\Phi_k, \Phi) = 0$, where ρ is a metric defined in the following way: Given any n-dimensional distributions Φ_1, Φ_2, we put $\varepsilon_{ij} = \inf\{\varepsilon \mid \Phi_j(F) < \Phi_i(F^\varepsilon) + \varepsilon$ for every closed $F\}$ (F^ε is the ε-neighborhood of F) and define $\rho(\Phi_1, \Phi_2) = \max(\varepsilon_{12}, \varepsilon_{21})$. The metric ρ, called the **Lévy distance**, was introduced by Lévy [6] in one dimension and by Yu. V. Prokhorov in metric spaces [10]. Each of these conditions except (2) is still necessary and sufficient for Φ_n to converge weakly to Φ when Φ_n and Φ are probability measures on a †complete separable metric space. It should also be noted that the probability measures on a complete separable metric space consti-

tute a complete separable metric space with respect to the Lévy distance.

A family $\Phi_\alpha (\alpha \in \Lambda)$ of probability measures on a complete separable metric space is said to be **tight** if for every $\varepsilon > 0$ there exists a †compact set $K = K(\varepsilon)$ such that $\Phi_\alpha(K^c) < \varepsilon$ for all $\alpha \in \Lambda$. A family Φ_α $(\alpha \in \Lambda)$ is tight if and only if it is †totally bounded with respect to the topology induced by the Lévy distance. Hence a tight family Φ_α $(\alpha \in \Lambda)$ has a weakly convergent subsequence.

We can give a criterion for the convergence of probability measures in terms of their characteristic functions. Suppose that Φ_k and Φ are n-dimensional probability measures with characteristic functions φ_k and φ. Then Φ_k converges weakly to Φ if and only if for every z, $\lim_k \varphi_k(z) = \varphi(z)$. Let φ_k be the characteristic function of an n-dimensional probability distribution Φ_k. If the sequence $\{\varphi_k\}$ converges pointwise to a limit function φ and the convergence of φ_k is uniform in some neighborhood of the origin, then φ is also the characteristic function of an n-dimensional probability distribution Φ and the sequence $\{\Phi_k\}$ converges weakly to Φ [7] (**Lévy's continuity theorem**).

For any probability distribution concentrated on $[0, \infty)$, the use of †Laplace transforms as a substitute for Fourier transforms provides a powerful tool. The method of **probability generating functions** is available for the study of arbitrary probability distribution concentrated on the nonnegative integers [14]. The method of moment-generating functions is also useful. There are many results on the relation between these functions, probability distributions, and their convergence [14–16].

Let $\Phi_1, \Phi_2, \ldots$ and Φ be 1-dimensional distributions. If all absolute moments exist, $\sum_{j=1}^{\infty} \beta_j^{-1/j} = \infty$ for $\beta_j = \int_{-\infty}^{\infty} |x|^j d\Phi(x) < \infty$, and $\lim_{k \to \infty} \int_{-\infty}^{\infty} x^j d\Phi_k(x) = \int_{-\infty}^{\infty} x^j d\Phi(x)$ $(j = 0, 1, 2, \ldots)$, then Φ_k weakly converges to Φ. This condition is sufficient but not necessary.

G. Infinitely Divisible Distributions

An n-dimensional probability distribution Φ is called **infinitely divisible** if for every positive interger k, there exists a probability distribution Φ_k such that $\Phi = \Phi_k * \Phi_k * \ldots * \Phi_k$ $(= \Phi_k^{*k})$. Both normal distributions and Poisson distributions are infinitely divisible. If an n-dimensional distribution Φ satisfies the condition $\int_{|x| > \varepsilon} \Phi(dx) < \varepsilon$, we say that $\Phi \in v(\varepsilon)$. Then Φ is an infinitely divisible distribution if and only if for every $\varepsilon > 0$ we can find $\Phi_1, \Phi_2, \ldots, \Phi_k \in v(\varepsilon)$ such that $\Phi = \Phi_1 * \Phi_2 * \ldots * \Phi_k$.

Let X_{ki}, $i = 1, 2, \ldots, n(k)$, be independent random variables for every k, and assume that the distribution of X_{ki} belongs to $v(\varepsilon_k)$, $i = 1, 2, \ldots, n(k)$, where $\varepsilon_k \to 0$ as $k \to \infty$. If the probability distributions of the sums $X_k = \sum_{i=1}^{n(k)} X_{ki}$ converge to a probability distribution as $k \to \infty$, then the limit distribution is infinitely divisible.

The characteristic function of a 1-dimensional infinitely divisible distribution can be written in the form

$$\varphi(z) = \exp\left(irz - \frac{v}{2} z^2 + \int_{-\infty}^{\infty} A(u, z) \frac{1+u^2}{u^2} dG(u) \right), \quad (3)$$

where γ is a constant, v is a nonnegative constant, $G(u)$ is a nondecreasing bounded function with $G(-\infty) = 0$, $A(u, z) = \exp(iuz) - 1 - izu/(1+u^2)$, and the value of $A(u, z)(1+u^2)/u^2$ at $u = 0$ is defined to be $-z^2/2$. Formula (3) is called **Khinchin's canonical form**. For the characteristic function of an infinitely divisible n-dimensional distribution, the canonical form is as follows:

$$\varphi(z) = \exp\left(i(m, z) - \sum_{p,q=1}^{n} c_{pq} z_p z_q + \int_{\mathbf{R}^n} \left(e^{i(z,x)} - 1 - \frac{i(z, x)}{1+|x|^2} \right) n(dx) \right),$$
$$z = (z_1, \ldots, z_n) \in \mathbf{R}^n, \quad (4)$$

where $m \in \mathbf{R}^n$, (c_{pq}) is a positive semidefinite matrix, and $n(dx)$ is a measure on $\mathbf{R}^n$ such that $n(\{0\}) = 0$ and

$$\int_{\mathbf{R}^n} \frac{|x|^2}{1+|x|^2} n(dx) < \infty.$$

Formula (4) is called **Lévy's canonical form.** If a 1-dimensional infinitely divisible distribution Φ satisfies $\int_{\mathbf{R}^1} x^2 d\Phi(x) < \infty$, then its characteristic function is given by

$$\varphi(z) = \exp\left(imz - \frac{v}{2} z^2 + \int_{-\infty}^{\infty} (e^{izu} - 1 - izu) \frac{1}{u^2} dK(u) \right), \quad (5)$$

where m is a real constant, v is a nonnegative constant, and $K(u)$ is a nondecreasing bounded function such that $K(-\infty) = 0$. It is called **Kolmogorov's canonical form.** (For infinitely divisible distributions on a homogeneous space → 5 Additive Processes.)

Let Φ and Ψ be n-dimensional distributions. If for some $\lambda > 0$, $\Psi(E) = \Phi(\lambda E)$ $(\lambda E = \{\lambda\xi \mid \xi \in E\})$ for every set E, we say that Φ and Ψ are equivalent. Let Φ and Ψ be probability distributions with distribution functions F and G

and characteristic functions φ and ψ. Then the following three statements are equivalent: (1) Φ and Ψ are equivalent; (2) $G(x)=F(\lambda x)$ for every x; and (3) $\psi(\lambda z)=\varphi(z)$ for every z. We call Φ a **stable distribution** if for every pair of distributions Φ_1, Φ_2 equivalent to Φ, the convolution $\Phi_1 * \Phi_2$ is equivalent to Φ. If Φ is stable, every distribution equivalent to Φ is also stable. We can characterize stable distributions in terms of their characteristic functions $\varphi(z)$ as follows: For every pair $\lambda_1, \lambda_2 > 0$, there exists a $\lambda = \lambda(\lambda_1, \lambda_2) > 0$ such that $\varphi(\lambda z) = \varphi(\lambda_1 z)\varphi(\lambda_2 z)$. We can restate this characterization as follows: Φ is stable if and only if for every pair of independent random variables X_1 and X_2 with identical distribution Φ and for any positive numbers λ_1 and λ_2, there exists a positive number λ such that $(\lambda_1 X_1 + \lambda_2 X_2)/\lambda$ has the distribution Φ. By the definition we see that all stable distributions are infinitely divisible.

In the 1-dimensional case, putting $\varphi(z) = \exp\psi(z)$, we have $\psi(\lambda z) = \psi(\lambda_1 z) + \psi(\lambda_2 z)$, which implies $\psi(z) = (-c_0 + i(z/|z|)c_1)|z|^\alpha$, where $c_0 \geqslant 0$, $-\infty < c_1 < \infty$, $0 < \alpha \leqslant 2$. The parameter α is called the **exponent** (or **index**) of the stable distribution. The stable distributions with exponent $\alpha = 2$ are the normal distributions, and the stable distributions with exponent $\alpha = 1$ are the Cauchy distributions. We have $\psi(z) = -c_0|z|^\alpha$ for a symmetric stable distribution. (For the stable distribution with exponent $1/2$ → Appendix A, Table 22).

Generalizing stable distributions, we can define **quasistable distributions**, which B. V. Gnedenko and A. N. Kolmogorov [17] called stable distributions also. Let F be the distribution function of a 1-dimensional distribution Φ. Φ is said to be quasistable if to every $b_1 > 0$, $b_2 > 0$ and real λ_1, λ_2 there correspond a positive number b and a real number λ such that we have the relation $F((x-\lambda_1)/b_1) * F((x-\lambda_2)/b_2) = F((x-\lambda)/b)$.

Let $\{X_i\}$ be a sequence of independent random variables with identical distribution. If for suitably chosen constants A_n and B_n the distributions of the sums $B_n^{-1}(\sum_{i=1}^n X_i) - A_n$ converge to a distribution, the limit distribution is a quasistable distribution (Lévy). A necessary and sufficient condition for a distribution to be quasistable is that its characteristic function $\varphi(z)$ satisfy the relation $\varphi(b_1 z)\varphi(b_2 z) = \varphi(bz)e^{i\gamma z}$ $(\gamma = \lambda - \lambda_1 - \lambda_2)$. The characteristic function of a quasistable distribution has the canonical representation

$$\varphi(z) = \exp\psi(z),$$

$$\psi(z) = imz - c|z|^\alpha(1 + i\beta(z/|z|)\omega(z,\alpha)),$$

where m is a real number, $c \geqslant 0$, $0 < \alpha \leqslant 2$, $|\beta| \leqslant 1$, and $\omega(z,\alpha) = \tan(\pi\alpha/2)$ $(\alpha \neq 1)$, $\omega(z,\alpha) = (2/\pi)\log|z|$ $(\alpha = 1)$. The parameter α is called the exponent of the quasistable distribution. A quasistable distribution with $\alpha \neq 1$ is obtained from a stable distribution by translation, but quasistable distributions with $\alpha = 1$ are not.

Semistable distributions are another generalization of stable distributions. A distribution is called semistable if its characteristic function $\varphi(z)$ satisfies the relation $\psi(qz) = q^\alpha\psi(z)$ for a positive number q $(\neq 1)$, where $\varphi(z) = \exp(\psi(z))$. Also in this case, the general form was obtained by Lévy [6].

A 1-dimensional probability distribution Φ is called an ***L*-distribution** if the distribution function F of Φ is the convolution of $F(x/a)$ and some other distribution function $F_a(x)$ for every $0 < a < 1$. Φ is an L-distribution if and only if there exists a sequence of independent random variables $\{X_k\}$ such that for suitably chosen constants $B_n > 0$ and A_n the distributions of the sums $B_n^{-1}(\sum_{k=1}^n X_k) - A_n$ converge to Φ and $\sup_{1 \leqslant k \leqslant n} P(|X_k/B_n| > \varepsilon) \to 0$ as $n \to \infty$ for every $\varepsilon > 0$. Quasistable distributions are L-distributions.

H. The Shape of Distributions

Let $F(x)$ be a 1-dimensional distribution function. The quantity ζ_p such that $F(\zeta_p - 0) \leqslant p \leqslant F(\zeta_p)$ $(0 < p < 1)$ is called the **quantile of order** p of F. In particular, the quantity $\zeta_{1/2}$ is called a **median**. If F satisfies the relation $1 - F(m+x) = F(m-x)$, it is called **symmetric**. In any 1-dimensional symmetric distribution, every moment of odd order about the mean (if it exists) is equal to zero.

The ratio $\gamma_1 = \mu_3/\sigma^3$ is used as a measure of departure from symmetry of a distribution and is called the **coefficient of skewness**. Furthermore, the ratio $\gamma_2 = \mu_3/\sigma^4 - 3$ is called the **coefficient of excess**. For the normal distribution, we have $\gamma_1 = \gamma_2 = 0$. If $\gamma_2 \neq 0$, γ_2 expresses the degree of deviation from the normal distribution.

A distribution function $F(x)$ is called **unimodal** if there exists one value $x = a$ such that $F(x)$ is convex for $x < a$ and concave for $x > a$. All L-distributions (and hence quasistable distributions) are unimodal [18].

I. Kolmogorov's Extension Theorem

Let $\Omega = \mathbf{R}^T$, where T is an arbitrary index set. We associate with $\mathbf{R}^T$ the σ-algebra $\mathfrak{B}^T$ generated by the cylinder sets, i.e., $\{\omega \in \Omega \mid \pi_{t_1}(\omega) \in E_1, \ldots, \pi_{t_n}(\omega) \in E_n\}$, where $\pi_t(\omega)$ denotes the tth coordinate of ω, $E_k \in \mathfrak{B}(\mathbf{R}^1)$, $1 \leqslant k \leqslant n$, $t_1 < t_2 < \ldots < t_n$, and $n = 1, 2, \ldots$. Given a probability

measure Φ on $(\mathbf{R}^T, \mathfrak{B}^T)$, we can define a finite-dimensional †marginal distribution Φ_S for any finite subset S of T by $\Phi_S(E)=\Phi(\pi_S^{-1}(E))$, $E\in\mathfrak{B}^S$, where π_S is the natural †projection π_S: $\mathbf{R}^T\to\mathbf{R}^S$. The measures $\{\Phi_S\}$ satisfy the following **consistency condition**: If $S_1\subset S_2$ $(\subset T)$ are finite and if $E\in\mathfrak{B}^{S_1}$, then

$$\Phi_{S_1}(E)=\Phi_{S_2}(\pi_{S_1S_2}^{-1}(E)), \qquad (6)$$

where $\pi_{S_1S_2}:\mathbf{R}^{S_2}\to\mathbf{R}^{S_1}$ is the natural projection.

Conversely, if we are given a family of finite-dimensional probability measures $\{\Phi_S\}$ which satisfies the consistency condition (6), then **Kolmogorov's extension theorem** [1] asserts that there exists a unique probability measure Φ on $(\mathbf{R}^T, \mathfrak{B}^T)$ such that $\Phi_S(E)=\Phi(\pi_S^{-1}(E))$, $E\in\mathfrak{B}^S$, for any finite $S\subset T$.

This theorem is useful in constructing †stochastic processes. For example, let $\Phi_{n_1,n_2,\ldots,n_k}$ be the †product measure of k copies of a given probability measure Φ on $\mathbf{R}^1$. Then the family $\{\Phi_{n_1,n_2,\ldots,n_k}; n_1,n_2,\ldots,n_k\in\mathbf{Z}, k\in\mathrm{N}\}$ satisfies the consistency condition and hence, by Kolmogorov's extension theorem, determines a probability measure on $\mathbf{R}^{\mathbf{Z}}$, which is denoted by $\Phi^{\mathbf{Z}}$. Thus $X_n(\omega)=\pi_n(\omega)$, $n\in\mathbf{Z}$, $(\omega\in(\mathbf{R}^{\mathbf{Z}}, \mathfrak{B}^{\mathbf{Z}}, \Phi^{\mathbf{Z}}))$, are independent identically Φ-distributed random variables.

Kolmogorov's extension theorem is generalized to the case where the component spaces are †standard measurable spaces (→ 270 Measure Theory) instead of $\mathbf{R}^1$, and also to the case where product spaces are replaced by †projective systems [19].

J. Characteristic Functionals on Infinite-Dimensional Spaces

Contrary to the finite-dimensional case, Bochner's theorem does not necessarily hold in infinite-dimensional spaces. For example, let $(T, \|\cdot\|)$ be an infinite-dimensional †Hilbert space and $\varphi(t)=\exp(-\|t\|^2)$. φ is continuous and positive definite, and $\varphi(0)=1$. But it is known that there is no probability measure on $T\doteq T^*$ (topological dual of T) which corresponds to φ. Bochner's theorem is generalized to infinite-dimensional spaces as follows.

Let T be a real †vector space endowed with the topology τ defined by a system of †Hilbertian seminorms $\{\|\cdot\|_\alpha, \alpha\in A\}$. Define a new topology $I(\tau)$ of T by all Hilbertian seminorms $\|\cdot\|$ each of which is HS-dominated by some $\|\cdot\|_\alpha$, $\alpha\in A$, i.e., $\sup\{(\sum_i\|e_i\|^2)^{1/2}; \{e_i\}:\alpha\text{-orthonormal}\}<\infty$. If $I(\tau)=\tau$, then (T,τ) is called a †nuclear space.

Let T_τ^* be the topological dual of (T,τ) (i.e., the set of all τ-continuous real valued linear functionals on T). Define a Borel structure $\mathfrak{B}(T_\tau^*)$ of T_τ^* as the σ-algebra generated by the system of half-spaces $\{x\in T_\tau^*; x(t)<a\}$, $t\in T$, $a\in\mathbf{R}^1$. For a probability measure Φ on $(T_\tau^*, \mathfrak{B}(T_\tau^*))$, define

$$\varphi(t)=\int_{T_\tau^*}\exp(i\langle x,t\rangle)\,d\Phi(x), \quad t\in T,$$

which is called the characteristic functional of Φ. A functional φ on T is the characteristic functional of a probability measure Φ on $(T_\tau^*, \mathfrak{B}(T_\tau^*))$ such that $\Phi^*(\bigcup_n T_{\alpha_n}^*)=1$ for some sequence $\{\alpha_n\}\subset A$, where Φ^* is the †outer measure (→ 270 Measure Theory) and T_α^* is the topological dual of $(T, \|\cdot\|_\alpha)$, if and only if φ is positive definite, $\varphi(0)=1$, and continuous with respect to the topology $I(\tau)$ [23].

As special cases of the foregoing theorem, we have the following. If (T,τ) is a nuclear space, then every positive definite τ-continuous functional φ with $\varphi(0)=1$ is the characteristic functional of a probability measure on T_τ^* (Minlos [24]). Schwartz's spaces $\mathcal{S}(\mathbf{R}^n)$ and $\mathcal{D}(\mathbf{R}^n)$ are nuclear. Let $(T, \tau=\|\cdot\|)$ be a Hilbert space. A †Hilbert-Schmidt operator U is, by definition, a †bounded linear operator on T such that $\sum_i\|Ue_i\|^2<\infty$, by any †complete orthonormal system $\{e_i\}$ (this quantity does not depend on the choice of $\{e_i\}$). Define a seminorm $\|\cdot\|_U$ by $\|t\|_U=\|Ut\|$, $t\in T$, for a Hilbert-Schmidt operator U. Then the topology $I(\tau)$ coincides with the topology induced by the system of seminorms $\|\cdot\|_U$, where U are Hilbert-Schmidt operators, which is called the **Sazonov topology**. Thus every functional φ on T, which is positive definite, $\varphi(0)=1$, and continuous with respect to the Sazonov topology, is the characteristic functional of a probability measure on T_τ^* (Sazonov [25]).

The probability measure on $\mathcal{S}'$ with the characteristic functional $\exp(-\int_{-\infty}^{\infty}|f(s)|^2\,ds)$, $f\in\mathcal{S}$, is the probability measure of a Gaussian †white noise on $\mathcal{S}'$.

References

[1] A. N. Kolmogorov, Grundbegriffe der Wahrscheinlichkeitsrechnung, Springer, 1933; English translation, Foundations of the theory of probability, Chelsea, 1950.

[2] H. Cramér, Mathematical methods of statistics, Princeton Univ. Press, 1946.

[3] J. L. Doob, Stochastic processes, Wiley, 1953.

[4] M. M. Loève, Probability theory, Van Nostrand, third edition, 1963.

[5] W. Feller, An introduction to probability theory and its applications II, Wiley, 1966.

[6] P. Lévy, Théorie de l'addition des variables aléatoires, Gauthier-Villars, 1937.

[7] P. Lévy, Calcul des probabilités, Gauthier-Villars, 1925.
[8] E. Lukacs, Characteristic functions, Hafner, 1960.
[9] Yu. V. Linnik, Decomposition of probability distributions, Oliver & Boyd, 1964. (Original in Russian, 1960.)
[10] Yu. V. Prokhorov, Convergence of random processes and limit theorems in probability theory, Theory of Prob. Appl., 1 (1956), 155–214. (Original in Russian, 1956.)
[11] L. LeCam, Convergence in distribution of stochastic processes, Univ. California Publ. Statist., 2 (1957), 207–236.
[12] K. R. Parthasarathy, Probability measures on metric spaces, Academic Press, 1967.
[13] P. Billingsley, Convergence of probability measures, Wiley, 1968.
[14] W. Feller, An introduction to probability theory and its applications I, Wiley, second edition, 1957.
[15] J. A. Shohat and J. D. Tamarkin, The problem of moments, Amer. Math. Soc. Math. Surveys, 1943.
[16] D. V. Widder, Laplace transform, Princeton Univ. Press, 1941.
[17] B. V. Gnedenko and A. N. Kolmogorov, Limit distributions for sums of independent random variables, Addison-Wesley, 1954. (Original in Russian, 1949.)
[18] M. Yamazato, Unimodality of infinitely divisible distribution functions of class L, Ann. Prob., 6 (1978), 523–531.
[19] S. Bochner, Harmonic analysis and the theory of probability, Univ. of California Press, 1955.
[20] E. Nelson, Regular probability measures on function space, Ann. Math., (2) 69 (1959), 630–643.
[21] I. M. Gel'fand and N. Ya. Vilenkin, Generalized functions IV, Applications of harmonic analysis, Academic Press, 1964. (Original in Russian, 1961.)
[22] Yu. V. Prokhorov, The method of characteristic functionals, Proc. 4th Berkeley Symp. Math. Stat. Prob. II, Univ. of California Press, 1961, 403–419.
[23] A. N. Kolmogorov, A note on the papers of R. A. Minlos and V. Sazonov, Theory of Prob. Appl., 4 (1959), 221–223. (Original in Russian, 1959.)
[24] R. A. Minlos, Generalized random processes and their extension to a measure, Selected Transl. Math. Stat. and Prob., Amer. Math. Soc., 3 (1962), 291–313. (Original in Russian, 1959.)
[25] V. Sazonov, A remark on characteristic functionals, Theory of Prob. Appl., 3 (1958), 188–192. (Original in Russian, 1958.)
[26] M. Kac, Probability and related topics in physical sciences, Interscience, 1959.

342 (XVII.1) Probability Theory

A. History

The origin of the **theory of probability** goes back to the mathematical problems connected with dice throwing that were discussed in letters exchanged by B. Pascal and P. de Fermat in the 17th century. These problems were concerned primarily with concepts such as †permutations, †combinations, and †binomial coefficients, whose theory was established at about the same time [1]. This elementary theory of probability was later enriched by the work of scholars such as Jakob Bernoulli [2], A. de Moivre [3], T. Bayes, L. de Buffon, Daniel Bernoulli, A. M. Legendre, and J. L. Lagrange. Finally, P. S. Laplace completed the classical theory of probability in his book *Théorie analytique des probabilités* (1812). In this work, Laplace not only systematized but also greatly extended previous important results by introducing new methods, such as the use of †difference equations and †generating functions. Since the 19th century, the theory of probability has been extensively applied to the natural sciences and even to the social sciences.

The definition of a priori probability due to Laplace provoked a great deal of argument when it was applied. For example, R. von Mises advocated an empirical theory of probability based on the notion of **Kollektiv (collective)**, which is a mathematical model of mass phenomena [5]. However, these arguments are concerned with philosophical rather than mathematical aspects. Nowadays, the main concern of mathematicians lies not in the intuitive or practical meaning of probability but in the logical setup governing probability. From this viewpoint the mathematical model of a random phenomenon is given by a probability measure space $(\Omega, \mathfrak{B}, P)$, where Ω is the set of all possible outcomes of the phenomenon, $P(E)$ represents the probability that an outcome belonging to E be realized, and $\mathfrak{B}$ is a σ-algebra consisting of all sets E for which $P(E)$ is defined. All probabilistic concepts, such as random variables, independence, etc., are defined on $(\Omega, \mathfrak{B}, P)$ in terms of measure theory. Such a measure-theoretic basis of probability theory is due to A. Kolmogorov [6], though similar considerations had been made before him for special problems, for example, in the work of E. Borel concerning the strong law of large numbers [7] and in the rigorous definition of Brownian motion by N. Wiener [8].

Ever since probability theory was given

solid foundations by Kolmogorov, it has made tremendous progress. The most important concept in today's probability theory is that of †**stochastic processes**, which correspond to functions in analysis. In applications a stochastic process is used as the mathematical model of a random phenomenon varying with time. The following types of stochastic processes have been investigated extensively: †additive processes, †Markov processes and †Markov chains, †martingales, †stationary processes, and †Gaussian processes. †Brownian motion and †branching processes are important special stochastic processes. In the same way as functions are often defined by differential equations, there are stochastic processes which can be defined by †stochastic differential equations. The theory of stochastic processes and stochastic differential equations can be applied to †stochastic control, †stochastic filtering, and †statistical mechanics. The †ergodic theory that originated in statistical mechanics is now regarded as an important branch of probability theory closely related to the theory of stationary processes.

B. Probability Spaces

Let Ω be an †abstract space and $\mathfrak{B}$ be a †σ-algebra of subsets of Ω. A **probability measure** (or **probability distribution**) over $\Omega(\mathfrak{B})$ is a set function $P(E)$ defined for $E\in\mathfrak{B}$ and satisfying the following conditions: (P1) $P(E)\geqslant 0$; (P2) for every sequence $\{E_n\}$ $(n=1,2,\dots)$ of pairwise disjoint sets in $\mathfrak{B}$,

$$P\left(\bigcup_n E_n\right)=\sum_n P(E_n);$$

(P3) $P(\Omega)=1$. The triple $(\Omega,\mathfrak{B},P)$ is called a **probability space**. The space Ω (resp. each element ω of Ω) is called the **basic space, space of elementary events**, or **sample space** (resp. **sample point** or **elementary event**). We say that a condition $\varepsilon(\omega)$ involving a generic sample point ω is an **event**; in particular, it is called a **measurable event** or **random event** if the set E of all sample points satisfying $\varepsilon(\omega)$ belongs to $\mathfrak{B}$. We assume that an event is always a measurable one, since we encounter only measurable events in the theory of probability. Because of the obvious one-to-one correspondence between measurable events and $\mathfrak{B}$-measurable sets (i.e., the correspondence of each event ε with the set E of all sample points ω satisfying ε), a $\mathfrak{B}$-measurable set itself is frequently called an event. If ε is an event and E is the $\mathfrak{B}$-measurable set corresponding to ε, we call $P(E)$ or $\Pr(\varepsilon)$ the **probability that the event ε occurs**, i.e., the **probability of the event** E. The **complementary event** (resp. **impossible event, sure event**) is the complementary set E^c (empty set $\varnothing$, whole space Ω). For a finite or infinite family $\{E_\lambda\}$ $(\lambda\in\Lambda)$, the **sum event** (resp. **intersection** or **product event**) of E_λ is the set $\bigcup_\lambda E_\lambda$ $(\bigcap_\lambda E_\lambda)$. If $E\cap F=\varnothing$, then we say that E and F are mutually exclusive or that they are **exclusive events**.

By the definition of P, we have $0\leqslant P(E)\leqslant 1$ for any event E, $P(\varnothing)=0$, and $P(\Omega)=1$. Moreover, if $\{E_n\}$ $(n=1,2,\dots)$ is a sequence of pairwise exclusive events and E is the sum event of E_n, we have

$$P(E)=\sum_{n=1}^{\infty}P(E_n).$$

This property is called the **additivity of probability**. If $P(E)=1$, the event E is said to occur **almost certainly (almost surely** (abbrev. a.s.), **for almost all** ω, or **with probability** 1).

Given a finite sequence $\{E_n\}$ $(n=1,2,\dots,N)$ of events, we say that the events E_n $(n=1,2,\dots,N)$ are mutually **independent** or that the sequence $\{E_n\}$ $(n=1,2,\dots,N)$ is **independent** if every subsequence satisfies

$$P(E_{i_1}\cap E_{i_2}\cap\dots\cap E_{i_k})=\prod_{j=1}^{k}P(E_{i_j}).$$

Given an infinite family $\{E_\lambda\}$ $(\lambda\in\Lambda)$ of events, we say that the events E_λ $(\lambda\in\Lambda)$ are mutually **independent** or that the family $\{E_\lambda\}$ $(\lambda\in\Lambda)$ is **independent** if every finite subfamily is independent. The concept of independence of events can be generalized to a family $\{\mathfrak{B}_\lambda\}$ $(\lambda\in\Lambda)$ of σ-subalgebras of $\mathfrak{B}$ as follows. A family $\{\mathfrak{B}_\lambda\}$ $(\lambda\in\Lambda)$ of σ-subalgebras of events is said to be independent if for every choice of $E_\lambda\in\mathfrak{B}_\lambda$, the family $\{E_\lambda\}$ $(\lambda\in\Lambda)$ of events is independent.

For a sequence $\{E_n\}$ $(n=1,2,\dots)$ of events, the sets $\limsup_n E_n$ and $\liminf_n E_n$ are called the **superior limit event** and **inferior limit event**, respectively. The superior limit event (inferior limit event) is the set of all ω for which infinitely many events among E_n (all events except finitely many E_n) occur. Therefore $P(\liminf_n E_n)$ is the probability that infinitely many events among E_n occur, and $P(\limsup_n E_n)$ is the probability that the events E_n occur for all n after some number n_0, where n_0 depends on ω in general. The **Borel-Cantelli lemma**, which is concerned with the evaluation of $P(\limsup_n E_n)$, reads as follows: Given a sequence $\{E_n\}$ $(n=1,2,\dots)$ of events, we have (i) whether the events E_n $(n=1,2,\dots)$ are mutually independent or not, $\sum_n P(E_n)<\infty$ implies that $P(\limsup_n E_n)=0$; and (ii) if the events E_n $(n=1,2,\dots)$ are mutually independent, $\sum_n P(E_n)=\infty$ imples that $P(\limsup_n E_n)=1$. Frequently, applications of part (ii) are greatly hampered by the requirement of independence; a number of sufficient conditions for depen-

dent events to have the same conclusion as (ii) have been discovered. The **Chung-Erdös theorem** [9] is quite useful in this connection.

C. Random Variables

Let $(\Omega, \mathfrak{B}, P)$ be a probability space. A **random variable** is a real-valued function X defined on Ω that is $\mathfrak{B}$-measurable (i.e., for every real number a, the set $\{\omega \mid X(\omega)<a\}$ is in $\mathfrak{B}$). If $X_1, X_2, \ldots, X_n$ are random variables, the mapping $X=(X_1, X_2, \ldots, X_n)$ from Ω into $\mathbf{R}^n$ is said to be an **n-dimensional** (or **$\mathbf{R}^n$-valued**) **random variable**. More generally, a mapping X from $(\Omega, \mathfrak{B})$ into another †measurable space $(S, \mathfrak{E})$ is called an **$(S, \mathfrak{E})$-valued random variable** if it is measurable, that is, for every set A of $\mathfrak{E}$, the set $\{\omega \mid X(\omega) \in A\}$ belongs to $\mathfrak{B}$.

Let $\mathfrak{B}^1$ be the σ-algebra of all †Borel subsets of the real line $\mathbf{R}$. Then each random variable X induces a probability measure Φ on $(\mathbf{R}, \mathfrak{B}^1)$ such that

$$\Phi(A)=P(\{\omega \mid X(\omega) \in A\}), \qquad A \in \mathfrak{B}^1.$$

The measure Φ is called the **(1-dimensional) probability distribution of the random variable** X or simply the **distribution of** X. The point function F defined by

$$F(a)=P(\{\omega \mid X(\omega) \leqslant a\}), \qquad a \in \mathbf{R},$$

is a monotone nondecreasing and right continuous function such that $\lim_{a \rightarrow \infty} F(a)=1$, $\lim_{a \rightarrow -\infty} F(a)=0$. The function F is called the **cumulative distribution function** (or simply the **distribution function**) of the random variable X. Similarly, an n-dimensional random variable $X=(X_1, \ldots, X_n)$ induces its **n-dimensional probability distribution** (or simply **n-dimensional distribution**) and its **n-dimensional distribution function** $F(a_1, \ldots, a_n)=P(\{\omega \mid X_1(\omega) \leqslant a_1, \ldots, X_n(\omega) \leqslant a_n\})$. If the X_n $(n=1,2, \ldots, N)$ are k_n-dimensional random variables $(n=1,2, \ldots, N)$, we say that the $l(=\sum_{n=1}^{N} k_n)$-dimensional random variable $X=(X_1, X_2, \ldots, X_N)$ is the **joint random variable** of X_n $(n=1,2, \ldots, N)$ and that the (l-dimensional) distribution Φ of X is the **joint distribution** (or **simultaneous distribution**) of X_n $(n=1,2, \ldots, N)$. On the other hand, the k_n-dimensional distribution Φ_n of X_n is called the **marginal distribution** of the l-dimensional distribution Φ.

Given a finite sequence $\{X_n\}$ $(n=1,2, \ldots, N)$ of random variables, if the relation

$$P(\{\omega \mid X_n(\omega) \in A_n\ (n=1,2, \ldots, N)\})$$

$$=\prod_{n=1}^{N} P(\{\omega \mid X_n(\omega) \in A_n\}) \qquad (1)$$

holds for every choice of 1-dimensional †Borel sets A_n $(n=1,2, \ldots, N)$, we say that the random variables X_n $(n=1,2, \ldots, N)$ are mutually **independent** or that the sequence $\{X_n\}$ $(n=1,2, \ldots, N)$ is **independent**. Given an infinite family $\{X_\lambda\}$ $(\lambda \in \Lambda)$ of random variables, we say that the random variables are mutually **independent** or that the family is independent if every finite subfamily is independent. The latter definition of independence of random variables is compatible with the previous definition of independence of σ-subalgebras of $\mathfrak{B}$; i.e., if $\mathfrak{B}[X_\lambda]$ denotes the σ-subalgebras of $\mathfrak{B}$ generated by the sets $\{\omega \mid X_\lambda(\omega) \in A_\lambda\}$, with A_λ an arbitrary 1-dimensional Borel set, the independence of the family $\{X_\lambda\}$ $(\lambda \in \Lambda)$ in the latter sense is equivalent to the independence of the family $\{\mathfrak{B}[X_\lambda]\}$ $(\lambda \in \Lambda)$ in the previous sense. If the X_n $(n=1,2, \ldots, N)$ (X_λ $(\lambda \in \Lambda)$) are k_n- (k_λ-) dimensional random variables, then the independence of the family $\{X_n\}$ $(n=1,2, \ldots, N)$ ($\{X_\lambda\}$ $(\lambda \in \Lambda)$) is defined similarly; it is enough to take k_n- (k_λ-) dimensional Borel sets A_n (A_λ) for 1-dimensional Borel sets in equation (1).

Given a family $\{X_\lambda\}$ $(\lambda \in \Lambda)$ of random variables, the smallest σ-algebra with respect to which every X_λ is measurable is called the σ-algebra generated by $\{X_\lambda\}$ $(\lambda \in \Lambda)$ and is denoted by $\mathfrak{B}[X_\lambda \mid \lambda \in \Lambda]$. Each element of this class is said to be **measurable with respect to the family $\{X_\lambda\}$ $(\lambda \in \Lambda)$ of random variables**.

Since a random variable X is a $\mathfrak{B}$-measurable function, we can speak of the †integral of X relative to the measure P on $\mathfrak{B}$. If X is integrable relative to P, the integral of X over A is denoted by $E(X; A)$. $E(X; \Omega)$, usually denoted by $E(X)$, is called the **mean**, **expectation**, or **expected value** of X, denoted also by $M(X)$ or m_X. If $(X-E(X))^2$ is integrable,

$$V(X)=E((X-E(X))^2)$$

is called the **variance** of X, denoted by $\sigma^2(X)$. The **standard deviation** of X is the nonnegative square root $\sigma(X)$ of the variance. If X and Y are two random variables for which $E((X-E(X))(Y-E(Y)))$ exists, the value $E((X-E(X))(Y-E(Y)))$ is called the **covariance** of X and Y. When X and Y have finite variances, the **correlation coefficient** of X and Y is defined by

$$\rho(X, Y)=\frac{E((X-E(X))(Y-E(Y)))}{\{E((X-E(X))^2) E((Y-E(Y))^2)\}^{1/2}}.$$

It follows that $E(aX+bY)=aE(X)+bE(Y)$, $V(aX+b)=a^2 V(X)$ for any real numbers a, b, and that, in particular, $E(XY)=E(X)E(Y)$, $V(X+Y)=V(X)+V(Y)$ for mutually independent random variables X and Y. It also follows from the definition that $-1 \leqslant \rho(X, Y) \leqslant 1$ in all cases. The independence of X and Y implies that $\rho(X, Y)=0$, but the converse is false in general. The variance is important because of the well-known **Chebyshev inequal-**

ity: If X is a random variable with finite variance σ^2,

$$P(|X-E(X)|\geqslant c)\leqslant\sigma^2/c^2$$

for every positive number c.

D. Convergence of Random Variables

If $P(\lim_{n\to\infty}X_n=X_\infty)=1$, the sequence $\{X_n\}$ is said to **converge almost everywhere (almost certainly, almost surely** (a.s.), or **with probability** 1) to X_∞. If $\lim_{n\to\infty}P(|X_n-X_\infty|>\varepsilon)=0$ for every positive number ε, the sequence $\{X_n\}$ is said to **converge in probability** to X_∞. For a given positive number p the sequence $\{X_n\}$ is said to **converge in the mean of order** p to X_∞ if $\lim_{n\to\infty}E(|X_n-X_\infty|^p)=0$. Finally, if the random variables X_n $(n=1,2,\ldots,\infty)$ have distributions Φ_n $(n=1,2,\ldots,\infty)$, respectively, and if

$$\lim_{n\to\infty}\int_{-\infty}^{\infty}f(x)d\Phi_n(x)=\int_{-\infty}^{\infty}f(x)d\Phi_\infty(x)$$

for every continuous function f with compact support, the sequence $\{X_n\}$ is said to **converge in distribution** to X_∞. Note that the sequence of random variables converging in distribution may not converge in any ordinary sense. For example, random variables converging in distribution may even be defined on different probability spaces. On one hand, almost sure convergence does not in general imply convergence in the mean. On the other hand, either almost sure convergence or convergence in the mean implies convergence in probability, and convergence in probability implies convergence in distribution. However, P. Lévy [10] proved that if the X_n $(n=1,2,\ldots)$ are mutually independent, the sequence

$$Y_k=\sum_{n=1}^{k}X_n,\quad k=1,2,\ldots,$$

is convergent almost everywhere if and only if it is convergent in distribution (or in probability). The famous **three-series theorem** of Khinchin and Kolmogorov [11] claims that the series $\sum_n X_n$ with $X_1,X_2,\ldots$ independent is convergent almost surely if and only if there exists a sequence of independent random variables $X_1',X_2',\ldots$ such that each of the three series

$$\sum_n P(X_n\neq X_n'),\quad \sum_n E(X_n'),\quad \sum_n V(X_n')$$

is convergent.

E. Conditional Probability and Conditional Expectation

Let $(\Omega,\mathfrak{B},P)$ be a probability space and $\mathfrak{F}$ a σ-subalgebra of $\mathfrak{B}$. If X is a random variable with finite mean, the function

$$\mu(E)=\int_E X(\omega)dP,\quad E\in\mathfrak{F},$$

defines on $(\Omega,\mathfrak{F})$ a completely additive set function which is †absolutely continuous with respect to P. Therefore, by the †Radon-Nikodym theorem, there is an $\mathfrak{F}$-measurable function f such that

$$\mu(E)=\int_E f(\omega)dP\quad\text{for every } E\in\mathfrak{F}.$$

This function is unique up to a set of P-measure zero and is called the **conditional expectation** (or **conditional mean**) of X relative to $\mathfrak{F}$, denoted by $E(X\mid\mathfrak{F})$. When $\mathfrak{F}$ is generated by a random variable Y, we also write $E(X\mid Y)$ for $E(X\mid\mathfrak{F})$ and call it the conditional expectation of X relative to Y. In this case, there is a †Borel measurable function f such that $E(X\mid Y)=f(Y(\omega))$, and we write $E(X\mid Y=y)$ for $f(y)$. The same fact holds when Y is a multidimensional random variable. It follows from the definition that the conditional expectation has the following properties, up to a set of P-measure zero: (i) if $X\geqslant 0$, then $E(X\mid\mathfrak{F})\geqslant 0$; (ii) $E(aX+bY\mid\mathfrak{F})=aE(X\mid\mathfrak{F})+bE(Y\mid\mathfrak{F})$; (iii) $E(E(X\mid\mathfrak{F}))=E(X)$; (iv) if X and $\mathfrak{F}$ are mutually independent, i.e., $\mathfrak{B}[X]$ and $\mathfrak{F}$ are mutually indepent, then $E(X\mid\mathfrak{F})=E(X)$; (v) if X is $\mathfrak{F}$-measurable, then $E(X\mid\mathfrak{F})=X$ and $E(XY\mid\mathfrak{F})=XE(Y\mid\mathfrak{F})$; (vi) if $\lim_{n\to\infty}X_n=X_\infty$ with $|X_n|\leqslant Y$ and Y is an integrable random variable, then $\lim_{n\to\infty}E(X_n\mid\mathfrak{F})=E(X_\infty\mid\mathfrak{F})$; (vii) if $\mathfrak{G}$ is a σ-subalgebra of $\mathfrak{F}$, then $E(E(X\mid\mathfrak{F})\mid\mathfrak{G})=E(X\mid\mathfrak{G})$; (viii) if X^2 is integrable and Y is any $\mathfrak{F}$-measurable random variable, then $E((X-E(X\mid\mathfrak{F}))^2)\leqslant E((X-Y)^2)$.

When X is the **indicator function** (i.e., the †characteristic function) χ_E of a set E in $\mathfrak{B}$, $E(\chi_E\mid\mathfrak{F})$ is called the **conditional probability** of E relative to $\mathfrak{F}$ and is denoted by $P(E\mid\mathfrak{F})$. In particular, if $\mathfrak{F}=\{F,F^c,\varnothing,\Omega\}$ with $1>P(F)>0$, $P(E\mid\mathfrak{F})$ is the simple function which takes the values $P(E\cap F)/P(F)$ on F and $P(E\cap F^c)/P(F^c)$ on F^c. These values are denoted respectively by $P(E\mid F)$ and $P(E\mid F^c)$. The definition of $P(E\mid Y)$ or $P(E\mid Y=y)$ is also the same as in the case of the conditional expectation.

Let $\mathfrak{F}$ be a σ-subalgebra of $\mathfrak{B}$ and Y a real random variable. According to the foregoing definition, $P(Y\in E\mid\mathfrak{F})$ or $P(Y^{-1}(E)\mid\mathfrak{F})$ is the conditional probability of the occurrence of the event $Y\in E$ under $\mathfrak{F}$. Since $P(Y\in E\mid\mathfrak{F})$ is determined except on a P-null set depending on E, an arbitrary version of $P(Y\in E\mid\mathfrak{F})$, viewed as a function of E, does not always satisfy the conditions of a probability measure. However, we can prove that there exists a nice version of $P(Y\in E\mid\mathfrak{F})$ which is a proba-

bility measure in $E \in \mathfrak{B}^1$ for every $\omega \in \Omega$ and that such a version is unique almost surely. This version is called a **regular conditional probability** of $Y \in E$ under $\mathfrak{F}$ or the **conditional probability distribution** of Y under $\mathfrak{F}$; this is written as $P_Y(E \mid \mathfrak{F})$. $P(Y \in E \mid X)$, $P(Y \in E \mid X = x)$, $P_Y(E \mid X)$, and $P_Y(E \mid X = x)$ are interpreted similarly. The conditional probability distribution can be defined not only for real random variables but also for every random variable which takes values in an †analytic measurable space.

F. Bayes's Formula

Let $E_1, E_2, \ldots, E_n$ be pairwise exclusive events, and assume that one of them must occur. If E is another random event, we have

$$P(E_i \mid E) = \frac{P(E_i)P(E \mid E_i)}{P(E_1)P(E \mid E_1) + \ldots + P(E_n)P(E \mid E_n)},$$

where $P(E_i)$ is the probability of the event E_i and $P(E \mid E_i)$ is the conditional probability of E under the assumption that the event E_i has occurred. This is called **Bayes's formula**. In practical applications $E_1, \ldots, E_n$ usually represent n unknown hypotheses. Suppose that the probabilities on the right-hand side of the formula are given. We then apply Bayes's formula to reevaluate the probability of each hypothesis E_i knowing that some event E has occurred as the result of a trial. This is why $P(E_i)$ ($P(E_i \mid E)$) is called the **a priori (a posteriori) probability**. However, the determination of the values of a priori probabilities is sometimes difficult, and we often set $P(E_i) = 1/n$ in practical applications, although this has caused a great deal of criticism.

When X is a random variable subject to the distribution with continuous probability density $f(x)$, Bayes's formula is extended to the following form:

$$f(x_0 \mid E) = \frac{f(x_0)P(E \mid X = x_0)}{\int_{-\infty}^{\infty} P(E \mid X = x) f(x)\, dx},$$

where $f(x \mid E)$ is the conditional probability density of the random variable X under the assumption that the event E has occurred, and $P(E \mid X = x_0)$ is the conditional probability of E relative to X.

G. Zero-One Laws

In probability theory there are many theorems claiming that an event with certain properties has probability 0 or 1. Such theorems are called **zero-one laws**. Here, we mention two famous examples, Kolmogorov's zero-one law [6] and the Hewitt-Savage zero-one law. Let $\alpha = \alpha(X_1, X_2, \ldots)$ be an event concerning a sequence of random variables $\{X_n\}$. α is called a **tail event** concerning $\{X_n\}$ if for every n, occurrence or nonoccurrence of α depends only on $\{X_n, X_{n+1}, \ldots\}$. For example, $\{\lim_{n\to\infty} X_n = 0\}$ is a tail event. α is called a **symmetric event** concerning $\{X_n\}$ if occurrence or nonoccurrence of α is invariant under every finite permutation of $X_1, X_2 \ldots$. For example, the event that $\sum_{k=1}^{n} X_k > 0$ for infinitely many n's is a symmetric event. **Kolmogorov's zero-one law**: Every tail event concerning a sequence of independent random variables has probability 0 or 1. **Hewitt-Savage zero-one law**: Every symmetric event concerning a sequence of independent and identically distributed random variables has probability 0 or 1.

Kolmogorov's zero-one law can be extended as follows: Let $\mathfrak{F}_n$, $n = 1, 2, \ldots$, be a sequence of independent σ-subalgebras of $\mathfrak{B}$. Then the σ-algebra $\mathfrak{T} = \bigcap_k \bigcup_{n>k} \mathfrak{F}_n$, called the **tail σ-algebra** of $\{\mathfrak{F}_n\}$, is trivial, i.e., $P(A) = 0$ or 1 for every $A \in \mathfrak{T}$. Kolmogorov's zero-one law is a special case where $\mathfrak{F}_n$ is the σ-algebra generated by X_n for every n.

References

[1] I. Todhunter, A history of the mathematical theory of probability from the time of Pascal to that of Laplace, Macmillan, 1865 (Chelsea, 1949).
[2] J. Bernoulli, Ars conjectandi, 1713.
[3] A. de Moivre, The doctrine of chances, 1718.
[4] P. S. Laplace, Théorie analytique des probabilités, Paris, 1812.
[5] R. von Mises, Vorlesungen aus dem Gebiete der angewandten Mathematik I, Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik, Franz Deuticke, 1931.
[6] A. Kolmogorov (Kolmogoroff), Grundbefriffe der Wahrscheinlichkeitsrechnung, Erg. Math., Springer, 1933; English translation, Foundations of the theory of probability, Chelsea, 1950.
[7] E. Borel, Les probabilités dénombrables et leur application arithmetiques, Rend. Circ. Mat. Palermo, 27 (1909), 247–271.
[8] N. Wiener, Differential space, J. Math. Phys., 2 (1923), 131–174.
[9] K. L. Chung and P. Erdös, On the application of the Borel-Cantelli lemma, Trans. Amer. Math. Soc., 72 (1952), 179–186.
[10] P. Lévy, Théorie de l'addition des variables aléatoires, Gauthier-Villars, 1937.
[11] J. L. Doob, Stochastic processes, Wiley, 1953.

[12] M. M. Loève, Probability theory, Van Nostrand, third edition, 1963.
[13] W. Feller, An introduction to probability theory and its applications, Wiley, I, second edition, 1957; II, 1966.
[14] L. Breiman, Probability, Addison-Wesley, 1968.
[15] J. Lamperti, Probability, Benjamin, 1966.

343 (VI.14) Projective Geometry

A. Introduction

Projective geometry is the most fundamental of classical geometries and one of the first examples of axiomatized mathematics.

B. Construction of Projective Geometry

We construct projective geometry axiomatically [4]. Given two sets P, Q and a †relation $\Gamma \subset P \times Q$, consider the triple $\mathfrak{P} = \{P, Q, \Gamma\}$. We call each element of P a **point** and each element of Q a **line**. If $(p, l) \in \Gamma$ holds for a point p and a line l, then we say that the line l contains the point p. When two lines l_1 and l_2 contain a point p, we say that they intersect at p. When several points are contained in the same line, these points are said to be **collinear**, and when several lines contain the same point, these lines are said to be **concurrent**. For $\mathfrak{P}$ we impose the following axioms:

(I) There exists one and only one line that contains two given distinct points.

(II) Suppose that we are given noncollinear points p_0, p_1, and p_2, and distinct points q_1, q_2. Now suppose that $\{p_0, p_1, q_1\}$ and $\{p_0, p_2, q_2\}$ are collinear triples. Then the line containing p_1, p_2 and the line containing q_1, q_2 necessarily intersect (Fig. 1).

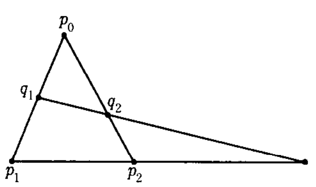


Fig. 1

(III) Every line contains at least three distinct points.

The $\mathfrak{P}$ that satisfy axioms (I) and (II) and axioms (I), (II), and (III) are called **general projective geometry** and **projective geometry**, respectively. The set of all points that are contained in a line is called the **point range** with the line as its **base**. In projective geometry, there exists a one-to-one correspondence between the set of lines and the set of point ranges, so we can identify every line with a point range. In this case a line $l \in Q$ is represented as a subset of P, and the relation $(p, l) \in \Gamma$ means that the point p belongs to the set l.

Let S be a subset of P and p_1, p_2 be any two distinct points of S. If the line that contains p_1 and p_2 is always contained in S, then S is called a **subspace**. Points and lines are subspaces. Now we impose the following axiom:

(IV) There exist a finite number of points such that any subspace that contains all of them contains P.

We call a projective geometry satisfying axiom (IV) a **finite-dimensional projective geometry**, which from this point on will be the sole object of our consideration. We call P a **projective space**. Consider sequences of subspaces of the type $P \supsetneqq P_{n-1} \supsetneqq \ldots \supsetneqq P_1 \supsetneqq P_0 \neq \varnothing$, where $\varnothing$ is the empty set. The number n of the longest sequence is called the **dimension** of P. If P is of dimension n, we write P^n instead of P. We call P^1 a **projective line** and P^2 a **projective plane**. Each subspace S of P, together with the set of lines of P contained in S, gives a finite-dimensional projective geometry, and so S is a projective space. Lines and points are projective spaces of dimensions 1 and 0, respectively. By convention, the empty set is a (-1)-dimensional projective space. We call each 2-dimensional subspace a **plane** and each $(n-1)$-dimensional subspace in P^n a **hyperplane**.

Let M, N be subspaces of P, and for a pair of points $p \in M$, $q \in N$ consider the set $p \cup q$ of all points on the line that contains p and q. The set $\{p \cup q \mid p \in M, q \in N\}$ is denoted by $M \cup N$, and we call it the set spanned by M and N. By convention, we put $\varnothing \cup M = M$ and $p \cup p = p$. Then $P^r \cup P^s$ is the projective space of the lowest dimension which contains P^r and P^s. On the other hand, if we denote the intersection of P^r and P^s by $P^r \cap P^s$, then it is the projective space of highest dimension that is contained in both of them. We call $P^r \cup P^s$ and $P^r \cap P^s$ the **join** and the **intersection** of P^r and P^s, respectively. When the dimension of the space spanned by $r+1$ points is r, we say that these points are **independent**; otherwise they are **dependent**. If any $r+1$ points of a given subset M of P^n are independent for each $r \leqslant n$, we say that points of M lie in a **general position**. The space P^r necessarily contains $r+1$ independent points, and there necessarily exists a P^r that contains $r+1$ arbitrary given points in a projective space; it is unique if the points are indepen-

dent. If $P^r \cup P^s = P^t$ and $P^r \cap P^s = P^u$, then $r+s=t+u$. We call the latter the **dimension theorem** (or **intersection theorem**) of projective geometry.

The set Σ_1 of all hyperplanes that contain a P_0^{n-2} in P^n is called a **pencil of hyperplanes**, and the P_0^{n-2} common to them is called the **center** of Σ_1. If a pencil of hyperplanes contains two distinct hyperplanes of P^n, then the pencil is determined uniquely by these two. When $n=2$ and 3, it is called a **pencil of lines** and a **pencil of planes**, respectively. Each pencil of hyperplanes of P^n, or more generally, each pencil of hyperplanes of a subspace of an arbitrary dimension in P^n, is called a **linear fundamental figure** of P^n or simply a **fundamental figure**. In P^n, the set Σ_r of all $P^{n-r}, P^{n-r+1}, \ldots, P^{n-1}$ that contain the same P_0^{n-r-1} is called the **star** with center P_0^{n-r-1}. Each set that consists of the totality of subspaces of an arbitrary demension in the same P^r or a subset of it is called a **P^r-figure**.

Under the assumption that P^r and P^s do not have points in common, the operation of constructing $P^r \cup P^s$ from P^r and P^s is called **projecting** P^s from P^r. Assuming that P^r and P^s have points in common the operation of constructing $P^r \cap P^s$ from P^r and P^s is called **cutting** P^s by P^r. Suppose that we are given spaces P_0, P_1, and P_2, and a fundamental figure Σ in the space P_1. By projecting Σ from P_0 and then cutting it by P_2, we can construct a fundamental figure Σ' on P_2. This operation is called **projection** of Σ from P_0 onto P_2, and we call P_0 the **center of projection** (Fig. 2). In this case, we say that Σ and Σ' are **in perspective** and denote the relation by $\Sigma \overline{\wedge} \Sigma'$. If for two fundamental figures Σ and Σ' there exist a finite number of fundamental figures F_i $(1 \leqslant i \leqslant l)$ such that $\Sigma \overline{\wedge} F_1 \overline{\wedge} \ldots \overline{\wedge} F_l \overline{\wedge} \Sigma'$, then we say that Σ and Σ' are **projectively related** to each other and denote this by $\Sigma \barwedge \Sigma'$ (Fig. 3). Now for arbitrary subspaces P_1^r, P_2^r $(0 \leqslant r \leqslant n)$, we take P_0^{n-r-1} that

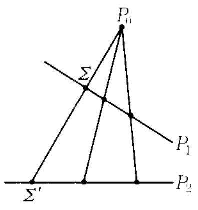


Fig. 2

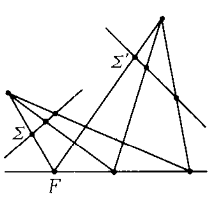


Fig. 3

have no points in common with them and project each point of P_1^r onto P_2^r from P_0^{n-r-1}. The one-to-one correspondence $P_1^r \to P_2^r$ thus obtained is called a **perspective mapping**. If a one-to-one correspondence $P_1^r \to P_2^r$ is represented as the composite of a finite number of perspective mappings, then we call it a **projective mapping**. These mappings are extended to those of fundamental figures, too.

Suppose that in a proposition or a figure in P^n, we interchange P^r and P^{n-r-1} $(0 \leqslant r \leqslant n)$ and also interchange *contains* and *is contained* (and related terms). The proposition or the figure thus obtained is said to be **dual** to the original one. In projective geometry, if a proposition is true, then its dual proposition is also true (**duality principle**). This is assured because propositions dual to axioms (I)–(IV) hold; and P^r and Σ_r are dual to each other. The projective space P_*^n obtained by the principle of duality, by regarding the hyperplanes of P^n as points of P_*^n, is called the **dual space** of P^n.

C. Projective Coordinates

Here we introduce **projective coordinates** in P^n. Consider **Desargues's theorem**: Suppose that p_1, p_2, p_3 and q_1, q_2, q_3, are two sets of points in P^n, each of which is independent and satisfies $p_i \neq q_i$ $(i=1,2,3)$. If the three lines $p_i \cup q_i$ $(i=1,2,3)$ are concurrent, then the three points $(p_2 \cup p_3) \cap (q_2 \cup q_3)$, $(p_3 \cup p_1) \cap (q_3 \cup q_1)$, $(p_1 \cup p_2) \cap (q_1 \cup q_2)$ are collinear. The converse is also true. This theorem holds for $n \geqslant 3$ generally. However, when $n=2$, there exist projective geometries for which it does not hold; we call these **non-Desarguesian geometries**. In such cases it is impossible to introduce coordinates, so we assume Desargues's theorem for $n=2$.

When four points p_i $(1 \leqslant i \leqslant 4)$ in P^n lie on the same plane and in general position, we call the figure that consists of these four points and the six straight lines $g_{ij} = p_i \cup p_j$ $(1 \leqslant i \leqslant j \leqslant 4)$ a **complete quadrangle** $p_1 p_2 p_3 p_4$; each p_i is called a **vertex**, and each g_{ij} is called a **side**. If six points q_i $(1 \leqslant i \leqslant 6)$ on a line l are points of intersection of six sides $g_{12}, g_{13}, g_{14}, g_{34}, g_{24}, g_{23}$ of a complete quadrangle with l, we call

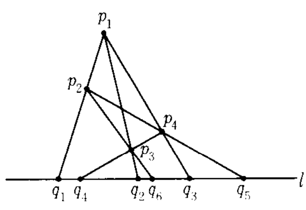


Fig. 4

them a **quadrangular set of six points** (Fig. 4). By Desargues's theorem, we can show that if there are given three fixed distinct points on a line l, then any pair of distinct points on l determines uniquely a point on l such that the six points thus obtained constitute a quadrangular set. The quadrangular property is invariant under projective mappings. On a line l we fix three mutually distinct points p_0, p_1, p_∞. For any two points p_x, p_y different from p_∞ on l, we take the point s such that $p_\infty, p_x, p_0, p_\infty, p_y, s$ constitute a quadrangular set of six points and call s the **sum** of p_x and p_y with respect to $[p_0, p_\infty, p_1]$ (Fig. 5). On the other hand, the point t such that $p_0, p_x, p_1, p_\infty, p_y, t$ constitute a quadrangular set of six points is called the **product** of p_x and p_y with respect to $[p_0, p_\infty, p_1]$ (Fig. 6). When we are given a fixed triple $[p_0, p_\infty, p_1]$ on a line l, as before, the set of points on l not equal to p_∞ is called a **point range of the number system**, provided that we exclude p_∞ from the point range. We call the set of three points $[p_0, p_\infty, p_1]$ a **frame** (or **projective frame**) of l, and we call p_0 the **origin**, p_1 the **unit point**, and p_∞ the **supporting point**.

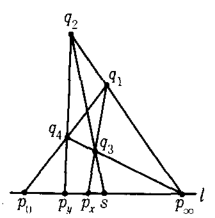


Fig. 5

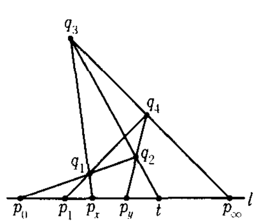


Fig. 6

A point range of the number system constitutes a †field (which may be noncommutative) with respect to the previously defined sum and product. We call the field a **Staudt algebra**, and an abstract algebra isomorphic to it is called a **coefficient field** of P^n. We denote by $K(p_0, p_\infty, p_1)$ the Staudt algebra that is determined by a frame $[p_0, p_\infty, p_1]$. A projective mapping of l onto itself that leaves invariant each of three distinct points p_0, p_∞, p_1 on l is necessarily an †inner automorphism of the field $K(p_0, p_\infty, p_1)$. Denoting by $[p_0, p_\infty, p_1]$ a frame on a line l of P^n with coefficient field K, we call each isomorphism $\theta: K(p_0, p_\infty, p_1) \rightarrow K$ a **coordinate system** of l. For each point p on l we call the element $\xi = \theta(p)$ of K the **inhomogeneous coordinate** of p with respect to this frame. Also, we call the pair (x^0, x^1) such that $x^0, x^1 \in K$ and $x^1(x^0)^{-1} = \xi$ **homogeneous coordinates** of p. Since the supporting point p_∞ is excluded from $K(p_0, p_\infty, p_1)$, we fix $(0, x^1)$ such that $x^1 \neq 0$ as the homogeneous coordinates of P_∞. In order for (x^0, x^1) and (y^0, y^1) to be homogeneous coordinates of the same point, it is necessary and sufficient that there exist an element $\lambda \neq 0$ of K such that $y^\alpha = x^\alpha \lambda$ $(\alpha = 0, 1)$.

In conformity with these results, we now introduce coordinates in P^n. A set $\mathfrak{F} = [a_0, a_1, \ldots, a_n, u]$ of ordered $n+2$ points in a general position is called a **frame** (or **projective frame**) of P^n; each of a_α $(0 \leqslant \alpha \leqslant n)$ is called a **fundamental point**, and u is called a **unit point**. For $A = a_{\alpha_0} \cup \ldots \cup a_{\alpha_r}$ $(0 \leqslant \alpha_0 < \ldots < \alpha_r \leqslant n)$, we denote by A^* the space spanned by the remaining fundamental points. For any point p that is not contained in A^*, we put $p_A = A \cap (p \cup A^*)$ and call it the **component** of p on A. Then $\mathfrak{F}_A = [a_{\alpha_0}, \ldots, a_{\alpha_r}, u_A]$ is a frame of A. Hereafter, we shall omit u_A for brevity. Suppose that isomorphisms $\theta_{\alpha\beta}: K(a_\alpha, a_\beta) \rightarrow K$ are assigned for each pair α, β $(0 \leqslant \alpha < \beta \leqslant n)$. Under a certain condition, the system $\{\theta_{\alpha\beta}\}$ is determined by one of the $\theta_{\alpha\beta}$. In this case we denote $\{\theta_{\alpha\beta}\}$ by θ and call $\{\mathfrak{F}, \theta\}$ a **projective coordinate system** of P^n. For any point p of P^n not contained in $A_0 = a_1 \cup \ldots \cup a_n$, we denote by p_i the component of p on $a_0 \cup a_i$ $(1 \leqslant i \leqslant n)$, and we put $\xi^i = \theta(p_i)$. The elements of the ordered set $(\xi^1, \xi^2, \ldots, \xi^n)$ are called the **inhomogeneous coordinates** of p with respect to $\mathfrak{F}$, and those of the set $(x^0, x^1, \ldots, x^n)$ such that $x^i(x^0)^{-1} = \xi^i$ are called the **homogeneous coordinates** of p. When p is contained in A_0, we define $(0, x^1, \ldots, x^n)$ as homogeneous coordinates of p with respect to $\mathfrak{F}$, provided that $(x^1, \ldots, x^n)$ are homogeneous coordinates of p with respect to $\mathfrak{F}_{A_0}$.

Now we represent the point whose coordinates are $(x^0, x^1, \ldots, x^n)$ simply by x. In P^n, when coordinates are introduced, a necessary and sufficient condition for points z to be on the line that passes through two distinct points x and y is that $z^\alpha = x^\alpha \lambda + y^\alpha \mu$ $(0 \leqslant \alpha \leqslant n)$, when $\lambda, \mu \in K$ are parameters. More generally, a point z is contained in the space spanned by $r+1$ independent points x_β $(0 \leqslant \beta \leqslant r)$ in P^n if and only if $z^\alpha = \sum_{\beta=0}^{r} x_\beta^\alpha \lambda^\beta$ $(0 \leqslant \alpha \leqslant n, \lambda^\beta \in K)$. In particular, the equation of a hyperplane is represented in the form $\sum_{\alpha=0}^{n} X_\alpha z^\alpha = 0$ $(X_\alpha \in K)$ with respect to variable coordinates z^α. Therefore each hyperplane is uniquely determined by the ratio of $X_0, X_1, \ldots, X_n$. We call $X_0, X_1, \ldots, X_n$ **hyperplane coordinates** of the hyperplane. If $n = 2$, they are called **coordinates** of a line, and if $n = 3$, **plane coordinates** of a plane. (For coordinates of P^r in P^n → 90 Coordinates B.)

D. Projective Transformations

A one-to-one correspondence φ between the point sets of two projective spaces P^n and $\bar{P}^n$ is called a **collineation in the wider sense** if for any three points p_1, p_2, p_3 that are collinear, $\varphi(p_i)$ $(i=1,2,3)$ are also collinear and vice versa. If $\bar{P}^n = P^n_*$, we call φ a **correlation**; if $\bar{P}^n = P^n$, we call φ a **collineation**. If we denote a correlation by τ_0, any other correlation is obtained as a composite of τ_0 and a collineation. If τ is a correlation, it naturally induces a mapping $P^n_* \to P^n$, which we also denote by τ. Then $\tau \circ \tau$ is a collineation. If $\tau \circ \tau$ is an identity, we call τ an **involutive correlation**. Suppose that $\varphi: P^n \to \bar{P}^n$ is a collineation in the wider sense and $0 \leqslant r \leqslant n-1$. Then φ induces a one-to-one correspondence between the set of r-dimensional subspaces of P^n and the set of r-dimensional subspaces of $\bar{P}^n$; and if $P^r \supset P^s$ in P^n, then $\varphi P^r \supset \varphi P^s$.

Next, suppose that we are given two projective spaces P^n and $\bar{P}^n$ that are subspaces of a space P^N $(n<N)$. (When Desargues's theorem holds, any two projective spaces of the same dimension can be identified with subspaces of a projective space of higher dimension.) In this case, when a collineation in the wider sense $\varphi: P^n \to \bar{P}^n$ is a projective mapping, we call it a **projective collineation in the wider sense**. A **projective collineation** is also called a **projective transformation**. The totality of collineations of P^n constitutes a †transformation group and is called the **group of collineations** of P^n; we denote it by $\mathfrak{C}(P^n)$. The totality of projective transformations of P^n constitutes a †normal subgroup of $\mathfrak{C}(P^n)$; we denote it by $\mathfrak{G}(P^n)$ and call it the **group of projective transformations**. The totality of projective transformations that leave invariant a frame $\mathfrak{F}$ of P^n constitutes a subgroup $\mathfrak{G}_0^{n+1}(\mathfrak{F})$. It is isomorphic to the group of †inner automorphisms $\mathfrak{J}(K)$ of the coefficient field K of P^n. A collineation is not necessarily a projective transformation. The former is obtained as a composite of a projective transformation and an automorphism of the coefficient field. Specifically, if we denote the group of †automorphisms of K by $\mathfrak{A}(K)$, then $\mathfrak{C}(P^n)/\mathfrak{G}(P^n) \cong \mathfrak{A}(K)/\mathfrak{J}(K)$. Hence in order for all collineations to be projective transformations, it is necessary and sufficient that all automorphisms of the coefficient field be inner automorphisms. If the coefficient field is the real number field, then collineations are always projective transformations. For the complex number field, however, this is not necessarily true.

Now, we consider the following three propositions: (1) The coefficient field of P^n is commutative. (2) Given frames $\mathfrak{F}$ and $\mathfrak{F}'$ of P^n, there exists a unique projective transformation sending $\mathfrak{F}$ onto $\mathfrak{F}'$. (3) Given two distinct lines l_1 and l_2 contained in a plane in P^n and two sets of three distinct points p_i $(i=1,2,3)$ and q_i $(i=1,2,3)$ that lie on l_1 and l_2 respectively, then the three points $(p_2 \cup q_3) \cap (p_3 \cup q_2)$, $(p_3 \cup q_1) \cap (p_1 \cup q_3)$, and $(p_1 \cup q_2) \cap (p_2 \cup q_1)$ are collinear. These three propositions are mutually equivalent. We call proposition (2) the **fundamental theorem of projective geometry** and proposition (3) the **theorem of Pappus**. If the coefficient field is the real (complex) number field, we call the projective space a **real (complex) projective space**. In classical geometry, only these cases were studied.

Suppose that the coefficient field is commutative. Then, if we assign an isomorphism $\theta_0: K(p_0, p_\infty, p_1) \to K$ for the Staudt algebra $K(p_0, p_\infty, p_1)$ on a line in a space, then the isomorphism θ of the Staudt algebra $K(q_0, q_\infty, q_1)$ on an arbitrary line onto K can be uniquely determined so that $\theta^{-1} \circ \theta_0$ is a projective mapping. Utilizing such isomorphisms, we can determine homogeneous coordinates in an arbitrary subspace of P^n by a frame on it.

Suppose that the coefficient field is a commutative field whose characteristic is not 2. For four collinear points p_i $(1 \leqslant i \leqslant 4)$ in P^n, where p_1, p_2, p_3 are distinct and $p_4 \neq p_1$, we consider a frame such that p_1, p_2, and p_3 are, respectively, the supporting point, the origin, and the unit point. The inhomogeneous coordinate λ of p_4 with respect to this frame is called the **anharmonic ratio (cross ratio** or **double ratio)** of these four points and is denoted by $[p_1, p_2; p_3, p_4]$. If we denote the inhomogeneous coordinates of p_i with respect to a general frame by (x_i^0, x_i^1) $(i=1,2,3,4)$, then λ can be expressed as

$$\lambda = [p_1, p_2; p_3, p_4] = \frac{(x_1^0 x_3^1 - x_3^0 x_1^1)(x_2^0 x_4^1 - x_4^0 x_2^1)}{(x_1^0 x_4^1 - x_4^0 x_1^1)(x_2^0 x_3^1 - x_3^0 x_2^1)}.$$

Moreover, if we interchange the order of the four points, then we have

$$\lambda = [p_2, p_1; p_4, p_3] = [p_3, p_4; p_1, p_2] = [p_4, p_3; p_2, p_1],$$

$$[p_1, p_2; p_4, p_3] = \frac{1}{\lambda},$$

$$[p_1, p_3; p_2, p_4] = 1 - \lambda,$$

$$[p_1, p_3; p_4, p_2] = \frac{1}{1-\lambda},$$

$$[p_1, p_4; p_2, p_3] = \frac{\lambda - 1}{\lambda},$$

$$[p_1, p_4; p_3, p_2] = \frac{\lambda}{\lambda - 1}.$$

In general, these six values are different; however, there are the following two exceptions: when $\lambda=-1$, $1/2$, and 2; and when λ is a root of $\lambda^2-\lambda+1=0$. When $\lambda=-1$, these four points are called a **harmonic range of points**, and the points p_3, p_4 are called **harmonic conjugates** with respect to p_1, p_2; or p_1, p_2 and p_3, p_4 are said to be **harmonically separated** from each other. When $\lambda^2-\lambda+1=0$, these four points are said to be an **equianharmonic range of points**. For the dual of these, we can consider the anharmonic ratio of four hyperplanes of a pencil of hyperplanes. The concept of the anharmonic ratio can be extended further to the case of four elements of fundamental figures in general. The anharmonic ratio is a quantity that is invariant under projective transformations.

Each projective transformation $x \rightarrow \bar{x}$ is expressed with respect to homogeneous coordinates x^α $(0 \leqslant \alpha \leqslant n)$ of P^n as

$$\rho\bar{x}^\alpha=\sum_{\beta=0}^{n} t^\alpha_\beta x^\beta, \qquad \rho, t^\alpha_\beta \in K,$$

$$\rho \neq 0, \qquad \det(t^\alpha_\beta) \neq 0. \tag{1}$$

Conversely, if $T=(t^\alpha_\beta)$ is a †regular matrix $(t^\alpha_\beta \in K)$, then (1) determines a projective transformation. So there is a one-to-one correspondence between projective transformations and †equivalence classes of the regular matrices $T=(t^\alpha_\beta)$ with the †equivalence relation $T \sim \lambda T$ $(\lambda \in K \setminus \{0\})$. Therefore, when K is commutative, the group of projective transformations $\mathfrak{G}(P^n)$ of P^n is isomorphic to the factor group $PGL(n+1, K)$ of the †general linear group $GL(n+1, K)$ with the coefficient field K by its center $\{\rho I \mid \rho \in K \setminus \{0\}\}$; that is, $\mathfrak{G}(P^n) \cong PGL(n+1, K)$.

Extending the definition of projective transformations, we call the transformation represented by (1), with an arbitrary square matrix that is not necessarily regular, a projective transformation. When T is regular, it is called a **regular projective transformation**, and when T is not regular, it is called a **singular projective transformation**. In particular, if the †rank of T is $n+1-h$, then we say that the projective transformation is **singular of the hth species**. If (1) is singular of the hth species, $n+1$ hyperplanes $\sum_{\beta=0}^{n} t^\alpha_\beta x^\beta=0$ $(0 \leqslant \alpha \leqslant n)$ have a space P^{h-1} in common. We call P^{h-1} the **singular subspace** of this transformation. A projective transformation is not defined on its singular subspace. A singular projective transformation of the hth species is the composite of the projection of P^n onto some P^{n-h} with the singular subspace as its center and a regular projective transformation of P^{n-h}.

If the coordinates of a point are denoted by (x^α) and hyperplane coordinates with respect to some frame by (X_α), then the linear transformation

$$\tau_*: \rho X_\alpha=\sum_{\beta=0}^{n} t^*_{\alpha\beta} x^\beta, \qquad \rho, t^*_{\alpha\beta} \in K, \tag{2}$$

is a correlation. (Here also, we extend the definition of correlation and include the case where $T_*=(t^*_{\alpha\beta})$ is not regular.) The condition that τ_* is an involutive correlation is given by $T_*=\pm{}^tT_*$. When $T_*=-{}^tT_*$, the involutive correlation τ_* is called a **null system**. The correlation τ_* is a null system if and only if any point x of P^n is contained in the hyperplane $\tau_*(x)$. When $T_*={}^tT_*$, we call the involutive correlation τ_* a **polar system**. For a polar system τ_*, the set of points x that are contained in hyperplanes $\tau_*(x)$ constitutes a **quadric hypersurface** (or **hyperquadric**).

E. Quadric Hypersurfaces

Let τ_* be a polar system, and let Q_2^{n-1} be the totality of points x contained in $\tau_*(x)$. Then the equation of the quadric hypersurface Q_2^{n-1} is given by

$$\sum_{\alpha,\beta=0}^{n} t^*_{\alpha\beta} x^\alpha x^\beta=0. \tag{3}$$

For such a correlation τ_* we call a relation between the set of points x of P^n and the set of hyperplanes $\tau_*(x)$ a **polarity** with respect to Q_2^{n-1}. We call $\tau_*(x)$ the **polar** of x with respect to Q_2^{n-1}, and x the **pole** of $\tau_*(x)$ with respect to Q_2^{n-1}. If the points of intersection of a line passing through a point x with Q_2^{n-1} and $\tau_*(x)$ are denoted by z_1, z_2; y, then x, y; z_1, z_2 is a harmonic range of points. When a point x lies on the polar of a point y, we say that x and y are mutually **conjugate**. Each point on Q_2^{n-1} is conjugate with itself, and the converse is also true. We call the polar of a point on Q_2^{n-1} the **tangent hyperplane** of Q_2^{n-1} at that point.

If τ_* is regular or singular of the hth species, we call the corresponding quadric hypersurface **regular** or **singular of the hth species**. If τ_* is singular of the hth species, its singular subspace is contained in Q_2^{n-1}. We call points on this singular subspace **singular points** of Q_2^{n-1}. Q_2^{n-1}, which is singular of the first species (i.e., Q_2^{n-1} with just one singular point), is called a **cone**.

We call a subspace contained in Q_2^{n-1} a **generating space**. If it is a line we call it a **generating line**. We put $q=(n-2)/2$ or $(n-1)/2$ according as n is even or odd. Then, if the coefficient field is an †algebraically closed field, for each regular Q_2^{n-1} there necessarily exist q-dimensional generating spaces. Also, Q_2^2 is a †ruled surface covered by two families

of generating lines, and Q_2^{2k} is covered by two families of k-dimensional generating spaces.

If p_i $(1 \leqslant i \leqslant 5)$ are five points in a general position in a plane, then there exists one and only one Q_2^1 passing through these points; we call Q_2^1 a **conic**. In order that six points p_i $(1 \leqslant i \leqslant 6)$ in a plane lie on Q_2^1, it is necessary and sufficient that the three points $(p_1 \cup p_2) \cap (p_4 \cup p_5)$, $(p_2 \cup p_3) \cap (p_5 \cup p_6)$, and $(p_3 \cup p_4) \cap (p_6 \cup p_1)$ be collinear (**Pascal's theorem**). The dual of the last theorem is called **Brianchon's theorem**.

Given two hypersurfaces Q_2^{n-1}, $\bar{Q}_2^{n-1}$ in P^n, we consider another $\bar{\bar{Q}}_2^{n-1}$ such that the polar of an arbitrary point x with respect to $\bar{\bar{Q}}_2^{n-1}$ belongs to the pencil of hyperplanes determined by polars of x with respect to Q_2^{n-1} and $\bar{Q}_2^{n-1}$. The set of all such $\bar{\bar{Q}}_2^{n-1}$ is called a **pencil of quadric hypersurfaces**. It is the set of all $\bar{\bar{Q}}_2^{n-1}$ that pass through the intersection of Q_2^{n-1} and $\bar{Q}_2^{n-1}$. In the cases $n=2$ and 3, we call it a **pencil of conics** and a **pencil of quadrics**, respectively.

Denoting by l^{ij} and $\bar{l}^{ij}$ $(0 \leqslant i \leqslant j \leqslant 3)$ the †Plücker coordinates of two straight lines l and $\bar{l}$ in P^3, we put

$$(l,\bar{l}) = l^{01}\bar{l}^{23} - l^{02}\bar{l}^{13} + l^{03}\bar{l}^{12} + \bar{l}^{01}l^{23} - \bar{l}^{02}l^{13} + \bar{l}^{03}l^{12}. \tag{4}$$

Then $(l,l)=0$ holds. If we regard these l^{ij} as homogeneous coordinates of P^5, then there exists a one-to-one correspondence between the points on the regular quadric hypersurface Q_2^4 defined by $(l,l)=0$ and the lines in P^3. We say that each point of Q_2^4 is the **image** of the line corresponding to it in P^3. Two lines l and $\bar{l}$ intersect if and only if $(l,\bar{l})=0$. Geometrically, this means that the images of l and $\bar{l}$ are conjugate with respect to Q_2^4. Therefore the line passing through the images of l and $\bar{l}$ is a generating line of Q_2^4. The image of a pencil of lines in P^3 is a generating line of Q_2^4. Quadric hypersurfaces and sets of lines in P^3 are important objects of study in both projective and algebraic geometry. In particular, **linear line congruences (linear line complexes)** that are families of lines dependent upon two (three) parameters are of great interest. In these theories, quadric hypersurfaces play a fundamental role. When the coefficient field is noncommutative, the above theory has to be greatly modified.

F. Projective Geometry and the Erlangen Program

From the standpoint of the †Erlangen program of F. Klein, the aim of projective geometry is to study properties that are invariant under the group of projective transformations. Utilizing various subgroups of this group, we can reconstruct various classical geometries. For example, consider the projective space P^n whose coefficient field is the real number field, and fix a hyperplane Π_∞. Let $\mathfrak{S}(P^n)$ be the subgroup of projective transformations formed by all projective transformations that leave Π_∞ invariant. Then the geometry that belongs to this group is †affine geometry. Similarly, fix an imaginary regular quadric hypersurface Q_2^{n-2} in Π_∞ and consider the geometry that belongs to the subgroup of $\mathfrak{S}(P^n)$ leaving this Q_2^{n-2} invariant. We thus obtain Euclidean geometry. Moreover, if we assign some regular quadric hypersurface Q_2^{n-1}, then the geometry belonging to the subgroup of $\mathfrak{G}(P^n)$ that leaves the Q_2^{n-1} invariant is a †non-Euclidean or †conformal geometry according as the transformation space is the set of inner points of Q_2^{n-1} or the whole Q_2^{n-1}.

G. Projective Geometry and Modular Lattices

†Lattices (lattice-ordered sets) and projective geometry are intimately related. The totality of subspaces of each dimension in general projective geometry $\mathfrak{P}$ constitutes a †complete †modular lattice $L(\mathfrak{P})$ with respect to the inclusion relation. If $\mathfrak{P}$ is a finite-dimensional projective geometry, then it is an †irreducible complemented modular lattice of finite †height. Conversely, suppose that L is a modular lattice with †minimum element Φ, and denote by P the totality of elements p †prime over Φ (i.e., †atomic elements) and by Q the totality of elements l prime over atomic elements. Then, if $p<l$ and $(p,l) \in \Gamma$, $\mathfrak{P}(L)=\{P,Q,\Gamma\}$ is a general projective geometry. If L is an irreducible complemented modular lattice of finite height, then $\mathfrak{P}(L)$ is a finite-dimensional projective geometry; in this case we have $\mathfrak{P} \approx \mathfrak{P}(L(\mathfrak{P}))$ and $L \approx L(\mathfrak{P}(L))$. So we may consider projective geometry and irreducible complemented modular lattices as having the same mathematical structure. If a lattice L is an n-dimensional projective geometry, its †dual lattice is also an n-dimensional projective geometry, and this is the principle of duality.

H. Analytic Representations of Projective Geometry

Let K be an arbitrary field, commutative or noncommutative. For an arbitrary natural number n, we consider an $(n+1)$-dimensional (for the noncommutative case, right or left) linear space $V^{n+1}(K)$ over K. The totality of linear subspaces in it constitutes an irreducible complemented modular lattice $P^n(K)$ with

respect to the inclusion relation, and $P^n(K)$ gives rise to an n-dimensional projective geometry. We call it a **right** or **left projective space**. Points of $P^n(K)$ correspond to (right or left) 1-dimensional linear subspaces. Conversely, it can be shown that an n-dimensional projective geometry over K is isomorphic to $P^n(K)$. Therefore projective geometries can be completely classified by means of the natural number n and the field K except when $n=2$ and the geometry is non-Desarguesian. We may restate this fact as follows: We consider a space $\tilde{P}=V^{n+1}(K)-\{0\}$. If we fix a basis of $V^{n+1}(K)$, then we can represent $\tilde{P}=\{x=(x^0, x^1, \dots, x^n) \mid x^\alpha \in K, 0 \leqslant \alpha \leqslant n\}$, $x \neq (0,0,\dots,0)$. If there exists a nonzero element λ of K such that $y=x\lambda$, then the elements x and y are called equivalent; we write $x \sim y$. We denote by $\mathbf{P}^n(K)$ the factor set of $\tilde{P}$ under the foregoing equivalence relation, and by $[x]$ the equivalence class that contains x. We put $l([x],[y])=\{[z] \mid z^\alpha = x^\alpha\lambda + y^\alpha\mu, \forall\lambda,\mu \in K\}$ and $Q=\{l([x],[y]) \mid [x], [y] \in \mathbf{P}^n(K)\}$. We call each element of $\mathbf{P}^n(K)$ a point and each element of Q a line. Then these points and lines and the natural inclusion relation satisfy axioms (I)–(IV) and give an n-dimensional projective geometry. When K is a †topological field (e.g., the real number field, the complex number field, or the †quaternion field), we may define the topology of $\mathbf{P}^n(K)$ as the factor space $\mathbf{P}^n(K)=\tilde{P}/\sim$. In particular, if K is the real number field $\mathbf{R}$, then $\mathbf{P}^n(\mathbf{R})$ is homeomorphic to the factor space obtained from the n-dimensional hypersphere $S^n:(x^0)^2+\dots+(x^n)^2=1$ in the $(n+1)$-dimensional Euclidean space E^{n+1} by identifying the end points of each diameter. Hence $\mathbf{P}^n(\mathbf{R})$ is compact. Similar facts hold for the cases of the complex and quaternion number fields. Since the group of projective transformations $\mathfrak{G}(\mathbf{P}^n(K))$ acts †transitively on $\mathbf{P}^n(K)$, if K is a topological field we can regard $\mathbf{P}^n(K)$ as a †homogeneous space of the topological group $\mathfrak{G}(\mathbf{P}^n(K))$. Moreover, the totality of r-dimensional subspaces in $\mathbf{P}^n(K)$ constitutes a †Grassmann manifold. In algebraic geometry the †direct product of two projective spaces is important; we call it a **biprojective space**.

I. Tits's Theory of Buildings (Generalization of Projective Geometry)

In a situation when a triple (G,B,N) consisting of a group G and its subgroups B, N satisfies the axioms of a **BN-pair** or **Tits system** (→ 13 Algebraic Groups R), a new geometric object, called a "building," was introduced by J. Tits [9]. His theory contains projective geometry as a particular case. The theory of buildings has deep connection with algebraic groups. The Tits system corresponds to a projective geometry in the following case. Let k be any commutative field, and let G be the general linear group of degree n over k, i.e., G consists of all nonsingular square matrices of degree n with entries in k. Let B be the subgroup of G consisting of all upper triangular matrices (i.e., matrices whose entries below the principal diagonal are all zero). Let N be the subgroup of G consisting of all monomial matrices (i.e., matrices such that each column and each row contain just one nonzero entry). Then (G,B,N) forms a Tits system called type (A_{n-1}). The corresponding **theory of buildings** of the type above is nothing but the projective geometry. Thus by means of Tits's theory of buildings the relationships among projective geometry and other geometries have been clarified [9].

References

[1] G. Birkhoff, Lattice theory, Amer. Math. Soc. Colloq. Publ., revised edition, 1967.
[2] W. V. D. Hodge and D. Pedoe, Methods of algebraic geometry I, Cambridge, 1947.
[3] O. Schreier and E. Sperner, Einführung in die analytische Geometrie und Algebra, Vandenhoeck & Ruprecht, II, 1951; English translation, Projective geometry of n-dimensions, Chelsea, 1961.
[4] O. Veblen and J. W. Young, Projective geometry I, II, Ginn, 1910–1938.
[5] E. Artin, Geometric algebra, Interscience, 1957.
[6] S. Iyanaga and K. Matsuzaka, Affine geometry and projective geometry, J. Fac. Sci. Univ. Tokyo, 14 (1967), 171–196.
[7] A. Seidenberg, Lectures in projective geometry, Van Nostrand, 1962.
[8] R. Hartshorne, Foundations of projective geometry, Benjamin, 1967.
[9] J. Tits, Buildings of spherical type and finite BN-pairs, Lecture notes in math. 386, Springer, 1974.

344 (VII.22) Pseudoconformal Geometry

A. Definitions

Let A and A' be subsets (with relative topology) of †complex manifolds X and X' of dimension n, respectively. A homeomorphism f of A onto A' is called a **pseudoconformal transformation** if there exists a †biholomorphic mapping $\tilde{f}$ of an open neighborhood of A in X onto an open neighborhood of A' in X' such

that $\tilde{f}(x)=f(x)$ for $x\in A$. If there exists such a mapping f, A is said to be **pseudoconformally equivalent** to A'. **Pseudoconformal geometry** is a geometry that studies geometric properties invariant under the pseudoconformal equivalence. However, most studies in pseudoconformal geometry so far have concentrated mainly on the investigation of smooth hypersurfaces in a complex manifold—more specifically, the smooth (or real analytic) boundaries of bounded domains in $\mathbf{C}^n$. In fact, to pseudoconformal geometry on hypersurfaces we can apply the methods of differential geometry as well as those of the theory of functions of several complex variables.

H. Poincaré [1] studied perturbations of the boundary of the unit ball in $\mathbf{C}^2$ that are pseudoconformally equivalent. E. Cartan [2] studied the equivalence problem of hypersurfaces in $\mathbf{C}^2$ and gave the complete list of all simply connected hypersurfaces on which the group of pseudoconformal automorphisms acts transitively. Such a hypersurface is called **homogeneous**.

Let M be a smooth hypersurface in a complex manifold X with the †almost complex structure tensor J, i.e., $J_x: T_xX\rightarrow T_xX$ is an involutive linear automorphism of the tangent space T_xX of X at x induced by the complex structure of X. Put $H_xM=T_xM\cap J_xT_xM$ for $x\in M$. The union of all H_xM is called the bundle of holomorphic tangent vectors of M and is denoted by $H(M)$. $H(M)$ is also called the **CR** (Cauchy-Riemann) **structure** of M. Let M' be a smooth hypersurface in a complex manifold X'. A diffeomorphism $f: M\rightarrow M'$ is called a **CR-equivalence** if the †differential mapping $Tf: TM\rightarrow TM'$ of f preserves the CR-structures, where TM denotes the †tangent bundle of M. If $f: M\rightarrow M'$ is a pseudoconformal transformation, then f is clearly a CR-equivalence. Let E_x be the annihilator of H_xM in $T_x^*(M)$. Then the union of E_x $(x\in M)$ defines a †line bundle E over M. The **Levi form** L_x at $x\in M$, defined only up to a multiplier, is the quadratic form on H_xM defined by $L_x(u,v)=d\theta(u,v)$ for $u, v\in H_xM$, where θ is a nonvanishing section of E in a neighborhood of x. If the Levi form is nondegenerate at every point of M, M is called a **nondegenerate hypersurface**. In particular, if the Levi form is definite, then M is called **strictly pseudoconvex**.

B. Equivalence Problem

Cartan studied the equivalence problem for the case $n=2$, and obtained a criterion for two hypersurfaces in $\mathbf{C}^2$ to be pseudoconformally equivalent. N. Tanaka (1965) generalized the method of Cartan for the case $n\geqslant 3$ and obtained a criterion in terms of †Cartan connections in some fiber bundle over the hypersurfaces. However, he did not publish the proof of his result until S. S. Chern and J. Moser [4], independently of Tanaka [3], obtained a similar result and gave the first proof of this result. Let M be a real analytic hypersurface in $\mathbf{C}^{n+1}$ $(n\geqslant 1)$ whose Levi form has p positive and q negative eigenvalues $(p+q=n)$. Let H be the subgroup of $SU(p+1,q+1)$ leaving the point $(1,0,\dots,0)\in\mathbf{C}^{n+2}$ fixed. According to the Cartan-Tanaka-Chern-Moser result, we can construct functorially a principal fiber bundle Y over M with structure group H and a Cartan connection ω on Y with values in the Lie algebra of $SU(p+1,q+1)$ such that if M and M' are pseudoconformally equivalent, then there is a bundle isomorphism φ of Y to Y' preserving the Cartan connections: $\varphi^*\omega'=\omega$, where Y' is the corresponding principal fiber bundle over M' and ω' is the Cartan connection on Y'. Conversely, if there is a bundle isomorphism φ of Y to Y' such that $\varphi^*\omega'=\omega$, then M and M' are pseudoconformally equivalent. By using this solution of the equivalence problem, we can prove that the group $A(M)$ of all pseudoconformal automorphisms of a nondegenerate real analytic hypersurface M in a complex manifold X of dimension n is a Lie transformation group of dimension not exceeding n^2+2n. H. Jacobowitz [5] constructed a similar bundle B over M and a Cartan connection on B in a different way from that of Chern and Moser. We do not know whether B and Y actually coincide.

C. Classification

Cartan (1932) classified all simply connected homogeneous hypersurfaces in $\mathbf{C}^2$. In particular, he proved that if M is a compact homogeneous strictly pseudoconvex hypersurface with $\dim M=3$, then M is pseudoconformally equivalent to either (1) S^3 or its quotient by the action of a root of unity or (2) the hypersurface given in the 2-dimensional projective space by the equation in homogeneous coordinates: $(z_1\bar{z}_1+z_2\bar{z}_2+z_3\bar{z}_3)^2=m^2|z_1^2+z_2^2+z_3^2|^2$ $(m>1)$ or the double covering of such a surface. A. Morimoto and T. Nagano [6] and later H. Rossi [7] tried to generalize this result and obtained a partial classification of simply connected compact homogeneous hypersurfaces with dimension $\geqslant 5$. D. Burns and S. Shnider [8] classified all simply connected compact homogeneous strictly pseudoconvex hypersurfaces M with $\dim M=2n+1\geqslant 5$. They proved that M is pseudoconformally equivalent to S^{2n+1} or the tangent sphere bundle of a rank one †symmetric space

or the unit circle bundle of a homogeneous negative line bundle over a homogeneous algebraic manifold.

On the other hand, a real hypersurface M in a complex manifold X of complex dimension $n+1$ is called **spherical** if at every point $p \in M$, there is a neighborhood of p in X such that $U \cap M$ is pseudoconformally equivalent to an open submanifold of S^{2n+1}. The hyperquadric $Q^{n+1} = \partial U_{n+1}$ is spherical, where $U_{n+1} = \{(z_1, \ldots, z_{n+1}) \in \mathbf{C}^{n+1} \mid \operatorname{Im}(z_{n+1}) > |z_1|^2 + \ldots + |z_n|^2\}$. If M is spherical, then the universal covering space $\tilde{M}$ of M is also spherical. If M is a homogeneous spherical hypersurface, then there is a covering into mapping $f: \tilde{M} \to S^{2n+1}$, and $f(\tilde{M})$ is a homogeneous domain in S^{2n+1}. We know that the only compact simply connected spherical M is S^{2n+1}. Burns and Shnider classified all homogeneous domains M in S^{2n+1}: M is pseudoconformally equivalent to (I) or (II) of the following: (Ia) $S^{2n+1} - V \cap S^{2n+1}$, where V is a complex vector subspace of $\mathbf{C}^{n+1}$ with $0 \leqslant \dim_{\mathbf{C}} V \leqslant n$. (Ib) $Q^{n+1} - L_m \cdot 0$, $0 \leqslant m \leqslant n$, where L_m is a certain subgroup of $SU(n+1, 1)/(\text{center})$. (II) $S^{2n+1} - S^{2n+1} \cap \mathbf{R}^{n+1}$. At present, it seems difficult to extend Cartan's classification of all simply connected homogeneous hypersurfaces to higher dimensions. K. Yamaguchi (1976) treated a hypersurface M in a complex manifold of dimension n with a large automorphism group $A(M)$. He showed that if $\dim A(M) = n^2 + 2n$, then M is a real hyperquadric in the n-dimensional complex projective space $P_n\mathbf{C}$ (→ Section B). He then showed that the second largest dimension for $A(M)$ is equal to $n^2 + 1$ except when $n = 3$ and the index $r = 1$, for which $\dim A(M) = 11 (= n^2 + 1)$. Under the additional assumption that M is homogeneous, he showed that if $\dim A(M) = n^2 + 1$, then M is the affine part of a real hyperquadric in $P_n\mathbf{C}$ (except when $n = 5$ and $r = 2$). He also obtained a similar result in the nonhomogeneous case.

D. Relations to Other Equivalences

Let D_1 and D_2 be bounded domains in $\mathbf{C}^n$ with smooth boundary $\partial D_i = M_i$ $(i = 1, 2)$ for which we denote by $H(M_i)$ the CR structures. We consider the following propositions (A)–(D) for these domains: (A) D_1 is biholomorphically equivalent to D_2. (B) M_1 is CR equivalent to M_2. (C) M_1 is pseudoconformally equivalent to M_2. (D) There is a diffeomorphism $f: \bar{D}_1 \to \bar{D}_2$ such that $f|_{D_1}: D_1 \to D_2$ is biholomorphic.

It is clear that (C) implies (D) and that (D) implies (A). On the other hand, we can prove that (B) is equivalent to (D). When does (A) imply (B) and when does (B) imply (C)?

C. Fefferman [10] proved that (A) implies (D) when D_1 and D_2 are strictly pseudoconvex. S. Bell generalized the result of Fefferman in the case when one of D_1 and D_2 is strictly pseudoconvex. Bell and E. Ligocka [11] proved that if M_1 and M_2 are real analytic and if D_1 and D_2 are pseudoconvex, then (A) implies (D). When D_1 and D_2 are not strictly pseudoconvex and M_i is not real analytic, we do not know whether (A) implies (D) or not. As remarked by Burns, Shnider, and Wells (1978), by using the theorem of Fefferman, we can prove that (A) implies (C) when M_1 and M_2 are real analytic and if D_1 and D_2 are strictly pseudoconvex. I. Naruki [12] obtained the same result. We do not know whether (A) implies (C) when M_1 and M_2 are real analytic and D_1 and D_2 are pseudoconvex, though we know that (A) implies (B). We do not know whether (B) implies (C) in general.

S. I. Pinchuk [13] proved the following: Let D, D' be strictly pseudoconvex domains in $\mathbf{C}^n$ with simply connected real analytic boundaries ∂D, $\partial D'$. Let $f: U \to \mathbf{C}^n$ be a nonconstant holomorphic mapping from a connected neighborhood U of a point $p \in \partial D$ in $\mathbf{C}^n$ into $\mathbf{C}^n$ such that $f(U \cap \partial D) \subset \partial D'$. Then we can find a holomorphic mapping $\tilde{f}: D \to D'$ such that $\tilde{f}(x) = f(x)$ for $x \in D \cap U$. Combining this theorem with Fefferman's result we see that for two domains as above, D is biholomorphically equivalent to D' if and only if ∂D is locally pseudoconformally equivalent to $\partial D'$, i.e, there are neighborhoods U and V of a point $p \in \partial D$ and $q \in \partial D'$, respectively, such that $U \cap \partial D$ and $V \cap \partial D'$ are pseudoconformally equivalent.

Concerning the †proper holomorphic mappings rather than diffeomorphisms, Burns and Shnider (1979) proved the following theorem: Let $M_i = \partial D_i$ $(i = 1, 2)$ be strictly pseudoconvex, and let $f: D_1 \to D_2$ be a proper holomorphic mapping. (a) If $D_1 = D_2$, then f extends smoothly up to the boundary D_1. (b) If ∂D_i is real analytic for $i = 1, 2$, then f extends holomorphically past the boundary.

E. Deformations of Domains

Let M be a compact connected strictly pseudoconvex real hypersurface in a complex manifold X of dimension $n+1$. Let φ be a smooth strictly †plurisubharmonic function defined on a neighborhood V of M such that $M = \{x \in V \mid \varphi(x) = 0\}$ and $d\varphi \neq 0$ on M. Let $U = \{x \in V \mid -\varepsilon < \varphi(x) < \varepsilon\}$ for small $\varepsilon > 0$ such that $\bar{U}$ is compact and ∂U is smooth. Let $\mathscr{P}(\bar{U})$ be the open set in $\mathbf{C}^\infty(\bar{U})$ of strictly plurisubharmonic functions ψ with $d\psi \wedge \bar{d}\psi \wedge (d\bar{d}\psi)^n \neq 0$ on $\bar{U}$. Let $B \subset \mathbf{R}^k$ be a small open ball around 0. We denote by $\mathscr{P}(\bar{U} \times \bar{B})$ the set of $\psi \in \mathbf{C}^\infty(\bar{U}$

$\times \bar{B}$) such that $\psi(x,t)=\psi_t(x)\in\mathscr{P}(\bar{U})$ for all $t\in\bar{B}$. For $\psi\in\mathscr{P}(\bar{U}\times\bar{B})$ we set $M_{t,\delta}=\{x\in U\mid\psi_t(x)=\delta\}$. After introducing these notations, Burns, Shnider, and Wells (1978) proved the following theorem. There exists an open dense set $\mathscr{V}\subset\mathscr{P}(\bar{U}\times\bar{B})$ with $\varphi\in\mathscr{V}$ and a set of †second category $\mathscr{R}\subset\mathscr{V}$ such that for every $\psi\in\mathscr{R}$, $t_i\in\bar{B}$ and $\delta_i\in\mathbf{R}$ small enough, (i) M_{t_1,δ_1} is CR-equivalent to M_{t_2,δ_2} if and only if $t_1=t_2$, $\delta_1=\delta_2$. (ii) The group of CR-automorphisms of M_{t_1,δ_1} reduces to the identity only. For $\psi\in\mathscr{P}(\bar{U}\times\bar{B})$, taking $t\in B$, $\delta\in\mathbf{R}$ small enough, $M_{t,\delta}$ is a compact connected strictly pseudoconvex hypersurface in X. If M bounds the relatively compact region D in X then $M_{t,\delta}$ also bounds a relatively compact region $D_{t,\delta}$. In particular, there exist smooth families of deformations of the unit ball in $\mathbf{C}^{n+1}$ of arbitrary high dimension. There are arbitrary small perturbations of the unit sphere in $\mathbf{C}^{n+1}$ that admit no pseudoconformal transformations other than the identity.

F. Topics Related to Pseudoconformal Geometry

(1) Pinchuk (1975) proved the following: Let D_1, D_2 be strictly pseudoconvex domains in $\mathbf{C}^n$ with C^ω boundary ∂D_1, ∂D_2. Let U be a neighborhood of a point $p\in\partial D_1$ in $\mathbf{C}^n$. If there is a C^1-mapping $f: U\cap\bar{D}_1\to C^n$ such that f is holomorphic on $U\cap D_1$ and $f(U\cap\partial D_1)\subset\partial D_2$, then there is a holomorphic mapping $\tilde{f}: U'\to\mathbf{C}^n$ of a neighborhood U' of $U\cap\partial D_1$ into C^n such that $\tilde{f}(x)=f(x)$ for $x\in U\cap\bar{D}_1\cap U'$. This result is related to the implication (B)$\Rightarrow$(C) in Section D.

(2) H. Alexander [14] proved the following: Let U be a connected neighborhood of a point $p\in S^{2n-1}$ in $\mathbf{C}^n$ and $f: U\to C^n$ a holomorphic mapping such that $f(U\cap S^{2n-1})\subset S^{2n-1}$. Then either f is a constant mapping or there is a biholomorphic automorphism $\tilde{f}: B_n\to B_n$ of the unit open ball B_n such that $\tilde{f}(x)=f(x)$ for $x\in U\cap B_n$. He also proved that every proper holomorphic mapping $f: B_n\to B_n$ is necessarily an automorphism of B_n if $n>1$. G. M. Henkin (1973) proved that every proper holomorphic mapping $f: D_1\to D_2$ of a strictly pseudoconvex domain D_1 into a strictly pseudoconvex D_2 can be extended continuously to a function $\tilde{f}: \bar{D}_1\to\bar{D}_2$. More precisely, there is a constant $c>0$ such that $|f(z_1)-f(z_2)|\leqslant c|z_1-z_2|^{1/2}$ for every $z_1, z_2\in D_1$.

(3) Let M be a real hypersurface in $\mathbf{C}^{n+1}$ with $H(M)$ the bundle of holomorphic tangent vectors to M. We take a real nonvanishing 1-form θ that annihilates $H(M)$. S. M. Webster (1978) called the pair (M,θ) a **pseudo-Hermitian manifold**. He considered the equivalence problem of pseudo-Hermitian manifolds by applying Cartan's method of equivalence. He proves, among other things, that the group of all pseudo-Hermitian transformations of the nondegenerate pseudo-Hermitian manifold (M,θ) of dimension $2n+1$ is a Lie transformation group of dimension not exceeding $(n+1)^2$. Webster considered the relation between pseudo-Hermitian manifolds and pseudoconformal geometry and proved that for $n\geqslant 2$ the ellipsoid E given by the equation $A_1x_1^2+B_1y_1^2+\ldots+A_{n+1}x_{n+1}^2+B_{n+1}y_{n+1}^2=1$, where $z_k=x_k+iy_k$ $(k=1,\ldots,n+1)$ is pseudoconformally equivalent to the hypersphere S^{2n+1} if and only if $A_k=B_k$ $(k=1,\ldots,n+1)$. This result gives, by virtue of Fefferman's theorem, a necessary and sufficient condition for an ellipsoidal domain to be biholomorphically equivalent to the unit ball.

References

[1] H. Poincaré, Les fonctions analytiques de deux variables et la représentation conforme, Rend. Circ. Mat. Palermo, 23 (1907), 185–220.
[2] E. Cartan, Sur la géométrie pseudo-conforme des hypersurfaces de l'espace de deux variables complexes I, II, Ann. Mat. Pura Appl., 11 (1932), 17–90; Ann. Scuola Norm. Sup. Pisa, 1 (1932), 333–354.
[3] N. Tanaka, On nondegenerate real hypersurfaces, graded Lie algebras and Cartan connections, Japan. J. Math., 2 (1976), 131–190.
[4] S. S. Chern and J. Moser, Real hypersurfaces in complex manifolds, Acta Math., 133 (1974), 219–271.
[5] H. Jacobowitz, Induced connections on hypersurfaces in $\mathbf{C}^n$, Inventiones Math., 43 (1977), 109–123.
[6] A. Morimoto and T. Nagano, On pseudo-conformal transformations of hypersurfaces, J. Math. Soc. Japan, 14 (1963), 289–300.
[7] H. Rossi, Homogeneous strongly pseudo-convex hypersurfaces, Rice Univ. Studies, 59 (1973), 131–145.
[8] D. Burns and S. Shnider, Spherical hypersurfaces in complex manifolds, Inventiones Math., 33 (1976), 223–246.
[9] K. Yamaguchi, Non-degenerate hypersurfaces in complex manifolds admitting large groups of pseudoconformal transformations I, Nagoya Math. J., 62 (1976), 55–96.
[10] C. Fefferman, The Bergman kernel and biholomorphic mappings of pseudoconvex domains, Inventiones Math., 26 (1974), 1–65.
[11] S. Bell and E. Ligocka, A simplification and extension of Fefferman's theorem on biholomorphic mappings, Inventiones Math., 57 (1980), 283–289.

[12] I. Naruki, On extendability of isomorphisms of Cartan connections and biholomorphic mappings of bounded domains, Tôhoku Math. J., 28 (1976), 117–122.
[13] S. I. Pinchuk, On holomorphic mappings of real analytic hypersurfaces, Math. USSR-Sb., 34 (1978), 503–519.
[14] H. Alexander, Proper holomorphic mappings in $\mathbf{C}^n$, Indiana Math. J., 26 (1977), 137–146.
[15] S. M. Webster, Pseudo-Hermitian structure on a real hypersurface, J. Differential Geometry, 13 (1978), 25–41.

345 (XIII.33) Pseudodifferential Operators

A. Pseudodifferential Operators

Pseudodifferential operators are a natural extension of linear partial differential operators. The theory of pseudodifferential operators grew out of the study of singular integral operators, and developed rapidly after 1965 with the systematic studies by J. J. Kohn and L. Nirenberg [1], L. Hörmander [2], and others. The term "pseudodifferential operator" first appeared in Kohn and Nirenberg [1].

Let P be a †linear partial differential operator of the form

$$P=p(x,D_x)=\sum_{|\alpha|\leqslant m} a_\alpha(x)D_x^\alpha, \qquad (1)$$

and let $u(x)$ be a function of class $C_0^\infty(\Omega)$ $(\subset C_0^\infty(\mathbf{R}^n))$. Then by means of the †Fourier inversion formula, $Pu(x)$ can be written in the form

$$Pu(x)=(2\pi)^{-n/2}\int_{\mathbf{R}^n}\exp(ix\cdot\xi)p(x,\xi)\hat{u}(\xi)\,d\xi, \qquad (2)$$

where $\hat{u}(\xi)$ denotes the †Fourier transform of $u(x)$ (→ 160 Fourier Transform H). But this representation of $Pu(x)$ has a meaning even if $p(x,\xi)$ is not a polynomial in ξ. Thus, for a general function $p(x,\xi)$, the **pseudodifferential operator** $P=p(x,D_x)$ with the **symbol** $p(x,\xi)$ is defined by (2). A symbol class is determined in accordance with various purposes, but it is always required that the corresponding operators have essential properties in common with partial differential operators. Hörmander [3] defined a symbol class $S^m_{\rho,\delta}(\Omega)$ for real numbers m, ρ, and δ with $\rho\geqslant 0$ and $\delta\geqslant 0$ in the following way: Let $p(x,\xi)$ be a C^∞-function defined in $\Omega\times\mathbf{R}^n$. If for any pair of multi-indices α, β and any compact set $K\subset\mathbf{R}^n$, there exists a constant $C_{\alpha,\beta,K}$ such that

$$|D_x^\alpha D_\xi^\beta p(x,\xi)|\leqslant C_{\alpha,\beta,K}(1+|\xi|)^{m+\delta|\alpha|-\rho|\beta|}, \qquad x\in K, \quad \xi\in\mathbf{R}^n,$$

then $p(x,\xi)$ is said to be of class $S^m_{\rho,\delta}(\Omega)$. The operator P defined by (2) is called a pseudodifferential operator (of order m) of class $S^m_{\rho,\delta}(\Omega)$ and is often denoted by $P=p(x,D_x)\in S^m_{\rho,\delta}(\Omega)$. When $\Omega=\mathbf{R}^n$ and constants $C_{\alpha,\beta,K}=C_{\alpha,\beta}$ are independent of K, we denote $S^m_{\rho,\delta}(\mathbf{R}^n)$ simply by $S^m_{\rho,\delta}$, and set

$$S^{-\infty}=\bigcap_{-\infty<m<\infty}S^m_{1,0}\left(=\bigcap_{-\infty<m<\infty}S^m_{\rho,\delta}\right),$$

$$S_{\rho,\delta}=\bigcup_{-\infty<m<\infty}S^m_{\rho,\delta}.$$

Differential operators (1) with coefficients of class $\mathscr{B}$ (→ 168 Function Spaces B(13)) belong to $S^m_{1,0}$. The complex power $(1-\Delta)^{z/2}$ of $1-\Delta=1-\sum_{j=1}^n\partial^2/\partial x_j^2$ is defined as a pseudodifferential operator of class $S^{\operatorname{Re}z}_{1,0}$ by the symbol $(1+|\xi|^2)^{z/2}$. Operators of class $S^m_{\rho,\delta}$ are continuous mappings of $\mathscr{S}$ into $\mathscr{S}$. Therefore, for any real s, the operator $(1-\Delta)^{s/2}$ can be uniquely extended to be a mapping of $\mathscr{S}'$ into $\mathscr{S}'$ by the relation

$$\langle(1-\Delta)^{s/2}u,v\rangle=\langle u,(1-\Delta)^{s/2}v\rangle, \qquad u\in\mathscr{S}', \quad v\in\mathscr{S}.$$

For any $1\leqslant r\leqslant\infty$ and real s, the †Sobolev space $H^{s,r}$ is defined by

$$H^{s,r}=\{u\in\mathscr{S}'\,|\,(1-\Delta)^{s/2}u\in L_r(\mathbf{R}^n)\},$$

which is a Banach space provided with the norm $\|u\|_{s,r}=\|(1-\Delta)^{s/2}u\|_{L_r}$. In particular, $H^s=H^{s,2}$ is a Hilbert space with the norm $\|u\|_s=\|u\|_{s,2}$. Set

$$H^{-\infty,r}=\bigcup_{-\infty<s<\infty}H^{s,r}, \quad H^{-\infty}=H^{-\infty,2},$$

$$H^{\infty,r}=\bigcap_{-\infty<s<\infty}H^{s,r}, \quad H^{\infty}=H^{\infty,2}.$$

Then

$$\mathscr{S}'\supset H^{-\infty}\supset\mathscr{E}', \quad H^{-\infty}\supset L_2(\mathbf{R}^n)\supset H^\infty(\subset\mathscr{B}).$$

Choosing the Hörmander class $S^m_{\rho,\delta}$ in the case $0\leqslant\delta\leqslant\rho\leqslant 1$ and $\delta<1$ as a model class, we here list the main results of the theory of pseudodifferential operators:

(i) Pseudolocal property. The operator P of class $S^m_{\rho,\delta}$ in general does not have the **local property** $u\in\mathscr{S}'\Rightarrow\operatorname{supp}Pu\subset\operatorname{supp}u$, but if $\rho>0$, then P has the **pseudolocal property** $u\in\mathscr{S}'\Rightarrow\operatorname{sing\,supp}Pu\subset\operatorname{sing\,supp}u$ [3].

(ii) Algebra of pseudodifferential operators. Let $P=p(x,D_x)\in S^m_{\rho,\delta}$ and $P_j=p_j(x,D_x)\in S^{m_j}_{\rho,\delta}$, $j=1,2$. Then there exist $P^*=p^*(x,D_x)\in S^m_{\rho,\delta}$ and $Q=q(x,D_x)\in S^{m_1+m_2}_{\rho,\delta}$ such that $(Pu,v)=(u,P^*v)$ for $u,v\in\mathscr{S}$, i.e., P^* is the formal ad-

joint of P, and $Q=P_1P_2$. Furthermore, if we set $p_\alpha^*(x,\xi)=D_x^\alpha(iD_\xi)^\alpha\overline{p(x,\xi)}$ and $q_\alpha(x,\xi)=(iD_\xi)^\alpha p_1(x,\xi)D_x^\alpha p_2(x,\xi)$, then for any integer N we have

$$p^*(x,\xi)-\sum_{|\alpha|<N}\frac{1}{\alpha!}p_\alpha^*(x,\xi)\in S_{\rho,\delta}^{m-(\rho-\delta)N} \tag{3}$$

and

$$q(x,\xi)-\sum_{|\alpha|<N}\frac{1}{\alpha!}q_\alpha(x,\xi)\in S_{\rho,\delta}^{m_1+m_2-(\rho-\delta)N}. \tag{4}$$

Hence the operator class $S_{\rho,\delta}^m$ is an algebra in the sense

$$P\in S_{\rho,\delta}^m,\quad P_j\in S_{\rho,\delta}^{m_j},\quad j=1,2,$$

$$\Rightarrow\quad P^*\in S_{\rho,\delta}^m,\quad P_1+P_2\in S_{\rho,\delta}^{m_0},\quad P_1P_2\in S_{\rho,\delta}^{m_1+m_2},$$

where $m_0=\max(m_1,m_2)$. In particular, if $0\leqslant\delta<\rho\leqslant 1$, we have $m-(\rho-\delta)N\to-\infty$ and $m_1+m_2-(\rho-\delta)N\to-\infty$ as $N\to\infty$. Then, we say that $p^*(x,\xi)$ and $q(x,\xi)$ have **asymptotic expansions** in the sense of (3) and (4), respectively:

$$p^*(x,\xi)\sim\sum_\alpha\frac{1}{\alpha!}p_\alpha^*(x,\xi)$$

and

$$q(x,\xi)\sim\sum_\alpha\frac{1}{\alpha!}q_\alpha(x,\xi)$$

[3,4].

(iii) H^s-boundedness. For $P\in S_{\rho,\delta}^m$ and any real s there exists a constant C_s such that

$$\|Pu\|_s\leqslant C_s\|u\|_{s+m},\qquad u\in H^{s+m} \tag{5}$$

[4].

(iv) A sharp form of Gårding's inequality. Let $p(x,\xi)=(p_{jk}(x,\xi);j,k=1,\dots,l)$ be a Hermitian symmetric and nonnegative matrix of $p_{j,k}(x,\xi)\in S_{\rho,\delta}^m$. Then for $P=p(x,D_x)$ there exists a constant C such that

$$\mathrm{Re}(Pu,u)\geqslant -C\|u\|_{(m-(\rho-\delta))/2}^2, \tag{6}$$

where $u=(u_1,\dots,u_l)$ with $u_j\in H^{m/2}$, $j=1,\dots,l$, and $\|u\|_{m/2}^2=\sum_{j=1}^l\|u_j\|_{m/2}^2$.

(v) Invariance under coordinate transformations. Assume that $0\leqslant 1-\rho\leqslant\delta\leqslant\rho\leqslant 1$. Let $x(y)=(x_1(y),\dots,x_n(y))$ be a C^∞-coordinate transformation from $\mathbf{R}_y^n$ onto $\mathbf{R}_x^n$ such that $\partial x_k(y)/\partial y_j\in\mathscr{B}$, $j,k=1,\dots,n$, and $C^{-1}\leqslant|\det(\partial_y x(y))|\leqslant C$ for a constant $C>0$, where $\det(\partial_y x(y))$ denotes the determinant of the Jacobian matrix $(\partial_y x(y))=(\partial x_k(y)/\partial y_j)$. Then for any $P=p(x,D_x)\in S_{\rho,\delta}^m$ in $\mathbf{R}_x^n$, there exists an operator $Q=q(x,D_x)\in S_{\rho,\delta}^m$ in $\mathbf{R}_y^n$ such that $(Qw)(y)=(Pu)(x(y))$ for $w(y)=u(x(y))\in\mathscr{S}$. This fact enables us to define pseudodifferential operators on C^∞-manifolds [3,4].

(vi) Parametrix. For a given $P\in S_{\rho,\delta}^m$, an operator $E\in S_{\rho,\delta}^\infty$ is called a **left** (resp. **right**) **parametrix** of P if $EP-I$ (resp. $PE-I$) is of class $S^{-\infty}$. If E is a left and right parametrix of P, we call it a **parametrix** of P. For a differential operator P, the existence of a left (resp. right) parametrix is a sufficient condition for P to be †hypoelliptic if $\rho>0$, (resp. the equation $Pu=f\in\mathscr{D}'$ is locally solvable).

The estimate (5), in particular, when $m=s=0$, has been obtained by Hörmander [3], V. V. Grushin (*Functional Anal. Appl.*, 4 (1970)), H. Kumano-go (*J. Fac. Sci. Univ. Tokyo*, 17 (1970)) when $0\leqslant\delta<\rho\leqslant 1$, and A. P. Calderón and R. Vaillancourt [5], H. O. Cordes (*J. Functional Anal.* 18 (1975)), T. Kato (*Osaka J. Math.*, 13 (1976)), Kumano-go [4], and others when $0\leqslant\delta\leqslant\rho\leqslant 1$ and $\delta<1$. A sharp form of Gårding's inequality has been proved by Hörmander [6], P. D. Lax and L. Nirenberg [7], and sharpened by A. Melin (*Ark. Mat.*, 9 (1971)), C. Fefferman and D. H. Phong (*Proc. Nat. Acad. Sci. US*, 76 (1979)), and Hörmander [8]. A general sufficient condition for the existence of a parametrix for an operator of class $S_{\rho,\delta}^m$ was obtained by Hörmander [3]. Let $P=p(x,D_x)$ belong to $S_{\rho,\delta}^m$ with $0\leqslant\delta<\rho\leqslant 1$. Assume that the symbol $p(x,\xi)$ satisfies the following conditions: (i) for some $C_0>0$, real m' ($\leqslant m$), and $R>0$, we have $|p(x,\xi)|\geqslant C_0|\xi|^{m'}$ ($|\xi|\geqslant R$); (ii) for any α, β there exists a constant $C_{0,\alpha,\beta}$ such that

$$|D_x^\beta D_\xi^\alpha p(x,\xi)/p(x,\xi)|\leqslant C_{0,\alpha,\beta}|\xi|^{\delta|\beta|-\rho|\alpha|},\qquad |\xi|\geqslant R.$$

Then there exists a parametrix $Q=q(x,D_x)$ of P in the class $S_{\rho,\delta}^{-m'}$.

By means of operators of class $S_{1,0}^m(\Omega)$ we can define the wave front set of $u\in\mathscr{D}'(\Omega)$, which enables us to resolve $\operatorname{sing\,supp} u$ on $T^*(\Omega)\setminus 0$, the cotangent bundle of Ω minus its zero section. An operator $P=p(x,D_x)\in S_{1,0}^m(\Omega)$ is said to be **microlocally elliptic** at $(x^0,\xi^0)\in T^*(\Omega)\setminus 0$ if $\lim_{\tau\to\infty}|p(x^0,\tau\xi^0)|/|\tau\xi^0|^m>0$. For a distribution $u\in\mathscr{D}'(\Omega)$, we say that a point (x^0,ξ^0) of $T^*(\Omega)\setminus 0$ does not belong to the **wave front set** (or the **singular spectrum**) of u, denoted by $\mathrm{WF}(u)$, if there exist $a(x),b(x)\in C_0^\infty(\Omega)$, $a(x^0)\neq 0$, $b(x^0)\neq 0$, and $P\in S_{1,0}^m(\Omega)$, which is microlocally elliptic at (x^0,ξ^0), such that $aPbu\in C_0^\infty(\Omega)$. Then we easily see that $\mathrm{WF}(u)$ is a closed conic subset of $T^*(\Omega)\setminus 0$. An important fact is that the relation $\operatorname{sing\,supp} u=\mathrm{Proj}_x\mathrm{WF}(u)$ (the projection of $\mathrm{WF}(u)$ on Ω) holds, from which we can perform a so-called **microlocal analysis**, the analysis on $T^*(\Omega)\setminus 0$, of $\operatorname{sing\,supp} u$. As the sharp form of the pseudolocal property of an operator $P\in S_{\rho,\delta}^m$, if $0\leqslant\delta<\rho\leqslant 1$, P has the **micro-pseudolocal property**: $u\in\mathscr{S}'\Rightarrow\mathrm{WF}(Pu)\subset\mathrm{WF}(u)$.

Pseudodifferential operators of multiple symbol have been defined by K. O. Friedrichs (*Courant Inst.*, 1968) and Kumano-go [4]. More refined and useful classes of pseudo-

differential operators have been defined by R. Beals (*Duke Math. J.*, 42 (1975)) Hörmander [8], and others.

The theory of pseudodifferential operators has found many fields of application, such as M. F. Atiyah and R. Bott (*Ann. Math.*, 86 (1967)) on the †Lefschetz fixed-point formula; Friedrichs and P. D. Lax (*Comm. Pure Appl.*, 18 (1965)) on symmetrizable systems; Hörmander [6], Yu. V. Egorov (*Russian Math. Surveys*, 30 (1975)) on subelliptic operators; Kumano-go (*Comm. Pure Appl. Math.*, 22 (1969)), F. Treves (*Amer. J. Math.*, 94 (1972)), S. J. Alinhac and M. S. Bouendi (*Amer. J. Math.*, 102 (1980)) on uniqueness of the Cauchy problem; S. Mizohata and Y. Ohya (*Publ. Res. Inst. Math. Sci.*, 4 (1968)), Hörmander (*J. Analyse Math.*, 32 (1977)) on †weakly hyperbolic equations; C. Morawetz, J. V. Ralston, and W. A. Strauss (*Comm. Pure Appl. Math.*, 30 (1977)), M. Ikawa (*Pub. Res. Inst. Math. Sci.*, 14 (1978)) on the exponential decay of solutions; and Nirenberg and Treves [16], Beals and Fefferman [17] on local solvability theory.

For recent developments in the theory of pseudodifferential operators and its applications → Kumano-go [4], M. Taylor (Princeton Univ. Press, 1981), Treves (Plenum, 1981), and others.

B. Fourier Integral Operators

A **Fourier integral operator** A: $C_0^\infty(\mathbf{R}^n)\to\mathscr{D}'(\mathbf{R}^n)$ is a locally finite sum of linear operators of the type

$$Af(x)=(2\pi)^{-(n+N)/2}\int_{\mathbf{R}^{N+n}}\exp(i\varphi(x,\theta,y))\times a(x,\theta,y)f(y)\,dy\,d\theta. \quad (7)$$

Here $a(x,\theta,y)$ is a C^∞-function satisfying the inequality

$$|D_x^\alpha D_\theta^\beta D_y^\gamma a(x,\theta,y)|\leqslant C(1+|\theta|)^{m-\rho|\beta|+(1-\rho)(|\alpha|+|\gamma|)}$$

for some fixed m and ρ, $1/2\leqslant\rho\leqslant1$, and any triple of multi-indices α, β, γ, and for $\varphi(x,\theta,y)$ a real-valued function of class C^∞ for $\theta\neq0$ and homogeneous of degree 1 in θ there. The function φ is called the **phase function** and a the **amplitude function**.

Let $C_\varphi=\{(x,\theta,y)\,|\,d_\theta\varphi(x,\theta,y)=0,\ \theta\neq0\}$ and $W=\{(x,y)\in\mathbf{R}^n\times\mathbf{R}^n\,|\,\exists\theta\neq0$ such that $(x,\theta,y)\in C_\varphi\}$. If $d_{x,\theta,y}\varphi(x,\theta,y)\neq0$ for $\theta\neq0$, then the kernel distribution $k(x,y)$ of A is of class C^∞ outside W. There have been detailed studies of the case where the $d_{x,\theta,y}(\partial\varphi(x,\theta,y)/\partial\theta_j)$, $j=1,2,\dots,N$, are linearly independent at every point of C_φ. In this case, C_φ is a smooth manifold in $\mathbf{R}^n\times(\mathbf{R}^N\setminus0)\times\mathbf{R}^n$, and the mapping $\Phi: C_\varphi\ni(x,\theta,y)\to(x,y,\xi,\eta)$, $\xi=d_x\varphi(x,\theta,y)$, $\eta=d_y\varphi(x,\theta,y)$, is an immersion of C_φ to $T^*(\mathbf{R}^n\times\mathbf{R}^n)\setminus0$, the cotangent bundle of $\mathbf{R}^n\times\mathbf{R}^n$ minus its zero section. The image $\Phi C_\varphi=\Lambda_\varphi$ is a **conic Lagrange manifold**, i.e., the canonical 2-form $\sigma=\sum_j d\xi_j\wedge dx_j-\sum_j d\eta_j\wedge dy_j$ vanishes on Λ_φ and the multiplicative group of positive numbers acts on Λ_φ. Let $\lambda_1,\dots,\lambda_{2n}$ be a system of local coordinates in Λ_φ. These, together with $\partial\varphi/\partial\theta_1,\dots,\partial\varphi/\partial\theta_N$, constitute a system of local coordinate functions of $\mathbf{R}^n\times(\mathbf{R}^N\setminus0)\times\mathbf{R}^n$ in a conic neighborhood of C_φ. Let J denote the Jacobian determinant

$$\frac{D\left(\lambda_1,\dots,\lambda_{2n},\dfrac{\partial\varphi}{\partial\theta_1},\dots,\dfrac{\partial\varphi}{\partial\theta_N}\right)}{D(x,\theta,y)}.$$

The function $a_{\Lambda_\varphi}=\sqrt{J}\,a|_{C_\varphi}\Phi^{-1}$ is called the local **symbol** of A. Here $a|_{C_\varphi}$ is the restriction of a to C_φ. The conic Lagrange manifold $\Lambda_\varphi=\Lambda_\varphi(A)$ and the symbol $a_{\Lambda_\varphi}=a_{\Lambda_\varphi}(A)$ essentially determine the singularity of the kernel distribution $k(x,y)$ of the Fourier integral operator A. Conversely, given a conic Lagrange manifold Λ in $T^*(\mathbf{R}^n\times\mathbf{R}^n)\setminus0$ and a function a_Λ on it, one can construct a Fourier integral operator A such that $\Lambda_\varphi(A)=\Lambda$ and $a_{\Lambda_\varphi}(A)=a_\Lambda$. Those Fourier integral operators whose associated conic Lagrange manifolds are the graphs of homogeneous †canonical transformations of $T^*(\mathbf{R}^n)\setminus0$ are most frequently used in the theory of linear partial differential equations. Let A be a Fourier integral operator such that $\Lambda_\varphi(A)$ is the graph of a homogeneous canonical transformation χ. Then the adjoint of A is a Fourier integral operator such that the associated conic Lagrange manifold is the graph of the inverse transformation χ^{-1}. Let A_1 be another such operator; if the associated conic Lagrange manifold is the graph of χ_1, then the composed operator A_1A is also a Fourier integral operator and the associated conic Lagrange manifold is the graph of the composed homogeneous canonical transformation $\chi_1\chi$.

A pseudodifferential operator of class $S^m_{\rho,1-\rho}(\mathbf{R}^n)$ is a particular type of Fourier integral operator. In fact, a Fourier integral operator A is a pseudodifferential operator of class $S^m_{\rho,1-\rho}(\mathbf{R}^n)$ if and only if $\Lambda_\varphi(A)$ is the graph of the identity mapping of $T^*(\mathbf{R}^n)\setminus0$. Hence for any Fourier integral operator A, A^*A and AA^* are pseudodifferential operators.

The following theorem is due to Egorov [11]: Let P and Q be pseudodifferential operators of class $S^m_{\rho,1-\rho}(\mathbf{R}^n)$ with the symbols $p(x,\xi)$ and $q(x,\xi)$, respectively, and let A be a Fourier integral operator such that the associated conic Lagrange manifold $\Lambda_\varphi(A)$ is the graph of a homogeneous canonical transformation χ of $T^*(\mathbf{R}^n)\setminus0$. If the equality $PA=AQ$

holds, then $q(x,\xi)-p(\chi(x,\xi))$ belongs to the class $S^{m+1-2\rho}_{\rho,1-\rho}(\mathbf{R}^n)$.

Assume that $m=1$, $\rho=1$, and that $p_1(x,\xi)$ is a real-valued C^∞-function, homogeneous of degree 1 in ξ for $|\xi|>1$, such that $p(x,\xi)-p_1(x,\xi)\in S^0_{1,0}(\mathbf{R}^n)$ and $d_\xi p_1(x^0,\xi^0)\neq 0$ at (x^0,ξ^0), $|\xi^0|>1$, where $p_1(x^0,\xi^0)=0$. Then one can find a Fourier integral operator A such that the function $q(x,\xi)$ of Egorov's theorem satisfies the relation $q(x,\xi)-\xi_1\in S^0_{1,0}(\mathbf{R}^n)$.

The theory of Fourier integral operators has its origin in the asymptotic representation of solutions of the wave equation (→ 325 Partial Differential Equations of Hyperbolic Type L; also, e.g., [12, 13, 14]). G. I. Eskin (*Math. USSR-Sb.*, 3 (1976)) used a type of Fourier integral operator in deriving the energy estimates and constructing the fundamental solutions for strict hyperbolic operators. Hörmander (*Acta Math.*, 121 (1968)) introduced the term "Fourier integral operators," and applied these operators to the derivation of highly accurate asymptotic formulas for spectral functions of elliptic operators. Egorov (*Math. USSR-Sb.*, 11 (1970)) applied his theorem and the corollary stated above to the study of hypoellipticity and local solvability for pseudodifferential operators of principal type. Using Egorov's theorem and the same corollary, Nirenberg and Treves [16] obtained decisive results concerning local solvability for linear partial differential operators of principal type; these results were completed by Beals and Fefferman [17]. Hörmander and J. J. Duistermaat [9, 15] constructed a general global theory of Fourier integral operators making use of Maslov's theory [14], which was originally published in 1965. By virtue of this research, the Fourier integral operator has come to be recognized as a powerful tool in the theory of linear partial differential operators. An interesting application of the global theory of Fourier integral operators appeared in J. Chazarain (*Inventiones Math.*, 24 (1974)). The boundedness of Fourier integral operators in the spaces $L^2(\mathbf{R}^n)$ (or the space H^s) has been studied in several cases. Some sufficient conditions for boundedness have been obtained by Eskin (*Math. USSR-Sb.*, 3 (1967)), Hörmander [9], D. Fujiwara [18], Kumano-go (*Comm. Partial Diff. Eq.*, 1 (1976)), K. Asada and Fujiwara (*Japan. J. Math.*, 4 (1978)), and others. A calculus of Fourier integral operators in $\mathbf{R}^n$ was given in Kumano-go [4].

The propagation of wave front sets by means of a Fourier integral operator is described as follows. Let us consider a phase function of the form $\varphi(x,\xi,y)=S(x,\xi)-y\cdot\xi$ in $\mathbf{R}^n_x\times\mathbf{R}^n_\xi\times\mathbf{R}^n_y$, and let $a(x,\xi)$ be an amplitude function independent of y of class $S^m_{1,0}$. Then by (7) the Fourier integral operator $A=A_S$ is defined by

$$A_Su(x)=(2\pi)^{-n/2}\int_{\mathbf{R}^n}\exp(iS(x,\xi))a(x,\xi)\hat{u}(\xi)d\xi. \tag{8}$$

Let T be the canonical transformation with the †generating function $S(x,\xi)$, i.e., T is defined by $y=\nabla_\xi S(x,\eta)$, $\xi=\nabla_x S(x,\eta)$. Then for the Fourier integral operator A_S we have

$$\mathrm{WF}(A_Su)\subset\{(x,\xi)=T(y,\eta)\,|\,(y,\eta)\in\mathrm{WF}(u)\},\qquad u\in\mathscr{S}'. \tag{9}$$

Next consider a hyperbolic operator $L=D_t+p(t,x,D_x)$ for a real-valued symbol $p(t,x,\xi)\in\mathscr{B}^0([0,T_0];S^1_{1,0})$ with some $T_0>0$. For a small $0<T\leqslant T_0$ the solution $S(t,x,\xi)$ of the eikonal equation $\partial_t S+p(t,x,\nabla_x S)=0$ on $[0,T]$ with the initial condition $S|_{t=0}=x\cdot\xi$ exists in $\mathscr{B}^1([0,T];S^1_{1,0})$. Consider the Cauchy problem $Lu=0$ on $[0,T]$, $u|_{t=0}=u_0$. Then there exists an amplitude function $e(t,x,\xi)\in\mathscr{B}^1([0,T];S^0_{1,0})$ such that the solution $u(t)$ is found in the form $u(t)=E_S(t)u_0$. On the other hand, let $(x,\xi)=(X(t,y,\eta),\Xi(t,y,\eta))$ be the bicharacteristic strip defined by †Hamilton's canonical equation $dx/dt=\nabla_\xi p(t,x,\xi)$, $d\xi/dt=-\nabla_x p(t,x,\xi)$ with $(x,\xi)|_{t=0}=(y,\eta)$. Then $(X(t,y,\eta),\Xi(t,y,\eta))$ can be solved by means of the relations $y=\nabla_\xi S(t,X,\eta)$, $\Xi=\nabla_x S(t,X,\eta)$, as a family of canonical transformations with a parameter $t\in[0,T]$. Thus by means of (9) we have

$$\mathrm{WF}(u(t))\subset\{(x,\xi)=(X(t,y,\eta),\Xi(t,y,\eta))\,|\,(y,\eta)\in\mathrm{WF}(u_0)\}, \tag{10}$$

which is the fundamental result in the study of the propagation of wave front sets as solutions of general hyperbolic equations (→ 325 Partial Differential Equations of Hyperbolic Type M).

The works of Egorov, Nirenberg and Treves, and Hörmander motivated the theory of hyperfunctions developed by M. Sato and gave rise to the concept of †quantized contact transformations, which correspond to Fourier integral operators in the theory of distributions. The above-stated transformation theorem of Egorov has been studied in detail with reference to systems of pseudodifferential equations with analytic coefficients [19] (→ 274 Microlocal Analysis).

References

[1] J. J. Kohn and L. Nirenberg, An algebra of pseudo-differential operators, Comm. Pure Appl. Math., 18 (1965), 269–305.

[2] L. Hörmander, Pseudo-differential operators, Comm. Pure Appl. Math., 18 (1965), 501–517.
[3] L. Hörmander, Pseudo-differential operators and hypoelliptic equations, Amer. Math. Soc. Proc. Symp. Pure Math., 10 (1967), 138–183.
[4] H. Kumano-go, Pseudo-differential operators, MIT Press, 1981. (Original in Japanese, 1974.)
[5] A. P. Calderón and R. Vaillancourt, A class of bounded pseudo-differential operators, Proc. Nat. Acad. Sci. US, 69 (1972), 1185–1187.
[6] L. Hörmander, Pseudo-differential operators and non-elliptic boundary problems, Ann. Math., 83 (1966), 129–209.
[7] P. D. Lax and L. Nirenberg, On stability for difference schemes, a sharp form of Gårding's inequality, Comm. Pure Appl. Math., 19 (1966), 473–492.
[8] L. Hörmander, The Weyl calculus of pseudo-differential operators, Comm. Pure Appl. Math., 32 (1979), 359–443.
[9] L. Hörmander, Fourier integral operators I, Acta Math., 128 (1971), 79–183.
[10] J. J. Duistermaat, Fourier integral operators, Lecture notes, Courant Institute, 1973.
[11] Yu. V. Egorov, On canonical transformations of pseudo-differential operators (in Russian), Uspekhi Mat. Nauk, 25 (1969), 235–236.
[12] P. D. Lax, Asymptotic solutions of oscillatory initial value problems, Duke Math. J., 24 (1957), 627–646.
[13] D. Ludwig, Exact and asymptotic solutions of the Cauchy problem, Comm. Pure Appl. Math., 13 (1960), 473–508.
[14] V. P. Maslov, Théorie des perturbations et méthodes asymptotiques, Gauthier-Villars, 1972.
[15] J. J. Duistermaat and L. Hörmander, Fourier integral operators II, Acta Math., 128 (1972), 183–269.
[16] L. Nirenberg and F. Treves, On local solvability of linear partial differential equations I, II, Comm. Pure Appl. Math., 23 (1970), 1–38, 459–510.
[17] R. Beals and C. Fefferman, On local solvability of linear partial differential equations, Ann. Math., (2) 97 (1973), 482–498.
[18] D. Fujiwara, On the boundedness of integral transformations with highly oscillatory kernels, Proc. Japan Acad., 51 (1975), 96–99.
[19] H. Komatsu (ed.), Hyperfunctions and pseudo-differential equations, Lecture notes in math. 287, Springer, 1973.

346 (XVIII.17) Psychometrics

A. General Remarks

Psychometrics is a collection of methods for drawing statistical conclusions from various psychological phenomena which are expressed numerically or quantitatively. It consists chiefly of statistical methods to deal with psychological measurements and of theories dealing with mathematical models concerning learning processes, social attitudes, and mental abilities.

B. Sensory Tests

A measurement wherein human senses are taken as the gauge is called a **sensory test**. The panel of judges must be composed appropriately, and the examining circumstances must be controlled. Various methods of psychological measurements are applied. In the following sections we describe the basic statistical procedures used in sensory testing.

C. Paired Comparison

When there are t objects (treatments or stimuli in some cases) $O_1, O_2, \ldots, O_t$, the method of comparing them two at a time in every possible way is called **paired comparison**. The following are typical mathematical models of this method.

(1) Thurstone-Mosteller Model. Suppose that the probability that O_i is preferred to O_j for a pair (O_i, O_j) is p_{ij}. Of the n judges who compare this pair, the number who prefer O_i is n_{ij}, and the number who prefer O_j is $n_{ji}=n-n_{ij}$. In this comparison it is assumed that the strengths of the stimuli O_i, O_j to the senses are random variables X_i, X_j, and O_i is preferred when $X_i > X_j$. Furthermore, it is assumed that the joint probability distribution of X_i and X_j is the 2-dimensional †normal distribution with μ_i and σ^2 as mean and variance of X_i, and ρ as correlation coefficient of X_i and X_j. There is no loss of generality in assuming that $2\sigma^2(1-\rho)=1$ and $\sum_{i=1}^{t}\mu_i=0$. Let $\Phi(x)$ be the standardized normal distribution function and $p_{ij}=\Phi(\mu_i-\mu_j)$. Using $p'_{ij}=n_{ij}/n$ as estimates of the true p_{ij}, we can obtain the estimates $\hat{\mu}_i$. Using $p''_{ij}=\Phi(\hat{\mu}_i-\hat{\mu}_j)$ and p'_{ij} we can test the hypothesis that $\mu_1=\mu_2=\ldots=\mu_t$.

(2) The Bradley-Terry Model. The experimental method in the Bradley-Terry model

is the same as in the Thurstone-Mosteller model. It is postulated that, associated with $O_1, O_2, \ldots, O_t$, there exist parameters π_i for O_i ($\pi_i \geqslant 0$, $\sum_{i=1}^{t} \pi_i = 1$) such that $p_{ij} = \pi_i/(\pi_i + \pi_j)$. Obtaining the †maximum likelihood estimator of π_i, we can test the appropriateness of the models.

(3) Scheffé's Model. Each pair O_i, O_j is presented to $2n$ judges; n of them examine O_i first and O_j next, and the remaining n examine the pair in the opposite order. A judgment is recorded on a 7-point (or 9-, 5-, or 3-point) scale. In the 7-point scaling system a judge presented with the ordered pair (O_i, O_j) marks one of the seven points 3, 2, 1, 0, -1, -2, -3, meaning, respectively, that O_i is strongly preferable to O_j, O_i is moderately preferable to O_j, O_i is slightly preferable to O_j, no preference, O_j is slightly preferable to O_i, etc. The mark given by the kth judge on his preference of O_i to O_j is denoted by X_{ijk}, which can be regarded as the sum of a main effect, deviation of subtractivity, order effect, and error. Significance tests for these effects and estimates of various parameters are given by using statistical †linear models. †BIBD, †PBIBD, etc., can also be applied to paired comparisons.

D. The Pair Test, Triangle Test, and Duo-Trio Test

The pair test, triangle test, and duo-trio test are sensory difference tests. The methods are as follows. **Pair test**: A judge is requested to designate a preference between the paired samples A and B. **Triangle test**: A judge is requested to select two samples of the same kind out of A, A, B. **Duo-trio test**: A judge is first acquainted with a sample A and then is requested to choose from A and B the one he has seen in the previous step. In all the above cases, the hypothesis that A and B are different and that the judge has no ability to determine the difference between them is tested by using the †binomial distribution.

E. Scaling

(1) One-Dimensional Case. Psychometric **scaling methods** are procedures for constructing scales for psychological phenomena. Some of them require judgments concerning a particular attitude that is considered unidimensional. Under the assumption that a psychological phenomenon is a random variable with some distribution law and the parameters of the distribution law determine psychological scales, a psychological scaling is given by estimating the parameters. The Thurstone-Mosteller model is a method for scaling a set of stimuli by means of observable proportions.

(2) Multidimensional Scaling (MDS). Multidimensional scaling is a collection of methods to deal with data consisting of many measurements on many objects and to characterize the mutual distance (dissimilarity), or closeness (affinity), by representing those objects by a small number of indices or by points in a small-dimensional Euclidean space. It has seen useful applications in the analysis of people's attitude and perception and their characterizations by means of a few numbers or points in a space of low dimension.

Historically, MDS was first developed by Torgerson (1958) and refined further by Shepard (1962) and Kruskal (1964). The method developed by Torgerson and also the INDSCAL method by Carrol and Chang (1970) are called **metric multidimensional scaling**, while the method by Shepard and Kruskal is called the **nonmetric MDS**. The former is applied when the data are represented in continuous scales and the latter when the data are in discrete nominal or ordinal scales. Techniques of multidimensional scaling are closely related and sometimes actually equivalent to various methods of multivariate analysis, especially principal component analysis, canonical correlation analysis, and discriminant analysis (→ 280 Multivariate Analysis).

F. Factor Analysis

Though factor analysis can be considered to be a method to deal with multivariate data in general (→ 280 Multivariate Analysis), it has had close connections with psychometric studies, in both theoretical developments and applications. Historically, it was initiated by Spearman (1927) and developed further by Thurston (1945) in order to measure human abilities from test scores. Mathematically, the model of factor analysis is formulated as follows: Let z_{jk} be the standardized score of the jth test achieved by the kth subject, $j = 1, \ldots, p$; $k = 1, \ldots, N$; then it is assumed that it can be represented as a linear combination of r common factors and one specific factor as

$$z_{jk} = a_{j1} f_{1k} + a_{j2} f_{2k} + \ldots + a_{jr} f_{rk} + u_j v_{jk}, \qquad (1)$$

where f_{ik}, $i = 1, \ldots, r$, $k = 1, \ldots, N$, represents the magnitude of the ith common factor (ability) in the kth subject and a_{ji} is the size of contribution of the ith factor to the score of the jth test.

Usually it is assumed that (i) $V(z_j) = 1$, (ii) $V(f_i) = 1$ and $\mathrm{Cov}(f_i, f_{i'}) = 0$ for $i \neq i'$, (iii) $V(v_j) = 1$ and $\mathrm{Cov}(v_j, v_{j'}) = 0$, $j \neq j'$, (iv) $\mathrm{Cov}(f_i, v_j) = 0$.

Then it follows that

$$r_{jj'} = \mathrm{Cor}(z_j, z_{j'}) = a_{j1}a_{j'1} + a_{j2}a_{j'2} + \ldots + a_{jr}a_{j'r},$$

and

$$1 = a_{j1}^2 + a_{j2}^2 + \ldots + a_{jr}^2 + u_j{}^2 = h_j^2 + u_j^2,$$

where h_j^2 is called the **communality** of the jth variable z_j and u_j^2 is called the **specificity**. It follows easily from (1) that any orthogonal transformation of the scores does not affect the model.

Problems of factor analysis are classified into three types:
(1) Estimation of communality: There are several methods of determining communality or initial estimates of it when some iterative procedure is used.
(2) Determination of **factor loadings**, which is the estimation of the a_{ji}: A number of methods have been proposed, among which those often used are the MINRES by Harman (1967), the varimax by Kaiser (1958), and the †maximum likelihood based on the normal model by Lawley and Maxwell.
(3) Estimation of **factor scores** f_{ik}: Usually factor scores are estimated after factor loadings have been determined. Thurstone proposed $\hat{F} = ZR^{-1}A$ and Harman $\hat{F} = ZA(A'A)^{-1}$, where F, Z, A are the matrices of factor scores, test scores, factor loadings, respectively, and R is the correlation matrix of the z's.

G. Learning Theory

(1) General Description. Assume that a sequence of trials is done in order to study some given behavior and that on each trial particular events occur (stimuli, responses, reinforcements, etc.) that influence the ensuing behavior. Then the behavior itself is modified by such a sequence of trials. **Learning models** refer to such processes of behavior modification, and they are frequently represented by recursive formulas for response probabilities.

Assume that two mutually exclusive response alternatives A_1 and A_2 occur on the nth trial $(n = 1, 2, \ldots)$ with respective probabilities P_n and $1 - P_n$, and that an event E_i occurs on the nth trial with $\Pr(\mathscr{E}_n = E_i) = \pi_i$ $(i = 1, 2, \ldots, t;$ $\sum_{i=1}^t \pi_i = 1)$. Then the recursive formula for P_n is of the form $P_{n+1} = f(P_n; \mathscr{E}_n, \mathscr{E}_{n-1}, \ldots, \mathscr{E}_1)$. If the formula can be written as $f = f(P_n; \mathscr{E}_n, \mathscr{E}_{n-1}, \ldots, \mathscr{E}_{n-d})$, then the response probability is called **d-trial path dependent**. In the special case $d = 0$, it is called **path independent**. For simplicity we write $f(P_n; \mathscr{E}_n = E_i, \mathscr{E}_{n-1} = E_j, \ldots, \mathscr{E}_1 = E_k) = f_{ij\ldots k}(P_n)$ $(i, j, \ldots, k = 1, 2, \ldots, t)$. If the response probability is path independent, then $f_{ij\ldots k}(P_n) = f_i f_j \ldots f_k(P_1)$, where $f_v(P_n) = f(P_n; \mathscr{E}_n = E_v)$ $(v = i, j, \ldots, k)$. When the recursive formula can be expressed as $f = f(P_n; \mathscr{E}_n, n)$, the response probability is said to be **quasi-independent of path** [11]. In the recursive formula f, two events E_i and E_j $(i \neq j)$ are said to be commutative if $f_{\ldots i \ldots j \ldots}(P_n) = f_{\ldots j \ldots i \ldots}(P_n)$. If any two events are commutative, the condition of **event commutativity** is satisfied. By making f explicit with respect to n, we write $P_n = F(n; \mathscr{E}_{n-1}, \ldots, \mathscr{E}_1; P_1)$. Under the condition of event commutativity, the explicit formula can be written $P_n = F(N_1, N_2, \ldots, N_t; P_1)$, where N_i is the frequency of occurrence of E_i in the first $(n-1)$ trials $(\sum_{i=1}^t N_i = n - 1)$. If both event commutativity and path independence of response probability are satisfied, the explicit formula $P_n = f_1^{N_1} f_2^{N_2} \ldots f_t^{N_t}(P_1)$ can be obtained.

(2) Linear Models. In a linear model, the recursive formula is written as a linear function of P_n.

Example (1). Bush-Mosteller model [6]. The Bush-Mosteller model assumes the response probability to be path independent. The recursive formula is expressed as $f_i(P_n) = \alpha_i P_n + (1 - \alpha_i)\lambda_i$, $\mathscr{E}_n = E_i$. Here α_i $(0 \leqslant \alpha_i \leqslant 1)$ represents the degree of ineffectiveness of E_i for learning and λ_i $(0 \leqslant \lambda_i \leqslant 1)$ is the †fixed point of f_i. A necessary and sufficient condition for E_i and E_j $(i \neq j)$ to be commutative is that either f_i or f_j be an †identity operator or $\lambda_i = \lambda_j$.

Example (2). Estes's stimulus-sampling model [7]. We can consider the stimulus as a set composed of m elements, each of which corresponds to either response A_1 or A_2; the manner of their correspondences depends on each trial. If J_n elements correspond to A_1 on the nth trial, then we have $P_n = J_n/m$. Suppose that on the nth trial s $(\leqslant m)$ elements are sampled, among which X_n elements correspond to A_2, and the remaining X_n' $(= s - X_n)$ elements correspond to A_1. As a result of the nth trial, if A_1 is reinforced, we set $Y_n = 1$; otherwise, we set $Y_n = 0$. Furthermore, assume that $J_{n+1} = J_n + X_n Y_n - X_n'(1 - Y_n)$. Hence, we obtain the recursive formula $P_{n+1} = P_n + \{X_n Y_n - X_n'(1 - Y_n)\}/m$. In this model, the response probability is path independent and $\mathscr{E}_n = (X_n, Y_n)$. Other linear models have been proposed in which the response probability is either quasi-independent of path [8] or path dependent [10].

(3) Nonlinear Models. In nonlinear models the recursive formula cannot be written as a linear function of P_n.

Example (3). Luce's β-model [9]. Let the response strengths of A_1 and A_2 on the nth trial be v_n and v_n', respectively (both positive), and assume that P_n, the response probability of A_1, is expressed as $v_n/(v_n + v_n')$. The response

strengths v_n and v'_n depend on each trial. Under the assumption that the response strength is path independent and that v_{n+1} changes independently from v'_n, the recursive formula of v_n is written as $v_{n+1} = \varphi_i(v_n)$, $\mathscr{E}_n = E_i$. Here, if we assume $\varphi_i(v) > 0$ for $v > 0$ and $\varphi_i(cv) = c\varphi_i(v)$ for $v > 0$ and $c > 0$, then $\varphi_i(v_n) = \beta_i v_n$ with $\beta_i > 0$. In a similar way, the recursive formula for v'_n can be expressed as $\varphi'_i(v'_n) = \beta'_i v'_n$ ($\beta'_i > 0$). Therefore we have $P_{n+1} = P_n / \{P_n + b_i(1-P_n)\}$ ($b_i = \beta'_i / \beta_i$), $\mathscr{E}_n = E_i$. This model is nonlinear, and the response probability is path independent. By making the recursive formula explicit, we obtain $P_n = P_1 / \{P_1 + (1-P_1)\exp(\sum_{i=1}^{t} N_i \log b_i)\}$. Hence it is clear that the events are commutative. Other nonlinear models in which the response probability is either quasi-independent of path [5] or path dependent [9] have also been proposed.

Here we have taken up only the case in which the number of response alternatives is 2, but we can generalize to the case of more than two alternatives. For fitting a model and experimental data, expected response probabilities and various other statistics deduced from the model (total error, trial number of first success or last error, and sequential statistics such as length of response †run or †autocorrelation between responses) are used. Estimation methods have also been devised for the parameters involved.

References

For sensory tests,
[1] H. A. David, The method of paired comparisons, Hafner, 1963.
[2] J. P. Guilford, Psychometric methods, McGraw-Hill, second edition, 1954.
[3] M. G. Kendall, Rank correlation methods, Griffin, third edition, 1962.
[4] W. S. Torgerson, Theory and method of scaling, Wiley, 1958.
For learning theory,
[5] R. J. Audley and A. R. Jonckheere, The statistical analysis of the learning process, Brit. J. Statist. Psychol., 9 (1956), 87–94.
[6] R. R. Buch and F. Mosteller, Stochastic models for learning, Wiley, 1955.
[7] W. K. Estes, Component and pattern models with Markovian interpretations, Studies in Mathematical Learning Theory, R. R. Bush and W. K. Estes (eds.), Stanford Univ. Press, 1959, 9–52.
[8] M. I. Hanania, A generalization of the Bush-Mosteller model with some significance tests, Psychometrika, 24 (1959), 53–68.
[9] R. D. Luce, Individual choice behavior: A theoretical analysis, Wiley, 1959.
[10] S. H. Sternberg, A path-dependent linear model, Studies in Mathematical Learning Theory, R. R. Bush and W. K. Estes (eds.), Stanford Univ. Press, 1959, 308–339.
[11] S. H. Sternberg, Stochastic learning theory, Handbook of Mathematical Psychology II, R. D. Luce, R. R. Bush, and E. Galanter (eds.), Wiley, 1963, 1–120.
For MDS,
[12] J. B. Kruskal, Nonmetric multidimensional scaling: A numerical method, Psychometrika, 29 (1964), 115–129.
[13] R. N. Shepard, The analysis of proximities: Multidimensional scaling with an unknown distance function, Psychometrika, 27 (1962), 125–219.
[14] L. Guttman, Multiple rectinear prediction and the resolution into components, Psychometrika, 5 (1940), 75–99.
[15] J. B. Carrol and J. J. Chang, Analysis of individual difference in multidimensional scaling via an N-way generalization of Eckart-Young decomposition, Psychometrika, 35 (1970), 283–319.
For factor analysis,
[16] C. Spearman, The abilities of man, Macmillan, 1927.
[17] H. H. Harman, Modern factor analysis, Univ. of Chicago Press, 1967.
[18] H. F. Kaiser, The varimax criterion for analytic rotation in factor analysis, Psychometrika, 23 (1958), 187–240.
[19] D. N. Lawley and A. E. Maxwell, Factor analysis as a statistical criterion, Butterworth, 1963.
[20] K. Takeuchi, H. Yanai, and B. N. Mukherjee, Foundations of multivariate analysis, Wiley, 1982.

Q

347 (V.12) Quadratic Fields

A. General Remarks

Any †extension field of the rational number field $\mathbf{Q}$ of degree 2 is called a **quadratic field**. Any quadratic field k is obtained from $\mathbf{Q}$ by adjoining a square root of a **square-free integer** (i.e., an integer $\neq 0, 1$ with no square factor) $m: k=\mathbf{Q}(\sqrt{m})$. If m is positive (negative), k is called a **real (imaginary** or **complex) quadratic field** (→ 14 Algebraic Number Fields). Let

$$\omega=\begin{cases}(1+\sqrt{m})/2 & \text{for } m\equiv 1 \pmod 4,\\ \sqrt{m} & \text{for } m\equiv 2,3 \pmod 4.\end{cases}$$

Then $(1,\omega)$ is a †minimal basis of k. That is, any †algebraic integer α of k has the unique expression $\alpha=a+b\omega$ with $a,b\in\mathbf{Z}$. The †discriminant d of k is given by $d=m$ in case $m\equiv 1$ (mod 4) and $d=4m$ in case $m\equiv 2,3$ (mod 4). The conjugate element of an element $\alpha=a+b\sqrt{m}$ $(a,b\in\mathbf{Q})$ of k over $\mathbf{Q}$ is given by $\alpha'=a-b\sqrt{m}$. The mapping $\sigma:\alpha\to\alpha'$ is an †automorphism of the field k.

B. Units

Let k be an imaginary quadratic field. The †units of k are ± 1, $\pm i$ in case $m=-1$; ± 1, $\pm\omega_0$, $\pm\omega_0^2$ $(\omega_0=(1+\sqrt{-3})/2)$ in case $m=-3$; and ± 1 in all other cases.

Let k be a real quadratic field. There exists a unit ε_0 that is the smallest one among the units (>1) of k. Any unit ε of k can be uniquely expressed in the form $\varepsilon=\pm\varepsilon_0^n$ $(n\in\mathbf{Z})$. That is, $\pm\varepsilon_0^{\pm 1}$ is a †fundamental unit of k. The fundamental unit $\varepsilon_0=(x+y\sqrt{d})/2$ (>1) can be calculated by finding a minimal positive integral solution (x,y) of †Pell's equation $x^2-dy^2=\pm 4$ by using continued fractions (→ 83 Continued Fractions; for a table of the fundamental unit of k for $m<100$ → [1]).

C. Prime Ideals

The decomposition of a prime number p in the †principal order $\mathfrak{o}$ of k is given as follows: (i) Let $p\mid d$, where d is the discriminant of k. Then p is decomposed in $\mathfrak{o}$ in the form $(p)=\mathfrak{p}^2$. (ii) Let $p\neq 2$ and $(p,m)=1$. If $(m/p)=1$, then $(p)=\mathfrak{p}\mathfrak{p}'$ in $\mathfrak{o}$ $(\mathfrak{p}\neq\mathfrak{p}')$ and $N(\mathfrak{p})=N(\mathfrak{p}')=p$. If $(m/p)=-1$, then $(p)=\mathfrak{p}$ in $\mathfrak{o}$ and $N(\mathfrak{p})=p^2$. Here $\mathfrak{p}$, $\mathfrak{p}'$ are prime ideals of $\mathfrak{o}$, N means the †norm, and (m/p) is the †Legendre symbol. (iii) Let $2\nmid d$, that is, $m\equiv 1$ (mod 4). If $m\equiv 1$ (mod 8), then $(2)=\mathfrak{p}\mathfrak{p}'$ $(\mathfrak{p}\neq\mathfrak{p}')$ and $N(\mathfrak{p})=N(\mathfrak{p}')=2$. If $m\equiv 5$ (mod 8), then $(2)=\mathfrak{p}$ and $N(\mathfrak{p})=4$.

D. The Kronecker Symbol

Let $k=\mathbf{Q}(\sqrt{m})$, and let d be the discriminant of k. We define the symbol χ for k as follows: (i) $\chi(p)=1$ if $(p)=\mathfrak{p}\mathfrak{p}'$ $(\mathfrak{p}\neq\mathfrak{p}')$ in $\mathfrak{o}$; (ii) $\chi(p)=-1$ if $(p)=\mathfrak{p}$ in $\mathfrak{o}$; (iii) $\chi(p)=0$ if $(p)=\mathfrak{p}^2$ in $\mathfrak{o}$; and (iv) $\chi(n)=\prod_i\chi(p_i)^{e_i}$ for $n=\prod_i p_i^{e_i}>0$. In particular, we define $\chi(1)=1$. If $(n,d)=1$, the symbol χ can also be defined using the †Jacobi symbol as follows: If $m\equiv 1$ (mod 4), then $\chi(n)=(n/|m|)$; and if $m\equiv 2,3$ (mod 4), then $\chi(n)=\chi_2^*(n)(n/|m'|)$ for $d=2^e m'$, where (i) $\chi_2^*(n)=(-1)^{(n-1)/2}$ for $e=2$, $m'\equiv 3$ (mod 4); (ii) $\chi_2^*(n)=(-1)^{(n^2-1)/8}$ for $e=3$, $m'\equiv 1$ (mod 4); and (iii) $\chi_2^*(n)=(-1)^{(n^2-1)/8+(n-1)/2}$ for $e=3$, $m'\equiv 3$ (mod 4). If $(n,d)\neq 1$, then $\chi(n)=0$. For a negative integer $-n$, we define $\chi(-n)=(\operatorname{sgn} d)\chi(n)$. The symbol $\chi(n)$ for $n\in\mathbf{Z}$ is called the **Kronecker symbol** for k.

The Kronecker symbol for k has the following four properties: (1) $\chi(n)=0$ if $(n,d)\neq 1$, and $\chi(n)=\pm 1$ if $(n,d)=1$; (2) $\chi(m)=\chi(n)$ if $m\equiv n$ (mod d); (3) $\chi(mn)=\chi(m)\chi(n)$; (4) $\chi(n)=1$ if and only if $n\equiv N(\mathfrak{a})$ (mod d) for some integral ideal $\mathfrak{a}$ of k such that $(\mathfrak{a},(d))=1$. (Property (4) shows that a quadratic field provides a class field; → 59 Class Field Theory.)

E. Ideal Classes

The †class number h of k was calculated by P. G. L. †Dirichlet (1840) by analytical methods as follows:

$$h\log\varepsilon_0=-\frac{1}{2}\sum_{r=1}^{d-1}\chi(r)\log\sin\frac{r\pi}{d}\quad\text{for } m>0,$$

$$h=\frac{-w}{2|d|}\sum_{r=1}^{|d|-1}\chi(r)r\quad\text{for } m<0,$$

where d is the discriminant of k, ε_0 is the positive fundamental unit (>1) of k, w is the number of roots of unity in k, and χ is the Kronecker symbol.

Denote by $h(d)$ the class number of the imaginary quadratic field with discriminant d. It was conjectured from the time of C. F. †Gauss that $h(d)\to\infty$ as $|d|\to\infty$. This conjecture was proved by H. Heilbronn (1934). More precisely, C. L. Siegel (*Acta Arith.*, 1 (1935)) proved

$$\lim_{|d|\to\infty}(\log h(d))/(\log\sqrt{|d|})=1.$$

$h(d)=1$ holds for $|d|=3,4,7,8,11,19,43,67,163$. In 1934, Heilbronn and E. H. Linfoot proved that there can be at most one more such d, and finally A. Baker and H. M. Stark independently proved that these nine numbers are the only ones for which $h(d)=1$ (Baker, *Mathematica*, 13 (1966); Stark, *Michigan Math*

J., 14 (1967)). Also, Baker and Stark proved independently (*Ann. Math.*, (2) 94 (1971); *Math. Comp.*, 29 (1975)) that $h(d)=2$ holds only for $|d|=15, 20, 24, 35, 40, 51, 52, 88, 91, 115, 123, 148, 187, 232, 235, 267, 403$, and 427.

For real quadratic fields, we have

$$\lim_{d\to\infty}(\log(h(d)\log\varepsilon_d))/(\log\sqrt{d})=1,$$

where ε_d is the fundamental unit of $k=\mathbf{Q}(\sqrt{d})$ ($\varepsilon_d>1$). However, it is not yet determined whether there exist infinitely many d with $h(d)=1$ (→ Appendix B, Table 4).

F. Genera

Let G be the group of all (†fractional) ideals of k, and let H be the group of all †principal ideals (α) of k such that $N(\alpha)>0$. Each coset of G modulo H is called an **ideal class in the narrow sense**. (This notion is a special case of the notion of ideal classes in the narrow sense of algebraic number fields; → 14 Algebraic Number Fields G.) In the cases (i) $m<0$ and (ii) $m>0$, $N(\varepsilon_0)=-1$, the usual classification of ideals and classification of ideals in the narrow sense are identical. When $m>0$, $N(\varepsilon_0)=1$, each ideal class is divided into two ideal classes in the narrow sense. We call an ideal class in the narrow sense simply an ideal class.

Let $p_1, \ldots, p_t$ be the set of all prime numbers dividing d. For $n\in\mathbf{Z}$ with $(n,d)=1$, we define $\chi_1(n), \ldots, \chi_t(n)$ as follows: For $p_i\neq 2$, we define $\chi_i(n)=(n/p_i)$; for $p_i=2$, we identify χ_i with χ_2^* in the definition of the Kronecker symbol. In order that $\chi_1(n)=\ldots=\chi_t(n)=1$ for $n\in\mathbf{Z}$ with $(n,d)=1$, it is necessary and sufficient that $n\equiv N(\alpha) \pmod d$ for an integer α of k (where $N(\alpha)=\alpha\alpha'$). Since $\chi_1\chi_2\ldots\chi_t$ is equal to the Kronecker symbol, it follows that $n\equiv N(\mathfrak{a}) \pmod d$ for an integral ideal $\mathfrak{a}$ of k is a necessary and sufficient condition for $\chi_1(n)\ldots\chi_t(n)=1$ to hold for $n\in\mathbf{Z}$ with $(n,d)=1$. Put $\varepsilon_i=\chi_i(N(\mathfrak{a}))$ $(i=1,\ldots,t)$, where $\mathfrak{a}$ is an integral ideal with $(\mathfrak{a},(d))=1$. Then $(\varepsilon_1,\ldots,\varepsilon_t)$ is uniquely determined for the ideal class C containing $\mathfrak{a}$ and does not depend on the choice of $\mathfrak{a}$. The set $\mathfrak{H}$ of all ideal classes of k such that $(\varepsilon_1,\ldots,\varepsilon_t)=(1,\ldots,1)$ is called the **principal genus** of k, and $\mathfrak{H}$ is a subgroup of the ideal class group $\mathfrak{C}$ of k. Each coset of $\mathfrak{C}$ modulo $\mathfrak{H}$ is called a **genus** of k. For each genus, the values $\varepsilon_i=\chi_i(N(\mathfrak{a}))$ $(i=1,\ldots,t)$ are uniquely determined. We call $(\varepsilon_1,\ldots,\varepsilon_t)$ the **character system** of this genus, and each genus is uniquely determined by its character system. A necessary and sufficient condition for $(\varepsilon_1,\ldots,\varepsilon_t)$ to be a character system for some genus is that $\varepsilon_i=\pm 1$ and $\varepsilon_1\varepsilon_2\ldots\varepsilon_t=1$. Hence there are 2^{t-1} genera of k, and $\mathfrak{C}/\mathfrak{H}$ is an Abelian group of type $(2,2,\ldots,2)$.

In order that an ideal class C belong to the principal genus, it is necessary and sufficient that $C=C_1^{1-\sigma}=C_1^2$ hold for some ideal class C_1. From this it follows that there are $t-1$ †invariants of the ideal class group $\mathfrak{C}$ of k that are powers of 2. Each ideal class C such that $C^\sigma=C$ is called an **ambig class** of k. There are 2^{t-1} ambig classes, and they form an Abelian group of type $(2,2,\ldots,2)$. Each ideal $\mathfrak{a}$ of k with $\mathfrak{a}^\sigma=\mathfrak{a}$ is called an **ambig ideal** of k. Let $(p_i)=\mathfrak{p}_i^2$ $(i=1,\ldots,t)$. Then each ambig ideal is uniquely expressed in the form $\mathfrak{a}=\mathfrak{p}_1^{\nu_1}\ldots\mathfrak{p}_t^{\nu_t}(a)$ by some $a\in\mathbf{Q}$ and $\nu_i=0, 1$. Each ambig class contains exactly two ambig ideals of the form $\mathfrak{p}_1^{\nu_1}\ldots\mathfrak{p}_t^{\nu_t}$. For example, for $k=Q(\sqrt{-65})$ we have $d=-2^2\cdot 5\cdot 13$, $t=3$, $h=8$, $\mathfrak{C}=\{E, A, A^2, A^3, B, AB, A^2B, A^3B\}$, where $A^4=E$, $B^2=E$, $A^\sigma=A^3$, and $B^\sigma=B$; the principal genus is $\mathfrak{H}=\{E, A^2\}$, and the ambig classes are $\{E, A^2, B, A^2B\}$.

G. Norm Residues

A quadratic field k is the †class field over $\mathbf{Q}$ for an ideal group H. The †conductor of H is said to be the **conductor** of $k/\mathbf{Q}$ or simply the conductor of k. The conductor $f=\prod_p f_p$ of $k=\mathbf{Q}(\sqrt{m})$ is given by $f=d$ for $m>0$ and $f=dp_\infty$ for $m<0$. That is, the p-conductor $f_p=p$ for $p\,|\,d$, $p\neq 2$; and $f_2=2^e$ for $2\,|\,d$, $d=2^e m'$ $((2,m')=1)$. By means of the †Hilbert norm-residue symbol, the Kronecker symbol χ is expressed by

$$\chi(a)=\begin{cases}\prod_{p|d}\left(\dfrac{a,d}{p}\right) & \text{for } (a,d)=1,\\ 0 & \text{for } (a,d)\neq 1.\end{cases}$$

H. History

The arithmetic of quadratic fields was originally developed in terms of the theory of binary quadratic forms with rational integral coefficients by Gauss and Dirichlet [2]. The theory was then translated into the terms of ideal theory by J. W. R. Dedekind [2] (→ 348 Quadratic Forms M). For example, the theory of genera for quadratic fields explained in Section F was first developed by Gauss in terms of binary quadratic forms, and the class number formula was obtained by Dirichlet as a formula for binary quadratic forms. Hilbert [4, ch. 2] developed the arithmetic of quadratic fields systematically by introducing the Hilbert norm-residue symbol (→ [1, 5, 6]). Later the arithmetic of quadratic fields assumed the aspect of a simple example of class field theory (→ 59 Class Field Theory).

References

[1] Z. I. Borevich and I. R. Shafarevich, Number theory, Academic Press, 1966. (Original in Russian, 1964.)
[2] P. G. L. Dirichlet, Vorlesungen über Zahlentheorie, herausgegeben und mit Zusätzen versehen von R. Dedekind, Viewig, fourth edition, 1894 (Chelsea, 1969).
[3] H. Hasse, Vorlesungen über Zahlentheorie, Springer, 1950.
[4] D. Hilbert, Die Theorie der algebraischen Zahlkörper, Jber. Deutsch. Math. Verein. 4, I–XVIII (1897), 175–546. (Gesammelte Abhandlungen I, Springer, 1932, 63–363; Chelsea, 1967.)
[5] B. W. Jones, The arithmetic theory of quadratic forms, Math. Assoc. of America and Wiley, 1950.
[6] E. G. H. Landau, Vorlesungen über Zahlentheorie I, II, III, Hirzel, 1927 (Chelsea, 1969).

348 (III.15) Quadratic Forms

A. General Remarks

A **quadratic form** Q is a quadratic homogeneous polynomial with coefficients in a †field K, written

$$Q(x_1, \ldots, x_n) = \sum_{1 \leqslant i \leqslant k \leqslant n} c_{ik} x_i x_k.$$

If the coefficients c_{ik} belong to the field of real (complex) numbers, we call Q a **real (complex) quadratic form**. Let V be an n-dimensional vector space over K. For a vector x in V whose coordinates are $x_1, \ldots, x_n$, we put $Q(x) = Q(x_1, \ldots, x_n)$. This gives rise to a mapping $x \rightarrow Q(x)$ of V into K. Such a mapping satisfies the following two conditions: (i) $Q(ax) = a^2 Q(x)$ $(a \in K)$; and (ii) $Q(x+y) - Q(x) - Q(y) = B(x, y)$ is a †symmetric bilinear form on V. Conversely, if a mapping $Q: V \rightarrow K$ satisfies these two conditions, then Q must come from a quadratic form (→ 256 Linear Spaces). $B(x, y)$ is called the **symmetric bilinear form associated with** Q.

We assume that the †characteristic of K is not 2. Putting $a_{ik} = a_{ki} = c_{ik}/2$ $(i<k)$, $a_{ii} = c_{ii}$ $(i = 1, \ldots, n)$, we have $Q(x) = \sum_{i,k=1}^{n} a_{ik} x_i x_k$. The matrix $A = (a_{ik})$ is the **matrix of the quadratic form** Q, and the determinant $|A|$ is the **discriminant** of Q, denoted by $\Delta(Q)$. (Sometimes, instead of $|A|$, we call $(-1)^{n(n-1)/2} 2^n |A|$, the discriminant of Q.) The rank of the matrix A is called the **rank** of Q. Using the notation for the †inner product of vectors, we can write $Q(x) = (x, Ax) = {}^t x A x$; $2^{-1} B(x, y) = (x, Ay) = {}^t x A y$. We say that Q is **nondegenerate** if $B(x, y)$ is nondegenerate (i.e., if $|A| \neq 0$).

Consider a linear substitution $x_i = \sum_{j=1}^{m} p_{ij} x'_j$ (i.e., $x = Px'$, with an $n \times m$ matrix P). Then we get a new quadratic form $Q'(x')$ with the matrix ${}^t PAP$. If each p_{ij} belongs to the field K (to a subring R of K that contains the unit element of K), we say that Q represents Q' over K (resp. R). A basic problem in the theory of quadratic forms is to determine the exact conditions under which a given quadratic form Q represents another quadratic form Q'. The problem of representing numbers by a quadratic form (representation problem) is the particular case corresponding to $m = 1$. Any quadratic form Q represents 0 by taking the zero matrix as P. Hence, by the expression "Q represents 0 over K," we usually mean the nontrivial representation of 0 over K by Q, i.e., $Q(x) = 0$ for some nonzero vector x. If Q is nondegenerate and represents 0, then it represents any element of K^*. Given an element μ in K^*, we consider the quadratic form Q' defined by $Q'(x_1, \ldots, x_{n+1}) = Q(x) - \mu x_{n+1}^2$. Then Q represents μ if and only if Q' represents 0.

Another important special case is that of $n = m$. Then the discriminant of Q' is given by $|P|^2 |A|$. In particular, if $|P| \neq 0$ ($|P|$ is an invertible element of R), we say that Q is **equivalent** to Q' over K (resp. R). This gives rise to an equivalence relation. Equivalent forms have the same rank. On the other hand, if the rank of Q is r, then Q is equivalent to a form $\sum_{i=1}^{r} a_i x_i^2$ over K $(a_i \neq 0, i = 1, \ldots, r)$. Generally, for elements a and b in $K^* = K - \{0\}$, we write $a \sim b$ if $a \cdot b^{-1} \in (K^*)^2$. Then if Q is equivalent to Q', we have $\Delta(Q) \sim \Delta(Q')$.

When we specify a field K, we assume that the coefficients of the quadratic forms and the coordinates of linear transformations are all contained in the field K. In particular, the equivalence of the forms is equivalence over K.

B. Complex Quadratic Forms

If K is the field of complex numbers $\mathbf{C}$, then a form of rank r is equivalent to the form $\sum_{i=1}^{r} x_i^2$; hence over $\mathbf{C}$ two forms of the same dimension are equivalent if and only if they are of the same rank.

C. Real Quadratic Forms

Now let K be the field of real numbers $\mathbf{R}$. If Q is of rank r, then it is equivalent to the form $\sum_{i=1}^{p} x_i^2 - \sum_{j=1}^{q} x_{p+j}^2$ $(p+q=r)$. Here, p and q are uniquely determined by Q (**Sylvester's law of inertia**). We call (p, q) the **signature** of Q. Two quadratic forms of the same dimension

are equivalent if and only if they have the same signature. A quadratic form with the signature $(n, 0)$ (resp. $(0, n)$) is called a **positive (negative) definite quadratic form.** Q is called a **definite quadratic form** if it is either positive or negative definite; otherwise it is an **indefinite quadratic form.** Each of the following conditions is necessary and sufficient for a form Q to be positive definite: (i) for any nonzero real vector x we have $Q(x)>0$; (ii) all the †principal minors of the matrix of Q are positive. Q is negative definite if and only if $-Q$ is positive definite. A form with n variables is called a **positive (negative) semidefinite quadratic form** if its signature is $(r, 0)$ (resp. $(0, r)$), where $1 \leqslant r < n$.

A linear transformation $x' \rightarrow x = Px'$ that leaves invariant the unit form $\sum_{i=1}^{n} x_i^2$ is an **orthogonal transformation.** Then P is an †orthogonal matrix. Any quadratic form can be transformed into a diagonal form $\sum_{i=1}^{n} a_i x_i^2$ via an orthogonal transformation. Here $a_1, \ldots, a_n$ are the †eigenvalues of the matrix of the form. Two forms Q and Q' are equivalent with respect to an orthogonal transformation if and only if the corresponding matrices have the same eigenvalues.

D. Quadratic Forms over Finite Fields and p-adic Number Fields

Let Q and Q' be nondegenerate quadratic forms over the †finite field $\mathbf{F}_q$. They are equivalent if and only if they have the same rank and $\Delta(Q) \sim \Delta(Q')$. Moreover, if the rank of Q is not less than 3, then Q represents 0.

Next, suppose that Q and Q' are nondegenerate quadratic forms over the †p-adic number field K. They are equivalent if and only if they have the same rank, $\Delta(Q) \sim \Delta(Q')$, and they have the same **Minkowski-Hasse character** χ, where χ is defined as follows: Let $C(Q)$ be the †Clifford algebra of Q, and let $C^*(Q)$ denote $C(Q)$ if n is even and $C^+(Q)$ if n is odd. Then $\chi(Q) = 1$ or -1 according as $C^*(Q) \cong M_t(K)$ or $M_t(K) \otimes D(K)$. (Here M_t is the total matrix algebra of degree t over K and $D(K)$ is the unique †quaternion algebra over K.) Also, if Q has rank not less than 5, then Q represents 0 in K.

E. Quadratic Forms over a General Field K

The following facts are valid on any field K whose characteristic is not 2. Given a quadratic form Q_1 with variables $x_1, \ldots, x_n$ and another form Q_2 with variables $x_{n+1}, \ldots, x_{n+m}$, we get a new quadratic form $Q_1 \oplus Q_2$ or $Q_1 + Q_2$ defined by $Q_1 \oplus Q_2(x_1, \ldots, x_{n+m}) = Q_1(x_1, \ldots, x_n) + Q_2(x_{n+1}, \ldots, x_{n+m})$. $Q_1 \oplus Q_2$ is called the **direct sum** of Q_1 and Q_2. The matrix of $Q_1 \oplus Q_2$ is the direct sum of the matrices of Q_1 and Q_2. If Q_1 and Q_1' are equivalent and $Q_1 \oplus Q_2$ and $Q_1' \oplus Q_2'$ are also equivalent, then Q_2 and Q_2' are equivalent (**Witt's theorem**).

The quadratic form $x_1x_2 + x_3x_4 + \ldots + x_{2r-1}x_{2r}$ is called the **kernel form** and is denoted by N_r. Any nondegenerate quadratic form $Q(x_1, \ldots, x_n)$ is equivalent to the direct sum of a kernel form $N_r(x_1, \ldots, x_{2r})$ and a form $Q_0(x_{2r+1}, \ldots, x_n)$, where if $Q_0 \neq 0$, i.e., $n > 2r$, we have $Q_0(x_{2r+1}, \ldots, x_n) = 0$ only if $x_{2r+1} = \ldots = x_n = 0$. N_r and Q_0 are uniquely determined by Q up to equivalence. The decomposition $N_r \oplus Q_0$ is called the **Witt decomposition** of Q (E. Witt [5]). The number r is called the **index** of Q. An element x in V is said to be **singular** with respect to Q if $Q(x) = 0$. A subspace W of V is said to be **totally singular** if all the elements in W are singular. Let B be the symmetric bilinear form associated with Q. Then x is singular with respect to Q if and only if $B(x, x) = 0$ (characteristic of $K \neq 2$). We say that x is **isotropic** if $B(x, x) = 0$. Thus a subspace W is totally singular if and only if it is **totally isotropic** (i.e., $B(x, y) = 0$ for all $x, y \in W$). The index r of Q is the dimension of a maximal totally singular subspace of V. In particular, if $K = \mathbf{R}$ and (p, q) is the signature of Q, then the index $r = \min(p, q)$. Here we must be careful, since some authors call the number $p - q$ or p or q the index of Q. To make the distinction clear, we also call our r the **index of total isotropy**, and the number $p - q$ the **index of inertia.**

Necessary and sufficient conditions for a nondegenerate Q to be a kernel form are: $n =$ the rank of $Q \equiv 0 \pmod 2$ when $K = \mathbf{C}$; $n \equiv 0 \pmod 2$ and $p - q = 0$ when $K = \mathbf{R}$; $n \equiv 0 \pmod 2$ and $\Delta(Q) \sim 1$ when $K = \mathbf{F}_q$; $n \equiv 0 \pmod 2$, $\Delta(Q) \sim 1$, and $\chi(Q) = 1$ when K is a p-adic number field.

Let $N_r \oplus Q_0$, $N_s \oplus Q_0'$ be Witt decompositions of Q and Q', respectively. We say that Q and Q' belong to the same **type** if Q_0 and Q_0' are equivalent and denote the set of types of nondegenerate quadratic forms over K by $\mathbf{W}$. We define the sum of the types of Q and Q' as the type of $Q \oplus Q'$, and this gives $\mathbf{W}$ the structure of a commutative group. The type of a kernel form is the identity element of this group $\mathbf{W}$, called the **Witt group.** The structure of the Witt group depends on K. If $K = \mathbf{C}$, then $\mathbf{W} \cong \mathbf{Z}/2\mathbf{Z}$; if $K = \mathbf{R}$, then $\mathbf{W} \cong \mathbf{Z}$; if K is a †local field with a †non-Archimedean valuation, then $\mathbf{W}$ is a finite group; if $K = \mathbf{F}_q$, then $\mathbf{W} \cong (\mathbf{Z}/2\mathbf{Z}) + (\mathbf{Z}/2\mathbf{Z})$ if $q \equiv 1 \pmod 4$, $\mathbf{W} \cong \mathbf{Z}/4\mathbf{Z}$ if $q \equiv 3 \pmod 4$, and $\mathbf{W} \cong \mathbf{Z}/2\mathbf{Z}$ if q is a power of 2.

F. Hermitian Forms

An expression $H(x)=\sum_{i,k=1}^{n} a_{ik}\bar{x}_i x_k$ is called a **Hermitian form** if $a_{ik}\in\mathbf{C}$, $a_{ik}=\bar{a}_{ki}$. (Here $\bar{a}_{ik}$, $\bar{x}_i$ are the complex conjugates of a_{ik}, x_i, respectively.) The value $H(x)$ is a real number. As for quadratic forms, we define the notions of the matrix of H and the discriminant, rank, and †sesquilinear form associated with H. The matrix A of H is a †Hermitian matrix whose principal minors are real numbers. If we apply a linear transformation $P(x')=x$, we obtain a Hermitian form with respect to x' whose matrix is given by ${}^t\bar{P}AP$. Any Hermitian form is equivalent to a form $\sum_{i=1}^{p}\bar{x}_i x_i - \sum_{j=1}^{q}\bar{x}_{p+j}x_{p+j}$; (p,q) is called the **signature** of H. We define the notions of **positive definite**, **negative definite**, and **indefinite Hermitian forms** as we did for quadratic forms over the field of real numbers. Each of the following conditions is necessary and sufficient for H to be positive definite: (i) $H(x)>0$ for any nonzero complex vector x; (ii) all the principal minors of the matrix of H are positive. The definition of a **semidefinite Hermitian form** is given in the same manner as for a quadratic form.

A linear transformation that leaves the Hermitian form $\sum_{i=1}^{n}\bar{x}_i x_i$ invariant is called a **unitary transformation**, and its matrix is a †unitary matrix. Any Hermitian form can be transformed via a unitary transformation into a diagonal form $\sum_{i=1}^{n} a_i\bar{x}_i x_i$, where $a_1,\ldots,a_n$ are the eigenvalues of the matrix of the Hermitian form.

The notion of Hermitian forms can be generalized as follows: Suppose that K is a †division ring with an **involution** $a\to\bar{a}$ ($a\in K$) (i.e., $a\to\bar{a}$ is a linear mapping of K onto itself such that $\bar{\bar{a}}=a$, $\overline{ab}=\bar{b}\bar{a}$). Then a Hermitian form H over K is defined by

$$H(x)=\sum_{i,k=1}^{n}\bar{x}_i a_{ik}x_k,$$

where $x_i\in K$, $a_{ik}\in K$, $a_{ik}=\bar{a}_{ki}$. In particular, if we have, for any given vector x whose coordinates belong to K, an element a in K such that $H(x)=a+\bar{a}$, then we have a Witt decomposition for H. Two examples of such K having involutions that differ from the identity mapping are a separable quadratic extension K of a field L and a †quaternion algebra K over a field L.

G. Quadratic Forms over Algebraic Number Fields

Let K be an algebraic number field of finite degree, $\mathfrak{p}$ be an †Archimedean or non-Archimedean †prime divisor of K, $K_\mathfrak{p}$ be the $\mathfrak{p}$-completion of K, and Q and Q' be nondegenerate quadratic forms over K. Then Q represents Q' over K if and only if Q represents Q' over $K_\mathfrak{p}$ for all $\mathfrak{p}$, and Q represents 0 in K (i.e., there exists a nonzero vector x whose coordinates belong to K such that $Q(x)=0$) if and only if it represents 0 in $K_\mathfrak{p}$ for all $\mathfrak{p}$ [3,4]. In particular, Q and Q' are equivalent over K if and only if they are equivalent over $K_\mathfrak{p}$ for all $\mathfrak{p}$. Hence the invariants with respect to equivalence over K of a nondegenerate quadratic form Q over K are $n=$ the rank of Q, Δ $=$ the discriminant of Q, the Minkowski-Hasse character $\chi_\mathfrak{p}$ for †prime divisors $\mathfrak{p}$ of K, and the index of inertia j_λ of Q over $K_{\mathfrak{p}_{\infty,\lambda}}$ for each †real infinite prime divisor $\mathfrak{p}_{\infty,\lambda}$ ($\lambda=1,\ldots,r_1$) of K. Here the following properties hold: (i) $\chi_\mathfrak{p}=1$ for all but finitely many $\mathfrak{p}$; (ii) $\prod_\mathfrak{p}\chi_\mathfrak{p}=1$ (this is equivalent to the †product formula of norm-residue symbols); (iii) $\Delta\sim(-1)^{(n^2+j_\lambda)/2}$ in $K_{\mathfrak{p}_{\infty,\lambda}}$; and (iv) $\chi_{\mathfrak{p}_{\infty,\lambda}}=1$ if $j_\lambda\equiv 0, 1, 2, 7 \pmod 8$, $=-1$ if $j_\lambda\equiv 3, 4, 5, 6 \pmod 8$ [3,4]. Conversely, if the system $\{n,\chi_\mathfrak{p},\chi_{\mathfrak{p}_{\infty,\lambda}},j_\lambda,\Delta\}$ satisfies conditions (i)–(iv), then it is the set of invariants of a quadratic form over K (**Minkowski-Hasse theorem**). In general, if a property concerning K holds if and only if it holds for all $K_\mathfrak{p}$, we say that the **Hasse principle** holds for the property.

H. Class and Genus of a Quadratic Form

Let K be an algebraic number field of finite degree. Quadratic forms Q and Q' over K are said to be of the same **class** if they are equivalent over the †principal order $\mathfrak{o}$ in K. On the other hand, Q and Q' are said to be of the same **genus** if (i) they are equivalent over the principle order $\mathfrak{o}_\mathfrak{p}$ in $K_\mathfrak{p}$ for all non-Archimedean prime divisors $\mathfrak{p}$ of K and (ii) they are equivalent over $K_\mathfrak{p}$ for all the Archimedean prime divisors $\mathfrak{p}$ of K. A genus is decomposed into a finite number of classes. For example, if K is the field of rational numbers, the number of classes in the genus of $\sum_{i=1}^{m}x_i^2$ is 1 for $m\leqslant 8$, while it is $\geqslant 2$ for $m>8$.

I. Reduction of Real Quadratic Forms

Let A be an $m\times m$ matrix and X an $m\times n$ matrix. We put $A[X]={}^tXAX$. Then we can write $Q(x)=S[x]={}^txSx$, where S is the matrix of the quadratic form Q. In this section we put $K=\mathbf{R}$ and define two forms to belong to the same **class** if they are equivalent over the ring of rational integers. We identify the form Q with its matrix $S=(s_{ij})$. Let S be a positive definite form in m variables. Then S is said to be a **reduced quadratic form** if $S[\mathfrak{g}_k]\geqslant s_{kk}$ and $s_{l\,l+1}\geqslant 0$ ($1\leqslant k\leqslant m$, $1\leqslant l\leqslant m-1$), where $\mathfrak{g}_k$ is an

arbitrary vector whose coordinates $g_1, \dots, g_m$ are integers such that $(g_k, \dots, g_m) = 1$. Any class of positive definite quadratic forms contains at least one (and generally only one) reduced form. For a reduced form $R = (r_{kl})$, the following inequalities hold: $0 < r_{11} \leqslant r_{22} \leqslant \dots \leqslant r_{mm}$; $\pm 2r_{kl} \leqslant r_{ll}$, $k < l$; $r_{11} r_{22} \dots r_{mm} < c(m)|R|$, where $c(m)$ depends only on m. The set of all symmetric matrices of degree m forms a linear space of dimension $m(m+1)/2$ in which the subset $\mathfrak{P}$ formed by the positive definite symmetric matrices is a convex open subset. Moreover, the subset $\mathfrak{R}$ formed by all reduced positive definite symmetric matrices is a convex cone whose boundary consists of finitely many hypersurfaces and whose vertex is the origin. Let S be an indefinite quadratic form whose signature is $(n, m-n)$. The set of positive definite quadratic forms H such that $S^{-1}[H] = S$ forms a variety of dimension $n(m-n)$, which is denoted by $H(S)$. We say that S is reduced if $H(S) \cap \mathfrak{R} \neq \varnothing$. Given a natural number D, there are only a finite number of definite or indefinite reduced quadratic forms with rational integral coefficients whose discriminant is $\pm D$. Hence the number of classes of quadratic forms with rational integral coefficients and discriminant $\pm D$ is also finite.

J. Units

Let S be a symmetric matrix of degree m with rational coordinates. Let $O(S)$ be the set of all real $m \times m$ matrices W for which $S[W] = S$, and $\Gamma(S)$ be the subset of $O(S)$ consisting of the integral matrices. An element of $\Gamma(S)$ is called a **unit** of S. $\Gamma(S)$ is a finite group if S is definite, but otherwise it is infinite (except for the case $m=2$, $-|S| = r^2$, with rational r). $O(S)$ is a †Lie group, and $\Gamma(S)$ is a discrete subgroup with a finite number of generators. The †homogeneous space $O(S)/\Gamma(S)$ is of finite measure with respect to a †Haar measure defined on the space.

K. Minkowski-Siegel-Tamagawa Theory

Let S and T be rational integral positive definite symmetric matrices of degree m and n, respectively $(m \geqslant n)$. Let $A(S, T)$ be the number of rational integral solutions for the equation $S[X] = T$, and $E(S)$ be the order of the group of units $\Gamma(S)$. We put

$$M(S, T) = \frac{A(S_1, T)}{E(S_1)} + \frac{A(S_2, T)}{E(S_2)} + \dots,$$

$$M(S) = \frac{1}{E(S_1)} + \frac{1}{E(S_2)} + \dots,$$

$$A_0(S, T) = \frac{M(S, T)}{M(S)},$$

where $S_1, S_2, \dots$ is a complete system of representatives of the classes in the genus of S. $M(S)$ is called the **measure of genus** of S. On the other hand, for a natural number q, we denote by $A_q(S, T)$ the number of the solutions of the congruence equation $S[X] \equiv T \pmod{q}$. If q is a prime power p^a, then the ratio $\varepsilon_{m,n} q^{-mn + n(n+1)/2} A_q(S, T)$ takes a constant value $\alpha_p(S, T)$ for sufficiently large a (where $\varepsilon_{m,n} = 1/2$ if $m = n \geqslant 2$; $= 1$ otherwise). Furthermore, let us consider a domain B in the Euclidean space of dimension $n(n+1)/2$ formed by the set of $n \times n$ symmetric matrices containing T, and let B_1 be the domain formed by the matrices X such that $S[X] \in B$. Then B_1 is a domain in the space of dimension mn formed by all $m \times n$ matrices. Let $\alpha_\infty(S, T)$ be the limit of the ratio $v(B_1)/v(B)$ of the volumes of B and B_1 as the domain B shrinks toward the point T. Then **Siegel's theorem** states: $\alpha_\infty(S, T) \prod_p \alpha_p(S, T) = \varepsilon A_0(S, T)$, where $\varepsilon = 2$ if $m = n+1$ or $m = n \geqslant 2$ and $\varepsilon = 1$ otherwise. The infinite product of the left-hand side of this equation does not converge absolutely if either $m = n = 2$ or $m = n+2$, and in those cases the order of the product $\prod_p$ is considered to be the natural order of the primes p.

A special case of Siegel's theorem was proved by H. Minkowski, but it was C. L. Siegel [6] who proved it in its general form. Except for a finite number of p, the numbers $\alpha_p(S, T)$ have been calculated. The explicit form of $\alpha_\infty(S, T)$ is also known. In particular, if we take the identity matrix $E^{(m)}$ of degree m as S, then the formula in Siegel's theorem is related to the problem of expressing natural numbers as sums of m squares. For $m = 2, 3, \dots, 8$, the genus of $E^{(m)}$ contains only one class. Hence, putting $n = 1$, $T = t$ ($=$a natural number), we obtain from Siegel's theorem the number of ways in which we can express t as the sum of m squares [6, pt. I]. Siegel's result was generalized by Siegel himself to the case where the form S is indefinite [6, pt. II] and where the coefficients of the forms are elements of an algebraic number field of finite degree [6, pt. III]. Also, regarding the number of possible ways to express a natural number t as a sum of m squares, the following formula was obtained by C. G. J. Jacobi for the case where $m = 4$, $n = 1$:

$$A(E^{(4)}, t) = 8\left(\sum_{d|t} d - \sum_{4d|t} d\right).$$

For the case $m = 3$, $n = 1$, it is known that if t is odd and $A(E^{(3)}, t) > 0$, then $t \not\equiv 7 \pmod{8}$ (for details → P. T. Batemann, *Trans. Amer. Math. Soc.*, 71 (1951)).

T. Tamagawa used the theory of †adelized algebraic groups and proved that the †Tamagawa number $\tau(SO(n, S))$ of the special ortho-

gonal group is 2. He also showed that from this fact, Siegel's theory in this section can be deduced (→ 13 Algebraic Groups P) [10].

L. Theta Series

Let $Q(x_1, \ldots, x_m)$ be a positive definite form with integral coefficients. For a complex number z, we put

$$F(z, Q) = \sum_{x_1, \ldots, x_m} \exp(2\pi i Q(x_1, \ldots, x_m)z),$$

where $x_1, \ldots, x_m$ run over all the integers. If $\operatorname{Im} z > 0$, the series converges and represents an †entire function of z. These series are called **theta series**. If we denote by $A(n)$ the number of integral solutions of the equation $Q(x_1, \ldots, x_m) = n$, we have

$$F(z, Q) = \sum_{n=0}^{\infty} A(n) e^{2\pi i n z}.$$

Moreover, if $m = 2k$, we have the following **transformation formula**:

$$F\left(\frac{az+b}{cz+d}, Q\right) = \varepsilon(d)(cz+d)^k F(z, Q),$$

where a, b, c, d are integers such that $ad - bc = 1$, $c \equiv 0 \pmod N$, N is a natural number determined by Q, and ε is a character mod N. In other words, $F(z, Q)$ is a †modular form with respect to the †congruence subgroup of level N. Using the theory of modular forms, E. Hecks showed that $A(n) = A_0(n) + O(n^{k/2})$, where $A_0(n)$ is a number-theoretic function of n determined by the genus of Q.

M. Binary Quadratic Forms with Integral Coefficients

Now put $m = 2$. Given a form $Q(x, y) = ax^2 + bxy + cy^2$, we put $D(Q) = b^2 - 4ac$ and call it the **discriminant** of Q (i.e., $D(Q) = -4\Delta(Q)$). Q is said to be **primitive** if $(a, b, c) = 1$. When $D(Q)$ is not a square, the theory concerning Q is closely related to the arithmetic theory of the †quadratic field $\mathbf{Q}(\sqrt{D}) = k$. Let d be the †discriminant of k and put $D = df^2$. When $f = 1$, there is a one-to-one correspondence between the †ideal classes of k and the classes of quadratic forms with discriminant d (when $D < 0$, we consider the classes of positive definite forms). The correspondence is given in the following manner: If $\mathfrak{a}$ is an ideal in k with a basis α_1, α_2, then the corresponding form is given by $Q(x, y) = N(\mathfrak{a})^{-1} N(\alpha_1 x + \alpha_2 y)$, where N is the absolute †norm. If $f > 1$, we must replace the ring of integers $\mathfrak{o}$ by the †order of the †conductor f. That is, if we consider the ring formed by the elements $x + fy\omega$ (x, y are rational integers; the meaning of ω is explained shortly), we again have a one-to-one correspondence between the classes of ideals of this order and the classes of quadratic forms [3]. We let $\omega = \sqrt{d}/2$ if $d \equiv 0 \pmod 4$; $= (1 + \sqrt{d})/2$ if $d \equiv 1 \pmod 4$. When $D > 0$, we can introduce the notion of proper equivalence as follows: Q and Q' are **properly equivalent** if the matrix of Q is transformed to the matrix of Q' by a linear transformation P whose determinant is 1. Then, in the correspondence for the case $f = 1$, if we take $\alpha_1 = r > 0$, $\alpha_2 = s + t\omega$, $t > 0$ ($r, s, t \in \mathbf{Q}$), then we get a relation between the classification of the forms and the classification in the finer sense of the ideals in k.

Suppose that $D > 0$ is not a square. Let (t, u) be an integral solution of †Pell's equation $t^2 - Du^2 = \pm 4$. Then the units of the form $Q(x, y) = ax^2 + bxy + cy^2$ with discriminant D are given by

$$\pm \begin{pmatrix} (t-bu)/2 & -cu \\ au & (t+bu)/2 \end{pmatrix}.$$

Let (t_0, u_0) be the smallest positive integral solution of $t^2 - Du^2 = 4$, put $\varepsilon_D = (t_0 + u_0\sqrt{D})/2$, and let h_D be the class number in the finer sense of the forms of discriminant D. Then the following formula holds (Dirichlet):

$$\frac{1}{\sqrt{D}} h_D \log \varepsilon_D = \sum_{n=1}^{\infty} \frac{1}{n}\left(\frac{D}{n}\right),$$

where (D/n) is the †Kronecker symbol (here $D = f^2 d$; we put $(D/n) = 0$ if $(f, n) \neq 1$, and $(D/n) = (d/n)$ if $(f, n) = 1$). For $D < 0$, the order w_D of the units is known: it is 6 if $D = -3$; 4 if $D = -4$; and 2 otherwise. We also have

$$\frac{2\pi}{\sqrt{|D|}} \frac{h_D}{w_D} = \sum_{n=1}^{\infty} \frac{1}{n}\left(\frac{D}{n}\right).$$

With respect to the numbers h_D and ε_D, little else is known.

References

[1] B. W. Jones, The arithmetic theory of quadratic forms, Wiley, 1950.
[2] N. Bourbaki, Eléments de mathématique, Algèbre, ch. 9, Actualités Sci. Ind., 1272a, Hermann, 1959.
[3] M. Eichler, Quadratische Formen und orthogonale Gruppen, Springer, 1952.
[4] O. T. O'Meara, Introduction to quadratic forms, Springer, 1963.
[5] E. Witt, Theorie der quadratischen Formen in beliebigen Körpern, J. Reine Angew. Math., 176 (1937), 31–44.
[6] C. L. Siegel, Über die analytische Theorie der quadratischen Formen I, II, III, Ann. Math., (2) 36 (1935), 527–606; 37 (1936), 230–263; 38 (1937), 212–291 (Gesammelte Abhand-

lungen, Springer, 1966, vol. 1, 326–405, 410–443, 469–548).
[7] C. L. Siegel, On the theory of indefinite quadratic forms, Ann. Math., (2) 45 (1944), 577–622. (Gesammelte Abhandlungen, Springer, 1966, vol. 2, 421–466.)
[8] E. Hecke, Analytische Arithmetik der positiven quadratischen Formen, Kgl. Danske Videnskab. Selskab, Mat.-Fys. Medd., XIII, 12 (1940). (Mathematische Werke, Vandenhoeck & Ruprecht, 1959, 789–918.)
[9] G. L. Watson, Integral quadratic forms, Cambridge Univ. Press, 1960.
[10] T. Tamagawa, Adèles, Amer. Math. Soc. Proc. Symp. Pure Math., 9 (1966), 113–121.
For binary quadratic forms,
[11] P. G. L. Dirichlet, Vorlesungen über Zahlentheorie, Vieweg, fourth edition, 1894 (Chelsea, 1969).

349 (XIX.3) Quadratic Programming

A. Problems

A **quadratic programming problem** is a special type of mathematical programming (→ 264 Mathematical Programming) where the objective function is quadratic while the constraints are linear. A typical formulation of the problem is as follows.

(Q) Maximize $\mathbf{z}=\mathbf{c}'\mathbf{x}-\frac{1}{2}\mathbf{x}'D\mathbf{x}$ under the condition $A\mathbf{x}\leqslant\mathbf{b}$ and $\mathbf{x}\geqslant 0$, $\mathbf{x}\in\mathbf{R}^n$.

Let the Lagrangian form for this problem be

$$\varphi(\mathbf{x},\boldsymbol{\lambda})=\mathbf{c}'\mathbf{x}-\tfrac{1}{2}\mathbf{x}'D\mathbf{x}+\boldsymbol{\lambda}'(\mathbf{b}-A\mathbf{x}).$$

Then, from the general properties of Lagrangian forms, the following theorem can be proved.

Theorem: If $\mathbf{x}=\mathbf{x}^*$ is an optimal solution of the problem (Q), there exists a vector λ^* satisfying the conditions

$$-D\mathbf{x}^*+\mathbf{c}\leqslant\boldsymbol{\lambda}^*,\qquad \boldsymbol{\lambda}^*\geqslant 0;$$

$$\mathbf{b}'\boldsymbol{\lambda}^*=(\mathbf{c}-D\mathbf{x}^*)'\mathbf{x}^*.$$

Moreover, if the matrix D is nonnegative definite, the above conditions are also sufficient for $\mathbf{x}=\mathbf{x}^*$ to be optimal. The second condition can be shown to be equivalent to

$$\boldsymbol{\lambda}^{*\prime}(\mathbf{b}-A\mathbf{x}^*)=0 \ \text{ and } \ \mathbf{x}^{*\prime}(A'\boldsymbol{\lambda}^*+D\mathbf{x}^*-\mathbf{c})=\mathbf{0}.$$

By introducing the slack vectors $\mathbf{u}\geqslant\mathbf{0}$ and $\mathbf{v}\geqslant\mathbf{0}$, the conditions can be expressed as

$$\mathbf{x}\geqslant\mathbf{0},\quad \mathbf{y}\geqslant\mathbf{0},\quad \mathbf{u}\geqslant\mathbf{0},\quad \mathbf{v}\geqslant\mathbf{0};$$

$$\text{(C)}\quad A\mathbf{x}+\mathbf{u}=\mathbf{b},\quad A'\mathbf{y}+D\mathbf{x}-\mathbf{v}=\mathbf{c};$$

$$\mathbf{y}'\mathbf{u}=0,\quad \mathbf{x}'\mathbf{v}=0.$$

When D is nonnegative definite, any feasible solution of the above system of equalities gives an optimal solution of the primary problem (Q), and when D is positive definite, the solution is unique. When D is not nonnegative definite, the optimal solution of (Q), if it exists, is one of the feasible solutions of (C). The last line of (C) implies that the solution must be a basic solution of the linear system of equalities, and it also restricts the possible combinations of the basic variables. Since there exist only a finite number of possible combinations of the basic variables, the quadratic programming problem can be solved in a finite number of steps, if it has an optimal solution.

B. Duality

The **dual problem** of (Q) is the following.

(QD) Minimize $w=\mathbf{b}'\mathbf{y}+\frac{1}{2}\mathbf{x}'D\mathbf{x}$ under the condition $A'\mathbf{y}+D\mathbf{x}\geqslant\mathbf{c}$ and $\mathbf{x}\geqslant 0$, $\mathbf{y}\geqslant 0$.

If D is nonnegative definite, the following theorem holds.

Theorem: If the primary problem (Q) has a solution $\mathbf{x}=\mathbf{x}^*$, then the dual problem has a solution $\mathbf{x}=\mathbf{x}^*$ and $\mathbf{y}=\mathbf{y}^*$, and $\max z=\min w$.

A more general form of the quadratic programming problem can be given as follows.

(Q) Maximize $z=\mathbf{c}'\mathbf{x}-\frac{1}{2}\mathbf{x}'D\mathbf{x}$ under the condition $\mathbf{x}\in V$ and $\mathbf{b}-A\mathbf{x}\in W$, where V and W are closed convex cones in $\mathbf{R}^n$ and $\mathbf{R}^m$, respectively.

Then the dual problem is expressed as follows.

(QD) Minimize $w=\mathbf{b}'\mathbf{y}-\frac{1}{2}\mathbf{x}'D\mathbf{x}$ under the condition $\mathbf{x}\in V$, $\mathbf{y}\in W^*$ and $A'\mathbf{y}+D\mathbf{x}-\mathbf{c}\in V^*$, where V^* and W^* are the dual cones of V and W.

The above theorem holds for both (Q) and (QD).

C. Algorithms

Various algorithms have been proposed for quadratic programming [1, 2, 4], most of which are based on condition (C). Wolfe [4] proposed a method based on the simplex method for linear programming. If we introduce the artificial vectors ξ and η, we can find a feasible solution of (C) by solving the following linear programming problem.

(LQ) Maximize $z=-1'\boldsymbol{\xi}-1'\boldsymbol{\eta}$ under the condition that

$$A\mathbf{x}+\mathbf{u}-\boldsymbol{\xi}=\mathbf{b},\quad A'\mathbf{y}+D\mathbf{x}-\mathbf{v}+\boldsymbol{\eta}=\mathbf{c};$$

$$\mathbf{x}\geqslant\mathbf{0},\quad \mathbf{y}\geqslant\mathbf{0},\quad \mathbf{u}\geqslant\mathbf{0},\quad \mathbf{v}\geqslant\mathbf{0},\quad \boldsymbol{\xi}\geqslant\mathbf{0},\quad \boldsymbol{\eta}\geqslant\mathbf{0};$$

$$\mathbf{y}'\mathbf{u}=0,\quad \mathbf{x}'\mathbf{v}=0.$$

(LQ) can be solved by applying the simplex algorithm with the only modification being

that the last line of the condition restricts the possible changes in the basic variables. When D is positive definite, we can always obtain a solution if there is a feasible solution of the original problem, and Wolfe proposed a modification of the foregoing algorithm for the case when D is nonnegative definite that tells whether or not it has an optimal solution, and gives it if it has one. Some other algorithms are also effective when D is positive or nonnegative definite, but when D is not nonnegative definite, no simple effective method has been found to reach the optimal solution even when its existence has been established.

References

[1] E. M. L. Beale, On quadratic programming, Naval Res. Logistic Quart., 6 (1959), 227–243.
[2] C. Hildreth, A quadratic programming procedure, Naval Res. Logistic Quart., 4 (1957), 79–85.
[3] W. S. Dorn, Duality in quadratic programming, Quart. Appl. Mat., 18 (1960), 155–162.
[4] P. Wolfe, The simplex method for quadratic programming, Econometrica, 27 (1959), 382–398.

350 (VI.10) Quadric Surfaces

A. Introduction

A subset F of a 3-dimensional Euclidean space E^3 is called a **quadric surface (surface of the second order** or simply **quadric)** if F is the set of zeros of a quadratic equation $G(x, y, z)=0$, where the coefficients of G are real numbers. The equation $G(x, y, z)=0$ is written as

$$ax^2+by^2+cx^2+d+2fyz+2gzx+2hxy+2f'x + 2g'y+2h'z=0. \tag{1}$$

In general, a straight line intersects a quadric surface at two points. If it intersects the surface at more than two points, then the whole straight line lies on the surface. Suppose that we are given a quadric surface and a point O. Suppose further that we are given a straight line passing through the point O and intersecting the quadric surface at two points A and A'. If $AO=OA'$ for all such straight lines, then the point O is called the **center** of the quadric surface.

B. Classification

The subset defined by equation (1) may be empty; for example, $x^2+y^2+z^2+1=0$. In this article, we consider only quadric surfaces that are not the empty set. When a quadric surface F without †singular points has a center or centers, we say that F is **central.**

If we choose a suitable rectangular coordinate system, the equation of a central quadric surface is written in one of the following forms:

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1, \tag{2}$$

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=1, \tag{3}$$

$$-\frac{x^2}{a^2}-\frac{y^2}{b^2}+\frac{z^2}{c^2}=1, \tag{4}$$

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1, \tag{5}$$

$$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1, \tag{6}$$

$$\frac{x^2}{a^2}=1. \tag{7}$$

When the equation takes the form (2), (3), (4), (5), or (6), we call the quadric surface an **ellipsoid, hyperboloid of one sheet, hyperboloid of two sheets, elliptic cylinder** (or **elliptic cylindrical surface**), or **hyperbolic cylinder (hyperbolic cylindrical surface)**, respectively. When the equation takes the form (7), the surface coincides with a pair of parallel planes. If $a=b$ in (2), (3), (4), or (5), the surface is a †surface of revolution with the z-axis as the axis of revolution. In this case, we call the surface an **ellipsoid of revolution, hyperboloid of revolution of one sheet, hyperboloid of revolution of two sheets,** or **circular cylinder** (or **circular cylindrical surface**), respectively. If $a=b=c$ for an ellipsoid of revolution, then the surface is a sphere with radius a.

If we choose a suitable rectangular coordinate system, the equation of a noncentral surface of the second order is written in one of the following forms:

$$2z=\frac{x^2}{a^2}+\frac{y^2}{b^2}, \tag{8}$$

$$2z=\frac{x^2}{a^2}-\frac{y^2}{b^2}, \tag{9}$$

$$2z=\frac{x^2}{a^2}. \tag{10}$$

When the equation takes the form (8), (9), or (10), we call the surface an **elliptic paraboloid, hyperbolic paraboloid,** or **parabolic cylinder** (or **parabolic cylindrical surface**), respectively. If a

$= b$ in (8), the surface is called an **elliptic paraboloid of revolution.**

Among these, (2), (3), (4), (8), and (9) are sometimes called **proper quadric surfaces**, and the others **degenerate quadric surfaces**.

Equations (2)–(10) are called the **canonical forms of the equations** of these surfaces (a, b, c in canonical forms should not be confused with a, b, c in (1)). The planes $x=0$, $y=0$, and $z=0$ in surfaces (2), (3), and (4) and the planes $x=0$ and $y=0$ in surfaces (8) and (9) are called **principal planes** of the respective surfaces; and lines of intersection of principal planes are called **principal axes**. For a surface of revolution, positions of principal planes and principal axes are indeterminate. We call a, b, c in equations of canonical form the lengths of the principal axes, or simply the principal axes. If F is a hyperboloid of one sheet or a hyperbolic paraboloid, there are two systems of straight lines lying on F; two straight lines belonging to the same system never meet (and are not parallel), and two straight lines belonging to different systems always meet (or are parallel). If F satisfies (3), these systems of straight lines are given by

$$\begin{cases} \dfrac{x}{a}-\dfrac{z}{c}=\lambda\left(1-\dfrac{y}{b}\right) \\ \dfrac{x}{a}+\dfrac{z}{c}=\dfrac{1}{\lambda}\left(1+\dfrac{y}{b}\right) \end{cases} \quad \begin{cases} \dfrac{x}{a}-\dfrac{z}{c}=\mu\left(1+\dfrac{y}{b}\right) \\ \dfrac{x}{a}+\dfrac{z}{c}=\dfrac{1}{\mu}\left(1-\dfrac{y}{b}\right). \end{cases}$$

If F satisfies (9), then two such systems are given by

$$\begin{cases} \dfrac{x}{a}-\dfrac{y}{b}=\lambda \\ \dfrac{x}{a}+\dfrac{y}{b}=\dfrac{2z}{\lambda} \end{cases} \quad \begin{cases} \dfrac{x}{a}+\dfrac{y}{b}=\mu \\ \dfrac{x}{a}-\dfrac{y}{b}=\dfrac{2z}{\mu} \end{cases}$$

(λ and μ are parameters). We call these straight lines **generating lines** of the respective surfaces. A hyperboloid of one sheet and a hyperbolic paraboloid are †ruled surfaces described by these generating lines.

When a quadric surface has singular points, they are double points. The set of double points of a quadric surface F is either a single point O, a straight line l, or a plane π. In the second case, F consists of two planes passing through l or l itself, and in the third case, F coincides with π. In the first case, we say that F is a **quadric conical surface** (or **quadric cone**) with vertex O. Its equation is written in the form $Ax^2+By^2+Cz^2=0$ ($ABC\neq0$). When A, B, C are of the same sign, F consists of only one point O. Otherwise, we can assume that A, $B>0$, $C=-1$. In this case, if $A=B$, F is called a **right circular cone**, and if $A\neq B$, F is called an **oblique circular cone**.

Given hyperboloids (3) and (4), we call the quadric cones

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=0 \qquad (3')$$

and

$$-\frac{x^2}{a^2}-\frac{y^2}{b^2}+\frac{z^2}{c^2}=0 \qquad (4')$$

asymptotic cones of (3) and (4), respectively.

C. Poles and Polar Planes

Suppose that we are given a straight line S passing through a fixed point P not contained in a quadric surface F, and S intersects the surface at two points X, Y. The locus of the point Q that is the †harmonic conjugate of P with respect to X and Y is a plane. We call this plane π the **polar plane** of P with respect to the quadric surface F, and P the **pole** of the plane π. If the polar plane of a point P contains a point Q, then the polar plane of Q contains P. In this case, we say that the two points P and Q are **conjugate** to each other with respect to the quadric surface. When the point P is on the quadric surface, the tangent plane at P is regarded as the polar plane of P. If the polar plane (with respect to a quadric surface) of each vertex of a tetrahedron is the face corresponding to that vertex, we call this tetrahedron a **self-polar tetrahedron**. If the polar planes (with respect to a quadric surface) of four vertices of a tetrahedron A are four faces of a tetrahedron B, the same property holds when we interchange A and B. We say that such tetrahedrons are **polar tetrahedrons** with respect to the quadric surface. Suppose that we are given a quadric surface and two planes. If the pole (with respect to the quadric surface) of one plane is on the other plane, these two planes are said to be **conjugate** with respect to the surface.

When we are given two †pencils of planes in †projective correspondence, the locus of lines of intersection of two corresponding planes is generally a hyperboloid of one sheet or a hyperbolic paraboloid. In particular, if the axes of these pencils of planes intersect, the locus is a quadric conical surface, and if the axes are parallel, the locus is a quadric cylindrical surface (i.e., an elliptic or hyperbolic cylinder). When there exists a projective correspondence between two straight lines not on a plane, the locus of lines joining corresponding points is a quadric surface (M. Chasles).

D. Surfaces of the Second Class

A surface F in $\mathbf{E}^3$ is called a **surface of the second class** if it admits two tangent planes

passing through an arbitrary straight line L provided that $F \cap L = \varnothing$. This surface can be represented as the set of zeros of a homogeneous equation of the second order in †plane coordinates u_1, u_2, u_3, u_4. It is possible that a surface of the second class degenerates into a conic or two points. In general, quadrics are surfaces of the second class, and vice versa.

As in the case of quadrics, we can define poles, polar planes, and polar tetrahedrons with reference to surfaces of the second class. Four straight lines joining corresponding vertices of two tetrahedrons polar with respect to a surface of the second class are on the same quadric. We say that two such tetrahedrons are in **hyperboloid position**.

E. Confocal Quadrics

A family of central quadrics represented by the following equations is called a **family of confocal quadrics**:

$$\frac{x^2}{a+k}+\frac{y^2}{b+k}+\frac{z^2}{c+k}=1, \qquad a>b>c>0, \tag{11}$$

where k is a parameter. For a quadric belonging to this family, any point on the ellipse $x^2/(a-c)+y^2/(b-c)=1$, $z=0$ or the hyperbola $x^2/(a-b)-z^2/(b-c)=1$, $y=0$ is called a **focus**. This ellipse and hyperbola are called **focal conics** of the quadric.

Given an ellipsoid F and a point $X(x,y,z)$ not contained in the principal plane, we can draw three quadrics F', F'', F''' passing through X and confocal with F. These surfaces F', F'', F''' intersect each other and are mutually perpendicular. One of them is an ellipsoid, another one a hyperboloid of one sheet, and the third a hyperboloid of two sheets. Let k_1, k_2, k_3 be the values of the parameter k in (11) corresponding to these three surfaces. Then the coordinates x, y, z of the point X are given by

$$x=\sqrt{\frac{(a+k_1)(a+k_2)(a+k_3)}{(b-a)(c-a)}},$$

$$y=\sqrt{\frac{(b+k_1)(b+k_2)(b+k_3)}{(a-b)(c-b)}},$$

$$z=\sqrt{\frac{(c+k_1)(c+k_2)(c+k_3)}{(a-c)(b-c)}}.$$

We call k_1, k_2, k_3 the **elliptic coordinates** of the point X.

Two points (x,y,z), (x',y',z') are called **corresponding points** if they belong to confocal quadrics of the same kind,

$$\frac{x^2}{a}+\frac{y^2}{b}+\frac{z^2}{c}=1,$$

$$\frac{(x')^2}{a+k}+\frac{(y')^2}{b+k}+\frac{(z')^2}{c+k}=1,$$

and satisfy

$$\frac{x}{\sqrt{a}}=\frac{x'}{\sqrt{a+k}}, \qquad \frac{y}{\sqrt{b}}=\frac{y'}{\sqrt{b+k}},$$

$$\frac{z}{\sqrt{c}}=\frac{z'}{\sqrt{c+k}}.$$

If $P_1, P_2; Q_1, Q_2$ are corresponding points, then $P_1Q_2=P_2Q_1$ (J. Ivory).

F. Circular Sections

When the intersection of a plane and a quadric is a circle, the intersection is called a **circular section**. In general, circular sections are cut off by two systems of parallel planes through a quadric. The point of contact on the tangent plane parallel to these is an †umbilical point of the quadric.

G. Quadric Hypersurfaces

A subset F of an n-dimensional Euclidean space E^n is called a **quadric hypersurface** (or simply **hyperquadric**) if it is the set of points $(x_1, \dots, x_n)$ satisfying the following equation of the second degree:

$$\sum_{i,k=1}^{n} a_{ik}x_ix_k+2\sum_{i=1}^{n} b_ix_i+c=0, \tag{12}$$

where a_{ik}, b_i, c are all real numbers. We can assume without loss of generality that the matrix $A=(a_{ik})$ is symmetric. Assume that A is not a zero matrix. In the case $n=2$, F is a conic, and in the case $n=3$, it is a quadric surface. The theory of classification of quadric surfaces can be generalized to the n-dimensional case as follows: Let $r(A^*)=r^*$ be the rank of the $(n+1)\times(n+1)$ matrix

$$A^*=\begin{bmatrix} a_{11} & \dots & a_{1n} & b_1 \\ & \dots & & \\ a_{n1} & \dots & a_{nn} & b_n \\ b_1 & \dots & b_n & c \end{bmatrix}
=\begin{bmatrix} & & & b_1 \\ & A & & \vdots \\ & & & b_n \\ b_1 & \dots & b_n & c \end{bmatrix},$$

and put $r(A)=r$. Then we have the following three cases: (I) $r=r^*$; (II) $r+1=r^*$; and (III) $r+2=r^*$. Corresponding to each case, equation (12) can be simplified (by a coordinate transformation in E^n) to the following canonical forms, respectively:

(I) $\quad \sum_{i=1}^{r} \lambda_i x_i^2=0,$

(II) $\quad \sum_{i=1}^{r} \lambda_i x_i^2+1=0,$

(III) $\quad \sum_{i=1}^{r} \lambda_i x_i^2+2x_{r+1}=0,$

where $(\lambda_1, \ldots, \lambda_r, 0, \ldots, 0)$ (with $n-r$ zeros) is proportional to the †eigenvalues of the matrix A. In general, we have $1 \leqslant r \leqslant n$. In the cases where $r=n$ in forms (I) and (II) and $r+1=n$ in (III), the hypersurface is called a **properly $(n-1)$-dimensional quadric hypersurface**, and in other cases, a **quadric cylindrical hypersurface**. In cases (I) and (II), the quadric cylindrical hypersurface is the locus of $(n-r)$-dimensional subspaces passing through each point of a properly $(r-1)$-dimensional quadric hypersurface and parallel to a fixed $(n-r)$-dimensional subspace. In case (III), the quadric cylindrical hypersurface is the locus of $(n-r-1)$-dimensional subspaces passing through each point of a properly r-dimensional quadric hypersurface and being parallel to a fixed $(n-r-1)$-dimensional subspace. For form (I) with $\lambda_i > 0$ $(i=1, \ldots, n)$, a properly $(n-1)$-dimensional quadric hypersurface reduces to a point in E^n; for form (II) with $\lambda_i > 0$ $(i=1, \ldots, n)$ it becomes the empty set. Suppose that we are given a quadric hypersurface F that is neither a point nor empty. Then the system $\{\lambda_1, \ldots, \lambda_r\}$ associated with F in its canonical equation is unique up to order (and signature in form (III)). Suppose that we are given a quadric surface F and a point P on F. Suppose further that if a point X other than P is on F, then the whole straight line PX lies on F. In this case, the hypersurface is called a **quadric conical hypersurface** (or simply **quadric cone**). For example, for case (I), we can take $P=O$ (the origin), and the hypersurface is a quadric cone. In cases (I) and (II), the hypersurface is symmetric with respect to the origin. In these cases, a hypersurface is called a **central quadric hypersurface**, and in case (III), it is called a **noncentral quadric hypersurface** or **parabolic quadric hypersurface**. When we cut a parabolic quadric hypersurface by a (2-dimensional) plane containing the x_{n+1}-axis, the section is a parabola. If $\lambda_i < 0$ $(i=1, \ldots, r)$ in form (II), then the surface is called an **elliptic quadric hypersurface**, and if there are both positive and negative numbers among the λ_i, the surface is called a **hyperbolic quadric hypersurface**. The section of an elliptic quadric hypersurface by a plane is always an ellipse. The section of a hyperbolic quadric hypersurface by a plane is an ellipse, a hyperbola, or two straight lines. In general, the section of a quadric hypersurface by a subspace is a quadric hypersurface on that subspace.

H. Quadric Hypersurfaces in an Affine Space

In Section G we considered a quadric hypersurface defined by (12) in an n-dimensional Euclidean space E^n and transformed the equation to canonical form by an orthogonal transformation of coordinates in E^n. If we regard E^n as an n-dimensional †affine space over the real number field and reduce (12) to the simplest form by a coordinate transformation in the affine space, we have the following canonical forms corresponding to cases (I), (II), and (III) discussed in Section G:

(I) $(s,t)\colon \sum_{i=1}^{s} x_i^2 - \sum_{j=s+1}^{r} x_j^2 = 0,$

(II) $(s,t)\colon \sum_{i=1}^{s} x_i^2 - \sum_{j=s+1}^{r} x_j^2 + 1 = 0,$

(III) $(s,t)\colon \sum_{i=1}^{s} x_i^2 - \sum_{j=s+1}^{r} x_j^2 + 2x_{r+1} = 0,$

where $0 \leqslant s \leqslant r$ and $r-s=t$. The terms *properly $(n-1)$-dimensional, cylindrical, conical, parabolic, elliptic*, and *hyperbolic* can be defined in terms of this affine classification. For example, a cone is of type (I), a parabolic hypersurface is of type (III), an elliptic hypersurface is of type (II) $(0,r)$, a hyperbolic hypersurface is of type (II) (s,t) $(s,t>0)$, and type (II) $(s,0)$ represents the empty set. A necessary and sufficient condition for a (nonempty) hypersurface to be represented by two canonical forms $N(s,t)$, $N'(s',t')$ is that (i) $N=N'$ and (ii) if $N=$(I) or $N=$(III), then $s=s'$, $t=t'$ or $s=t'$, $t=s'$, and if $N=$(II), then $s=s'$, $t=t'$.

I. Quadric Hypersurfaces in a Projective Space

Suppose that we are given a field K of characteristic not equal to 2 and an n-dimensional †projective space $\mathbf{P}^n$ over K. A subset F of $\mathbf{P}^n$ is called a **quadric hypersurface** (or simply **hyperquadric**) if F is represented by a homogeneous equation of the second degree $\sum_{i,k=0}^{n} a_{ik} x_i x_k = 0$, where $(x_0, x_1, \ldots, x_n)$ are homogeneous coordinates in $\mathbf{P}^n$ and $a_{ik} \in K$; $A=(a_{ik})$ is a nonzero symmetric matrix. The problem of classifying such surfaces is reduced to that of †quadratic forms or, equivalently, to that of symmetrix matrices in K. Two symmetric matrices A and B are equivalent if there exists a regular matrix T such that $B={}^tTAT$ (→ 348 Quadratic Forms). In particular, when K is an †algebraically closed field or a †real closed field, a simple result is obtained. If K is an algebraically closed field, then the equation of the quadric hypersurface is reduced to the canonical form $\sum_{i=0}^{r} x_i^2 = 0$, where $r=r(A)=$ the rank of A. When K is a real closed field, then the canonical form is $\sum_{i=0}^{s} x_i^2 - \sum_{j=s+1}^{r} x_j^2 = 0$.

References

[1] G. Salmon, A treatise on the analytic geometry of three dimensions, Hodges, Figgis & Co., seventh edition, 1928 (Chelsea, 1958).

[2] G. Salmon and W. Fiedler, Analytische Geometrie des Raumes I, Teubner, fourth edition, 1898.
[3] H. F. Baker, Principles of geometry III, Solid geometry, Cambridge Univ. Press, 1922.
[4] O. Schreier and E. Sperner, Einführung in die analytische Geometrie und Algebra II, Vandenhoeck & Ruprecht, 1951; English translation, Projective geometry of n-dimensions, Chelsea, 1961.
[5] M. Protter and C. Morrey, Analytic geometry, Addison-Wesley, 1975.

351 (XX.23) Quantum Mechanics

A. Historical Remarks

†Newtonian mechanics (classical mechanics) successfully explained the motion of mechanical objects, both celestial and terrestrial, on a macroscopic scale. It failed, however, to explain blackbody radiation, which was discovered in the last decade of the 19th century. M. Planck introduced a hypothesis of discrete energy quanta, each of which contains an amount of energy E equal to the frequency of the radiation ν multiplied by a universal constant h (called **Planck's constant**). He applied this hypothesis to derive a new formula for radiation that gives predictions in good agreement with observations. A. Einstein proposed the hypothesis of the photon as a particlelike discrete unit of light rays. Assuming that many physical quantities, including energy, have only discrete values, N. H. Bohr explained the stability of electronic states in atoms. As illustrated in these examples, **quantum mechanics** is applied to study the motion of microscopic objects, including molecules, atoms, nuclei, and elementary particles.

B. Quantum-Mechanical Measurement

Fundamental differences between the new mechanics and classical mechanics are due to the facts that many physical quantities, for example, energy, can take only discrete values in the microscopic world, and that states of microscopic objects are disturbed by observation.

A (pure) **state** at a certain time is expressed by a unit vector ψ in a †Hilbert space $\mathscr{H}$, and **observables**, or physical quantities, are expressed by †self-adjoint operators in such a space.

Let a_n $(n=1,2,\dots)$ be †eigenvalues of an observable A, and let P_n be a †projection operator onto the eigenspace spanned by †eigenvectors belonging to the eigenvalue a_n. Suppose that $A=\sum_n a_n P_n$. Then the hypothesis on measurement in quantum mechanics is given as follows.

When an observable A is observed in a state ψ, one of the eigenvalues a_n is found with probability proportional to $(\psi, P_n\psi)^2$. When an eigenvalue a_n is once observed, a state jumps from ψ to an eigenstate $P_n\psi$ which belongs to the eigenvalue a_n. Quantum mechanics predicts only a probability p_n with which a certain value a_n is found when an observable A is observed. This probability, given by $(\psi, P_n\psi)$, is not changed even if ψ is replaced by $e^{i\theta}\psi$, $0\leqslant\theta\leqslant 2\pi$. Therefore $e^{i\theta}\psi$ represents the same state as ψ. The set of $e^{i\theta}\psi$, $0\leqslant\theta\leqslant 2\pi$, for a fixed ψ ($\|\psi\|=1$) is called a **unit ray**.

If P_n is 1-dimensional and $P_n\varphi=\varphi$, $\|\varphi\|=1$, then $(\psi, P_n\psi)=|(\psi,\varphi)|^2$ is called the **transition probability** between the two states.

The **expectation value** (or **expectation**) of an operator A in a state ψ, usually normalized to $(\psi,\psi)=1$, is defined to be $\langle A\rangle=(\psi, A\psi)=\sum_n a_n p_n$.

A general self-adjoint operator A can be written as $A=\int\lambda\, dP(\lambda)$. When A is observed in a state ψ, the probability for a value to be found between λ_1 and λ_2 ($\lambda_2>\lambda_1$; λ_2 included and λ_1 excluded if $P(\lambda)$ is right continuous) is $(\psi(P(\lambda_2)-P(\lambda_1)),\psi)$ (→ 390 Spectral Analysis of Operators).

The quantity $(\varphi, A\psi)$ is called the **matrix element** of A between the two states φ and ψ. A state ψ can be viewed as a functional $\psi(A)=\langle A\rangle$ on the set of all observables A (its value being the expectation), which is linear in A, positive in the sense $\psi(A^*A)\geqslant 0$ for any operator A, and normalized: $\psi(1)=1$. If $0\leqslant\lambda\leqslant 1$ and $\psi(A)=\lambda\psi_1(A)+(1-\lambda)\psi_2(A)$ for all observables A, then the state ψ is called a **mixture** of states ψ_1 and ψ_2 with weights λ and $(1-\lambda)$. If a state is not a nontrivial (i.e., $\lambda\neq 0,1,\psi_1\neq\psi_2$) mixture, it is called **pure**. The state $\langle A\rangle=(\psi, A\psi)$ on the set of all self-adjoint operators A given by a vector ψ is pure in this sense. If $\sup_\alpha\psi(A_\alpha)=\psi(A)$ whenever A_α is an increasing net of positive operators with A as its limit, then ψ is called **normal**. Any normal positive linear functional on the set of all self-adjoint operators can be described by a trace-class positive operator ρ, called the **density matrix**, as $\langle A\rangle=\mathrm{tr}(A\rho)$. If $\{\psi_n\}$ is a complete orthonormal set, where each ψ_n is an eigenvector of ρ belonging to the eigenvalue λ_n, then $\langle A\rangle=\sum_n\lambda_n(\psi_n, A\psi_n)$.

C. Canonical Commutation Relations

In quantum mechanics, canonical variables are represented by the self-adjoint operators

Q_k (coordinates) and P_k (momenta), $k=1,\ldots,N$, which satisfy the **canonical commutation relations**

$$[Q_k, P_l]=i\hbar\delta_{kl}\mathbf{1},$$

where $\mathbf{1}$ is the identity operator, $[A,B]$ denotes the commutator $AB-BA$, $\hbar=h/(2\pi)$, and the relation is supposed to hold on a certain dense domain of vectors. Self-adjoint operators Q_k and P_k satisfying the above relations are unique up to quasi-equivalence under a suitable domain assumption, e.g., if $\sum_k Q_k^2+\sum_k P_k^2$ is essentially self-adjoint on a dense domain invariant under multiplication of the Q's and P's and on which the above relations are satisfied. Under such an assumption, Q_k and P_k are unitarily equivalent to a direct sum of the **Schrödinger representation** on $L_2(\mathbf{R}^N, dx_1\ldots dx_N)$, where Q_k is multiplication by the kth coordinate x_k and P_k is the differentiation $-i\hbar(\partial/\partial x_k)$ (**Rellich-Dixmier theorem**).

The above Schrödinger representation is called the **position representation** (or q-representation). The formulation using the function space L_2 of real variables p_k, $k=1,2,\ldots,N$, on which the operators P_k act as multiplications by p_k, is called the **momentum representation** (or p-representation).

If Hermitian operators A and B satisfy the canonical commutation relations in the form $(A\psi,B\psi)-(B\psi,A\psi)=i\hbar(\psi,\psi)$, then the following **Heisenberg uncertainty relation** holds for the expectation:

$$\langle(A-\langle A\rangle)^2\rangle\langle(B-\langle B\rangle)^2\rangle\geqslant\frac{\hbar^2}{4}.$$

This gives the **uncertainty** in observations, which means that two observables A and B cannot simultaneously be observed with accuracy. This is another important property of microscopic motion that cannot be found in macroscopic motion.

In a direct sum of the Schrödinger representation of the canonical commutation relations, the unitary operators

$$U(a)=\exp i\sum_k a_kQ_k,\qquad V(b)=\exp i\sum_k b_kP_k$$

with real parameters a_k and b_k, $k=1,\ldots,N$, satisfy the following **Weyl form** of the canonical commutation relations:

$$U(a)U(a')=U(a+a'),\qquad V(b)V(b')=V(b+b'),$$

$$U(a)V(b)=V(b)U(a)\exp\left(-i\sum_k a_kb_k\right).$$

Conversely, any pair of families of unitary operators $U(a)$ and $V(b)$ satisfying these relations and depending continuously on parameters a and b are unitarily equivalent to those obtained as above (**von Neumann uniqueness theorem**).

D. Time Evolution and the Schrödinger Equation

The time t of an observation is fixed in the foregoing discussion. A state changes, however, as the time t changes, in such a way that the transition probability between states is preserved. By Wigner's theorem (→ Section H), this time evolution of states can be implemented by unitary operators U_t defined by the transformation of vectors $\psi\to U_t\psi=\psi_t$. Furthermore, under some continuity assumption, such as that of $(\varphi, U_t\psi)$, U_t can be made a continuous one-parameter group. By Stone's theorem, $U_t=e^{-iHt/\hbar}$ for a self-adjoint operator H. This operator is called the **Hamiltonian operator** (or simply **Hamiltonian**) determined by the structure of a system. An infinitesimal change in ψ corresponding to an infinitesimal change in t can be generated by this operator H as follows:

$$i\hbar\frac{\partial\psi_t}{\partial t}=H\psi_t.$$

This equation is called the **time-dependent Schrödinger equation**.

A state ψ changes but observables do not change with time in the **Schrödinger picture** above. The other picture, known as the **Heisenberg picture**, is equally possible. In this picture, the state is expressed by a time-independent vector, while operators A vary with time as follows: $A\to U_t^*AU_t=A(t)$. Rates of change of operators $A(t)$ can be calculated by means of the equation

$$\frac{dA(t)}{dt}=\frac{i}{\hbar}[H,A(t)],$$

which is called the **Heisenberg equation of motion**. When time t changes, the expectation value of an operator A in a state ψ changes in both pictures according to

$$d\langle A\rangle/dt=i\langle[H,A]\rangle/\hbar.$$

According to classical analytical dynamics, a change of a dynamical quantity that is a function of †canonical variables q_i (positions) and p_i (momenta) is given by

$$dA/dt=-(H,A),$$

where H is a †Hamiltonian function and the parentheses (,) denote the †Poisson bracket. A replacement of the Poisson bracket (A,B) by $[A,B]/i\hbar$ transforms this classical equation into the quantum-mechanical equation above. It should be noticed that the mathematical structure of the Poisson bracket is similar to that of commutator. In this transition from classical to quantum mechanics the **correspondence principle** can be used. This requires that the laws of quantum mechanics must lead

to the equations of classical mechanics in the classical situation, where many quanta are involved and $\hbar$ can be regarded as infinitesimally small in the commutation relation. The correspondence principle suggests that Hamiltonian operators in quantum mechanics can be obtained from Hamiltonian functions $H(p_k, q_k)$ of canonical variables p_k and q_k in classical mechanics after replacing p_k and q_k by the operators P_k and Q_k in the Schrödinger representation (up to uncertainty of about the order of operators). This process of moving from canonical variables and the Hamiltonian function in classical mechanics to canonical operators and the Hamiltonian in quantum mechanics is called **quantization**. Taking a system of s particles and letting x_k, y_k, and z_k be the Cartesian coordinates of the kth particle, we usually write the equation of motion as

$$i\hbar\frac{\partial\psi}{\partial t}=\left(-\sum_{k=1}^{s}\frac{\hbar^2}{2m_k}\Delta_k + V(x_1,y_1,z_1,\ldots,x_s,y_s,z_s)\right)\psi,$$

which is a second-order partial differential equation. Here m_k is the mass of the kth particle; Δ_k is the †Laplacian of x_k, y_k, and z_k; and V is a real function called the potential energy. This equation is the time-dependent **Schrödinger equation**. The partial differential operator on the right-hand side is called the (s-body) **Schrödinger operator** and $\psi(x_1, y_1, z_1, \ldots, x_s, y_s, z_s, t)$ is called a **wave function**. The probability of finding a particle in the volume $dx_k dy_k dz_k$ bounded by x_k, x_k+dx_k, y_k, y_k+dy_k, and z_k, z_k+dz_k is proportional to $|\psi(x_1,y_1,z_1,\ldots,x_s,y_s,z_s,t)|^2$. Usually $|\psi|^2$ is normalized so that its integral over the whole space is 1. We sometimes call ψ the **probability amplitude**. When ψ is given by $e^{-iEt/\hbar}\varphi(x_1,\ldots,z_s)$, the expectation value of an operator A in a state ψ, $\langle A\rangle=\int\psi^*A\psi\,dx_1\ldots dz_s$, does not depend on time. When this is the case, ψ is called a **stationary state**.

A real value E and a function $\varphi(x_1,\ldots,z_s)$ are found by solving an †eigenvalue problem $H\varphi=E\varphi$. This equation is the **time-independent Schrödinger equation**, which is a second-order partial differential equation. Since the Hamiltonian H stands for the energy of this system, the eigenvalues E are the energy values that this system can take.

When a potential function V is given, it is a nontrivial matter to prove that the (s-body) Schrödinger operator with the given V is essentially self-adjoint on the set of, for example, all C^∞-functions with compact supports so that its closure H defines mathematically the continuous one-parameter group of unitaries $U_t=e^{-iHt/\hbar}$ for the time evolution of the quantum system of s particles with the given interaction potential V. If V satisfies an estimate $\|V\psi\|\leqslant\lambda\|H_0\psi\|+\mu\|\psi\|$ (called the **Kato perturbation** on H_0) for some nonnegative $\lambda<1$ and $\mu\geqslant 0$ and for all ψ in a dense domain on which H_0 is essentially self-adjoint, where H_0 denotes the Schrödinger operator with $V=0$ (called the **free Hamiltonian** or the **kinetic energy** term), then H_0+V is essentially self-adjoint on the same domain. For the case where V consists of Coulomb interactions between electrons and Coulomb potentials on electrons by fixed nuclei, for example, such an estimate and hence the essential self-adjointness of Hamiltonians for atoms and molecules were established first by T. Kato (*Trans. Amer. Math. Soc.*, 70 (1951)).

For a 1-body Schrödinger operator (or 2-body Schrödinger operator after the center of mass motion has been separated out), the point spectrum is that of the particle trapped by the potential, and the state represented by its eigenvector is called a **bound state**. The eigenvalue is nonpositive for a reasonable class of potentials V (for example, if $V(x)$ $(x\in\mathbf{R}^3)$ is continuous and $O(|x|^{-1-\varepsilon})$ as $|x|\to\infty$ for some $\varepsilon>0$), and its absolute value is called the **binding energy**. The eigensolutions of the Schrödinger equation are what have been called stationary states above.

There are also stationary solutions that do not correspond to the point spectrum and hence are not square integrable. They are used in the stationary methods of scattering theory (→ 375 Scattering Theory).

E. Some Exact Solutions for the 1-Body Schrödinger Equation

(1) **Harmonic oscillator**. First consider the case in which the space is of 1 dimension, so that the Laplacian Δ is $(d/dx)^2$. Let m be the mass of the particle and $V(x)=m\omega^2x^2/2$ for a positive constant ω (called the **angular frequency**). The Hamiltonian

$$H=-(\hbar^2/2m)(d/dx)^2+m\omega^2x^2/2$$

has simple eigenvalues

$$E_n=\hbar\omega(n+(1/2)),\qquad n=0,1,2,\ldots,$$

with a complete orthonormal set of eigenfunctions

$$\psi_n(x)=c_nH_n(q)e^{-q^2/2},\qquad q=(m\omega/\hbar)^{1/2}x,$$

where $H_n(q)$ is a †Hermite polynomial and c_n is the normalization constant:

$$H_n(q)=\sum_{k=0}^{[n/2]}(-1)^kn!(2q)^{n-2k}/\{(n-2k)!k!\},$$

$$c_n=\{2^{2n}(n!)^2\pi\hbar/(m\omega)\}^{-1/4}.$$

When the space is of r dimensions, n in E_n is replaced by $n_1+\ldots+n_r$ with nonnegative integers $n_1, \ldots, n_r$, and the corresponding eigenfunction is $\prod_{j=1}^r \psi_{n_j}(x_j)$.

(2) **One-dimensional square-well potential.** Let $V(x)=V$ for $|x| \leqslant a/2$ and $V(x)=0$ for $|x|>a/2$ $(x \in \mathbf{R})$. If $V \geqslant 0$, there are no point spectra. If $V<0$ and

$$N-1<a(-2mV)^{1/2}/(\pi\hbar) \leqslant N,$$

then there are N eigenvalues $(N=1,2,\ldots)$ obtained as the roots $E<0$ of one of the following equations:

$$\{(V/E)-1\}^{1/2}=\tan\{a(-2mE)^{1/2}/(2\hbar)\},$$

$$\{(V/E)-1\}^{1/2}=-\cot\{a(-2mE)^{1/2}/(2\hbar)\}.$$

(3) **Separation of angular dependence for central potential.** If $V(x)$ $(x \in \mathbf{R}^3)$ depends only on $r=|x|=(\sum_{j=1}^3 (x_j)^2)^{1/2}$ (called a **central potential**), then all eigenvalues E and a basis for eigenfunctions $\psi(x)$ can be obtained in terms of the polar coordinate r, θ, φ $(x_1=r\sin\theta\cos\varphi$, $x_2=r\sin\theta\sin\varphi$, $x_3=r\cos\theta)$ as

$$\psi(x)=Y_{lm}(\theta,\varphi)r^{-1}u(r),$$

$$-(\hbar^2/(2m))u''(r)+\{\hbar^2 l(l+1)/(2mr) + V(r)-E\}u(r)=0,$$

$$\|\psi\|=\int_0^\infty |u(r)|^2\,dr<\infty,$$

where the angular function Y_{lm} is an eigenfunction of the square $\mathbf{L}^2$ of the **orbital angular momentum** $\mathbf{L}=-ix\times\nabla$:

$$Y_{lm}=c_{lm}P_l^m(\cos\theta)e^{im\varphi},$$

$$c_{lm}=(-1)^m\{(2l+1)(l-m)!/(4\pi(l+m)!)\}^{1/2},$$

$$m=l, l-1, \ldots, -l+1, -l, \qquad l=0,1,2,\ldots.$$

Here $P_l^m(x)$ is an †associated Legendre polynomial:

$$P_l^m(x)=(1-x^2)^{m/2}(d^{l+m}(x^2-1)^l/dx^{l+m})/(2^l l!).$$

The above equation for $u(r)$ is called the **radial equation**. The nonnegative integer l is the **azimuthal quantum number**, and the integer m is the **orbital magnetic quantum number**. The wave function $\psi(x)$ with the angular dependence $Y_{lm}(\theta,\varphi)$ is called the **S-wave**, **P-wave**, **D-wave**, ... according as $l=0,1,2,\ldots$.

(4) **Hydrogen-type atom.** Let $V(r)=-Ze^2/r$ $(Z>0)$. For each l and m, there are eigenvalues $-e^2Z^2/(2an^2)$ with eigenfunctions $\psi_{nlm}=r^{-1}u_{nl}(r)Y_{lm}(\theta,\varphi)$, where $n=l+1, l+2, \ldots$ is the **principal quantum number**,

$$u_{nl}(r)=c_{nl}L_{n+l}^{(2l+1)}(s)s^{l+1}e^{-s/2}, \qquad s=2Zr/(na),$$

$$c_{nl}=-\{(n-l-1)![2Z/(na(n+l)!)]^3/(2n)\}^{1/2},$$

and $L_N^\mu(x)$ is the μth derivative of the †Laguerre polynomial

$$L_N(x)=\sum_{\nu=0}^N (N!)^2(-x)^\nu/\{(N-\nu)!(\nu!)^2\}.$$

The eigenvalue is determined by n, and its multiplicity is n^2, corresponding to the different possible values of l and m.

F. Path Integrals

R. P. Feynman (*Rev. Mod. Phys.*, 20 (1948)) has given the solution of the Schrödinger equation as an integral of $e^{iL/\hbar}$ over all possible paths $q(t)$, where $L=L(q,\dot{q})$ $(\dot{q}=(d/dt)q(t))$ is the classical Lagrangian for the Hamiltonian system. This integral is called the Feynman path integral. Mathematical reformulation of the formula in terms of the Wiener measure has been given by M. Kac (*Proc. 2nd Berkeley Symp. Math. Statist. Probability*, 1950; *Probability and related topics in the physical sciences*, Wiley, 1959).

Consider the 1-body Schrödinger operator $H=H_0+V$ (form sum), where V is the sum of a locally integrable function bounded below and a Kato perturbation on H_0. Let $b(t)$ $(t \geqslant 0)$ be the Wiener process and $q(t)=\hbar b(t)/(2m)^{1/2}$. For any L_2 functions f,

$$(e^{-tH/\hbar}f)(x)$$

$$=E\left(f(x+q(t))\exp\left\{-\int_0^t V(x+q(s))\,ds/\hbar\right\}\right)$$

for almost all x, where E denotes the expectation for the Wiener process. If V is a sum of L_2 and L_∞ functions (for spatial dimension $\leqslant 3$), then the right-hand side is continuous in x for $t>0$. This is called the **Feynman-Kac formula**.

Let L_0 be the Hamiltonian for a 1-dimensional harmonic oscillator with $m=\omega=\hbar=1$ and ψ_0 be the eigenfunction $\psi_0(x)=\pi^{-1/4}\exp(-x^2/2)$. Consider $H=L_0+V$ (form sum), where V is a sum of a locally integrable function bounded below and a Kato perturbation on L_0. Let $q(t)$ $(t \in \mathbf{R})$ be Gaussian random variables with mean 0 and covariance $E(q(t)q(s))=2^{-1}\exp(-|t-s|)$ (called the **oscillator process**). For any $f_j(x)$ in $L_2(\mathbf{R}, \psi_0^2\,dx)$ $(j=1,\ldots,n)$ and $t_0 \leqslant t_1 \ldots \leqslant t_n \leqslant t_{n+1}$,

$$(\psi_0, e^{-(t_1-t_0)H}f_1e^{-(t_2-t_1)H}f_2\ldots f_n e^{-(t_{n+1}-t_n)H}\psi_0)$$

$$=E\left(\left\{\prod_{j=1}^n f_j(q(t_j))\right\}\exp\left\{-\int_{t_0}^{t_{n+1}} V(q(s))\,ds\right\}\right).$$

The above path integral formulas are closely related to the **Trotter product formula**

$$e^{-t(A+B)}=\lim_{n\to\infty}(e^{-tA/n}e^{-tB/n})^n \qquad (t \geqslant 0),$$

where A and B are self-adjoint operators

bounded below and $A+B$ is essentially self-adjoint. (The same formula holds without the boundedness assumption when $t \in i\mathbf{R}$.)

G. The Dirac Equation

The Schrödinger equation is not relativistically invariant. The **Klein-Gordon equation**

$$(\square-\kappa^2)\psi=0, \qquad \square\equiv\Delta-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}, \qquad \kappa=\frac{mc}{\hbar}$$

is obtained by replacing p_k by $(\hbar/i)\partial/\partial x_k$ $(k=1,2,3)$ and E by $i\hbar\partial/\partial t$ in the relativistic identity

$$E^2=m^2c^4+\mathbf{p}^2c^2,$$

where c is the speed of light. Wave functions of free particles are believed to satisfy this equation. P. A. M. Dirac assumed that the ψ of a free electron is expressed in terms of a †spinor with four components satisfying a linear differential equation that automatically implies the Klein-Gordon equation. Relativity requires the equal handling of space and time. The **Dirac equation**

$$\sum_{\mu=0}^{3}\gamma^\mu\frac{\partial\psi}{\partial x^\mu}+i\kappa\psi=0 \tag{8}$$

$(x^0=ct)$ satisfies these requirements. The coefficients γ^μ can be so determined that every component of ψ also satisfies the Klein-Gordon equation. Thus the γ^μ are found to be 4×4 matrices satisfying the commutation relations $\gamma^\mu\gamma^\nu+\gamma^\nu\gamma^\mu=2g^{\mu\nu}$ $(\mu,\nu=0,1,2,3)$, where $g^{\mu\nu}=0$ for $\mu\neq\nu$ and $g^{00}=-g^{kk}=1$ $(k=1,2,3)$. Sixteen linearly independent matrices are obtained by repeated multiplication of five matrices, which include the four matrices γ^0, γ^1, γ^2, γ^3 and the identity matrix. Any 4×4 matrix can be expressed as a linear combination of these sixteen matrices.

The Dirac equation has †plane wave solutions

$$\psi(\mathbf{x},t)=u\exp((i/\hbar)\mathbf{p}\cdot\mathbf{x}-(i/\hbar)Et),$$

where the energy eigenvalues E are

$$\pm\sqrt{m^2c^4+p^2c^2}.$$

There are four independent eigensolutions $u^{(1)}$, $u^{(2)}$, $u^{(3)}$, and $u^{(4)}$, because u has four components. Two of them are of positive energy and the other two are of negative energy. Although the negative energy case is physically undesirable, it has to be taken into account in order to obtain a mathematically complete set. To solve this difficulty, Dirac proposed the hypothesis (**Dirac's hole theory**) that all the negative energy states are filled up by an infinite number of electrons in the normal state of the vacuum. The absence from the vacuum of a negatively charged electron in a negative energy state could then be expected to manifest itself as a positively charged particle (positron) with positive mass and kinetic energy. If gamma rays are absorbed to excite an electron from a negative energy state into a positive energy state, an electron-position pair must be created. Y. Nishina and O. Klein calculated the cross section of Compton scattering (the **Klein-Nishina formula**) by use of the Dirac equation and found good agreement with observations, thus providing evidence that the Dirac equation is correct. The existence of negative energy states, however, forces us to give up considering the Dirac equation as an equation of one electron. The positron theory is introduced, and the Dirac equation is considered as the classical field of electron waves and is second-quantized (→ 377 Second Quantization).

The Klein-Gordon equation can also be considered to be the classical wave equation of matter and can be second-quantized. Motions of particles with zero spin, pi mesons (π) for example, obey this equation.

We can rewrite the Dirac equation as $i\hbar\partial\psi/\partial t=H\psi$, $H=c\alpha\cdot\mathbf{p}+mc^2\beta$, where $\gamma^k=\beta\alpha_k$ $(k=1,2,3)$, and $\gamma^0=\beta$. Since H cannot commute with the orbital angular momentum of an electron $\mathbf{L}=\mathbf{r}\times\mathbf{p}=-i\hbar\mathbf{r}\times\nabla$, $\mathbf{L}$ is not conserved. However, the total angular momentum $\mathbf{J}=\mathbf{L}+(\hbar/2)\boldsymbol{\sigma}$ can be conserved when $\boldsymbol{\sigma}$ is a vector whose components can be given as $\begin{pmatrix}\sigma_k & 0\\ 0 & \sigma_k\end{pmatrix}$, where $\beta=\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}$, $\alpha_k=\begin{pmatrix}0 & \sigma_k\\ \sigma_k & 0\end{pmatrix}$, and the σ_k, called **Pauli spin matrices**, are given by $\sigma_x=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$, $\sigma_y=\begin{pmatrix}0 & -i\\ i & 0\end{pmatrix}$, $\sigma_z=\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}$. (The γ^μ are called **Dirac's γ matrices**.) The quantity $\mathbf{S}=\frac{\hbar}{2}\boldsymbol{\sigma}$ is the **intrinsic angular momentum** of the electron, also called the **spin**. Many particles besides the electron, the neutron for example have spin. The matrix $\mathbf{S}^2=S_x^2+S_y^2+S_z^2$ is diagonal and is equal to $s(s+1)\hbar^2 I$ (I is the identity matrix). For the electron $s=1/2$, and $\hbar/2$ is called the absolute value of the spin. Therefore we say that electrons are particles of spin $\hbar/2$. This was predicted in the theory of light spectra.

When the speed of an electron is very small, so that $(v/c)^2$ can be neglected, states of the electron can be expressed in terms of two-component wave functions. This approximation is called the **Pauli approximation**. If the spin-orbit term that appears in the Pauli approximation is also neglected, these two components become independent of each

other and individually satisfy the Schrödinger equation.

H. Application of Representation Theory of Lie Groups

A **symmetry** (with active interpretation) is a bijective mapping of pure states (represented by unit rays in a Hilbert space) preserving transition probabilities between them. **Wigner's theorem** says that any symmetry can be implemented by either a unitary or an antiunitary mapping of the underlying Hilbert space as a mapping of unit rays onto themselves. Furthermore a connected Lie group of symmetries is implemented by unitaries, which form a projective representation. Eigenstates ψ of the Schrödinger equation are functions of the coordinates of each particle. Let these coordinates be denoted together as x.

Suppose that an operator T operates on the x, as, for example, a rotation of the coordinate system or a permutation of the labels of the particles. If T commutes with H or H is invariant under the transformation $x \rightarrow x' = Tx$, then $T\psi(x) = \psi'(x) = \psi(T^{-1}x)$ satisfies the same Schrödinger equation as ψ, where the transformation of the function is defined by $\psi'(x') = \psi(x)$. The set of transformations $x \rightarrow x' = Tx$ forms a †group $\{T\}$, and the corresponding transformations $\psi \rightarrow T\psi$ give a (generally infinite-dimensional) representation of this group, which should be †unitary on L_2-space relative to the Lebesgue measure dx if T leaves the measure invariant. There are **degeneracies** of the energy eigenvalues, each of which is equal to the dimension of the corresponding representation of the group $\{T\}$. If the representation for each eigenvalue is decomposed into irreducible ones, then the decomposed stationary state can be labeled by an †irreducible representation.

When H is spherically symmetric, i.e., H is invariant under the 3-dimensional rotation group, states are classified by the irreducible representations D_L of the †rotation group (→ 258 Lorentz Group). The square of the sum $\mathbf{L}$ of all orbital angular momenta has eigenvalues $L(L+1)\hbar^2$, where L must be 0 or a positive integer. There are $2L+1$ degenerate states, each of which belongs to a different M, the z-component of $\mathbf{L}$, where M ranges from L to $-L$ by unit steps. Even when there is an interaction between the orbital angular momentum and the spin angular momentum, states are labeled by the irreducible representations D_J of the rotation group, where $\mathbf{J}$ is the sum of the orbital angular momentum $\mathbf{L}$ and the spin angular momentum $\mathbf{S}$ ($\mathbf{J} = \mathbf{L} + \mathbf{S}$) and eigenvalues of $\mathbf{J}^2$ are given by $J(J+1)\hbar^2$. Each J must be zero, a positive integer, or a half-integer. Adding inversions to the pure rotations, we obtain the 3-dimensional orthogonal group (→ 60 Classical Groups I). Irreducible representations of this group are written as $D_J^{\pm}$, where $\pm$ corresponds to the characters of the inversion relative to the origin. States with $+$ are called **even states** and those with $-$, **odd states**. For example, energy levels of atoms and nuclei can be classified by $D_J^{\pm}$.

To obtain matrix elements of observables between two stationary states, group representation theory is useful. The transformation of every observable obeys a certain rule under the transformation of coordinates. The scalar is transformed according to D_0^+, the vector according to D_1^-, the pseudovector according to D_1^+, and the traceless tensor according to D_2^+. If the transformation of an observable is given by D_J, then a matrix element of this observable between the states belonging to $D_{J'}$ and $D_{J''}$ vanishes unless the tensor product representation $D_J \otimes D_{J'}$ contains as a factor a representation equivalent to $D_{J''}$. In electromagnetic transitions in atoms or nuclei, $D_1 \otimes D_{J'} = D_{J'+1} + D_{J'} + D_{J'-1}$ ($J' \geqslant 1$) if the electric dipole transition dominates ($J = 1$). This implies the **selection rule** $J' \rightarrow J'+1, J', J'-1$. When $J' = 0$, only the transition $0 \rightarrow 1$ is possible. More general selection rules can be obtained in the same way for general multipole transitions. Representation theory is useful in determining general formulas of transition strengths.

There is a class of particles, many of which can occupy the same state, called **bosons**. There is another class of particles, of which only one can occupy a given state, called **fermions**. For example, the electron, neutron, and proton are fermions, while the photon and pi meson are bosons. Two identical particles, both of which are either fermions of the same kind or bosons of the same kind, cannot be distinguished. Therefore the Hamiltonian should be invariant under permutations between identical fermions, or between identical bosons. A system consisting of N identical particles can be classified by the irreducible representations of the †symmetric group S_N of N elements. When the particles are fermions, two of them cannot occupy the same state (this law is called the **Pauli principle**), so that only totally antisymmetric states are permissible for fermions. When a system consists only of fermions of the same kind with spin $\hbar/2$ and a Hamiltonian of this system does not include terms depending on spins, then the wave functions are just products of spin and orbital parts. In order to make wave functions totally

antisymmetric, orbital wave functions with a total spin $v/2$ should be limited to those corresponding to the †Young diagram $[2^{N-v}, 1^v] = T(2, 2, \dots, 1, 1)$.

I. Polyatomic Molecules

The group generated by the 2-dimensional rotations about the axis connecting two atoms and the reflections with respect to the planes containing this axis is used to classify states of diatomic molecules. Stationary states can be classified by the absolute value Λ of the angular momentum of the diatomic system around this axis, which can be zero or a positive integer. When $\Lambda \geqslant 1$, the corresponding state has twofold degeneracy, whereas when $\Lambda = 0$, two states (labeled by $\pm$) arise, depending on the character of the reflections. If these two atoms are identical, the molecular states can be classified further as even and odd according to the character of reflections with respect to the plane containing the center of mass and perpendicular to the axis. The classification of the spectral terms of a polyatomic molecule is related to its symmetry, described by the set of all transformations that interchange identical atoms. For example, stationary states of methane molecules CH_4 are classified by the irreducible representations of the group T_d (which is generated by adjoining reflection symmetry to the †tetrahedral group T). Level structures of a crystal are classified by the irreducible representations of its symmetry groups.

In the approximation of many-body problems by means of independent particles, the wave function of the total system is constructed by multiplying the wave functions of the individual particles. To construct such a wave function, it is useful to consider the reduction to irreducible parts of the †tensor products of representations of the groups attached to the individual particles. For example, an atom with two electrons carrying the angular momenta J_1 and J_2 has $2J'+1$ different angular momentum states, where $J' = \min(J_1, J_2)$ corresponding to the decomposition $D_{J_1} \otimes D_{J_2} = D_{J_1+J_2} + D_{J_1+J_2-1} + \dots + D_{|J_1-J_2|}$. The right-hand side of this equation gives all possible states of the atom.

J. Charge Symmetry

The proton and the neutron can be considered to be different states of the same particle, called the nucleon, because these two particles have very similar natures except for their charges, masses, etc. As an approximation, the Hamiltonian of a system consisting of protons and neutrons may be taken to be invariant under the interchange of protons and neutrons. This invariance is called **charge symmetry**. Analogous to ordinary spin, **isospin** can be introduced to describe the two states of nucleons. The up state of the isospin corresponds to the proton and the down state corresponds to the neutron. Consider transformations belonging to the †special unitary group $SU(2)$ in the 2-dimensional space spanned by the proton state and the neutron state. If a Hamiltonian of N nucleons is invariant under any transformation belonging to $SU(2)$, then the eigenstate of these N nucleons is classified by its irreducible representations D_T, where T stands for the total isospin of each state. This invariance, called **isospin invariance**, holds in the nucleus and elementary particles if electromagnetic and weak interactions, and possibly the interaction responsible for the proton-neutron mass difference, are neglected. When a state of N nucleons has an isospin $T = v/2$, the orbital-spin wave function of this state must correspond to the Young diagram $[2^{N-v}, 1^v]$, since the isospin wave function multiplied by the orbital-spin wave function is a totally antisymmetric wave function. If this Hamiltonian is also independent of spin, it is invariant under unitary transformations in the 4-dimensional space spanned by the four internal states of the nucleon: up and down spins, up and down isospins. Therefore the states of N nucleons can be classified by the irreducible representations of the group $U(4)$ (Wigner's **supermultiplet theory**).

K. The C^*-Algebra Approach

The uniqueness of operators satisfying the canonical commutation relations (representations of CCR's) up to quasi-equivalence (→ Section C) no longer holds if the number of canonical variables become infinite (a so-called system of infinitely many degrees of freedom), a point first emphasized by K. O. Friedrichs (*Mathematical aspects of the quantum theory of fields*, Interscience, 1953), and physical examples illustrating this point were given by L. van Hove (*Physica*, 18 (1952)) and R. Haag (*Mat. Fys. Medd. Danske Vid. Selsk.*, 29 (1955)). The use of C^*-algebras in physics was first advocated by I. E. Segal (*Ann. Math.*, 48 (1947)), and the physical relationship among all inequivalent representations of a C^*-algebra was first discussed by R. Haag and D. Kastler (*J. Math. Phys.*, 5 (1964)).

In C^*-algebra approach, a physical observable is an element of a C^*-algebra $\mathfrak{A}$ and a state is a functional φ on $\mathfrak{A}$ (its value $\varphi(A)$ is

the expectation value of the observable $A \in \mathfrak{A}$ when measured in that state) that is linear, positive in the sense $\varphi(A^*A) \geqslant 0$ for any $A \in \mathfrak{A}$, and normalized, i.e., $\|\varphi\| = 1$ or equivalently $\varphi(1) = 1$ if $1 \in \mathfrak{A}$. The †GNS construction associates with every state φ a Hilbert space H_φ, a representation $\pi_\varphi(A)$, $A \in \mathfrak{A}$, of $\mathfrak{A}$ by bounded linear operators on H_φ, and a unit cyclic vector Ω_φ in H_φ such that $\varphi(A) = (\Omega_\varphi, \pi_\varphi(A)\Omega_\varphi)$. Two states φ and ψ (or rather π_φ and π_ψ) are called disjoint if there is no nonzero mapping T from H_φ to H_ψ such that $T\pi_\varphi(A) = \pi_\psi(A)T$ for all $A \in \mathfrak{A}$. Abundant disjoint states occur for a system of infinitely many degrees of freedom, e.g., equilibrium states of a infinitely extended system with different temperatures (→ 402 Statistical Mechanics), superselection sectors explained below (→ 150 Field Theory) and equilibrium or ground states with broken symmetry.

Because actual measurement can be performed only on a finite number of observables (though chosen at will from an infinite number of possibilities) and only with nonzero experimental errors, information on any state φ can be obtained by measurements only up to a neighborhood in the weak topology: $|\varphi(A_i) - a_i| < \varepsilon_i$, $i = 1, \ldots, n$. A set K_1 of states can describe measured information on states at least equally well as another set K_2 (K_1 **physically contains** K_2) if the closure of K_1 in the weak topology contains K_2. From another viewpoint, all states of K_2 are weak limits of states in K_1 and are physically relevant if states in K_1 are physically relevant. The set of all mixtures of vector states $(\psi, \pi(A)\psi)$ for any fixed faithful representation π of $\mathfrak{A}$ is weakly dense in the set of all states of $\mathfrak{A}$, a point emphasized by Haag and Kastler as a foundation of the algebraic viewpoint in the formulation of quantum theory.

Under 360° rotation a vector representing a state of a particle with spin $\hbar/2$ acquires a factor -1 (→ 258 Lorentz Group), while the vacuum vector would be unchanged. A nontrivial linear combination (superposition) of these two would then be changed to a vector in a different ray. If the 360° rotation is not to produce a physically observable effect, then we should either forbid nontrivial superpositions of states of the two classes or, equivalently, restrict observables to those leaving the subspace spanned by vectors in each class invariant so that the relevant linear combinations of vectors, when considered as states on the algebra of observables in the form of expectation functionals $(\psi, A\psi)$, are actually mixtures (rather than superpositions) of states in two classes and are invariant under the 360° rotation. This is called the **univalence superselection rule** and has been pointed out by A. S. Wightman, G. C. Wick, and E. P. Wigner (*Phys. Rev.*, 88 (1952)).

In quantum field theory, the vacuum state can be taken to be pure (by central decomposition if necessary) and in the associated GNS representation (called the vacuum sector) all vectors can be assumed to be physically relevant pure states. In principle, all physically relevant information is in the vacuum representation; for example, a particle with spin $\hbar/2$ can also be discussed in the vacuum sector if we consider a state of this particle in the presence of its antiparticle at a far distance, such as behind the moon (the **behind-the-moon argument**). However, it is mathematically more convenient to consider the states of the particle without any compensating object (in the same way that an infinitely extended gas is more convenient for some purposes than a finitely extended gas surrounded by walls), which can be obtained as weak limits of states in the vacuum sector by removing the compensating particle to spatial infinity and which produce inequivalent representations called **superselection sectors**.

L. Foundation of Quantum Mechanics

Hilbert spaces and the underlying field of complex numbers, which constitute a mathematical background for quantum mechanics, are not immediately discernible from physical observations, and hence there are various attempts to find axioms for quantum mechanics that imply the usual mathematical structure and at the same time allow direct physical interpretation.

One approach of this kind focuses attention on the set of all observables that have only two possible measured values 1 (yes) and 0 (no), called **questions**, together with their order structure (logical implications) and associated lattice structure (join, meet, and orthocomplementation as logical sum, product, and negation). This is called **quantum logic** in contradistinction to the situation in classical physics, where it would form a Boolean lattice. The lattice of all orthogonal projections (corresponding to all closed subspaces of a Hilbert space) in quantum mechanics is a †complete, †orthocomplemented, **weakly modular** (also called **orthomodular**) †atomic lattice satisfying the **covering law**, where weak modularity means $c \wedge (c' \vee b) = b$ and $b \vee (b' \wedge c) = c$ whenever $b \leqslant c$, and the covering law means that every $b \neq 0$ possesses an atom p under it ($p \leqslant b$) and that if an atom q satisfies $q \wedge b = 0$, then any c between $q \vee b$ and b ($q \vee b \geqslant c \geqslant b$) is b or $q \vee b$. Conversely any such lattice is a di-

rect sum of irreducible ones, each of which, if of dimension (length of longest chain) > 3, can be obtained as the lattice of subspaces V satisfying $(V^{\perp})^{\perp}=V$ in a vector space over a (generally noncommutative) field with an anti-automorphic involution $*$, equipped with a nondegenerate Hermitian form. In this approach, an additional requirement is needed to restrict the underlying field and its $*$ to be more familiar ones, such as real, complex, or quaternion fields and their usual conjugations $*$. If that is done, then the set of all probability measures on the lattice (i.e., assignment of expectation values $0 \leqslant \mu(a) \leqslant 1$ for all elements a in the lattice such that $\mu(\bigvee_i a_i)=\sum_i \mu(a_i)$ if $a_i \perp a_j$ for all pairs $i \neq j$, $\mu(a) \geqslant 0$, $\mu(1)=1$) is exactly the restriction to questions of states $\rho(A)=\mathrm{tr}(\rho A)$ given by the density matrices ρ (**Gleason's theorem**).

It is also possible to characterize the set of all states equipped with the convex structure (mixtures) geometrically. The set of all states (without the normalization condition) of a finite-dimensional, formally real, irreducible †Jordan algebra over the field of reals (the positivity of a state φ is defined by $\varphi(a^2) \geqslant 0$) has been characterized as a transitively homogeneous self-dual cone in a finite-dimensional real vector space (a cone V is transitively homogeneous if the group of all nonsingular linear transformations leaving V invariant is transitive on the topological interior of V) by E. B. Vinberg (*Trans. Moscow Math. Soc.*, 12 (1963); 13 (1965)), where the relevant Jordan algebras were completely classified earlier by P. Jordan, J. von Neumann, and E. P. Wigner (*Ann. Math.*, 36 (1934)) as direct sums of the following irreducible ones: the Jordan algebra (with the product $A \circ B=(AB+BA)/2$) of all Hermitian $n \times n$ matrices over the real, complex, or quaternion field, all 3×3 Hermitian matrices over octanions, or the so-called **spin balls** (the set of all normalized states being a ball) linearly generated by the identity and γ_j $(j=1, \ldots, n)$ satisfying $\gamma_j \circ \gamma_k=0$ if $j \neq k$ and $\gamma_j^2=1$.

In infinite-dimensional cases, this type of characterization extends to the "natural" positive cones of vectors (A. Connes, *Ann. Inst. Fourier*, 24 (1974); J. Bellissard and B. Iochum, *Ann. Inst. Fourier*, 28 (1978)); while the convex cone of all states (without normalization) of Jordan algebras and C^*-algebras have been characterized in terms of a certain class of projections associated with faces of the cone, called P-projections, by E. M. Alfsen, F. W. Shultz, and others (*Acta Math.*, 140 (1978); 144 (1980)). In finite-dimensional cases, Araki (*Commun. Math. Phys.*, 75 (1980)) has given a characterization allowing direct physical interpretation by replacing P-projection with a notion of filtering corresponding to quantum-mechanical measurement.

Due to some features of quantum-mechanical measurement not in conformity with common sense, there have arisen **hidden variable theories** that are deterministic and reproduce the quantum-mechanical prediction. For a situation where a pair of (correlated) particles in states a and b are created and their spins (1 or -1, i.e., up or down spin) measured at positions distant from each other, the expectation value $E(a,b)$ for the product would be given in a hidden variables theory by $E(a,b)=\int A_a(\lambda)B_b(\lambda)\,d\rho(\lambda)$ for a probability measure ρ and the functions A_a and B_b of hidden variables λ, representing spins and hence satisfying $|A_a| \leqslant 1$ and $|B_b| \leqslant 1$. Then the following **Bell's inequality** holds:

$$|E(a,b)-E(a,b')+E(a',b)+E(a',b')| \leqslant 2.$$

This contradicts both quantum-mechanical predictions and experimental results, so that hidden variable theories of this type have been rejected.

References

[1] P. A. M. Dirac, Principles of quantum mechanics, Clarendon Press, fourth edition, 1958.
[2] A. Messiah, Quantum mechanics I, II, North-Holland, 1961, 1962. (Original in French, 1959.)
[3] W. Pauli, General principles of quantum mechanics, Springer, 1980. (Original in German, 1958.)
[4] S. Tomonaga, Quantum mechanics I, II, North-Holland, 1962–1966.
[5] J. von Neumann, Mathematical foundations of quantum mechanics, Princeton Univ. Press, 1955. (Original in German, 1932.)
[6] B. L. van der Waerden, Sources of quantum mechanics, Dover, 1967.
[7] B. L. van der Waerden, Group theory and quantum mechanics, Springer, 1974.
[8] H. Weyl, The theory of groups and quantum mechanics, Dover, 1949. (Original in German, 1928.)
[9] E. P. Wigner, Group theory and its application to the quantum mechanics of atomic spectra, Academic Press, 1959.
[10] B. Simon, Functional integration and quantum mechanics, Academic Press, 1979.
[11] G. G. Emch, Algebraic methods in statistical mechanics and quantum field theory, Wiley, 1972.
[12] T. Bastin (ed.), Quantum theory and beyond, Cambridge Univ. Press, 1971.

[13] B. d'Espagnat, Conceptual foundations of quantum mechanics, Benjamin, second edition, 1976.
Also → references to 150 Field Theory, 375 Scattering Theory, 377 Second Quantization, and 386 S-Matrices.

352 (XI.15) Quasiconformal Mappings

A. History

H. Grötzsch (1928) introduced quasiconformal mappings as a generalization of conformal mappings. Let $f(z)$ be a continuously differentiable homeomorphism with positive Jacobian between plane domains. The image of an infinitesimal circle $|dz|=$constant is an infinitesimal ellipse with major axis of length $(|f_z|+|f_{\bar z}|)|dz|$ and minor axis of length $(|f_z|-|f_{\bar z}|)|dz|$. When the ratio $K(z)=(|f_z|+|f_{\bar z}|)/(|f_z|-|f_{\bar z}|)$ is bounded, f is called **quasiconformal**. If $K\equiv 1$, then f is conformal. Grötzsch noticed that Picard's theorem still holds under the weaker condition; he determined the quasiconformal mappings between two given domains, which are not conformally equivalent to each other, providing the smallest $\sup K$, that is, those closest to conformality [1].

We cannot speak of the history of quasiconformal mappings without mentioning the discovery of extremal length by A. Beurling and L. V. Ahlfors (→ 143 Extremal Length), which has led to the precise definition for quasiconformality itself.

Quasiconformal mappings have less rigidity than conformal mappings, and for this reason they have been utilized for the type problem or the classification of open Riemann surfaces (Ahlfors, S. Kakutani, O. Teichmüller, K. I. Virtanen, Y. Tôki; → 367 Riemann Surfaces). Quasiconformal mappings have important applications in other fields of mathematics, e.g., in the theory of †partial differential equations of elliptic type (M. A. Lavrent'ev [2]) and especially in the problem of moduli of Riemann surfaces, including the theory of Teichmüller spaces (→ 416 Teichmüller Spaces). These applications are explained in Sections C and D.

B. Definitions

The current definitions of quasiconformality, which dispense with continuous differentiability, are due to Ahlfors [3], A. Mori [4], and L. Bers [5] (→ C. B. Morrey [6]). Consider an orientation-preserving topological mapping f of a domain D on the $z(=x+iy)$-plane. The quasiconformality of f is defined as follows. (1) (the geometric defintion) Let Q be a curvilinear quadrilateral, i.e., a closed Jordan domain with four specified points on the boundary, and let the interior of Q be mapped conformally onto a rectangular domain I. The ratio ($\geqslant 1$) of the sides of I, called the modulus of Q and denoted by $\operatorname{mod} Q$, is uniquely determined. If $\operatorname{mod} f(Q)\leqslant K \operatorname{mod} Q$ for any curvilinear quadrilateral Q in D, then f is called a **K-quasiconformal mapping** of D. This is equivalent to: (2) (the analytic definition) f is absolutely continuous on almost every line segment parallel to the coordinate axes contained in D (this condition is often referred to as ACL in D) and satisfies the inequality $|f_{\bar z}|\leqslant\dfrac{K-1}{K+1}|f_z|$ almost everywhere in D with some constant $K\geqslant 1$. When the value of K is irrelevant to the problem considered, K-quasiconformal mappings are simply said to be quasiconformal.

The K-quasiconformal mapping f satisfies the so-called **Beltrami differential equation**

$$f_{\bar z}=\mu f_z$$

almost everywhere in D with the measurable coefficient μ. The **maximal dilatation** $(1+\|\mu\|_\infty)/(1-\|\mu\|_\infty)$ does not exceed K. Sometimes f is called, for short a μ-conformal mapping. These notions are also defined for mappings between †Riemann surfaces, where the $(-1,1)$-form $\mu\, d\bar z\, dz^{-1}$ is independent of the choice of the local parameter z.

If in the above statements f is not necessarily topological but merely a continuous function satisfying the same requirements, we call it a **μ-conformal function**. (If in addition $\|\mu\|_\infty<(K-1)/(K+1)$, we call it a **$K$-quasiregular function** or **K-pseudoanalytic function**.) A μ-conformal function is represented as the composite $g\circ h$ of an analytic function g with a μ-conformal mapping h.

C. Principal Properties and Results

The inverse mapping of a K-quasiconformal mapping is also K-quasiconformal. The composite mapping $f_2\circ f_1$ of a K_1-quasiconformal mapping f_1 with a K_2-quasiconformal mapping f_2, if it can be defined, is K_1K_2-quasiconformal. A 1-quasiconformal mapping is conformal. Every quasiconformal mapping is †totally differentiable a.e. (almost everywhere), its Jacobian is positive a.e., and $(|f_z|+|f_{\bar z}|)/(|f_z|-|_{\bar z}|)\leqslant K$ a.e.

Let f be a K-quasiconformal mapping of $|z|<1$ onto $|w|<1$. Then f extends to a homeo-

morphism of $|z|\leqslant 1$ onto $|w|\leqslant 1$. If, furthermore, $f(0)=0$, then the Hölder condition

$$\left(\frac{|z_1-z_2|}{16}\right)^K \leqslant |f(z_1)-f(z_2)| \leqslant 16|z_1-z_2|^{1/K}$$

holds for $|z_1|\leqslant 1$, $|z_2|\leqslant 1$, and 16 is the best coefficient obtainable independently of K (Mori). This shows that any family of K-quasiconformal mappings of $|z|<1$ onto $|w|<1$ is †normal. For further properties and bibliography → O. Lehto and Virtanen [7] and Ahlfors [8].

(a) Boundary Correspondences and Extensions. Ahlfors and Beurling characterized the correspondence between $|z|=1$ and $|w|=1$ induced by f [9]. What amounts to the same thing, the following theorem holds: Let $\mu(x)$ be a real-valued monotone increasing continuous function on R such that $\lim_{x\to\pm\infty}\mu(x)=\pm\infty$. Then there exists a quasiconformal mapping of the upper half-plane $y>0$ onto itself with boundary correspondence $x\mapsto\mu(x)$ if and only if

$$\frac{1}{\rho}\leqslant\frac{\mu(x+t)-\mu(x)}{\mu(x)-\mu(x-t)}\leqslant\rho$$

for some constant $\rho\geqslant 1$ and for all $x, t\in\mathbf{R}$.

Theorem of quasiconformal reflection (Ahlfors [10]). Let L denote a curve which passes through ∞ and divides $\mathbf{C}\cup\{\infty\}$ into two domains Ω, Ω^* such that $\Omega\cup L\cup\Omega^*=\mathbf{C}\cup\{\infty\}$. Then there exists an orientation-reversing quasiconformal mapping of Ω onto Ω^* which keeps every point of L fixed if and only if some constant C exists satisfying $|\zeta_3-\zeta_1|/|\zeta_2-\zeta_1|\leqslant C$ for any three points $\zeta_1, \zeta_2, \zeta_3$ on L such that $\zeta_3\in\widehat{\delta_1\xi_2}$.

(b) Mapping Problem. Given a measurable function μ in a simply connected domain D with $\|\mu\|_\infty<1$, there exists a μ-conformal mapping of D onto a plane domain Δ which is unique up to conformal mappings of Δ [8]. When μ is real analytic and the derivatives of functions are defined in the usual manner, a classical result concerning the †conformal mapping of surfaces asserts the existence of a solution of Beltrami's differential equation $f_{\bar z}=\mu f_z$.

Concerning the dependence of μ-conformal mapping on μ, Ahlfors and Bers [11] obtained the following important result: Denote by f^μ a μ-conformal mapping of the whole finite plane onto itself that preserves 0 and 1. The space of functions μ has the structure of a Banach space with L_∞-norm, and the space of mappings f^μ also has the structure of a Banach space with respect to a suitable norm. If $\{\mu(t)=\mu(z;t)\}$ is a family of μ depending on the parameter t with $\|\mu(t)\|_\infty\leqslant k<1$ and $\mu(t)$ is continuous (resp. continuously differentiable, real analytic, complex analytic) in t, then $f^{\mu(t)}$ is also continuous (continuously differentiable, real analytic, complex analytic). For the proofs of these important results, which have opened up a new way to study †Teichmüller space, the extension and reflection of quasiconformal mappings are made essential use of.

(c) Extremal Quasiconformal Mappings. Let $K(f)$ denote the maximal dilatation of a quasiconformal mapping f. Suppose that a family $\mathscr{F}=\{f\}$ of quasiconformal mappings is given. If some $f_0\in\mathscr{F}$ exists such that $K(f_0)$ attains the infimum of $K(f)$ for all $f\in\mathscr{F}$, then f_0 is called an **extremal quasiconformal mapping** in $\mathscr{F}$.

Let $R=\{(x,y)\,|\,0<x<a, 0<y<b\}$, $R'=\{(x',y')\,|\,0<x'<a', 0<y'<b'\}$ be a pair of rectangular domains. Let $\mathscr{F}$ be the family of all quasiconformal mappings of R onto R' which map each vertex to a vertex with $(0,0)\mapsto(0,0)$. Then the unique extremal quasiconformal mapping for $\mathscr{F}$ is the affine mapping $x'=(a'/a)x$, $y'=(b'/b)y$ (Grötzsch [1]).

Next suppose that we are given two homeomorphic closed Riemann surfaces R, S and a †homotopy class $\mathscr{F}$ of orientation-preserving homeomorphisms of R onto S. Then $\mathscr{F}$ contains a unique extremal quasiconformal mapping. More precisely, either R and S are conformally equivalent to each other or else R admits an essentially unique analytic $(2,0)$-form Φ such that the respective local coordinates z, w of R, S satisfy the differential equation

$$(\partial w/\partial\bar z)/(\partial w/\partial z)=[(K-1)/(K+1)]\bar\Phi/|\Phi| \qquad (1)$$

with some constant $K>1$ everywhere on R, at which $\Phi\neq 0$ (Teichmüller [12], Ahlfors [3]). This turns out to be a generalization of Grötzsch's extremal affine mapping. The extremal mapping f satisfying equation (1) is sometimes referred to as the **Teichmüller mapping**.

Consider again a μ-conformal mapping g of the unit disk $D:|z|<1$ onto itself which induces a topological automorphism of the boundary $|z|=1$. If we define $\mathscr{F}$ as the family of all quasiconformal automorphisms f of D satisfying $f(e^{i\theta})=g(e^{i\theta})$, then the extremal quasiconformal mapping in $\mathscr{F}$ exists but is not always determined uniquely (K. Strebel [13]). As to the Teichmüller mapping, the uniqueness theorem is as follows: If the norm $\|\Phi\|=\iint_D|\Phi(z)|\,dx\,dy$ of Φ in (1) is finite, the Teichmüller mapping is the unique extremal quasiconformal mapping in $\mathscr{F}$. Otherwise, the uniqueness does not hold in general (Strebel [13]). On the other hand, a necessary and sufficient condition is proved for the Beltrami

coefficient μ of a quasiconformal mapping of $\mathscr{F}$ to be extremal (R. S. Hamilton [14], E. Reich and Strebel in [15]). Moreover, this last result can be extended to the extremal quasiconformal mapping between arbitrary Riemann surfaces.

D. Applications

In the earlier stage of development of this theory, quasiconformal mappings were applied only to the †type problem of simply connected Riemann surfaces and to the classification of Riemann surfaces of infinite genus (→ 367 Riemann Surfaces). This application is based on the fact that it is often possible to find a quasiconformal mapping with the prescribed boundary correspondence even when no equivalent conformal mapping exists and the fact that the classes O_G and O_{HD} (→ 367 Riemann Surfaces) of Riemann surfaces are invariant under quasiconformal mappings, as they are under conformal mappings.

It is worth remarking that the investigation of quasiconformal mappings is intimately connected with the recent development of the theory of †Kleinian groups via Teichmüller spaces.

The theory of quasiconformal mappings was also applied by Lavrent'ev [16] and Bers [2] to partial differential equations, particularly to those concerning the behavior of fluids. They utilized the fact that if the density and its reciprocal are bounded in a steady flow of a 2-dimensional †compressible fluid, then the mapping of the physical plane to the potential plane (the plane on which the values of the †velocity potential and the †stream function are taken as coordinates) is quasiconformal, and that if in addition the supremum of the †Mach number is smaller than 1, then the mapping from the physical plane to the †hodograph plane is pseudoanalytic.

E. Similar Notions

The term *quasiconformal* was used differently by Lavrent'ev, as follows: A topological mapping $f=u+iv$ is called **quasiconformal** with respect to a certain system of linear partial differential equations when u and v satisfy the system. This is a generalized definition because the system may not be equivalent to a Beltrami equation. However, it is reduced to a quasiconformal mapping if the system is uniformly elliptic. Bers used the term **pseudoanalytic** to describe a certain function related to linear partial differential equations of elliptic type. This function is pseudoanalytic in the sense of Section B on every relatively compact subset and has properties similar to those of analytic functions.

Analytic transformations in the theory of functions of several variables are called pseudoconformal by some mathematicians, and there is a similar term *quasi-analytic*. The latter is an entirely different notion from the one discussed in this article.

F. Generalization to Higher Dimensions

Let f be a continuous ACL-mapping of a subdomain G of $\mathbf{R}^n$ into $\mathbf{R}^n$ whose Jacobian matrix is denoted by $f'(x)$. Furthermore, the operator norm and the determinant of f' are denoted by $\|f'\|$ and $\det f'$, respectively. Then f is said to be quasiregular if all the partial derivatives of f are locally of class L^n on G and if there exists a constant $K\geqslant 1$ such that $(\|f'\|(x))^n\leqslant K\cdot\det f'(x)$ almost everywhere in G. The smallest $K\geqslant 1$ for which this inequality is true is called the outer dilatation of f and is denoted by $K_0(f)$. If f is quasiregular, then the smallest $K\geqslant 1$ for which the inequality $\det f'(x)\leqslant K\cdot[\min_{|y|=1}|f'(x+y)|^n]$ holds almost everywhere in G is called the inner dilatation of f and is denoted by $K_I(f)$. If $\max(K_I(f),K_0(f))\leqslant K'$, then f is said to be K'-quasiregular. An orientation-preserving mapping is called K-quasiconformal (J. Väisälä [17]) if it is a K-quasiregular homeomorphism. When $n=2$, these definitions agree with those given in Section B.

For $n\geqslant 3$ the following properties also still hold: A quasiregular mapping is discrete, open, totally differentiable a.e. and is absolutely continuous (O. Martio, S. Rickman, and Väisälä [18]). Quasiconformal extension of higher-dimensional half-spaces have been studied by Ahlfors and L. Carleson [15].

References

[1] H. Grötzsch, Über möglichst konforme Abbildungen von schlichten Bereichen, Ber. Verh. Sächs. Akad. Wiss. Leipzig, 84 (1932), 114–120.
[2] M. A. Lavrent'ev, Varational methods for boundary value problems for systems of elliptic equations, Noordhoff, 1963. (Original in Russian, 1962.)
[3] L. V. Ahlfors, On quasiconformal mappings, J. Analyse Math., 3 (1954), 1–58, 207–208.
[4] A. Mori, On quasi-conformality and pseudo-analyticity, Trans. Amer. Math. Soc., 84 (1957), 56–77.
[5] L. Bers, On a theorem of Mori and the definition of quasi-conformality, Trans. Amer. Math. Soc., 84 (1957), 78–84.

[6] C. B. Morrey, On the solutions of quasi-linear elliptic partial differential equations, Trans. Amer. Math. Soc., 43 (1938), 126–166.
[7] O. Lehto and K. I. Virtanen, Quasikonforme Abbildungen, Springer, 1965; English translation, Quasiconformal mappings in the plane, Springer, second edition, 1973.
[8] L. V. Ahlfors, Lectures on quasiconformal mappings, Van Nostrand, 1966.
[9] A. Beurling and L. V. Ahlfors, The boundary correspondence under quasiconformal mappings, Acta Math., 96 (1956), 125–142.
[10] L. V. Ahlfors, Quasiconformal reflections, Acta Math., 109 (1963), 291–301.
[11] L. V. Ahlfors and L. Bers, Riemann's mapping theorem for variable metrics, Ann. Math., (2) 72 (1960), 385–404.
[12] O. Teichmüller, Extremale quasikonforme Abbildungen und quadratische Differentiale, Abh. Preuss. Akad. Wiss. Math.-Nat. Kl., 22 (1939), 1–197.
[13] K. Strebel, Zur Frage der Eindeutigkeit extremaler quasikonformer Abbildungen des Einheitskreises, Comment. Math. Helv., 36 (1962), 306–323; 39 (1964), 77–89.
[14] R. S. Hamilton, Extremal quasiconformal mappings with prescribed boundary values, Trans. Amer. Math. Soc., 138 (1969), 399–406.
[15] L. V. Ahlfors et al., Contributions to analysis, Academic Press, 1974.
[16] L. Bers, Mathematical aspects of subsonic and transonic gas dynamics, Wiley, 1958.
[17] J. Väisälä, Lectures on n-dimensional quasiconformal mappings, Lecture notes in math. 229, Springer, 1971.
[18] O. Martio, S. Rickman, and J. Väisälä, Definitions for quasiregular mappings, Ann. Acad. Sci. Fenn., 448 (1969), 1–40.
[19] L. Bers, A new proof of a fundamental inequality for quasiconformal mappings, J. Analyse Math., 36 (1979), 15–30.

R

353 (XX.25) Racah Algebra

A. General Remarks

Racah algebra is a systematic method of calculating the †matrix element $(\psi, A\psi')$ in †quantum mechanics, where A is a dynamical quantity and ψ and ψ' are irreducible components of the state obtained by combining n †angular momenta. The angular momentum **j** has x-, y-, z-components j_x, j_y, j_z, respectively. Each component is i times the infinitesimal rotation around the respective axis and is the generator of the infinitesimal rotation for every irreducible component ψ. The addition of two angular momenta leads to a †tensor representation $D(j_1)\otimes D(j_2)$ of two †irreducible representations of the 3-dimensional rotation group. The problem is to decompose this tensor representation into the direct sum of irreducible representations.

B. Irreducible Representations of the Three-Dimensional Rotation Group

Irreducible representations of the group $SO(3)$ of 3-dimensional rotations can be obtained from irreducible representations $D(j)$ ($j=0,1,2,\ldots$) of its †universal covering group $SU(2)$ of 2×2 matrices with determinant 1, through the 2-fold covering isomorphism $SO(3)\cong SU(2)/\{\pm I\}$ (→ 60 Classical Groups I). The representation $D(j)$ ($j=0,1/2,1,3/2,\ldots$) of $SU(2)$ is the $2j$-fold tensor product $A\otimes\ldots\otimes A$ of $A\in SU(2)$ restricted to the totally symmetric part of the $2j$-fold tensor product space. Let $u=\binom{1}{0}$ and $v=\binom{0}{1}$ be a basis for the complex 2-dimensional space on which $SU(2)$ operates. The symmetrized tensor product of $(j+m)$-fold u and $(j-m)$-fold v multiplied by a positive normalization constant ($m=j,j-1,\ldots,-j$) defines an orthonormal basis of the representation space of $D(j)$, which we shall denote by $\psi(jm)$.

Decomposition of the tensor product of two irreducible representations $D(j_1)$ and $D(j_2)$ into irreducible components leads to

$$D(j_1)\otimes D(j_2)=\sum D(j),$$
$$j=j_1+j_2,\quad j_1+j_2-1,\ldots,|j_1-j_2|.$$

For the basis we can write

$$\psi(jm)=\sum_{m_1,m_2}\psi(j_1m_1)\psi(j_2m_2)(j_1m_1j_2m_2\,|\,j_1j_2jm),$$

and the coefficients are called the **Clebsch-Gordan coefficients** or **Wigner coefficients**. The vectors $\psi(jm)$ in each irreducible representation space are determined only up to an overall phase factor (a complex number of modulus 1). By a suitable choice of the resulting arbitrary phase (which may depend on j_1, j_2, j), the coefficients are given by

$$j_1+j_2\geqslant j\geqslant |j_1-j_2|,$$
$$(j_1m_1j_2m_2\,|\,j_1j_2jm)=\delta(m_1+m_2,m)$$
$$\times\sqrt{\frac{(2j+1)(j_1+j_2-j)!(j+j_1-j_2)!(j+j_2-j_1)!}{(j_1+j_2+j+1)!}}$$
$$\times\sum_\nu\left((-1)^\nu\frac{\sqrt{(j_1+m_1)!(j_1-m_1)!(j_2+m_2)!}}{\nu!(j_1+j_2-j-\nu)!(j_1-m_1-\nu)!}\right.$$
$$\left.\times\frac{\sqrt{(j_2-m_2)!(j+m)!(j-m)!}}{(j_2+m_2-\nu)!(j-j_2+m_1+\nu)!(j-j_1-m_2+\nu)!}\right).$$

They satisfy †orthogonality relations. Another concrete expression for the same coefficients, but of a different appearance, was obtained earlier by Wigner. Wigner introduced the **3j-symbol**, given by

$$\begin{pmatrix} j_1j_2j_3\\ m_1m_2m_3\end{pmatrix}=(-1)^{j_1-j_2-m_3}(2j_3+1)^{-1/2}$$
$$\times(j_1m_1j_2m_2\,|\,j_1j_2j_3\;\;-m_3)$$

for $m_1+m_2+m_3=0$ and zero otherwise. This is invariant under cyclic permutations of 1, 2, 3 and is multiplied by $(-1)^{j_1+j_2+j_3}$ under transpositions of indices as well as under the simultaneous sign change of all the m's. The $3j$-symbol multiplied by $(-1)^{j_2+j_3-j_1}$ is the V-coefficient of Racah.

There are two ways, $(D(j_1)\otimes D(j_2))\otimes D(j_3)$ and $D(j_1)\otimes(D(j_2)\otimes D(j_3))$, to reduce the tensor product of three irreducible representations, and two corresponding sets of basis vectors. The transformation coefficient for the two ways of reduction is written in the form

$$\langle j_1j_2(j_{12})j_3;j\,|\,j_1,j_2j_3(j_{23});j\rangle$$
$$=\sqrt{(2j_{12}+1)(2j_{23}+1)}\;W(j_1j_2j\;j_3;j_{12}j_{23}).$$

Here $W(abcd;ef)$, called the **Racah coefficient**, can be written as the sum of products of four Wigner coefficients. W has the following symmetry properties:

$$W(abcd;ef)=W(badc;ef)$$
$$=W(cdab;ef)$$
$$=W(acbd;fe)$$
$$=(-1)^{e+f-a-d}W(ebcf;ad)$$
$$=(-1)^{e+f-b-c}W(aefd;bc)$$

and satisfies an orthogonality relation. The **6j-symbol** $\left\{\begin{smallmatrix}abe\\cdf\end{smallmatrix}\right\}$ is related to the Racah coefficient by

$$W(abdc;ef)=(-1)^{a+b+c+d}\begin{Bmatrix}abe\\cdf\end{Bmatrix}.$$

C. Irreducible Tensors

A dynamical quantity T_q^k $(q=k,k-1,\ldots,-k)$ that transforms in the same way as the basis of $D(k)$ under rotation of coordinates is called an **irreducible tensor of rank** k. That is, it satisfies

$$[j_x \pm ij_y, T_q^k] = \sqrt{(k\mp q)(k\pm q+1)}\; T_{q\mp 1}^k,$$

$$[j_z, T_q^k] = qT_q^k.$$

Here $[a,b]=ab-ba$. The matrix element of this quantity between two irreducible components can be written in the form

$$(\alpha jm|T_q^k|\alpha' j' m')$$
$$=(1/\sqrt{2j+1})(\alpha j\|T^{(k)}\|\alpha' j')(j'm'kq|j'kjm),$$

where α is a parameter to distinguish multiple components with the same j, and components of different α are assumed to be orthogonal. In this formula the Clebsch-Gordan coefficients are determined from group theory, while $(\alpha j\|T^{(k)}\|\alpha' j')$ depends on the dynamics of the system.

When $T^{(k)}$ and $U^{(k)}$ operate only on the state vectors in the subspaces H_1 and H_2, respectively, of the total space (†Hilbert space) $H=H_1\times H_2$, their scalar product $(T^{(k)}, U^{(k)})=\sum_q(-1)^q T_q^k U_q^k$ has the matrix element

$$(\alpha_1\alpha_2 j_1 j_2 jm|(T^{(k)}, U^{(k)})|\alpha_1'\alpha_2' j_1' j_2' jm)$$
$$=(-1)^{j_1+j_2'-j}W(j_1 j_2 j_1' j_2'; jk)$$
$$\times(\alpha_1 j_1\|T^{(k)}\|\alpha_1' j_1')(\alpha_2 j_2\|U^{(k)}\|\alpha_2' j_2').$$

For an irreducible component of the tensor product of two irreducible tensors,

$$[T^{(k_1)}\otimes U^{(k_2)}]_q^k$$
$$=\sum_{q_1+q_2=q} T_{q_1}^{k_2} U_{q_2}^{k_2}(k_1 q_1 k_2 q_2|k_1 k_2 kq),$$

the matrix can be written as

$$(\alpha j_1 j_2 j\|[T^{(k_1)}\otimes U^{(k_2)}]^{(k)}\|\alpha' j_1' j_2' j')$$
$$=\sqrt{(2k+1)(2j+1)(2j'+1)}$$
$$\times\sum_{\alpha''}(\alpha j_1\|T^{(k_1)}\|\alpha'' j_1')(\alpha'' j_2\|U^{(k_2)}\|\alpha' j_2')$$
$$\times\begin{Bmatrix} j_1 & j_2 & j \\ j_1' & j_2' & j' \\ k_1 & k_2 & k \end{Bmatrix}.$$

The last factor, the **9*j*-symbol**, is defined as the matrix element between basis vectors of $[D(j_1)\times D(j_2)]\times[D(j_3)\times D(j_4)]$ and $[D(j_1)\times D(j_3)]\times[D(j_2)\times D(j_4)]$:

$$\langle j_1 j_2(j_{12}) j_3 j_4(j_{34}) jm| j_1 j_3(j_{13}) j_2 j_4(j_{24}) jm\rangle$$
$$=\sqrt{(2j_{12}+1)(2j_{34}+1)(2j_{13}+1)(2j_{24}+1)}$$
$$\times\begin{Bmatrix} j_1 & j_2 & j_{12} \\ j_3 & j_4 & j_{34} \\ j_{13} & j_{24} & j \end{Bmatrix}.$$

The 9*j*-symbol can be written as a weighted sum of the products of the three *W*'s.

See [6] and [7] for explicit formulas of Clebsch-Gordan coefficients and [8] for Racah coefficients.

References

[1] E. P. Wigner, Group theory and its application to the quantum mechanics of atomic spectra, Academic Press, 1959.

[2] M. E. Rose, Elementary theory of angular momentum, Wiley, 1957.

[3] U. Fano and G. Racah, Irreducible tensorial sets, Academic Press, 1959.

[4] A. R. Edmonds, Angular momentum in quantum mechanics, Princeton Univ. Press, 1959.

[5] F. Bloch, S. G. Cohen, A. De-Shalit, S. Sambutsky, and I. Talmi, Spectroscopic and group-theoretical methods in physics (Racah memorial volume), North-Holland, 1968.

[6] E. U. Condon and G. H. Shortley, Theory of atomic spectra, Cambridge Univ. Press, 1935 (reprinted with corrections, 1951).

[7] M. Morita, R. Morita, T. Tsukada, and M. Yamada, Clebsch-Gordan coefficients for $j_2=5/2$, 3, and 7/2, Prog. Theoret. Phys., Suppl., 26 (1963), 64–74.

[8] L. C. Biedenharn, J. M. Blatt, and M. E. Rose, Some properties of the Racah and associated coefficients, Rev. Mod. Phys., 24 (1952), 249–257.

354 (XVI.5) Random Numbers

A. General Remarks

A sequence of numbers that can be regarded as realizations of independent and identically distributed †random variables is called a sequence or table of **random numbers**. It is a basic tool for the †Monte Carlo method, †simulation of stochastic phenomena in nature or in society, and †sampling or †randomization techniques in statistics. Random numbers used in practice are pseudorandom numbers (→ Section B); theoretically, the definition of random numbers leads to an algorithmic approach to the foundations of probability [1,2].

B. Pseudorandom Numbers

Tables of numbers generated by random mechanisms have been statistically tested and

published. To generate random numbers on a large scale, electronic devices based on stochastic physical phenomena, such as thermoelectron noise or radioactivity, can be used. For digital computers, however, numbers generated by certain simple algorithms can be viewed practically as a sequence of random numbers; this is called a sequence of **pseudorandom numbers**.

Distribution of random numbers that are easily generated and suitable for general use is the continuous uniform distribution on the interval (0, 1), which is approximated by the discrete distribution on $\{0, 1, \ldots, N-1\}$ $(N \gg 1)$. Random numbers with distribution function $F(\cdot)$ are obtained by transforming uniform distributions by $F^{-1}(\cdot)$. For typical distributions, computation tricks avoiding the direct computation of $F^{-1}(\cdot)$ have been devised. Among them the use of †order statistics and acceptance-rejection techniques have wide applicability.

For the generation of uniform pseudorandom numbers on $\{0, 1, \ldots, N-1\}$, $N = n^s$ (n = a computer word length), the following algorithms are used. Each of them is written in terms of simple computer instructions. (1) The **middle-square method** was proposed by von Neumann. We square an integer of s digits of radix (or base) n and take out the middle s digits as the next term. We repeat this process and obtain a sequence of pseudorandom numbers. The sequence thus generated might be cyclic with a short period, possibly after many repetitions. The lengths from initial values to the terminal cycles are empirically checked. (2) The †Fibonacci sequence $\{u_n\}$ defined by $u_{k+1} \equiv u_{k-1} + u_k \pmod{n^s}$ is apparently regular, but it is uniformly distributed. (3) The **congruence method** [3]: Define a sequence by $u_{k+1} \equiv au_k + c \pmod{n^s}$ or $(\mathrm{mod}\, n^s \pm 1)$. If $c = 0$, the procedure is called the multiplicative congruence method, otherwise the mixed congruence method. The cycle, that is, the minimum k such that $u_k = u_0$, and the constants a, c, and u_0 that make the cycle maximum for given n and s are determined by number theory. The points $(u_{kl}, u_{kl+1}, \ldots, u_{kl+l-1})$, $k = 0, 1, 2, \ldots$, lie on a small number of parallel hyperplanes in the l-dimensional cube. Good choice of the constant a makes the sequence quite satisfactory. (4) H. Weyl considered sequences $f(k) = k\alpha \pmod 1$, where α is an irrational number and $k = 1, 2, \ldots$, whose values are uniformly distributed on the interval (0, 1). They are not independent, though they can be used for some special purposes. A modified sequence $x_k = k^2\alpha \pmod 1$ is known to be random for any irrational α in the sense that the †serial correlation $N^{-1}\sum_{k=1}^{N} x_k x_{k+l} - 1/4$ converges to 0 uniformly in l as $N \to \infty$.

C. Statistical Tests

To check uniform random numbers on (0, 1) the following tests are used: (1) Divide (0, 1) into subintervals; then the frequency of random numbers falling into these is a multinomial sample. †Goodness of fit can be tested by the †chi-square test; independence can be tested by observing the frequency of transitions of subintervals in which a pair of consecutive numbers falls, as well as by observing the overall properties, such as uniformity of the frequency of patterns of subintervals in which a set of random numbers falls. (2) For a set of random numbers, the distance of the empirical distribution function from that of the theoretical one is tested by the †Kolmogorov-Smirnov test. (3) Observe the rank orders of a set of random numbers, and test the randomness of their permutations (test the number of runs up and down).

D. Kolmogorov-Chaitin Complexity and Finite Random Sequences

As Shannon's entropy is a quantity for measuring the randomness of random variables, the Kolmogorov-Chaitin complexity [4, 5] is that of individual objects based on logic instead of probability. For constructive objects $x \in X$, $y \in Y$ and a partial recursive function $A: Y \times \{1, 2, \ldots\} \to X$, define

$$K_A(x \mid y) = \begin{cases} \min(\log_2 n \mid A(y, n) = x), \\ \infty \quad (\text{if } A(y, n) = x \text{ for no } n). \end{cases}$$

The function A is said to be **asymptotically optimal** if for any B there exists a constant C such that $K_A(x \mid y) \leqslant K_B(x \mid y) + C$ for any $x \in X$ and $y \in Y$. For an asymptotically optimal A, which is known to exist, $K_A(x \mid y)$ is simply denoted by $K(x \mid y)$ and is called the **Kolmogorov-Chaitin complexity** of x given y.

P. Martin-Löf [6] discussed a relation between complexity and randomness. Consider any statistical test for the randomness on the set of (say) finite decimal sequences which is effective in the sense that it has a finite algorithm. Then there exists a constant C independent of L and M such that

$$K(\xi_1, \ldots, \xi_L \mid L) \geqslant L \log_2 10 - M$$

implies the acceptance of the decimal sequence $\xi_1, \ldots, \xi_L$ by the test at the level $1 - 2^{-M-C}$. This condition on the complexity is satisfied by at least $(1 - 2^{-M})10^L$ sequences among the decimal sequences of length L.

E. Collective and Infinite Random Sequences

For finite sequences, the notion of randomness is obscure by nature. For infinite sequences,

however, clearer definition is possible. Based on the notion of collectives by R. von Mises, a definition of infinite random sequences has been given by A. Church [7]. A **selection function** is a $\{0,1\}$-valued function on the set of (say) finite decimal sequences such that $\{n, \varphi(\xi_1, \xi_2, \ldots, \xi_{n-1})=1\}$ is an infinite set for any infinite sequence $\xi_1, \xi_2, \ldots$. For a selection function φ and an infinite sequence $\xi_1, \xi_2, \ldots$, the **φ-subsequence** is defined as $\xi_{n_1}, \xi_{n_2}, \ldots$, where $\{n_1<n_2<\ldots\}=\{n, \varphi(\xi_1, \ldots, \xi_{n-1})=1\}$. For a class ψ of selection functions, an infinite decimal sequence is called a **ψ-collective** if each of the numbers $0, 1, \ldots, 9$ appears in it with a limiting relative frequency of 1/10, and the same thing holds for any φ-subsequence with $\varphi \in \psi$. By definition, a **random sequence** is a ψ-collective for the class ψ of recursive selection functions. Almost all real numbers are random in their decimal expansions.

F. Normal Numbers

Let $x-[x]=\sum x_n r^{-n}$ be the r-adic expansion of the fractional part of a real number x. For any ordered set $B_k=(b_1, \ldots, b_k)$ of numbers $0, 1, \ldots, r-1$, let $N_n(x, B_k)$ be the number of occurrences of the block B_k in the sequence $x_1, \ldots, x_n$. If $N_n(x, B_k)/n \rightarrow r^{-k}$ as $n \rightarrow \infty$ for every k and every B_k, then x is said to be **normal** to base r. Almost all real numbers are normal to any r. D. G. Champernowne [8] constructed a normal number given by the decimal expansion 0.1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, ... No one has so far been able to prove or disprove the normality of such irrational numbers as π, e, $\sqrt{2}$, $\sqrt{3}$, ... W. Schmidt [9] proved that the normality to base r implies the normality to base p if and only if $\log r/\log p$ is rational. A real number whose decimal expansion is random in the above sense is normal to base 10. For the converse, a necessary and sufficient condition for a selection function φ (for which $\varphi(\xi_1, \ldots, \xi_L)$ depends only on L) to have the property that the normality implies the $\{\varphi\}$-collectiveness has been obtained in [10].

References

[1] D. E. Knuth, The art of computer programming II, second edition, Addison-Wesley, 1981, ch. 3.
[2] E. R. Sowey, A second classified bibliography on random number generation and testing, Int. Statist. Rev., 46 (1978), 89–102.
[3] D. H. Lehmer, Mathematical methods in large-scale computing units, Proc. 2nd Symp. on Large-Scale Digital Calculating Machinery, Harvard Univ. Press, 1951, 141–146.
[4] A. N. Kolmogrov, Logical basis for information theory and probability theory, IEEE Trans. Information Theory, IT-14 (1968), 662–664.
[5] G. J. Chaitin, Algorithmic information theory, IBM J. Res. Develop., 21 (1977), 350–359.
[6] P. Martin-Löf, The definition of random sequences, Information and Control, 9 (1966), 602–619.
[7] A. Church, On the concept of a random sequence, Bull. Amer. Math. Soc., 47 (1940), 130–135.
[8] L. Kuipers and H. Niederreiter, Uniform distribution of sequences, Wiley-Interscience, 1974.
[9] W. Schmidt, On normal numbers, Pacific J. Math., 10 (1960), 661–672.
[10] T. Kamae, Subsequences of normal sequences, Israel J. Math., 16 (1973), 121–149.

355 (II.10) Real Numbers

A. Axioms for the Real Numbers

The set **R** of all **real numbers** has the following properties:

(1) Arithmetical properties: (i) For each pair of numbers $x, y \in \mathbf{R}$, there exists one and only one number $w \in \mathbf{R}$, called their **sum** and denoted by $x+y$, for which $x+y=y+x$ (commutative law) and $(x+y)+z=x+(y+z)$ (associative law) hold. Furthermore, there exists a unique number 0 (**zero**) such that $x+0=x$ for every $x \in \mathbf{R}$ (existence of †zero element). Also, for each x, there exists one and only one number $-x \in \mathbf{R}$ for which $x+(-x)=0$. (ii) For each pair of numbers $x, y \in \mathbf{R}$, there exists one and only one number $w \in \mathbf{R}$, called their **product** and denoted by xy, for which $xy=yx$ (commutative law), $(xy)z=x(yz)$ (associative law), and $(x+y)z=xz+yz$ (distributive law) hold. Furthermore, there exists a unique number 1 (**unity**) $\in \mathbf{R}$ such that $1x=x$ for every $x \in \mathbf{R}$ (existence of †unity element). Also, for each $x \neq 0$ ($x \in \mathbf{R}$) there exists one and only one number $x^{-1} \in \mathbf{R}$ for which $xx^{-1}=1$. Owing to properties (i) and (ii), all †four arithmetic operations obey the usual laws (with the single exception of division by zero); in other words, **R** is a †field.

(2) Order properties: (i) For each $x, y \in \mathbf{R}$, one and only one of the following three relations holds: $x<y$, $x=y$, or $x>y$. With $x \leqslant y$ meaning $x<y$ or $x=y$, the relation $\leqslant$ obeys the transitive law: $x \leqslant y$ and $y \leqslant z$ imply $x \leqslant z$, which makes **R** †totally ordered. (ii) Order and

arithmetical properties are related by: $x \leqslant y$ implies $x+z \leqslant y+z$ for any $z \in \mathbf{R}$, and $x \leqslant y$ and $0 \leqslant z$ imply $xz \leqslant yz$; in other words, **R** is an †ordered field.

In particular, $x \in \mathbf{R}$ with $x>0$ is called a **positive number**, and $x \in \mathbf{R}$ with $x<0$ a **negative number**. We write $|x|=x$ if $x \geqslant 0$ and $|x|=-x$ if $x<0$, and call $|x|$ the **absolute value** of x.

(3) Continuity property: If nonempty subsets A and B of **R**, with $a<b$ for each pair $a \in A$ and $b \in B$, satisfy $\mathbf{R}=A \cup B$ and $A \cap B=\varnothing$ (empty set), then the pair (A, B) of sets is called a **cut** of **R**. For each cut (A, B) of **R**, there exists a number $x \in \mathbf{R}$ (necessarily unique) such that for every $a \in A$, $a \leqslant x$, and for every $b \in B$, $b \geqslant x$ (i.e., $x=\sup A=\inf B$). This property of **R** is called **Dedekind's axiom of continuity** (→ 294 Numbers).

The set **R** of all real numbers is determined uniquely up to an isomorphism, with respect to arithmetic operations and ordering, by properties (1)–(3). **R** forms an additive Abelian group; its subgroup $\{0, \pm 1, \pm 2, \ldots, \pm n, \ldots\}$ generated by 1 can be identified with the group **Z** of integers. The subset of all positive integers $\{1, 2, \ldots, n, \ldots\}$ may be identified with the set **N** of all natural numbers. The subset $\{m/n \mid m, n \in \mathbf{Z}, n \neq 0\}$ of **R** forms the subfield of **R** generated by 1. It can be identified with the field **Q** of all rational numbers. A real number that is not rational is called an **irrational number**.

B. Properties of Real Numbers

(1) For each pair of positive numbers a and $b>a$, there exists a natural number n with $a<nb$ (**Archimedes' axiom**).

(2) For each pair of positive real numbers a and b with $a<b$, there exists a rational number x such that $a<x<b$ (**denseness of rational numbers**).

(3) For any subset A in **R** †bounded from above (below), the †least upper bound of A: $a=\sup A$ (†greatest lower bound of A: $b=\inf A$) exists.

Let $\{a_n\}$ be a †sequence of real numbers. Assume that for each arbitrary positive number ε there exists a number n_0 such that $|a_n-b|<\varepsilon$ for all $n>n_0$. Then we write $\lim_{n\to\infty} a_n=b$ (or $a_n \to b$) and call b the **limit** of $\{a_n\}$. We also say that $\{a_n\}$ is a **convergent sequence** or that a_n **converges** to b.

(4) If for two sequences $\{a_n\}$, $\{b_n\}$, we have $a_1 \leqslant a_2 \leqslant \ldots \leqslant a_n \leqslant \ldots \leqslant b_n \leqslant \ldots \leqslant b_2 \leqslant b_1$ and $\lim(b_n-a_n)=0$, then there exists one and only one number $c \in \mathbf{R}$ with $\lim a_n=\lim b_n=c$ (**principle of nested intervals**).

(5) Let $\{a_n\}$ be a sequence of real numbers. If for any arbitrary positive ε there exists a natural number n_0 satisfying $|a_m-a_n|<\varepsilon$ for all $m, n>n_0$, then $\{a_n\}$ is called a **fundamental sequence** or **Cauchy sequence**. Any fundamental sequence of real numbers is convergent (**completeness of real numbers**).

For a set with properties 1 and 2 of Section A, it can be proved that property 3 of Section A is equivalent to property 3, or properties 1 and 4, or properties 1 and 5 of this section.

C. Intervals

For two numbers $a, b \in \mathbf{R}$ with $a<b$, we write

$$(a, b)=\{x \mid a<x<b\},$$

$$(a, b]=\{x \mid a<x \leqslant b\},$$

$$[a, b)=\{x \mid a \leqslant x<b\},$$

$$[a, b]=\{x \mid a \leqslant x \leqslant b\},$$

and call them (**finite**) **intervals**, of which a and b are their **left** and **right endpoints**, respectively. Specifically, (a, b) is called an **open interval** and $[a, b]$ a **closed interval**. The symbols ∞ and $-\infty$ are introduced as satisfying $\infty>x$, $x>-\infty$, $\infty>-\infty$ for all $x \in \mathbf{R}$. Writing $+\infty$ for ∞, we call $+\infty$ and $-\infty$ **positive infinity** and **negative infinity**, respectively. To extend the concept of intervals, we define $(-\infty, b)=\{x \mid x<b, x \in \mathbf{R}\}$, $(-\infty, b]=\{x \mid x \leqslant b, x \in \mathbf{R}\}$, $(a, \infty)=\{x \mid a<x, x \in \mathbf{R}\}$, $[a, \infty)=\{x \mid a \leqslant x, x \in \mathbf{R}\}$, and $(-\infty, \infty)=\mathbf{R}$, and call them **infinite intervals**.

Let $\{a_n\}$ be a sequence of real numbers. If for each infinite interval (a, ∞) ($(-\infty, a)$) there exists a number n_0 such that $a_n \in (a, \infty)$ ($a_n \in (-\infty, a)$) for all $n>n_0$, then we write $a_n \to +\infty$ ($a_n \to -\infty$) and call ∞ ($-\infty$) the limit of a_n, denoted as before by $\lim a_n$.

D. Topology of R

With the collection of all its open intervals (a, b) as an †open base, **R** is a †topological space (†order topology) that satisfies the †separation axioms $\mathbf{T}_2$, $\mathbf{T}_3$, $\mathbf{T}_4$. In **R** every (finite or infinite) interval (including **R** itself) is †connected, and the set **Q** of rational numbers is dense. A necessary and sufficient condition for a subset F of **R** to be †compact is that F be bounded and closed (**Weierstrass's theorem**). In particular, any finite closed interval is compact. **R** is a locally compact space satisfying the second †countability axiom. Further, any (finite or infinite) open interval is homeomorphic to **R**. The topology of **R** may also be

defined by the notion of convergence (→ 87 Convergence).

Arithmetic operations in **R** are all continuous: If $a_n \to a$ and $b_n \to b$, then $a_n + b_n \to a + b$, $a_n - b_n \to a - b$, $a_n b_n \to ab$, and $a_n/b_n \to a/b$ (where $b \neq 0, b_n \neq 0$). Hence **R** is a †topological field (regarding the characterization of **R** as a topological group or a topological field → 422 Topological Abelian Groups).

R as a topological Abelian group (with respect to addition) is isomorphic to the topological Abelian group $\mathbf{R}^+$ of all positive real numbers with respect to multiplication. To be precise, there exist topological mappings $f: \mathbf{R} \to \mathbf{R}^+$ with $f(x+y) = f(x)f(y)$ and $g: \mathbf{R}^+ \to \mathbf{R}$ with $g(xy) = g(x) + g(y)$. If $f(1) = a$, $g(b) = 1$, then f, g are uniquely determined and are written $f(x) = a^x$, $g(x) = \log_b x$.

Regarding **R** as a topological Abelian group (with respect to addition), any proper closed subgroup Γ of **R** is discrete and isomorphic to the additive group **Z** of integers. That is, for some $\varepsilon > 0$ we have $\Gamma = \{n\varepsilon \mid n \in \mathbf{Z}\}$. In particular, the quotient group $\mathbf{R}/\mathbf{Z}$ as a topological group is isomorphic to the rotation group of a circle (1-dimensional †torus group). Elements of $\mathbf{R}/\mathbf{Z}$ are called **real numbers** mod 1.

E. The Real Line

Let l be a Euclidean straight line considered to lie horizontally, say from left to right. Let p_0, p_1 be two distinct points on l, with p_0 situated to the left of p_1. Then there exists one and only one bijection φ from the set L of all points of l to **R** satisfying (i) $\varphi(p_0) = 0$, $\varphi(p_1) = 1$; (ii) if p lies to the left of q, then $\varphi(p) < \varphi(q)$; and (iii) for two line segments pq and $p'q'$ (where p and p' are to the left of q and q', respectively), $pq \equiv p'q'$ (pq and $p'q'$ are †congruent) $\Leftrightarrow \varphi(q) - \varphi(p) = \varphi(q') - \varphi(p')$. Then $\varphi(p)$ is called the **coordinate** of the point p, and (p_0, p_1) the **frame** of the line l. A Euclidean straight line with a fixed frame is called a **real line** (identified with **R** by the mapping φ) and is usually denoted by the same notation **R** or $\mathbf{R}^1$.

References

[1] N. Bourbaki, Eléments de mathématique III. Topologie générale, ch. 4. Nombres réels, Actualités Sci. Ind., 1143c, Hermann, third edition, 1960; English translation, General topology pt. 1, Addison-Wesley, 1966.
[2] G. Cantor, Gesammelte Abhandlungen, Springer, 1932.
[3] R. Dedekind, Gesammelte mathematische Werke I–III, Braunschweig, 1930–1932.
[4] J. Dieudonné, Foundations of modern analysis, Academic Press, 1960, enlarged and corrected printing, 1969.
[5] K. Weierstrass, Gesammelte Abhandlungen 1–7, Mayer & Miler, Akademische Verlag., 1894–1927.
Also → references to 381 Sets.

356 (I.9) Recursive Functions

A. General Remarks

A function whose †domain and †range are both the set of natural numbers $\{0, 1, 2, \ldots\}$ is called a **number-theoretic function**. In this article, the term *natural number* is used to mean a non-negative integer. Hilbert (1926) and K. Gödel [1] considered certain number-theoretic functions, called recursive functions by them and now called **primitive recursive functions** after S. C. Kleene [2] (the definition is given in Section B). Gödel introduced an efficient method of arithmetizing metamathematics based on representing certain finitary procedures in metamathematics by primitive recursive functions. Then the following problem naturally arises: How shall we define a finitary method? In other words, how shall we characterize a number-theoretic function that is effectively computable, or provided with an algorithm of computation? Gödel defined the notion of general recursive function by introducing a formal system for the elementary calculation of functions, following the suggestion given by J. Herbrand. Kleene later improved Gödel's definition and developed the theory of general recursive functions [2]. Furthermore, A. Church and Kleene defined λ-calculable functions using the λ-notation (Church [4]), and E. L. Post and A. M. Turing defined the notion of computable functions by introducing the concept of Turing machines. These notions, introduced independently and almost simultaneously, were found to be equivalent. Hence such functions are now simply called recursive functions. Here, instead of giving the definition of recursive functions in the original style (the Herbrand-Gödel-Kleene definition), we give it by utilizing the idea of introducing schemata, a natural extension of the notion of primitive recursive functions. We employ the letters x, y, z, x_1, x_2, ... for variables ranging over the natural numbers.

B. Primitive Recursive Functions

Consider the following five definition schemata:

(I) $\varphi(x)=x' \quad (=x+1)$,

(II) $\varphi(x_1,\ldots,x_n)=q$ (q a given natural number),

(III) $\varphi(x_1,\ldots,x_n)=x_i \quad (1\leqslant i\leqslant n)$,

(IV) $\varphi(x_1,\ldots,x_n)=\psi(\chi_1(x_1,\ldots,x_n),\ldots,\chi_m(x_1,\ldots,x_n))$,

(V) $\varphi(0,x_2,\ldots,x_n)=\psi(x_2,\ldots,x_n)$,

$\varphi(y',x_2,\ldots,x_n)=\chi(y,\varphi(y,x_2,\ldots,x_n),x_2,\ldots,x_n)$,

where $\psi(\ )$ is a constant natural number if $n=1$. A function is called **primitive recursive** if it is definable by a finite series of applications of the operations (IV) and (V) ($\psi,\chi,\chi_1,\ldots,\chi_m$ are already-introduced functions) starting from functions each of which is given by (I), (II), or (III). Given the functions $\psi_1,\ldots,\psi_l$, we define the **relativization** (with respect to $\psi_1,\ldots,\psi_l$) of the definition of primitive recursive functions as follows: A function is called **primitive recursive in** $\psi_1,\ldots,\psi_l$ if it is definable by a finite series of applications of (IV) and (V) starting from $\psi_1,\ldots,\psi_l$ and from functions each of which is given by (I), (II), or (III).

We say that a function $\varphi(x_1,\ldots,x_n)$ is the **representing function** of a †predicate $P(x_1,\ldots,x_n)$ if φ takes only 0 and 1 as values and satisfies

$$P(x_1,\ldots,x_n)\Leftrightarrow\varphi(x_1,\ldots,x_n)=0.$$

Then we call P a **primitive recursive predicate** if its representing function φ is primitive recursive. The following functions and predicates are examples of primitive recursive ones: $a+b$, $a\cdot b$, a^b, $a!$, $\min(a,b)$, $\max(a,b)$, $|a-b|$, $a=b$, $a<b$, $a|b$ (a divides b), $\Pr(a)$ (a is a prime number), p_i (the $(i+1)$st prime number, $p_0=2$, $p_1=3,\ldots$), $(a)_i$ (the exponent of p_i of the unique factorization of a into prime numbers if $a\neq0$; otherwise, 0).

Whenever we are given a concept or a theorem, we always transform it by replacing the predicates contained in it (if any) by corresponding representing functions. Then an operation Ω is called **primitive recursive** if the function or the predicate $\Omega(\psi_1,\ldots,\psi_l,Q_1,\ldots,Q_m)$ that results from the application of Ω to functions $\psi_1,\ldots,\psi_l$ and predicates $Q_1,\ldots,Q_m$ ($l,m\geqslant0,l+m>0$) is primitive recursive in $\psi_1,\ldots,Q_m$. Put $\varphi(x_1,\ldots,x_n,z)=\sum_{y<z}\psi(x_1,\ldots,x_n,y)$. Then φ is primitive recursive in ψ, and the finite sum $\sum_{y<z}$ is a primitive recursive operation in this sense.

Similarly, the following operations are primitive recursive: the finite product $\prod_{y<z}$, the logical connectives $\neg$, $\vee$, $\wedge$, $\rightarrow$ (→ 411 Symbolic Logic), definitions by cases, the **bounded quantifiers** $\exists y_{y<z}$, $\forall y_{y<z}$, and the **bounded μ-operator** $\mu y_{y<z}$ defined as follows: $\mu y_{y<z}R(x,y)$ is the least y such that $y<z$ and $R(x,y)$ holds, if there exists such a number y; otherwise, it equals z. The following operation is also primitive recursive:

$$\varphi(y,x_2,\ldots,x_n)=\chi(y,\tilde{\varphi}(y;x_2,\ldots,x_n),x_2,\ldots,x_n),$$

where

$$\tilde{\varphi}(y;x_2,\ldots,x_n)=\prod_{i<y}p_i^{\varphi(i,x_2,\ldots,x_n)}.$$

A function φ is said to be primitive recursive **uniformly** in $\psi_1,\ldots,\psi_l$ when φ is definable by applying a primitive recursive operation to $\psi_1,\ldots,\psi_l$.

Almost all results mentioned in this section were given by Gödel [1]. There are further investigations on primitive recursive functions by R. Péter (1934), R. Robinson (1947), and others. Note that a function definable by a †double recursion is not necessarily primitive recursive. Péter (1935, 1936) investigated in detail functions that are definable, in general, by k-fold recursions for every positive interger k [8].

C. General Recursive Functions

The following μ-operator is used to define general recursive functions by extending primitive recursive functions. For a predicate $R(y)$ on the natural numbers, $\mu yR(y)$ is the least y such that $R(y)$, if $\exists yR(y)$; otherwise, $\mu yR(y)$ is undefined. Generally, $\mu y(\psi(x_1,\ldots,x_n,y)=0)$ is not necessarily defined for each n-tuple $(x_1,\ldots,x_n)$ of natural numbers. Now, a function is called a **general recursive function** (or simply **recursive function**) if it is definable by a series of applications of schemata including a new schema

(VI) $\varphi(x_1,\ldots,x_n)=\mu y(\psi(x_1,\ldots,x_n,y)=0)$

for the definition of φ from any function ψ that satisfies

$$\forall x_1\ldots\forall x_n\exists y(\psi(x_1,\ldots,x_n,y)=0),$$

in addition to those used to define the primitive recursive functions. Thus, by definition, a primitive recursive function is general recursive. **A general recursive predicate** is a predicate such that its representing function is general recursive. The facts, including the ones concerning relativization, that are valid for pri-

mitive recursive functions are also valid for general recursive functions.

Kleene's Normal Form Theorem. For each n, we can construct a primitive recursive predicate $T_n(z, x_1, \ldots, x_n, y)$ and a primitive recursive function $U(y)$ such that given any general recursive function $\varphi(x_1, \ldots, x_n)$, a natural number e can be found such that

$$\forall x_1 \ldots \forall x_n \exists y T_n(e, x_1, \ldots, x_n, y), \tag{1}$$

$$\varphi(x_1, \ldots, x_n) = U(\mu y T_n(e, x_1, \ldots, x_n, y)). \tag{2}$$

Any natural number e for which (1) and (2) hold is said to **define** φ **recursively** or to be a **Gödel number** of a recursive function φ. Let $\psi_1, \ldots, \psi_l$ (abbreviated Ψ) be any given functions. We can relativize Kleene's normal form theorem with respect to them as follows: For each n, we can construct a predicate $T_n^{\Psi}(z, x_1, \ldots, x_n, y)$ that is primitive recursive in Ψ such that given any function φ that is general recursive in Ψ, a natural number e can be found such that

$$\forall x_1 \ldots \forall x_n \exists y T_n^{\Psi}(e, x_1, \ldots, x_n, y), \tag{3}$$

$$\varphi(x_1, \ldots, x_n) = U(\mu y T_n^{\Psi}(e, x_1, \ldots, x_n, y)), \tag{4}$$

where $U(y)$ is the primitive recursive function mentioned in Kleene's normal form theorem. A natural number e for which (3) and (4) hold is said to **define** φ **recursively in** Ψ or to be a **Gödel number** of φ from Ψ. In particular, a Gödel number e of φ from Ψ can be found independently of Ψ (except for l and the respective numbers of variables of $\psi_1, \ldots, \psi_l$) when φ is general recursive uniformly in Ψ.

Now let S be a †formal system containing ordinary number theory. A number-theoretic predicate $P(x_1, \ldots, x_n)$ is said to be **decidable** within S if there is a formula $\mathrm{P}(\mathrm{a}_1, \ldots, \mathrm{a}_n)$ (with no †free variables other than the distinct variables $\mathrm{a}_1, \ldots, \mathrm{a}_n$) of S such that for each n-tuple $(x_1, \ldots, x_n)$ of natural numbers (the symbol $\vdash$ means provable in S),

(i) $\vdash \mathrm{P}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ or $\vdash \neg \mathrm{P}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$

and

(ii) $P(x_1, \ldots, x_n) \Leftrightarrow \vdash \mathrm{P}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$,

where $\mathbf{x}_1, \ldots, \mathbf{x}_n$ designate the numerals corresponding to $x_1, \ldots, x_n$ in S. If S is a consistent system such that primitive recursive predicates are decidable within S and the predicates Pf_A (for any formula A, $Pf_A(x_1, \ldots, x_n, y)$ means that y is the Gödel number of a proof of $\mathrm{A}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$) are primitive recursive, then a necessary and sufficient condition for P to be decidable within S is that P is a general recursive predicate (A. Mostowski, 1947).

Church (1936) proposed the following statement: Every **effectively calculable function** is a general recursive function. The converse of this is evidently true by the definition of recursiveness. So **Church's thesis** and its converse provide the exact definition of the notion of effectively computable functions. Though this notion is somewhat vague and intuitive, the definition seems to be satisfactory, as mentioned at the beginning of this article. Therefore, any function with a computation procedure or **algorithm** can be assumed to be general recursive. Utilizing this, various decision problems have been negatively solved (→ 97 Decision Problem). Furthermore, traditional descriptive set theory can be reinvestigated from this point of view, and the concept of effectiveness used in †semi-intuitionism is clarified using general recursive functions (→ 22 Analytic Sets).

D. Recursive Enumerability

A set $\{\varphi(0), \varphi(1), \varphi(2), \ldots\}$ enumerated by a general recursive function φ (allowing repetitions) is called a **recursively enumerable set**. The empty set is also considered recursively enumerable. It is known that in this definition "general recursive" can be replaced by "primitive recursive" (J. B. Rosser, 1936). A set E of natural numbers is recursively enumerable if and only if there is a primitive recursive predicate $R(x, y)$ such that $x \in E \Leftrightarrow \exists y R(x, y)$ (Kleene [2]).

Generally, a predicate $E(x_1, \ldots, x_m)$ is called a **recursively enumerable predicate** if there is a general recursive predicate $R(x_1, \ldots, x_m, y_1, \ldots, y_n)$ such that $E(x_1, \ldots, x_m) \Leftrightarrow \exists y_1 \ldots \exists y_n R(x_1, \ldots, x_m, y_1, \ldots, y_n)$. (Here "general recursive" can be replaced by "primitive recursive.")

We call a set E a **recursive set** if the predicate $x \in E$ is general recursive. The set $C = \{x \mid \exists y T_1(x, x, y)\}$ is an example of a set that is recursively enumerable but not recursive, and it has the following remarkable property: For every recursively enumerable set E, there is a primitive recursive function such that $x \in E \Leftrightarrow \varphi(x) \in C$. In this sense, the set C is said to be **complete** for the class of recursively enumerable sets. Post's problem, which asked whether the sets that are recursively enumerable but not recursive have the same †degree of (recursive) unsolvability as that of C, was negatively solved simultaneously by R. M. Friedberg (1957) and A. A. Muchnik (1956–1958). A recursively enumerable set E is general recursive if and only if there is a general recursive predicate $R(x, y)$ such that $x \notin E \Leftrightarrow \exists y R(x, y)$ (Kleene [5], Post [6]).

E. Partial Recursive Functions

A function $\varphi(x_1, \ldots, x_n)$ is called a **partial function** if it is not necessarily defined for all n-tuples $(x_1, \ldots, x_n)$ of natural numbers. For two partial functions $\psi(x_1, \ldots, x_n)$ and $\chi(x_1, \ldots, x_n)$, $\psi(x_1, \ldots, x_n) \simeq \chi(x_1, \ldots, x_n)$ means that if either $\psi(x_1, \ldots, x_n)$ or $\chi(x_1, \ldots, x_n)$ is defined for $x_1, \ldots, x_n$, so is the other, and the values are the same. For any given natural number e, $\varphi(x_1, \ldots, x_n) \simeq U(\mu y T_n(e, x_1, \ldots, x_n, y))$ (or $\simeq U(\mu y T_n^{\Psi}(e, x_1, \ldots, x_n, y)))$ is a partial function, in general. We say that such a function is **partial recursive (partial recursive (uniformly) in** Ψ) and that a natural number e defines φ **recursively ((uniformly) in** Ψ) or is a **Gödel number** of a partial recursive function φ (from Ψ). When a natural number e is a Gödel number of $\varphi(x_1, \ldots, x_n)$ (a Gödel number of φ from Ψ), $\varphi(x_1, \ldots, x_n)$ is sometimes written as $\{e\}(x_1, \ldots, x_n)(\{e\}^{\Psi}(x_1, \ldots, x_n))$. If a predicate $R(x_1, \ldots, x_n, y)$ is general recursive, then $\mu y R(x_1, \ldots, x_n, y)$ is partial recursive. Therefore, $\{z\}(x_1, \ldots, x_n)$ is a partial recursive function of the variables of z and of $x_1, \ldots, x_n$.

On the partial recursive functions, the following two theorems, given by Kleene [3], are most useful. (1) For natural numbers m, n, a primitive recursive function $S_n^m(z, y_1, \ldots, y_m)$ can be found such that, for any natural number e, $\{e\}(y_1, \ldots, y_m, x_1, \ldots, x_n) \simeq \{S_n^m(e, y_1, \ldots, y_m)\}(x_1, \ldots, x_n)$. (2) For any partial recursive function $\psi(z, x_1, \ldots, x_n)$, a natural number e can be found such that $\{e\}(x_1, \ldots, x_n) \simeq \psi(e, x_1, \ldots, x_n)$.

The notion of partial recursive functions appeared first in the theory of †constructive ordinal numbers of Church and Kleene (1963). Partial recursive functions can be defined in the Herbrand-Gödel-Kleene style as a natural extension of general recursive functions, and they are also definable by a finite series of applications of the schemata (IV), (V), and (VI) (in each schema, $=$ used for the definition of φ should be replaced by $\simeq$) starting from functions given by (I), (II), and (III).

F. Extension of Recursive Functions to Number-Theoretic Functionals

Let $\alpha_1, \ldots, \alpha_m$ be number-theoretic functions of one variable. If $\varphi(x_1, \ldots, x_n)$ is (partial) recursive uniformly in $\alpha_1, \ldots, \alpha_m$, then a Gödel number e of φ is found independently of $\alpha_1, \ldots, \alpha_m$, and $\varphi(x_1, \ldots, x_n)$ is expressed as $U(\mu y T_n^{\alpha_1 \cdots \alpha_m}(e, x_1, \ldots, x_n, y))$. Now suppose that $\alpha_1, \ldots, \alpha_m$ range over the set N^N of all number-theoretic functions of one variable, and put

$$\varphi(\alpha_1, \ldots, \alpha_m, x_1, \ldots, x_n)$$
$$\simeq U(\mu y T_n^{\alpha_1 \cdots \alpha_m}(e, x_1, \ldots, x_n, y)).$$

We call such a functional $\varphi(\alpha_1, \ldots, \alpha_m, x_1, \ldots, x_n)$ (partial) recursive, and with it we can develop a theory of recursive functions of variables of two types.

Extending the notion of recursive functionals, Kleene introduced and investigated the recursive functionals of variables of arbitrary (finite) types [10, 11]. The natural numbers are the **objects of type** 0, and the one-place functions from type-j objects to natural numbers are **objects of type** $j+1$. Denote variables ranging over the type-j objects by α^j, $\beta^j, \gamma^j, \ldots$, or $\alpha_1^j, \alpha_2^j, \alpha_3^j, \ldots$, etc. Consider a functional (simply called a function) of a given finite number of such variables of types taking natural numbers as values. A function φ is called a **primitive recursive function** if it is definable by a finite series of applications of the following schemata (I)–(VIII), where a is a variable of type 0, $\mathfrak{b}$ is any list (possibly empty) of variables that are mutually distinct and different from the other variables of the schema, and ψ, χ are given functions of the indicated variables. (I) $\varphi(a, \mathfrak{b}) = a'$; (II) $\varphi(\mathfrak{b}) = q$ (q is a natural number); (III) $\varphi(a, \mathfrak{b}) = a$; (IV) $\varphi(\mathfrak{b}) = \psi(\chi(\mathfrak{b}), \mathfrak{b})$; (V) $\varphi(0, \mathfrak{b}) = \psi(\mathfrak{b})$, $\varphi(a', \mathfrak{b}) = \chi(a, \varphi(a, \mathfrak{b}), \mathfrak{b})$; (VI) $\varphi(\mathfrak{a}) = \psi(\mathfrak{a}_1)$ ($\mathfrak{a}_1$ is a list of variables from which $\mathfrak{a}$ is obtained by changing the order of two variables of the same type); (VII) $\varphi(\alpha^1, a, \mathfrak{b}) = \alpha^1(a)$; (VIII) $\varphi(\alpha^j, \mathfrak{b}) = \alpha^j(\lambda \alpha^{j-2} \chi(\alpha^j, \alpha^{j-2}, \mathfrak{b}))$ ($\lambda \alpha^{j-2}$ designates that χ is a function of the variables α^{j-2}).

We assign to each function $\varphi(\mathfrak{a})$ a natural number called an **index** (which plays the same role as a Gödel number) in such a way that it reflects the manner of application of the schemata used to introduce $\varphi(\mathfrak{a})$. Now, we write $\varphi(\mathfrak{a})$ with an index e as $\{e\}(\mathfrak{a})$. We call a function $\varphi(\mathfrak{a})$ **partial recursive** if it is definable by a finite series of applications of the schemata (I)–(VIII) ($\simeq$ is employed instead of $=$ in (IV)–(VI) and (VIII)), and (IX) $\varphi(a, \mathfrak{b}, \mathfrak{c}) \simeq \{a\}(\mathfrak{b}, \mathfrak{c})$ ($\mathfrak{c}$ is a finite list of variables of arbitrary types). In particular, $\varphi(\mathfrak{a})$ is called a **general recursive function** if it is defined for all values of the argument $\mathfrak{a}$. These notions can be **relativized** also with respect to any given functions. Note that for the case of types $\leqslant 1$, primitive recursive functions, partial recursive functions, and also general recursive functions in the present sense are equivalent to the corresponding notions (introduced via relativization with uniformity) in the ordinary sense already described. The following theorem is important: Let r be the maximum type of $\mathfrak{a}$. Then there

are two primitive recursive predicates M, N such that

$$\{e\}(\mathfrak{a})\simeq$$

$$w\Leftrightarrow\forall\xi^{r-1}\exists\eta^{r-2}M(e,\mathfrak{a},w,\xi^{r-1},\eta^{r-2}),\qquad r\geqslant 2,$$

$$w\Leftrightarrow\exists\xi^{r-1}\forall\eta^{r-2}N(e,\mathfrak{a},w,\xi^{r-1},\eta^{r-2}),\qquad r>2.$$

Every function definable using (IX′) $\psi(\mathfrak{a})\simeq\mu x(\psi(\mathfrak{a},x)=0)$ instead of (IX) is partial recursive. However, not all the partial recursive functions of variables of types $\geqslant 2$ can be obtained by applying schemata (I)–(VIII) and (IX′).

Further developments have been pursued by J. E. Fenstad, J. Moldestad, and others in abstract computation theory [20–23].

G. Recursive Functions of Ordinal Numbers

G. Takeuti introduced a notion of primitive recursiveness for functions from a segment of the ordinal numbers to ordinal numbers. Using this, he constructed a model of set theory in ordinal number theory. In connection with recursive functions of ordinal numbers, there are also investigations by A. Lévy, M. Machover, Takeuti and A. Kino, T. Tugué, S. Kripke, and others.

Early treatments of recursive functions of ordinal numbers dealt only with functions on infinite cardinals. For example, Takeuti considered functions with a fixed infinite cardinal κ as a domain and a range, and defined κ-recursive functions using schemata similar to the abovementioned (I)–(VI). Subsequently Kripke observed that the assumption that κ is a cardinal is not necessary, and introduced the notion of **admissible ordinals**. An admissible ordinal κ has the closure properties required for the construction of the calculus, and whenever α, $\beta<\kappa$ and $\beta=f(\alpha)$ is computable, then $\beta=f(\alpha)$ is computable in fewer than κ stages. Given an admissible ordinal κ, **κ-recursiveness** can be defined, as in the case of general recursiveness, by various equivalent methods, e.g., schemata, the equation calculus, and definability in both quantifier forms. Most of the elementary properties of general recursive functions (e.g., the normal form theorem, parametrization theorem, enumeration theorem, etc.) are also valid for κ-recursive functions. The notions of degrees of unsolvability and recursive enumerability can also be generalized, yielding the notions of κ-degrees and κ-recursive enumerability, respectively. The fine structures of these properties are currently the objects of intensive research.

Every infinite cardinal is admissible. The least admissible ordinal is ω, and the next is the ordinal ω_1 of Church and Kleene, i.e., the first nonconstructive ordinal. In fact, for every $n\geqslant 1$, the first ordinal not expressible as the order type of a Δ_n^1 predicate is admissible (→ Section H). For each infinite cardinal κ there are κ^+ admissible ordinals of power κ. Platek investigated recursion theory in a still wider setting. He dealt with functions defined, not on a segment of ordinal numbers, but on a set, and introduced the notion of admissible sets, i.e., sets on which a well-behaved recursion theory can be developed. An admissible set is a transitive ε-model of a certain weak set theory, and an ordinal κ is admissible if and only if there exists an admissible set A such that $A\cap O_n=\kappa$, where O_n is the class of all ordinal numbers.

Recent developments have shown that generalized recursion theory, set theory, and infinitary logic are closely related. In addition to the abovementioned, there are some investigations by Y. N. Moschovakis and others [14–27].

H. Hierarchies

Utilizing the theory of †recursive functions, S. C. Kleene succeeded in establishing a theory of hierarchies that essentially contains classical descriptive set theory as an extreme case [5, 10, 31, 32]. Although research following a similar line had also been done by M. Davis, A. Mostowskiĭ, and others, it was Kleene who succeeded in bringing the theory to its present form.

Sets or functions are described by †predicates, which we classify as follows. Let $a, b, \ldots$, $a_1, a_2, \ldots, x, y, \ldots$, be variables ranging over the set $\mathbf{N}$ of natural numbers, and $\alpha, \beta, \ldots$, $\alpha_1, \alpha_2, \ldots, \xi, \eta, \ldots$ be variables ranging over the set $\mathbf{N}^{\mathbf{N}}$ of all †number-theoretic functions with one argument. Let $\psi_1, \ldots, \psi_l (l\geqslant 0)$ be number-theoretic functions. A predicate $P(\alpha_1,\ldots,\alpha_m,a_1,\ldots,a_n)$ $(m,n\geqslant 0, m+n>0)$ with variables of two †types is called **analytic** in $\psi_1,\ldots,\psi_l$ $(l\geqslant 0)$ if it is expressible syntactically by applying a finite number of logical symbols: $\rightarrow$, $\vee$, $\wedge$, $\neg$, $\exists x$, $\forall x$, $\exists\xi$, $\forall\xi$, to †general recursive predicates in $\psi_1,\ldots,\psi_l$. In particular, when P is expressible without function quantifiers $\exists\xi$, $\forall\xi$, it is called **arithmetical** in $\psi_1,\ldots,\psi_l$ $(l\geqslant 0)$. When $l=0$, they are called simply analytic and arithmetical, respectively.

For brevity, consider the case $l=0$, and denote by $\mathfrak{a}$ a finite list of variables $(\alpha_1,\ldots,\alpha_m,a_1,\ldots,a_n)$. Every arithmetical predicate $P(\mathfrak{a})$ is expressible in a form contained

in the following table (a):

(a) $R(\mathfrak{a})$;

$\exists x R(\mathfrak{a}, x)$, $\quad \forall x \exists y R(\mathfrak{a}, x, y), \ldots,$

$\forall x R(\mathfrak{a}, x)$, $\quad \exists x \forall y R(\mathfrak{a}, x, y), \ldots,$

where each R is †general recursive. In order to obtain such an expression we first transform the given predicate into its †prenex normal form and then contract successive quantifiers of the same kind by the formula

$$\exists x_1 \ldots \exists x_n A(x_1, \ldots, x_n) \Leftrightarrow \exists x A((x)_0, \ldots, (x)_{n-1}) \quad (1)$$

and its "dual form." Each form in (a) (or the class of all predicates with that form) is denoted by Σ_k^0 or Π_k^0, where the suffix k refers to the number of quantifiers prefixed, and Σ or Π shows that the outermost quantifier is existential or universal, respectively. A predicate that is expressible in both forms Σ_k^0 and Π_k^0 (or the class of such predicates) is denoted by Δ_k^0. A predicate belongs to Δ_1^0 if and only if it is general recursive (an analog of †Suslin's theorem).

For $k \geqslant 1$, there exists in Σ_k^0 (or Π_k^0) an **enumerating predicate** that specifies every predicate in Σ_k^0 (Π_k^0). For example, for Π_2^0 and $m = n = 1$, there is a †primitive recursive predicate $S(\alpha, z, a, x, y)$ such that, given a general recursive predicate $R(\alpha, a, x, y)$, we have a natural number e such that

$$\forall x \exists y R(\alpha, a, x, y) \Leftrightarrow \forall x \exists y S(\alpha, e, a, x, y)$$

(**enumeration theorem**). In this theorem, we can take $T_2^\alpha(z, a, x, y)$ (→ Section F) as $S(\alpha, z, a, x, y)$. For each $k \geqslant 0$, there exists a Σ_{k+1}^0 (Π_{k+1}^0) predicate that is not expressible in its dual form Π_{k+1}^0 (Σ_{k+1}^0) (hence, of course, in neither Σ_k^0 not Π_k^0) (**hierarchy theorem**). Therefore, table (a) gives the classification of the arithmetical predicates in a hierarchy. This hierarchy is called the **arithmetical hierarchy**. For each $k \geqslant 1$, there exists a **complete** predicate with respect to $\Sigma_k^0(\Pi_k^0)$, that is, a $\Sigma_k^0(\Pi_k^0)$ predicate with only one variable such that any $\Sigma_k^0(\Pi_k^0)$ predicate is expressible by substituting a suitable general (or more strictly, primitive) recursive function for its variable (**theorem on complete form**). When $m = 0$, all the general recursive predicates in Σ_k^0 exhaust Δ_{k+1}^0 (**Post's theorem**).

Concerning the function quantifiers, we have

$$\exists \alpha_1 \ldots \exists \alpha_m A(\alpha_1, \ldots, \alpha_m) \Leftrightarrow \exists \alpha A(\lambda t(\alpha(t))_0, \ldots, \lambda t(\alpha(t))_{m-1}) \quad (2)$$

$$\exists x A(x) \Leftrightarrow \exists \alpha A(\alpha(0)), \quad (3)$$

$$\forall x \exists \alpha A(x, \alpha) \Leftrightarrow \exists \alpha \forall x A(x, \lambda t \alpha(2^x \cdot 3^t)), \quad (4)$$

and their dual forms. For any general recursive predicate R, there is a primitive recursive predicate S such that

$$\exists \alpha R(\alpha, \mathfrak{a}) \Leftrightarrow \exists \alpha \exists x S(\bar{\alpha}(x), \mathfrak{a}) \Leftrightarrow \exists y S(y, \mathfrak{a}) \quad (5)$$

and its dual hold. Using these facts, we can classify the forms of all analytic predicates by the table (b):

(b) $A(\mathfrak{a})$;

$\forall \alpha \exists x R(\mathfrak{a}, \alpha, x)$, $\quad \exists \alpha \forall \beta \exists x R(\mathfrak{a}, \alpha, \beta, x), \ldots,$

$\exists \alpha \forall x R(\mathfrak{a}, \alpha, x)$, $\quad \forall \alpha \exists \beta \forall x R(\mathfrak{a}, \alpha, \beta, x), \ldots,$

where A is arithmetical and each R is general recursive. Similarly, denote by Σ_k^1, Π_k^1 each form of predicate in (b) (or the class of all predicates reducible to that form), where k is the number of function quantifiers prefixed; also, denote by Δ_k^1 the (class of) predicates expressible in both forms Σ_k^1 and Π_k^1. For Σ_k^1, Π_k^1 ($k \geqslant 1$), we have the enumeration theorem, the hierarchy theorem, and the theorem on complete form. The hierarchy given by table (b) is called the **analytic hierarchy**.

For $l > 0$ (namely, when predicates are arithmetical or analytic in $\psi_1, \ldots, \psi_l$), we can †uniformly †relativize the above results with respect to $\psi_1, \ldots, \psi_l$. Now let $\{\Sigma_k^{r, \psi_1, \ldots, \psi_l}, \Pi_k^{r, \psi_1, \ldots, \psi_l}\}_k$ ($r = 0, 1$) be the corresponding hierarchy relative to $\psi_1, \ldots, \psi_l$. Given a set C ($\subset \mathbf{N}^{\mathbf{N}}$) of functions with one argument, we can consider hierarchies of predicates which are arithmetical or analytic in a finite number of functions in C. Such a hierarchy is called a C-**arithmetical** or C-**analytic hierarchy** and denoted by $\{\Sigma_k^0[C], \Pi_k^0[C]\}_k$ or $\{\Sigma_k^1[C], \Pi_k^1[C]\}_k$, respectively. That is, when we regard $\Sigma_k^r[C]$ as a class of predicates (or sets) P, it is the family $\{P \mid P \in \Sigma_k^{r, \xi_1, \ldots, \xi_l}, \xi_1, \ldots \xi_l \in C, l = 0, 1, 2, \ldots\}$. These notations have been given by J. W. Addison [28, 29]. The $\mathbf{N}^{\mathbf{N}}$-arithmetical hierarchy and the $\mathbf{N}^{\mathbf{N}}$-analytic hierarchy for sets correspond respectively to the finite Borel hierarchy and the projective hierarchy in the †space of irrational numbers. Addison called the theory of those hierarchies **classical descriptive set theory**, and in contrast to this, the theory of arithmetical and analytic hierarchies for sets ($C = \varnothing$) **effective descriptive set theory** [28].

We now restrict our consideration to predicates for natural numbers (i.e., to the case $m = 0$). Define the predicates L_k by $L_0(a) \Leftrightarrow a = a$, $L_{k+1}(a) \Leftrightarrow \exists x T_1^{L_k}(a, a, x)$. For each $k \geqslant 0$, $L_{k+1}(a)$ is a Σ_{k+1}^0 predicate which is of the highest †degree of recursive unsolvability among the Σ_{k+1}^0 predicates, and its degree is properly higher than that of $L_k(a)$. Thus L_k, $k = 0, 1, 2, \ldots$, determine the **arithmetical hierarchy of degrees of recursive unsolvability**. Kleene has extended the series of L_k by using

the system S_3 (→ 81 Constructive Ordinal Numbers) of notations for the constructive ordinal numbers as follows [6]: $H_1(a) \Leftrightarrow a=a$; for $y \in O$, $H_{2^y}(a) \Leftrightarrow \exists x T_1^{H_y}(a,a,x)$; for $3 \cdot 5^y \in O$, $H_{3 \cdot 5^y}(a) \Leftrightarrow H_{y_{(a)_1}}((a)_0)$, where $y_n = \{y\}(n_0)$. This H_y is defined for each $y \in O$, and it is of a properly higher degree than that of H_z when $z <_O y$. If $|y|=|z|$ ($|y|$ is the ordinal number represented by y), then H_y and H_z are of the same degree (C. Spector [34]). Thus a hierarchy of degrees is uniquely determined by constructive ordinal numbers. This hierarchy is called the **hyperarithmetical hierarchy of degrees of recursive unsolvability**. A function or predicate is said to be **hyperarithmetical** if it is recursive in H_y for some $y \in O$. These concepts and the results mentioned below can be relativized to any given functions or predicates.

A necessary (Kleene [31]) and sufficient (Kleene [32]) condition for a predicate to be hyperarithmetical is that it be expressible in both one-function quantifier forms Δ_1^1 (an effective version of Suslin's theorem). Denote by Hyp the set ($\subset \mathbf{N}^{\mathbf{N}}$) of all hyperarithmetical functions α. For an arithmetical predicate $A(\alpha,a)$, $\exists \alpha_{\alpha \in \mathrm{Hyp}} A(\alpha,a)$ is always a Π_1^1 predicate (Kleene [33]). Conversely, for any Π_1^1 predicate P, there is a general recursive predicate R such that $P(a) \Leftrightarrow \exists \alpha_{\alpha \in \mathrm{Hyp}} \forall x R(a,\alpha,x)$ (Spector [35]). As to †uniformization, for a Π_1^1 predicate $P(a,b)$, we have $\forall x \exists y P(x,y) \Rightarrow \exists \alpha_{\alpha \in \mathrm{Hyp}} \forall x P(x,\alpha(x))$ (G. Kreisel, 1962). Let $\mathbf{E}$ be an object of type 2 defined by: $\mathbf{E}(\alpha)=0$ if $\exists x(\alpha(x)=0)$, otherwise $\mathbf{E}(\alpha)=1$. A function $\varphi(a_1,\ldots,a_n)$ is hyperarithmetical if and only if it is general recursive in $\mathbf{E}$ (Kleene [10]). A predicate that is hyperarithmetical relative to Π_k^1 predicates ($k \geqslant 0$) is of Δ_{k+1}^1 (Kleene [32]), but the converse does not hold in general (Addison and Kleene, 1957).

Kleene extended his theory of hierarchy to the case of predicates of variables of any type by utilizing the theory of general recursive functions with variables of finite types $0,1,2,\ldots$ [10]. Let $\mathfrak{a}^t$ be a list of variables of types $\leqslant t$. We say a predicate $P(\mathfrak{a}^t)$ is of order r in completely defined functions $\psi_1,\ldots,\psi_l$ ($l \geqslant 0$) (for brevity, denote them by Ψ) if P is syntactically expressible in terms of variables of finite types, predicates that are general recursive in Ψ, and symbols of the †predicate calculus with quantification consisting only of variables of types $<r$. The predicates of order 0 in Ψ are exactly the general recursive ones in Ψ. When $t \geqslant 1$, and Ψ are functions of variables of type 0, a predicate $P(\mathfrak{a}^t)$ is of order 1 (of order 2) in Ψ if and only if P is arithmetical (analytic).

We have theorems similar to (2)–(4) and the following theorem and its dual for $r \geqslant 2$: For any given general recursive predicate $P(\mathfrak{a}^r,\sigma^r,\xi^{r-2})$, there is a primitive recursive predicate $R(\mathfrak{a}^r,\eta^{r-1},\xi^{r-2})$ such that

$$\exists \sigma^r \forall \xi^{r-2} P(\mathfrak{a}^r,\sigma^r,\xi^{r-2}) \Leftrightarrow \exists \eta^{r-1} \forall \xi^{r-2} R(\mathfrak{a}^r,\eta^{r-1},\xi^{r-2}). \quad (6)$$

Using these equivalences, each predicate $P(\mathfrak{a}^t)$ of order $r+1$ ($r \geqslant 0$) is expressible in one of the following forms:

(c) $B(\mathfrak{a})$;

$\forall \alpha^r \exists \xi^{r-1} R(\mathfrak{a},\alpha^r,\xi^{r-1})$,

$\exists \alpha^r \forall \beta^r \exists \xi^{r-1} R(\mathfrak{a},\alpha^r,\beta^r,\xi^{r-1}),\ldots,$

$\exists \alpha^r \forall \xi^{r-1} R(\mathfrak{a},\alpha^r,\xi^{r-1})$,

$\forall \alpha^r \exists \beta^r \forall \xi^{r-1} R(\mathfrak{a},\alpha^r,\beta^r,\xi^{r-1}),\ldots,$

where B is of order r and each R is general recursive. When $t=r+1$, table (c) gives the classification of the predicates of order $r+1$ into the hierarchy. In fact, for the predicates $P(\mathfrak{a}^{r+1})$ in each form, we have the enumeration theorem, the hierarchy theorem, and the theorem on complete form (Kleene [10]).
D. A. Clarke [30] has published a detailed review of the general theory of hierarchies.

References

[1] K. Gödel, Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I, Monatsh. Math. Phys., 38 (1931), 173–198.
[2] S. C. Kleene, General recursive functions of natural numbers, Math. Ann., 112 (1936), 727–742.
[3] S. C. Kleene, On notation for ordinal numbers, J. Symbolic Logic, 3 (1938), 150–155.
[4] A. Church, The calculi of lambda-conversion, Ann. Math. Studies, Princeton Univ. Press, 1941.
[5] S. C. Kleene, Recursive predicates and quantifiers, Trans. Amer. Math. Soc., 53 (1943), 41–73.
[6] E. L. Post, Recursively enumerable sets of positive integers and their decision problems, Bull. Amer. Math. Soc., 50 (1944), 284–316.
[7] S. C. Kleene, Introduction to metamathematics, Van Nostrand, 1952.
[8] R. Péter, Rekursive Funktionen, Akademische Verlag., 1951.
[9] A. A. Markov, Theory of algorithms (in Russian), Trudy Mat. Inst. Steklov., 42 (1954).
[10] S. C. Kleene, Recursive functionals and quantifiers of finite types I, Trans. Amer. Math. Soc., 91 (1959), 1–52.
[11] S. C. Kleene, Recursive functionals and quantifiers of finite types II, Trans. Amer. Math. Soc., 108 (1963), 106–142.
[12] H. Hermes, Enumerability, decidability, computability, Springer, 1965.

[13] H. Rogers, Theory of recursive functions and effective computability, McGraw-Hill, 1967.
[14] J. Barwise, Infinitary logic and admissible sets, J. Symbolic Logic, 34 (1969), 226–252.
[15] R. B. Jensen and C. Karp, Primitive recursive set functions, Proc. Symposia in Pure Math., XIII (1971), 143–176.
[16] S. Kripke, Transfinite recursions on admissible ordinals, J. Symbolic Logic, 29 (1964), 161–162.
[17] M. Machover, The theory of transfinite recursion, Bull. Amer. Math. Soc., 67 (1961), 575–578.
[18] G. Takeuti, A formalization of the theory of ordinal numbers, J. Symbolic Logic, 30 (1965), 295–317.
[19] T. Tugué, On the partial recursive functions of ordinal numbers, J. Math. Soc. Japan, 16 (1964), 1–31.
[20] J. Moldestad, Computation in higher types, Springer, 1977.
[21] J. E. Fenstad, R. O. Gandy, and G. E. Sacks (eds.), Generalized recursion theory, North-Holland, 1974.
[22] J. E. Fenstad, R. O. Gandy, and G. E. Sacks (eds.), Generalized recursion theory II, North-Holland, 1978.
[23] J. E. Fenstad, General recursion theory, Springer, 1980.
[24] Y. N. Moschovakis, Elementary induction on abstract structures, North-Holland, 1974.
[25] Y. N. Moschovakis, Descriptive set theory, North-Holland, 1980.
[26] J. Barwise, Admissible sets and structures, Springer, 1975.
[27] P. G. Hinman, Recursion theoretic hierarchies, Springer, 1978.
[28] J. W. Addison, Separation principles in the hierarchies of classical and effective descriptive set theory, Fund. Math., 46 (1958–1959), 123–135.
[29] J. W. Addison, Some consequences of the axiom of constructibility, Fund. Math., 46 (1958–1959), 337–357.
[30] D. A. Clarke, Hierarchies of predicates of finite types, Mem. Amer. Math. Soc., 1964.
[31] S. C. Kleene, Arithmetical predicates and function quantifiers, Trans. Amer. Math. Soc., 79 (1955), 312–340.
[32] S. C. Kleene, Hierarchies of number-theoretic predicates, Bull. Amer. Math. Soc., 61 (1955), 193–213.
[33] S. C. Kleene, Quantification of number-theoretic functions, Compositio Math., 14 (1959), 23–40.
[34] C. Spector, Recursive well-orderings, J. Symbolic Logic, 20 (1955), 151–163.
[35] C. Spector, Hyperarithmetical quantifiers, Fund. Math., 48 (1959–1960), 313–320.

357 (VI.6) Regular Polyhedra

A. Regular Polygons

A †polygon in a Euclidean plane bounding a †convex cell whose sides and interior angles are all respectively congruent is called a **regular polygon**. When the number of vertices (which equals the number of sides) is n, it is called a **regular n-gon**. There exist a circle (**circumscribed circle**) passing through all the vertices of a regular n-gon and a concentric circle (**inscribed circle**) tangent to all the sides. We call the center of these circles the **center** of the regular n-gon. The n vertices of a regular n-gon are obtained by dividing a circle into n equal parts. (When a polygon in a Euclidean plane bounds a †convex cell, this 2-cell is sometimes called a convex polygon. Thus *regular polygon* sometimes means the convex cell bounded by a regular polygon as described above.) A necessary and sufficient condition for a regular n-gon to be geometrically constructible is that n be decomposable into the product of prime numbers $n = 2^m p_1 \ldots p_k$ ($m \geqslant 0$), where the p_i ($i = 1, 2, \ldots$) are different †Fermat numbers (→ 179 Geometric Construction).

B. Regular Polyhedra

Consider a regular polygon on a plane, and take a point on the line perpendicular to the plane at the center of the polygon. The set of points on all half-lines joining this point and points on the polygon (considered as a convex cell) is called a **regular polyhedral angle** having this point as vertex (Fig. 1).

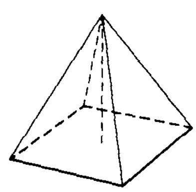


Fig. 1

When a †convex polyhedron $\mathfrak{F}$ in E^3 satisfies the following two conditions, we call it a **regular polyhedron**: (1) Each face of $\mathfrak{F}$, which is a 2-dimensional cell, is a regular polygon, and all faces of $\mathfrak{F}$ are congruent to each other. (2) Its vertices are all surrounded alike. That is, by the projection of $\mathfrak{F}$ from each vertex of $\mathfrak{F}$, we obtain a regular polyhedral angle; these regular polyhedral angles are all congruent to each other. From (2) we see that the number of edges emanating from each vertex of $\mathfrak{F}$ is

independent of the vertex. It has been known since Plato's time that there are only five kinds of regular polyhedra: **tetrahedrons** (Fig. 2), **octahedrons** (Fig. 3), **icosahedrons** (Fig. 4), **cubes** or **hexahedrons** (Fig. 5), and **dodecahedrons** (Fig. 6) (see also see Table 1).

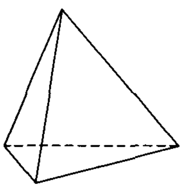


Fig. 2
Regular tetrahedron.

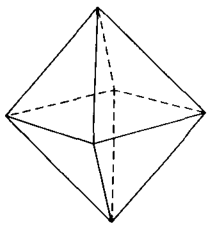


Fig. 3
Regular octahedron.

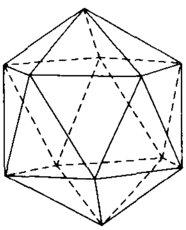


Fig. 4
Regular icosahedron.

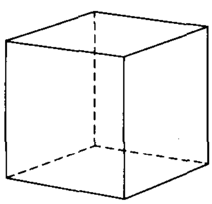


Fig. 5
Regular hexahedron (or cube).

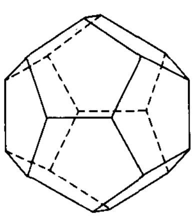


Fig. 6
Regular dodecahedron.

From a given regular polyhedron, we can obtain another one by taking as vertices the centers of all the faces of the given polyhedron (Fig. 7). We say that the given regular polyhedron and the one obtained in this way are **dual** to each other. The octahedron and hexahedron are dual to each other, as are the icosahedron and dodecahedron. The tetrahedron is dual to itself. For a regular polyhedron $\mathfrak{F}$, there exist concentric circumscribed and inscribed spheres whose center is the center of symmetry of $\mathfrak{F}$ and is called the **center** of $\mathfrak{F}$. Drawing tangent planes to the circumscribed sphere at each vertex, we can obtain a regular polyhedron dual to the given one (Fig. 8).

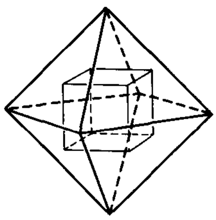


Fig. 7

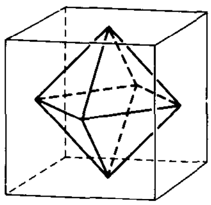


Fig. 8

In a regular polyhedron, let a be the length of an edge, θ the magnitude of the dihedral angle at each edge, and R and r the radii of circumscribed and inscribed spheres, respectively. Then the following relations hold (we assume that each face is a regular p-gon and q faces meet at each vertex):

$$\sin\frac{\theta}{2}=\cos\frac{\pi}{q}\Big/\sin\frac{\pi}{p},$$

$$R=\frac{a}{2}\sin\frac{\pi}{q}\Big/\sin\frac{\pi}{p}\cos\frac{\theta}{2}, \tag{1}$$

$$r=\frac{a}{2}\cot\frac{\pi}{p}\tan\frac{\theta}{2},\qquad \frac{R}{r}=\tan\frac{\pi}{p}\tan\frac{\pi}{q}$$

(see Table 2). Corresponding to these polyhedra, we have finite subgroups of $O(3)$ called regular polyhedral groups (→ 151 Finite Groups).

C. Higher-Dimensional Cases

It is possible to generalize these considerations to higher dimensions to define inductively

Table 1. Regular Polyhedra in 3-Dimensional Euclidean Space E^3

Figure	Face	Number of Vertices	Number of Edges	Number of Faces	Number of Faces around a Vertex
Regular tetrahedron	Equilateral triangle	4	6	4	3
Regular octahedron	Equilateral triangle	6	12	8	4
Regular icosahedron	Equilateral triangle	12	30	20	5
Regular hexahedron	Square	8	12	6	3
Regular dodecahedron	Regular pentagon	20	30	12	3

Table 2. Numerical Values for Eqs. (1)

Number of Faces	$\sin\theta$	θ	R/a	r/a
4	$2\sqrt{2}/3$	70°31′43.6″	$\sqrt{6}/4$	$\sqrt{6}/12$
6	1	90°	$\sqrt{3}/2$	$1/2$
8	$2\sqrt{2}/3$	109°28′16.4″	$1/\sqrt{2}$	$1/\sqrt{6}$
12	$2/\sqrt{5}$	116°33′54.2″	$\dfrac{\sqrt{3}(\sqrt{5}+1)}{4}$	$\dfrac{\sqrt{25+11\sqrt{5}}}{2\sqrt{10}}$
20	$2/3$	138°11′22.8″	$\dfrac{\sqrt{(5+\sqrt{5})}}{2\sqrt{2}}$	$\dfrac{3+\sqrt{5}}{4\sqrt{3}}$

Table 3. Regular Polyhedra in 4-Dimensional Euclidean Space E^4

Figure	3-Dimensional Regular Polyhedra Kind	3-Dimensional Regular Polyhedra Number	Number of Vertices	Duality
Regular 5-hedron	Tetrahedron	5	5	a
Regular 8-hedron	Cube	8	16	b
Regular 16-hedron	Tetrahedron	16	8	b
Regular 24-hedron	Octahedron	24	24	a
Regular 120-hedron	Dodecahedron	120	600	b
Regular 600-hedron	Tetrahedron	600	120	b

a: dual to itself; b: dual to each other

Table 4. Regular Polyhedra in n-Dimensional Euclidean Space ($n \geqslant 5$)

Figure	Regular Polyhedron in $\mathbf{R}^{n-1}$ Kind	Regular Polyhedron in $\mathbf{R}^{n-1}$ Number	Number of Vertices	Duality
Regular $(n+1)$-hedron	Regular n-hedron	$n+1$	$n+1$	a
Regular $2n$-hedron	Regular $(2n-2)$-hedron	$2n$	2^n	b
Regular 2^n-hedron	Regular n-hedron	2^n	$2n$	b

a: dual to itself; b: dual to each other

regular polyhedra in E^n, $n \geqslant 4$. When $n=4$ we have 6 kinds of regular polyhedra (Table 3). For $n \geqslant 5$ we have only 3 kinds (Table 4) (→ 70 Complexes).

References

[1] J. S. Hadamard, Leçons de géométrie élémentaire II, Armand Colin, second edition, 1906, 425–427.
[2] D. Hilbert and S. Cohn-Vossen, Anschauliche Geometrie, Springer, 1932; English translation, Geometry and the imagination, Chelsea, 1952.
[3]E. Steinitz and H. Rademacher, Vorlesungen über die Theorie der Polyeder, Springer, 1934.
[4] H. S. M. Coxeter, Regular polytopes, Methuen, 1948; third edition, Dower, 1973.
[5] H. S. M. Coxeter, Regular complex polytopes, Cambridge Univ. Press, 1974.

358 (II.2) Relations

A. General Remarks

In its wider sense the term *relation* means †n-ary relation ($n=1,2,3,\ldots$) (→ 411 Symbolic Logic G), but in this article we restrict ourselves to its most ordinary meaning, i.e., to the case $n=2$. Let X, Y be two sets and x, y be two variables taking their values in X, Y, respectively. A proposition $R(x,y)$ containing x, y is called a **relation** or a **binary relation** if it can be determined whether $R(a,b)$ is true or false for each pair (a,b) in the †Cartesian product $X \times Y$. For example, if both X and Y are the set of rational integers, then the following propositions are relations: $x \leqslant y$, $x-y$ is even, x divides y. A relation $R(x,y)$ is sometimes written as xRy.

For a given relation R, we define its **inverse relation** R^{-1} by $yR^{-1}x \Leftrightarrow xRy$. Then R is the

inverse relation of R^{-1}. In the example above, the inverse relation of $x \leqslant y$ is $y \geqslant x$, and the inverse relation of x is a divisor of y is y is a multiple of x. A relation R is called **reflexive** if xRx holds. R is called **symmetric** if $xRy \Leftrightarrow yRx$ (namely, if R and R^{-1} are identical). R is called **transitive** if xRy and yRz imply xRz. R is called **antisymmetric** if xRy and yRx imply $x = y$. A reflexive, symmetric, and transitive relation is called an **equivalence relation** (→ 135 Equivalence Relations). A reflexive and transitive relation is called a †preordering. A reflexive, transitive, and antisymmetric relation is called an †ordering (→ 311 Ordering).

Suppose that we are given a relation xRy ($x \in X, y \in Y$). Then the set $G = \{(x, y) \mid xRy\}$, which consists of elements (x, y) of the Cartesian product $X \times Y$ satisfying xRy, is called the **graph** of the relation R. Conversely, for any subset G of $X \times Y$, there exists a unique relation R with the graph G given by $xRy \Leftrightarrow (x, y) \in G$.

B. Correspondences

For a subset G of the Cartesian product $X \times Y$, the triple $\Gamma = (G, X, Y)$ is called a **correspondence** from X to Y. The set X is called the **initial set** of the correspondence Γ, and Y the **final set** of Γ. A relation xRy ($x \in X, y \in Y$) determines a correspondence $\Gamma = (G, X, Y)$ by its graph G, and conversely, a correspondence Γ determines a relation R. Given a correspondence $\Gamma = (G, X, Y)$, the sets $A = \mathrm{pr}_X G$ and $B = \mathrm{pr}_Y G$, where $\mathrm{pr}_X : X \times Y \rightarrow X$ and $\mathrm{pr}_Y : X \times Y \rightarrow Y$ are the †canonical projections, are called the **domain** and **range** of the correspondence Γ, respectively. For $x \in X$, the set $\{y \in Y \mid (x, y) \in G\}$ is denoted by $G(x)$ or $\Gamma(x)$, and we say that any element y of $G(x)$ **corresponds** to x by Γ.

For a subset G of $X \times Y$, we define a subset G^{-1} of $Y \times X$ by $(x, y) \in G \Leftrightarrow (y, x) \in G^{-1}$. Given a correspondence $\Gamma = (G, X, Y)$, the correspondence (G^{-1}, Y, X) is denoted by Γ^{-1} and is called the **inverse correspondence** of Γ. If G is the graph of a relation R, then G^{-1} is the graph of the inverse relation R^{-1}. The domain of the correspondence Γ is the range of Γ^{-1}, and vice versa. We have $(\Gamma^{-1})^{-1} = \Gamma$.

Suppose that we are given correspondences $\Gamma_1 = (G_1, X, Y)$ and $\Gamma_2 = (G_2, Y, Z)$. We define a subset G of $X \times Z$ by: $(x, z) \in G \Leftrightarrow$ there exists $y \in Y$ satisfying $(x, y) \in G_1$ and $(y, z) \in G_2$. Then the correspondence $\Gamma = (G, X, Z)$ is denoted by $\Gamma_2 \circ \Gamma_1$ and is called the **composite** of Γ_1 and Γ_2. We have the **associative law** $(\Gamma_3 \circ \Gamma_2) \circ \Gamma_1 = \Gamma_3 \circ (\Gamma_2 \circ \Gamma_1)$ and the law $(\Gamma_2 \circ \Gamma_1)^{-1} = \Gamma_1^{-1} \circ \Gamma_2^{-1}$.

Let Γ be a correspondence from X to Y, and assume that to any x belonging to the domain A of Γ there corresponds one and only one $y \in Y$, namely, $\Gamma(x) = \{y\}$ for any $x \in A$. Then Γ is called a **univalent correspondence**. If Γ and Γ^{-1} are both univalent correspondences, Γ is called a **one-to-one correspondence**. For given sets A and B, a univalent correspondence with domain A and range B is called a †mapping (or †function) with domain A and range B (→ 381 Sets C).

References

See references to 381 Sets.

359 (XX.21) Relativity

A. History

The theory of relativity is a system of theoretical physics established by A. Einstein and is composed of special relativity and general relativity. Toward the end of the 19th century, it was believed that electromagnetic waves propagate through the ether, a hypothetical medium. A number of experimenters tried to find the motion of the earth relative to the ether, but all these attempts were unsuccessful (A. A. Michelson, E. W. Morley). Studying these results, in 1905 Einstein proposed the **special theory of relativity**, which extended Galileo's relativity principle of †Newtonian mechanics to †electromagnetism and radically revised the concepts of space and time. Almost all the conclusions of special relativity theory are now confirmed by experiments, and this theory has even become a guiding principle for developing new theories in physics. By extending special relativity, Einstein established (1915) the **general theory of relativity**. Its principal part is a new theory of gravitation containing Newton's theory as a special case. Its conclusions about the solar system are compatible with observed facts that are regarded as experimental support for the theory. Effects due to general relativity other than those just described have been studied to a considerable extent, but it is hard at present to test theoretical results experimentally, and there are some doubts about the limit of its applicability.

B. Special Relativity

In Newtonian mechanics, natural phenomena are described in a 3-dimensional Euclidean

space considered independent of time. In special relativity, however, it is postulated that space and time cannot be separated but are unified into a 4-dimensional pseudo-Euclidean space with the †fundamental form

$$ds^2=\sum_{a,b} g_{ab}\,dx^a dx^b \equiv c^2 dt^2-dx^2-dy^2-dz^2,$$

where $a, b=0, 1, 2, 3$ and

$$(x^0, x^1, x^2, x^3)=(ct, x, y, z).$$

Here (x, y, z) are spatial Cartesian coordinates, t is time, and c is the speed of light. This space was introduced by H. Minkowski and is called **Minkowski space-time**. By means of it Minkowski gave an ingenious geometric interpretation to special relativity.

A nonzero vector V is called timelike, null (or lightlike), or spacelike according as $V^2>0$, $=0$, or <0, where $V^2=\sum_{a,b} g_{ab} V^a V^b$.

The group of motions in Minkowski space-time is called the **inhomogeneous Lorentz group**. Its elements can be written as

$$x^{i\prime}=\sum_a c^i_a x^a+c^i, \qquad \sum_{a,b} g_{ab} c^a_i c^b_j=g_{ij},$$

where c^i_j and c^i are constants. The transformations with $c^i=0$ are usually called **Lorentz transformations**, and the group G composed of these transformations is called the **homogeneous Lorentz group** or simply the **Lorentz group**. These are important concepts in special relativity. If G_0 denotes the †connected component of the identity element of G, the factor group G/G_0 is an Abelian group of type (2, 2) and of order 4. We call G_0 the **proper Lorentz group**. A frequently used element of G_0 is

$$L_v : x'=\frac{x-vt}{\sqrt{1-v^2/c^2}}, \qquad t'=\frac{t-(v/c^2)x}{\sqrt{1-v^2/c^2}},$$

$$y'=y, \qquad z'=z; \qquad |v|<c. \tag{1}$$

Such transformations form a 1-dimensional subgroup of G_0 with v as a parameter, and the composition law of the subgroup is given by

$$L_u \cdot L_v=L_w, \qquad w=\frac{u+v}{1+uv/c^2}.$$

Elements of G not belonging to G_0 are

$$T : x^{0\prime}=-x^0, \quad x^{i\prime}=x^i; \qquad i=1,2,3,$$

$$S : x^{0\prime}=x^0, \quad x^{i\prime}=-x^i; \qquad i=1,2,3.$$

Both T (**time reversal**) and S (**space reflection** or **parity transformation**) have aroused much interest among physicists.

Historically, the transformation formula (1) was first obtained by H. A. Lorentz, under the assumption of contraction of a rod in the direction of its movement in order to overcome the difficulties of the ether hypothesis, but his theoretical grounds were not satisfactory. On the other hand, Einstein started with the following two postulates: (i) **Special principle of relativity**: A physical law should be expressed in the same form in all **inertial systems**, namely, in all coordinate systems that move relative to each other with uniform velocity. (ii) **Principle of invariance of the speed of light**: The speed of light in a vacuum is the same in all inertial systems and in all directions, irrespective of the motion of the light source. From these assumptions Einstein derived (1) as the transformation formula between inertial systems $\mathbf{x}=(ct, x, y, z)$ and $\mathbf{x}'=(ct', x', y', z')$ that move relative to each other with uniform velocity v along the common x-axis. This was the first step in special relativity, and along this line of thought, Einstein solved successively the problems of the Lorentz-Fitzgerald contraction, the dilation of time as measured by moving clocks, the aberration of light, the Doppler effect, and Fresnel's dragging coefficient.

C. Relativity and Electromagnetism

In special relativity, a physical quantity is represented by a †tensor (or a scalar or a vector) in Minkowski space-time, and physical laws are written in tensor form and are invariant under Lorentz transformations of coordinates. This is the mathematical expression of the special principle of relativity. Since the transformation (1) tends to a **Galileo transformation** in Newtonian mechanics as $c\to\infty$, the special principle of relativity is a generalization of the Newton-Galileo principle of relativity. To summarize mathematically, it may safely be said that special relativity is a theory of invariants with respect to the Lorentz group. To illustrate this conclusion we consider electromagnetic theory.

The electric field $\mathbf{E}$ is usually represented by a "polar vector" and the magnetic field $\mathbf{H}$ by an "axial vector" in a 3-dimensional Euclidean space. Even if the magnetic field does not exist in one inertial system, the field can arise in another system that moves uniformly relative to the original system. In view of this, electric and magnetic fields are considered in relativity to form one physical quantity with components

$$(F_{ij})=\begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & H_z & -H_y \\ E_y & -H_z & 0 & H_x \\ E_z & H_y & -H_x & 0 \end{pmatrix}.$$

This quantity transforms as an †alternating tensor of degree 2 under Lorentz transformations. In like manner, the electric charge density ρ and the electric current density $\mathbf{J}$ are unified into a †contravariant vector with

respect to Lorentz transformations:

$$s=(s^0,s^1,s^2,s^3)$$
$$=(\rho,J_x/c,J_y/c,J_z/c).$$

Such a vector **s** in Minkowski space-time is sometimes called a **four-vector** as distinguished from an ordinary vector such as **J**. If the electromagnetic field F_{ij} and the current four-vector s are thus defined, the †Maxwell equations, the basic equations of electromagnetism, can be written in tensor form:

$$\sum_a \frac{\partial F^{ia}}{\partial x^a}=s^i, \qquad \frac{\partial F_{jk}}{\partial x^i}+\frac{\partial F_{ki}}{\partial x^j}+\frac{\partial F_{ij}}{\partial x^k}=0,$$

where

$$F^{ij}=\sum_{a,b} g^{ia}g^{jb}F_{ab}, \qquad \sum_a g^{ia}g_{ja}=\delta^i_j.$$

In the same way, the equation of motion for a charged particle in an electromagnetic field can be expressed as

$$\frac{d^2x^i}{ds^2}=\frac{e}{mc}\sum_a F^i_a\frac{dx^a}{ds}, \qquad F^i_j=\sum_a g^{ia}F_{ja},$$

where e and m are the charge and mass of the particle, respectively, and s is the arc length along the particle trajectory (→ 130 Electromagnetism).

Though special relativity originated in studies of electromagnetic phenomena, it has gradually become clear that the theory is valid also for other phenomena. One interesting result is that the energy of a particle moving with uniform velocity v is given by

$$E=mc^2/\sqrt{1-v^2/c^2},$$

and accordingly even a particle at rest has energy mc^2 (**rest energy**). This shows the equivalence of mass and energy, with the conversion formula given by $E=mc^2$. This conclusion has been verified experimentally by studies of nuclear reactions and has become the basis of the development of nuclear power. The special principle of relativity also showed its validity in the electron theory of P. A. M. Dirac (1928) and the quantum electrodynamics of S. Tomonaga (1943) and others. It has, however, been shown that the invariance for space reflection (namely for the coset SG_0 of the Lorentz group G) is violated in the decay of elementary particles (T. D. Lee and C. N. Yang, 1956; C. S. Wu et al., 1957). Similar results have been obtained for time reversal (J. H. Christenson, J. W. Cromin, V. L. Fitch, and R. Turlay, 1964).

D. General Relativity

Special relativity has its origin in studies of electromagnetic phenomena, while the central part of general relativity is a theory of gravitation founded on the **general principle of relativity** and the **principle of equivalence**. The first principle is an extension of the special principle of relativity to accelerated systems in general. It requires that a physical law should be independent of the choice of local coordinates in a 4-dimensional †differentiable manifold representing space and time (**space-time manifold**). Since a physical quantity is represented by a tensor on the space-time manifold, physical laws are expressed in tensor form, in agreement with the first principle. The second principle claims that gravitational and inertial mass are equal, and accordingly fictitious forces due to acceleration (such as centrifugal force) cannot be distinguished from gravitational force. This had been shown with high accuracy by the experiments of R. von Eötvös (1890) and others.

Starting from these two principles, Einstein was led to the following conclusion. If a gravitational field is produced by matter, the corresponding space-time structure is altered; namely, flat Minkowski space-time is changed into a curved 4-dimensional manifold with †pseudo-Riemannian metric of †signature (1, 3). The †fundamental tensor g_{ij} of this manifold represents the gravitational potential, and the gravitational equation satisfied by g_{ij} can be expressed as a geometric law of the manifold. Gravitational phenomena are thus reduced to properties of the geometric structure of the space-time manifold. This idea, which was not seen in the older physics, became the motif in the development of †unified field theories.

Now the gravitational law proposed by Einstein is an analog of the †Poisson equation in Newtonian mechanics. Let R_{ij} and R be the †Ricci tensor and the †scalar curvature formed from g_{ij}. Then outside the source of a gravitational field, g_{ij} must satisfy

$$G_{ij}\equiv R_{ij}-g_{ij}R/2=0, \quad \text{that is,} \quad R_{ij}=0, \tag{2}$$

and inside the source,

$$G_{ij}=\kappa T_{ij}, \tag{3}$$

where κ is the gravitational constant. Here the **energy-momentum tensor** T_{ij} is a †symmetric tensor representing the dynamical state of matter (energy, momentum, and stress). Usually (2) and (3) are called the **exterior** and **interior field equations**, respectively.

Next, the **equation of motion** of a particle in a gravitational field is given by

$$\frac{\delta^2x^i}{\delta s^2}=0, \qquad \sum_{a,b} g_{ab}\frac{dx^a}{ds}\frac{dx^b}{ds}=+1, \tag{4}$$

if the particle mass is so small that its effect on the field is negligible. Here $\delta/\delta s$ stands for †covariant differentiation with respect to the

arc length s along the particle trajectory. In other words, a particle in a gravitational field moves along a timelike †geodesic in the space-time manifold. Similarly, the path of light is represented by a **null geodesic**, whose equation is formally obtained from (4) by replacing the right-hand side of the second equation by zero.

Experimental verification of the theory of general relativity has been obtained by detection of the following effects: the shift of spectral lines due to the gravity of the earth and of white dwarfs, the deflection of light or radio waves passing near the sun, the time delay of radar echo signals passing near the sun, and the advance of the perihelion of Mercury. All the observational data are compatible with the theoretical results. Time delay and the advance of the perihelion have been observed in a binary system of neutron stars. It is generally accepted that these results are experimental verifications of general relativity.

It should be noted that (2) has wave solutions, which have no counterpart in Newton's gravitational theory. This fact implies that gravitational effects propagate with the velocity of light. The gravitational waves transport energy and momentum, and the gravitational mass is decreased by the emission of the waves. Experiments to detect gravitational waves generated in the universe have been planned, and the decrease of the orbital period of a binary star system due to the emission of gravitational waves has been observed.

The concept of gravitational waves suggests the existence of a quantum of the gravitational field (graviton); however, the detection of the graviton is far from feasible.

In the interior equation (3), the matter producing a gravitational field is represented by a tensor T_{ij} of class C^0. But there is also a way of representing it by singularities of a solution of the exterior equation (2). From this point of view, the equation of motion of a material particle (i.e., singularity) is not assumed a priori as in (4), but is derived as a result of (2) (A. Einstein, L. Infeld, and B. Hoffman, 1938).

In the static and weak field limits, the fundamental form is approximately given by Newton's gravitational potential φ as

$$ds^2 \fallingdotseq c^2(1-\varphi/c^2)\,dt^2 - (1+\varphi/c^2)(dx^2+dy^2+dz^2),$$

and (2) and (3) reduce in this limit to Laplace's and Poisson's equations, respectively. Newton's theory of gravity is valid in the limit $\varphi/c^2 \ll 1$.

Stimulated by the discoveries of neutron stars and black holes and by the big-bang theory of the universe in the 1960s, numerous studies of general relativity have been carried out on such problems as the gravitational field of a spinning mass, the dynamical process of gravitational collapse, the space-time structure of black holes, the generation of gravitational waves, the global structure and dynamics of the universe, and so on. A comparison of the theoretical predictions and the observations is generally favorable, but the phenomena in the universe are so complex that the effects of general relativity cannot always be isolated.

E. Solutions of Einstein's Equations

The isometric symmetry of space-time is described by †Killing vectors. The stationary metric is characterized by a timelike Killing vector, in which case equation (2) reduces to an †elliptic partial differential equation on a 3-dimensional manifold. If the space-time is axially symmetric as well as stationary, (2) reduces to the **Ernst equation**:

$$(\varepsilon+\varepsilon^*)\nabla^2\varepsilon = 2\nabla\varepsilon\cdot\nabla\varepsilon, \tag{5}$$

where ∇ represents divergence in a flat space. The metric tensors are derived from the complex potential ε. The solutions of (5) can be obtained using techniques developed for the soliton problem.

One example of stationary and axially symmetric solutions is the **Kerr metric**, which is written as

$$ds^2 = c^2\,dt^2 - \frac{2mr}{\rho^2}(a\sin^2\theta\,d\varphi - c\,dt)^2 - \rho^2\left(\frac{dr^2}{\Delta}+d\theta^2\right)^2 - (r^2+a^2)\sin^2\theta\,d\varphi^2, \tag{6}$$

with $\rho^2 = r^2 + a^2\cos^2\theta$ and $\Delta = r^2 - 2mr + a^2$. This metric solution represents a gravitational field around a spinning mass with mass $M = mc^2/G$ and angular momentum $J = Mac$. When $a = 0$, this metric reduces to the Schwarzschild metric.

Applying a Bäcklund transformation to the Kerr metric, an infinite series of stationary and axially symmetric solutions can be derived. All these solutions belong to the space-time metric with, in general, two Killing vectors.

The dynamical evolution of space-time structure has been studied by means of the †Cauchy problem of general relativity. Choosing appropriate dynamical variables, equation (2) or (3) is divided into constraint equations in terms of the initial data and evolution equations in terms of the dynamical variables. The latter †hyperbolic equations may also be written in Hamiltonian form.

A typical example of such a problem is the dynamics of a spatially homogeneous 3-

dimensional manifold; this has been studied as a cosmological model. The space-time with a constant scalar curvature R is called **de Sitter space**, and reduces to the Minkowski space if $R=0$. If the 3-dimensional space is isotropic as well as homogeneous, the metric takes the form

$$ds^2=c^2dt^2-a(t)^2\{dx^2+f(x)^2 \times(d\theta^2+\sin^2\theta\, d\varphi^2)\},$$

where $f(x)=\sin x$, x, or $\sinh x$. These are called **Robertson-Walker metrics** and are considered to describe a realistic expanding universe.

F. Global Structure of Space-Time

Following the advances of modern differential geometry, manifestly coordinate-independent techniques to analyze space-time properties have been applied to general relativity. The mathematical model of space-time is a connected 4-dimensional †Hausdorff C^∞-manifold endowed with a metric of signature (1,3). The metric allows the physical description of local causality and of local conservation of energy and momentum. The metric functions obey the Einstein field equation (2) or (3).

In order to clarify the global structure of the solutions of Einstein's equations, maximally analytic extension of the solutions has been studied. The maximal extension of the Schwarzschild metric is given as

$$ds^2=\frac{32m^3}{r}e^{-r/2m}(dv^2-du^2) -r^2(d\theta^2+\sin^2\theta\, d\varphi^2),$$

using the **Kruskal coordinates**, which are related to the coordinates of (6) by

$$\left(\frac{r}{2m}-1\right)e^{r/2m}=u^2-v^2,$$

$$\tanh t/4m=u/v \quad \text{or} \quad v/u.$$

To study the global structure at infinity, a †conformal mapping of the metric is used. For example, the Minkowski metric is written in the form $ds^2=\Omega^2 d\bar{s}^2$, where

$$d\bar{s}^2=dp\,dq-\tfrac{1}{4}\sin^2(p-q)(d\theta^2+\sin^2\theta\, d\varphi^2)$$

and $\Omega=\sec p\sec q$, $\tan p=t+r$, $\tan q=t-r$. By means of this mapping, all points, including infinity, are assigned finite p, q coordinate values in $-\pi/2\leqslant q\leqslant p\leqslant\pi/2$.

Singularities in space-time are one of the major problems concerning the global structure of the manifold. For some Cauchy problems relevant to cosmology and gravitational collapse, the inevitable occurrence of a singularity has been proved (**singularity theorem**). A sufficient condition for occurrence of a singularity is that there be some point p such that all the null geodesics starting from p converge to p again. In addition to this condition, for the proof of the singularity theorem it is presumed that the space-time is free of closed nonspacelike curves, that a Cauchy surface exists, and that the energy-momentum tensor satisfies the condition

$$\sum_{i,j}\left(T_{ij}-\frac{1}{2}Tg_{ij}\right)V^iV^j\geqslant 0$$

for any timelike vector V^i. The singularity whose existence is implied by this theorem means that the space-time manifold is geodesically incomplete (the space-time is complete if every geodesic can be extended to arbitrary values of its affine parameter).

The causal structure of space-time is also related to the global structure of the manifold. In this regard, black holes have been introduced as the final state of gravitational collapse. In the black-hole structure of space-time, there exists a closed surface called an **event horizon** in an asymptotically flat space. The event horizon is the boundary (the set of points) in space-time from which one can escape to infinity, or the boundary of the set of points that one can see from the infinite future. Then the black hole is a region from which no signal can escape to the exterior of the event horizon.

If we assume that singularities do not exist in the exterior of the event horizon, a stationary black-hole structure is uniquely described by the Kerr metric [6]. In the case of spherically symmetric collapse, this assumption is verified and the final metric is given by the Schwarzschild metric. However, it is not known whether this assumption is true in more general gravitational collapse.

References

[1] A. Einstein, The meaning of relativity, Princeton Univ. Press, fifth edition, 1956.
[2] H. Weyl, Space, time, matter, Dover, 1950. (Original in German, 1923.)
[3] A. Lichnerowicz, Théories relativistes de la gravitation et de l'électromagnétisme, Masson, 1955.
[4] C. W. Misner, K. S. Thorne, and J. A. Wheeler, Gravitation, Freeman, 1973.
[5] L. Witten (ed.), Gravitation, an introduction to current research, Wiley, 1962.
[6] C. DeWitt and B. S. DeWitt (eds.), Black holes, Gordon & Breach, 1972.
[7] S. W. Hawking and G. F. R. Ellis, The large scale structure of space-time, Cambridge Univ. Press, 1973.

[8] M. Carmeli, Group theory and general relativity, McGraw-Hill, 1977.
[9] S. W. Hawking and W. Israel (eds.), General relativity, an Einstein centenary survey, Cambridge Univ. Press, 1979.
[10] A. Held (ed.), General relativity and gravitation I, II, Plenum, 1980.

360 (XXI.8) Renaissance Mathematics

Toward the middle of the 13th century, scholastic theology and philosophy were at their height with the *Summa theologiae* of Thomas Aquinas (1225?–1274); but in the latter half of the century, the English philosopher Roger Bacon (1214–1294) attacked Aquinian philosophy in his *Opus majus*, insisted on the importance of experimental methods in science, and strongly urged the study of mathematics. The Renaissance flourished first in Italy, then in other European countries in the 15th and 16th centuries, in the domains of the arts and literature. Newer ideas in mathematics and the natural sciences dominated the 17th century. However, it was the invention of printing in the 15th century, the translation of the Greek texts of Euclid and Archimedes into European languages, and the importation of Arabian science into Europe during the Renaissance that prepared for this development.

In the 15th century, the German priest Nicolaus Cusanus (1401–1464) discussed infinity, the convergence of infinite series, and some problems of quadrature. During the same period, the German scholar Regiomontanus (1436–1476) wrote the first systematic treatise on trigonometry independent of astronomy. Leonardo da Vinci (1452–1519), the all-encompassing genius born in the same century, left manuscripts in which he wrote about mechanics, geometric optics, and perspective. Da Vinci's contemporary, the German painter A. Dürer (1471–1528), wrote a textbook on perspective. In 1494, L. Pacioli (1445?–1514) published *Summa de arithmetica*, one of the first printed books on mathematics. Its content, influenced by Arabian mathematics, includes practical arithmetic and double-entry bookkeeping. The book enjoyed wide popularity.

The best known result of 16th-century mathematics is the solution of algebraic equations of degrees 3 and 4 by the Italian mathematicians Scipione del Ferro, N. Tartaglia (1506–1557), G. Cardano (1501–1576), and L. Ferrari (1522–1565). Cardano published the solution of equations of the third degree in his book *Ars magna* (1545). The solution was due to Tartaglia, to whom acknowledgment was made, although publication of the method was against his will. This constitutes a famous episode in the history of mathematics, but what is historically more important is the fact that essential progress beyond Greek mathematics was made by mathematicians of this period, since the Greeks were able to solve equations only of degrees 1 and 2. Algebra was subsequently systematized by the French mathematician F. Viète (1540–1603).

By the end of the 15th century, practical mathematics (influenced by the Arabians) had become popular in Europe, and more advanced mathematics began to be studied in European universities, especially in Italy. In 1543, N. Copernicus (1473–1543) published his heliocentric theory (1543); G. Galilei (called Galileo) (1564–1642), the indomitable proponent of this theory, was also born in the 16th century. Copernicus studied at the Universities of Bologna, Padua, and Ferrara; Galileo studied at the University of Pisa and taught at the Universities of Pisa, Padua, and Florence. A system of numeration was imported from Arabia to Europe in the 13th century; by the time of S. Stevin (1548?–1620?) it took the definite form of a decimal system, and with the development and acceptance of printing, the forms of the numerals became fixed.

References

[1] M. Cantor, Vorlesungen über Geschichte der Mathematik II, Teubner, second edition, 1900.
[2] J. Cardan (G. Cardano), The book of my life (trans. J. Stoner), Dover, 1963.
[3] G. Cardano, Opera, 10 vols., Lyon, 1663.
[4] N. Bourbaki, Eléments d'histoire des mathématiques, Hermann, 1960.

361 (XX.32) Renormalization Group

A. Introduction

The concept of renormalization was introduced by S. Tomonaga, J. S. Schwinger, M. Gell-Mann, and F. E. Low in order to overcome the difficulty of divergence in field theory. If the upper bound of the momentum is limited to a finite cutoff value Λ, then physical quantities, for example, the mass m of an electron, can be obtained as finite quantities by letting

Λ go to infinity after summing all divergent terms. This is called the **renormalization method** using **subtraction**. Since the cutoff Λ is arbitrary insofar as it is finite, the Green's functions are indefinite because they depend on Λ. This dependence on the cutoff Λ corresponds to the response for the scale transformation of length, and this transformation is a certain (semi) group, called a **renormalization group**. Several kinds of renormalization group have been used in field theory, as well as in the statistical mechanics of phase transition.

B. Renormalization Group in Field Theory [1-3]

A typical method to resolve the ultraviolet divergence is to add **subtraction terms** in the Lagrangian so that they cancel the divergence. This cancellation is usually performed in each order of the perturbation expansion. When the addition of a finite number of subtraction terms cancels the divergence, the relevant Hamiltonian is said to be **renormalizable**. Otherwise it is called **unrenormalizable**. The Lagrangian density

$$\mathscr{L}_0=\frac{1}{2}\partial^\mu\varphi_0\partial_\mu\varphi_0-\frac{1}{2}m_0^2\varphi_0^2-\frac{1}{4!}g_0\varphi_0^4,$$

for example, can be renormalized by the transformation $\varphi_0=Z_3^{1/2}\varphi$, $g_0=Z_1Z_3^{-2}g$, and $m_0^2=m^2+\delta m^2$. All the divergences are taken into the renormalization constants Z_3, Z_1, and δm^2, so that the renormalized quantities φ, g, and m are finite. These renormalization constants can be calculated by means of a perturbation method, but the requirement of circumventing the divergences alone is not sufficient to determine them explicitly. This indeterminacy is usually expressed as $Z_3(\mu)$ and $Z_1(\mu)$, i.e., in terms of a parameter μ, called the renormalization point. The μ-dependence of these functions can be determined by means of the following renormalization conditions:

$$\lim_{p^2+\mu^2\to 0}(p^2+m^2)G^{(2)}(p,\mu)=1;$$

$$\Gamma(p_i,\mu)=1,\quad p_ip_j=\frac{\mu^2}{3}(1-4\delta_{ij}).$$

Since the renormalization constants depend on the continuous parameter μ, the renormalized Green's function and coupling constant g are also functions of μ, and consequently the quantity $d^{(2N)}(\{p_i\},\mu,g(\mu))$ defined by

$$d^{(2N)}(\{p_i\},\mu,g(\mu))$$
$$\equiv\left[\prod_{i=1}^{2N}(p_i^2+m^2)\right]G^{(2N)}(p_i,\mu,g(\mu))$$

satisfies the **renormalization equation**

$$\left(\mu\frac{\partial}{\partial\mu}+\beta\frac{\partial}{\partial g}-2N\gamma\right)d^{(2N)}(\{p_i\},\mu,g)=0,$$

where $\beta=\mu\,dg/d\mu$ and $\gamma=\frac{1}{2}\mu\,d\log Z_3(\mu)/d\mu$. If the coefficients β and γ are calculated perturbationally up to a certain order, the renormalized Green's function is obtained up to the same order by solving the foregoing renormalization equation. This is the first kind of renormalization group. A second kind expresses the response of the renormalized Green's function to the change in the mass and coupling constant, and is expressed by the **Callan-Symanzik equation** [4, 5]

$$\left(m\frac{\partial}{\partial m}+\beta(g)\frac{\partial}{\partial g}+2N\gamma_\varphi(g)\right)G_R^{(2N)}=\Delta G_G^{(2N)},$$

where $\beta(g)=Zm_0\partial g/\partial m_0$ and $\gamma_\varphi(g)=\frac{1}{2}Zm_0\cdot\partial\log Z_3/\partial m_0$, and where Z is determined by $Zm_0\partial m/\partial m_0=m$. The inhomogeneous term is defined by

$$\Delta G_R^{(2N)}=Zm_0\frac{\partial}{\partial m_0}G_R^{(2N)}$$
$$+Zm_0\frac{\partial}{\partial m_0}(N\log Z_3)G_R^{(2N)}.$$

The irreducible Green's function $\Gamma_R^{(2N)}$ satisfies

$$\left(m\frac{\partial}{\partial m}+\beta(g)\frac{\partial}{\partial g}-2N\gamma_\varphi(g)\right)\Gamma_R^{(2N)}=i\Delta\Gamma_R^{(2N)}.$$

Since the inhomogeneous term can be neglected in the high-energy region, the foregoing equation becomes homogeneous, and consequently its solution is

$$\Gamma_R^{(2N)}\left(p_i,\frac{m}{\lambda},g\right)$$
$$=\exp\left[-2N\int_g^{g(\lambda)}\gamma(g')\beta^{-1}(g')dg'\right]$$
$$\times\Gamma_R^{(2N)}(p_i,m,g(\lambda)).$$

Here $g(\lambda)$ is the solution of the equation $\int_g^{g(\lambda)}\beta^{-1}(g')dg'=\log\lambda$. Since dimensional analysis yields $\Gamma_R^{(2N)}(\lambda p_i,m,g)=\lambda^{4-2N}\Gamma_R(p_i,m/\lambda,g)$, the foregoing solution shows that high-energy phenomena can be described by the low-energy phenomena whose coupling constant is given by $g_\infty\equiv g(\infty)$. In particular, when $g_\infty=0$, high-energy phenomena are described by the asymptotically free field. This circumstance is called **asymptotic freedom**.

C. Renormalization Group Theory in Statistical Physics

The renormalization group technique has proved to be powerful in statistical physics,

particularly in studies of phase transitions and critical phenomena [6–10]. The correlation length ξ diverges like $\xi(T) \sim (T-T_c)^{-\nu}$ near the critical point T_c, where ν is called the **critical exponent** of ξ. Similarly, the correlation function $C(R)$ for the distance R behaves like $C(R) \sim R^{-(d-2+\eta)} \times \exp(-\kappa R)$, $\kappa = \xi^{-1}$, where d is the dimensionality of the system and η is the exponent describing the deviation of the singular behavior of $C(R)$ from classical theory. Renormalization is useful in evaluating these critical exponents systematically. The fundamental idea is to eliminate some degrees of freedom, to find recursion formulas for interaction parameters, and then to evaluate critical exponents from their asymptotic behavior near the fixed point. There are many different ways of carrying out this idea explicitly. Roughly classifying these into two groups, we have (i) momentum-space renormalization group theories [6–10] and (ii) real-space renormalization group theories [10, 11].

The common fundamental structure of these renormalization group techniques is explained as follows. First the momentum space or real space is divided into cells and rapidly fluctuating parts, namely, small-momentum parts inside each cell are integrated or eliminated, and consequently the remaining slowly fluctuating parts, namely, long-wave parts, are renormalized. The original Hamiltonian $\mathscr{H}_0$ is transformed into $\mathscr{H}_1$ by means of this elimination process and by some scale transformation that preserves the phase space volume. This renormalization operation is written as $\mathbf{R}_b$: i.e., $\mathscr{H}_1 = \mathbf{R}_b \mathscr{H}_0$. Similarly we have $\mathscr{H}_2 = \mathbf{R}_b \mathscr{H}_1 = \mathbf{R}_b^2 \mathscr{H}_0, \ldots, \mathscr{H}_n = \mathbf{R}_b \mathscr{H}_{n-1} = \mathbf{R}_b^n \mathscr{H}_0, \ldots$. This transformation $\mathbf{R}_b$ has the (semi) group property $\mathbf{R}_{bb'} = \mathbf{R}_b \mathbf{R}_{b'}$. A generator $\mathbf{G}$ is defined by $\mathbf{G} = \lim_{b \to 1+0} (\mathbf{R}_b - 1)/(b-1)$. That is, $\mathbf{R}_b = \exp(l\mathbf{G})$, $e^l = b$. The transformation of $\mathscr{H}$ is expressed as $d\mathscr{H}/dl = \mathbf{G}[\mathscr{H}]$. The fixed point $\mathscr{H}^* = \mathbf{R}_b \mathscr{H}^*$ is the solution of $\mathbf{G}[\mathscr{H}^*] = 0$. In order to find critical exponents from the asymptotic behavior of $\mathbf{G}$ near $\mathscr{H}^*$, we consider a Hamiltonian of the form $\mathscr{H} = \mathscr{H}^* + wQ$ and expand $\mathbf{G}[\mathscr{H}]$ as $\mathbf{G}[\mathscr{H}^* + wQ] = w\mathbf{K}Q + O(w^2)$. If the operator $\mathbf{K}$ thus defined has a negative eigenvalue λ_i, the corresponding physical quantity Q_i becomes irrelevant after repeating the renormalization procedure, and the physical quantity Q_j corresponding to a positive eigenvalue $\lambda_j > 0$ becomes relevant. Thus, by introducing a field h_j conjugate to the relevant operator Q_j, we study the Hamiltonian $\mathscr{H} = \mathscr{H}^* + \sum_j h_j Q_j$. The free energy per unit volume $f[\mathscr{H}] \equiv f(h_1, h_2, \ldots)$ is found to have the scaling property

$$f(h_1, \ldots, h_j, \ldots) \simeq b^{-d} f(b^{\lambda_1} h_1, \ldots, b^{\lambda_j} h_j, \ldots).$$

By taking Q_1 as the energy operator, we have $h_1 \sim T - T_c \equiv t$. By the normalization $b^{\lambda_1} h_1 = 1$, we obtain the scaling law

$$f(t, \ldots, h_j, \ldots) \simeq t^{d/\lambda_1} f(1, \ldots, h_j/t^{\varphi_j}, \ldots).$$

The critical exponent of the specific heat defined by $C \sim t^{-\alpha}$ is given by the formula $\alpha = 2 - d/\lambda_1$. Other scaling exponents $\{\varphi_j\}$ can be obtained via the formula $\varphi_j = \lambda_j/\lambda_1$ from the eigenvalues of $\mathbf{K}$. The simplest example of R_b is the case where a single interaction parameter K is transformed into a new parameter K' by $K' = f_b(K)$. The fixed point K^* is given by the solution of $K^* = f_b(K^*)$. The correlation exponent ν defined by $\xi \sim (K - K^*)^{-\nu}$ is given by the Wilson formula $\nu = \log b / \log \Lambda$, $\Lambda = (df_b/dK)_{K=K^*}$. In most cases, $\mathbf{R}_b$ is constructed perturbationally, and critical exponents are usually calculated in power series of $\varepsilon \equiv d_c - d$, as $\varphi = \varphi_0 + \varphi_1 \varepsilon + \varphi_2 \varepsilon^2 + \ldots$, where d_c denotes the critical dimension. This is called the **ε-expansion**. The first few terms are calculated explicitly for specific models, such as the φ^4-model. By applying the Borel sum method to these ε-expansions, one can estimate critical exponents [10, 11].

The renormalization group method can be applied to other many-body problems, such as the Kondo effect [12].

References

[1] E. C. G. Stueckelberg and A. Petermann, La normalisation des constantes dans la théorie des quanta, Helv. Phys. Acta, 26 (1953), 499–520.

[2] M. Gell-Mann and F. E. Low, Quantum electrodynamics at small distances, Phys. Rev., 95 (1954), 1300–1312.

[3] N. N. Bogolyubov and D. V. Shirkov, Introduction to the theory of quantized fields, Interscience, 1959.

[4] C. G. Callan, Jr., Broken scale invariance in scalar field theory, Phys. Rev., D2 (1970), 1541–1547.

[5] K. Symanzik, Small distance behaviour in field theory and power counting, Comm. Math. Phys., 18 (1970), 227–246; Small distance-behaviour analysis and Wilson expansions, Comm. Math. Phys., 23 (1971), 49–86.

[6] C. DiCastro and G. Jona-Lasinio, On the microscopic foundation of scaling laws, Phys. Lett., 29A (1969), 322–323; C. DiCastro, The multiplicative renormalization group and the critical behavior in $d = 4 - \varepsilon$ dimensions, Lett. Nuovo Cimento, 5 (1972), 69–74.

[7] K. G. Wilson, Renormalization group and critical phenomena. I, Renormalization group and the Kadanoff scaling picture, Phys. Rev.,

B4 (1971), 3174–3183; II, Phase-space cell analysis of critical behavior, Phys. Rev., B4 (1971), 3184–3205.
[8] K. G. Wilson and M. E. Fisher, Critical exponents in 3.99 dimensions, Phys. Rev. Lett., 28 (1972), 240–243.
[9] K. G. Wilson and J. Kogut, The renormalization group and the ε expansion, Phys. Rep., 12C (1974), 75–200.
[10] C. Domb and M. S. Green (eds.), Phase transitions and critical phenomena VI, Academic Press, 1976.
[11] A. A. Migdal, Recursion equations in gauge field theories, Sov. Phys. JETP, 42 (1975), 413–418; Phase transitions in gauge and sublattice systems, Sov. Phys. JETP, 42 (1975), 743–746. (Original in Russian, 1975.)
[12] K. G. Wilson, The renormalization group: Critical phenomena and the Kondo problem, Rev. Mod. Phys., 47 (1975), 773–840.
[13] L. P. Kadanoff, Notes on Migdal's recursion formulas, Ann. Phys., 100 (1976), 359–394.

362 (IV.16) Representations

A. General Remarks

For a mathematical system A, a mapping from A to a similar (but in general "more concrete") system preserving the structure of A is called a **representation** of A. In this article, we consider the representations of †groups and †associative algebras. For representations of other algebraic systems → 42 Boolean Algebras; 231 Jordan Algebras; 248 Lie Algebras. For topological, analytic, and algebraic groups → 13 Algebraic Groups; 69 Compact Groups; 249 Lie Groups; 422 Topological Abelian Groups; 423 Topological Groups; 437 Unitary Representations. For specific groups → 60 Classical Groups; 61 Clifford Algebras.

B. Permutation Representations of Groups

We denote by $\mathfrak{S}_M$ the group of all †permutations of a set M (→ 190 Groups B). A **permutation representation** of a group G in M is a homomorphism $G \to \mathfrak{S}_M$. We denote by a_M the permutation of M corresponding to $a \in G$ and write $a_M(x) = ax$ ($x \in M$). Then we have a condition $(ab)x = a(bx)$, $1x = x$ ($a, b \in G$, 1 is the identity element, $x \in M$). In general, if the product $ax \in M$ of $a \in G$ and $x \in M$ is defined and satisfies this condition, then G is said to **operate on** M **from the left**, and M is called a **left** G**-set**. Giving a permutation representation of G in M is equivalent to giving the structure of a left G-set to M. A **reciprocal permutation representation** of G in M is an †antihomomorphism $G \to \mathfrak{S}_M$, which becomes a homomorphism if we define the multiplication in $\mathfrak{S}_M$ by the right notation $x(fg) = (xf)g$. If the product $xa \in M$ of $a \in G$ and $x \in M$ is defined and satisfies the conditions $x(ab) = (xa)b$ and $x1 = x$, then, as before, G is said to **operate on** M **from the right**, and M is called a **right** G**-set**. Giving a reciprocal permutation representation of G in M is equivalent to giving the structure of a right G-set to M.

A (reciprocal) permutation representation is said to be **faithful** if it is injective; the corresponding G-set is also said to be faithful. In particular, we can take G itself as M and define the left (right) operation by the multiplication from the left (right). Then we have a faithful permutation representation (reciprocal permutation representation), which is called the **left (right) regular representation** of G. For $a \in G$, the induced permutation $a_G: x \to ax$ (xa) is called the **left (right) translation** by a.

We call a left G-set simply a G-set. If a subset N of a G-set M satisfies the condition that $a \in G$, $x \in N$ implies $ax \in N$, then N forms a G-set, which is called a G**-subset** of M. If a G-set M has no proper G-subset (i.e., one different from M itself and the empty subset), then for any two elements $x, y \in M$ there exists an element $a \in G$ satisfying $ax = y$. In this case, the operation of G on M is said to be **transitive**, and the corresponding permutation representation is also said to be transitive. If an equivalence relation R in a G-set M is compatible with the operation of G (i.e., R satisfies the condition that $a \in G$, $R(x, y)$ implies $R(ax, ay)$), then the quotient set M/R forms a G-set in the natural way, called the **quotient** G**-set** of M by R. If a G-set M has no nontrivial quotient G-set, i.e., if the only equivalence relations compatible with the operation are

$$R(x, x') \quad \text{for any} \quad x, x' \in M$$

and

$$R(x, x') \quad \text{if and only if} \quad x = x',$$

then the operation of G on M and the corresponding permutation representation are said to be **primitive**.

A mapping $f: M \to M'$ of G-sets is called a G**-mapping** (G**-map**) if the condition $f(ax) = af(x)$ ($a \in G, x \in M$) is satisfied. G-injection, G-surjection, and G-bijection are defined naturally. The inverse mapping of a G-bijection is also a G-bijection. Two permutation representations are said to be **similar** if there exists a G-bijection between the corresponding G-sets.

Let M be a transitive G-set, and fix any element $x \in M$. If we view G as a G-set, the mapping $f: G \to M$ defined by $f(a) = ax$ is a G-surjection and induces a G-bijection $\bar{f}: G/R \to M$. Here an equivalence class of R is precisely a left coset of the **stabilizer (stability group** or **isotropy group)** $H_x = \{a \in G \mid ax = x\}$. Hence we have a G-bijection $G/H_x \to M$. Conversely, for any subgroup H of G, G/H is a transitive left G-set. A transitive G-set is called a **homogeneous space** of G.

For a family $\{M_\lambda\}_{\lambda \in \Lambda}$ of G-sets, the Cartesian product $\prod_{\lambda \in \Lambda} M_\lambda$ and the †direct sum $\sum_{\lambda \in \Lambda} M_\lambda$ become G-sets in the natural way; they are called the **direct product** of G-sets and the **direct sum** (i.e., disjoint union) of G-sets, respectively. Every G-set M is the direct sum of a family $\{M_\lambda\}$ of transitive G-subsets, and each M_λ is called an **orbit** (or **system of transitivity**). For a G-set M, the direct product G-set $M^k = M \times \ldots \times M$ (k times) contains a G-subset $M^{(k)} = \{(x_1, \ldots, x_k) \mid i \neq j \text{ implies } x_i \neq x_j\}$. If $M^{(k)}$ is transitive, M is said to be **k-ply transitive**. If M is transitive and the stabilizer of each point of M consists of the identity element alone, M is said to be **simply transitive**.

If M has n elements, a permutation representation of a group G in M is said to be of **degree** n. When G is a group of permutations of M, the canonical injection $G \to \mathfrak{S}_M$ is a faithful permutation representation; this case has been studied in detail (→ 151 Finite Groups G).

C. Linear Representations of Groups and Associative Algebras

Let K be a †commutative ring with unity element and M be a K-module. Though we shall mainly treat the case where K is a field and M is a finite-dimensional †linear space over K, the case where K is an †integral domain and M is a †free module over K of finite rank is also important. Since K is commutative, we can write $\lambda x = x\lambda$ ($\lambda \in K, x \in M$). Let $\mathscr{E}_K(M)$ be the †associative algebra over K consisting of all K-endomorphisms of M, and let $GL(M)$ be the group of all †invertible elements in $\mathscr{E}_K(M)$, where we assume $M \neq \{0\}$. Let A be an associative algebra over K. A **linear representation of the algebra** A in M is an algebra homomorphism $A \to \mathscr{E}_K(M)$. We always assume that A has a unity element and the homomorphisms are unitary. For convenience, we can also consider a linear representation in the trivial space $M = \{0\}$, which is called the **zero representation.** A **reciprocal linear representation** is an antihomomorphism $A \to \mathscr{E}_K(M)$. A **linear representation of a group** G in M is a group homomorphism $G \to GL(M)$. This can be extended uniquely to a linear representation of the †group ring $K[G]$ in M, and conversely, the restriction of a linear representation of $K[G]$ in M to G is a linear representation of G; and similarly for reciprocal linear representations. Thus the study of (reciprocal) linear representations of a group G in M can be reduced to the study of (reciprocal) linear representations of the group ring $K[G]$ in M.

We now consider the linear representation of associative algebras, which we call simply "algebras." (Note that a group ring has a canonical basis—the group itself—and allows a more detailed investigation; → Sections G, I.)

Given a commutative ring K with unity and a linear representation ρ of a K-algebra A in a K-module M, we introduce the structure of a left A-module into M by defining $ax = \rho(a)x$ ($a \in A, x \in M$); the structure of a K-module in M obtained by the canonical homomorphism $K \to A$ coincides with the original one. This A-module is called the **representation module** of ρ. Conversely, for any left A-module M we can define a linear representation ρ of A in M (with M viewed as a K-module via $K \to A$) by putting $\rho(a)x = ax$; the representation module of ρ coincides with the original one. This representation ρ is called the **linear representation associated with** M. A reciprocal linear representation of A corresponds to a right A-module. Thus the study of (reciprocal) linear representations of A is equivalent to the study of left (right) A-modules. For instance, if the operation of a group G on M is trivial: $\sigma x = x$ ($\sigma \in G, x \in M$), the corresponding representation of G in M assigns the identity mapping I_M to every $\sigma \in G$. Furthermore, if $M = K$, this representation is called the **unit representation** of G (over K).

Let ρ, ρ' be linear representations of A in K-modules M, M', respectively. Then an A-homomorphism $M \to M'$ is precisely a K-homomorphism $f: M \to M'$ satisfying the condition $f \circ \rho(a) = \rho'(a) \circ f$ ($a \in A$); this is sometimes called a **homomorphism** from ρ to ρ'. In particular, an A-isomorphism is a K-isomorphism $f: M \to M'$ satisfying the condition $f \circ \rho(a) \circ f^{-1} = \rho'(a)$ ($a \in A$); in this case we say that ρ and ρ' are **similar (isomorphic** or **equivalent)** and write $\rho \cong \rho'$.

Let M be the representation module of a linear representation ρ of an algebra A. If ρ is injective, ρ and the corresponding M are said to be **faithful**. For example, the linear representation associated with the left A-module A is faithful; this is called the **(left) regular representation** of A. If M is †simple as an A-module, ρ is said to be **irreducible** (or **simple**). A homomorphism from an irreducible representation ρ to ρ must be an isomorphism or the zero

homomorphism (**Schur's lemma**). In particular, if K is an †algebraically closed field and M is finite-dimensional, then such a homomorphism is a scalar multiplication. A linear representation is said to be **reducible** if it is not irreducible. If M is †semisimple as an A-module, ρ is said to be **completely reducible** (or **semisimple**). If A is a semisimple ring, any linear representation of A is completely reducible. The converse also holds (→ 368 Rings G).

The linear representations associated with a submodule and a quotient module of M as an A-module are called a **subrepresentation** and a **quotient representation**, respectively. The linear representation associated with the direct sum $M_1+\ldots+M_r$ of the representation modules $M_1,\ldots,M_r$ of linear representations $\rho_1,\ldots,\rho_r$ is written $\rho_1+\ldots+\rho_r$ and called the **direct sum** of representations. If ρ is never similar to the direct sum of two nonzero linear representations, then ρ is said to be **indecomposable**; this means that M is †indecomposable as an A-module.

For linear representations ρ, ρ' of a group G in M, M', we define the linear representation $\rho\otimes\rho'$ in $M\otimes M'$ by $(\rho\otimes\rho')(g)=\rho(g)\otimes\rho'(g)$ $(g\in G)$; this is called the **tensor product** of representations ρ and ρ'.

D. Matrix Representations

Let K^n be the K-module consisting of all n-tuples (ξ_i) of elements in a commutative ring K. $\mathscr{E}_K(K^n)$ is identified with the K-algebra $M_n(K)$ of all $n\times n$ matrices (λ_{ij}) over K: $(\lambda_{ij})(\xi_j)=(\sum_{j=1}^n\lambda_{ij}\xi_j)$. Thus a linear representation of A in K^n, i.e., a homomorphism $A\rightarrow M_n(K)$, is called a **matrix representation** of A over K, and n is called its **degree**. A matrix representation of a group G over K of degree n is a homomorphism $G\rightarrow GL(n,K)$, where $GL(n,K)$ is the group of all $n\times n$ invertible matrices. If $(e_1,\ldots,e_n)$ is a †basis of a K-module M, then by the K-isomorphism $K^n\rightarrow M$ given by the assignment $(\xi_i)\rightarrow\sum_{i=1}^n e_i\xi_i$, we have a bijective correspondence between the matrix representations of A of degree n and the linear representations of A in M, and the corresponding representations are similar. Explicitly, the linear representation ρ corresponding to a matrix representation $a\rightarrow(\lambda_{ij}(a))$ is given by

$$\rho(a)e_j=\sum_{i=1}^n e_i\lambda_{ij}(a),\quad a\in A.$$

Hence giving the finite-dimensional linear representations over a field K is equivalent to giving the matrix representations over K. Let T, T' be matrix representations of degree n, n'. Then a homomorphism from T to T' is an $n'\times n$ matrix P satisfying $PT(a)=T'(a)P$ $(a\in A)$. Therefore T and T' are similar if and only if $n=n'$ and $PT(a)P^{-1}=T'(a)$ $(a\in A)$ for some $n\times n$ invertible matrix P. For a representation of a group G, it suffices that this equation is satisfied by all $a\in G$.

We always assume that K is a field. Then a K-module is a linear space over K. A linear representation ρ over K of a K-algebra A is said to be of **degree** n if its representation module M is of dimension n over K. Suppose that a sequence $\{0\}=M_0\subset M_1\subset\ldots\subset M_r=M$ of A-submodules of M is given. We take a basis $(e_1,\ldots,e_n)$ of M over K such that $(e_1,\ldots,e_{m_i})$ forms a basis of M_i over K $(1\leqslant i\leqslant r)$. Then the matrix representation corresponding to ρ relative to the basis $(e_1,\ldots,e_n)$ has the form

$$a\rightarrow T(a)=\begin{bmatrix} T_{11}(a) & T_{12}(a) & \ldots & T_{1r}(a)\\ & T_{22}(a) & \ldots & T_{2r}(a)\\ & & \ddots & \vdots\\ 0 & & & T_{rr}(a)\end{bmatrix}$$

where, if we put $n_i=\dim M_i/M_{i-1}=m_i-m_{i-1}$, $T_{ij}(a)$ is an $n_i\times n_j$ matrix and $T_{ij}(a)=0$ for $i>j$. The residue classes of $e_{m_{i-1}+1},\ldots,e_{m_i}$ form a basis of the quotient space M_i/M_{i-1} over K, and the matrix representation corresponding to the linear representation ρ_i associated with M_i/M_{i-1} relative to this basis is given by T_{ii}. The sequence $\{M_i\}$ is a †composition series if and only if each ρ_i (hence T_{ii}) is irreducible. In this case, $\rho_1,\ldots,\rho_r$ are uniquely determined by ρ up to their order and similarity (**Jordan-Hölder theorem**). An irreducible representation ρ' similar to some ρ_i is called an **irreducible component** of ρ. The number $p>0$ of ρ_i similar to ρ' is called the **multiplicity** of ρ' as an irreducible component of ρ. We also say that ρ contains ρ' p times or ρ' appears p times as an irreducible component of ρ. The representation ρ is completely reducible if and only if it is similar to the direct sum of its irreducible components (admitting repetition). In this case, ρ is similar to the matrix representation

$$a\rightarrow\begin{bmatrix} T_{11}(a) & & 0\\ & \ddots & \\ 0 & & T_{rr}(a)\end{bmatrix}.$$

E. Coefficients and Characters of Linear Representations

We consider the linear representations of an algebra over a field K. A right (left) A-module M is regarded as a linear space over K. In its dual space M^*, we introduce the structure of a left (right) A-module using the inner product $\langle\ ,\ \rangle$ as follows: $\langle x,ax^*\rangle=\langle xa,x^*\rangle$ $(\langle x,x^*a\rangle$

$=\langle ax, x^*\rangle$), where $a\in A$, $x\in M$, $x^*\in M^*$. If ρ is the representation associated with M, the representation associated with M^* is called the **transposed representation (dual representation** or **adjoint representation)** of ρ, and is denoted by ${}^t\rho$. The linear mapping ${}^t\rho(a)$ is the †transposed mapping of $\rho(a)$. If M is finite-dimensional over K, we have $(M^*)^*=M$ as an A-module. For a linear representation ρ of a group G, the mapping $g\rightarrow{}^t\rho(g)^{-1}$ $(g\in G)$ is called the **contragredient representation** of ρ. The reciprocal linear representation associated with the right A-module A is called the **right regular representation** of A, and its transposed representation (i.e., the representation associated with the left A-module A^*) is called the **coregular representation** of A. For any finite-dimensional semisimple algebra and group ring of a finite group, the regular representation and the coregular representation are similar (→ 29 Associative Algebras H).

Let ρ be a linear representation of A over K and M be its representation module. For any $x\in M$, $x^*\in M^*$, we define a †linear form $\rho_{x,x^*}\in A^*$ on A by $\rho_{x,x^*}(a)=\langle ax, x^*\rangle$ $(a\in A)$. This is called the **coefficient** of ρ relative to x, x^* and is determined by its values at generators of A as a linear space. In particular, a coefficient of a linear representation ρ of a group G can be regarded as a function on G taking values in K. For a fixed $x^*\in M^*$ the assignment $x\rightarrow\rho_{x,x^*}$ gives an A-homomorphism $M\rightarrow A^*$, where A^* is considered as a left A-module. Therefore any nonzero coefficient ρ_{x,x^*} of an irreducible representation ρ generates an A-submodule of A^* isomorphic to M. In other words, any irreducible representation of A is similar to some subrepresentation of the coregular representation of A. In particular, any irreducible representation of a finite-dimensional semisimple algebra or a finite group is an irreducible component of the regular representation. Let A_ρ^* be the subspace of A^* generated by all coefficients ρ_{x,x^*} $(x\in M, x^*\in M^*)$ for a given linear representation ρ. Then $\rho\cong\rho'$ implies $A_\rho^*=A_{\rho'}^*$. If $\rho_1,\dots,\rho_r$ are irreducible representations of A such that ρ_i and ρ_j are not similar unless $i=j$, the sum $A_{\rho_1}^*+\dots+A_{\rho_r}^*$ in A^* is direct. In particular, for a semisimple algebra A, let the ρ_i $(1\leqslant i\leqslant r)$ be the irreducible representations associated with the minimal left ideals of the †simple components A_i of A. Then any irreducible representation is similar to one and only one of $\rho_1,\dots,\rho_r$, and A^* can be decomposed into the direct sum of $A_{\rho_1}^*,\dots,A_{\rho_r}^*$. In addition, each $A_{\rho_i}^*$ is canonically identified with A_i^*.

We shall treat finite-dimensional representations exclusively. Let $(e_1,\dots,e_n)$ be a basis of the representation module M of ρ over K, and let $a\rightarrow T(a)=(\lambda_{ij}(a))$ be the matrix representation that corresponds to ρ with respect to this basis. Then $\lambda_{ij}=\rho_{e_j,e_i^*}$ $(1\leqslant i,j\leqslant n)$, where $(e_1^*,\dots,e_n^*)$ is the dual basis. If K is algebraically closed and ρ is irreducible (or more generally, †absolutely irreducible), then $\{\lambda_{ij}\}$ is linearly independent; therefore we have $\dim A_\rho^*=n^2$ (G. Frobenius and I. Schur). We take a matrix representation T corresponding to ρ and put $\chi_\rho(a)=\operatorname{tr}T(a)$ $(a\in A)$. Then χ_ρ is a function on A that is uniquely determined by ρ and belongs to A_ρ^*; χ_ρ is called the **character** of ρ. For a linear representation ρ of a group G, the character of ρ can be regarded as a function on G. Moreover, it can be viewed as a function on the set of all †conjugate classes of G. The character of ρ is equal to the sum of the characters of the irreducible components of ρ taken with their multiplicities. The character of an irreducible representation is called an **irreducible character** (or **simple character**). If K is of characteristic 0, then $\rho\cong\rho'$ is equivalent to $\chi_\rho=\chi_{\rho'}$, and the different irreducible characters are linearly independent. The character of an absolutely irreducible representation (→ Section F) is called an **absolutely irreducible character**. If we consider absolutely irreducible characters only, the statement holds irrespective of the characteristic of K.

The sum of all absolutely irreducible characters of A is called the **reduced character** (or **reduced trace**) of A. The direct sum of all absolutely irreducible representations of A is called the **reduced representation** of A, and its character is equal to the reduced character. The determinant of the reduced representation is called the **reduced norm** of A.

F. Scalar Extension of Linear Representations

Let K, L be commutative rings with unity element, and fix a homomorphism $\sigma: K\rightarrow L$. We denote by M^σ the scalar extension $\sigma^*(M)=M\otimes_K L$ of a K-module M relative to $\sigma: x\lambda\otimes\mu=x\otimes\lambda^\sigma\mu$ $(x\in M; \lambda\in K, \mu\in L)$ (→ 277 Modules L). For an algebra A over K, the scalar extension A^σ of the K-module A has the natural structure of an algebra over $L:(a\otimes\lambda)(b\otimes\mu)=ab\otimes\lambda\mu$ $(a,b\in A; \lambda,\mu\in L)$. For a group G, we can regard $(K[G])^\sigma=L[G]: g\otimes\lambda=g\lambda$ $(g\in G, \lambda\in L)$. If M is a left A-module, then M^σ has the natural structure of a left A^σ-module; $(a\otimes\lambda)(x\otimes\mu)=ax\otimes\lambda\mu$ $(a\in A, x\in M; \lambda,\mu\in L)$. For the linear representation ρ associated with M, the linear representation ρ^σ over L associated with M^σ is called the **scalar extension** of ρ relative to $\sigma: \rho^\sigma(a\otimes 1)=\rho(a)\otimes 1_L$. Let $(e_1,\dots,e_n)$ be a basis of M over K. If the matrix representation $a\rightarrow(\lambda_{ij}(a))$ corresponds to M relative to this basis, then the matrix representation corre-

sponding to M^σ relative to the basis $(e_1 \otimes 1, \ldots, e_n \otimes 1)$ over L is given by $a \otimes 1 \rightarrow (\lambda_{ij}(a)^\sigma)$. A linear representation over L is said to be **realizable** in K if it is similar to the scalar extension ρ^σ of some linear representation ρ over K.

In particular, if $\sigma: K \rightarrow L$ is an isomorphism, ρ^σ is called the **conjugate representation** of ρ relative to σ. The conjugate representation relative to the automorphism $\sigma: \lambda \rightarrow \bar{\lambda}$ (complex conjugation) of the complex number field is called the **complex conjugate representation.** If $\mathfrak{m}$ is an ideal of K and $\sigma: K \rightarrow K/\mathfrak{m}$ (†residue class ring) is the canonical homomorphism, then the construction of ρ^σ from ρ is called the **reduction modulo** $\mathfrak{m}$ (→ Section I). If $\mathfrak{p}$ is a prime ideal of K and $\sigma: K \rightarrow K_{\mathfrak{p}}$ (†local ring) is the canonical homomorphism, then the construction of ρ^σ from ρ is called the **localization** relative to $\mathfrak{p}$. If K is an integral domain and $\mathfrak{p} = K - \{0\}$, then $K_{\mathfrak{p}}$ is the †field of quotients of K. We can also consider the "completion of representation" with respect to $\mathfrak{p}$.

Let K be a field, L a field extension, and $\sigma: K \rightarrow L$ the canonical injection. Then for a linear representation ρ of A and its representation module M, the scalar extensions ρ^σ, M^σ are written ρ^L, M^L, respectively. In view of $M \subset M^L$, $A \subset A^L$, $\mathscr{E}_K(M) \subset \mathscr{E}_L(M^L)$ by the natural injections, we can regard ρ^L as an extension of the mapping ρ. We shall consider finite-dimensional representations exclusively. For linear representations ρ_1, ρ_2 over K, $\rho_1 \cong \rho_2$ is equivalent to $\rho_1^L \cong \rho_2^L$. An irreducible representation ρ over K is said to be **absolutely irreducible** if its scalar extension ρ^L to any field extension L is irreducible; an equivalent condition is that the scalar extension $\rho^{\bar{K}}$ to the †algebraic closure $\bar{K}$ is irreducible. Another equivalent condition is that every endomorphism of the representation module M of ρ must be a scalar multiplication. If every irreducible representation of A over K is absolutely irreducible, K is called a **splitting field** for A. For a group G, if the field K is a splitting field for the group ring $K[G]$, then K is called a splitting field for G. Let A be finite-dimensional over K. If K is a splitting field for A, any irreducible representation of A^L is realizable in K for any field extension L of K. For an arbitrary field K, the scalar extension ρ^L of an irreducible representation ρ to a †separable algebraic extension L of K is completely reducible. For simplicity, we assume that K is †perfect and $L = \bar{K}$. Then the multiplicities of all irreducible components of ρ^L are the same; this multiplicity is called the **Schur index** of ρ.

The set $S(K)$ of †algebra classes over K, each of which is represented by a (central) simple component of the group algebra $K[G]$ of some finite group G, is a subgroup of the †Brauer group $B(K)$ of K, known as the **Schur subgroup** of $B(K)$. Recent research has clarified considerably the structure of this group [19].

G. Linear Representations of Finite Groups

Let G be a finite group of order g. The linear representation of G over K is equivalent to the linear representation of the group ring $K[G]$, concerning which we have already stated the general facts. If K is the ring $\mathbf{Z}$ of rational integers, a linear representation over K is sometimes called an **integral representation**. We assume that K is a field. If the characteristic of K is zero or more generally not a divisor of g, every linear representation of G over K is completely reducible (H. Maschke). Such a representation is called an **ordinary representation.** If g is divisible by the characteristic of K, we have a **modular representation** (→ Section I).

The **exponent** of G is the smallest positive integer n satisfying $a^n = 1$ for every element $a \in G$. A field containing all the nth roots of unity is a splitting field for G (R. Brauer, 1945). Consequently, for such a field K, any scalar extension of an irreducible representation over K is irreducible, and any irreducible representation over any field extension of K is realizable in K. We fix a splitting field K for G and assume that K is of characteristic 0; for example, we can assume $K = \mathbf{C}$.

The number of nonsimilar irreducible representations of G is equal to the number of conjugate classes in G. Each irreducible representation appears as an irreducible component of the regular representation with multiplicity equal to the degree. In addition, each degree is a divisor of the order g of G. Let ρ be a linear representation of a subgroup H of G and M be its representation module. Then the linear representation of G associated with the $K[G]$-module $K[G] \otimes_{K[H]} M$ is called the **induced representation** and is denoted by ρ^G. If the matrix representation T corresponds to ρ, then using the partition of G into the cosets $G = a_1 H \cup \ldots \cup a_r H$ we can write the matrix representation corresponding to ρ^G as

$$a \rightarrow \begin{bmatrix} T(a_1^{-1} a a_1) & \ldots & T(a_1^{-1} a a_r) \\ \ldots & \ldots & \ldots \\ T(a_r^{-1} a a_1) & \ldots & T(a_r^{-1} a a_r) \end{bmatrix},$$

where we define $T(b) = 0$ for $b \notin H$. The induced representation from a representation of degree 1 of a subgroup is called a **monomial representation.** To such a representation corresponds a matrix representation T such that $T(a)$ has exactly one nonzero entry in each row and column for every $a \in G$. For the trivial sub-

group $H=\{e\}$, we obtain the regular representation of G. In general, for an irreducible representation σ of G and an irreducible representation ρ of a subgroup H, the multiplicity of σ in ρ^G coincides with that of ρ in the restriction σ_H of σ to H (the **Frobenius theorem**). The following **orthogonality relations** hold for irreducible characters χ and ψ of G:

$$\sum_{a\in G}\chi(a)\psi(a^{-1})=\begin{cases} g, & \chi=\psi,\\ 0, & \chi\neq\psi,\end{cases}$$

$$\sum_{\chi}\chi(a)\chi(b^{-1})=\begin{cases} g/g_a, & C(a)=C(b),\\ 0, & C(a)\neq C(b).\end{cases}$$

In the second formula, χ ranges over all the irreducible characters of G, $C(a)$ denotes the conjugate class of G containing a, and g_a is the number of elements in $C(a)$.

H. Linear Representations of Symmetric Groups

All irreducible representations of a †symmetric group $\mathfrak{S}_n$ over the field $\mathbf{Q}$ of rational numbers are absolutely irreducible. Hence the representation theory of $\mathfrak{S}_n$ over a field of characteristic zero reduces to that over $\mathbf{Q}$. Since the group algebra $A=\mathbf{Q}[\mathfrak{S}_n]$ is semisimple, to obtain an irreducible representation of $\mathfrak{S}_n$ it is sufficient to find a †primitive idempotent (i.e., an idempotent that is not the sum of two orthogonal nonzero idempotents) of A. Such an idempotent can be obtained in the following way. As in Fig. 1, we draw a diagram T consisting of n squares arranged in rows of decreasing lengths, the left ends of which are arranged in a single column. Such a diagram T is called a **Young diagram**; if it has k rows of lengths $f_1\geqslant f_2\geqslant\ldots\geqslant f_k>0$, $f_1+f_2+\ldots+f_k=n$, then it is written $T=T(f_1,\ldots,f_k)$. We put the numerals 1 to n in any order into the n squares of $T=T(f_1,f_2,\ldots,f_k)$, as in Fig. 1, for example. We then denote by σ any permutation of $\mathfrak{S}_n$ that preserves each row and construct an element of $A: s=\sum\sigma$. Similarly, we denote by τ any permutation of $\mathfrak{S}_n$ that preserves each column and set $t=\sum(\operatorname{sgn}\tau)\tau$. If we set $u=t\cdot s=\sum\pm\tau\cdot\sigma$, then u is a primitive idempotent of A except for a numerical factor. This implies that u yields an irreducible representation of $\mathfrak{S}_n$. The element u of A is called the **Young symmetrizer** associated with T.

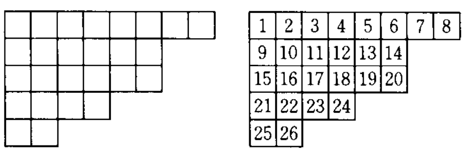


Fig. 1
Young diagram. $n=26$; $f_1=8$, $f_2=6$, $f_3=6$, $f_4=4$, $f_5=2$.

If we put the numerals 1 to n into the n squares of T in a different order, we obtain another symmetrizer u' associated with T. However, these two irreducible representations associated with u and u' are similar. Hence there corresponds to T a fixed class of irreducible representations of $\mathfrak{S}_n$, i.e., a fixed irreducible character of $\mathfrak{S}_n$. Moreover, any two different Young diagrams yield different irreducible characters, and any irreducible character is obtained by a suitable Young diagram. Thus there exists a one-to-one correspondence between the Young diagrams and the irreducible characters of $\mathfrak{S}_n$.

The method of determining the character associated with a given diagram was found by Schur and H. Weyl (→ 60 Classical Groups).

I. Modular Representations of Finite Groups

Let G be a finite group of order g, and let K be a splitting field of G of characteristic $p\neq 0$. If p is a divisor of g, we have the case of modular representation, in which the situation is quite different from the case of ordinary representation. The theory of modular representations of a finite group was developed mainly by Brauer after 1935.

The elements of G whose orders are prime to p are called **p-regular**. Let k be the number of p-regular classes of G, i.e., conjugate classes of G containing the p-regular elements. Then there exist exactly k nonsimilar absolutely irreducible modular representations $F_1, F_2, \ldots, F_k$. The number of nonsimilar indecomposable components of the regular representation R of G is also equal to k, and we denote these representations by $U_1, U_2, \ldots, U_k$. We can number them in such a way that F_κ appears in U_κ as both its top and bottom component. If the degree of F_κ is f_κ and that of U_κ is u_κ, then U_κ appears f_κ times in R and F_κ appears u_κ times in R. The multiplicities $c_{\kappa\lambda}$ of F_λ in U_κ are called the **Cartan invariants** of G.

Take an algebraic number field Ω that is a splitting field of G. Let $\mathfrak{p}$ be a prime ideal in Ω dividing p, and let $\mathfrak{o}$ be the domain of †$\mathfrak{p}$-integers of Ω. Then the residue class field $\mathfrak{o}/\mathfrak{p}$ is a finite field of characteristic p and a splitting field of G. Hence we can assume that $\mathfrak{o}/\mathfrak{p}=K$, where K is the field considered at the beginning of this section. Let $Z_1, Z_2, \ldots, Z_n$ be the nonsimilar irreducible representations of G in Ω. We can assume that all the coefficients of Z_i are contained in $\mathfrak{o}$. Replacing every coefficient in Z_i by its residue class mod $\mathfrak{p}$, we obtain a modular representation $\bar{Z}_i$. The modular representations $\bar{Z}_1, \ldots, \bar{Z}_n$ thus obtained may be reducible. The multiplicities $d_{i\kappa}$ of F_κ in $\bar{Z}_i$ are called the **decomposition numbers** of G.

They are related to the Cartan invariants by the fundamental relations

$$c_{\kappa\lambda}=\sum_{i=1}^{n} d_{i\kappa}d_{i\lambda}.$$

The determinant $|c_{\kappa\lambda}|$ of degree k is a power of p. We set $g=p^e g'$, $(p,g')=1$. Then we may assume that Ω contains a primitive g'th root of unity $\delta(\in \mathfrak{o})$. Since $(p,g')=1$, the residue class $\bar{\delta}(\in K)$ of δ is a primitive g'th root of unity. Let M be a modular representation of G. The characteristic roots of $M(a)$ for a p-regular element a are powers $\bar{\delta}^\nu$ of $\bar{\delta}$. We replace each $\bar{\delta}^\nu$ by δ^ν and obtain an element $\xi(a)$ of Ω as the sum of these δ^ν. In this manner we define a complex-valued function ξ on the set of p-regular elements of G. We call ξ the **modular character** (or **Brauer character**) of M. Two modular representations have the same irreducible components if and only if their modular characters coincide. Denoting by φ_κ the modular character of F_κ and by η_κ that of U_κ, we have the following orthogonality relations for the modular characters:

$$\sum_a \varphi_\kappa(a)\eta_\lambda(a^{-1})=\begin{cases} g, & \kappa=\lambda,\\ 0, & \kappa\neq\lambda,\end{cases}$$

$$\sum_\kappa \varphi_\kappa(a)\eta_\kappa(b^{-1})=\begin{cases} g/g_a, & C(a)=C(b),\\ 0, & C(a)\neq C(b).\end{cases}$$

In the first sum, a ranges over all p-regular elements of G.

We say that F_κ and F_λ belong to the same **block** if there exists a sequence of indices κ, α, $\beta, \ldots, \gamma$, λ such that $c_{\kappa\alpha}\neq 0$, $c_{\alpha\beta}\neq 0, \ldots, c_{\gamma\lambda}\neq 0$. This is obviously an equivalence relation, and $F_1, F_2, \ldots, F_k$ are classified into a finite number, say s, of blocks $B_1, B_2, \ldots, B_s$. If F_κ belongs to a block B_τ, we say by a stretch of language that the corresponding U_κ also belongs to B_τ. All the irreducible components of $\bar{Z}_i$ belong to the same block since $c_{\kappa\lambda}\neq 0$ if $d_{i\kappa}\neq 0$ and $d_{i\lambda}\neq 0$. If the irreducible components of $\bar{Z}_i$ belong to B_τ, we say that Z_i belongs to B_τ. Let x_τ be the number of Z_i belonging to B_τ and y_τ the number of F_κ belonging to B_τ. Then $x_\tau \geqslant y_\tau$. If χ_i is the ordinary character of Z_i, then χ_i can be considered as the modular character of $\bar{Z}_i$. If we denote the degree of Z_i by z_i, then $g_a\chi_i(a)/z_i$ for $a\in G$ is an algebraic integer and hence belongs to $\mathfrak{o}$. Now Z_i and Z_j belong to the same block if and only if $g_a\chi_i(a)/z_i \equiv g_a\chi_j(a)/z_j \pmod{\mathfrak{p}}$ for all p-regular elements a of G.

If p^α is the highest power of p that divides all the degrees z_i of Z_i belonging to B_τ, then it is also the highest power of p dividing all the degrees f_κ of F_κ belonging to B_τ. We call $d=e-\alpha$ the **defect** of B_τ; obviously $0\leqslant d\leqslant e$. If Z_i belongs to a block of defect d, then the power of p dividing z_i is p^{e-d+h_i} $(h_i\geqslant 0)$. A block of defect 0 contains exactly one ordinary representation Z_i, hence also exactly one modular representation F_κ $(x_\tau=y_\tau=1)$. Moreover, we have $Z_i=F_\kappa=U_\kappa$. It follows that all the degrees z_i of Z_i belonging to a block of defect 1 are exactly divisible by p^{e-1}; the converse is also true. Z_i belongs to a block of defect 0 if and only if $\chi_i(a)=0$ for any element a of G whose order is divisible by p.

Let D be any p-Sylow subgroup of the †centralizer $C_G(a)$ of an element a of G, and let $(D:1)=p^d$. Then d is called the **defect** of the class $C(a)$, and D is called a **defect group** of $C(a)$. The number of blocks of defect e is equal to the number of p-regular classes of defect e. Let B_τ be a block of defect d. Then there exists a p-regular class of defect d containing an element a such that $g_a\chi_i(a)/z_i\not\equiv 0 \pmod{\mathfrak{p}}$ for any Z_i in B_τ. The defect group D of $C(a)$ is called the **defect group** of B_τ, and D is uniquely determined up to conjugacy in G. The number of blocks of G with defect group D is equal to the number of blocks of the †normalizer $N_G(D)$ with defect group D.

An arbitrary element x of G can be written uniquely as a product $x=sr=rs$, where s, called the **p-factor** of x, is an element whose order is a power of p, and r is a p-regular element. We say that two elements of G belong to the same **section** if and only if their p-factors are conjugate in G. This is an equivalence relation. Obviously, each section is the union of conjugate classes of G. If the p-factor of x is not conjugate to any element of the defect group D of B_τ, then $\chi_i(x)=0$ for all Z_i in B_τ. Let $\varphi_1^s, \varphi_2^s, \ldots, \varphi_{k(s)}^s$ be the absolutely irreducible modular characters of $C_G(s)$, and let χ'_l be the absolutely irreducible ordinary characters of $C_G(s)$. Since

$$\chi'_l(sr)=\varepsilon_l\chi'_l(r)=\varepsilon_l\sum_\sigma d'_{l\sigma}\varphi_\sigma^s(r), \quad r\in C_G(s),$$

we have

$$\chi_i(sr)=\sum_l r_{il}\chi'_l(sr)=\sum_\sigma d_{i\sigma}^s\varphi_\sigma^s(r).$$

The $d_{i\sigma}^s$ are called the **generalized decomposition numbers** of G. If the order of s is p^l, then the $d_{i\sigma}^s$ are algebraic integers of the field of the p^lth roots of unity. Let s be conjugate to an element of D. There corresponds to B_τ a union $\tilde{B}_\tau$ of blocks of $C_G(s)$, and if $\sigma\neq\rho$, then $\tilde{B}_\tau$ and $\tilde{B}_\rho$ contain no irreducible modular representations in common. We have $d_{i\sigma}^s=0$ for any Z_i in B_τ (i.e., $\varphi_\sigma^s\notin\tilde{B}_\tau$). Brauer's original proof of this result was considerably complicated; simpler proofs were given independently by K. Iizuka and H. Nagao. From these relations we get the following refinement of the orthogonality relations for group characters. If Z_i and Z_j belong to different blocks of G, then $\sum_{a\in S}\chi_i(a)\chi_j(a^{-1})=0$, where a ranges over all

the elements belonging to a fixed section S of G. If elements a and b of G belong to different sections, then $\sum_{\chi_i}\chi_i(a)\chi_i(b^{-1})=0$, where χ_i ranges over all the characters of G belonging to a fixed block B_τ.

J. Projective Representations of Finite Groups

Let V be a finite-dimensional linear space over a field K, and let $P(V)$ be the †projective space associated with V (→ 343 Projective Geometry). The set of all projective transformations of $P(V)$ forms the group $PGL(V)$, which can be identified with the quotient group $GL(V)/K^*1_V$. Here $K^*=K-\{0\}$ and K^*1_V is the set of all scalar multiples of the identity transformation 1_V of V and is the center of $GL(V)$. A homomorphism $G\rightarrow PGL(V)$ is called a **projective representation** of G in V or simply a projective representation of G over K. Two projective representations (ρ, V) and (ρ', V') of G are said to be **similar** if there exists an isomorphism $\varphi: PGL(V)\rightarrow PGL(V')$ induced by a suitable isomorphism $V\rightarrow V'$ such that $\varphi\circ\rho(a)\circ\varphi^{-1}=\rho'(a)$ $(a\in G)$. Let $V_1\neq\{0\}$ be a subspace of V. We can assume that $P(V_1)\subset P(V)$. If (ρ, V) is a projective representation of G such that each $\rho(a)$ $(a\in G)$ leaves $P(V_1)$ invariant, we get a projective representation (ρ_1, V_1) by restricting the $\rho(a)$ to $P(V_1)$. In this case (ρ_1, V_1) is called a **subrepresentation** of ρ. A projective representation is said to be **irreducible** if there exists no proper subrepresentation of ρ.

A mapping $\sigma: G\rightarrow GL(V)$ is called a section for (ρ, V) if $\pi(\sigma(a))=\rho(a)$ for each $a\in G$, where π is the natural projection of $GL(V)$ onto $PGL(V)$. Any section σ defines a mapping $f: G\times G\rightarrow K^*$ satisfying $\sigma(a)\sigma(b)=f(a,b)\sigma(ab)$ $(a,b\in G)$. The set $\{f(a,b)\}_{a,b\in G}$ is called the **factor set** of ρ with respect to σ. The mapping f is a †2-cocycle of G with values in K^*. The 2-cohomology class $c_\rho\in H^2(G,K^*)$ of f is determined by ρ and is independent of the choice of sections for ρ. A projective representation ρ has a section σ which is a linear representation of G in V if and only if $c_\rho=1$. If G is a finite group, for any $c\in H^2(G,K^*)$ there exists an irreducible projective representation ρ of G over K which belongs to c, i.e., $c_\rho=c$. If ρ and ρ' are similar, then $c_\rho=c_{\rho'}$. The tensor product $\rho\otimes\rho'$ of two projective representations ρ and ρ' can be defined as in the case of linear representations, and we have $c_{\rho\otimes\rho'}=c_\rho\cdot c_{\rho'}$. If K is algebraically closed, then $H^2(G,K^*)$ is determined by the characteristic of K. When K is the complex field $\mathbf{C}$, the group $H^2(G,\mathbf{C}^*)=\mathfrak{M}(G)$ is called the **multiplier** of G. If $\mathfrak{M}(G)=1$, then G is called a **closed group**, and any projective representation of G is induced by a linear representation of G. In general, if ρ is a projective representation of G over $\mathbf{C}$, then the order of c_ρ is a divisor of the degree of ρ (dimension of V). Moreover, if ρ is irreducible, then both the degree of ρ and the square of the order of c_ρ are divisors of the order of G. K. Yamazaki, among others, studied the projective representations of finite groups in detail.

K. Integral Representations

Every complex matrix representation of G is equivalent to a matrix representation in the ring of algebraic integers. If an algebraic number field K is specified, every $K[G]$-module V contains G-invariant R-†lattices (briefly, G-lattices), where R is the ring of integers in K. A G-lattice L is characterized as an $R[G]$-module, which is finitely generated and †torsion free (hence †projective) as an R-module. It provides an **integral representation** of G as an automorphism group of the R-projective module L.

$R[G]$-modules L and M need not be isomorphic even when the $K[G]$-modules $K\otimes L$ and $K\otimes M$ are isomorphic. The set of G-lattices in a fixed $K[G]$-module V is divided into a finite number of $R[G]$-isomorphism classes (**Jordan-Zassenhaus theorem**). Let $\mathfrak{p}$ be a prime ideal of R and $R_\mathfrak{p}$ be the localization of R at $\mathfrak{p}$. The study of $R_\mathfrak{p}$-representations is intimately related with modular representation theory. For any $R[G]$-module L there is an associated family of $R_\mathfrak{p}[G]$-modules $L_\mathfrak{p}=R_\mathfrak{p}\otimes L$, where $\mathfrak{p}$ ranges over all primes of R. G-lattices L and M in a $K[G]$-module V are said to be of the same **genus** if $L_\mathfrak{p}\cong M_\mathfrak{p}$ for every $\mathfrak{p}$. The number of genera of G-lattices in V is given by $\prod_{\mathfrak{p}|g}h_\mathfrak{p}$ ($g=$ order G), where $h_\mathfrak{p}$ denotes the number of $R_\mathfrak{p}[G]$-equivalence classes of $R_\mathfrak{p}[G]$-lattices in V. When V is absolutely irreducible, the number of $R[G]$-equivalence classes in a genus equals the (ideal) †class number of K (J. M. Maranda and S. Takahashi).

The †Krull-Schmidt theorem, asserting the uniqueness of a direct sum decomposition into indecomposable $R[G]$-modules, holds if R is a complete discrete valuation ring or if R is a discrete valuation ring and K is a splitting field of G. The condition for the finiteness of the number of nonisomorphic indecomposable G-lattices is known. In particular, for $R=\mathbf{Z}$ it reduces to the requirement that the Sylow p-subgroup of G be cyclic of order p or p^2 for every $p|g$. Regarding projective $\mathbf{Z}[G]$-modules → 200 Homological Algebra G.

The **isomorphism problem**, i.e., the question of whether the isomorphism $\mathbf{Z}[G]\cong\mathbf{Z}[H]$ of integral group algebras implies the isomor-

phism $G \cong H$ of groups, has been answered affirmatively for certain special cases such as †meta-Abelian groups.

$R[G]$ is an R-†order in $K[G]$, and in this context, a considerable portion of the integral representation theory has been extended to more general orders in separable algebras [14–16].

References

[1] B. L. van der Waerden, Gruppen von linearen Transformationen, Springer, 1935 (Chelsea, 1948).
[2] B. L. van der Waerden, Algebra II, Springer, fifth edition, 1967.
[3] H. Boerner, Darstellungen von Gruppen, Springer, 1955.
[4] N. Bourbaki, Eléments de mathématique, Algèbre, ch. 8, Actualités Sci. Ind., 1261a, Hermann, 1958.
[5] C. W. Curtis and I. Reiner, Representation theory of finite groups and associative algebras, Interscience, 1962.
[6] R. Brauer and C. Nesbitt, On the modular characters of groups, Ann. Math., (2) 42 (1941), 556–590.
[7] R. Brauer, Zur Darstellungstheorie der Gruppen endlicher Ordnung I, II, Math Z., 63 (1956), 406–444; 72 (1959–1960), 25–46.
[8] M. Osima, Notes of blocks of group characters, Math. J. Okayama Univ., 4 (1955), 175–188.
[9] I. Schur, Über die Darstellung der endlichen Gruppen durch gebrochene lineare Substitutionen, J. Reine Angew. Math., 127 (1904), 20–50.
[10] I. Schur, Untersuchungen über die Darstellung der endlichen Gruppen durch gebrochene lineare Substitutionen, J. Reine Angew. Math., 132 (1907), 85–137.
[11] I. Schur, Über die Darstellung der symmetrischen und der alternierenden Gruppe durch gebrochene lineare Substitutionen, J. Reine Angew. Math., 139 (1911), 155–250.
[12] R. Brauer, Representations of finite groups, Lectures on modern mathematics, T. L. Saaty (ed.), Wiley, 1963, vol. 1, 133–175.
[13] W. Feit, Characters of finite groups, Benjamin, 1967.
[14] I. Reiner, A survey of integral representation theory, Bull. Amer. Math. Soc., 76 (1970), 159–227.
[15] R. Swan, K-theory of finite groups and orders, Lecture notes in math. 149, Springer, 1970.
[16] K. W. Roggenkamp (with V. Huber-Dyson), Lattices over orders I, II, Lecture notes in math. 115, 142, Springer, 1970.
[17] L. Dornhoff, Group representation theory, Dekker, A, 1971; B, 1972.
[18] J.-P. Serre, Représentations linéaires des groupes finis, second edition, Hermann, 1972.
[19] T. Yamada, The Schur subgroup of the Brauer group, Lecture notes in math. 397, Springer, 1974.

363 (XXI.40) Riemann, Georg Friedrich Bernhard

Georg Friedrich Bernhard Riemann (September 17, 1826–July 20, 1866) was born the son of a minister in Breselenz, Hanover, Germany. He attended the universities of Göttingen and Berlin. In 1851 he received his doctorate at the University of Göttingen and in 1854 became a lecturer there. In 1857 he rose to assistant professor, and in 1859 succeeded P. G. L. †Dirichlet as full professor. In 1862 he contracted tuberculosis, and he died at age 40. Despite his short life, his contributions encompassed all aspects of mathematics.

His doctoral thesis (1851) stated the basic theorem on †conformal mapping and became the foundation for the geometric theory of functions. In his paper presented for the position of lecturer (1854), he defined the †Riemann integral and gave the conditions for convergence of trigonometric series. In his inaugural lecture in the same year, he discussed the foundations of geometry, introduced n-dimensional manifolds, formulated the concept of †Riemannian manifolds, and defined their curvature. In his paper of 1857 on †Abelian functions, he systematized the theory of †Abelian integrals and Abelian functions. In his paper of 1858 on the distribution of prime numbers, he considered the †Riemann zeta function as a function of a complex variable and stated †Riemann's hypothesis concerning the distribution of its zeros. It remains for modern mathematics to investigate whether this hypothesis is correct. In his later years, influenced by W. Weber, Riemann became interested in theoretical physics. He gave lectures on the uses of partial differential equations in physics that were edited and published by H. Weber.

References

[1] G. F. B. Riemann, Gesammelte mathematische Werke und wissenschaftlicher Nachlass, R. Dedekind and H. Weber, (eds.), Teubner, second edition, 1892 (Nachträge, M. Noether and W. Wirtinger (eds.), 1902) (Dover, 1953).

[2] F. Klein, Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert I, Springer, 1926 (Chelsea, 1956).

364 (VII.3) Riemannian Manifolds

A. Riemannian Metrics

Let M be a †differentiable manifold of class C^r $(1 \leqslant r \leqslant \omega)$, and g be a †Riemannian metric of class C^{r-1} on M. Then (M, g) or simply M is called a **Riemannian manifold** (or **Riemannian space**) of class C^r (→ 105 Differentiable Manifolds). The metric g is a †covariant tensor field of order 2 and of class C^{r-1}; it is called the **fundamental tensor** of M. Using the value g_p of g at each point $p \in M$, a positive definite inner product $g_p(X, Y)$, X, $Y \in T_p$, is introduced on the †tangent vector space T_p to M at p, and hence T_p can be considered as a †vector space over $\mathbf{R}$ with inner product that can be identified with the Euclidean space E^n of dimension $n = \dim M$. Utilizing the properties of the space E^n, we can introduce various notions on T_p and M. (For example, given a tangent vector $L \in T_p$, we define the length $\|L\| = \|L\|_{g,p}$ of L to be the quantity $g_p(L, L)^{1/2}$. A **normal vector** at a point p of a submanifold N of M is well defined as an element of the orthogonal complement of the subspace $T_p(N)$ of $T_p(M)$ with respect to g_p; a differential form of degree 1 is identified with a tangent vector field.) A necessary and sufficient condition for a differentiable manifold M of class C^r to have a Riemannian metric is that M be †paracompact. A Euclidean space E^n has a Riemannian metric expressed by $\sum_{i=0}^n dx^i \otimes dx^i$ in terms of an orthogonal coordinate system (x^i).

We assume that M is connected and of class C^∞. A curve $x: [a, b] \to M$ is called **piecewise smooth** or of **class** D^∞ if x is continuous and there exists a partition of $[a, b]$ into finite subintervals $[t_{i-1}, t_i]$ such that the restrictions $x | [t_{i-1}, t_i]$ are †immersions of class C^∞. The length $\|x\|$ of such a curve x is defined to be $\int_a^b \|x'(t)\| \, dt$, where $x'(t)$ is the tangent vector of x defined for almost all values of t. As in a Euclidean space, the length $\|x\|$ is independent of the choice of parameter t, and the concepts of †canonical parameter and orientation of x can be defined (→ 111 Differential Geometry of Curves and Surfaces). A function $d: M \times M \to [0, \infty)$ is defined so that the value $d(p, q)$, p, $q \in M$, is the infimum of the lengths of curves of class D^∞ joining p and q. The function d is a †distance function on M, and the topology of M defined by d coincides with the original topology of M. There exists an essentially unique structure of a Riemannian manifold on (real or complex) †elliptic or †hyperbolic space (→ 285 Non-Euclidean Geometry), and d is the distance function of these spaces.

If there exists an immersion φ of a differentiable manifold N in a Riemannian manifold (M, g), then a Riemannian metric $\varphi^* g$ is defined on N by the †pullback process ($\|L\|_{\varphi^* g} = \|d\varphi(L)\|_g$). (For example, a submanifold and a †covering manifold of M have Riemannian manifold structures induced by the natural mappings (→ 365 Riemannian Submanifolds).) If $M = E^3$ and N is a 2-dimensional submanifold of M, then $\varphi^* g$ is the †first fundamental form of N. Assume further that φ is a diffeomorphism and N has a Riemannian metric h. If $\varphi^* g = h$, then (N, h) is said to be **isometric** to (M, g), and φ is called an **isometry**. The set $I(M)$ of all isometries (isometric transformations) of M onto M is a group. A necessary and sufficient condition for a mapping ψ: $N \to M$ to be an isometry is that $d_N(p, q) = d_M(\psi(p), \psi(q))$, p, $q \in N$. In particular, $I(E^n)$ is the †congruent transformation group.

If a differentiable manifold M is the product manifold of Riemannian manifolds (M_1, g_1) and (M_2, g_2), then $(M, \pi_1^* g_1 + \pi_2^* g_2)$ is called the **Riemannian product** of M_1 and M_2, where π_α, $\alpha = 1, 2$, are projections from M to M_α.

Let F be the †tangent n-frame bundle over M and $B = B_g(M)$ be the subset of F consisting of all orthonormal frames with respect to g. Then B is an $O(n)$-subbundle of F of class C^∞, called the **tangent orthogonal n-frame bundle** (or **orthogonal frame bundle**). In this way we get a one-to-one correspondence between the set of all $O(n)$-subbundles of F and the set of all Riemannian metrics of M.

B. Riemannian Connections

There exists a unique †affine connection in the orthogonal frame bundle B whose †torsion tensor is zero. This connection is called the **Riemannian connection** (or **Levi-Civita connection**; → 80 Connections K). Let ∇ denote the †covariant differential operator defined by this connection (→ 80 Connections, 417 Tensor Calculus). (For a vector field X, the covariant differential operator ∇_X acts on any tensor field T defined on a submanifold having X as a tangent vector field.) The covariant differential ∇g of the fundamental tensor g vanishes identically. The †connection form of the Riemannian connection is expressed by n^2 differential 1-forms $(\omega_j^i)_{1 \leqslant i,j \leqslant n}$ on B, and we have $\omega_j^i + \omega_i^j = 0$. Let $(\omega^i)_{1 \leqslant i \leqslant n}$ be the †canonical 1-forms on B. Then $(\omega_i^j)_{1 \leqslant i < j \leqslant n}$ together with (ω^i) give rise to an absolute parallelism on B (that is, they are linearly independent at

every point). Let (θ^i) and (θ_i^j) be the corresponding set of differential 1-forms on the orthogonal frame bundle B_N over another Riemannian manifold N with $\dim N = \dim M$. If there exists an isometry $\psi: M \to N$, then the differential $d\psi$ is a diffeomorphism from $B = B_M$ to B_N, and we get $(d\psi)^*(\theta^i) = \omega^i$, $(d\psi)^*(\theta_i^j) = \omega_i^j$. Conversely, if there exists a diffeomorphism $\Psi: B_M \to B_N$ satisfying $\Psi^*(\theta^i) = \omega^i$, $\Psi^*(\theta_i^j) = \omega_i^j$ and M is †orientable, then there exists an isometry $\psi: M \to N$ such that $d\psi = \Psi$ holds on a connected component B_0 of B. Moreover, ψ is uniquely determined if we choose one B_0. In this way the problem of the existence of an isometry from M may be reduced to one of the existence of a diffeomorphism from B preserving absolute parallelism (as well as the order of the basis (ω^i, ω_i^j)).

According to the general theory of affine connections, the Riemannian connection on M determines a †Cartan connection uniquely with $E^n = I(E^n)/O(n)$ as †fiber, which is called the **Euclidean connection**. As a consequence, every tangent vector space $T_p(M)$ is regarded as a Euclidean space E_p^n, and for a given curve $x: [a, b] \to M$ of class D^∞ and for $t \in [a, b]$ there exists an isometry $I_{x,t}: E_{x(t)}^n \to E_{x(a)}^n$ satisfying the following three conditions (we denote $I_{x,b}$ by I_x): (1) If x is a composite of two curves y and z, then $I_x = I_y \cdot I_z$. (2) Differentiability: If x is of class C^∞ at t_0, then $t \to I_{x,t}$ is of class C^∞ at t_0. (3) I_x depends on the orientation of x but not on the choice of its parameter. The **development** $\bar{x}$ of x is the curve in $E_{x(a)}^n$ defined by $\bar{x}(t) = I_{x,t}(x(t))$, and we get $\|\bar{x}\| = \|x\|$. (I_x is sometimes called the **development along** x.) Utilizing the concept of development, the theory of curves in E^n can be used to study curves on M (→ 111 Differential Geometry of Curves and Surfaces). For example, if $\bar{x}$ is a segment, then x or $x([a, b])$ is called the **geodesic arc** (→ 80 Connections L); the †Frenet formula is automatically formulated and proved. The rotation part I_x^R of I_x (the composite of I_x and the parallel displacement of E_p^n translating $I_x(x(b))$ to $x(a)$) is regarded as an isomorphism of the inner product space $T_{x(b)}$ to $T_{x(a)}$. I_x^R is extended to an isomorphism of the †tensor algebra $\mathscr{T}(T_{x(b)})$ to $\mathscr{T}(T_{x(a)})$, which is denoted by the same symbol I_x^R and called the **parallel displacement** or **parallel translation along** x. Given a tensor field K on M, we have $\nabla_{x'(a)} K = [dI_{x,t}^R(K(x(t)))/dt]_{t=a}$. In particular, a necessary and sufficient condition for $\nabla K = 0$ is that $I_x^R(K(x(b))) = K(x(a))$ for any x, in which case K is said to be **parallel**.

C. Exponential Mapping (→ 178 Geodesics)

A curve x on M or the image of x is called a **geodesic** if any subarc $x | [a, b]$ of x is a geodesic arc. Let $N(S)$ be the **normal bundle** of a submanifold S of M, that is, the differentiable vector bundle over S consisting of all normal vectors at all points of S. Then S is contained in $N(S)$ as the set of zero vectors at all points of S. There exist a neighborhood U of S in $N(S)$ and a mapping $\mathrm{Exp}_S: U \to M$ of class C^∞ with the following property: There exists a geodesic arc x with the initial tangent vector $L \in U$, length $\|x\| = \|L\|$, and final point $\mathrm{Exp}_S(L)$. Let U_S be the largest U with this property. Then $\mathrm{Exp}_S: U_S \to M$ is determined uniquely by S. The mapping Exp_S is called the **exponential mapping** on S. If the rank of the Jacobian matrix of Exp_S is less than n at $L \in U_S$, then L or $\mathrm{Exp}_S(L)$ is called the **focal point** of S on the geodesic $s \to \mathrm{Exp}_S(sL)$ ($0 \leqslant s$, $sL \in U_S$). If S is compact, then S has an open neighborhood V_S in $N(S)$ satisfying the following three conditions: (i) $V_S \subset U_S$; (ii) $\|L\| = d(\mathrm{Exp}_S(L), S)$ for $L \in V_S$, where the right-hand member expresses the infimum of the distance between the point $\mathrm{Exp}_S(L)$ and points of S; (iii) the restriction $\mathrm{Exp}_S | V_S$ is an embedding. The image $\mathrm{Exp}_S(V_S)$ is the **tubular neighborhood** of S. In the special case where S consists of only one point p, $N(\{p\})$ coincides with the tangent vector space $T_p(M)$, and the focal point of p is called the **conjugate point** of p, given as the zero point of the †Jacobi field (→ 178 Geodesics, 279 Morse Theory). In this case, V_S is denoted by V_p. If T_p is identified with $\mathbf{R}^n$ (or E^n) by means of an orthonormal basis of T_p, then $(\mathrm{Exp}_p)^{-1}$ defined on $\mathrm{Exp}_p(V_p)$ is a coordinate mapping, called the **normal coordinate mapping**. Furthermore, $\mathrm{Exp}_p(V_p)$ contains a neighborhood W_p of p such that there exists a unique geodesic arc x joining any two points q and r of W_p with $\|x\| = d(q, r)$ and contained in W_p. W_p is called a **convex neighborhood** of p.

D. Curvature

The set of differential 1-forms (ω^i, ω_i^j), by means of which absolute parallelism is given in the orthogonal frame bundle B of M, satisfies the †structure equation $d\omega^i = -\sum_j \omega_j^i \wedge \omega^j$, $d\omega_j^i = -\sum_k \omega_k^i \wedge \omega_j^k + \Omega_j^i$, and (Ω_j^i) is called the **curvature form** of the Riemannian connection of M. This form is expressed by a tensor field R (→ 80 Connections; 417 Tensor Calculus) of type $(1, 3)$ on M, called the **curvature tensor**; if R_{jkl}^i are the components of R with respect to an orthonormal frame $b \in B$ of the tangent vector space T_p of M, then $\Omega_j^i = (1/2) \sum R_{jkl}^i \omega^k \wedge \omega^l$ at b. Let (X, Y) be an orthonormal basis of a 2-dimensional subspace P of T_p. Then the inner product $K_p(P)$ of X and $R(X, Y)Y$ is determined by P independently of the choice of the basis (X, Y), where the i-component of

$R(X, Y)Z$ with respect to the basis b of T_p is given by $\sum R^i_{jkl} Z^j X^k Y^l$. $K_p(P)$ is the †Gaussian curvature of the surface $\mathrm{Exp}_p(V_p \cap P)$ and is called the **sectional curvature** (or **Riemannian curvature**) of P. The curvature tensor R is uniquely determined by the function $K_p(P)$ of p and P. If $\dim M \geqslant 3$ and if at every point p of M, $K_p(P)$ has a constant value M_p independent of the choice of P, then M_p is a constant independent of the choice of p (F. Schur). If $K_p(P)$ is constant, then M is called a **space of constant curvature**. If $\nabla R = 0$, then M is called a **locally symmetric space** (→ 412 Symmetric Riemannian Spaces and Real Forms; 413 Symmetric Spaces). In a local sense, Riemannian metrics of these spaces are uniquely determined by the curvature tensor R up to a constant factor. If M is of constant curvature K, complete, and simply connected, then M is isometric to E^n, the sphere (which is the universal covering Riemannian manifold of a real †elliptic space), or a real †hyperbolic space according as K is 0, positive, or negative. The compact spaces of positive constant curvature, that is, the Riemannian manifolds having the sphere as the universal covering Riemannian manifold, were completely classfied by J. A. Wolf [1]. A complete, simply connected, and locally symmetric space is a †symmetric Riemannian space. The **Ricci tensor** (R_{ij}) is defined by $R_{ij} = -\sum_k R^k_{ijk}$. Let Q be the quadratic form on T_p given by (R_{ij}). Then the value $Q(L)$ for a unit vector $L \in T_p$ is the mean of $K_p(P)$ for all sections P (2-dimensional subspaces of T_p) containing L and is called the **Ricci curvature** (or **mean curvature**) of the direction L at p. The mean R of $Q(L)$ for all the unit vectors L at p is called the **scalar curvature** at p (→ 417 Tensor Calculus). $Q(L)$ and R are expressed by $Q(L) = \sum_i g_p(R(X_i, L)L, X_i)$ and $R = \sum_i Q(X_i)$, up to positive constant factors, in terms of an orthonormal basis (X_i) of T_p. If the Ricci tensor of M is a scalar multiple of the fundamental tensor, then M is called an **Einstein space**. (When $\dim M \geqslant 3$, this scalar is constant.) If M is a †Kähler manifold and P is restricted to a complex plane (invariant under the almost complex structure), then $K(P)$ is called the **holomorphic sectional curvature**. A Kähler manifold M of constant holomorphic sectional curvature is locally isometric to a complex Euclidean space, elliptic space, or hyperbolic space.

The properties of the sectional curvature and the Ricci curvature are closely related to the behavior of geodesics of Riemannian manifolds, and these properties reflect those of the topological structures of the manifolds (→ 178 Geodesics). The compact simply connected homogeneous Riemannian manifolds of strictly positive sectional curvature have been classified [2–4]. Related to algebraic geometry, as the solution of the Frankel conjecture, the following holds: If a compact Kähler manifold has strictly positive sectional curvature, then it is biholomorphic to the complex projective space [5, 6] (→ 232 Kähler Manifolds). Furthermore, curvature tensors are related to †characteristic classes. For example, we have the **Gauss-Bonnet formula**: If M is an even-dimensional compact and oriented Riemannian manifold, the integral of $a_n K_{(n)} \omega$ on M is equal to the †Euler-Poincaré characteristic, where

$$a_n = n!/(2^n \pi^{n/2} (n/2)!),$$

ω is the volume element of M, and $K_{(n)}$ is defined as follows: For a positive even number s, $K_{(s)}$ is a real-valued function of the s-dimensional subspaces P of the tangent vector spaces T_p of M, which is given by

$$K_{(s)}(P) = b_s \sum \varepsilon_{i_1 \ldots i_s} \varepsilon_{j_1 \ldots j_s} \langle R_p(X_{i_1}, X_{i_2}) X_{j_1}, X_{j_2} \rangle \ldots \langle R_p(X_{i_{s-1}}, X_{i_s}) X_{j_{s-1}}, X_{j_s} \rangle$$

in terms of an orthonormal basis $(X_1, \ldots, X_s)$ of P, where $b_s = (-1)^{s/2}/(2^{s/2} s!)$, $\sum$ is summation over all pairs of s-tuples satisfying $\{i_1, \ldots, i_s\}$, $\{j_1, \ldots, j_s\} \subset \{1, 2, \ldots, n\}$, $\varepsilon_{i_1 \ldots i_s}$ is the sign of $(i_1, \ldots, i_s)$, $\langle\ ,\ \rangle$ is the inner product in T_p with respect to g_p, R_p is the value of the tensor R at p, and $R_p(X_i, X_j)X_k$ is as already defined at the beginning of this section. In particular, $K_{(2)} = K$. If $K_{(s)}$ of a compact and orientable M is constant for a certain s, then the kth †Pontryagin class of M (with real coefficients) vanishes for all $k \geqslant s/2$.

E. Holonomy Groups

Let p be a fixed point of M, and let Ω_p be the set of all closed oriented curves of class D^∞ with initial and final points p and with parameters neglected. The set $H = \{I_x \mid x \in \Omega_p\}$, called the **holonomy group** of M, is a subgroup of $I(T_p)$ (T_p is identified with E^n) independent of the choice of p (→ 80 Connections), and $x \to I_x$ is a homomorphism from Ω_p to H. The restriction H_0 of this homomorphism to all closed curves homotopic to zero is called the **restricted holonomy group**. The rotation part h of H, called the **homogeneous holonomy group**, is a subgroup of the orthogonal group $O(n)$ of T_p. The rotation part h_0 of H_0, called the **restricted homogeneous holonomy group**, is a connected component of h and a †compact Lie group. The †Lie algebra of h_0 is spanned by $\{I_x(R_{x(b)}(X, Y)) \mid x : [a, b] \to M$ is of class D^∞, $x(a) = p$, and $X, Y \in T_{x(b)}\}$, where $R_{x(b)}(X, Y)$ is the endomorphism of the linear space $T_{x(b)}$ defined in Section D.

If $M=E^n$, then $H=\{e\}$, where e is the identity element. If $h=\{e\}$ ($h_0=\{e\}$), then M is called **flat (locally flat)** (→ 80 Connections E). Local flatness is equivalent to M being locally isometric to E^n. If M is complete and H (regarded as a transformation group of E^n) has a fixed point, then M is isometric to E^n. Any finite rotation group h is the homogeneous holonomy group of some locally flat and compact Riemannian manifold.

With respect to the linear group h of T_p, we get a unique decomposition $T_p=V_{(0)}\oplus V_{(1)}\oplus\ldots\oplus V_{(r)}$ of mutually orthogonal subspaces, where $V_{(0)}(\dim V_{(0)}\geqslant 0)$ consists of all h-invariant vectors and $V_{(i)}$, $i=1,\ldots,r$, are irreducible h-invariant subspaces. If h or h_0 is irreducible (reducible) on T_p, then M is called **irreducible (reducible)**. If M is complete and simply connected (hence $h=h_0$), then M is the Riemannian product of closed submanifolds $M_{(\alpha)}$, $\alpha=0,1,\ldots,r$, satisfying $V_{(\alpha)}=T(M_{(\alpha)})$. This decomposition $M=\prod M_{(\alpha)}$ is determined uniquely by M and called the **de Rham decomposition** of M [7]. In this case h is the direct product of closed subgroups $h_{(\alpha)}$, where every $h_{(\alpha)}$ acts on $V_{(\beta)}$, $\beta\neq\alpha$ as the identity, and can be regarded as the homogeneous holonomy group of $M_{(\alpha)}$. If h_0 is irreducible and M is not locally symmetric, then h_0 acts †transitively on the unit sphere of T_p. The classification of possible candidates for such h_0 has been made [8,9]. For example, if n is even and h_0 is the †unitary group $U(n/2)$, then h acts transitively on the unit sphere. A necessary and sufficient condition for h to be contained in $U(n/2)$ is that M have a †complex structure and the structure of a Kähler manifold.

The group h acts naturally on the †tensor algebra $\mathscr{T}(T_p)$ of T_p. If a tensor field A on M is parallel, then A_p is invariant under h. Conversely, if $A_0\in\mathscr{T}(T_p)$ is invariant under h, there exists a unique parallel tensor field A satisfying $A_p=A_0$. The orthogonal frame bundle B is †reducible to the h-bundle.

F. Transformation Groups

The group $I(M)$ consisting of all isometries of M with the †compact-open topology is a †Lie transformation group. The isotropy subgroup at any point is compact. In particular, if M is compact, so is $I(M)$. The differential $d\varphi$ of $\varphi\in I(M)$ is a transformation of the orthogonal frame bundle B. If b_0 is a fixed point of B, then the mapping β defined by $\varphi\rightarrow d\varphi(b_0)$ embeds $I(M)$ as a closed submanifold of B, and the differentiable structure of $I(M)$ is thus determined. If β is surjective, it follows from the structure equation that M is of constant curvature and equals E^n, a real †hyperbolic space, or a real †elliptic space (or a sphere). A necessary and sufficient condition for the image of β to be a subbundle of B is that $I(M)$ be transitive. If the image of β contains the h-bundle, then M is a symmetric space. If M is compact and $I(M)$ is transitive, then the image of β is contained in the h-bundle (→ 191 G-Structures). If $I(M)$ is transitive on M, then M is complete and is the †homogeneous space of $I(M)$. Conversely, a homogeneous space $M=G/K$ of a Lie group G by a compact subgroup K has a Riemannian metric invariant under G. In general, an element of $I(M)$ preserves quantities uniquely determined by the Riemannian metric g, such as the Riemannian connection, its curvature, the set of all geodesics, etc. Furthermore, any element of $I(M)$ commutes with ∇ and the †Laplace-Beltrami operator. If M is compact and oriented, then the connected component $I_0(M)$ of $I(M)$ preserves any †harmonic differential form. If M is complete and simply connected, then $I_0(M)$ is clearly decomposed into a direct product by the de Rham decomposition of M. An element of the Lie algebra of $I(M)$ is regarded as a vector field X on M, called the **infinitesimal motion**, which satisfies the equation $L_X g=0$; that is, $\nabla_j\xi_i+\nabla_i\xi_j=0$, where L_X denotes †Lie derivation and the ξ_i are †covariant components of X with respect to a natural frame $(\partial/\partial X_i)$, $i=1,\ldots,n$ (→ 417 Tensor Calculus). This equation is called **Killing's differential equation**, and a solution X of this equation is called a **Killing vector field**. The set of all Killing vector fields is a Lie algebra of finite dimension ($\leqslant\dim B$). If M is complete, then this Lie algebra coincides with that of $I(M)$. If M is compact and the Ricci tensor is negative definite, then $I(M)$ is discrete. If, furthermore, the sectional curvature is nonpositive, then an isometry of M homotopic to the identity transformation is the identity transformation itself.

It is known that $\dim I(M)\leqslant n(n+1)/2$ if $\dim M=n$, and the maximum dimension is attained only when M is a space of constant curvature. For Riemannian manifolds with large $I(M)$, extensive work on the structures of M and $I(M)$ has been done by I. P. Egorov, S. Ishihara, N. H. Kuiper, L. N. Mann, Y. Muto, T. Nagano, M. Obata, H. Wakakuwa, K. Yano, and others [10, 11].

The fixed point set of a family of isometries has interesting differential geometric properties [10]. For example, let G be any subset of $I(M)$ and F the set of points of M which are left fixed by all the elements of G. Then each connected component of F is a closed †totally geodesic submanifold of M. If M is compact and f is an isometry of M, then $\Lambda_f=\chi(F)$, where Λ_f denotes the †Lefschetz number and

$\chi(F)$ the †Euler characteristic of the fixed point set F of f. As for the existence of fixed points of an isometry, the following are known: Let f be an isometry of a compact, orientable Riemannian manifold M with positive sectional curvature. If $\dim M$ is even and f is orientation preserving, or if $\dim M$ is odd and f is orientation reversing, then f has a fixed point. In the case of nonpositive curvature, the following is basic: Every compact group of isometries of a complete, simply connected Riemannian manifold with nonpositive sectional curvature has a fixed point (E. Cartan). If a compact, orientable Riemannian manifold admits a fixed-point-free 1-parameter group of isometries, then its †Pontryagin numbers vanish.

On a Riemannian manifold M, a transformation of M which preserves the Riemannian connection, or equivalently which commutes with covariant differentiation ∇ is called an **affine transformation**. Let $A(M)$ denote the group of all affine transformations of M. A transformation preserving the set of all geodesics is called a **projective transformation**. Let $P(M)$ denote the group of all projective transformations of M. A transformation preserving the angle between tangent vectors is called a **conformal transformation**. Let $C(M)$ denote the group of all conformal transformations. They are Lie transformation groups with respect to suitable topologies. Clearly, $I(M)\subset A(M)\subset P(M)$, $I(M)\subset C(M)$ (→ 191 G-Structures).

$A_0(M)$, the connected component of $A(M)$, is decomposed into a direct product according to the de Rham decomposition of M when M is complete and simply connected (J. Hano). If M is complete and irreducible, then $A(M)=I(M)$ except when M is a 1-dimensional Euclidean space. If M is complete and its restricted homogeneous holonomy group h_0 leaves no nonzero vectors, then $A_0(M)=I_0(M)$. If M is compact, then $A_0(M)=I_0(M)$ always.

If M is complete and has a parallel Ricci tensor, then the connected component $P_0(M)=A_0(M)$, unless M is a space of positive constant sectional curvature ($n>2$) (Nagano, N. Tanaka, Y. Tashiro). If M is compact, simply connected, and has constant scalar curvature, then $P_0(M)=I_0(M)$, unless M is a sphere ($n>2$) (K. Yamauchi).

Similarly to the case of $P(M)$, it is known that if M is complete and has a parallel Ricci tensor, then the connected component $C_0(M)=I_0(M)$, unless M is a sphere ($n>2$) (Nagano). A conformal transformation remains conformal if the Riemannian metric g is changed conformally, namely, to $e^{2f}g$, f being any smooth function on M. A subset of $C(M,g)$ is called **essential** if it cannot be reduced to a subset of $I(M,\bar{g})$ for any metric $\bar{g}$ conformal to g. When M is compact, $C(M)$ or $C_0(M)$ is essential if and only if it is not compact. If $C_0(M)$ is essential, then M is conformally diffeomorphic to a sphere or a Euclidean space ($n>2$) [12–15]. When M is compact and has constant scalar curvature and $C_0(M)\neq I_0(M)$, sufficient conditions for M to be isometric to a sphere have been obtained by S. I. Goldberg and S. Kobayashi, C. C. Hsiung, S. Ishihara, A. Lichnerowicz, Obata, S. Tanno, Tashiro, K. Yano, and others. For example, if $C_0(M)$ is essential, then M is a sphere [14]. In general, however, there are compact Riemannian manifolds with constant scalar curvature for which $C_0(M)\neq I_0(M)$ (N. Ejiri).

G. Spheres as Riemannian Manifolds

A Euclidean n-sphere S^n ($n\geqslant 2$) has the properties of a Riemannian manifold. It is a space of positive constant sectional curvature $1/r^2$ ($r=$ radius) with respect to the natural Riemannian metric as a hypersurface of the Euclidean $(n+1)$-space $\mathbf{E}^{n+1}$. A sphere is characterized by the existence of solutions of certain differential equations on a Riemannian manifold. On a unit sphere S^n in $\mathbf{E}^{n+1}$, the eigenvalues of the †Laplace-Beltrami operator Δ on smooth functions are given by $0<\lambda_1<\ldots<\lambda_k<\ldots$, $\lambda_k=k(n+k-1)$. It is known that eigenfunctions f corresponding to λ_k, $\Delta f=\lambda_k f$, are the restrictions to S^n of harmonic homogeneous polynomial functions F of degree k on $\mathbf{E}^{n+1}$. On a compact Riemannian manifold M, if the Ricci curvature of M is not less than that of S^n, then the first eigenvalue $\bar{\lambda}_1$ of Δ on M satisfies $\bar{\lambda}_1\geqslant\lambda_1=n$ [16]. Conversely, under the same assumption on the Ricci curvature, if $\bar{\lambda}_1=n$, then M is a sphere (Obata). On the other hand, if g is the standard metric on S^n, then $\Delta f=nf$ is equivalent to the system of differential equations

$$\nabla_j\nabla_i f+fg_{ji}=0. \qquad (\mathrm{E}_1)$$

A complete Riemannian manifold M ($n\geqslant 2$) admits a nontrivial solution of (E_1) if and only if M is a sphere (Obata, Tashiro). In general, the restriction f to S^n of a harmonic homogeneous polynomial of degree k satisfies $\Delta f=k(n+k-1)f$ as well as a certain system (E_k) of differential equations of degree $k+1$ involving the Riemannian metric. For example,

$$\nabla_k\nabla_j\nabla_i f+2g_{ji}\nabla_k f+g_{ik}\nabla_j f+g_{kj}\nabla_i f=0. \qquad (\mathrm{E}_2)$$

If a complete Riemannian manifold M admits a nontrivial solution of (E_k) for some integer $k\geqslant 2$, then M is locally isometric to a sphere (Obata, Tanno, S. Gallot [17]). The gradient of a solution of (E_1) is an infinitesimal conformal

transformation and that of (E_2) is an infinitesimal projective transformation.

As the Kähler or quaternion Kähler version of (E_2), there is a system of differential equations characterizing the complex projective space or the quaternion projective space as a Kähler manifold (Obata, Tanno, D. E. Blair, Y. Maeda).

On a sphere, a Riemannian metric which is conformal to the standard metric and has the same scalar curvature as the standard one is always standard, namely, it has a positive constant sectional curvature [14].

H. Scalar Curvature

On a 2-dimensional Riemannian manifold M, the sectional curvature, the Ricci curvature, and the scalar curvature all coincide with the †Gaussian curvature, which is a function on M. If M is compact, by the †Gauss-Bonnet theorem the Gaussian curvature K of M must satisfy the following sign condition in terms of the †Euler characteristic $\chi(M)$:
if $\chi(M)>0$, then K is positive somewhere;
if $\chi(M)=0$, then K changes sign unless it is identically zero;
if $\chi(M)<0$, then K is negative somewhere.
This sign condition is also sufficient for a given function K to be the Gaussian curvature of some metric on M. More precisely, starting with a Riemannian metric with constant Gaussian curvature, one can say that a smooth function K is the Gaussian curvature of some metric conformally equivalent to the original metric if and only if K satisfies the foregoing sign condition [18].

H. Yamabe [19] announced that on every compact Riemannian manifold (M,g) of dimension $n\geqslant 3$, there exists a strictly positive function u such that the Riemannian metric $\bar{g}=u^{4/(n-2)}g$ has constant scalar curvature. N. S. Trudinger, however, pointed out that his original proof contains a gap in some cases. The problem reduces to the following nonlinear partial differential equation on a compact manifold M:

$$\bar{R}u^{(n+2)/(n-2)}=4\frac{n-1}{n-2}\Delta u+Ru,$$

where R is the scalar curvature of g and $\bar{R}$ a constant which should be the scalar curvature of $\bar{g}=u^{4/(n-2)}g$ (→ 183 Global Analysis). Nevertheless, Yamabe's original proof can be pushed to cover a large class of metrics with $\int_M R\,dM<0$. Furthermore, it has since been solved for a wider class: namely, if M is not conformally flat and $n\geqslant 6$, or if it is conformally flat and its fundamental group is finite, then the problem has been solved affirmatively [20].

On the other hand, any smooth function on a compact manifold M of dimension $n\geqslant 3$ that is negative somewhere is the scalar curvature of some metric on M. In particular, on a compact manifold $(n\geqslant 3)$ there always exists a Riemannian metric with constant negative scalar curvature [18]. Any smooth function can be the scalar curvature if and only if M admits a metric of constant positive scalar curvature. The foregoing results show that there is no topological obstruction to the existence of metrics with negative scalar curvature of a compact manifold of dimension $n\geqslant 3$.

For positive scalar curvature, there is a topological obstruction. A compact †spin structure (spin manifold) having nonvanishing †$\hat{A}$-genus cannot carry a Riemannian metric of positive scalar curvature. The existence of such a manifold has been shown. If a compact, simply connected manifold M of dimension $n\geqslant 5$ is not a spin manifold, then there exists a Riemannian metric of positive scalar curvature. Furthermore, if M is a spin manifold and spin †cobordant to M' with positive scalar curvature, then M carries a Riemannian metric of positive scalar curvature [22]. A torus T^n cannot carry a metric of positive scalar curvature. In fact, any metric of nonnegative scalar curvature on T^n must be flat [22].

Let K_g and R_g denote the sectional curvature and the scalar curvature, respectively, of a Riemannian metric g. Then the following are known for a compact manifold M of dimension $\geqslant 3$: If M carries a metric g with $K_g\leqslant 0$, then it carries no metric with $R>0$. If M carries a metric g with $K_g<0$, then it carries no metric with $R\geqslant 0$. If M carries metrics g_1, g_2 with $K_{g_1}\leqslant 0$ and $R_{g_2}\geqslant 0$, then both metrics are flat [22].

If the assignment of the scalar curvature to a Riemannian metric is viewed as a mapping of a space of Riemannian metrics into a space of functions on a manifold M, then locally it is almost always surjective when M is compact (A. E. Fischer and J. E. Marsden, O. Kobayashi, J. Lafontaine).

I. Ricci Curvature and Einstein Metrics

In this paragraph the manifolds under consideration are assumed to be of dimension $n\geqslant 3$. The Ricci tensor (R_{ij}) is a symmetric tensor field of type $(0,2)$ on a Riemannian manifold. The problem of finding a Riemannian metric g which realizes a given Ricci tensor reduces to the one of solving a system of nonlinear second-order partial differential equations for g. The Bianchi identity (→ 417

Tensor Calculus)

$$g^{kj}\nabla_k R_{ji} - \frac{1}{2}\nabla_i R = 0$$

must be satisfied. There is a symmetric (0, 2)-tensor on $\mathbf{R}^n$ which cannot be the Ricci tensor for any Riemannian metric in a neighborhood of $0 \in \mathbf{R}^n$. However, if a C^∞ (or C^ω) symmetric tensor field (R_{ij}) of type (0, 2) is invertible at a point p, then in a neighborhood of p there exists a C^∞ (or C^ω) Riemannian metric g such that (R_{ij}) is the Ricci tensor of g [24].

The positivity of the Ricci curvature on a Riemannian manifold puts rather strong restrictions on the topology of the manifold (→ 178 Geodesics). However, nonnegative Ricci curvature and positive Ricci curvature are not too far from each other. If, on a complete Riemannian manifold M with nonnegative Ricci curvature, there is a point at which the Ricci curvature is positive, then there exists a complete metric on M with positive Ricci curvature [25–27].

If a Riemannian manifold (M, g) is an Einstein space, then g is called an **Einstein metric** on the manifold M. Let v_g denote the volume element determined by g. When M is compact, $\mathcal{M}$ denotes the space of Riemannian metrics on M with total volume 1. The integral of the scalar curvature $\mathcal{G}(g) = \int_M R_g v_g$ is a functional on $\mathcal{M}$. The critical points of $\mathcal{G}$ are Einstein metrics (D. Hilbert). Let $\mathcal{M}_1 (\subset \mathcal{M})$ denote the space of metrics with constant scalar curvature. Then if $\mathcal{G}$ is restricted to $\mathcal{M}_1$, then the †nullity and †coindex at the critical point are finite [28, 29].

An Einstein metric is always real analytic in some coordinate system. In particular, if two simply connected Einstein spaces have neighborhoods on which metrics are isometric, then they are isometric [30]. Though S^n with standard Riemannian metric is a typical example of an Einstein space, S^{4k+3} $(k \geqslant 1)$ carries an Einstein metric that is not standard [31].

References

[1] J. A. Wolf, Spaces of constant curvature, Publish or Perish, 1977.
[2] M. Berger, Les variétés riemanniennes homogènes simplement connexes à courbure strictment positive, Ann. Scuola Norm. Sup. Pisa, (3) 15 (1961), 179–246.
[3] N. R. Wallach, Compact homogeneous Riemannian manifolds with strictly positive curvature, Ann. Math., (2) 96 (1972), 277–295.
[4] L. Berard-Bergery, Les variétés riemanniennes homogènes simplement connexes de dimension impaire à courbure strictment positive, J. Math. Pures Appl., (9) 55 (1976), 47–67.
[5] S. Mori, Projective manifolds with ample tangent bundles, Ann. Math., (2) 110 (1979), 593–606.
[6] Y. T. Siu and S. T. Yau, Compact Kähler manifolds of positive bisectional curvature, Inventiones Math., 59 (1980), 189–204.
[7] G. de Rham, Sur la reductibilité d'un espace de Riemann, Comment. Math. Helv., 26 (1952), 328–344.
[8] M. Berger, Sur les groupes d'holonomie homogène des variétés à connexion affine et des variétés riemanniennes, Bull. Soc. Math. France, 83 (1955), 279–330.
[9] J. Simons, On the transitivity of holonomy systems, Ann. Math., (2) 76 (1962), 213–234.
[10] S. Kobayashi, Transformation groups in differential geometry, Erg. Math., 70, 1972.
[11] K. Yano, The theory of Lie derivatives and its applications, North-Holland, 1957.
[12] J. Lelong-Ferrand, Transformations conformes et quasi conformes des variétés riemanniennes compactes (démonstration de la conjecture de Lichnerowicz), Acad. Roy. Belg., Cl. Sc. Mém. Colloq. 39, no. 5 (1971), 1–44.
[13] A. J. Ledger and M. Obata, Compact Riemannian manifolds with essential group of conformorphisms, Trans. Amer. Math. Soc., 150 (1970), 645–651.
[14] M. Obata, The conjectures on conformal transformations of Riemannian manifolds, J. Differential Geometry, 6 (1971), 247–258.
[15] D. V. Alekseeviskiĭ, Groups of conformal transformations of Riemann spaces, Math. USSR-Sb., 18 (1972), 285–301.
[16] A. Lichnerowicz, Géométrie des groupes de transformations, Dunod, 1958.
[17] S. Gallot, Équations différentielles caractéristiques de la sphère, Ann. Sci. Ecole Norm. Sup., 12 (1979), 235–267.
[18] J. L. Kazdan and F. W. Warner, Existence and conformal deformation of metrics with prescribed Gaussian and scalar curvature, Ann. Math., (2) 101 (1975), 317–331.
[19] H. Yamabe, On a deformation of Riemannian structures on compact manifolds, Osaka Math. J., 12 (1960), 21–37.
[20] T. Aubin, Equations différentielles non linéaires et problème de Yamabe concernant la courbure scalaire, J. Math. Pure Appl., 55 (1976), 269–296.
[21] N. Hitchin, Harmonic spinors, Advances in Math., 14 (1974), 1–55.
[22] M. Gromov and L. B. Lawson, The classification of simply connected manifold of positive scalar curvature, Ann. Math., (2) 111 (1980), 423–434.
[23] R. Schoen and S. T. Yau, The structure

of manifolds with positive scalar curvature, Manuscripta Math., 28 (1979), 159–183.
[24] D. DeTurck, Existence of metrics with prescribed Ricci curvature: Local theory, Inventiones Math., 65 (1981), 179–207.
[25] T. Aubin, Metriques riemanniennes et courbure, J. Differential Geometry, 4 (1970), 383–424.
[26] P. Ehrlich, Metric deformations of curvature I: local convex deformations, Geometriae Dedicata, 5 (1976), 1–23.
[27] J. P. Bourguignon, Ricci curvature and Einstein metrics, Lecture notes in math. 838, Springer, 1981, 42–63.
[28] Y. Muto, On Einstein metrics, J. Differential Geometry, 9 (1974), 521–530.
[29] N. Koiso, On the second derivative of the total scalar curvature, Osaka J. Math., 16 (1979), 413–421.
[30] D. M. DeTurck and J. L. Kazdan, Some regularity theorems in Riemannian geometry, Ann. Sci. Ecole Norm. Sup., 14 (1981), 249–260.
[31] J. E. D'Atri and W. Ziller, Naturally reductive metrics and Einstein metrics on compact Lie groups, Mem. Amer. Math. Soc., 18 (1979).
See also references to 80 Connections, 105 Differentiable Manifolds, 109 Differential Geometry, 111 Differential Geometry of Curves and Surfaces, 178 Geodesics, 191 *G*-Structures, 365 Riemannian Submanifolds, 417 Tensor Calculus.

365 (VII.13) Riemannian Submanifolds

A. Introduction

If an †immersion (or an †embedding) f of a †Riemannian manifold (M, g) into a Riemannian manifold $(\tilde{M}, \tilde{g})$ satisfies the condition $f^*\tilde{g} = g$, then f is called an **isometric immersion** (or embedding) and M is called a **Riemannian submanifold** of $\tilde{M}$. In this article, $f(M)$ will be identified with M except where there is danger of confusion. Suppose $\dim M = n$ and $\dim \tilde{M} = n+p$. Then the †bundle $F(M)$ of orthonormal tangent frames of M, the bundle $F_\nu(M)$ of orthonormal normal frames of M, and their †Whitney sum $F(M) \oplus F_\nu(M)$ are †principal fiber bundles over M with †structure groups $O(n)$, $O(p)$, and $O(n) \times O(p)$, respectively. These are subbundles of the restriction to M of the bundle $F(\tilde{M})$ of orthonormal frames of $\tilde{M}$. The vector bundles associated with $F(M)$, $F_\nu(M)$, and $F(M) \oplus F_\nu(M)$ are, respectively, the †tangent bundle $T(M)$, the †normal bundle $\nu(M)$, and their Whitney sum $T(M) \oplus \nu(M)$.

B. General Results for Immersibility

An n-dimensional real analytic Riemannian manifold can be locally isometrically embedded into any real analytic Riemannian manifold of dimension $n(n+1)/2$ (M. Janet (1926), E. Cartan (1927)). The generalization to the C^∞ case is an open question even when the ambient space is Euclidean.

An n-dimensional compact C^r Riemannian manifold $(3 \leqslant r \leqslant \infty)$ can be isometrically embedded into an $(n(3n+11)/2)$-dimensional Euclidean space (J. F. Nash (1956)). An n-dimensional noncompact C^r Riemannian manifold $(3 \leqslant r \leqslant \infty)$ can be isometrically embedded into a $2(2n+1)(3n+7)$-dimensional Euclidean space (Nash (1956), R. E. Greene (1970)).

Let M be an n-dimensional Riemannian manifold with †sectional curvature K_M and $\tilde{M}$ an $(n+p)$-dimensional Riemannian manifold with sectional curvature $K_{\tilde{M}}$. Then M cannot be isometrically immersed into $\tilde{M}$ in the following cases:
(1) $p \leqslant n-2$ and $K_M < K_{\tilde{M}}$ (T. Otsuki (1954));
(2) $p \leqslant n-1$, $K_M \leqslant K_{\tilde{M}} \leqslant 0$, M is compact, and $\tilde{M}$ is complete and simply connected (C. Tompkins (1939), S. S. Chern and N. H. Kuiper (1952), B. O'Neill (1960));
(3) $p \leqslant n-1$, $K_M \leqslant 0$, $K_{\tilde{M}}$ is constant ($\leqslant 0$), M is compact, and $\tilde{M}$ is complete and simply connected [2].

C. Fundamental Equations

Let $f:(M,g) \to (\tilde{M}, \tilde{g})$ be an isometric immersion. Let ∇ and $\tilde{\nabla}$ denote the †covariant differentiations with respect to the †Riemannian connections of M and $\tilde{M}$, respectively. For vector fields X and Y on M, the tangential component of $\tilde{\nabla}_X Y$ is equal to $\nabla_X Y$. Put

$$\sigma(X, Y) = \tilde{\nabla}_X Y - \nabla_X Y. \qquad (1)$$

Then σ is a $\nu(M)$-valued symmetric $(0,2)$ tensor field on M, which is called the **second fundamental form** of M (or of f). For a normal vector ξ at $x \in M$, put $g(A_\xi X, Y) = \tilde{g}(\sigma(X, Y), \xi)$. Then A_ξ defines a symmetric linear transformation on $T_x(M)$, which is called the second fundamental form in the direction of ξ. The eigenvalues of A_ξ are called the **principal curvatures** in the direction of ξ. The connection on $\nu(M)$ induced from the Riemannian connection of $\tilde{M}$ is called the **normal connection** of M (or of f). Let $\nabla^\perp$ denote the covariant differentiation with respect to the normal con-

nection. For a tangent vector field X and a normal vector field ξ on M, the tangential (resp. normal) component of $\tilde{\nabla}_X \xi$ is equal to $-A_\xi X$ (resp. $\nabla^\perp_X \xi$), that is to say, the relation

$$\tilde{\nabla}_X \xi = -A_\xi X + \nabla^\perp_X \xi \qquad (2)$$

holds. (1) is called the **Gauss formula**, and (2) is called the **Weingarten formula**.

Let R, $\tilde{R}$, and $R^\perp$ be the †curvature tensors of ∇, $\tilde{\nabla}$, and $\nabla^\perp$, respectively. Then the †integrability condition for (1) and (2) implies

$$\tilde{R}(X,Y)Z = R(X,Y)Z + A_{\sigma(X,Z)}Y - A_{\sigma(Y,Z)}X + (\nabla'_X\sigma)(Y,Z) - (\nabla'_Y\sigma)(X,Z) \qquad (3)$$

for vector fields X, Y, Z tangent to M, where ∇' denotes covariant differentiation with respect to the connection in $T(M) \oplus \nu(M)$. (3) is called the equation of Gauss and Codazzi. More precisely, the tangential component of (3) is given by the **equation of Gauss** and the normal component of (3) is given by the **equation of Codazzi**. Similarly, for vector fields ξ and η normal to M, the relation

$$\tilde{g}(\tilde{R}(X,Y)\xi,\eta) = \tilde{g}(R^\perp(X,Y)\xi,\eta) - g([A_\xi, A_\eta]X, Y) \qquad (4)$$

holds, which is called the **equation of Ricci**. Formulas (1)–(4) are the fundamental equations for the isometric immersion $f: M \to \tilde{M}$.

As a particular case, suppose $\tilde{M}$ is a †space form of constant curvature c. Then the equations of Gauss, Codazzi, and Ricci reduce respectively to

$$R(X,Y)Z = c(g(Y,Z)X - g(X,Z)Y) + A_{\sigma(Y,Z)}X - A_{\sigma(X,Z)}Y, \qquad (3_c)_t$$

$$(\nabla'_X\sigma)(Y,Z) - (\nabla'_Y\sigma)(X,Z) = 0, \qquad (3_c)_n$$

$$\tilde{g}(R^\perp(X,Y)\xi,\eta) = g([A_\xi, A_\eta]X, Y). \qquad (4_c)$$

Conversely, let (M,g) be an n-dimensional simply connected Riemannian manifold, and suppose there is given a p-dimensional †Riemannian vector bundle $\nu(M)$ over M with curvature tensor $R^\perp$ and a $\nu(M)$-valued symmetric $(0,2)$ tensor field σ on M. For a †cross section ξ of $\nu(M)$, define A_ξ by $g(A_\xi X, Y) = \langle \sigma(X,Y), \xi \rangle$, where $\langle\ ,\ \rangle$ is the fiber metric of $\nu(M)$. If they satisfy $(3_c)_t$, $(3_c)_n$, and (4_c), then M can be immersed isometrically into an $(n+p)$-dimensional complete and simply connected space form $M^{n+p}(c)$ of curvature c in such a way that $\nu(M)$ is the normal bundle and σ is the second fundamental form. Moreover, such an immersion is unique up to an †isometry of $M^{n+p}(c)$.

Let $(e_A)_{1 \leq A \leq n+p}$ be a local cross section of $F(\tilde{M})$ such that its restriction to M gives a local cross section of $F(M) \oplus F_\nu(M)$, and let (ω^A) be its dual. Then $f^*\omega^\alpha = 0$ for $n+1 \leq \alpha \leq n+p$. Let $(\tilde{\omega}^A_B)_{1 \leq A,B \leq n+p}$ and $(\tilde{\Phi}^A_B)_{1 \leq A,B \leq n+p}$ be the †connection form and the †curvature form of $\tilde{M}$ with respect to (e_A), and put $\omega^A_B = f^*\tilde{\omega}^A_B$. Then $(\omega^i_j)_{1 \leq i,j \leq n}$ is the connection form of M with respect to $(e_i)_{1 \leq i \leq n}$. $(\omega^\alpha_i)_{1 \leq i \leq n < \alpha \leq n+p}$ gives the second fundamental form, that is,

$$\sigma(e_i, e_j) = \sum \omega^\alpha_i(e_j) e_\alpha. \qquad (1')$$

Put $\omega^\alpha_i = \sum h^\alpha_{ij}\omega^j$. Then (h^α_{ij}) is the matrix representing the symmetric linear transformation A_{e_α} with respect to (e_i), that is,

$$A_{e_\alpha} e_i = \sum h^\alpha_{ij} e_j. \qquad (2')$$

Moreover, $(\omega^\alpha_\beta)_{n+1 \leq \alpha,\beta \leq n+p}$ is the connection form of the normal connection with respect to $(e_\alpha)_{n+1 \leq \alpha \leq n+p}$. Let $(\Phi^i_j)_{1 \leq i,j \leq n}$ and $(\Phi^\alpha_\beta)_{n+1 \leq \alpha,\beta \leq n+p}$ be the curvature forms of (ω^i_j) and (ω^α_β), respectively. Then the equations of Gauss, Codazzi, and Ricci are given respectively by

$$f^*\tilde{\Phi}^i_j = \Phi^i_j + \sum_{\alpha=n+1}^{n+p} \omega^i_\alpha \wedge \omega^\alpha_j \quad (1 \leq i,j \leq n), \qquad (3')_t$$

$$f^*\tilde{\Phi}^\alpha_i = d\omega^\alpha_i + \sum_{A=1}^{n+p} \omega^\alpha_A \wedge \omega^A_i \quad (1 \leq i \leq n < \alpha \leq n+p), \qquad (3')_n$$

$$f^*\tilde{\Phi}^\alpha_\beta = \Phi^\alpha_\beta + \sum_{k=1}^{n} \omega^\alpha_k \wedge \omega^k_\beta \quad (n+1 \leq \alpha,\beta \leq n+p). \qquad (4')$$

D. Basic Notions

Let M be a Riemannian submanifold of $\tilde{M}$. A point $x \in M$ is called a **geodesic point** if $\sigma = 0$ at x. If every point of M is a geodesic point, then M is called a **totally geodesic submanifold** of $\tilde{M}$. M is a totally geodesic submanifold of $\tilde{M}$ if and only if every geodesic of M is a geodesic of $\tilde{M}$.

A mapping $\mathfrak{h}: M \to \nu(M)$ defined by $x \to \frac{1}{n}\sum_{i=1}^{n} \sigma(e_i, e_i)$ is independent of the choice of an orthonormal basis (e_i). $\mathfrak{h}$ is called the **mean curvature vector** and $\|\mathfrak{h}\|$ is called the **mean curvature**. M is called a **minimal submanifold** of $\tilde{M}$ if $\mathfrak{h} \equiv 0$ (→ 275 Minimal Submanifolds).

A point $x \in M$ is called an **umbilical point** if $\sigma = g \otimes \mathfrak{h}$ at x. $x \in M$ is an umbilical point if and only if A_ξ is proportional to the identity transformation for all $\xi \in \nu_x(M)$. If every point of M is an umbilical point, then M is called a **totally umbilical submanifold** of $\tilde{M}$.

A point $x \in M$ is called an **isotropic point** if $\|\sigma(X,X)\|/\|X\|^2$ does not depend on $X \in T_x(M)$. If every point of M is an isotropic point, then M is called an **isotropic submanifold** of $\tilde{M}$. It is clear that an umbilical point is an isotropic point.

$\mu(x) = \dim \bigcap_{\xi \in \nu_x(M)} \ker A_\xi$ is called the **index of relative nullity** at $x \in M$.

E. Rigidity

An isometric immersion $f: M \to \tilde{M}$ is said to be **rigid** if it is unique up to an isometry of $\tilde{M}$, that is, if $f': M \to \tilde{M}$ is another isometric immersion, then there exists an isometry φ of $\tilde{M}$ such that $f' = \varphi \circ f$. If $f: M \to \tilde{M}$ is rigid, then every isometry of M can be extended to an isometry of $\tilde{M}$.

An isometric immersion $f: M \to M^{n+1}(c)$ of an n-dimensional Riemannian manifold into an $(n+1)$-dimensional complete and simply connected space form is rigid in each of the following cases:
(1) $n=2$, $c=0$, and M is compact and of positive curvature (S. Cohn-Vossen (1929)).
(2) The index of relative nullity is $\leqslant n-3$ at each point (R. Beez (1876); → [8]).
(3) $n \geqslant 5$, $c>0$, M is complete, and the index of relative nullity is $\leqslant n-2$ at each point (D. Ferus (1970)).
(4) $n \geqslant 4$, $c \neq 0$, and the †scalar curvature of M is constant ($\neq n(n-1)c$) (C. Harle (1971)).

A generalization of (2) for the case of higher codimension was obtained by C. Allendoerfer (1939). Various rigidity conditions have been studied by S. Dolbeault-Lemoine, R. Sacksteder, E. Kaneda and N. Tanaka, and others.

F. Totally Geodesic and Totally Umbilical Submanifolds

A totally geodesic submanifold of a space form is also a space form of the same curvature. Totally geodesic submanifolds of compact †symmetric spaces of rank 1 were completely classified by J. A. Wolf (1963), and totally geodesic submanifolds of symmetric spaces of rank $\geqslant 2$ were studied by Wolf and B. Y. Chen and T. Nagano [4].

Let $f: M \to M^{n+p}(\tilde{c})$ be a totally umbilical immersion of an n-dimensional Riemannian manifold M into an $(n+p)$-dimensional space form. Then M is a space form $M^n(c)$ with $c \geqslant \tilde{c}$, and $f(M)$ is contained in a certain $(n+1)$-dimensional totally geodesic submanifold $M^{n+1}(\tilde{c})$ of $M^{n+p}(\tilde{c})$. If $\tilde{c}>0$, then $f(M)$ is locally a hypersphere; if $\tilde{c}=0$, then $f(M)$ is locally a hyperplane or a hypersphere; if $\tilde{c}<0$, then $f(M)$ is locally a geodesic sphere, a horosphere, or a parallel hypersurface of a totally geodesic hypersurface [2].

G. Minimal Submanifolds

For general properties of minimal submanifolds → 275 Minimal Submanifolds.

There is no compact minimal submanifold in a simply connected Riemannian manifold with nonpositive sectional curvature (O'Neill (1960)). On the contrary, a sphere has plenty of compact minimal submanifolds.

For each positive integer s, an n-dimensional sphere of curvature $\dfrac{n}{s(s+n-1)}$ can be minimally immersed into a $\left\{(2s+n-1)\dfrac{(s+n-2)!}{s!(n-1)!}-1\right\}$-dimensional unit sphere and the immersion is rigid if $n=2$ or $s \leqslant 3$ (E. Calabi (1967), M. do Carmo and N. Wallach (1971)).

Among all n-dimensional compact minimal submanifolds of an $(n+p)$-dimensional unit sphere, the totally geodesic submanifold is isolated in the sense that it is characterized by each of the following conditions:
(1) sectional curvature $> \dfrac{n}{2(n+1)}$ (T. Itoh (1978)),
(2) Ricci curvature $> n-2$ (N. Ejiri (1979)),
(3) scalar curvature $> n(n-1) - \dfrac{n}{2-1/p}$ (J. Simons (1968)).

H. Submanifolds of Constant Mean Curvature

A manifold of constant mean curvature is a solution to a variational problem. In particular, with respect to any volume-preserving variation of a domain D in a Euclidean space, the mean curvature of $M = \partial D$ is constant if and only if the volume of M is critical.

The interesting question "If the mean curvature of an isometric immersion $f: M \to M^{n+1}(c)$ of an n-dimensional compact Riemannian manifold M into an $(n+1)$-dimensional space form $M^{n+1}(c)$ is constant, is M a sphere?" has not yet been completely solved, where $M^{n+1}(c)$ denotes a Euclidean space, a hyperbolic space, or an open hemisphere according as $c=0$, <0, or >0. The answer is affirmative in the following cases: (1) $\dim M = 2$, and the †genus of M is zero (H. Hopf (1951), Chern (1955)). (2) f is an embedding (A. D. Alexandrov (1958); → [8]).

These results remain true even if the assumption "the mean curvature is constant" is replaced by the weaker condition "the principal curvatures $k_1 \geqslant \ldots \geqslant k_n$ satisfy a relation $\varphi(k_1, \ldots, k_n) = 0$ such that $\partial\varphi/\partial k_i > 0$."

Unlike an open hemisphere, a sphere S^{n+1} admits many compact hypersurfaces of constant mean curvature, among which totally umbilical hypersurfaces and the product of two spheres are the only ones with nonnegative sectional curvature (B. Smyth and K. Nomizu (1969)).

A nonnegatively or nonpositively curved complete surface of nonzero constant mean curvature in a 3-dimensional space form $M^3(c)$ is either a sphere or a †Clifford torus if $c>0$ and is either a sphere or a right circular

cylinder if $c \leqslant 0$ (T. Klotz and R. Osserman (1966–1967), D. Hoffman (1973)).

I. Isoparametric Hypersurfaces

A hypersurface M of $\tilde{M}$ is said to be **isoparametric** if M is locally defined as the †level set of a function f on (an open set of) $\tilde{M}$ with property

$$df \wedge d\|df\|^2 = 0 \quad \text{and} \quad df \wedge d(\Delta f) = 0.$$

A hypersurface M of a complete and simply connected space form $M^{n+1}(c)$ is isoparametric if and only if M has locally constant principal curvatures (Cartan). If $c \leqslant 0$, M has at most two distinct principal curvatures (Cartan). If $c > 0$, the number of distinct principal curvatures of M is 1, 2, 3, 4, or 6 (H. Münzner (1980)). If $c = 0$, then M is locally $S^k \times E^{n-k}$, and if $c < 0$, then M is locally E^n or $S^k \times H^{n-k}$ (Cartan). Isoparametric hypersurfaces of S^{n+1} having at most three distinct principal curvatures were completely classified by Cartan. R. Takagi, T. Takahashi, and H. Ozeki and M. Takeuchi obtained several results for isoparametric hypersurfaces of S^{n+1} with four or six distinct principal curvatures [7].

If a subgroup of the isometry group of $M^{n+1}(c)$ acts transitively on M, then M is isoparametric. The converse is true if $c \leqslant 0$, or if $c > 0$ and M has at most three distinct principal curvatures (Cartan), but not true in general (Ozeki and Takeuchi [7]).

J. Isometric Immersions between Space Forms

Let $f: M^n(c) \to M^{n+p}(\tilde{c})$ be an isometric immersion of an n-dimensional space form into an $(n+p)$-dimensional space form.
(1) If $n=2$, $p=1$, $c>0$, $c \geqslant \tilde{c}$, $M^2(c)$ is complete, and $M^3(\tilde{c})$ is complete and simply connected, then f is totally umbilical (H. Liebmann (1901); → [2]).
(2) If $n=2$, $p=1$, $c<0$, $c<\tilde{c}$, $M^2(c)$ is complete, and $M^3(\tilde{c})$ is complete and simply connected, then f does not exist (D. Hilbert (1901); → [2]).
(3) If $n=2$, $p=1$, $c=0<\tilde{c}$, $M^2(0)$ is complete, and $M^3(\tilde{c})$ is complete and simply connected, then there exist infinitely many f (L. Bianchi (1896); → [2]).
(4) If $n=2$, $p=1$, $c=0>\tilde{c}$, $M^2(0)$ is complete, and $M^3(\tilde{c})$ is complete and simply connected, then $f(M^2(0))$ is either a horosphere or a set of points at a fixed distance from a geodesic (J. Volkovand S. Vladimirova, S. Sasaki; → [2]).
(5) If $n \geqslant 3$, $p=1$, and $c>\tilde{c}$, then f is totally umbilical.
(6) If $p=1$, $c=\tilde{c}=0$, $M^n(0)$, is complete, and $M^{n+1}(0)$ is complete and simply connected, then f is cylindrical (A. Pogorelov (1956), P. Hartman and L. Nirenberg (1959), and others).
(7) If $p \leqslant n-1$, $c=\tilde{c}>0$, and both $M^n(c)$ and $M^{n+p}(\tilde{c})$ are complete, then f is totally geodesic (D. Ferus (1975)).
(8) If $p \leqslant n-1$, $\tilde{c}>c>0$, $M^n(c)$ is complete, and $M^{n+p}(\tilde{c})$ is complete and simply connected, then f does not exist (J. Moore (1972)).

K. Homogeneous Hypersurfaces

Let M be an n-dimensional †homogeneous Riemannian manifold which is isometrically immersed into an $(n+1)$-dimensional complete and simply connected space form $M^{n+1}(c)$.
(1) If $c=0$, then M is isometric to $S^k \times E^{n-k}$ (S. Kobayashi (1958), Nagano (1960), Takahashi [9]).
(2) If $c>0$, then M is isometric to E^2 or else is given as an orbit of a subgroup of the isometry group of $M^{n+1}(c)$ (W. Y. Hsiang, H. B. Lawson, Takagi; → [7]).
(3) If $c<0$, then M is isometric to E^n, $S^k \times H^{n-k}$, or a 3-dimensional group manifold

$$B = \left\{ \begin{pmatrix} e^t & 0 & x \\ 0 & e^{-t} & y \\ 0 & 0 & 1 \end{pmatrix} \middle| x, y, t \in \mathbf{R} \right\}$$

with the metric $ds^2 = e^{-2t}dx^2 + e^{2t}dy^2 + dt^2$ (Takahashi (1971)). Each of the hypersurfaces above except E^2 in (2) and B in (3) is given as an orbit of a certain subgroup of the isometry group of $M^{n+1}(c)$.

L. Kähler Submanifolds

A †complex submanifold of a †Kähler manifold is a Kähler manifold with respect to the induced Riemannian metric. A complex analytic and isometric immersion of a Kähler manifold (M, J, g) into a Kähler manifold $(\tilde{M}, \tilde{J}, \tilde{g})$ is called a **Kähler immersion**, and M is called a **Kähler submanifold** of $\tilde{M}$. A Kähler submanifold is a minimal submanifold. A compact Kähler submanifold M of a Kähler manifold $\tilde{M}$ can never be homologous to 0, that is, there exists no submanifold M' of $\tilde{M}$ such that $M = \partial M'$. If $[M] \in H_*(\tilde{M}, Z)$ denotes the †homology class represented by a Kähler submanifold M of $\tilde{M}$, then $\mathrm{vol}(M) \leqslant \mathrm{vol}(M')$ holds for any submanifold $M' \in [M]$ with equality if and only if M' is a Kähler submanifold (W. Wirtinger (1936)).

A Kähler manifold of constant †holomorphic sectional curvature is called a **complex space form**. An n-dimensional complete and simply connected complex space form is either $P_n(C)$, C^n, or D_n. Every Kähler submanifold of a complex space form is rigid (Calabi [10]).

Kähler immersions of complex space forms

into complex space forms were completely determined by Calabi [10] and by H. Nakagawa and K. Ogiue (1976).

C^n (resp. D_n) is the only †Hermitian symmetric space that can be immersed in C^m (resp. D_m) as a Kähler submanifold (Nakagawa and Takagi [11]), and Kähler immersions of Hermitian symmetric spaces into $P_m(C)$ were precisely studied by Nakagawa, Y. Sakane, Takagi, Takeuchi, and others. More generally, Kähler immersions of homogeneous Kähler manifolds into $P_m(C)$ were determined by Takeuchi (1978).

$Q_n=\{[z_i]\in P_{n+1}(C)\mid\sum z_i^2=0\}$ in $P_{n+1}(C)$ is the only Einstein-Kähler hypersurface of a complex space form that is not totally geodesic (B. Smyth (1967), S. S. Chern (1967)). The result remains true even if "Einstein" is replaced by "parallel Ricci tensor" (Takahashi (1967)). Besides linear subspaces, Q_n is the only Einstein-Kähler submanifold of $P_m(C)$ that is a complete intersection (J. Hano (1975)).

Integral theorems and pinching problems with respect to various curvatures for compact Kähler submanifolds of $P_m(C)$ have been studied by K. Ogiue, S. Tanno (1973), S. T. Yau (1975), and others [12]. For example, if the holomorphic sectional curvature of $P_{n+p}(C)$ is 1, then each of the following is sufficient for an n-dimensional compact Kähler submanifold to be totally geodesic:
(1) holomorphic sectional curvature $>1/2$ (A. Ros (1985)),
(2) sectional curvature $>1/8$ (A. Ros and L. Verstraelen (1984)),
(3) Ricci curvature $>n/2$ [12],
(4) embedded and scalar curvature $>n^2$ (J. H. Cheng (1981)).

The index of relative nullity $\mu(x)$ of an n-dimensional complete Kähler submanifold M of $P_m(C)$ satisfies $\mathrm{Min}_{x\in M}\,\mu(x)=0$ or $2n$ (K. Abe (1973)).

M. Totally Real Submanifolds

An isometric immersion of a Riemannian manifold (M,g) into a Kähler manifold $(\tilde{M},\tilde{J},\tilde{g})$ satisfying $\tilde{J}T_x(M)\subset \nu_x(M)$ at each point $x\in M$ is called a **totally real immersion**, and M is called a **totally real submanifold** of $\tilde{M}$. A totally geodesic submanifold $P_n(R)$ in $P_n(C)$, $S^1\times S^1$ in $P_2(C)$ and an immersion $P_n(C)\to P_{n(n+2)}(C)$ defined by $[z_i]\to[z_i\bar{z}_j]$ give typical examples of totally real submanifolds.

N. Submanifolds with Planar Geodesics

A surface in E^3 whose geodesics are all plane curves is (a part of) a plane or sphere. More generally, let $f:M\to M^m(c)$ be an isometric immersion of M into a complete and simply connected space form $M^m(c)$. If the image of each geodesic of M is contained in a 2-dimensional totally geodesic submanifold of $M^m(c)$, then f is either a totally geodesic immersion, a totally umbilical immersion or a minimal immersion of a compact symmetric space of rank 1 by harmonic functions of degree 2; the last case occurs only when $c>0$ (S. L. Hong (1973), J. Little (1976), K. Sakamoto [14]).

A Kähler submanifold of a complete and simply connected complex space form with the same property as above is either a totally geodesic submanifold or a Veronese submanifold (a Kähler immersion of $P_n(C)$ into $P_{n(n+3)/2}(C)$) (K. Nomizu (1976)).

Submanifolds with the above property are closely related to isotropic submanifolds with $\nabla'\sigma=0$. Submanifolds with $\nabla'\sigma=0$ in symmetric spaces have been studied by Ferus, H. Naito, Takeuchi, K. Tsukada, and others.

O. Total Curvature

Let $f:M\to E^m$ be an isometric immersion of an n-dimensional compact Riemannian manifold M into a Euclidean space. Let $\nu_1(M)$ be the unit normal bundle, S^{m-1} the unit sphere centered at $O\in E^m$, and let $f_\perp:\nu_1(M)\to S^{m-1}$ be the parallel translation. Let ω and Ω be the †volume elements of $\nu_1(M)$ and S^{m-1}, respectively. Then for each $\xi\in\nu_1(M)$, $f_\perp^*\Omega=(\det A_\xi)\omega$ holds. As a generalization of the †total curvature for a space curve, the **total curvature** of the immersion f is defined as

$$\tau(f)=\frac{1}{\mathrm{vol}(S^{m-1})}\int_{\nu_1(M)}|f_\perp^*\Omega|$$
$$=\frac{1}{\mathrm{vol}(S^{m-1})}\int_{\nu_1(M)}|\det A_\xi|\omega.$$

If $\beta(M)$ is the least number of critical points of a †Morse function on M, then

$$\inf_f\tau(f)=\beta(M)\geqslant 2$$

holds (Chern and R. Lashof (1957, 1958), Kuiper (1958)). $\tau(f)=2$ if and only if f is an embedding and $f(M)$ is a convex hypersurface of some E^{n+1} in E^m (Chern and Lashof (1958)). If $\tau(f)<3$, then M is homeomorphic to S^n (Chern and Lashof (1958)). These results generalize theorems for space curves by W. Fenchel (1929), I. Fary (1949), J. Milnor (1950), and others.

An isometric immersion f which attains $\inf(f)$ is called a **minimum immersion** or a **tight immersion**. †Exotic spheres do not have minimum immersions (Kuiper (1958)). †R-

spaces have minimum immersions and, in particular, a minimum immersion of a symmetric R-space is a †minimal immersion into a hypersphere (Kobayashi and Takeuchi (1968)).

The **total mean curvature** of an isometric immersion $f: M \to E_m$ of an n-dimensional compact Riemannian manifold into a Euclidean space is, by definition,

$$\int_M \|\mathfrak{h}\|^n * 1.$$

It satisfies

$$\int_M \|\mathfrak{h}\|^n * 1 \geqslant \mathrm{vol}(S^n),$$

where S^n is the n-dimensional unit sphere. The equality holds if and only if f is totally umbilical (T. J. Willmore (1968), Chen [3]).

References

[1] S. Kobayashi and K. Nomizu, Foundations of differential geometry II, Interscience, 1969.
[2] M. Spivak, A comprehensive introduction to differential geometry III, IV, V, Publish or Perish, 1975.
[3] B. Y. Chen, Geometry of submanifolds, Dekker, 1973.
[4] B. Y. Chen and T. Nagano, Totally geodesic submanifolds of symmetric spaces I, II, III, Duke Math. J., 44 (1977), 745–755; 45 (1978), 405–425.
[5] S. S. Chern, M. doCarmo, and S. Kobayashi, Minimal submanifolds of a sphere with second fundamental form of constant length, Functional Analysis and Related Fields, Springer, 1970.
[6] A. D. Alexandrov, A characteristic property of spheres, Ann. Mat. Pura Appl., 58 (1962), 303–315.
[7] H. Ozeki and M. Takeuchi, On some types of isoparametric hypersurfaces I, II, Tôhoku Math. J., 27 (1975), 515–559; 28 (1976), 7–55.
[8] P. Ryan, Homogeneity and some curvature conditions for hypersurfaces, Tôhoku Math. J., 21 (1969), 363–388.
[9] T. Takahashi, Homogeneous hypersurfaces in spaces of constant curvature, J. Math. Soc. Japan, 22 (1970), 395–410.
[10] E. Calabi, Isometric imbedding of complex manifolds, Ann. Math., (2) 58 (1953), 1–23.
[11] H. Nakagawa and R. Takagi, On locally symmetric Kähler submanifolds in a complex projective space, J. Math. Soc. Japan, 28 (1976), 638–667.
[12] K. Ogiue, Differential geometry of Kähler submanifolds, Advances in Math., 13 (1974), 73–114.
[13] B. Y. Chen and K. Ogiue, On totally real submanifolds, Trans. Amer. Math. Soc., 193 (1974), 257–266.
[14] K. Sakamoto, Planar geodesic immersions, Tôhoku Math. J., 29 (1977), 25–56.
[15] D. Ferus, Totale Absolutkrümmung in Differentialgeometrie und Topologie, Springer, 1968.

366 (VIII.6) Riemann-Roch Theorems

A. General Remarks

The †Riemann-Roch theorem (abbreviation: R. R. theorem) is one of the most significant results in the classical theory of †algebraic functions of one variable. Let X be a compact †Riemann surface of †genus g, and let $D = \sum m_i P_i$ be a †divisor on X. We denote by $\deg D$ the degree of D, which is defined to be $\sum m_i$. The divisor D is said to be positive if $D \neq 0$ and $m_i \geqslant 0$ for all i. A nonzero †meromorphic function f on X determines a divisor $(f) = \sum a_i Q_i - \sum b_j R_j (a_i, b_j > 0)$, where the Q_i are the zeros of order a_i and the R_j are the poles of order b_j. The set of meromorphic functions f such that $(f) + D$ is positive, together with the constant $f = 0$, forms a finite-dimensional linear space $L(D)$ over $\mathbf{C}$. The R. R. theorem asserts that $\dim L(D) = \deg D - g + 1 + r(D)$, where $r(D)$ is a nonnegative integer determined by D. If K is the †canonical divisor of X, then $r(D) = \dim L(K - D)$ (→ 9 Algebraic Curves C; 11 Algebraic Functions D). (For the R. R. theorem for algebraic surfaces → 15 Algebraic Surfaces D.)

Generalizations of this important theorem to the case of higher-dimensional compact †complex manifolds were obtained by K. Kodaira, F. Hirzebruch, A. Grothendieck, M. F. Atiyah and I. M. Singer, and others. Let X be a compact complex manifold, B be a †complex line bundle over X, and $\mathcal{O}(B)$ be the †sheaf of germs of holomorphic cross sections of B. When B is determined by a divisor D of X, we have $H^0(X, \mathcal{O}(B)) \cong L(D)$. Hence a desirable generalization of the R. R. theorem will provide a description of $\dim_{\mathbf{C}} H^0(X, \mathcal{O}(B))$ in terms of quantities relating to the properties of X and B. Following this idea, various theorems of Riemann-Roch type have been obtained.

B. Hirzebruch's Theorem of R. R. Type

Keeping the notation given in Section A, we put $\chi(X, \mathcal{O}(B)) = \sum_q (-1)^q \dim H^q(X, \mathcal{O}(B))$.

Generally, if $\mathscr{F}$ is an arbitrary †coherent analytic sheaf on X, we can define $\chi(X, \mathscr{F})$ using the same formula (replacing $\mathcal{O}(B)$ by $\mathscr{F}$). The quantity $\chi(X, \mathscr{F})$ has simple properties in various respects. For example, if the sequence $0 \rightarrow \mathscr{F}' \rightarrow \mathscr{F} \rightarrow \mathscr{F}'' \rightarrow 0$ is exact, we have $\chi(X, \mathscr{F}) = \chi(X, \mathscr{F}') + \chi(X, \mathscr{F}'')$. If an †analytic vector bundle F depends continuously on auxiliary parameters, then $\chi(X, \mathcal{O}(F))$ remains constant. Let X be a projective algebraic manifold of complex dimension n. We consider the †Chern class $c = 1 + c_1 + \ldots + c_n$ of X and express it formally as the product $\prod_{i=1}^{n}(1+\gamma_i)$. Thus the ith Chern class c_i is expressed as the ith elementary symmetric function of $\gamma_1, \ldots, \gamma_n$. Consider the formal expression $T(X) = \prod_{i=1}^{n} \gamma_i / (1 - e^{-\gamma_i})$. $T(X)$ can be expanded as a formal power series in the γ_i, and each homogeneous term, being symmetric in the γ_i, can be expressed as a polynomial in $c_1, \ldots, c_n$, and thus determines a cohomology class of X. Similarly, we consider the formal expression of the Chern class of the vector bundle F as $1 + d_1 + \ldots + d_q = \prod_{j=1}^{q}(1+\delta_j)$, where q is the dimension of the fiber of F. We put $ch(F) = \sum_{j=1}^{q} e^{\delta_j}$. The formal series $ch(F)$ is also an element of the cohomology ring of X whose $(\nu+1)$st term consists of a 2ν-dimensional cohomology class. We call $ch(F)$ the †Chern character of F (→ 237 K-Theory B), and define $T(X, F)$ to be the value of $ch(F)\ T(X)$ at the †fundamental cycle X. (The multiplication $ch(F) \cdot T(X)$ is formal. $T(X, F)$ is determined by the term of dimension $2n$ alone.) $T(X, F)$ is called the **Todd characteristic** with respect to F. **Hirzebruch's theorem of R. R. type** asserts that $\chi(X, \mathcal{O}(F)) = T(X, F)$. In particular, when $n = 1$, $F = [D]$ (the line bundle determined by the divisor D), Hirzebruch's formula yields the classical R. R. theorem. If F satisfies the conditions for the vanishing theorem of cohomologies, the formula gives an estimate for $\dim H^0(X, \mathcal{O}(F))$ [11].

In 1963, Atiyah and Singer developed a theory on indices of elliptic differential operators on a compact orientable differentiable manifold and obtained a general result that includes the proof of Hirzebruch's theorem for an arbitrary compact complex manifold [4, 5] (→ 237 K-Theory H).

C. R. R. Theorem for Surfaces

If X is a compact complex surface, i.e., a compact complex manifold of dimension 2, then for complex line bundles F_1 and F_2, the intersection number $(F_1 F_2)$ is defined to be $c_1(F_1) \cup c_1(F_2)[X]$. The **R. R. theorem for a line bundle** F on X is stated as follows: $\chi(X, \mathcal{O}(F)) = (F^2)/2 - (KF)/2 + ((K^2) + c_2(X))/12$, where K is the canonical line bundle of X and $c_2(X)$ denotes the value at X of the 2nd Chern class of X, that is, the Euler number of X.

The Noether formula $\chi_X = ((K^2) + c_2(X))/12$ follows from the above identity. The R. R. theorem for surfaces is a powerful tool for the study of compact complex surfaces.

D. Grothendieck's Theorem of R. R. Type

Grothendieck took an entirely new point of view in generalizing Hirzebruch's theorem. The following is a description of his idea as reformulated by A. Borel and J.-P. Serre [8]. We consider a nonsingular quasiprojective algebraic variety X (→ 16 Algebraic Varieties) over a ground field of arbitrary characteristic. Namely, X is a closed subvariety of an open set in a projective space (over an algebraically closed ground field). Consider the group $K(X)$, which is the quotient of the free Abelian group generated by the equivalence classes of algebraic vector bundles over X modulo the subgroup generated by the elements of the form $F - F' - F''$, where F, F', F'' are classes of bundles such that there exists an exact sequence $0 \rightarrow F' \rightarrow F \rightarrow F'' \rightarrow 0$. A similar construction for the (equivalence classes of the) †coherent algebraic sheaves instead of the vector bundles yields another Abelian group $K'(X)$. It can be shown that $K(X)$ is isomorphic to $K'(X)$ by the correspondence $F \rightarrow \mathcal{O}(F)$ (= the sheaf of germs of regular cross sections of F). Addition in $K(X)$ is induced by the †Whitney sum of the bundles, and $K(X)$ has the structure of a ring with multiplication induced by the tensor product. For a vector bundle F, its Chern class $c(F) = 1 + c_1(F) + \ldots + c_q(F)$ (q = the dimension of the fiber) is defined as an element of the †Chow ring $A(X)$ with appropriate properties. ($A(X)$ is the ring of the rational equivalence classes of algebraic cycles on X, and $c_i(F)$ is the class of a cycles of codimension i.) We define $ch(F)$ as before. It can be shown that $c(F)$ and $ch(F)$ are determined by the image of F in $K(X)$, and we have $c(\xi + \eta) = c(\xi)c(\eta)$, $ch(\xi + \eta) = ch(\xi) + ch(\eta)$, $ch(\xi\eta) = ch(\xi)ch(\eta)$ $(\xi, \eta \in K(X))$. If we have a †proper morphism $f: Y \rightarrow X$ between quasiprojective algebraic varieties Y and X, we have homomorphisms $f^!: K(X) \rightarrow K(Y)$ and $f_!: K(Y) \rightarrow K(X)$. The former is defined by taking the induced vector bundle and the latter by the correspondence

$$\mathscr{F} \rightarrow \sum (-1)^q (\mathscr{R}^q f) \mathscr{F},$$

where $\mathscr{F}$ is a coherent algebraic sheaf on Y and $(\mathscr{R}^q f)\mathscr{F}$ is the qth †direct image of $\mathscr{F}$ under f. (Since f is proper, $(\mathscr{R}^q f)\mathscr{F}$ is coherent.) Between Chow rings we have homomor-

phisms $f^*: A(X) \to A(Y)$ and $f_*: A(Y) \to A(X)$, defined by taking inverse and direct images of cycles. With this notation, the theorem asserts that if X and Y are quasiprojective and $f: Y \to X$ is a proper morphism, then $f_*(ch(\xi)T(Y)) = ch(f_!(\xi))T(X)$. This is called **Grothendieck's theorem of R. R. type**. If X consists of a single point, the theorem gives Hirzebruch's theorem for algebraic bundles. Since algebraic and analytic theories of coherent sheaves on a complex projective space are isomorphic, this result covers Hirzebruch's theorem (→ 237 K-Theory).

The subgroup $R(Y)$ of $A(Y)$ given by $R(Y) = \{ch(\xi) \cdot T(Y) \mid \xi \in K(Y)\}$ is called the **Riemann-Roch group** of Y. Thus, using the notions developed by Grothendieck, the R. R. theorem can be expressed as follows: $R(Y)$ is mapped into $R(X)$ by a proper morphism $f: Y \to X$. Generalizations to †almost complex manifolds and †differentiable manifolds were made by Atiyah and Hirzebruch in this latter form [1]. One of the remarkable results is that an element of $R(Y)$ takes an integral value at the fundamental cycle. This theorem is obtained by taking X to be a single point.

E. R. R. Theorem for Singular Varieties

Let X be a projective variety over $\mathbf{C}$ and let $H_.(X)$ (resp. $H^{\cdot}(X)$) denote the singular homology (resp. cohomology) group with rational coefficients. Note that $K(X)$ may not agree with $K'(X)$ when X is singular. The R. R. theorem for X formulated by P. Baum, W. Fulton, and R. MacPherson [6] says that there exists a unique natural transformation $\tau: K'(X) \to H_.(X)$ such that (1) if $\xi \in K(X)$ and $\eta \in K'(X)$, then $\xi \otimes \eta \in K'(X)$ and $\tau(\xi \otimes \eta) = ch(\xi)(\tau(\eta))$; (2) whenever X is nonsingular, $\tau(\mathcal{O}_X) = T(X)(X)$. Note that the naturality of τ means that for any $f: X \to Y$ and any $\eta \in K'(X)$, $f_*\tau(\eta) = \tau(f_*\eta)$.

References

[1] M. F. Atiyah and F. Hirzebruch, Riemann-Roch theorems for differentiable manifolds, Bull. Amer. Math. Soc., 65 (1959), 276–281.
[2] M. F. Atiyah, R. Bott, and V. K. Patodi, On the heat equation and the index theorem, Inventiones Math., 19 (1973), 279–330.
[3] M. F. Atiyah, V. K. Patodi, and I. M. Singer, Spectral asymmetry and Riemann geometry, Bull. Lond. Math. Soc., 5 (1973), 229–234.
[4] M. F. Atiyah, G. B. Segal, and I. M. Singer, The index of elliptic operators, Ann. Math., (2) 87 (1968), 484–604 (I by Atiyah and Singer, II by Atiyah and Segal, III by Atiyah and Singer).
[5] M. F. Atiyah and I. M. Singer, The index of elliptic operators on compact manifolds, Bull. Amer. Math. Soc., 69 (1963), 422–433.
[6] P. Baum, W. Fulton, and R. MacPherson, Riemann-Roch for singular varieties, Publ. Math. Inst. HES, no. 45 (1975), 101–145.
[7] P. Berthelot, A. Grothendieck, L. Illusie, et al., Théorie des intersections et théorème de Riemann-Roch, SGA 6, Lecture notes in math. 225, Springer, 1971.
[8] A. Borel and J.-P. Serre, La théorème de Riemann-Roch, Bull Soc. Math. France, 86 (1958), 97–136.
[9] W. Fulton, Rational equivalence on singular varieties, Publ. Math. Inst. HES, no. 45 (1975), 147–167.
[10] W. Fulton and R. MacPherson, Intersecting cycles on an algebraic variety, Real and Complex Singularities, Sijthoff & Noordhoff, 1976, 179–197.
[11] F. Hirzebruch, Topological methods in algebraic geometry, Springer, third edition, 1966.
[12] K. Kodaira, On compact analytic surfaces I, Ann. Math., (2) 71 (1961), 111–152.
[13] R. MacPherson, Chern classes of singular varieties, Ann. Math., (2) 100 (1974), 423–432.
[14] B. Segre, Nuovi metodi e risultati nella geometria sulle varieta algebriche, Ann. Mat. (4) 35 (1953), 1–128.
[15] J. Todd, The geometrical invariants of algebraic loci, Proc. London Math. Soc. (2) 43 (1937), 127–138.

367 (XI.12) Riemann Surfaces

A. General Remarks

Riemann considered certain surfaces, now named after him, obtained by modifying in a suitable manner the domains of definition of multiple-valued †analytic functions on the complex plane in order to obtain single-valued functions defined on the surfaces. For example, consider the function $z = f(w) = w^2$, where w varies in the complex plane, and its inverse function $w = g(z) = \sqrt{z}$. Then $g(0) = 0$ and $g(\infty) = \infty$, whereas if $z \neq 0, \infty$, there exist two values of w satisfying $g(z) = w$. By setting $z = re^{i\theta}$ $(r > 0, 0 \leqslant \theta < 2\pi)$, the corresponding two values of w are $w_1 = \sqrt{r}\, e^{(\theta/2)i}$ and $w_2 = \sqrt{r}\, e^{(\theta/2 + \pi)i}$. Now consider how we should modify the complex z-plane so that we can obtain a single-valued function on the modified surface representing the same relationship

between z and w. Let π_1 and π_2 be two copies of the complex plane. Delete the nonnegative real axes from π_1 and π_2 and patch them crosswise along the slits (Fig. 1). The surface R thus obtained is locally homeomorphic to the complex plane except for the origin and the point at infinity, and situations at the origin and the point at infinity are as indicated in Fig. 1. For $z \neq 0, \infty$, there are two points z_1 and z_2 in π_1 and π_2, respectively, with the same coordinate z. Let w_1 and w_2 correspond to z_1 and z_2, respectively. Then the function $w=\sqrt{z}$ becomes single-valued on R, and w_1 and w_2 are holomorphic functions of z_1 and z_2, respectively. The surface R is called the Riemann surface determined by $w=\sqrt{z}$.

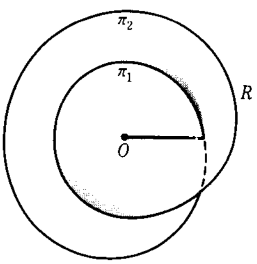


Fig. 1

Working from the idea illustrated by this example, H. Weyl and T. Radó gave rigorous definitions of Riemann surfaces. The usual definition nowadays is as follows: Let $\mathfrak{A}$ be a set of pairs (U, ψ_U) of open sets U in a †connected †Hausdorff space R and topological mappings ψ_U of U onto plane regions satisfying the following two conditions: (i) $R=\bigcup_{(U,\psi_U)\in\mathfrak{A}} U$; (ii) for each (U_1, ψ_{U_1}), $(U_2, \psi_{U_2}) \in \mathfrak{A}$ with $V= U_1 \cap U_2 \neq \varnothing$, $\psi_{U_1} \circ \psi_{U_2}^{-1}$ gives an (orientation-preserving) †conformal mapping of each connected component of $\psi_{U_2}(V)$ onto a corresponding one of $\psi_{U_1}(V)$. Two such sets $\mathfrak{A}_1$ and $\mathfrak{A}_2$ are equivalent, by definition, if $\mathfrak{A}_1 \cup \mathfrak{A}_2$ also satisfies conditions (i) and (ii). The equivalence class $(\mathfrak{A})$ of such $\mathfrak{A}$ is called a **conformal structure (analytic structure** or **complex structure)** on R (→ 72 Complex Manifolds). A pair $(R, (\mathfrak{A}))$ consisting of a connected Hausdorff space R and a conformal structure $(\mathfrak{A})$ is called a **Riemann surface**, with R its **base space** and $(\mathfrak{A})$ its **conformal structure.** (A Riemann surface in this sense is sometimes called an **abstract Riemann surface.**) It is a complex manifold of †complex dimension 1. For (U, ψ_U) in $\mathfrak{A} \in (\mathfrak{A})$, (U, ψ_U) (or sometimes U itself) is called an **analytic neighborhood**, and ψ_U is called a **local uniformizing parameter** (or simply a **local parameter**). In the remainder of this article we call R itself a Riemann surface.

From condition (i) it follows that a Riemann surface R is a real 2-dimensional †topological manifold. Moreover, by condition (ii) it can be deduced that R satisfies the †second countability axiom and consequently is a †surface and admits a †simplicial decomposition (T. Radó, 1925). It is also orientable (→ 410 Surfaces). Therefore R is a †locally compact metric space. It is not possible to define curve lengths on R compatible with the conformal structure, but since angles can be defined, R is considered to be a real 2-dimensional space with a †conformal connection. It is customary in the theory of functions to call R **closed** if it is compact and **open** otherwise. A plane region D is considered an open Riemann surface with the conformal structure $\mathfrak{A}=(D, 1: D \to D)$. A Riemann sphere is also considered to be a closed Riemann surface whose analytic neighborhoods are given by $\{U_1, \varphi\}$ and $\{U_2, 1/\varphi\}$, where $U_1(U_2)$ is the domain corresponding to $\{|z|<2\}(\{|z|>1/2\} \cup \{\infty\})$ under the stereographic projection φ.

A function f on a Riemann surface is said to be meromorphic, holomorphic, or harmonic on R if $f \circ \psi_U^{-1}$ is †meromorphic, †holomorphic, or †harmonic in the usual sense on $\psi_U(U)$ for every analytic neighborhood (U, ψ_U). More generally, suppose that for mappings between plane regions we are given a property $\mathfrak{P}$ that is invariant under conformal mappings. A mapping T of a Riemann surface R_1 onto another Riemann surface R_2 is said to have the property $\mathfrak{P}$ if the mapping $\psi_{U_2} \circ T \circ \psi_{U_1}^{-1}$ of $\psi_{U_1}(U_1)$ into $\psi_{U_2}(U_2)$ has the property $\mathfrak{P}$ for every pair of analytic neighborhoods (U_1, ψ_{U_1}) nd (U_2, ψ_{U_2}) in R_1 and R_2, respectively. Thus such a mapping T may be conformal, †analytic, †quasiconformal, harmonic, etc. If there exists a one-to-one conformal mapping of a Riemann surface R_1 onto another Riemann surface R_2, then R_1 and R_2 are said to be **conformally equivalent**. Two such Riemann surfaces are sometimes identified with each other.

B. Covering Surfaces

One of the main themes of the theory of functions is the study of analytic mappings of a Riemann surface R into another Riemann surface R_0, i.e., the theory of covering surfaces of Riemann surfaces.

Suppose, in general, that there are two surfaces R and R_0 and a continuous open mapping T of R into R_0 such that the inverse image of a point in R_0 under T is an †isolated set in R. Then T is called an **inner transformation** in the sense of Stoïlow and $(R, R_0; T)$ a **covering surface** with R_0 its **basic surface** and T its **projection.** Often R is called simply a covering surface of R_0. A point p_0 with $p_0=$

$T(p)$ is called the projection of p, and p is said to lie above p_0. In this case, there exist surface coordinates (U, ψ) and (U_0, ψ_0) at p and p_0, respectively, such that $\psi(U)=\{z \mid |z|<1\}$, $\psi(p)=0$, $\psi_0(U_0)=\{w \mid |w|<1\}$, $\psi_0(p_0)=0$, and $w=(\psi_0 \circ T \circ \psi^{-1})(z)=z^n$, with the positive integer n independent of the choice of coordinate neighborhoods. If $n>1$, then p is called a **branch point**, n the **multiplicity**, and $n-1$ the **degree of ramification**. The set of all branch points forms an at most countably infinite set of isolated points. If there is no branch point, then the covering surface $(R, R_0; T)$ and the projection T are said to be **unramified**. For a given curve C_0 in R_0 and a point p in R lying above the initial point of C_0, a curve C in R with p its initial point satisfying $T(C)=C_0$ is called the **prolongation** along C_0 (or the **lift** of C_0) starting from p. If any proper subarc of C_0 sharing the initial point with C_0 admits a prolongation along itself starting from p but C_0 does not admit a prolongation along itself starting from p, then R is said to have a **relative boundary** above the endpoint of C_0. A covering surface without a relative boundary is called **unbounded**. A †simply connected surface R^∞ that is an unramified unbounded covering surface of R_0 is said to be a **universal covering surface** of R_0. The universal covering surface exists for every R.

Suppose that R is an unbounded covering surface of R_0. Then the number of points on R that lie above each point of R_0 is always constant, say n ($\leqslant +\infty$), where the branch points of R are counted with their multiplicities. n is called the **number of sheets** of R over R_0, or R is said to be ***n*-sheeted** over R_0. If R and R_0 are compact surfaces with †Euler characteristic χ and χ_0, respectively, and if R is an n-sheeted covering surface of R_0, then we have the **Riemann-Hurwitz relation**: $\chi=n\chi_0-V$, V being the sum of the degrees of ramification. A topological mapping S of an unramified unbounded covering surface R of R_0 onto itself such that $T \circ S=T$ is called a **covering transformation**. The group of all covering transformations of a universal covering surface of R_0 is isomorphic to the †fundamental group (i.e., the 1-dimensional homotopy group) of R_0.

In a covering surface $(R, R_0; T)$ whose basic surface R_0 is a Riemann surface, T can be regarded as an analytic mapping of R onto R_0 by giving R a conformal structure in a natural manner. In particular, if R_0 is the sphere, then its covering surface is a Riemann surface. Conversely, any Riemann surface can be regarded as a covering surface of the sphere. This fact had long been known for closed Riemann surfaces; for open Riemann surfaces, it can be deduced from the existence theorem of analytic functions proved by H. Behnke and K. Stein, which states that an open Riemann surface is a †Stein manifold. Historically, by a Riemann surface mathematicians meant either the abstract Riemann surface or a covering surface of the sphere, until the two notions were proved identical.

Suppose that R is a covering surface of the z-sphere R_0 with the projection $T: R \rightarrow R_0$, and denote by R_a the region that lies above $0<|z-a|<r_0$. If there exists a topological mapping ψ of R_a onto $\{(r, \theta) \mid 0<r<r_0, -\infty<\theta<\infty\}$ such that $a+re^{i\theta}=T(\psi^{-1}(r, \theta))$, then R is said to have a **logarithmic branch point** above a; in contrast, a branch point of multiplicity n of the type defined previously is sometimes called an **algebraic branch point**.

Ahlfors's theory of covering surfaces, which treats covering surfaces $(R, R_0; T)$ not only from the topological viewpoint but also from the metrical one, is particularly important. Let R and R_0 be either compact surfaces with simplicial decompositions or their closed subregions with boundaries consisting of 1-simplexes and vertices such that T preserves simplicial decompositions. Here we call the part of the boundary of R whose projection is in the interior of R_0 the relative boundary. With respect to a suitable metric on R_0, let S be the ratio of areas of R and R_0, L the length of the relative boundary, and $-\rho$ and $-\rho_0$ the †Euler characteristics of R and R_0, respectively. Then **Ahlfors's principal theorem** asserts that $\max(0, \rho) \geqslant \rho_0 S - hL$, where $h>0$ is a constant determined only by R_0. This has been applied widely in various branches of mathematics, including the theory of distribution of values of analytic mappings between Riemann surfaces.

C. Uniformization

Suppose that we are given a correspondence between the z-plane and w-plane determined by a †function element $p_0=(z_0^*, w_0^*)$ $(z_0^*=P_0(t)$, $w_0^*=Q_0(t))$. This correspondence generally gives rise to a multiple-valued analytic function $w=f(z)$. We show how to construct a Riemann surface so that the function $w=f(z)$ can be considered a single-valued function on it. We use f again to mean the connected component of the set of function elements $p=(z^*, w^*)$ $(z^*=P(t), w^*=Q(t))$ in the wider sense containing p_0, where the analytic neighborhood of each point p is defined to be the set of elements that are direct analytic continuations of p. Then f is a Riemann surface. For a point $p=(z^*, w^*)$ $(z^*=P(t), w^*=Q(t))$ in f, set $z=F(p)=P(0)$, $w=G(p)=Q(0)$. Then two meromorphic functions $z=F(p)$ and $w=G(p)$

are defined on f, and f can be considered as a covering surface of the z-sphere and the w-sphere. We call f an †analytic function in the wider sense. Thus we obtain a single-valued function $w=G(p)$ on the Riemann surface f that can be regarded as a modification of the original function $w=f(z)$. Suppose that there exist two meromorphic functions $z=\varphi(\zeta)$ and $w=\psi(\zeta)$ on a region D in the ζ-plane, and let $z=P(\zeta-\zeta_0)$ and $w=Q(\zeta-\zeta_0)$ be †Laurent expansions of φ and ψ at each point ζ_0 in D. If the function element $p_{\zeta_0}=(z^*, w^*)$ $(z^*=P(t), w^*=Q(t))$ belongs to the Riemann surface f, then the correspondence $w=f(z)$ determined by the function element p_{ζ_0} is said to be **locally uniformized** on D by $z=\varphi(\zeta)$ and $w=\psi(\zeta)$. In particular, if $\{p_\zeta \mid \zeta\in D\}=f$, then f is said to be **uniformized** by $z=\varphi(\zeta)$ and $w=\psi(\zeta)$. If an analytic function f in the wider sense, considered as a Riemann surface, is conformally equivalent to a region D in the ζ-plane, then f can be uniformized by $z=F(p)$ and $w=G(p)$. In general, f is not conformally equivalent to a plane region, but if an unramified unbounded covering surface $(\hat{f}, f; T)$ of f is conformally equivalent to a region D in the ζ-plane, then f is uniformized by $z=F\circ T(\zeta)$ and $w=G\circ T(\zeta)$ (**Schottky's uniformization**). In particular, since the universal covering surface $(f^\infty, f; T)$ of f is simply connected, it is conformally equivalent to the sphere $|\zeta|\leqslant\infty$, the finite plane $|\zeta|<\infty$, or the unit disk $|\zeta|<1$. Consequently, f is uniformized by $z=F\circ T(\zeta)$ and $w=G\circ T(\zeta)$. Therefore analytic functions in the wider sense are always uniformizable.

For example, an †algebraic function f considered as a Riemann surface is always closed. If its †genus $g=0$, then f is the sphere and is thus uniformized by rational functions $z=F(\zeta)$ and $w=G(\zeta)$. If $g>0$, then $(f^\infty, f; T)$ is conformally equivalent to $|\zeta|<1$ or $|\zeta|<\infty$, and hence f is uniformized by $z=F\circ T(\zeta)$ and $w=G\circ T(\zeta)$. When $|\zeta|<1$, z and w are †automorphic functions with respect to the group of linear transformations preserving $|\zeta|<1$, while if $|\zeta|<\infty$, they are †elliptic functions.

D. The Type Problem

A simply connected Riemann surface R is conformally equivalent to the sphere, the finite plane, or the unit disk. Then R is said to be **elliptic**, **parabolic**, or **hyperbolic**, respectively. The problem of determining the types of simply connected covering surfaces of the sphere by their structures, such as the distributions of their branch points, is called the **type problem** for Riemann surfaces. For example, if a simply connected covering surface does not cover three points on the sphere, it must be hyperbolic (Picard's theorem). The Nevanlinna theory of meromorphic functions stimulated this type problem. However, it is difficult to measure the ramifications of covering surfaces, and many detailed reuslts of the type problem obtained in the 1930s are limited mainly to the case where all branch points lie above a finite number of points on the sphere. A sufficient condition for R to be of parabolic type, given by Z. Kobayashi (using the so-called Kobayashi net, and a sufficient condition for R to be of hyperbolic type, given by S. Kakutani (using quasiconformal mappings), are significant results on the type problem. The type problem had by then been extensively generalized to the following classification theory.

E. Classification Theory of Riemann Surfaces

Riemann surfaces are, as pointed out by Weyl, "not merely devices for visualizing the many-valuedness of analytic functions, but rather an essential component of the theory ... the only soil in which the functions grow and thrive." So the problem naturally arises of how to extend various results in the theory of analytic functions of a complex variable to the theory of analytic mappings between Riemann surfaces. In general, open Riemann surfaces can have infinite genus and are quite complicated. So to obtain fruitful results and systematic development, one usually sets certain restrictions on the properties of the Riemann surfaces. In connection with this, R. Nevanlinna, L. Sario, and others initiated the **classification theory of Riemann surfaces**, which classifies Riemann surfaces by the existence (or nonexistence) of functions with certain properties.

Denote by $\mathfrak{X}(R)$ the totality of functions on a Riemann surface R with a certain property $\mathfrak{X}$. The set of all Riemann surfaces R for which $\mathfrak{X}(R)$ does not contain any function other than constants is denoted by $O_{\mathfrak{X}}$. The family of analytic functions and the family of harmonic functions are denoted by $A(R)$ and $H(R)$, respectively. The family of positive functions, that of bounded functions, and that of functions with finite Dirichlet integrals are denoted by $P(R)$, $B(R)$, and $D(R)$, respectively. From these families, various new families are created, e.g., $ABD(R)=A(R)\cap B(R)\cap D(R)$. Usually O_{HB}, O_{HD}, O_{HBD}, O_{AB}, O_{AD}, O_{ABD}, and also O_G, the family of Riemann surfaces on which there are no Green's functions, are considered. P. J. Myrberg found an example of a Riemann surface of infinite genus which has a large boundary but belongs to O_{AB}. The idea behind Myberg's example is often used to construct examples in classification theory. From works

of Y. Tôki, L. Sario, K. I. Virtanen, H. L. Royden, M. Parreau, M. Sakai, and others, it can be seen that there are inclusion relations, as indicated in Fig. 2, among the classes just mentioned. There is no inclusion relation between O_{AB} and O_{HD}. For Riemann surfaces of finite genus, $O_G = O_{HD}$. Closed Riemann surfaces are all in O_G.

Open Riemann surfaces in O_G are also said to be **parabolic** (or **of null boundary**), and those not in O_G, **hyperbolic** (or **of positive boundary**). Several characterizations for parabolic Riemann surfaces are known.

$$O_G \subsetneqq O_{HP} \subsetneqq O_{HB} \begin{array}{c} \subsetneqq O_{HD} = O_{HBD} \subsetneqq \\ \subsetneqq O_{AB} \subsetneqq \end{array} O_{AD} = O_{ABD}$$

Fig. 2

From a similar point of view, the classification problem for subregions was studied in detail by Parreau, A. Mori, T. Kuroda, and others. We call a noncompact region Ω which is the complement of a compact subset of a Riemann surface a **Heins's end**. M. Heins called the minimal number ($\leqslant \infty$) of generators of the semigroup of the additive class of HP-functions that vanish continuously at the relative boundary of a Heins's end Ω the **harmonic dimension** of Ω. Its properties were investigated by Z. Kuramochi, M. Ozawa, and others. Generally, a function f is said to be $\mathfrak{X}$-**minimal** if f is positive and contained in $\mathfrak{X}(R)$ and there exists a constant C_g for every g in $\mathfrak{X}(R)$ with $f \geqslant g \geqslant 0$ such that $g \equiv C_g f$. The family of Riemann surfaces R not belonging to O_G and admitting $\mathfrak{X}$-minimal functions on R is denoted by $U_{\mathfrak{X}}$. C. Constantinescu and A. Cornea and others studied Riemann surfaces in U_{HB} and $U_{\underline{HD}}$ where $\underline{HD}$ is the class of positive functions in HD or limits of monotone decreasing sequences of such functions. There are inclusion relations $U_{HB} \supsetneqq O_{HB} - O_G$, $U_{\underline{HD}} \supsetneqq O_{HD} - O_G$, and $U_{\underline{HD}} \subsetneqq U_{UB}$. One of the interesting results in classification theory is the fact, discovered by Kuramochi, that $U_{HB} \cup O_G \subsetneqq O_{AB}$ and $U_{\underline{HD}} \cup O_G \subsetneqq O_{AD}$.

Classification theory has a very deep connection with the theory of †ideal boundaries. A. Pfluger and Royden showed that the classes O_G and O_{HD} are invariant under quasiconformal mappings. However, it is still an open problem whether O_{HB} is invariant in this sense.

F. Prolongations of Riemann Surfaces

As classification theory shows, pathological phenomena occur for Riemann surfaces from the viewpoint of function theory in the plane. These are caused by the complexity of ideal boundaries of Riemann surfaces, and in particular by the complexity of the set of ideal boundary points at which handles of Riemann surfaces (i.e., parts with cycles not homologous to 0) accumulate. Hence it is desirable to find larger Riemann surfaces so that ideal boundary points of original surfaces that are not accumulating points of handles become interior points. Suppose that a Riemann surface R is conformally equivalent to a proper subregion R' of another Riemann surface R_1. Then R_1 is said to be a **prolongation** of R, and R is **prolongable**. A nonprolongable Riemann surface is said to be **maximal**. Closed Riemann surfaces are always maximal, but there also exist maximal open Riemann surfaces (Radó). However, every open Riemann surface is homeomorphic to a prolongable Riemann surface (S. Bochner). Characterizations of prolongable Riemann surfaces and relationships between the various null classes mentioned in Section E and prolongations were investigated from several viewpoints by R. de Possel, J. Tamura, and others.

G. Analytic Mappings of Riemann Surfaces

Apart from the development of the classification theory of Riemann surfaces, efforts have been made to extend various results in the theory of analytic functions of a complex variable to the case of analytic mappings between Riemann surfaces. L. Sario studied the method of **normal operators**, which is utilized to construct harmonic functions on Riemann surfaces with given singularities at their ideal boundaries; and he extended the main theorems of Nevanlinna to analytic mappings between arbitrary Riemann surfaces (→ 124 Distributions of Values of Functions of a Complex Variable). M. Heins introduced the notions of Lindelöf type, Blaschke type, and others which are special classes of analytic mappings. Utilizing these notions, Constantinescu and Cornea, Kuramochi, and others extended various results in the theory of cluster sets (→ 62 Cluster Sets) by studying the behavior of analytic mappings at ideal boundaries. The theory of †capacities on ideal boundaries has also been developed.

On every open Riemann surface R there exists a nonconstant holomorphic function (Behnke and Stein). Furthermore, Gunning and Narasimhan proved that there exists a holomorphic function on R whose derivative never vanishes. In other words, R is conformally equivalent to an unramified covering surface of the sphere. Such a locally homeomorphic analytic mapping is called the **immersion** of R. The proof is based on the fol-

lowing deep result (S. Mergelyan's theorem): Suppose that K is a compact set on R such that $R-K$ has no relatively compact connected component. Then every continuous function which is holomorphic on the interior of K can be approximated uniformly on K by a holomorphic function on R [22].

A surface of genus 0 is said to be **planar** (or **of planar character** or **schlichtartig**). A simply connected surface is planar. Using the Dirichlet principle, P. Koebe proved the following **general uniformization theorem**: Every planar Riemann surface R can be mapped conformally to the canonical slit regions on the extended complex plane **C**. More precisely, given a point p on R, there exists the **extremal horizontal (vertical) slit mapping** $F_1(F_2)$ such that (i) $F_1(F_2)$ maps R conformally to a region on **C** whose boundary consists of horizontal (vertical) slits and possibly points; (ii) F_1 and F_2 have a simple pole at p with residue 1; (iii) the total area of the slits and points is zero. Suppose that $F_i=1/z+a_i+c_iz+\ldots$ $(i=1,2)$ in terms of local parameter z at p. Then $s=(c_1-c_2)/2$ is called the **span** of R. It is known that $s\,(=\|d(F_1-F_2)/2\|^2)\geqslant 0$, where the equality holds if and only if R belongs to O_{AD}. In the case of finite genus g, there also exist the conformal mappings of R onto the parallel slit regions on the $(g+1)$-sheeted covering surface of **C** (Z. Nehari, Y. Kusunoki, and others). L. Ahlfors proved that a Riemann surface of genus g bounded by m contours can be mapped conformally to an at most $(2g+m)$-sheeted unbounded covering surface of the unit disk.

The structures of closed Riemann surfaces are determined by the algebraic structures of meromorphic function fields on them. H. Iss'sa obtained a noteworthy result which established that open Riemann surfaces are also determined by their meromorphic function fields [24]. It is known too that open Riemann surfaces are determined by their rings of holomorphic functions.

H. Differential Forms on Riemann Surfaces

Since Riemann surfaces are considered real 2-dimensional differentiable manifolds of class C^∞, differential 1-forms $\omega=u\,dx+v\,dy$ and differential 2-forms $\Omega=c\,dx\wedge dy$ are defined on them, and operations such as the exterior derivative $d\omega=(\partial v/\partial x-\partial u/\partial y)\,dx\wedge dy$ and exterior product can be defined (→ 105 Differentiable Manifolds). Since coordinate transformations satisfy the Cauchy-Riemann differential equation, the **conjugate differential** $*\omega=-v\,dx+u\,dy$ of ω can be defined. A differential form ω satisfying $d\omega=d*\omega=0$ is called a **harmonic differential**, and one with $*\omega=-i\omega$ is said to be **pure**. A pure differential is expressed as $\omega=f\,dz$. Here, if f is a holomorphic function, then ω is called a **holomorphic** (or **analytic**) **differential**, and if f is a meromorphic function, then ω is called a **meromorphic** (or **Abelian**) **differential**. The differential form ω is a holomorphic differential if and only if it is **closed** (i.e., $d\omega=0$) and pure. The differential ω is called **exact** (or **total**) if ω is written as $dF=F_x\,dx+F_y\,dy$ with a globally single-valued function F.

Next, the set of all measurable differentials ω with $\|\omega\|^2=\iint_R\omega\wedge\overline{*\omega}<\infty$ forms a †Hilbert space with respect to the norm $\|\omega\|$. The method of orthogonal decomposition in the theory of Hilbert spaces is the main device to study this space and also its suitable subspaces and to obtain the existence theorem of harmonic and holomorphic differentials with various properties (→ 194 Harmonic Integrals). However, in contrast to differentiable manifolds, it should be noted that finer orthogonal decompositions into subspaces with specific properties hold for open Riemann surfaces. For instance, let $\Gamma_a(\Gamma_h)$ be the Hilbert space of analytic (harmonic) differentials with finite norm, and set $\Gamma_{ae}=\{\omega\in\Gamma_a\mid\omega\text{ is exact}\}$, $\Gamma_{hse}=\{\omega\in\Gamma_h\mid\omega\text{ is †semiexact}\}$; then we have the orthogonal decompositions

$$\Gamma_a=\Gamma_{ae}\dot{+}\Gamma_{aS},$$

$$\Gamma_h={}^*\Gamma_{hse}\dot{+}\Gamma_{hm}=\Gamma_{hse}\dot{+}{}^*\Gamma_{hm},\quad\text{etc.},$$

where ${}^*\Gamma_x$ stands for the space $\{\omega\mid{}^*\omega\in\Gamma_x\}$ and Γ_{as} and Γ_{hm} are known as the space of analytic Schottky differentials and the space of differentials of harmonic measures, respectively. Both spaces Γ_{aS} and Γ_{hm} have remarkable properties [9].

I. Abelian Integrals on Open Riemann Surfaces

The systematic effort to extend the theory of Abelian integrals on closed Riemann surfaces to open Riemann surfaces was initiated by Nevanlinna in 1940. At the first stage of the development, only those Riemann surfaces with small boundaries (i.e., ones in O_G or O_{HD}) were treated; later a more general treatment was made possible by the discovery of the notion of semiexact differentials (K. Virtanen, Ahlfors).

Let R be an arbitrary open Riemann surface. A 1-dimensional cycle C is called a **dividing cycle** if for any compact set in R there exists a cycle outside the compact set homologous to C. A differential is said to be **semiexact** if its period along every dividing cycle vanishes.

Ahlfors defined the distinguished (complex harmonic) differentials with polar singularities and obtained in terms of them a generalization of Abel's theorem. Independently, Y. Kusunoki defined the semiexact canonical (meromorphic) differentials and gave in terms of them a formulation of Abel's theorem and the Riemann-Roch theorem on R [27]. It was proved that a meromorphic differential $df = du + idv$ is semiexact canonical if and only if du is (real) distinguished, and then u is almost constant on every ideal boundary component of R (in appropriate †compactification of R). Hence every meromorphic function f for which df is distinguished is almost constant on every ideal boundary component, and therefore f reduces to a constant by the boundary theorem of Riesz type if R has a large boundary. Whereas by the Riemann-Roch theorem above a nonconstant meromorphic function f such that df is (exact) canonical exists on any open Riemann surface R with finite genus, and f gives a canonical parallel slit mapping of R (→ Section G). H. L. Royden [28] and B. Rodin also gave generalizations of the Riemann-Roch theorem. M. Yoshida, H. Mizumoto, M. Shiba, and others further generalized the Kusunoki type theorems. The Riemann-Roch theorem for a closed Riemann surface can be deduced from that for open Riemann surfaces by considering an open Riemann surface obtained from a closed surface by deleting a point.

Riemann's period relation on R has been studied for various classes of differentials, but for the case of infinitely many nonvanishing periods, no definitive result has been obtained. The same is true for the theory of Abelian differentials with infinitely many singularities. On the other hand, the analogy to the classical theory is lost completely if no restriction is posed on the differentials on R. In this context the following results due to Behnke and Stein [26] are outstanding: (1) There exists an Abelian differential of the first kind on R with infinitely many given periods. (2) For two discrete sequences $\{p_n\}$ and $\{q_n\}$ of points in R, there exists a single-valued meromorphic function with zeros at p_n and poles at q_m. It is further proved that there exists an Abelian differential with prescribed divisor and periods (Kusunoki and Sainouchi). This generalizes the results above and the Gunning-Narasimhan theorem.

References

[1] H. Weyl, Die Idee der Riemannschen Fläche, Teubner, 1913, third edition, 1955; English translation, The concept of a Riemann surface, Addison-Wesley, 1964.
[2] S. Stoïlow, Leçons sur les principes topologiques de la théorie des fonctions analytiques, Gauthier-Villars, 1938.
[3] L. Ahlfors, Zur Theorie der Überlagerungsflächen, Acta Math., 65 (1935), 157–194.
[4] R. Nevanlinna, Uniformisierung, Springer, 1953.
[5] H. Behnke and F. Sommer, Theorie der analytischen Funktionen einer komplexen Veränderlichen, Springer, second revised edition, 1962.
[6] A. Pfluger, Theorie der Riemannschen Flächen, Springer, 1957.
[7] G. Springer, Introduction to Riemann surfaces, Addison-Wesley, 1957.
[8] L. V. Ahlfors (ed.), Contribution to the theory of Riemann surfaces, Ann. Math. Studies, Princeton Univ. Press, 1953.
[9] L. V. Ahlfors and L. Sario, Riemann surfaces, Princeton Univ. Press, 1960.
[10] M. Schiffer and D. C. Spencer, Functionals of finite Riemann surfaces, Princeton Univ. Press, 1954.
[11] L. Bers, Riemann surfaces, Lecture notes, New York Univ., 1957.
[12] M. Tsuji, Potential theory in modern function theory, Maruzen, 1959 (Chelsea, 1975).
[13] R. C. Gunning, Lectures on Riemann surfaces, Princeton Univ. Press, 1966.
For the classification of Riemann surfaces → [6, 9] and the following:
[14] C. Constantinescu and A. Cornea, Ideale Ränder Riemannscher Flächen, Springer, 1963.
[15] L. Sario and M. Nakai, Classification theory of Riemann surfaces, Springer, 1970.
[16] M. Sakai, Analytic functions with finite Dirichlet integrals on Riemann surfaces, Acta Math., 142 (1979), 199–220.
For analytic mappings of Riemann surfaces → [4, 6, 9] and the following:
[17] B. Rodin and L. Sario, Principal functions, Van Nostrand, 1968.
[18] L. Sario and K. Noshiro, Value distribution theory, Van Nostrand, 1966.
[19] L. Sario and K. Oikawa, Capacity functions, Springer, 1969.
[20] L. Ahlfors, Open Riemann surfaces and extremal problems on compact subregions, Comm. Math. Helv., 24 (1950), 100–134.
[21] R. Gunning and R. Narasimhan, Immersion of open Riemann surfaces, Math. Ann., 174 (1967), 103–108.
[22] E. Bishop, Subalgebras of functions on a Riemann surface, Pacif. J. Math., 8 (1958), 29–50.
[23] M. H. Heins, Algebraic structure and conformal mapping, Trans. Amer. Math. Soc., 89 (1958), 267–276.

[24] H. Iss'sa, On the meromorphic function field of a Stein variety, Ann. Math., (2) 83 (1966), 34–46.
For the generalization of algebraic functions and Abelian integrals → [4, 6, 9] and the following:
[25] R. Nevanlinna, Quadratisch integrierbare Differentiale auf einer Riemannschen Mannigfaltigkeit, Ann. Acad. Sci. Fenn., 1 (1941), 1–34.
[26] H. Behnke and K. Stein, Entwicklung analytischer Functionen auf Riemannschen Flächen, Math. Ann., 120 (1949), 430–461.
[27] Y. Kusunoki, Theory of Abelian integrals and its applications to conformal mappings, Mem. Coll. Sci. Univ. Kyôto, (A, Math.) 32 (1959), 235–258; 33 (1961), 429–433.
[28] H. L. Royden, The Riemann-Roch theorem, Comment. Math. Helv., 34 (1960), 37–51.
Also → reference to 11 Algebraic Functions, 416 Teichmüller Spaces.

368 (III.9) Rings

A. Definition

A nonempty set A is called a **ring** if the following conditions are satisfied.

(1) Two †operations, called **addition** and **multiplication** (the **ring operations**), are defined, which send an arbitrary pair of elements a, b of A to elements $a+b$ and ab of A.

(2) For arbitrary elements a, b, c of A, these operations satisfy the following four laws: (i) $a+b=b+a$ (**commutative law of addition**); (ii) $(a+b)+c=a+(b+c)$, $(ab)c=a(bc)$ (**associative laws**); (iii) $a(b+c)=ab+ac$, $(a+b)c=ac+bc$ (**distributive laws**); and (iv) for every pair a, b of elements of A, there exists a unique element c of A such that $a+c=b$. Thus a ring A is an †Abelian group under addition. Each element a of a ring A determines operations L_a and R_a defined by $L_a(x)=ax$, $R_a(x)=xa$ $(x\in A)$. Thus a ring A has the structure of a †left A-module and a †right A-module. Since the operations L_a and R_b commute for every pair a, b of elements of A, the ring A is also an A-A-bimodule (→ 277 Modules).

The identity element of A under addition is called the **zero element** and is denoted by 0. It satisfies the equation $a0=0a=0$ $(a\in A)$. An element $e\in A$ is called a **unity element (identity element** or **unit element)** of A if it satisfies $ae=ea=a$ $(a\in A)$. If A has such a unity element, it is unique and is often denoted by 1. A ring with unity element is called a **unitary ring**. Most of the important examples of rings are unitary. Hence we often call a unitary ring simply a ring. If a ring has only one member (namely, 0), then 0 is the unity element of the ring. Such a ring is called a **zero ring**. However, if a ring has more than one element, the unity element is distinct from the zero element. A ring is called a **commutative ring** if it satisfies (v) $ab=ba$ $(a,b\in A)$ (**commutative law for multiplication**) (→ 67 Commutative Rings).

In this article we shall discuss associative rings. Certain nonassociative rings are important; an example is †Lie algebra. (An algebra is a ring having a †ground ring.)

B. Further Definitions

An element $a\neq 0$ of a ring A is called a **zero divisor** if there exists an element $b\neq 0$ such that $ab=0$ or $ba=0$. A commutative unitary ring having more than one element is called an **integral domain** if it has no zero divisors (→ 67 Commutative Rings). Elements a and b of a ring are said to be **orthogonal** if $ab=ba=0$. An element a satisfying $a^n=0$ for some positive integer n is called a **nilpotent element**, and a nonzero element a satisfying $a^2=a$ is called an **idempotent element**. An idempotent element is said to be **primitive** if it cannot be represented as the sum of two orthogonal idempotent elements. For any subsets S and T of a ring A, let $S+T(ST)$ denote the set of elements $s+t(st)$ $(s\in S, t\in T)$. In particular, SS is denoted by S^2 (similarly for S^3, S^4, etc.), and furthermore, $\{a\}+S(\{a\}S)$ is denoted by $a+S(aS)$. If $ST=TS=\{0\}$, then subsets S and T are said to be **orthogonal**. A subset S of a ring is said to be **nilpotent** if $S^n=0$ for some positive integer n, and **idempotent** if $S^2=S$.

For an element a of a unitary ring A, an element a' such that $a'a=1$ $(aa'=1)$ is called a **left (right) inverse element** of a. There exists a left (right) inverse element of a if and only if A is generated by a as a left (right) A-module. If there exist both a right inverse and a left inverse of a, then they coincide and are uniquely determined by a. This element is called the **inverse element** of a and is denoted by a^{-1}. An element that has an inverse element is called an **invertible element (regular element** or **unit)**. The set of all invertible elements of a unitary ring forms a group under multiplication. A nonzero unitary ring is called a †**skew field** (or †**division ring)** if every nonzero element is invertible. Furthermore, a skew field that satisfies the commutative law is called a **commutative field** or simply a field (→ 149 Fields). In a general ring A, if we define a new operation $(a,b)\to a\circ b$ by setting $a\circ b=a+b-ab$, then A is a †semigroup with the identity element 0 with respect to this operation. The inverse

element under this operation is called the **quasi-inverse element**, and an element that has a quasi-inverse element is called a **quasi-invertible element** (or **quasiregular element**). An element a of a unitary ring is quasi-invertible if and only if the element $1-a$ is invertible.

C. Examples

(1) Rings of numbers. The ring **Z** of rational integers, the rational number field **Q**, the real number field **R**, and the complex number field **C** are familiar examples (→ 14 Algebraic Number Fields, 257 Local Fields).

(2) Rings of functions. The set K^I of functions defined on a set I and taking values in a ring K forms a commutative ring under pointwise addition and multiplication. In particular, let $K=\mathbf{R}$, and let I be an interval of **R**. Then the set $C^0(I)$ of continuous functions, the set $C^r(I)$ of functions that are r-times continuously differentiable, and the set $C^\omega(I)$ of analytic functions on I are subrings (→ Section E) of $\mathbf{R}^I$.

(3) Rings of expressions. The set $K[X_1, \ldots, X_n]$ of polynomials and the set $K[[X_1, \ldots, X_n]]$ of †formal power series in indeterminates $X_1, \ldots, X_n$ with coefficients in a commutative ring K are commutative rings (→ 369 Rings of Polynomials, 370 Rings of Power Series).

(4) **Endomorphism rings** of modules. The set $\mathscr{E}_K(M)$ of †endomorphisms of a †module M over a ring K is in general a noncommutative ring. In particular, if M is a finite-dimensional †linear space over a field K, then $\mathscr{E}_K(M)$ can be identified with a †full matrix ring (→ 256 Linear Spaces, 277 Modules).

(5) For other examples → 29 Associative Algebras, 36 Banach Algebras, 67 Commutative Rings, 284 Noetherian Rings, and 439 Valuations.

D. Homomorphisms

A mapping $f: A \rightarrow B$ of a ring A into a ring B satisfying conditions (i) $f(a+b)=f(a)+f(b)$ and (ii) $f(ab)=f(a)f(b)$ $(a, b \in A)$ is called a **homomorphism**. If a homomorphism f is †bijective, then the inverse mapping $f^{-1}: B \rightarrow A$ is also a homomorphism, and in this case f is called an **isomorphism**. More precisely, a homomorphism (isomorphism) of rings is often called a **ring homomorphism (ring isomorphism)**. There exists only one homomorphism of any ring onto the zero ring. For unitary rings A and B, a homomorphism $f: A \rightarrow B$ is said to be **unitary** if it maps the unity element of A to the unity element of B. By a homomorphism, a unitary homomorphism is usually meant. In this sense there exists a unique homomorphism of the ring **Z** of rational integers into an arbitrary unitary ring. A composite of homomorphisms is also a homomorphism. The identity mapping 1_A of a ring A is an isomorphism. A homomorphism of a ring A into itself is called an **endomorphism**, and an isomorphism of A onto itself is called an **automorphism** of A. If a is an invertible element of a unitary ring A, then the mapping $x \rightarrow axa^{-1}$ $(x \in A)$ is an automorphism of A, called an **inner automorphism**.

When condition (ii) for a homomorphism is replaced by (ii') $f(ab)=f(b)f(a)$ $(a, b \in A)$, a mapping satisfying (i) and (ii') is called an **antihomomorphism**. In particular, if an antihomomorphism f is bijective, then the inverse mapping f^{-1} is also an antihomomorphism, and f is called an **anti-isomorphism**. **Antiendomorphisms** and **antiautomorphisms** are defined similarly.

E. Subrings, Factor Rings, and Direct Products

A subset S of a ring A is called a **subring** of A if a ring structure is given on S and the canonical injection $S \rightarrow A$ is a ring homomorphism. Thus the ring operations of S are the restrictions of those of A. If we deal only with unitary rings and unitary homomorphisms, then a subring S necessarily contains the unity element of A. The smallest subring containing a subset T of a ring A is called the subring **generated** by T. The set of elements that commute with every element of T forms a subring and is called the **commutor** (or **centralizer**) of T. In particular, the commutor of A itself is called the **center** of A.

A †quotient set A/R of a ring A by an equivalence relation R is called a factor ring (quotient ring) of A if a ring structure is given on A/R and the canonical surjection $A \rightarrow A/R$ is a ring homomorphism. This is the case if and only if the equivalence relation R is compatible with the ring operations (i.e., aRa' and bRb' imply $(a+b)R(a'+b')$ and $(ab)R(a'b')$). Let α and β be elements of the factor ring A/R, represented by a and b, respectively. Then the definition of factor ring implies that $\alpha+\beta(\alpha\beta)$ is the equivalence class represented by $a+b(ab)$. Every ring A has two trivial factor rings, namely, A itself and the zero ring 0. If A has no nontrivial factor rings, then A is called a **quasisimple ring** (→ Section G). If $f: A \rightarrow B$ is a ring homomorphism, then the image $f(A)$ is a subring of B, and the equivalence relation R in A defined by $f(aRb \Leftrightarrow f(a)=f(b))$ is compatible with the ring operations of A. Thus f induces an isomorphism $A/R \rightarrow f(A)$ (→ Section F).

If $\{A_i\}_{i \in I}$ is a family of rings, the Cartesian

product $A=\prod_{i\in I}A_i$ forms a ring under the componentwise operations $(a_i)+(b_i)=(a_i+b_i)$ and $(a_i)(b_i)=(a_ib_i)$. This ring is called the **direct product** of the family of rings $\{A_i\}_{i\in I}$. The mapping $p_i: A\rightarrow A_i$ that assigns to each (a_i) its ith component a_i is called a **canonical homomorphism**. For any set of homomorphisms $f_i: B\rightarrow A_i$ $(i\in I)$, there exists a unique homomorphism $f: B\rightarrow A$ such that $f_i=p_i\circ f$ for each i.

F. Ideals

A subset of a ring A is called a **left (right) ideal** of A if it is a submodule of the left (right) A-module of A (→ 277 Modules). In other words, a left (right) ideal J of A is an additive subgroup of A such that $AJ\subset J$ $(JA\subset J)$. Under the operations induced from A, J is a ring (however, J is not necessarily unitary). A subset of A is called a **two-sided ideal** or simply an **ideal** of A if it is a left and right ideal.

For an ideal J of a ring A, we define a relation R in A by $aRb \Leftrightarrow a-b\in J$. Then R is an equivalence relation that is compatible with the operations of A. Each equivalence class is called a **residue class** modulo J, and the quotient ring A/R is denoted by A/J and called the **residue (class) ring** (or **factor ring**) **modulo** J. If it is a field, it is called a **residue (class) field.** Conversely, given an equivalence relation R that is compatible with the operations of A, the equivalence class of 0 forms an ideal J of A, and the equivalence relation defined by J coincides with R.

If $f: A\rightarrow B$ is a ring homomorphism, then the †kernel of f as a homomorphism of additive groups forms an ideal J of A, and f induces an isomorphism $A/J\rightarrow f(A)$. It S is a subring and J is an ideal of a ring A, then $S+J$ is a subring of A and $S\cap J$ is an ideal of S. Furthermore, the natural homomorphism $S\rightarrow (S+J)/J$ induces an isomorphism $S/S\cap J\rightarrow (S+J)/J$ **(isomorphism theorem)**.

A left (right) ideal of a ring A is said to be **maximal** if it is not equal to A and is properly contained in no left (right) ideal of A other than A. Similarly, a left (right) ideal of A is said to be **minimal** if it is nonzero and properly contains no nonzero left (right) ideal of A.

If e is an idempotent element of a unitary ring A, then $1-e$ and e are orthogonal idempotent elements, and $A=Ae+A(1-e)$ is the direct sum of left ideals. This is called **Peirce's left decomposition. Peirce's right decomposition** is defined similarly. A left ideal J of A can be expressed as $J=Ae$ for some idempotent element e if and only if there exists a left ideal J' such that $A=J+J'$ is a direct sum decomposition. More generally, if $e_1,\ldots,e_n$ are orthogonal idempotent elements whose sum is equal to 1, then $A=Ae_1+\ldots+Ae_n$ is the direct sum of left ideals. Conversely, if $A=J_1+\ldots+J_n$ is the direct sum of left ideals and $1=e_1+\ldots+e_n$ $(e_i\in J_i)$ is the corresponding decomposition of the unity element, then $e_1,\ldots,e_n$ are orthogonal idempotent elements. In particular, if $J_1,\ldots,J_n$ are two-sided ideals, then each J_i is a ring with unity element e_i, and by a natural correspondence, the ring A is isomorphic to the direct product $\prod_{i=1}^n J_i$. In this case, A is called the **direct sum** of ideals $J_1,\ldots,J_n$ and is denoted by $A=\bigoplus_{i=1}^n J_i$, or $A=\sum_{i=1}^n J_i$.

A ring A is called a **left (right) Artinian ring** if it is †Artinian as a left (right) A-module (i.e., if A satisfies the †minimal condition for left (right) ideals of A). A ring A is called a **left (right) Noetherian ring** if it is †Noetherian as a left (right) A-module (i.e., if A satisfies the †maximal condition for left (right) ideals of A). If A is commutative, *left* and *right* are omitted in these definitions. The property of being Artinian or Noetherian is inherited by quotient rings and the direct product of a finite family of rings, but not necessarily by subrings. For general rings, the maximal and minimal conditions for left (right) ideals are independent, but for unitary rings, left (right) Artinian rings are necessarily left (right) Noetherian (Y. Akizuki, C. Hopkins).

G. Semisimple Rings

The statement that a unitary ring A is †semisimple as a left A-module is equivalent to the statement that A is semisimple as a right A-module; in this case A is called a **semisimple ring** (→ Section H). Every module over a semisimple ring is also semisimple. A semisimple ring is left (right) Artinian and Noetherian. A semisimple ring is called a **simple ring** if it is nonzero and has no proper ideals except $\{0\}$, that is, if A is a quasisimple ring. Thus A is a simple ring if and only if A is a nonzero, unitary, quasisimple, left (right) Artinian ring. If A is a semisimple ring, then it has only a finite number of minimal ideals $A_1,\ldots,A_n$, and A is expressible as the direct sum $A=A_1+\ldots+A_n$, where each A_i is a simple ring, called a **simple component** of A. Any ideal of A is the direct sum of a finite number of simple components of A. Quotient rings of a semisimple ring and the direct product of a finite number of semisimple rings are also semisimple.

Any left (right) ideal of a semisimple ring A is expressible as Ae (eA) for some idempotent element e, and Ae (eA) is minimal if and only if e is primitive. In particular, a minimal left (right) ideal is a simple left (right) A-module that is contained in a certain simple component. Two minimal left (right) ideals are iso-

morphic as A-modules if and only if they are contained in the same simple component. If A_i, $1 \leqslant i \leqslant n$, are the simple components of A, then for each simple left A-module M there exists a unique i such that $A_i M \neq \{0\}$, and M is isomorphic to a minimal left ideal contained in A_i.

If M is a finite-dimensional linear space over a (skew) field D, then the endomorphism ring $A = \mathscr{E}_D(M)$ of M over D is a simple ring. Conversely, for any ring A, the endomorphism ring $D = \mathscr{E}_A(M)$ of a simple A-module M is a (skew) field (**Schur's lemma**). If A is a simple ring, then any simple A-module M can be considered as a finite-dimensional linear space over $D = \mathscr{E}_A(M)$, and A is isomorphic to $\mathscr{E}_D(M)$ (**Wedderburn's theorem**). Furthermore, if r is the dimension of M over D, then $\mathscr{E}_D(M)$ is isomorphic to the full matrix ring $M_r(D^\circ)$ of degree r over the field D°, which is anti-isomorphic to D. The dimension r is also equal to the †length of A as an A-module. The center of $A = \mathscr{E}_D(M)$ is isomorphic to the center of D, which is a commutative field. Thus a simple ring is an associative algebra over a commutative field (→ 29 Associative Algebras).

H. Radicals

Let A be a ring. Then among ideals consisting only of quasi-invertible elements of A, there exists a largest one, which is called the **radical** of A and denoted by $\mathfrak{R}(A)$. The radical of the residue ring $A/\mathfrak{R}(A)$ is $\{0\}$. A ring A is called a **semiprimitive ring** if $\mathfrak{R}(A)$ is $\{0\}$. On the other hand, A is called a left (right) **primitive ring** if it has a †faithful simple left (right) A-module. The radical $\mathfrak{R}(A)$ is equal to the intersection of all ideals J such that A/J is a left (right) primitive ring. In a unitary ring A, $\mathfrak{R}(A)$ coincides with the intersection of all maximal left (right) ideals of A. Furthermore, in a left (right) Artinian ring A, $\mathfrak{R}(A)$ is the largest nilpotent ideal of A, and the condition $\mathfrak{R}(A) = \{0\}$ is equivalent to the condition that A is a semisimple ring.

Among ideals consisting only of nilpotent elements of A, there exists a largest one, which is called the **nilradical** (or simply the radical) and denoted by $\mathfrak{N}(A)$ (→ 67 Commutative Rings). The nilradical of $A/\mathfrak{N}(A)$ is $\{0\}$. In general, $\mathfrak{N}(A)$ is contained in $\mathfrak{R}(A)$, and if A is left (right) Artinian, we have $\mathfrak{N}(A) = \mathfrak{R}(A)$. A ring A is called a **semiprimary ring** if $A/\mathfrak{N}(A)$ is left (right) Artinian and therefore semisimple. Furthermore, a ring A is called a **primary ring** (**completely primary ring**) if $A/\mathfrak{N}(A)$ is a simple ring (skew field). A primary ring is isomorphic to a full matrix ring over a completely primary ring.

References

[1] B. L. van der Waerden, Algebra, Springer, I, 1966; II, 1967.
[2] N. Jacobson, The theory of rings, Amer. Math. Soc. Math. Surveys, 1943.
[3] E. Artin, C. J. Nesbitt, and R. M. Thrall, Rings with minimum condition, Univ. of Michigan Press, 1944.
[4] N. Jacobson, Structure of rings, Amer. Math. Soc. Colloq. Publ., 1956.
[5] N. Bourbaki, Eléments de mathématique, Algèbre, ch. 1, 8, Actualités Sci. Ind., 1144b, 1261a, Hermann, 1964, 1958.
[6] C. Chevalley, Fundamental concepts of algebra, Academic Press, 1956.
[7] N. Jacobson, Lectures in abstract algebra I, Van Nostrand, 1951 (Springer, 1976).
[8] R. Godement, Cours d'algèbre, Hermann, 1963.
[9] J. P. Jans, Rings and homology, Holt, Rinehart and Winston, 1964.
[10] S. Lang, Algebra, Addison-Wesley, 1965.
[11] N. H. McCoy, The theory of rings, Macmillan, 1964.

369 (III.13) Rings of Polynomials

A. General Remarks

In this article, we mean by a *ring* a †commutative ring with †unity element. Let R be a ring, and let $X_1, \ldots, X_n$ be **variables** (**letters**, **indeterminates**, or **symbols**). Then the set of †polynomials in $X_1, \ldots, X_n$ with coefficients in R is called the **ring of polynomials** (or **polynomial ring**) in n variables $X_1, \ldots, X_n$ over R and is denoted by $R[X_1, \ldots, X_n]$ (→ 67 Commutative Rings; 284 Noetherian Rings; 337 Polynomials; 368 Rings). On the other hand, when R and R' are rings with common unity element and $R \subset R'$, then for a subset S of R' we denote the subring of R', generated by S over R, by $R[S]$. When $S = \{x_1, \ldots, x_n\}$, then there is a homomorphism φ of the polynomial ring $R[X_1, \ldots, X_n]$ onto $R[S]$ defined by

$$\varphi(\textstyle\sum a_{i_1 \ldots i_n} X_1^{i_1} \ldots X_n^{i_n}) = \sum a_{i_1 \ldots i_n} x_1^{i_1} \ldots x_n^{i_n} \qquad (a_{i_1 \ldots i_n} \in R).$$

If φ is an isomorphism, then $x_1, \ldots, x_n$ are said to be **algebraically independent** over R; and otherwise, **algebraically dependent** over R. Thus the ring of polynomials in n variables over R may be regarded as a ring $R[x_1, \ldots, x_n]$ generated by a finite system of algebraically independent elements $x_1, \ldots, x_n$ over R.

B. Ideals, Homogeneous Rings, and Graded Rings

Consider the polynomial ring $R[X]=R[X_1, \ldots, X_n]$ in n variables over a ring R. A polynomial $f\in R[X]$ is a †zero divisor if and only if there is a nonzero member a of R for which $af=0$. If $\mathfrak{a}$ is an †ideal of R, then $R[X]/\mathfrak{a}R[X]\cong(R/\mathfrak{a})[X_1, \ldots, X_n]$. Therefore, if $\mathfrak{p}$ is a †prime ideal of R, then $\mathfrak{p}R[X]$ is a prime ideal of $R[X]$. If R is a †unique factorization domain (u.f.d.), then $R[X]$ is also a u.f.d. If R is a normal ring, then so is $R[X]$. By the †Hilbert basis theorem, if R is †Noetherian, then $R[X]$ is also Noetherian. If m is the †Krull dimension of R, then Krull dim $R[X]\geqslant n+m$; the equality holds if R is Noetherian. If R is a field, then $R[X]$ is not only a u.f.d. but also a †Macaulay ring.

A **homogeneous ideal** of $R[X]$ is an ideal generated by a set of †homogeneous polynomials f_λ (the degree of f_λ may depend on λ). When $\mathfrak{a}$ is a homogeneous ideal, an element in $R[X]/\mathfrak{a}$ is defined to be a **homogeneous element** of degree d if it is the class of a homogeneous polynomial of degree d modulo $\mathfrak{a}$, and the quotient ring $R[X]/\mathfrak{a}$ is called a **homogeneous ring**. More generally, assume that a ring R is, as a †module, the †direct sum $\sum_{i=0}^{\infty} R_i$ of its submodules R_i $(i=0,1,2, \ldots)$ and that $R_iR_j\subset R_{i+j}$ for every pair (i,j). Then we call R a **graded ring**, and an element in R_d a **homogeneous element** of degree d. (In some literature the term graded ring is used in a wider sense; see below) In a graded ring R, if an ideal is generated by homogeneous elements, then the ideal is called a **homogeneous ideal** (or **graded ideal**). In a graded ring $R=\sum_{i=0}^{\infty} R_i$, if the ideal $\sum_{i=1}^{\infty} R_i$ has a finite basis, then R is generated (as a ring) by a finite number of elements over its subring R_0. Therefore, the graded ring $R=\sum R_i$ is Noetherian if and only if R_0 is Noetherian and R is generated by a finite number of elements over R_0. In this case, every homogeneous ideal is the intersection of a finite number of homogeneous †primary ideals, and every prime divisor of a homogeneous ideal is a homogeneous prime ideal.

The notion of a graded ring is generalized further as follows: A ring R is **graded** by an additive semigroup I (containing 0) if R is $\sum_{i\in I} R_i$ (direct sum) and if $R_iR_j\subset R_{i+j}$.

C. Zero Points

(1) Zero Points in an Affine Space. We consider the polynomial ring $K[X]=K[X_1, \ldots, X_n]$ in n variables over a field K and a field Ω containing K. A point $(a_1, \ldots, a_n)$ of an n-dimensional †affine space $\Omega^n=\{(a_1, \ldots, a_n) \mid a_i\in\Omega\}$ is called a **zero point** of a subset S of $K[X]$ if $f(a_1, \ldots, a_n)=0$ for every $f(X_1, \ldots, X_n)\in S$. A point $(a_1, \ldots, a_n)$ is called **algebraic** over K (**K-rational**) if every a_i is algebraic over K (is an element of K). In this way we define algebraic zero points and rational zero points. Zero points of S are zero points of the †ideals generated by S. Therefore, in order to investigate the set of zero points of S, we may restrict ourselves to the case where S is an ideal. Denote by $V(S)$ the set of zero points of S. If $\mathfrak{a}_1$, $\mathfrak{a}_2$ are ideals of $K[X]$, then (i) $V(\mathfrak{a}_1\cap\mathfrak{a}_2)=V(\mathfrak{a}_1\mathfrak{a}_2)=V(\mathfrak{a}_1)\cup V(\mathfrak{a}_2)$; (ii) $V(\mathfrak{a}_1+\mathfrak{a}_2)=V(\mathfrak{a}_1)\cap V(\mathfrak{a}_2)$; and (iii) if $\mathfrak{a}_1$ and $\mathfrak{a}_2$ have a common †radical, then $V(\mathfrak{a}_1)=V(\mathfrak{a}_2)$.

(2) Zero Points in a Projective Space. A point $(\lambda a_1, \ldots, \lambda a_n)$ of an $(n-1)$-dimensional †projective space over Ω (with $a_i\in\Omega$, some $a_i\neq 0$, $\lambda\in\Omega$, $\lambda\neq 0$) is called a **zero point** of a polynomial $f(X_1, \ldots, X_n)$ if, f being expressed as $\sum f_i$ with homogeneous polynomials f_i of degree i, $f_i(a_1, \ldots, a_n)=0$ for every i (this condition holds if and only if $f(\lambda a_1, \ldots, \lambda a_n)=0$ for any element λ in Ω, provided that Ω contains infinitely many elements). Therefore, zero points of a subset S of $K[X]$ are zero points of the smallest homogeneous ideal containing S. Thus, in order to study the sets of zero points, it is sufficient to consider sets of zero points of homogeneous ideals, and propositions similar to (i), (ii), and (iii) of part 1 of this section hold for homogeneous ideals $\mathfrak{a}_1$, $\mathfrak{a}_2$.

D. The Normalization Theorem

Let $\mathfrak{a}$ be an ideal of †height h in the polynomial ring $K[X]=K[X_1, \ldots, X_n]$ in n variables over a field K. Then there exist elements $Y_1, \ldots, Y_n$ of $K[X]$ such that (i) $K[X]$ is †integral over $K[Y]=K[Y_1, \ldots, Y_n]$ and (ii) $Y_1, \ldots, Y_h$ generate $\mathfrak{a}\cap K[Y]$ (**normalization theorem for polynomial rings**).

Using this theorem, we obtain the following important theorems on finitely generated rings.

(1) **Normalization theorem for finitely generated rings.** If a ring R is finitely generated over an integral domain I, then there exist an element a $(\neq 0)$ of I and algebraically independent elements $z_1, \ldots, z_t$ of R over I such that the †ring of quotients R_S (where $S=\{a^n \mid n=1,2, \ldots\}$) is integral over $I[a^{-1}, z_1, \ldots, z_t]$.

(2) If $\mathfrak{p}$ is a prime ideal of an integral domain R that is finitely generated over a field K, then (height of $\mathfrak{p}$)+(†depth of $\mathfrak{p}$)=(†transcendence degree of R over K), and the depth of $\mathfrak{p}$ coincides with the transcendence degree of $R/\mathfrak{p}$ over K. In particular, if $\mathfrak{m}$ is a maximal ideal of R, then $R/\mathfrak{m}$ is algebraic over K.

(3) **Hilbert's zero point theorem (Hilbert Nullstellensatz).** Let $\mathfrak{a}$ be an ideal of the polynomial ring $K[X]=K[X_1, \ldots, X_n]$ over the field K, and assume that the field Ω containing K is †algebraically closed. If $f \in K[X]$ satisfies the condition that every algebraic zero point (→ Section C) of $\mathfrak{a}$ is a zero point of f, then some power of f is contained in $\mathfrak{a}$.

E. Elimination Theory

Let $f_1, \ldots, f_N$ be elements of the polynomial ring $R=I[X_1, \ldots, X_m, Y_1, \ldots, Y_n]$ in $m+n$ variables over an integral domain I. For each maximal ideal $\mathfrak{m}$ of I, let $\varphi_{\mathfrak{m}}$ be the canonical homomorphism with modulus $\mathfrak{m}$, and let $\Omega_{\mathfrak{m}}$ be an algebraically closed field containing $I/\mathfrak{m}$. Let $W_{\mathfrak{m}}$ be the set of points $(a_1, \ldots, a_n)$ of the n-dimensional affine space $\Omega_{\mathfrak{m}}^n$ over $\Omega_{\mathfrak{m}}$ such that the system of equations $\varphi_{\mathfrak{m}}(f_i)(X_1, \ldots, X_m, a_1, \ldots, a_n)=0$ $(i=1,2, \ldots, N)$ has a solution in $\Omega_{\mathfrak{m}}^m$. To **eliminate** $X_1, \ldots, X_m$ from $f_1, \ldots, f_N$ is to obtain $g(Y_1, \ldots, Y_n) \in I[Y_1, \ldots, Y_n]$ such that every point of $W_{\mathfrak{m}}$ is a zero point of $\varphi_{\mathfrak{m}}(g)$ for every $\mathfrak{m}$; such a g (or an equation $g=0$) is called a **resultant** of $f_1, \ldots, f_N$. The set $\mathfrak{a}$ of resultants forms an ideal of $I[Y_1, \ldots, Y_n]$, and $\{g_1, \ldots, g_M\}$ is called a **system of resultants** if the radical of the ideal generated by it coincides with $\mathfrak{a}$. If I is finitely generated over a field, then, denoting by $\mathfrak{b}$ the radical of the ideal generated by $f_1, \ldots, f_N$, we have $\mathfrak{a}=\mathfrak{b} \cap I[Y_1, \ldots, Y_n]$. In particular, let I be a field. It is obvious that $W_{(0)}$ is contained in the set V of zero points of $\mathfrak{a}$. However, it is not necessarily true that $V=W_{(0)}$. If every f_i is homogeneous in $X_1, \ldots, X_m$ and also in $Y_1, \ldots, Y_n$, then we have $V=W_{(0)}$.

If we wish to write a system of resultants explicitly, we can proceed as follows: Regard the f_i as polynomials in X_1 with coefficients in $I[X_2, \ldots, X_m, Y_1, \ldots, Y_n]$, and obtain resultants $R(f_i, f_j)$ by eliminating X_1 from the pairs f_i, f_j. Then eliminate X_2 from these resultants, and so forth. To obtain $R(f_i, f_j)$, we may use **Sylvester's elimination method**. Namely, let f and g be polynomials in x with coefficients in I: $f=a_0x^m+a_1x^{m-1}+\ldots+a_m$, $g=b_0x^n+b_1x^{n-1}+\ldots+b_n$. Let $D(f,g)$ be the following determinant of degree $m+n$:

$$\begin{array}{l} n\left\{\begin{array}{l}\\ \\ \\ \\ \end{array}\right. \\ m\left\{\begin{array}{l}\\ \\ \\ \\ \end{array}\right.\end{array}\begin{vmatrix} a_0 & a_1 & \cdots & a_m & 0 & & \cdots & 0 \\ 0 & a_0 & \cdots & a_{m-1} & a_m & 0 & \cdots & 0 \\ & & & \cdots & & & & \\ & & & \cdots & & & & \\ 0 & & \cdots & 0 & a_0 & a_1 & \cdots a_{m-1} & a_m \\ b_0 & b_1 & \cdots & & b_{n-1} & b_n & 0 \cdots & 0 \\ & & & \cdots & & & & \\ & & & \cdots & & & & \\ 0 & \cdots & 0 & b_0 & b_1 & & \cdots b_{n-1} & b_n \end{vmatrix}.$$

Then $D(f,g)=0$ if and only if either f and g have a common root or $a_0=b_0=0$. Therefore, if I is a u.f.d. and a_0 and b_0 have no common factor, then $D(f,g)$ is the required resultant $R(f,g)$. (For other methods of elimination see B. L. van der Waerden, *Algebra*, vol. II. For criteria on whether a finitely generated ring over a field is †regular → 370 Rings of Power Series B.)

F. Syzygy Theory

(1) **Classical Case.** The notion of **syzygy** was introduced by Sylvester (*Phil. Trans.*, 143 (1853)), then generalized and clarified by Hilbert [3], whose definition can be formulated as follows: Let $R=k[X_1, \ldots, X_n]$ be a polynomial ring of n variables over a field k. R has the natural gradation (i.e., R is a graded ring in which each X_i $(1 \leqslant i \leqslant n)$ is of degree 1 and elements of k are of degree 0). Let M be a finitely generated graded R-module. If $f_1, \ldots, f_m$ form a minimal basis of M over R consisting of homogeneous elements, we introduce m indeterminates $u_1, \ldots, u_m$ and put $F=\sum_{1 \leqslant j \leqslant m} Ru_j$, the free R-module generated by $u_1, \ldots, u_m$. Set $\deg(u_j)=\deg(f_j)$ $(1 \leqslant j \leqslant m)$ and supply F with the structure of a graded R-module. Let φ be the graded R-homomorphism of F onto M defined by $\varphi(u_j)=f_j$. Then $N=\operatorname{Ker}(\varphi)$ is a graded R-module uniquely determined by M up to isomorphism (of graded R-modules); N is called the **first syzygy** of M. For a positive integer r, the r**th syzygy** of M is inductively defined as the first syzygy of the $(r-1)$st syzygy of M. The **Hilbert syzygy theorem** states that for any finitely generated graded R-module M, the nth syzygy of M is free. In other words, M admits a **free resolution** of length $\leqslant n$, i.e., an exact sequence of the form

$$0 \to F^{(v)} \to \ldots \to F^{(1)} \to F^{(0)} \to M \to 0,$$

where $v \leqslant n$ and each $F^{(i)}$ $(0 \leqslant i \leqslant v)$ is a finitely generated free graded R-module. It follows that if M_d denotes the homogeneous part of degree d in M, there exists a polynomial $P(X)$ of degree $\leqslant n-1$ such that $\dim_k(M_d)=P(d)$ for sufficiently large d; $P(X)$ is called the **Hilbert polynomial** (or **characteristic function**) of the graded R-module M.

(2) **Generalization by Serre.** The syzygy theory was generalized by J.-P. Serre [2] as follows: Let R be a Noetherian ring and M a finitely generated R-module. Then we can find a finitely generated free R-module F and an R-homomorphism φ of F onto M. The kernel of φ, called a **first syzygy** of M, is not uniquely

determined by M. However, if N_1 and N_2 are first syzygies of M, then there exist finitely generated †projective R-modules P_1 and P_2 such that $N_1 \oplus P_1 \cong N_2 \oplus P_2$ (→ 277 Modules K). For a positive integer r, an r**th syzygy** is defined inductively as in (1) of this section. An important result of Serre is that R is a †regular ring of †Krull dimension at most n if and only if an nth syzygy of every finitely generated R-module is †projective.

(3) Serre Conjecture. D. Quillen (*Invertiones Math.*, 36 (1976)) and A. Suslin (Dokl. Akad. Nauk SSSR (26 Feb. 1976)) solved the Serre conjecture by proving that every projective module over a ring of polynomials over a field is free.

(4) Special Cases. In the following special cases, we can define the first syzygy of M uniquely up to isomorphism: (i) R is a Noetherian †local ring and M is a finitely generated R-module; and (ii) R is a graded Noetherian ring $\sum_{d \geqslant 0} R_d$, where R_0 is a field and M is a finitely generated graded R-module [4].

References

[1] M. Nagata, Local rings, Wiley, 1962 (Krieger, 1975).
[2] J.-P. Serre, Algèbre locale, multiplicités, Lecture notes in math., Springer, 1965.
[3] D. Hilbert, Über die Theorie der algebraischen Formen, Math. Ann., 36 (1890), 473–534.
[4] J.-P. Serre, Sur la dimension homologique des anneaux et des modules noethériens, Proc. Intern. Symp. Alg. Number Theory, Tokyo and Nikko (1955), 175–189.
Also → references to 284 Noetherian Rings.

370 (III.14) Rings of Power Series

A. Rings of Formal Power Series (→ 67 Commutative Rings; 284 Noetherian Rings; 368 Rings)

Let R be a commutative ring with unity element 1. Let F_d be the module of †homogeneous polynomials of degree d in $X_1, \ldots, X_n$ with coefficients in R. A formal infinite sum $\sum_{d=0}^{\infty} a_d = a_0 + a_1 + \ldots + a_n + \ldots +$ of elements $a_d \in F_d$ is called a **formal power series** or simply **power series** in n variables $X_1, \ldots, X_n$ with coefficients in R, and a_d is called the **homogeneous part** of degree d of the power series. The homogeneous part a_0 of degree zero is called the **constant term**. Addition and multiplication are defined by $(\sum a_d) + (\sum b_d) = \sum (a_d + b_d)$, $(\sum a_d)(\sum b_d) = \sum_d (\sum_{i+j=d} a_i b_j)$. By these operations, the set of power series forms a commutative ring, which is called the **ring of (formal) power series** (or **(formal) power series ring**) in $X_1, \ldots, X_n$ over R and is denoted by $R[[X_1, \ldots, X_n]]$ or $R\{X_1, \ldots, X_n\}$. If there is a natural number N such that $a_d = 0$ for every $d > N$, then the power series $\sum a_d$ is identified with the polynomial $a_0 + a_1 + \ldots + a_N$. Thus $R[X_1, \ldots, X_n] \subset R[[X_1, \ldots, X_n]]$. Set $\mathfrak{X} = \sum_i X_i R[[X_1, \ldots, X_n]]$. Then $R[[X_1, \ldots, X_n]]$ is †complete under the $\mathfrak{X}$-adic topology (→ 284 Noetherian Rings B).

Assume that R' is a commutative ring containing R and having a unity element in common with R, $\mathfrak{a}'$ is an ideal of R' such that R' is complete under the $\mathfrak{a}'$-adic topology, and $x_1, \ldots, x_n$ are elements of $\mathfrak{a}'$. Then an infinite sum $\sum c_{i_1 \ldots i_n} x_1^{i_1} \ldots x_n^{i_n}$ (each i_j ranges over nonnegative rational integers and $c_{i_1 \ldots i_n} \in R$) has a well-defined meaning in R' (namely, if S_d is a finite sum of these terms such that $\sum i_j \leqslant d$, then the infinite sum is defined to be $\lim_{d \to \infty} S_d$). This element $\sum c_{i_1 \ldots i_n} x_1^{i_1} \ldots x_n^{i_n}$ is also called a **power series** in $x_1, \ldots, x_n$ with coefficients in R. The set of such power series in $x_1, \ldots, x_n$ is a subring of R', called the **power series ring** in $x_1, \ldots, x_n$ over R and denoted by $R[[x_1, \ldots, x_n]]$ or $R\{x_1, \ldots, x_n\}$. Defining φ by $\varphi(\sum c_{i_1 \ldots i_n} X_1^{i_1} \ldots X_n^{i_n}) = \sum c_{i_1 \ldots i_n} x_1^{i_1} \ldots x_n^{i_n}$, we obtain a ring homomorphism φ: $R[[X_1, \ldots, X_n]] \to R[[x_1, \ldots, x_n]]$. If φ is an isomorphism, then we say that $x_1, \ldots, x_n$ are **analytically independent** over R.

If $\overline{\mathfrak{m}}$ is a †maximal ideal of the formal power series ring $R[[X_1, \ldots, X_n]]$, then $\mathfrak{m} = \overline{\mathfrak{m}} \cap R$ is a maximal ideal of R and $\overline{\mathfrak{m}}$ is generated by $\mathfrak{m}$ and $X_1, \ldots, X_n$. An element f of the power series ring is †invertible if and only if its constant term f_0 is an invertible element of R, and in this case $f^{-1} = \sum_{d=0}^{\infty} f_0^{-d-1} \cdot (f_0 - f)^d$. If R is one of the following, then $R[[X_1, \ldots, X_n]]$ is also of the same kind: (i) †Noetherian ring, (ii) †local ring, (iii) †semilocal ring, (iv) †integral domain, (v) †regular local ring, (vi) Noetherian †normal ring. But even if R is a †unique factorization domain (u.f.d.), $R[[X_1, \ldots, X_n]]$ need not be a u.f.d. (If R is a field, or more generally, if R is a regular semilocal ring, then $R[[X_1, \ldots, X_n]]$ is a u.f.d.). In particular, a formal power series ring $k[[X]]$ in one variable X over a field k is an integral domain whose field of quotients is called the **field of (formal) power series** (or **(formal) power series field) in one variable** X over k and is denoted by $k((X))$; an element of $k((X))$ is expressed uniquely in the form $\sum_{n=r}^{\infty} a_n X^n$ $(a_n \in k, r \in \mathbf{Z})$.

B. Rings of Convergent Power Series

Let K be a field with multiplicative †valuation v (for instance, $K=\mathbf{C}$, the complex number field, and $v(\alpha)=|\alpha|$ for $\alpha\in\mathbf{C}$). A formal power series $f(X_1,\ldots,X_n)=\sum c_{i_1\ldots i_n}X_1^{i_1}\ldots X_n^{i_n}$ is said to be a **convergent power series** if there are positive numbers $r_1,\ldots,r_n$, M such that $v(c_{i_1\ldots i_n})r_1^{i_1}\ldots r_n^{i_n}\leqslant M$ for every $(i_1,\ldots,i_n)$. In this case, if $a_i\in K$ and $v(a_i)<r_i$, then $\sum c_{i_1\ldots i_n}a_1^{i_1}\ldots a_n^{i_n}$ has its sum in the †completion of K. The set of convergent power series is a subring of $K[[X_1,\ldots,X_n]]$. It is called the **ring of convergent power series** (or **convergent power series ring**) in n variables over K and is denoted by $K\langle\langle X_1,\ldots,X_n\rangle\rangle$ or $K\{X_1,\ldots,X_n\}$. It is a regular local ring of †Krull dimension n. Hence it is a u.f.d. and its completion is $K[[X_1,\ldots,X_n]]$. If v is a †trivial valuation, then $K\langle\langle X_1,\ldots,X_n\rangle\rangle=K[[X_1,\ldots,X_n]]$.

Weierstrass's Preparation Theorem. For an element $f=\sum c_{i_1\ldots i_n}X_1^{i_1}\ldots X_n^{i_n}\in K\langle\langle X_1,\ldots,X_n\rangle\rangle$, assume that $c_{0\ldots 0i}=0$ for $i=0,1,\ldots,r-1$ and $c_{0\ldots 0r}\neq 0$. Then for an arbitrary element g of $K\langle\langle X_1,\ldots,X_n\rangle\rangle$, there exists a unique $q\in K\langle\langle X_1,\ldots,X_n\rangle\rangle$ such that $g-qf\in\sum_{i=0}^{r-1}X_n^iK\langle\langle X_1,\ldots,X_{n-1}\rangle\rangle$. In particular (considering the case where $g=X_n^r$), there is an invertible element u of $K\langle\langle X_1,\ldots,X_n\rangle\rangle$ such that $fu=f_0+f_1X_n+\ldots+f_{r-1}X_n^{r-1}+X_n^r$ $(f_i\in K\langle\langle X_1,\ldots,X_{n-1}\rangle\rangle)$.

By this theorem, we see easily that if $\mathfrak{a}$ is an ideal of †height h of $K\langle\langle X_1,\ldots,X_n\rangle\rangle$, then $K\langle\langle X_1,\ldots,X_n\rangle\rangle/\mathfrak{a}$ is isomorphic to a ring that is a finite module over $K\langle\langle X_1,\ldots,X_{n-h}\rangle\rangle$.

If $\mathfrak{p}$ is a †prime ideal of $K\langle\langle X_1,\ldots,X_n\rangle\rangle$, then $\mathfrak{p}K[[X_1,\ldots,X_n]]$ is a prime ideal.

The Jacobian Criterion. Let K be a field, and let R be the ring of polynomials $K[X_1,\ldots,X_n]$, the ring of formal power series $K[[X_1,\ldots,X_n]]$, or the ring of convergent power series $K\langle\langle X_1,\ldots,X_n\rangle\rangle$ in n variables $X_1,\ldots,X_n$ over K. †Partial derivatives $\partial/\partial X_i$ are well defined in R. For $f_1,\ldots,f_t\in R$, a †Jacobian matrix $J(f_1,\ldots,f_t)$ is defined to be the $t\times n$ matrix whose (i,j)-entry is $\partial f_i/\partial X_j$. Let $\mathfrak{p}$ be a †prime divisor of the ideal $\sum_i f_iR$, and let $\mathfrak{q}$ be a prime ideal containing $\mathfrak{p}$. If the †rank of $(J(f_1,\ldots,f_t)$ modulo $\mathfrak{q})$ is equal to the height of $\mathfrak{p}$, then the ring $R_{\mathfrak{q}}/\sum f_iR_{\mathfrak{q}}$ is a regular local ring. The converse is also true if K is a †perfect field (if K is not a perfect field, then, modifying $J(f_1,\ldots,f_t)$, we can have a similar criterion [1]).

C. Hensel Rings

A **Hensel ring** (or **Henselian ring**) is a commutative ring with unity element satisfying the following two conditions: (i) R has only one maximal ideal $\mathfrak{m}$ (i.e., R is a †quasilocal ring); and (ii) if f, g_0, h_0 are monic polynomials in one variable x (here, a polynomial in x is called monic if the coefficient of the term of the highest degree is 1) such that $f-g_0h_0\in\mathfrak{m}R[x]$, $g_0R[x]+h_0R[x]+\mathfrak{m}R[x]=R[x]$, then there are monic polynomials g, $h\in R[x]$ such that $f=gh$ and $g\equiv g_0$, $h\equiv h_0$ modulo $\mathfrak{m}$.

Important examples of Hensel rings are complete local rings, rings of convergent power series, and †complete valuation rings. When R is a Hensel ring, a commutative ring R' with unity element such that R' is a finite R-module is the direct sum of a finite number of Hensel rings. For any quasilocal ring Q, there exists a Hensel ring $\tilde{Q}$, called the **Henselization** of Q (for details → [1]), for which the following statements hold: (i) $\tilde{Q}$ is a †faithfully flat Q-module; (ii) if $\mathfrak{m}$ is the maximal ideal of Q, then the maximal ideal of $\tilde{Q}$ is $\mathfrak{m}\tilde{Q}$, and $\tilde{Q}/\mathfrak{m}\tilde{Q}=Q/\mathfrak{m}Q$; (iii) if R is a Hensel ring that contains Q and has a maximal ideal $\mathfrak{n}$, and $\mathfrak{n}\cap Q=\mathfrak{m}$, then there is one and only one Q-homomorphism φ of $\tilde{Q}$ into R; (iv) if Q is a †normal ring, then the Q-homomorphism φ is an injection; (v) if Q is a local ring, then $\tilde{Q}$ is also a local ring, and Q is dense in $\tilde{Q}$.

References

[1] M. Nagata, Local rings, Interscience, 1962 (Krieger, 1975).
[2] O. Zariski and P. Samuel, Commutative algebra II, Van Nostrand, 1960; new edition, Springer, 1975.

371 (XVIII.10) Robust and Nonparametric Methods

A. General Remarks

Robust and **nonparametric** or **distribution-free methods** are statistical procedures specifically devised to deal with broad families of probability distributions.

In the theory of statistical inference it is usual to assume that the probability distribution of the population from which the observed values are chosen at random is specified exactly except for a small number of unknown parameters (→ 401 Statistical Inference). In practical applications, however, it often happens that the assumptions made for the model,

especially those about the shape of the distribution, may not hold for the actual data. In such cases robust and/or nonparametric procedures that do not require exact knowledge of the shape of the distribution and yet prove to be relatively efficient or valid are required. The term **nonparametric** or **distribution-free** is used for problems of testing hypotheses, and the term **robust** is mainly used for problems of point estimation (→ 396 Statistic; 399 Statistical Estimation; 400 Statistical Hypothesis Testing).

Although the idea of the sign test appears in the work of J. Arbuthnot (1710), the theoretical foundation for nonparametric tests was first given in the proposals for the permutation test by R. A. Fisher (1935), the rank test by F. Wilcoxon (1945), and the test based on U-statistics by H. B. Mann and D. R. Whitney (1947). In the years that followed two important ideas appeared: the concept of asymptotic relative efficiency by E. J. G. Pitman (1948) and the development of the theory of U-statistics by W. Hoeffding (1948). H. Chernoff and I. R. Savage (1958) showed, in studying the asymptotic distribution of a class of rank statistics, that the asymptotic efficiencies of nonparametric tests are incredibly high. These findings accelerated the studies of nonparametric tests; recent progress is summarized in the books by J. Hájek and Z. Šidák [1], M. L. Puri and P. K. Sen [2], R. H. Randles and D. A. Wolfe [3], and P. J. Huber [6].

On the other hand, G. E. P. Box (1953) first coined the term robustness in his sensitivity studies, in which he investigated how the standard statistical procedures obtained under certain assumptions are influenced when such assumptions are violated. Two papers by J. W. Tukey (1960, 1962) provided the initial foundation for robust estimation. J. L. Hodges and E. L. Lehmann (1963) noticed that estimators of location could be derived from nonparametric tests and that these estimators have sometimes much higher efficiency than the sample mean. A similar study for scale was made by S. Kakeshita and T. Yanagawa (1967). Huber (1964) proposed an estimator of location by generalizing the method of least squares. Along with the idea of the influence curve introduced by F. R. Hampel (1974) the estimator proposed by Huber has become a core of subsequent studies of robust estimation. K. Takeuchi (1971) proposed an adaptive estimate that is asymptotically fully efficient for a wide class of underlying distribution functions. The developments of the theory of robust estimation are reviewed by Huber [4–6] and R. V. Hogg [7]. Various proposed estimators are compared in the book by D. F. Andrews et al. [8].

B. The One-Sample Problem

Let $F(x)$ be a †distribution function of a †random variable X, $(X_1, \ldots, X_n)$ be an independent †random sample of size n from $F(x)$, and $(x_1, \ldots, x_n)$ be an observed sample value. The $100p$ percentile of F is denoted by ξ_p, i.e., $F(\xi_p)=p$. For testing the †hypothesis $H:\xi_p \leqslant \xi^0$ against the †alternative hypothesis $A: \xi_p > \xi^0$, the following procedure is proposed. Let $i(x_1, \ldots, x_n)$ be the number of x_i that are greater than ξ^0. A test procedure by the following †test function φ is †uniformly most powerful in some neighborhood of $\xi_p = \xi^0$ for the double exponential distribution, where $\varphi(x_1, \ldots, x_n)$ is defined by the equations

$$\varphi(x_1, \ldots, x_n) = \begin{cases} 1 & \text{when } i(x_1, \ldots, x_n) > c, \\ a & \text{when } i(x_1, \ldots, x_n) = c, \\ 0 & \text{when } i(x_1, \ldots, x_n) < c \end{cases}$$

$(0<a<1, 0 \leqslant c \leqslant n)$. This procedure is called the **sign test**.

Suppose that $F(x)$ is symmetric about $x = \xi_{1/2}$. Let R_i^+ be the rank of $|X_i - \xi^0|$ among $|X_1 - \xi^0|, \ldots, |X_n - \xi^0|$, and let $\Psi(t) = 1, 0$ according as $t>0, \leqslant 0$. Set

$$S_n(X_1, \ldots, X_n) = \sum_{i=1}^{n} a_n(R_i^+) \Psi(X_i - \xi^0)$$

for some weights $a_n(1), \ldots, a_n(n)$. The following procedure φ, called the **signed rank test**, is also used for testing the hypothesis $H:\xi_{1/2} \leqslant \xi^0$ against the alternative hypothesis $A:\xi_{1/2} > \xi^0$:

$$\varphi(x_1, \ldots, x_n) = \begin{cases} 1 & \text{when } S_n(x_1, \ldots, x_n) > c, \\ a & \text{when } S_n(x_1, \ldots, x_n) = c, \\ 0 & \text{when } S_n(x_1, \ldots, x_n) < c. \end{cases}$$

The procedure with $a_n(i) = i$ is frequently used and is called the **Wilcoxon signed rank test**, which is the uniformly most powerful rank test in a neighborhood of $\xi_{1/2} = \xi^0$ for $F(x) = 1/(1+e^{-x})$, the logistic distribution.

C. The Two-Sample Problem

Let F and G be continuous distribution functions of random variables X and Y, respectively, $(X_1, \ldots, X_m)$ and $(Y_1, \ldots, Y_n)$ be the corresponding random samples, and $(x_1, \ldots, x_m)$ and $(y_1, \ldots, y_n)$ be the respective sample values. Consider the problem of testing the hypothesis $H: F(x) \equiv G(x)$ against the alternative hypothesis $A_1: F(x) \not\equiv G(x)$ or $A_2: F(x) \geqslant G(x)$ for all x and $F(x) \not\equiv G(x)$. When the alternative hypothesis A_2 is true, we say that the random variable Y is **stochastically larger** than X and write $F > G$. A frequently used example of such an alternative hypothesis A_2 is $G(x) = F(x-\theta)$, θ

>0. Let $\mathscr{H}$ be the family of all strictly increasing continuous functions. Then the hypothesis H and the alternative hypothesis A_2 are invariant under the group of transformations of the form $x_i'=h(x_i)$, $y_j'=h(y_j)$ $(i=1,\ldots,m;\ j=1,\ldots,n;\ h\in\mathscr{H})$. The †maximal invariant statistic in this case is the rank $(R_1,\ldots,R_m)$ of $(X_1,\ldots,X_m)$ or the rank $(S_1,\ldots,S_n)$ of $(Y_1,\ldots,Y_n)$ when the combined sample $(X_1,\ldots,X_m;Y_1,\ldots,Y_n)$ is ordered in an ascending order. If a test function $\varphi(x_1,\ldots,x_m;y_1,\ldots,y_n)$ satisfies $P_\varphi(F,F)\leqslant\alpha$ and $P_\varphi(F,G)\geqslant\alpha$ for any $F>G$, then φ is considered a desirable test, where

$$P_\varphi(F,G)=\int\cdots\int\varphi(x_1,\ldots,x_m;y_1,\ldots,y_n)\times\prod_i dF(x_i)\prod_j dG(y_j).$$

Lehmann's Theorem. If φ satisfies the conditions stating that $y_j^*>y_j$ $(j=1,\ldots,n)$ yield $\varphi(x_1,\ldots,x_m;y_1^*,\ldots,y_n^*)\geqslant\varphi(x_1,\ldots,x_m;y_1,\ldots,y_n)$, then $P_\varphi(F,G)\geqslant P_\varphi(F,F)$. If in addition φ is a †similar test, then φ is unbiased (→ 400 Statistical Hypothesis Testing C).

The **Wilcoxon test** (or the **Mann-Whitney U-test**) is described by a test function $\varphi=1$ when $U\geqslant c$ and $\varphi=0$ when $U<c$, where U is a †U-statistic defined by

$$U=\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\psi(x_i,y_j)$$

with $\psi(x,y)=1$ when $x\leqslant y$ and $\psi(x,y)=0$ when $x>y$. This test is similar and unbiased.

Testing the hypothesis $H:F=G$ against the alternative hypothesis $A:F\neq G=F(x/\sigma)$, $\sigma>1$, is another two-sample problem, for which the following test was proposed by T. Tamura. The test function is given by $\varphi=1$ for $U\geqslant c$ and $\varphi=0$ for $U<c$, where $U=\binom{m}{2}^{-1}\binom{n}{2}^{-1}\sum_{i<i'}\sum_{j<j'}\psi(x_i,x_{i'};y_j,y_{j'})$ with $\psi(u,u';v,v')=1$ when $v<u<v'$, $v<u'<v'$ or $v'<u<v$, $v'<u'<v$ and $\psi(u,u';v,v')=0$ otherwise.

The following statistic T_N is used frequently in nonparametric problems. Let $x_1,\ldots,x_m$; $y_1,\ldots,y_n$ be arranged in order of magnitude. Set $z_k=+1$ or 0 when the kth value $(k=1,2,\ldots,n+m)$ in the arrangement is an x_i or y_j, respectively. For a given set of $N=n+m$ reals $\{e_k\}$, T_N is defined by $T_N=m^{-1}\sum_{k=1}^{N}e_kz_k$. Set $H_N(x)=\lambda_NF_m(x)+(1-\lambda_N)G_n(x)$, where $F_m(x)$ and $G_n(x)$ are the †empirical distribution functions based on $(x_1,\ldots,x_m)$ and $(y_1,\ldots,y_n)$, respectively, and $0<\lambda_0\leqslant\lambda_N=m/N\leqslant1-\lambda_0<1$. Then T_N is represented by the integral

$$\int J_N(H_N(x))\,dF_m(x)$$

with $e_k=J_N(k/N)$. Chernoff and Savage [9] proved that under some regularity conditions the asymptotic distribution of T_N is normal and that the asymptotic mean μ and the variance σ^2 of T_N are given by

$$\mu=\int J(H(x))\,dF(x),$$

$$N\sigma^2=2(1-\lambda)\left\{\iint_{x<y}G(x)(1-G(y))\times J'(H(x))J'(H(y))\,dF(x)\,dF(y)+\frac{1-\lambda}{\lambda}\iint_{x<y}F(x)(1-F(y))\times J'(H(x))J'(H(y))\,dG(x)\,dG(y)\right\},$$

where $J(H)=\lim_{N\to\infty}J_N(H)$, $H(x)=\lambda F(x)+(1-\lambda)G(x)$, and $\lambda=\lim\lambda_N$. When $e_k=k/N$, the statistic T_N is equivalent to the †U-statistic in the Wilcoxon test. When e_k is the mean $E(Z_k)$ of the kth order statistic Z_k in an independent sample of size N from $N(0,1)$, then the test by T_N is called the **Fisher-Yates-Terry normal score test**. When J is the inverse function of the distribution function of $N(0,1)$, then the test is called the **van der Waerden test**.

D. The k-Sample Problem

Let $(X_{ij},j=1,\ldots,n_i)$ be a random sample of size n_i from the population with a distribution function $F_i(x)$ for each $i=1,\ldots,k$. The **k-sample problem** is concerned with testing the hypothesis $H:F_1(x)=\ldots=F_k(x)$ against an alternative hypothesis A_1: not all the $F_i(x)$ are equal, $A_2:F_i(x)=F(x-\theta_i)$ with $\theta_i\not\equiv\theta$, or $A_3:F_i(x)=F(x/\sigma_i)$ with $\sigma_i\not\equiv\sigma$. Several tests have been proposed for this problem, using quadratic forms of the vector-valued U-statistic $\mathbf{U}=(U^1,\ldots,U^k)$ whose coordinates U^i are defined by means of a function

$$\psi^i(x_{11},\ldots,x_{1m_{1i}};\ldots;x_{k1},\ldots,x_{km_{ki}}),\qquad i=1,\ldots,k.$$

When $N=\sum n_i\to\infty$ with $n_i=\rho_iN$, $0<\rho_i<1$, and $\sum\rho_i=1$, then

$$\mathbf{V}=(\sqrt{N}(U^1-E(U^1)),\ldots,\sqrt{N}(U^k-E(U^k)))$$

has asymptotically a †multivariate normal distribution $N(0,\Sigma)$. Let B be the projection matrix corresponding to the eigenspace for the zero eigenvalues of the matrix Σ, and let Λ be a matrix such that $\Lambda B=0$, $\Sigma\Lambda=I-B$. The statistic $\mathbf{V}\Lambda^t\mathbf{V}$ has asymptotically a †noncentral chi-square distribution with degrees of freedom $=$ rank Σ. Several kinds of test represented by a critical region of the form $\mathbf{V}\Lambda\mathbf{V}'>c$ are proposed, among which the **Kruskal-Wallis test** is a particular one having

$$\psi^i(x_1;\ldots;x_k)=\sum_\alpha\frac{1}{n_in_\alpha}\delta(x_\alpha,x_i),\qquad i=1,\ldots,k,$$

as basic functions, where $\delta(x,y)=1$ when $x<y$ and $\delta(x,y)=0$ otherwise.

E. Asymptotic Relative Efficiency of Tests

If there is more than one test procedure for a given testing problem, then one may wish to compare these procedures. Let $\{\varphi_n\}$ and $\{\psi_n\}$ be two sequences of level α tests, where φ_n and ψ_n are test functions based on a sample of size n. The †power functions of φ_n and ψ_n are denoted by $\beta(\theta\mid\varphi_n)$ and $\beta(\theta\mid\psi_n)$, respectively. Let θ be a real parameter and $\{\theta_i\}$ be such that $\theta_i\to\theta_0$ as $i\to\infty$. Consider a hypothesis $\theta=\theta_0$ and a sequence $\{\theta_i\}$ of alternative hypotheses. If, for any increasing sequences $\{n_i\}$ and $\{n_i^*\}$ of positive integers satisfying $\alpha<\lim\beta(\theta_i\mid\varphi_{n_i})=\lim\beta(\theta_i\mid\psi_{n_i^*})<1$, $\lim n_i^*/n_i$ $(=e(\{\varphi_n\},\{\psi_n\})$, say) exists and is independent of α and $\lim\beta(\theta_i\mid\varphi_{n_i})$, then e is called Pitman's asymptotic relative efficiency of $\{\varphi_n\}$ against $\{\psi_n\}$. Suppose further that the tests $\{\varphi_n\}$ and $\{\psi_n\}$ are based on statistics $T_n=t_n(\mathbf{X})$ and $T_n^*=t_n^*(\mathbf{X})$, respectively, in the following manner:

$$\varphi_n(\mathbf{x})=\begin{cases}1 & \text{when } t_n(\mathbf{x})>c,\\ a & \text{when } t_n(\mathbf{x})=c,\\ 0 & \text{when } t_n(\mathbf{x})<c,\end{cases}$$

$$\psi_n(\mathbf{x})=\begin{cases}1 & \text{when } t_n^*(\mathbf{x})>c^*,\\ b & \text{when } t_n^*(\mathbf{x})=c^*,\\ 0 & \text{when } t_n^*(\mathbf{x})<c^*,\end{cases}$$

where $\mathbf{X}=(X_1,\ldots,X_n)$ and $\mathbf{x}=(x_1,\ldots,x_n)$. Put $\theta_0=0$ and $\theta_n=k/\sqrt{n}$ (k=constant) for simplicity. If T_n *and* T_n^* are asymptotically normal, then under some conditions e is given by the formula

$$e=\lim\frac{(dE_\theta(T_n)/d\theta\mid_{\theta=0})^2\sigma_0^2(T_n^*)}{(dE_\theta(T_n^*)/d\theta\mid_{\theta=0})^2\sigma_0^2(T_n)}.$$

As an example, consider a two-sample problem on a †location parameter. If the population distribution is normal and the Wilcoxon test is used to test the hypothesis of equality of means, then its asymptotic relative efficiency against Student's test is $3/\pi$. For the same problem, the asymptotic relative efficiencies of the Fisher-Yates-Terry normal score test and the van der Waerden test against Student's test are both unity. For the hypothesis of equality of means in the k-sample problem, the asymptotic relative efficiency of the Kruskal-Wallis test against the F-test is $3/\pi$, provided that the sample is distributed normally.

F. Kolmogorov-Smirnov Tests

Let $F_n(x)$ be the empirical distribution function based on a random sample of size n from $F_0(x)$. Set

$$d_n=\sup|F_n(x)-F_0(x)|,$$

$$D_n=\sup(F_n(x)-F_0(x)),$$

$$s_n=\sup_{a\leqslant F_0(x)}\left|\frac{F_n(x)-F_0(x)}{F_0(x)}\right|,$$

$$S_n=\sup_{a\leqslant F_0(x)}\frac{F_n(x)-F_0(x)}{F_0(x)}.$$

Then

$$\lim_{n\to\infty}P_r(d_n<z/\sqrt{n})=L(z)=\sum_{k=-\infty}^{\infty}(-1)^k e^{-2k^2z^2},$$

$$\lim_{n\to\infty}P_r(D_n<z/\sqrt{n})=K(z)=1-e^{-2z^2},$$

$$P_r(D_n<D)=1-D\sum_{j=0}^{[n(1-D)]}\binom{n}{j}\times\left(1-D-\frac{j}{n}\right)^{n-j}\left(D+\frac{j}{n}\right)^{j-1},$$

$$\lim_{n\to\infty}P_r(s_n<z/\sqrt{n})=\frac{4}{\pi}\sum_{k=0}^{\infty}(-1)^k\times(2k+1)^{-1}e^{-((2k+1)^2\pi^2/8)((1-a)/a)z^2},$$

$$\lim_{n\to\infty}P_r(S_n<z/\sqrt{n})=\sqrt{\frac{2}{\pi}}\int_0^{z\sqrt{a/(1-a)}}e^{-t^2/2}\,dt.$$

The statistics d_n, D_n, s_n, S_n are frequently used to test the hypothesis $F(x)=F_0(x)$. (This problem is called testing goodness of fit.)

In a two-sample problem, let $F_m(x)$ and $G_n(x)$ be two empirical distribution functions based on samples of sizes m and n from $F(x)$ and $G(x)$, respectively. Set

$$d_{m,n}=\sup|F_m(x)-G_n(x)|,$$

$$D_{m,n}=\sup(F_m(x)-G_n(x)).$$

If the hypothesis $F=G$ is true, we have

$$\lim P_r(d_{m,n}<z/\sqrt{N})=L(z),$$

$$\lim P_r(D_{m,n}<z/\sqrt{N})=K(z),$$

provided that m and n tend to ∞ so that $N=mn/(m+n)\to\infty$ and m/n is constant. Taking account of these facts, $d_{m,n}$ and $D_{m,n}$ are used to test the hypothesis $F=G$. The tests using the statistics d_n, D_n, $d_{m,n}$, $D_{m,n}$ are called **Kolmogorov-Smirnov tests.**

G. Interval Estimation

Let $(X_1,\ldots,X_n)$ be an independent random sample from the population with a distribution function $F(x-\theta)$, where θ is an unknown location parameter, and $(x_1,\ldots,x_n)$ be its observed value. Suppose that $F(x)$ is continuous and symmetric about the origin.

Using the statistic $S_n=S(x_1,\ldots,x_n)$ for the one-sample nonparametric test for testing the hypothesis $H:\theta=0$, a †confidence interval of θ is constructed as follows. For an appropriately given γ $(0<\gamma<1)$, select constants d_1 and d_2 in the range of $S(x_1,\ldots,x_n)$ that satisfy

$$P_0\{d_1<S(X_1,\ldots,X_n)<d_2\}=1-\gamma,$$

where P_0 means the probability under the hypothesis $H:\theta=0$. If there exist statistics $L_s(x_1,\ldots,x_n)$ and $U_s(x_1,\ldots,x_n)$ such that $L_s(x_1,\ldots,x_n)\leqslant\theta<U_s(x_1,\ldots,x_n)$ if and only if $d_1<S(x_1-\theta,\ldots,x_n-\theta)<d_2$ for all θ, then the confidence interval of θ with $100(1-\gamma)\%$ confidence coefficient is given by $(L_s(x_1,\ldots,x_n), U_s(x_1,\ldots,x_n))$. This interval is distribution-free, i.e., it holds that

$$P\{L_s(X_1,\ldots,X_n)<\theta<U_s(X_1,\ldots,X_n)\}=1-\gamma \quad \text{for all } F.$$

When S_n is the statistic for the Wilcoxon signed rank test, L_s and U_s are given by $L_s=W_{(M+1-d_2)}$ and $U_s=W_{(M-d_1)}$, where $M=n(n+1)/2$ and $W_{(1)}\leqslant\ldots\leqslant W_{(M)}$ are ordered values for M averages $(x_i+x_j)/2$ $(i\leqslant j=1,2,\ldots,n)$.

H. Point Estimation

Let $(X_1,\ldots,X_n)$ be an independent random sample from the population with a distribution function $F(x-\theta)$, where θ is an unknown location parameter, and let $(x_1,\ldots,x_n)$ be its observed value.

There are four methods of constructing robust estimators of θ. Let $X_{(1)}<\ldots<X_{(n)}$ be ordered values of $X_1,\ldots,X_n$. The first method is to use $T_n=a_1X_{(1)}+\ldots+a_nX_{(n)}$ for some given constants $a_1,\ldots,a_n$ such that $\sum_{i=1}^n a_i=1$. T_n is called the **L-estimator**. An example is

$$T_n(\alpha)=(pX_{([\alpha n]+1)}+X_{([\alpha n]+2)}+\ldots+pX_{(n-[\alpha n])})/n(1-2\alpha),$$

where $p=1+[\alpha n]-\alpha n$. This estimator is called the **α-trimmed mean**. Let J be a real-valued function such that $\int_0^1 J(t)dt=1$, and set $a_k=\int_{(k-1)/n}^{k/n}J(t)dt$; then as $n\to\infty$, T_n converges to $T(F)=\int_0^1 J(t)F^{-1}(t)dt$ in probability. Suppose that F is a distribution function having an †absolutely continuous density function f. Denote the derivative of f by f', and let $I(F)$ be the †Fisher information on θ. Set $\psi(t)=-f'(t)/f(t)$ and $J(t)=\psi'(F^{-1}(t))/I(F)$. Then Chernoff, J. L. Gastwirth, and M. V. Johns [10] proved that under some regularity conditions T_n is an †asymptotically efficient estimator of θ for F.

Let ρ be a real-valued (usually convex) function of a real parameter with derivative $\psi=\rho'$. The second method is to estimate θ by T_n by minimizing $\sum_{i=1}^n\rho(x_i-T_n)$ or by satisfying

$$\sum_{i=1}^n\psi(x_i-T_n)=0.$$

T_n is called the **M-estimator**. When $\rho(t)=t^2$, it agrees with the least squares estimator. Let $\Phi(x)$ be the distribution function of the standard normal distribution, $H(x)$ be an arbitrary continuous distribution function which is symmetric about the origin, $\mathscr{F}$ be a class of distribution functions of the form $F(x)=(1-\varepsilon)\Phi(x)+\varepsilon H(x)$ for a given ε $(0<\varepsilon<1)$, and $V(\rho,F)$ be the asymptotic variance of T_n. Huber [11] proved that ρ minimizing $\sup_{F\in\mathscr{F}}V(\rho,F)$ is given by $\rho_K(t)=t^2/2$, $K|t|-K^2/2$ defined for $|t|\leqslant K$, $>K$, respectively, for some constant K. Under quite general conditions, the M-estimator converges as $n\to\infty$ to $T(F)$ in probability, which is defined by $\int\psi(x-T(F))dF(x)=0$. If $\psi(t)=-f'(t)/f(t)$ is chosen for $\psi(x)$, then T_n is the †maximum likelihood estimator of θ for F and is asymptotically efficient under some regularity conditions. Generally, the M-estimator defined above is not scale invariant. A scale invariant version of the M-estimator is obtained by replacing the defining equation by

$$\sum_{i=1}^n\psi\left(\frac{x_i-T_n}{S_n}\right)=0, \tag{1}$$

where S_n is any robust estimate of scale, e.g., the median of $\{|x_i-M|/0.6745\}_{i=1,2,\ldots,n}$ where M is the sample median, or by solving the simultaneous equations (1) and

$$\sum_{i=1}^n\chi\left(\frac{x_i-T_n}{S_n}\right)=0$$

with respect to T_n and S_n. In the context of the maximum likelihood estimation, χ is chosen to be $\chi(t)=t\psi(t)-1$ (→ 399 Statistical Estimation P).

The third method employs nonparametric tests for testing the hypothesis $H:\theta=0$ against the alternative hypothesis $A:\theta>0$. Let J be a real-valued and nondecreasing function such that $\int_0^1 J(t)dt=0$ and $R_k^+(\theta)$ be the rank of $|X_k-\theta|$ among $|X_1-\theta|,\ldots,|X_n-\theta|$. Set $\Psi(t)=1,0$ according as $t>0$, $\leqslant 0$ and $S(X_1-\theta,\ldots,X_n-\theta)=\sum_{k=1}^n J((R_k^+(\theta)+n)/(2n+1))\Psi(X_k-\theta)$. Let

$$\theta^*=\sup\{\theta; S(X_1-\theta,\ldots,X_n-\theta)>\mu\},$$

$$\theta^{**}=\inf\{\theta; S(X_1-\theta,\ldots,X_n-\theta)<\mu\},$$

where μ is the expected value of $S(X_1,\ldots,X_n)$ under the hypothesis $H:\theta=0$. Then an estimator of θ is defined by $T_n=(\theta^*+\theta^{**})/2$. Hodges and Lehmann [12] first proposed this technique, and this estimator is called the **R-estimator**. When $F(x)$ is symmetric about the origin and $J(t)=t-\frac{1}{2}$, S tends to be the statistic for the Wilcoxon signed rank test, and the

R-estimator reduces to the median of $n(n+1)/2$ averages $(X_i+X_j)/2$ $(1 \leqslant i \leqslant j \leqslant n)$. As $n \to \infty$, the R-estimator converges in probability to $T(F)$, defined by

$$\int J\left(\frac{F(x)+1-F(2T(F)-x)}{2}\right)dF(x)=0.$$

For symmetric F, the R-estimator defined by the statistic S with $J(t)=-f'(F^{-1}(t))/f(F^{-1}(t))$ is asymptotically efficient under some regularity conditions.

Although the above three methods provide robust estimators, which are seldom affected by outlying observations or contamination by gross errors, their behavior still depends on F. The last method of constructing robust estimators consists of estimating θ adaptively by utilizing information on the shape of F. Among these, a striking one is the asymptotically fully efficient estimator for a wide class of F proposed by Takeuchi [13]. The estimator is constructed by using subsamples of size K $(K<n)$ drawn from the original sample, estimating the elements of the †covariance matrix of the order statistics by U-statistics, and selecting the best weights of the L-estimator. L. A. Jaeckel [14] made an α-trimmed mean adaptive estimator by selecting an α that minimizes the estimated asymptotic variance.

I. The Influence Curve

Let $T(F)$ be a functional of a distribution function F, and let an estimator T_n of θ calculated from an empirical distribution function F_n converge to $T(F)$ in probability as $n \to \infty$. A real-valued function $\mathrm{IC}(x; F, T)$ defined by

$$\mathrm{IC}(x;F,T)=\lim_{\varepsilon\to 0}\frac{T((1-\varepsilon)F+\varepsilon\delta_x)-T(F)}{\varepsilon} \qquad \text{for all } x$$

is called the **influence curve**, where δ_x is the distribution function of a point mass 1 at x. The curve was first introduced by Hampel [15] to study the stability aspect of estimators against a small change of F. As an example, when F is symmetric about the origin, the influence curve for the α-trimmed mean $\mathrm{IC}(x; F, T)$ is given by

$$\begin{array}{ll} F^{-1}(\alpha)/(1-2\alpha) & \text{when } x<F^{-1}(\alpha), \\ x/(1-2\alpha) & \text{when } F^{-1}(\alpha)\leqslant x < F^{-1}(1-\alpha), \\ F^{-1}(1-\alpha)/(1-2\alpha) & \text{when } x>F^{-1}(1-\alpha). \end{array}$$

By substituting the empirical distribution function F_n for F in $T(F)$, we can represent the robust estimators discussed in Section H, at least approximately, by $T_n=T(F_n)$. Under some conditions, it can be proved that as $n\to\infty$,

$$n^{1/2}(T(F_n)-T(F)-\frac{1}{n}\sum_{i=1}^{n}\mathrm{IC}(X_i;F,T))\to 0$$

in probability. Thus it follows that $n^{1/2}(T_n-T(F))$ is asymptotically normally distributed with asymptotic variance $\int(\mathrm{IC}(x;F,T))^2\,dF(x)$.

J. The Regression Problem

Consider the linear regression problem (→ 403 Statistical Models D)

$$X_i=\sum_{j=1}^{p}\theta_j a_{ij}+\varepsilon_i, \qquad i=1,2,\ldots,n,$$

where the X_i are observable variables, the θ_j are regression coefficients to be estimated, the a_{ij} are given constants, and $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are identically and independently distributed random errors whose distribution function is given by $F(x)$. The idea of the M-estimator is a direct generalization of the method of least squares; namely, to adopt $(\hat\theta_1,\ldots,\hat\theta_p)$ as an estimator for $(\theta_1,\ldots,\theta_p)$ that minimizes $\sum_{i=1}^{n}\rho(X_i-\sum_j\theta_j a_{ij})$ for some function ρ such as the one described above.

R-estimators of the regression coefficients are obtained by minimizing $\sum_{i=1}^{n}a_n(R_i)\Delta_i$, where $\Delta_i=X_i-\sum_j\theta_j a_{ij}$, R_i is the rank of Δ_i among $\Delta_1,\ldots,\Delta_n$, and $a_n(\cdot)$ is some monotone function satisfying $\sum_{i=1}^{n}a_n(i)=0$. It has been proved that minimizing $\sum_{i=1}^{n}a_n(R_i)\Delta_i$ is asymptotically equivalent to minimizing

$$\sum_{j=1}^{p}\left|\sum_{i=1}^{n}a_n(R_i)a_{ij}\right|,$$

the properties of which were first studied by J. Jurečková [16].

K. Dependence

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be random samples from a population with a bivariate distribution function, R_i be the rank of X_i among $X_1, \ldots, X_n$ when they are rearranged in an ascending order, and S_j be the rank of Y_j among $Y_1, \ldots, Y_n$ defined similarly as R_i. Various quantities are devised to measure the degree of dependence between X and Y.

(1) Spearman's Rank Correlation. Set $d_i=R_i-S_i$. Then $r_s=1-6\sum_i d_i^2/(n^3-n)$ is called **Spearman's rank correlation.** If there is no dependence between X and Y, i.e., if the X_i and Y_j are independent random variables, then $E(r_s)=0$ and $V(r_s)=(n-1)^{-1}$.

(2) Kendall's Rank Correlation. Take pairs (R_i, S_i) and (R_j, S_j). If $(R_j-R_i)(S_j-S_i)>0$, set

$\varphi_{ij}=1$; otherwise, $\varphi_{ij}=-1$. The statistic $r_k=\binom{n}{2}^{-1}\sum\varphi_{ij}$ is called **Kendall's rank correlation**, where $\sum$ runs over all possible pairs chosen from $(R_1, S_1), \ldots, (R_n, S_n)$.

If there is no dependence between X and Y, then $E(r_k)=0$ and $V(r_k)=2(2n+5)/(9n(n-1))$.

(3) Rankit Correlation. R_i and S_i, $i=1, \ldots, n$, are replaced by the corresponding normal scores, i.e., the means of order statistics in an independent sample of size n from $N(0,1)$; then the usual †sample correlation coefficient is calculated from these scores. This correlation coefficient r_R is called rankit correlation; and if there is no dependence between X and Y, then $E(z)=0$, $V(z)=(n-3)^{-1}$, asymptotically, where $z=\frac{1}{2}\log\frac{1+r_R}{1-r_R}$ (†Fisher's z-transformation).

These statistics are used to test the hypothesis of independence.

References

[1] J. Hájek and Z. Šidák, Theory of rank tests, Academic Press, 1967.
[2] M. L. Puri and P. K. Sen, Nonparametric methods in multivariate analysis, Wiley, 1971.
[3] R. H. Randles and D. A. Wolfe, Introduction to the theory of nonparametric statistics, Wiley, 1979.
[4] P. J. Huber, Robust statistics: A review, Ann. Math. Statist., 43 (1972), 1041–1067.
[5] P. J. Huber, Robust regression: Asymptotics, conjectures and Monte Carlo, Ann. Statist., 1 (1973), 799–821.
[6] P. J. Huber, Robust statistics, Wiley, 1981.
[7] R. V. Hogg, Adaptive robust procedures: A partial review and some suggestions for future applications and theory, J. Amer. Statist. Assoc., 69 (1974), 909–923.
[8] D. F. Andrews, P. J. Bickel, F. R. Hampel, P. J. Huber, W. H. Rogers, and J. W. Tukey, Robust estimates of location: Survey and advances, Princeton Univ. Press, 1972.
[9] H. Chernoff and I. R. Savage, Asymptotic normality and efficiency of certain nonparametric test statistics, Ann. Math. Statist., 29 (1958), 972–994.
[10] H. Chernoff, J. L. Gastwirth, and M. V. Johns, Asymptotic distribution of linear combinations of functions of order statistics with application to estimation, Ann. Math. Statist., 38 (1967), 52–72.
[11] P. J. Huber, Robust estimation of a location parameter, Ann. Math. Statist., 35 (1964), 73–101.
[12] J. L. Hodges and E. L. Lehmann, Estimates of location based on rank tests, Ann. Math. Statist., 34 (1963), 598–611.
[13] K. Takeuchi, A uniformly asymptotically efficient estimator of a location parameter, J. Amer. Statist. Assoc., 66 (1971), 292–301.
[14] L. A. Jaeckel, Some flexible estimates of location, Ann. Math. Statist., 42 (1971), 1540–1552.
[15] F. R. Hampel, The influence curve and its role in robust estimation, J. Amer. Statist. Assoc., 69 (1974), 383–393.
[16] J. Jurečková, Nonparametric estimate of regression coefficients, Ann. Math. Statist., 42 (1971), 1328–1338.

372 (XXI.3) Roman and Medieval Mathematics

The Romans were interested in mathematics for everyday use; their arithmetic consisted of computation (by means of the **abacus**), weights and measures, and money. For their monetary system, they developed a computational method using duodecimal fractions. Julius Caesar (102?–44 B.C.), known for his calendar reform in 46 B.C., also undertook to measure his territory, which aroused a demand for land surveying techniques. Books on practical geometry which provided this knowledge were called *gromatics* (a "groma" was a land surveying instrument). Toward the end of the Western Roman Empire (476), Greek mathematics was studied; during this period Boethius (c. 480–524) wrote his two books on arithmetic and geometry. The former was a summarized translation of Nichomachus' book, and the latter included propositions from the first three books of Euclid's *Elements* (without proof) and practical geometry.

Music, astronomy, geometry, and arithmetic, which constituted the "mathemata" of Plato's Academy (closed in 529), were treated as the "quadrivium" (the four major subjects) in the *Encyclopedia* of Martianus Capella, Cassiodorus, Isidorus, Hispalensis, and others. After the establishment of the Roman Church in the 5th century, the quadrivium was to be studied for the glory of God. Books on mathematics from this period laid emphasis on the computation of an ecclesiastical calendar and mystical interpretations of integers, as seen in books by Bede Venerabilis, Alcuin, and Maurus from the 7th through 9th centuries.

Arabian science was imported first through Spain—under Moorish influence beginning in 711, the year of the fall of the Visigoths—and then through the Crusades (1096–1270). Computation with figures, originating in India, replaced the abacus in the 12th century, when

Arabian books on arithmetic and algebra, along with Greek books on geometry and astronomy (such as books by Euclid and Ptolemy), were translated into Latin. Italian merchants, whose occupation necessitated computation, rapidly adopted the new system. Representative books of this period are *Liber abaci* (1202) and *Practica geometrica* (1220) by Leonardo da Pisa (also known as Fibonacci, c. 1170–1250). The former includes the four arithmetic operations, showing Indian influence, commercial arithmetic, and algebra. The new methods were not limited to merchants. The French bishop Nicole Oresme (c. 1323–1382), who influenced Leonardo da Vinci (1452–1519), introduced fractional exponents and conceived the graphic representation of temperature, a precursor of coordinates and functions.

From the 11th century, universities developed from theological seminaries, first in Italian cities such as Bologna and Palermo, and later in Paris, Oxford, and Cambridge. Mathematics was taught in these universities, although no remarkable creative contributions were made. However, theologians such as Albertus Magnus (c. 1193–1280) and Thomas Aquinas (1225?–1274) discussed infinity in a way that went beyond Greek thought and thus helped to lay a basis for the modern philosophy of mathematics.

References

[1] M. Cantor, Vorlesungen über Geschichte der Mathematik, Teubner, I, 1880; II, 1892.
[2] M. Clagett, Greek science in antiquity, Abelard-Schuman, 1955.
[3] M. Clagett, Archimedes in the Middle Ages I, II, Univ. of Wisconsin Press, 1964.
[4] M. Clagett, The science of mechanics in the Middle Ages, Univ. of Wisconsin Press, 1959.
[5] G. Sarton, Introduction to the history of science I, From Homer to Omar Khayyâm, Carnegie Institution, 1927.

S

373 (XVIII.13) Sample Survey

A. General Remarks

The sample survey is a means of getting statistical information about a certain aggregate from the observation of some but not all of it. The aggregate concerned is usually called the **finite population**, and the observed part is called the **sample**. Introducing a random mechanism, J. Neyman established a method of ascertaining objectively the reliability of such information. This method is mainly applied to demographic statistical surveys and opinion polls, but it is also applicable to random samples of physical materials. We briefly sketch the mathematical structure of this method without going into detail about technical problems that arise in the practical survey situation.

Suppose that the population consists of N units, where N is called its **size**. Each unit has some characteristic α, which is an element of some set Ω. The set of all characteristics of the units in the population is designated by $\theta = \{\alpha_1, \ldots, \alpha_N\}$, which we regard as a parameter. The set of all possible θ is denoted by $\Theta \subset \Omega^N$.

Suppose that one unit is chosen and observed according to some procedure, the index number of the unit in the population is J, and the observed characteristic is X. It is assumed that the observation is without error, hence $X = \alpha_J$.

Denote the whole sample by $(\mathbf{J}, \mathbf{X}) = (J_1, \ldots, J_n, X_1, \ldots, X_n)$, where n is the **sample size** and $X_i = \alpha_{J_i}$. The sample size n may be a random variable, and among $J_1, \ldots, J_n$, duplication may be allowed. The probabilistic scheme for $\mathbf{J}$ is called the **sampling procedure**, and if it satisfies the condition

(c) $\Pr\{J_i = j\}$ is independent of α_{J_i} and of $X_{i+1}, \ldots, X_n$ (but may depend on $X_1, \ldots, X_{i-1}$),

it is called a **random sampling procedure**. Moreover, if the †joint distribution of $\mathbf{J}$ is independent of θ, it is called **regular**. Specifically, if n is constant and $\Pr\{\mathbf{J}\}$ is symmetric in $\mathbf{J}$, it is called **uniform**.

The two main mathematical problems of sample surveys are to determine a random sampling procedure and to provide methods whereby statistical inferences can be made concerning θ (→ 401 Statistical Inference).

B. The Problem of Inference

Condition (c) is assumed. The probability of $(\mathbf{J}, \mathbf{X})$ is given by

$$\Pr\{(\mathbf{J}, \mathbf{X}) = (j_1, \ldots, j_n, X_1, \ldots, X_n)\}$$
$$= \Pr\{J_1 = j_1\}\chi_\theta(X_1)\Pr\{J_2 = j_2 \mid J_1, X_1\}\chi_\theta(X_2)$$
$$\ldots \Pr\{J_n = j_n \mid J_1, X_1, \ldots, J_{n-1}, X_{n-1}\}\chi_\theta(X_n),$$

where $\chi_\theta(X_i)$ is defined as 1 if $X_i = \alpha_{J_i}$ and 0 otherwise. This formula can be shortened to the form

$$\Pr\{(\mathbf{J}, \mathbf{X})\} = P(\mathbf{J}, \mathbf{X})\chi_\theta(\mathbf{X}, \mathbf{J}), \tag{1}$$

where

$$\chi_\theta(\mathbf{X}, \mathbf{J}) = \begin{cases} 1 & \text{if } X_i = \alpha_{J_i}, \ i = 1, \ldots, n, \\ 0 & \text{otherwise.} \end{cases}$$

Expression (1) is the fundamental model for the sample survey problem. Note that $P(\mathbf{J}, \mathbf{X})$ is independent of the parameter θ. Therefore if we let $\mathbf{I} = (I_1, \ldots, I_m)$ $(I_1 < \ldots < I_m)$ be the set of numbers in $(J_1, \ldots, J_n)$ after deleting duplications, and let $\mathbf{Y} = (Y_1, \ldots, Y_m)$ be the corresponding X values, then the joint distribution of $\mathbf{I}$ and $\mathbf{Y}$ can also be expressed as

$$\Pr\{(\mathbf{I}, \mathbf{Y})\} = P(\mathbf{I}, \mathbf{Y})\chi_\theta(\mathbf{Y}, \mathbf{I}),$$

where

$$\chi_\theta(\mathbf{Y}, \mathbf{I}) = \begin{cases} 1 & \text{if } Y_j = \alpha_{I_j}, j = 1, \ldots, m, \\ 0 & \text{otherwise.} \end{cases}$$

Since $\chi_\theta(\mathbf{Y}, \mathbf{I}) = \chi_\theta(\mathbf{X}, \mathbf{J})$ for all θ, the †conditional probability distribution of $(\mathbf{J}, \mathbf{X})$ for given $(\mathbf{I}, \mathbf{Y})$ is independent of $\chi_\theta(\mathbf{X}, \mathbf{J})$, and hence $(\mathbf{J}, \mathbf{X})$ is a †sufficient statistic. According to the general theory of sufficient statistics, we can restrict ourselves to the class of procedures depending only on $(\mathbf{I}, \mathbf{Y})$.

C. Estimation

Suppose that $g(\theta) = g(\alpha_1, \ldots, \alpha_N)$ is a real parameter whose unbiased estimators are under discussion.

Theorem. There exists an unbiased estimator of $g(\theta)$ if and only if there exists a decomposition

$$g(\theta) = \sum h_\nu(\alpha_{j_1(\nu)}, \ldots, \alpha_{j_n(\nu)}),$$
$$\Pr\{\mathbf{I} = (j_1(\nu), \ldots, j_n(\nu))\} > 0, \quad \nu = 1, 2, \ldots. \tag{2}$$

If the sampling procedure is regular, the second condition can be replaced by $\Pr\{\mathbf{I} \supset (j_1(\nu), \ldots, j_n(\nu))\} > 0$, $\nu = 1, 2, \ldots$. Hence, for example, if α_i is real and the sampling procedure is regular, $\bar{\alpha} = \sum \alpha_i / N$ is †estimable if and only if $\Pr\{\mathbf{I} \ni i\} > 0$ for all i, and $\sigma_\alpha^2 = \sum(\alpha_i - \bar{\alpha})^2/(N-1) = \sum\sum(\alpha_i - \alpha_j)^2/N(N-1)$ is estimable if and only if $\Pr\{\mathbf{I} \ni i, j\} > 0$ for all i and j. Also, $\prod_{i=1}^N \alpha_i$ is not estimable unless $\Pr\{\mathbf{I} = (1, \ldots, N)\} > 0$. The decomposition (2) is not unique, and corresponding to different decompositions, different unbiased estimators are derived. Also, for the case of regular sampling

procedures, for any $\theta=\theta_0$ it is always possible to construct an unbiased estimator $g(\theta)$ such that $\Pr\{\hat{g}(\theta)\equiv g(\theta_0)\mid\theta=\theta_0\}=1$ if $g(\theta)$ is estimable. Hence the variance of the locally best unbiased estimator is always 0, and the †uniformly minimum variance unbiased estimators exist only in the trival case.

If some kind of symmetry exists among the population units as well as the sampling procedure and the parameter, it would be natural to require the same kind of symmetry for the estimators. Let G be a group of permutations among N numbers. Assume that for any $\theta\in\Theta$ and $\gamma\in G$, we have $\gamma\theta\in\Theta$ and $g(\gamma\theta)=g(\theta)$. If $\Pr\{\gamma\mathbf{J}\}=\Pr\{\mathbf{J}\}$ for all γ, then the sampling procedure is said to be **invariant** with respect to G. An estimator is also called invariant if its value does not change under any permutation $\gamma\in G$ of the numbers of sample units. Thus if G is the set of all permutations (i.e., the †symmetric group), then the invariant estimator is a function of $\mathbf{Y}$ (or $\mathbf{X}$) only. Moreover, if the dimension m of $\mathbf{Y}$ is constant, $\mathbf{Y}$ is complete (under some mild conditions); hence there exists a unique minimum variance invariant unbiased estimator.

When there is some additional information, it can be represented by **auxiliary variables** $\beta_1,\ldots,\beta_N$, which are known and assumed to have some relation to $\alpha_1,\ldots,\alpha_N$. Assume that the α_i are real numbers and that the parameter to be estimated is $\theta=\bar{\alpha}=(\sum\alpha_i)/N$. If we can assume that α_i and β_i are approximately proportional, we can estimate the unknown population mean $\bar{\alpha}$ by $\hat{\bar{\alpha}}^*=(\bar{X}/\bar{Z})\times\bar{\beta}$ where $\bar{X}$ is the sample mean of the α's and $\bar{Z}$ the sample mean of the β's. Although $\hat{\bar{\alpha}}^*$ is not unbiased, we may expect that it has small error if the relation between two variables is close. $\hat{\bar{\alpha}}^*$ is usually called the ratio estimator.

In practical research, as an estimator of the population total $A=\sum\alpha_i$, we usually use $\hat{A}=\sum X_i/P_i$, where $P_i=\Pr\{\mathbf{J}\ni i\}$. Its variance is $V(\hat{A})=\sum\sum(P_iP_j-P_{ij})(\alpha_i/P_i-\alpha_j/P_j)^2$ and is estimated by $v(\hat{A})=\sum\sum\{(P_iP_j-P_{ij})/P_{ij}\}(X_i/P_i-X_j/P_j)^2$, where $P_{ij}=\Pr\{\mathbf{J}\supset(i,j)\}$. When N is unknown, it can be estimated by the same procedure as $\hat{A}$ (say $\hat{N}$), and the population mean $\bar{\alpha}$ can be estimated by $\hat{\bar{\alpha}}=\hat{A}/\hat{N}$, which is called a ratio estimator. $\hat{\bar{\alpha}}$ is biased except when N is known.

D. Asymptotic Confidence Intervals

It is usually impossible to obtain any meaningful †confidence interval based on exact small-sample theory. But when the α_i are real and the sampling procedure is uniform and without replacement, the sample mean $\bar{X}$ is asymptotically normal with mean $\bar{\alpha}$ and variance $\frac{1}{n}\left(1-\frac{n}{N}\right)\sigma_\alpha^2$, where $\sigma_\alpha^2=\frac{1}{N-1}\sum(\alpha_i-\bar{\alpha})^2$, as N and $n\to\infty$, and $\limsup n/N<1$, provided that

$$\limsup_{N\to\infty}\frac{\max_{1<i\leqslant N}(\alpha_i-\bar{\alpha})^2}{\sum(\alpha_i-\bar{\alpha})^2}=0.$$

Also, the sample variance converges to σ_α^2 as $n\to\infty$. From these results we can construct asymptotic confidence intervals for $\bar{\alpha}$.

E. The Problem of Sampling Procedures

In determining the sampling procedures, both the technical aspects of sampling and the accuracy of the estimators should be considered. The most commonly used methods are **multistage sampling** and **stratified sampling**, or some combination of the two. For example, the population is partitioned into several clusters. First we select some of them according to a probability scheme and then choose units from the selected clusters. This procedure is called **two-stage sampling**. The probabilities for the selection of clusters may be uniform or proportional to the size of the clusters. **Stratified sampling** is the method of dividing the population into several subpopulations, called **strata**, and selecting the sample units within each stratum. If the size of the ith stratum is N_i, the size of the sample chosen from this stratum is n_i, and the probability is uniform within each stratum, then the most common estimator for the population mean $\bar{\alpha}$ is given by

$$\hat{\bar{\alpha}}=\sum N_i\bar{X}_i/N,$$

where $\bar{X}_i$ is the mean of the sample values in the ith stratum. The variance of $\hat{\bar{\alpha}}$ is given by

$$V(\hat{\bar{\alpha}})=\left(\frac{1}{N}\right)^2\sum\frac{N_i^2}{n_i}\left(1-\frac{n_i}{N_i}\right)\sigma_i^2,$$

where σ_i^2 is the population variance within the ith stratum.

If the cost of drawing one sample unit in the ith stratum is equal to c_i, then for fixed cost, the variance of the estimator is minimized when

$$n_i/N_i\propto\sigma_i/\sqrt{c_i},$$

which is called the condition of **optimum allocation**.

F. Replicated Sampling Plan

W. E. Deming proposed an effective method in practical sample surveys, called a replicated sampling plan, following J. W. Tukey's hint. It enables us to easily evaluate variances of esti-

mates for any estimator and any sampling procedure. Let the sample be composed of k subordinate samples which are selected by the same random sampling procedure from the same population, and let $\hat{\theta}_i(\mathbf{J}_i, X_i)$ be the estimate from the ith subordinate sample by the same estimator and $\hat{\theta}$ be the estimate from the whole sample by the same estimator. Then, provided $\hat{\theta}=\sum\hat{\theta}_i/k$ approximately, the variance of $\hat{\theta}$ can be estimated by $v(\hat{\theta})=\sum(\hat{\theta}_i-\hat{\theta})^2/k(k-1)$. If the sample is selected by the simple random sampling procedure and is of large scale, $\hat{\theta}_i$ and $\hat{\theta}$ are approximately normal variates, and $v(\theta)$ is evaluated by using the sample range of the $\hat{\theta}_i$. In large-scale sample surveys, even when the random sampling procedure is not simple, the theory related to the normal distribution can be applied to the $\hat{\theta}_i$ and $\hat{\theta}$. It has been shown that $v(\hat{\theta})$ evaluated by the above method includes not only the sampling error but also the random part of the nonsampling errors.

G. Conceptual Problems

Although it has been established that the sample survey method is useful in large-scale social or economic surveys, there are difficult conceptual problems about the foundations of the method (especially when auxiliary information exists) that are still far from being settled.

References

[1] W. G. Cochran, Sampling techniques, Wiley, third edition, 1977.
[2] W. E. Deming, On simplifications of sampling design through replication with equal probabilities and without stages, J. Amer. Statist. Assoc., 51 (1956), 24–53.
[3] R. J. Jessen, Statistical survey techniques, Wiley, 1978.
[4] J. Neyman, Contribution to the theory of sampling human populations, J. Amer. Statist. Assoc., 33 (1938), 101–116.
[5] P. V. Sukhatme and B. V. Sukhatme, Sampling theory of surveys with applications, Asia Publishing House, second edition, 1970.
[6] K. Takeuchi, Some remarks on general theory for unbiased estimation of a finite population, Japan. J. Math., 35 (1966), 73–84.

374 (XVIII.4) Sampling Distributions

A. General Remarks

To perform statistical inference, it is necessary to find the †probability distribution of a †statistic involved in the situation (→ 396 Statistic, 401 Statistical Inference). In general, the probability distribution of a statistic is called the **sampling distribution**. A set $\{X_1, X_2, \ldots, X_n\}$ of †random variables that are independently and identically distributed according to a distribution F is called a **random sample** from F. A common sampling distribution described in this article is that of the statistic $Y=f(X_1, \ldots, X_n)$, where the set $\{X_1, \ldots, X_n\}$ is a random sample from a †normal distribution. Examples of such a statistic Y of dimension 1 are the †sample mean, †sample variance, linear or quadratic forms of $\{X_1, \ldots, X_n\}$, their ratios, and †order statistic, while examples of Y of higher dimensions are the sample mean vector and the sample covariance matrix and its eigenvalues. The †normal distribution with †mean μ and †variance σ^2 is denoted by $N(\mu, \sigma^2)$, while the †p-dimensional (p-variate) normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ is denoted by $N(\boldsymbol{\mu}, \Sigma)$ (→ Appendix A, Table 22).

B. Samples from Univariate Normal Distributions

If random variables $X_1, \ldots, X_n$ are distributed independently according to $N(\mu_1, \sigma_1^2), \ldots, N(\mu_n, \sigma_n^2)$, then a linear form $\sum_i a_i X_i$ has the distribution $N(\sum_i a_i\mu_i, \sum_i a_i^2\sigma_i^2)$. In particular, if $\{X_1, \ldots, X_n\}$ is a random sample from $N(\mu, \sigma^2)$, then the sample mean $\bar{X}=\sum_i X_i/n$ has the distribution $N(\mu, \sigma^2/n)$.

Let $\{X_1, \ldots, X_n\}$ be a random sample from the distribution $N(0, 1)$. The sampling distribution of the statistic $Y=\sum_i X_i^2$ is called the **chi-square distribution** with n **degrees of freedom** and is denoted by $\chi^2(n)$. It has the †probability density

$$f_n(y)=2^{-n/2}\left(\Gamma\left(\frac{n}{2}\right)\right)^{-1} y^{-1+n/2}e^{-y/2}$$

for $0<y<\infty$, $f_n(y)=0$ elsewhere, where Γ is the †gamma function. The distribution of $Y=\sum_{i=1}^n(X_i+\mu_i)^2$ depends only on n and $\lambda=\sum_i\mu_i^2$, and is called the **noncentral chi-square distribution** with n **degrees of freedom** and **noncentrality** λ and denoted by $\chi^2(n, \lambda)$. It has the probability density

$$f_{n,\lambda}(y)=\sum_{k=0}^{\infty} e^{-\lambda/2}\left(\frac{\lambda}{2}\right)^k\frac{1}{k!}f_{n+2k}(y)$$

$$=e^{-\lambda/2}{}_0F_1\left(\frac{n}{2};\frac{\lambda y}{4}\right)f_n(y)$$

for $0<y<\infty$, where f_{n+2k} and f_n are the probability densities of chi-square distributions and ${}_0F_1$ is an extended hypergeometric func-

tion. Noncentral chi-square distributions have the following †reproducing property: If $Y_1, \ldots, Y_k$ are distributed independently according to $\chi^2(n_1, \lambda_1), \ldots, \chi^2(n_k, \lambda_k)$, then $\sum_i Y_i$ has the distribution $\chi^2(\sum_i n_i, \sum_i \lambda_i)$. Also, we have **Cochran's theorem** (*Proc. Cambridge Philos. Soc.*, 30 (1934)): If $X_1, \ldots, X_n$ are distributed independently according to $N(\mu_1, 1), \ldots, N(\mu_n, 1)$ and if for quadratic forms $Q_m = \sum_i \sum_j a_{ij}^{(m)} X_i X_j$ for $m = 1, \ldots, k$ the matrices $A_m = (a_{ij}^{(m)})$ with †rank r_m satisfy the condition $A_1 + \ldots + A_k = I$ (unit matrix), then a necessary and sufficient condition for $Q_1, \ldots, Q_k$ to have independent noncentral chi-square distributions with $r_1, \ldots, r_k$ degrees of freedom, respectively, is that $r_1 + \ldots + r_k = n$. In particular, when $\mu_i = 0$ for all i, they have (central) chi-square distributions, and the theorem implies their reproducing property.

Let X and Y be independent random variables having distributions $N(\delta, 1)$ and $\chi^2(n)$, respectively. Then the sampling distribution of $T = X/\sqrt{Y/n}$ is called the **noncentral *t*-distribution** with n **degrees of freedom** and **noncentrality** δ and is denoted by $t(n, \delta)$. Its probability density is given by

$$f_{n,\delta}(t) = \sum_{k=0}^{\infty} e^{-\delta^2/2} \delta^k \frac{2^{k/2}}{k!} \frac{\Gamma((n+k+1)/2)}{\sqrt{\pi n}\,\Gamma(n/2)} \times \frac{(t/\sqrt{n})^k}{(1+t^2/n)^{(n+k+1)/2}}$$

for $-\infty < t < \infty$. In particular, when $\delta = 0$, the distribution is called the ***t*-distribution** with n **degrees of freedom** and is denoted by $t(n)$. Its probability density is simplified to

$$f_n(t) = \frac{\Gamma((n+1)/2)}{\sqrt{\pi n}\,\Gamma(n/2)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}$$

for $-\infty < t < \infty$.

Let $\{X_1, \ldots, X_n\}$ be a random sample from $N(\mu, \sigma^2)$. Exact sampling distributions of the †sample variance $S^2 = \sum (X_i - \bar{X})^2/(n-1)$ and of the ***t*-statistic** $T = \sqrt{n}(\bar{X} - \mu_0)/\sqrt{S^2}$, where $\mu = \mu_0$ is a given number, were essentially obtained by Student [5]: $(n-1)S^2/\sigma^2$ and T are distributed according to $\chi^2(n-1)$ and $t(n-1)$, respectively. His proof was made rigorous by R. A. Fisher (*Metron*, 5 (1925)), who proved in particular that $\bar{X}$ and S^2 are independent. If $\mu \neq \mu_0$, then T follows the distribution $t(n-1, \sqrt{n}(\mu - \mu_0)/\sigma)$.

Let X and Y be distributed independently according to $\chi^2(m, \lambda)$ and $\chi^2(n)$, respectively. The distribution of $Z = (X/m)/(Y/n)$ is called the **noncentral *F*-distribution** with m and n **degrees of freedom** and **noncentrality** λ. In the special case when $\lambda = 0$ it is called the ***F*-distribution** with m and n **degrees of freedom** and is denoted by $F(m, n)$. The probability densities $f_{m,n,\lambda}$ and $f_{m,n}$ of these distributions are given by

$$f_{m,n}(z) = \frac{(m/n)^{m/2}}{B(m/2, n/2)} \frac{z^{(m/2)-1}}{(1+mz/n)^{(m+n)/2}},$$

$$f_{m,n,\lambda}(z) = \sum_{k=0}^{\infty} e^{-\lambda/2} \left(\frac{\lambda}{2}\right)^k \frac{1}{k!} \frac{(m/n)^{(m/2)+k}}{B((m/2)+k, n/2)} \times \frac{z^{(m/2)+k-1}}{(1+mz/n)^{((m+n)/2)+k}}$$

$$= e^{-\lambda/2}\, {}_1F_1\left(\frac{m+n}{2}; \frac{m}{2}; \frac{\lambda}{2} \frac{mz/n}{1+mz/n}\right) f_{m,n}(z)$$

for $0 < z < \infty$, where B and ${}_1F_1$ are the †beta function and the confluent hypergeometric function (→ 167 Functions of Confluent Type), respectively.

Let X be a random variable having the distribution $F(m, n)$. The distribution of $Z = \frac{1}{2}\log X$ is called the ***z*-distribution** with m and n **degrees of freedom** and is denoted by $z(m, n)$. Its probability density is given by

$$\frac{2(m/n)^{m/2}}{B(m/2, n/2)} \frac{e^{mz}}{(1+me^{2z}/n)^{(m+n)/2}}$$

for $-\infty < z < \infty$. If $S_1^2 = \sum_{i=1}^m (X_i - \bar{X})^2/(m-1)$ and $S_2^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2/(n-1)$ are sample variances based on independent samples of sizes m and n taken from $N(\mu, \sigma^2)$ and $N(\nu, \tau^2)$, respectively, then the statistic $z = \frac{1}{2}\log(S_1^2/S_2^2)$, which was introduced by Fisher (*Proc. Int. Math. Congress*, 1924), is distributed according to $z(m-1, n-1)$ under the hypothesis $\sigma^2 = \tau^2$. Fisher [6] tabulated percent points of $z(m, n)$.

C. Samples from Multivariate Normal Distributions

Let $\mathbf{X}$ be a p-dimensional random vector, namely, a vector having real random variables as its components. $\mathbf{X}$ has the p-variate normal distribution $N(\boldsymbol{\mu}, \Sigma)$ if and only if for any real vector $\mathbf{a} = (a_1, \ldots, a_p)'$, the random variable $\mathbf{a}'\mathbf{X}$ has the normal distribution $N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\Sigma\mathbf{a})$. If $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are independent and have p-variate normal distributions $N(\boldsymbol{\mu}_1, \Sigma_1), \ldots, N(\boldsymbol{\mu}_n, \Sigma_n)$, respectively, and if $A_1, \ldots, A_n$ are $m \times p$ real matrices, then the random vector $A_1\mathbf{X}_1 + \ldots + A_n\mathbf{X}_n$ has the m-variate normal distribution $N(\boldsymbol{\mu}, \Sigma)$, where $\boldsymbol{\mu} = \sum_{j=1}^n A_j \boldsymbol{\mu}_j$ and $\Sigma = \sum_{j=1}^n A_j \Sigma_j A_j'$.

Suppose that $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ is a random sample from the p-variate normal distribution $N(\mathbf{0}, \Sigma)$, and let $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ be a $p \times n$ matrix. Then the probability distribution of $W = \mathbf{X}\mathbf{X}'$ is called the **Wishart distribution** with **scale matrix** Σ and n **degrees of freedom** and is denoted by $W_p(\Sigma, n)$ or simply $W(\Sigma, n)$. If $n > p - 1$, the joint probability density function

of $p(p+1)/2$ arguments of $W=(W_{ij})$ is

$$f_{\Sigma,n}(W)=(\Gamma_p(n/2)|2\Sigma|^{n/2})^{-1} \times \mathrm{etr}(-\Sigma^{-1}W/2)|W|^{(n-p-1)/2}$$

for $W>0$, where $W>0$ means that W is †positive definite, $\mathrm{etr}(A)=\exp(\mathrm{tr}\,A)$, and Γ_p, a **multidimensional gamma function**, is defined as

$$\Gamma_p(a)=\pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma\left(a-\frac{1}{2}(i-1)\right)$$

for $a>(p-1)/2$. When $n\leqslant p-1$ the distribution is **singular** and has no probability density.

Suppose that $\mathbf{X}_1,\dots,\mathbf{X}_n$ are independent and obey normal distributions $N(\boldsymbol{\mu}_1,\Sigma)$, $\dots,N(\boldsymbol{\mu}_n,\Sigma)$, and let $\mathbf{X}=(\mathbf{X}_1,\dots,\mathbf{X}_n)$, $M=(\boldsymbol{\mu}_1,\dots,\boldsymbol{\mu}_n)$. Then the distribution of $W=\mathbf{X}\mathbf{X}'$ is called the p**-dimensional noncentral Wishart distribution** with **scale matrix** Σ, n **degrees of freedom**, and **noncentrality matrix** $\Omega=\Sigma^{-1}MM'$ and is denoted by $W(\Sigma,n,\Omega)$. If $n>p-1$, the probability density function is

$$\mathrm{etr}(-\Omega/2)_0F_1(n/2;\Omega\Sigma^{-1}W/4)f_{\Sigma,n}(W)$$

for $W>0$. ${}_\alpha F_\beta$ is a hypergeometric function with matrix argument, which is defined by ${}_\alpha F_\beta(a_1,\dots,a_\alpha;b_1,\dots,b_\beta;S)={}_\alpha F_\beta^{(p)}(a_1,\dots,a_\alpha;b_1,\dots,b_\beta;S,I)$, where

$${}_\alpha F_\beta^{(p)}(a_1,\dots,a_\alpha;b_1,\dots,b_\beta;S,T)=\sum_{k=0}^{\infty}\sum_{\kappa}\left(\frac{\prod_{i=1}^{\alpha}(a_i)_\kappa}{\prod_{j=1}^{\beta}(b_j)_\kappa}\right)\frac{C_\kappa(S)C_\kappa(T)}{k!C_\kappa(I_p)},$$

$\kappa=(k_1,\dots,k_p)$ is an ordered set of integers such that $k_1+\dots+k_p=k$ and $k_1\geqslant\dots\geqslant k_p\geqslant 0$, and where $C_\kappa(S)$ is a **zonal polynomial** (→ [8]) of a symmetric matrix S. The multivariate hypergeometric coefficient $(a)_\kappa$ is given by

$$(a)_\kappa=\prod_{i=1}^{p}\left(a-\frac{1}{2}(i-1)\right)_{k_i},$$

$$(a)_k=a(a+1)\dots(a+k-1).$$

The noncentral Wishart distribution is singular when $n\leqslant p-1$. Similarly to the noncentral chi-square distribution, the noncentral Wishart distribution has the reproducing property with respect to both the number of degrees of freedom and the noncentrality matrix. Also, Cochran's theorem can be extended to the multivariate case: Let $\mathbf{X}_1,\dots,\mathbf{X}_n$ be p-variate random vectors independently distributed according to $N(\boldsymbol{\mu}_1,\Sigma),\dots,N(\boldsymbol{\mu}_n,\Sigma)$, respectively, and let $A_m=(a_{ij}^{(m)})$, $m=1,\dots,k$, be $p\times p$ real matrices of †rank r_m and such that $A_1+A_2+\dots+A_k=I$ (unit matrix). A necessary and sufficient condition for random matrices $\mathbf{Q}_m=\sum_{i,j}a_{ij}^{(m)}\mathbf{X}_i\mathbf{X}_j'$, $m=1,\dots,k$, to be independently distributed according to noncentral Wishart distributions with $r_1,\dots,r_k$ degrees of freedom, respectively, is that $r_1+\dots+r_k=n$. If, in particular, $\boldsymbol{\mu}_1=\dots=\boldsymbol{\mu}_n=0$ and if $r_m\geqslant p$, then $\mathbf{Q}_m$ is distributed according to $W(\Sigma,r_m)$.

If $\mathbf{W}$ has the distribution $W(\Sigma,n)$ with $n>p-1$, then the eigenvalues $\lambda_1,\dots,\lambda_p$ ($\lambda_1\geqslant\dots\geqslant\lambda_p>0$) of W have the joint probability density function $C_{p,n}|\Sigma|^{-n/2}{}_0F_0^{(p)}(-\Sigma^{-1}/2,\Lambda)|\Lambda|^{(n-p-1)/2}\prod_{i<j}(\lambda_i-\lambda_j)$, where Λ is a diagonal matrix with diagonal elements $\lambda_1,\dots,\lambda_p$ and $C_{p,n}=\pi^{p^2/2}(2^{pn/2}\Gamma_p(p/2)\Gamma_p(n/2))^{-1}$. If $\Sigma=I$, then the joint probability density function becomes

$$C_{p,n}\mathrm{etr}(-\Lambda/2)|\Lambda|^{(n-p-1)/2}\prod_{i<j}(\lambda_i-\lambda_j).$$

Suppose that $\mathbf{S}_1$ and $\mathbf{S}_2$ have independent Wishart distributions $W(\Sigma,n_1)$ and $W(\Sigma,n_2)$, respectively. The random matrix $\mathbf{B}=(\mathbf{S}_1+\mathbf{S}_2)^{-1/2}\mathbf{S}_1(\mathbf{S}_1+\mathbf{S}_2)^{-1/2}$ is called the beta matrix, and its distribution is denoted by $B(n_1/2,n_2/2)$. Its probability density function is

$$\left(\Gamma_p\left(\frac{n_1+n_2}{2}\right)\Big/\Gamma_p\left(\frac{n_1}{2}\right)\Gamma_p\left(\frac{n_2}{2}\right)\right)|\mathbf{B}|^{(n_1-p-1)/2}\times|I-\mathbf{B}|^{(n_2-p-1)/2}$$

for $0<\mathbf{B}<I$.

Suppose that $\mathbf{S}$ has the distribution $W(\Sigma,n)$ and $\mathbf{B}$ has $B(n_1/2,n_2/2)$; then, for any nonsingular symmetric matrix Ω,

$$P\{\mathbf{S}<\Omega\}=\left\{|2\Sigma\Omega^{-1}|^{-n/2}\Gamma_p\left(\frac{p+1}{2}\right)\Big/\Gamma_p\left(\frac{n+p+1}{2}\right)\right\}{}_1F_1\left(\frac{n}{2};\frac{n+p+1}{2};-\frac{1}{2}\Sigma^{-1}\Omega\right)$$

and

$$P\{\mathbf{B}<\Omega\}=\left\{\Gamma_p\left(\frac{n_1+n_2}{2}\right)\Gamma_p\left(\frac{p+1}{2}\right)\Big/\Gamma_p\left(\frac{n_1+p+1}{2}\right)\Gamma_p\left(\frac{n_2}{2}\right)\right\}|\Omega|^{n_1/2}\times{}_2F_1\left(\frac{n_1}{2},-\frac{n_2}{2}+\frac{p+1}{2};\frac{n_1+p+1}{2};\Omega\right).$$

If $\{\mathbf{X}_1,\dots,\mathbf{X}_n\}$ is a random sample from $N(\boldsymbol{\mu},\Sigma)$, then the sample mean $\overline{\mathbf{X}}=\sum_{\alpha=1}^{n}\mathbf{X}_\alpha/n$ and the sample covariance matrix $\mathbf{S}=\sum_{\alpha=1}^{n}(\mathbf{X}_\alpha-\overline{\mathbf{X}})(\mathbf{X}_\alpha-\overline{\mathbf{X}})'/(n-1)$ are distributed independently according to the respective distributions $N(\boldsymbol{\mu},\Sigma/n)$ and $W(\Sigma,n-1)$. If $n>p-1$, $T^2=n(\overline{\mathbf{X}}-\boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\overline{\mathbf{X}}-\boldsymbol{\mu}_0)$ is called the **noncentral Hotelling T^2 statistic** with $n-1$ **degrees of freedom** and **noncentrality** $\lambda=n(\boldsymbol{\mu}-\boldsymbol{\mu}_0)'\Sigma^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)$. $(n-p)T^2/p(n-1)$ has a †noncentral F-distribution with p and $n-p$ degrees of freedom and noncentrality λ.

Let $\mathbf{X}=(X_1,\dots,X_p)'$ and $\mathbf{Y}=(Y_1,\dots,Y_q)'$, $p\leqslant q$ denote two random vectors, Σ_{11} and Σ_{22} their respective covariance matrices, and Σ_{12} the $p\times q$ matrix of covariances between

the components of $\mathbf{X}$ and $\mathbf{Y}$. Each of the nonnegative roots $\rho_1, \ldots, \rho_p$ of the equation $|\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \rho^2\Sigma_{11}| = 0$ is called the **canonical correlation coefficient**. Let $(\mathbf{X}_1, \mathbf{Y}_1), \ldots, (\mathbf{X}_n, \mathbf{Y}_n)$ be a random sample from the $(p+q)$-variate normal distribution $N(\mathbf{0}, \Sigma)$, and let $\mathbf{S}_{11} = \Sigma_{\alpha=1}^n \mathbf{X}_\alpha \mathbf{X}_\alpha'/n$, $\mathbf{S}_{12} = \mathbf{S}_{21}' = \Sigma_{\alpha=1}^n \mathbf{X}_\alpha \mathbf{Y}_\alpha'/n$, $\mathbf{S}_{22} = \Sigma_{\alpha=1}^n \mathbf{Y}_\alpha \mathbf{Y}_\alpha'/n$, and $\mathbf{S} = (\mathbf{S}_{ij})_{i,j=1,2}$. The sample canonical correlation coefficients are the nonnegative roots $r_1, \ldots, r_p$ of $|\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21} - r^2\mathbf{S}_{11}| = 0$ and for $n \geqslant p+q$ the probability density function of $r_1^2, \ldots, r_p^2$ is

$$C_{p,q,n}|I - P^2|^{n/2}|R^2|^{(q-p-1)/2}|I - R^2|^{(n-q-p-1)/2} \times \prod_{i<j}(r_i^2 - r_j^2)\,{}_2F_1^{(p)}\left(\frac{n}{2}, \frac{n}{2}; \frac{q}{2}; R^2, P^2\right),$$

where

$$C_{p,q,n} = \pi^{p^2/2}\Gamma_p\left(\frac{n}{2}\right)\Big/\left\{\Gamma_p\left(\frac{n-q}{2}\right)\Gamma_p\left(\frac{p}{2}\right)\Gamma_p\left(\frac{q}{2}\right)\right\},$$

and R^2 and P^2 are diagonal matrices with elements r_i^2 and ρ_j^2, respectively. If, in particular, $p=1$, then ρ_1 and r_1 are, respectively, the population and the sample multiple correlation coefficients, and $(n-q)r_1^2/q(1-r_1^2)$ follows the distribution $F(q, n-q)$ whenever $\rho_1 = 0$.

Let $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ be a random sample from the 2-dimensional normal distribution with †correlation coefficient ρ. Then the sample correlation coefficient

$$R = \frac{\Sigma_i(X_i - \bar{X})(Y_i - \bar{Y})}{(\Sigma_i(X_i - \bar{X})^2\Sigma_i(Y_i - \bar{Y})^2)^{1/2}}$$

has probability density

$$f_n(r; \rho) = (2^{n-3}/\pi(n-3)!)(1-\rho^2)^{(n-1)/2}(1-r^2)^{(n-4)/2} \times \sum_{k=0}^{\infty}\Gamma^2\left(\frac{1}{2}(n+k-1)\right)\frac{(2\rho r)^k}{k!}$$

for $-1 < r < 1$. For the special case $\rho = 0$, the probability density becomes

$$f_n(r) = \frac{1}{\sqrt{\pi}}\frac{\Gamma\left(\frac{1}{2}(n-1)\right)}{\Gamma\left(\frac{1}{2}(n-2)\right)}(1-r^2)^{(n-4)/2},$$

which implies that $T = \sqrt{n-2}R/\sqrt{1-R^2}$ has the †t-distribution with $n-2$ degrees of freedom.

Given a random sample from a p-variate normal distribution, the probability density of the †sample partial correlation coefficient $R_{12\cdot3\ldots p}$ between the first and the second components with the remaining components fixed is given by $f_{n-p+2}(r; \rho_{12\cdot3\ldots p})$, where f is the density of R mentioned in the previous paragraph and $\rho_{12\cdot3\ldots p}$ is the population partial correlation.

D. Large-Sample Theory

So far we have dealt with random samples $\{X_1, \ldots, X_n\}$ composed of finitely many random variables (or vectors). The theory dealing with such finite cases is called small-sample theory, which is not always suitable for numerical applications. In comparison with this, in large-sample theory, where the sample size is assumed to be sufficiently large, an approximation of the sampling distribution can often be obtained easily by means of the †central limit theorem.

If for three sequences X_n, μ_n, and σ_n, $n = 1, 2, \ldots$, of random variables, real numbers, and positive numbers, respectively, the sequence $(X_n - \mu_n)/\sigma_n$ †converges in distribution to $N(0,1)$ as $n \to \infty$, then the sequence X_n is said to be **asymptotically distributed** according to $N(\mu_n, \sigma_n^2)$. The definition can be extended to higher dimensions. We write $X_n = o_p(r_n)$ for a sequence r_n of positive numbers if and only if X_n/r_n †converges in probability to zero as $n \to \infty$. The following theorem is useful: If $X_n = a + o_p(r_n)$, where a is a constant and $r_n = o(1)$, and if a real-valued function $f(x)$ is of class C^s in a neighborhood of $x = a$, then

$$f(X_n) = \sum_{k=0}^{s}\frac{1}{k!}f^{(k)}(a)(X_n - a)^k + o_p(r_n^s).$$

If X_n is asymptotically distributed according to $N(\mu, \sigma^2/n)$ and $f(x)$ is differentiable at $x = \mu$ with the derivative $f'(\mu) \neq 0$, then $f(X_n)$ is asymptotically distributed according to $N(f(\mu), (f'(\mu))^2\sigma^2/n)$. In higher-dimensional cases, if $\mathbf{X}_n$ is asymptotically distributed according to $N(\boldsymbol{\mu}, \Sigma/n)$ and $f(\mathbf{x})$ is continuously differentiable in a neighborhood of $\mathbf{x} = \boldsymbol{\mu}$ with nonzero vector $\mathbf{c} = (\partial f/\partial x_1, \ldots, \partial f/\partial x_p)_{\mathbf{x}=\boldsymbol{\mu}}$, then $f(\mathbf{X}_n)$ is asymptotically distributed according to $N(f(\boldsymbol{\mu}), \mathbf{c}\Sigma\mathbf{c}'/n)$.

Let $\{X_1, \ldots, X_n\}$ be a random sample from a univariate distribution with finite †moments $\nu_i = E(X^i)$ for $i = 1, \ldots, k$, and let $a_i = \Sigma_\alpha X_\alpha^i/n$ be its ith sample moment. Then the random vector $(a_1, \ldots, a_k)$ asymptotically follows the k-variate normal distribution as $n \to \infty$ with mean vector $(\nu_1, \ldots, \nu_k)$ and covariance matrix $n^{-1}(\sigma_{ij})$, where $\sigma_{ij} = \nu_{i+j} - \nu_i\nu_j$. Let $M_i = \Sigma_\alpha(X_\alpha - \bar{X})^i/n$ and $\mu_i = E(X - \nu_1)^i$ for $i = 2, \ldots, k$ be the sample central moment of order i and population central moment of order i, respectively. Then the random vector $(\bar{X}, M_2, \ldots, M_k)$ obeys the k-variate normal distribution asymptotically as $n \to \infty$ with mean vector $(\nu_1, \mu_2, \ldots, \mu_k)$

and covariance matrix $n^{-1}(\sigma_{ij})$, where $\sigma_{11}=\mu_2$, $\sigma_{1i}=\mu_{i+1}-i\mu_2\mu_{i-1}$, $\sigma_{ij}=\mu_{i+j}-i\mu_{i-1}\mu_{j+1}-j\mu_{i+1}\mu_{j-1}-\mu_i\mu_j+ij\mu_2\mu_{i-1}\mu_{j-1}$ for $i,j\geqslant 2$.

A random variable χ_n^2 that has a †chi-square distribution with n degrees of freedom obeys the distribution $N(n,2n)$ asymptotically as $n\to\infty$. Also, $\sqrt{2\chi_n^2}-\sqrt{2n-1}$ obeys $N(0,1)$ asymptotically. The latter distribution approximates χ_n^2 indirectly better than $N(n,2n)$ approximates χ_n^2 directly. The t-distribution with n degrees of freedom obeys $N(0,1)$ asymptotically as $n\to\infty$. If X_n obeys an F-distribution with m and n degrees of freedom, then mX_n obeys asymptotically the distribution $\chi^2(m)$ as $n\to\infty$. If X_n obeys a †binomial distribution $Bin(n,p)$, then X_n obeys asymptotically the distribution $N(np,np(1-p))$, and $\operatorname{Arc\,sin}\sqrt{X_n/n}$ obeys asymptotically $N(\operatorname{Arc\,sin}\sqrt{p},1/4n)$ as $n\to\infty$. This transformation is called the **arc sin** (or **angular**) **transformation**. If $(X_1,X_2,\ldots,X_k)$ obeys the multinomial distribution $Mu(n;p_1,p_2,\ldots,p_k)$, then it is asymptotically distributed according to the normal distribution $N(\boldsymbol{\mu}_n,\Sigma_n)$, where $\boldsymbol{\mu}_n=(np_1,\ldots,np_k)$, $\Sigma_n=(\sigma_{ij}^{(n)})$, $\sigma_{ii}^{(n)}=np_i(1-p_i)$, and $\sigma_{ij}^{(n)}=-np_ip_j$, $(i\neq j)$, and the random variable $\sum_{j=1}^{k+1}(X_j-np_j)^2/np_j$, where $X_{k+1}=n-(X_1+\ldots+X_k)$ and $p_{k+1}=1-(p_1+\ldots+p_k)$, obeys asymptotically the distribution $\chi^2(k)$ [11].

If X_n has the †Poisson distribution with mean λ_n, where $\lambda_n\to\infty$ as $n\to\infty$, then X_n and $\sqrt{X_n}$ obey the respective distributions $N(\lambda_n,\lambda_n)$ and $N(\sqrt{\lambda_n},1/4)$ asymptotically. If R is the sample correlation coefficient based on a random sample of size n from a 2-dimensional (bivariate) normal distribution with population correlation coefficient ρ, then R is asymptotically distributed according to $N(\rho,(1-\rho^2)^2/n)$ as $n\to\infty$, and therefore $z=\frac{1}{2}\log((1+R)/(1-R))$ obeys asymptotically the distribution $N(\frac{1}{2}\log((1+\rho)/(1-\rho)),1/n)$ asymptotically. This transformation is called **Fisher's** z-**transformation**. The distribution

$$N\left(\frac{1}{2}\log\frac{1+\rho}{1-\rho}+\frac{\rho}{2(n-1)},\frac{1}{n-3}\right)$$

gives a better approximation.

E. Empirical Distribution Function

Let $\{X_1,\ldots,X_n\}$ be a random sample from a distribution F. The random function

$$F_n(x)=\frac{1}{n}\{\text{number of } X\text{'s that are} \leqslant x\}$$

is called the **empirical distribution function**. For any collection of fixed x's $(-\infty=x_0<x_1<\ldots<x_k<\infty)$, the random vector $(nF_n(x_1), n(F_n(x_2)-F_n(x_1)),\ldots,n(F_n(x_k)-F_n(x_{k-1})))$ obeys the †multinomial distribution $Mu(n;p_1,\ldots,p_k)$, where $p_j=F(x_j)-F(x_{j-1})$, $j=1,\ldots,k$, provided that the p's are positive. In particular, the vector is asymptotically distributed according to the k-variate normal. The result is substantially strengthened as follows: the **Glivenko-Cantelli theorem** states that $\sup_x|F_n(x)-F(x)|$ converges to zero with probability 1 as n tends to infinity. If $F(x)$ is continuous, then the random function $\sqrt{n}(F_n(t)-F(t))$ converges in distribution to a †Gaussian process $X(t)$ such that $E(X(t))=0$ and $E(X(s)X(t))=F(s)(1-F(t))$ for $s\leqslant t$. A Gaussian process $X(t)$, $0\leqslant t\leqslant 1$, with this moment condition is called a **Brownian bridge** if $F(t)=t$, for $0\leqslant t\leqslant 1$. If $F(x)$ is continuous, then the distributions of the random variables $C_n=\sqrt{n}\sup_x(F_n(x)-F(x))$ and $D_n=\sqrt{n}\sup|F_n(x)-F(x)|$ do not depend on F. Asymptotically, they have identical distributions with $\sup_t B(t)$ and $\sup_t|B(t)|$, respectively, where $B(t)$ is a Brownian bridge. We have

$$P(\sup_t B(t)\leqslant x)=1-e^{-2x^2},$$

$$P(\sup_t|B(t)|\leqslant x)=1+2\sum_{k=1}^{\infty}(-1)^k e^{-2k^2x^2},$$

$$x>0.$$

Let $\{X_1,\ldots,X_m\}$ and $\{Y_1,\ldots,Y_n\}$ be random samples from continuous distributions F and G, respectively, and let $F_m(x)$ and $G_n(x)$ be their empirical distribution functions. Under the hypothesis $H_0:F\equiv G$, the distribution of the **Kolmogorov-Smirnov test statistic**

$$D_{m,n}=\sup_x|F_m(x)-G_n(x)|$$

does not depend on F (or G), and asymptotically, as $m\to\infty$ and $m/n\to\lambda\leqslant 1$, the random function $\sqrt{m}(F_m(t)-G_n(t))$ converges in distribution to a Gaussian process $X(t)$ such that $E(X(t))=0$ and $E(X(s)X(t))=(1+\lambda)F(s)(1-F(t))$, $s\leqslant t$.

F. Edgeworth and Cornish-Fisher Expansions

Let $\{X_1,X_2,\ldots,X_n\}$ be a sample from a distribution with mean μ and variance σ^2. The random variable $(X_1+X_2+\ldots+X_n-n\mu)/\sqrt{n}\sigma$ is called the normalized sum of the sample. The distribution function $F_n(x)$ of the normalized sum of a sample from an absolutely continuous distribution F with higher-order moments admits the **Edgeworth expansion** [15]:

$$F_n(x)$$

$$=\Phi(x)+\sum_{k=1}^{\nu-2}R_k(x)\left(\frac{1}{\sqrt{n}}\right)^k\phi(x)+B\left(\frac{1}{\sqrt{n}}\right)^{\nu-1},$$

where Φ and ϕ are the †cumulative distribution and the †probability density functions, respec-

tively, of $N(0, 1)$, and B is a quantity bounded by a constant depending on F and v. $R_k(x)$ is the polynomial given by

$$R_k(x) = -\Sigma H_{k+2l-1}(x) \prod_{j=1}^{k} \frac{1}{m_j!}\left(\frac{\gamma_{j+2}}{(j+2)!}\right)^{m_j},$$

where $H_k(x)$ is the †Hermite polynomial of degree k, γ_k is the †cumulant (†semi-invariant) of order k of the distribution of $(X_1-\mu)/\sigma$, the summation extends over all nonnegative m's such that $m_1+2m_2+\ldots+km_k=k$, and $l=m_1+m_2+\ldots+m_k$. In particular, we have $R_1(x)=-\gamma_3(x^2-1)/6$ and $R_2(x)=-\gamma_4(x^3-3x)/24-\gamma_3^2(x^5-10x^3+15x)/72$. For a †lattice distribution F concentrated on $0, \pm 1, \pm 2, \ldots$ but not on $0, \pm\rho, \pm 2\rho, \ldots$ for any $\rho>1$, the following expansion is valid for $x=0, \pm 1, \pm 2, \ldots$:

$$F_n\left(\frac{x-n\mu}{\sqrt{n}\sigma}\right) = \Phi(z) + \sum_{k=1}^{v-2} Q_k(z)\left(\frac{1}{\sqrt{n}}\right)^k \phi(z) + B\left(\frac{1}{\sqrt{n}}\right)^{v-1},$$

where $z=(x-n\mu+1/2)/\sqrt{n}\sigma$ and the Q's are suitable polynomials; $Q_1(z)=R_1(z)$ and $Q_2(z)=R_2(z)+z/24\sigma^2$.

The Edgeworth expansion makes it possible to derive asymptotic formulas for the relation between those u and v such that $F_n(v)=\Phi(u)$. If F is an absolutely continuous distribution with moments of order v ($\geqslant 3$), then we have the **Cornish-Fisher expansions** [16]:

$$u = v + \sum_{k=1}^{v-2} A_k(v)\left(\frac{1}{\sqrt{n}}\right)^k + O(n^{-(v-1)/2})$$

and

$$v = u + \sum_{k=1}^{v-2} B_k(u)\left(\frac{1}{\sqrt{n}}\right)^k + O(n^{-(v-1)/2}),$$

where the A's and B's are polynomials derived from the R's of the Edgeworth expansion; $A_1(v)=-\gamma_3(v^2-1)/6$, $A_2(v)=-\gamma_4(v^3-3v)/24+\gamma_3^2(4v^3-7v)/36$, $B_1(u)=\gamma_3(u^2-1)/6$, $B_2(u)=\gamma_4(u^3-3u)/24-\gamma_3^2(2u^3-5u)/36$.

The expansions imply, in particular, that the random variable $v+\Sigma_{k=1}^{v-2} A_k(v)n^{-k/2}$ with $v=(X_1+X_2+\ldots+X_n-n\mu)/\sqrt{n}\sigma$ is asymptotically distributed according to $N(0,1)$ and that the $100\alpha\%$ point v_α of F_n is approximated by $u_\alpha+\Sigma_{k=1}^{v-2} B_k(u_\alpha)n^{-k/2}$,where u_α is the $100\alpha\%$ point of $N(0, 1)$. These approximations can be improved further in some cases by a suitable transformation of the sum $X_1+X_2+\ldots+X_n$. Thus, for example, if X is distributed according to $\chi^2(n)$, then the Cornish-Fisher expansions with $v=3$ are

$$u = v - \frac{1}{3}\sqrt{\frac{2}{n}}(v^2-1) + O\left(\frac{1}{n}\right)$$

and

$$v = u + \frac{1}{3}\sqrt{\frac{2}{n}}(u^2-1) + O\left(\frac{1}{n}\right),$$

where $v=(X-n)/\sqrt{2n}$. However, the distribution of the random variable

$$\sqrt{\frac{9n}{2}}\left\{\left(\frac{X}{n}\right)^{1/3} - 1 + \frac{2}{9n}\right\}$$

is much better approximated by $N(0, 1)$, and

$$n\left(1-\frac{2}{9n}+\sqrt{\frac{2}{9n}}u_\alpha\right)^3$$

gives a more accurate approximation to the $100\alpha\%$ point of the distribution $\chi^2(n)$. These are called the **Wilson-Hilferty approximations** (*Proc. Nat. Acad. Sci. US*, 17 (1931)).

The Edgeworth expansion was shown to be valid in more general situations by R. N. Bhattacharya and J. K. Ghosh [17]. In particular, they obtained the following: Let $\{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n\}$ be a random sample from a p-variate distribution with a nonzero †absolutely continuous component w.r.t. †Lebesgue measure on $\mathbf{R}^p$. Let f_0 $(\equiv 1), f_1, \ldots, f_k$ be linearly independent, real-valued, and continuously differentiable functions. For $i=1, \ldots, n$, put $\mathbf{Z}_i=(f_1(\mathbf{X}_i), f_2(\mathbf{X}_i), \ldots, f_k(\mathbf{X}_i))$, and assume that the distribution of $\mathbf{Z}_1$ has moments up to the order v ($\geqslant 3$). Let H be a real-valued function on $\mathbf{R}^k$ such that the vth order derivatives are continuous in a †neighborhood of $\boldsymbol{\mu}\equiv E(\mathbf{Z}_1)$. Let $V=(v_{ij})$, $i, j=1, \ldots, k$, be the covariance matrix of the random vector $\mathbf{Z}_1$, and put $\sigma^2=\Sigma v_{ij}l_il_j$, where $l_i=\partial H(\mathbf{z})/\partial \mathbf{z}_i|_{\mathbf{z}=\boldsymbol{\mu}}$, and $\mathbf{z}=(z_1, \ldots, z_k)$. Then

$$\sup_B \left|\Pr\{\sqrt{n}(H(\overline{\mathbf{Z}})-H(\boldsymbol{\mu}))\in B\} - \int_B \psi_{v,n}(x)\,dx\right| = O(n^{-(v-2)/2}),$$

where $\overline{\mathbf{Z}}=\Sigma_{i=1}^n \mathbf{Z}_i/n$, the supremum is taken over all Borel measurable sets B,

$$\psi_{v,n}(x) = \left(1+\sum_{k=1}^{v-2}\left(\frac{1}{\sqrt{n}}\right)^k P_k\left(-\frac{d}{dx}\right)\right)\phi_\sigma(x),$$

$\phi_\sigma(x)$ is the probability density function of the normal distribution $N(0, \sigma^2)$, and the P's are polynomials whose coefficients are independent of n.

G. Order Statistics

Let $\{X_1, \ldots, X_n\}$ be a random sample from a univariate distribution with continuous probability density $f(x)$ and distribution function $F(x)$, and let $X_{(1)}\leqslant \ldots \leqslant X_{(n)}$ be †order statistics. The joint probability density of $Y_1=X_{(\alpha)}$, $Y_2=$

$X_{(\beta)}, \dots, Y_{p-1} = X_{(\varepsilon)}$, and $Y_p = X_{(\eta)}$ is given by

$$\frac{n!}{(\alpha-1)!(\beta-\alpha-1)!\dots(\eta-\varepsilon-1)!(n-\eta)!} \times (F(y_1))^{\alpha-1}(F(y_2)-F(y_1))^{\beta-\alpha-1}\dots \times (F(y_p)-F(y_{p-1}))^{\eta-\varepsilon-1}(1-F(y_p))^{n-\eta}\dots \times f(y_1)f(y_2)\dots f(y_p)$$

for $-\infty < y_1 < \dots < y_p < \infty$, where $\alpha < \beta < \dots < \varepsilon < \eta$. If for given constants $0 < \lambda_1 < \dots < \lambda_p < 1$, each of the subscripts $\alpha, \dots, \eta$ tends to infinity as $n \to \infty$ under the conditions $r_i = n\lambda_i + o(\sqrt{n})$, where $\alpha = r_1, \dots, \eta = r_p$, then the random vector $(Y_1, \dots, Y_p)$ asymptotically obeys the p-dimensional normal distribution with mean vector $(\xi_1, \dots, \xi_p)$ and covariance matrix $n^{-1}(\sigma_{ij})$, where $\sigma_{ij} = \sigma_{ji} = \lambda_i(1-\lambda_j)/f(\xi_i)f(\xi_j)$ for $i \leqslant j$ and ξ_i is the λ_i-quantile of the population, defined by $\lambda_i = F(\xi_i)$.

Suppose that there exist two sequences a_n and b_n of real and positive numbers, respectively, such that as $n \to \infty$ the sequence $(X_{(n)} - a_n)/b_n$ converges in distribution to a nondegenerate distribution G. The underlying distribution F is said to belong to the **domain of attraction** of the limiting distribution G (DA(G), for short). Except for the change of location and scale, only the following three distributions have nonempty domains of attraction:

$$G_1(x) = \begin{cases} 0 & x < 0 \\ e^{-x^{-\gamma}} & x \geqslant 0 \end{cases},$$

$$G_2 = \begin{cases} e^{-(-x)^\gamma} & x < 0 \\ 1, & x \geqslant 0 \end{cases},$$

$$G_3(x) = e^{-e^{-x}}.$$

Writing $\bar{F}(x) = 1 - F(x)$, $\alpha(F) = \inf\{x \mid F(x) > 0\}$ and $\omega(F) = \sup\{x \mid F(x) < 1\}$, we have the following theorem (B. V. Gnedenko, *Ann. Math.*, 44 (1943)):
$F \in \mathrm{DA}(G_1)$ iff $\omega(F) = \infty$ and there exists an a such that

$$\lim_{u \to \infty} \bar{F}(a+ux)/\bar{F}(a+u) = x^{-\gamma} \quad \text{for all } x;$$

$F \in \mathrm{DA}(G_2)$ iff $a \equiv \omega(F) < \infty$ and

$$\lim_{u \to 0} \bar{F}(a-ux)/\bar{F}(a-u) = x^{\gamma} \quad \text{for all } x > 0;$$

$F \in \mathrm{DA}(G_3)$ iff there exists a positive function $R(t)$ such that

$$\lim_{t \to \omega(F)} \bar{F}(t + xR(t))/\bar{F}(t) = e^{-x} \quad \text{for all } x.$$

If $F(x)$ is twice differentiable, $f(x) = F'(x)$ is positive for sufficiently large x, and $\lim_{x \to \omega(F)} d\{(1-F(x))/f(x)\}/dx = 0$, then $F \in \mathrm{DA}(G_3)$. Noticing the relation $X_{(1)} = -\max\{-X_1, -X_2, \dots, -X_n\}$, we can also derive the possible limiting distributions for the sequence $(X_{(1)} - a_n)/b_n$ and their domains of attraction. The statistics $R_n = X_{(n)} - X_{(1)}$ and $M_n = (X_{(1)} + X_{(n)})/2$ are called the range and the **midrange** of the sample, respectively. If, for some a_n, a_n', and b_n, both sequences $(X_{(n)} - a_n)/b_n$ and $(X_{(1)} - a_n')/b_n$ converge to nondegenerate distributions G and H, respectively, then they are asymptotically independent, and we have

$$\lim \Pr\{(R_n - a_n + a_n')/b_n \leqslant x\} = \int_{-\infty}^{\infty} (1 - H(y-x))\,dG(y),$$

$$\lim \Pr\{(2M_n - a_n - a_n')/b_n \leqslant x\} = \int_{-\infty}^{\infty} H(x-y)\,dG(y).$$

H. Characterization of the Distribution by means of a Property of the Sampling Distribution

A distribution or a family of distributions can be characterized by a property of the sampling distribution of a suitable statistic. Let $\{X_1, X_2, \dots, X_n\}$ be a random sample from a nondegenerate distribution F, and let $X_{(1)} \leqslant \dots \leqslant X_{(n)}$ be the †order statistics. The †sample mean $\bar{X}$ is independent of the †sample variance $S^2 = \Sigma(X_j - \bar{X})^2/(n-1)$ iff F is normal $N(\mu, \sigma^2)$ (Kawata and Sakamoto, *J. Math. Soc. Japan*, 1 (1949)). Let a_{ij}, $i, j = 1, \dots, n$, be real numbers such that $\Sigma a_{ij} = 0$ and $\Sigma a_{ii} \neq 0$. If F has a finite second moment, then the condition $E(\Sigma a_{ij} X_i X_j \mid \bar{X}) = \text{const.}$ implies that F is normal. Two linear statistics $L_1 = a_1 X_1 + \dots + a_n X_n$ and $L_2 = b_1 X_1 + \dots + b_n X_n$ are independent only if F is normal, provided that $a_j b_j \neq 0$ for some j. In fact, the X's need not be identically distributed: If L_1 and L_2 are independent and $a_j b_j \neq 0$, then the distribution of X_j is normal **(Skitovich-Darmois theorem)**. Yu. V. Linnik [23] gave a necessary and sufficient condition for the normality of F to be equivalent to the identity of the distributions of L_1 and L_2. The condition is stated in terms of the zeros of the entire function $\sigma(z) = |a_1|^z + \dots + |a_n|^z - |b_1|^z - \dots - |b_n|^z$. The result contains as a special case the following characterization theorem for the normal distribution: If $\Sigma a_j^2 = 1$ and L_1 has a distribution identical to that of X_1, then F is normal $N(\mu, \sigma^2)$ with $\mu(\Sigma a_j - 1) = 0$. R. Shimizu gave a complete description of the characteristic function of the distribution for which L_1 has the same distribution as X_1. In particular, it was proved that if $\log|a_1|/\log|a_2|$ is an irrational number, α is the positive number given by $\Sigma |a_j|^\alpha = 1$, and if L_1 has a distribution identical to that of X_1, then F is the †stable distribution with †characteristic

exponent α. The result was extended in [24] to the cases where the a's are random variables independent of the X's. If $E(X_i)=0$ and $E(\bar{X} | X_1-\bar{X}, X_2-\bar{X}, \ldots, X_n-\bar{X})=0$, then F is normal [26]. Let $\mu_1, \mu_2, \ldots, \mu_n$ be a set of real numbers. If the sampling distribution of the statistic $\sum(X_i+\mu_i)^2$ depends on the μ's only through $\sum\mu_i^2$, then F is normal. If X_i is positive and has finite mean, then the condition $E(\bar{X} | X_2/X_1, X_3/X_1, \ldots, X_n/X_1)=\text{const.}$ implies that F is the gamma distribution. If the distribution F is not concentrated on a lattice 0, $\rho, 2\rho, \ldots$, then $E(X_{(k+1)}-X_{(k)} | X_{(k)})=\text{const.}$ for some k implies that F is the †exponential distribution: $F(x)=1-e^{-\lambda x}$, $x>0$. $X_{(k+1)}-X_{(k)}$ has the identical distribution for some k with $\min\{X_1, X_2, \ldots, X_{n-k}\}$ iff F is exponential. If $a_1, a_2, \ldots, a_n$ are positive numbers such that $a_1+a_2+\ldots+a_n=1$ and such that $\log a_1/\log a_2$ is irrational, then the sampling distribution of $\min\{X_1/a_1, X_2/a_2, \ldots, X_n/a_n\}$ is the same as that of X_1 iff F is exponential. $X_{(1)}$ is independent of $X_{(1)}-\bar{X}$ iff F is exponential.

Suppose that the distribution F has a bounded density function and that the integral $\int_{-\infty}^{\infty} e^{tx} dF(x)$ is finite on a neighborhood of $t=0$. Let $\{X_1, X_2, \ldots, X_n\}$ be a random sample from a distribution of the family $\mathscr{P}=\{F((x-\mu)/\sigma) | -\infty<\mu<\infty, \sigma>0\}$. For $n\geqslant 9$, the sampling distribution of the statistics

$$\left\{\frac{(X_3-X_1)}{(X_9-X_7)}, \frac{(X_2-X_1)}{(X_8-X_7)}, \frac{(X_6-X_4)}{(X_9-X_7)}, \frac{(X_5-X_4)}{(X_8-X_7)}, \operatorname{sgn}(X_3-X_1), \operatorname{sgn}(X_2-X_1), \operatorname{sgn}(X_6-X_4), \operatorname{sgn}(X_5-X_4)\right\}$$

uniquely determines the family $\mathscr{P}$. If F is symmetric, then for $n\geqslant 6$ the distribution of $\{|(X_4-X_3)/(X_2-X_1)|, |(X_6-X_5)/(X_2-X_1)|\}$ uniquely determines $\mathscr{P}$ [25].

I. U-Statistics

Let $\{X_1, \ldots, X_n\}$ be a random sample from a certain distribution, and let $\varphi(x_1, \ldots, x_m)$ be a real-valued function that is symmetric with respect to the arguments $x_1, \ldots, x_m$. The statistic

$$U=\binom{n}{m}^{-1} \sum \varphi(X_{\alpha_1}, \ldots, X_{\alpha_m}),$$

where the summation extends over all combinations $(\alpha_1, \ldots, \alpha_m)$ taken from $(1, 2, \ldots, n)$, is called a **U-statistic**. Let $\theta=E(\varphi(X_1, \ldots, X_m))$, and assume that $E(\varphi(X_1, \ldots, X_m)^2)$ is finite. Then the mean and variance of U are given by

$$E(U)=\theta, \ V(U)=\binom{n}{m}^{-1} \sum_{i=1}^{m} \binom{m}{i}\binom{n-m}{m-i}\zeta_i,$$

where ζ_i is the covariance between $\varphi(X_1, \ldots, X_m)$ and $\varphi(X_1, \ldots, X_i, X'_{i+1}, \ldots, X'_m)$, with $X'_{i+1}, \ldots, X'_m$ an additional independent random sample of size $m-i$ from the same distribution. If $\zeta_1\neq 0$, then U obeys the distribution $N(\theta, m^2\zeta_1/n)$ asymptotically as $n\to\infty$ (W. Hoeffding, *Ann. Math. Statist.*, 19 (1948)).

These results can be generalized to the case of several populations and samples. Let $X_{11}, \ldots, X_{1n_1}; \ldots; X_{c1}, \ldots, X_{cn_c}$ be c independent random samples, each drawn from one of c populations. Assume that a real-valued function $\varphi(x_{11}, \ldots, x_{1m_1}; \ldots; x_{c1}, \ldots, x_{cm_c})$ is symmetric with respect to the arguments $x_{i1}, \ldots, x_{im_i}$ for each $i=1, \ldots, c$. Then

$$U=\prod_{i=1}^{c}\binom{n_i}{m_i}^{-1} \cdot \sum \varphi(X_{1\alpha(11)}, \ldots, X_{1\alpha(1m_1)}; \ldots; X_{c\alpha(c1)}, \ldots, X_{c\alpha(cm_c)}),$$

where the summation extends over all combinations $(\alpha(i1), \ldots, \alpha(im_i))$ of m_i numbers taken from $(1, 2, \ldots, n_i)$ for each $i=1, \ldots, c$, is called a **U-statistic**. The mean and variance of U can be obtained as before, while U is asymptotically normally distributed as the sample sizes $n_1, \ldots, n_c$ tend to infinity in fixed proportion. Furthermore, if there are given several U-statistics, their †joint distribution is asymptotically normal.

J. Distributions Having Monotone Likelihood Ratio, and Pólya-Type Distributions

Let $(\mathscr{X}, \mathfrak{B})$ be a sample space and

$$\mathscr{P}=\{p_\theta(x) | \theta\in\Omega\}$$

be a family of probability densities with respect to a fixed †σ-finite measure. The function $p_\theta(x)$ regarded as a function of θ with a fixed observed value of x is called the **likelihood function**, and its value at a particular point θ is called the **likelihood** of that point. The family $\mathscr{P}$ with $\Omega\subset\mathbf{R}$ is said to have **monotone likelihood ratio** with respect to a real-valued function $T(x)$ if and only if for any $\theta<\theta'$ such that θ and θ' belong to Ω the ratio $p_{\theta'}(x)/p_\theta(x)$ is a nondecreasing function of $T(x)$. Under the assumption that $\mathscr{X}\subset\mathbf{R}$ and $\partial^2\log p_\theta(x)/\partial x\partial\theta$ exists, a necessary and sufficient condition for $\mathscr{P}$ to have monotone likelihood ratio with respect to $T(x)\equiv x$ is that $\partial^2\log p_\theta(x)/\partial x\partial\theta\geqslant 0$ for any x and θ. If $\{X_1, \ldots, X_n\}$ is a random sample from a distribution that has a monotone likelihood ratio and if a real-valued function $\psi(x_1, \ldots, x_n)$ is nondecreasing in each of its arguments, then the expectation $E_\theta(\psi(X_1, \ldots, X_n))$ is a nondecreasing function of θ.

The family $\mathscr{P}$ is said to be of **Pólya type n** if and only if for any $m=1, 2, \ldots, n$ and any

real numbers $x_1 < \ldots < x_m$ and $\theta_1 < \ldots < \theta_m$, the determinant of the matrix $(p_{\theta_i}(x_j))$, $i, j = 1, \ldots, m$, is nonnegative, and $\mathscr{P}$ is said to be **strictly of Pólya type** n if the determinant is positive. Being of Pólya type 2 is equivalent to having a monotone likelihood ratio. If $\mathscr{P}$ is (strictly) of Pólya type n for any positive integer n, then it is said to be (strictly) of **Pólya type**. An †exponential family of distributions with probability density $p_\theta(x) = \exp(\theta x + \alpha(\theta) + s(x))$ for $x \in \mathscr{X} \subset \mathbf{R}$ and $\theta \in \Omega \subset \mathbf{R}$ is strictly of Pólya type. Each of the noncentral chi-square distribution, noncentral t-distribution, and noncentral F-distribution is strictly of Pólya type with respect to the noncentrality parameter.

References

[1] H. Cramér, Mathematical methods of statistics, Princeton Univ. Press, 1946.
[2] S. S. Wilks, Mathematical statistics, Wiley, 1962.
[3] M. G. Kendall and A. Stuart, The advanced theory of statistics I, Griffin, fifth edition, 1952.
[4] N. L. Johnson and S. Kotz, Distributions in statistics, 4 vols., Wiley, 1969–1972.
[5] Student (W. G. Gosset), The probable error of a mean, Biometrika, 6 (1908), 1–25.
[6] R. A. Fisher, Statistical methods for research workers, Oliver & Boyd, 1925.
[7] T. W. Anderson, An introduction to multivariate statistical analysis, Wiley, 1958.
[8] A. T. James, Distributions of matrix variates and latent roots derived from normal samples, Ann. Math. Statist., 35 (1964), 475–501.
[9] A. G. Constantine, Some non-central distribution problems in multivariate analysis, Ann. Math. Statist., 34 (1963), 1270–1285.
[10] H. Chernoff, Large-sample theory: Parametric case, Ann. Math. Statist., 27 (1956), 1–22.
[11] K. Pearson, On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, Philos. Mag., (5) 50 (1900). Reproduced in K. Pearson, Early statistical papers, Cambridge Univ. Press, 1948.
[12] K. R. Parthasarathy, Probability measures on metric spaces, Academic Press, 1967.
[13] P. Billingsley, Convergence of probability measures, Wiley, 1968.
[14] R. H. Randles and D. A. Wolfe, Introduction to the theory of nonparametric statistics, Wiley, 1979.
[15] F. Y. Edgeworth, The law of errors, Trans. Cambridge Philos. Soc., 20 (1904).
[16] E. A. Cornish and R. A. Fisher, Moments and cumulants in the specification of distributions, Rev. Int. Statist. Inst., 5 (1937). Reproduced in R. A. Fisher, Contributions to mathematical statistics, Wiley, 1950.
[17] R. N. Bhattacharya and J. K. Ghosh, On the validity of the formal Edgeworth expansion, Ann. Statist., 6 (1978), 434–451.
[18] R. N. Bhattacharya and R. Ranga Rao, Normal approximation and asymptotic expansions, Wiley, 1976.
[19] E. J. Gumbel, Statistics of extremes, Columbia Univ. Press, 1958.
[20] A. E. Sarhan and B. G. Greenberg (eds.), Contributions to order statistics, Wiley, 1962.
[21] H. A. David, Order statistics, Wiley, second edition, 1981.
[22] J. Galambos, The asymptotic theory of extreme order statistics, Wiley, 1978.
[23] Yu. V. Linnik, Linear forms and statistical criteria, Selected Transl. Math. Statist. Prob., 3 (1962), 1–90. (Original in Russian, 1953.)
[24] R. Shimizu and L. Davies, General characterization theorems for the Weibull and the stable distributions, Sankhyā, ser. A, 43 (1981), 282–310.
[25] Yu. V. Prokhorov, On a characterization of a class of probability distributions by distributions of some statistics, Theory of Prob. Appl., 10 (1965), 438–445. (Original in Russian, 1965.)
[26] A. M. Kagan, Yu. V. Linnik, and C. R. Rao, Characterization problems in mathematical statistics, Wiley, 1973.
[27] J. Galambos and S. Kotz, Characterizations of probability distributions, Springer, 1978.
[28] G. P. Patil, S. Kotz, and J. K. Ord (eds.), Statistical distributions in scientific work III, Characterizations and applications, Reidel, 1975.
[29] S. Karlin, Decision theory for Pólya type distributions, Case of two actions I, Proc. 3rd Berkeley Symp. Math. Stat. Prob. I, Univ. of California Press, 1956, 115–128.

375 (XX.26) Scattering Theory

A. General Remarks

The path of a moving (incident) particle is distorted when it interacts with another (target) particle, such as an atom or a molecule. Phenomena of this sort are generally called

scattering. Scattering is called **elastic** when the internal properties of the incident particle and the target remain unchanged after the collision, and **inelastic** when the internal properties change, other particles are emitted, or the two particles form a bound state.

The extent of scattering depends on the sizes of the incident and target particles. The **scattering cross section** is defined as the probability that the incident beam will be scattered per unit time (normalized to one particle per unit time crossing unit area perpendicular to the direction of incidence). In †classical mechanics the scattering cross section is equal to the cross section of the target perpendicular to the incoming beam, hence the term "cross section." The probability of scattering into a unit solid angle in a particular direction is called the **differential cross section**. The probability that the incoming particle is absorbed by the target, called the **absorption cross section**, is intimately connected with the scattering cross section. Analyses of scattering give information on the structure and interactions of atoms, molecules, and elementary particles. One can also study the scattering of acoustic and electromagnetic waves by inhomogeneous media and obstacles, by considering notions similar to the above.

Scattering theory may be dated back to Lord Rayleigh. Since the advent of quantum mechanics in the mid-1920s, scattering problems, mainly for central (spherically symmetric) potentials, have been investigated strenuously by physicists. It may be said, however, that a scattering theory having mathematically rigorous foundations began around the 1950s, when the pioneering work of K. Friedrichs (*Comm. Pure Appl. Math.*, 1 (1948)), A. Ya. Povzner (*Mat. Sb.*, 32 (1953)), T. Kato (*J. Math. Soc. Japan*, 9 (1957)), J. M. Cook (*J. Math. Phys.*, 36 (1957)), and J. M. Jauch (*Helv. Phys. Acta*, 31 (1958)), among others, appeared, and scattering theory has now grown into a branch of mathematical physics.

General references for mathematical scattering theory are, e.g., [1–5].

B. Wave and Scattering Operators

In †quantum mechanics the dynamics of an interacting system is given by a †one-parameter group of unitary operators e^{-itH}, where t denotes the time and H, called the Hamiltonian of the system, is a †self-adjoint operator acting in a †Hilbert space $\mathscr{H}$. Elements of $\mathscr{H}$ represent (pure) states of the system. Let H_0 be the "free" Hamiltonian of the corresponding "noninteracting" system. (There are at present no generally accepted definite criteria for "free" and "noninteracting.") H_0 is assumed to be †absolutely continuous, which is the case in most practical situations. Then the **outgoing** and **incoming wave operators** $W_\pm \equiv W_\pm(H, H_0)$ are defined, if they exist, by

$$W_\pm = \operatorname*{s\text{-}lim}_{t\to\pm\infty} e^{itH}e^{-itH_0} \qquad (\text{s-lim} = {}^\dagger\text{strong limit}).$$

This means that given any free motion $e^{-itH_0}u$ there is an initial $(t=0)$ state $u_\pm$ $(=W_\pm u)$ such that $e^{-itH}u_\pm$ and $e^{-itH_0}u$ are asymptotically equal at $t=\pm\infty$. $W_\pm$ are †isometric, intertwine the two dynamics: $e^{-itH}W_\pm = W_\pm e^{-itH_0}$, and map $\mathscr{H}$ (which is nothing but the †absolutely continuous subspace $\mathscr{H}_{ac}(H_0)$ for H_0) onto a closed subspace of $\mathscr{H}_{ac}(H)$. The **scattering operator** S is defined as $S = W_+^* W_-$ (A^* is the Hilbert-space †adjoint of A). S commutes with H_0 and maps states in the remote past into states in the distant future. One of the most important problems in scattering theory is to prove the †unitarity of S, or equivalently, $\operatorname{Ran} W_+ = \operatorname{Ran} W_-$ (Ran = †range = †image). $W_\pm$ is called **complete** if $\operatorname{Ran} W_\pm = \mathscr{H}_{ac}(H)$. The completeness of $W_\pm$ implies that S is unitary.

As a typical example we consider the 1-body problem. Note that the 2-body problem reduces to the 1-body problem by separating out the center-of-mass motion, which is free. Then $H_0 = -\Delta$ (the negative †Laplacian in $\mathbf{R}^3$), $H = -\Delta + V$, the operator V being multiplication by a real-valued function $V(x)$, called the **potential**, and $\mathscr{H} = L_2(\mathbf{R}^3)$. If V is **short range**, i.e., if, roughly speaking, $V(x) = O(|x|^{-1-\varepsilon})$ $(\varepsilon > 0)$ at ∞, the wave operators are known to exist and to be complete (S. Agmon, *Ann. Scuola Norm. Sup. Pisa*, (4) 2 (1975)). If the potential $V(x)$ is **long range**, i.e., if, roughly speaking, $V(x) = O(|x|^{-\varepsilon})$ $(\varepsilon > 0)$ at ∞, then the foregoing definition of the wave operators has to be modified. For the Coulomb potential $V(x) = c/x$, for instance, one can adopt the following definition of **modified wave operators**:

$$\tilde{W}_\pm = \operatorname*{s\text{-}lim}_{t\to\pm\infty} e^{itH}\exp(-itH_0 - i(c/2)H_0^{-1/2}\log t).$$

It can be shown that the $\tilde{W}_\pm$ exist (which implies that the ordinary wave operators do not exist) and are complete: $\operatorname{Ran} \tilde{W}_\pm = \mathscr{H}_{ac}(H)$ (J. D. Dollard, in [6]). The same result obtains for more general long-range potentials (H. Kitada, *J. Math. Soc. Japan*, 30 (1978); T. Ikebe and H. Isozaki, *Integral Equations and Operator Theory*, 5 (1982)).

If the wave operators exist and are complete, they give †unitary equivalence between the †absolutely continuous parts of H_0 and H (→ T. Kato [7]; 331 Perturbation of Linear Operators).

In the foregoing discussion it was tacitly assumed that in dealing with scattering phenomena we adhere to states in $\mathscr{H}_{ac}(H_0)$ and

$\mathscr{H}_{ac}(H)$. A more physically intuitive definition in the case of potential scattering ($H_0 = -\Delta$, $H = H_0 + V$) of **scattering states** $\Sigma_{\pm}(H)$ is the following: $f \in \Sigma_{\pm}(H)$ if and only if for any $r > 0$,

$$\lim_{T \to \pm\infty} \frac{1}{T} \int_0^T \|F_r e^{-itH} f\|^2 dt = 0$$

$$(\| \ \| = L_2\text{-norm}),$$

where F_r is the (projection) operator of multiplication by the characteristic function of $\{x \in \mathbf{R}^3 \mid |x| < r\}$. In general no inclusion relations between $\Sigma_{\pm}(H)$ and $\mathscr{H}_{ac}(H)$ are known. But for a wide class of potentials it is known that $\mathscr{H}_{ac}(H)$ as well as the †continuous subspace of H coincides with $\Sigma_{\pm}(H)$ (in this case there is no †singular continuous spectrum) (W. O. Amrein and V. Georgescu, *Helv. Phys. Acta*, 46 (1973); C. Wilcox, *J. Functional Anal.*, 12 (1973); Amrein, in [8]).

A purely abstract result in scattering theory may be noted. Let H_0 and H be self-adjoint operators in an abstract Hilbert space $\mathscr{H}$ such that $V = H - H_0$ is a †trace-class (†nuclear) operator. Then the **generalized wave operators** $W_{\pm}(H, H_0) = \text{s-lim}_{t \to \pm\infty} \exp(itH)\exp(-itH_0) P_{ac}(H_0)$ exist, where $P_{ac}(H_0)$ is the †projection onto $\mathscr{H}_{ac}(H_0)$. Since this statement is symmetric in H_0 and H, the "inverse" generalized wave operators $W_{\pm}(H_0, H)$ also exist, from which one can conclude that $W_{\pm}(H, H_0)$ are complete. Moreover, the **invariance principle** holds, which means roughly the following: If ϕ is a strictly increasing function on R, then $W_{\pm}(\phi(H), \phi(H_0))$ exists and is equal to $W_{\pm}(H, H_0)$. This result can be applied to potential scattering when $V(x) \in L_2(\mathbf{R}^3) \cap L_1(\mathbf{R}^3)$ (Kato [7]; D. B. Pearson, *J. Functional Anal.*, 28 (1978)).

C. Stationary (Time-Independent) Approach

We again consider the 1-body problem as in Section B. $V(x)$ is assumed to verify certain appropriate decay conditions at ∞ as the case may be. Consider the †resolvents $R_0(z) = (H_0 - z)^{-1}$ and $R(z) = (H - z)^{-1}$ for $z \in C - R$, which are well-defined bounded integral operators on $\mathscr{H} = L_2(\mathbf{R}^3)$. Here we note the following: $[0, \infty)$ is the †continuous spectrum of H_0 and H, $(-\infty, 0)$ is contained in the †resolvent set of H_0, and H has possibly †discrete †eigenvalues in $[-a, 0)$ with $(-\infty, -a)$ contained in the resolvent set, where a is a positive number. When z approaches a positive real value, $R_0(z)$ and $R(z)$ do not have limits as †bounded operators on $\mathscr{H}$. But if we regard them as operators from $L_{2,\gamma}$ to $L_{2,-\gamma}$ ($L_{2,\alpha} = \{u \mid (1+|x|)^{\alpha} u(x) \in L_2(\mathbf{R}^3)\}$), $\gamma > 1/2$, they can be shown to have boundary values $R_0(\lambda \pm i0)$ and $R(\lambda \pm i0)$, $\lambda > 0$ (**limiting absorption principle**; → Agmon, *loc. cit.*, for short-range potentials; for long-range potentials → R. Lavine, *J. Functional Anal.*, 12 (1973); T. Ikebe and Y. Saitō, *J. Math. Kyoto Univ.*, 12 (1972); and Saitō, *Publ. Res. Inst. Math. Sci.*, 9 (1974). With these boundary values we can define "stationary" wave operators whose range is easily proved to coincide with $\mathscr{H}_{ac}(H)$, and show their equality to the time-dependent wave operators discussed in Section B, thus obtaining the completeness of the latter.

H_0 is known to have **generalized** (improper) **eigenfunctions** $\varphi_0(x, \xi) = e^{ix \cdot \xi}$ with **generalized** (improper) **eigenvalues** $|\xi|^2$. The associated eigenfunction expansion is nothing but the Fourier integral expansion. An eigenfunction expansion, similar to the Fourier expansion, that diagonalizes H can be obtained by using generalized eigenfunctions

$$\varphi_{\pm}(x, \xi) = \varphi_0(x, \xi) - (R(|\xi|^2 \pm i0) V \varphi_0(\cdot, \xi))(x),$$

which are the solutions to the **Lippmann-Schwinger equation**

$$\varphi_{\pm}(x, \xi) = \varphi_0(x, \xi) - \frac{1}{4\pi} \int_{\mathbf{R}^3} \frac{e^{\pm i|\xi||x-y|}}{|x-y|} V(y) \varphi_{\pm}(y, \xi) dy.$$

A rough statement of this is the following: Let $\tilde{u}_{\pm}(\xi) = (2\pi)^{-3/2} \int \overline{\varphi_{\pm}(x, \xi)} u(x) dx$. Then $\|\tilde{u}\| = \|u\|$, $(Hu)^{\sim}(\xi) = |\xi|^2 \tilde{u}(\xi)$, and $u(x) = (2\pi)^{-3/2} \int \varphi_{\pm}(x, \xi) \tilde{u}(\xi) d\xi$ (→ e.g. [4, XI.6] for a more precise statement).

In view of the fact that S commutes with H_0, we can show that S admits the following representation: Let $\hat{u}(\xi) = (2\pi)^{-3/2} \int \overline{\varphi_0(x, \xi)} u(x) dx$ be the †Fourier transform of u. Then

$$(Su)^{\hat{}}(\xi) \equiv (\hat{S}\hat{u})(\xi) = \hat{u}(\xi) - \pi i \int_{S^2} |\xi| T(\xi, |\xi|\omega') \hat{u}(|\xi|\omega') d\omega',$$

where

$$T(\xi, \xi') = (2\pi)^{-3} \int \varphi_0(x, \xi) V(x) \varphi(x, \xi') dx$$

is the kernel of the so-called **T-operator**, which is a †compact operator on $L_2(S^2)$ under suitable conditions on $V(x)$. $T(\xi, \xi')$ is related to the experimentally measurable total cross section (for incident momentum ξ) $\sigma(\xi)$:

$$\sigma(\xi) = 2\pi \int_{S^2} |f(|\xi|; \omega, \omega')|^2 d\omega' \qquad (\omega = \xi/|\xi|),$$

$$f(\lambda; \omega, \omega') = -2\pi^2 T(\lambda\omega, \lambda\omega') \qquad (\lambda > 0).$$

The quantity $f(\lambda; \omega, \omega')$ is called the **scattering amplitude** and appears in the asymptotic

expansion of $\varphi_\pm(x,\xi)$ as

$$\varphi_\pm(x,\xi)\sim\varphi_0(x,\xi)+\frac{e^{\pm i|\xi||x|}}{|x|}f(|\xi|;\hat{x},\hat{\xi}),$$

where $\hat{a}=a/|a|$ for $a\in\mathbf{R}^3$.

An abstract version of the limiting absorption principle and eigenfunction expansion is known as the Kato-Kuroda theory, for which the reader is referred to Kato and S. T. Kuroda in [6] and in *Functional analysis and related fields*, Springer, 1970, and to Kuroda (*J. Math. Soc. Japan*, 25 (1973)).

D. Time-Dependent Approach

Consider the same situation as in Section C. Since scattering is a time-dependent phenomenon, it seems natural to develop scattering theory in a time-dependent fashion. Indeed, there is an approach to the completeness of wave operators that does not resort to any eigenfunction expansion results, but instead follows the temporal development of the wave packet $e^{-itH}u$. The completeness of $W_\pm$ will be established if one can show that any $u\in\mathscr{H}_{ac}(H)$ orthogonal to Ran $W_\pm$ is 0. A crucial step to prove this is to find a clever decomposition of a wave packet into an outgoing and an incoming one, or to find projections $P_\pm$ such that $P_+ + P_- = I$ and $P_\pm e^{-itH_0}$ goes to 0 as $t\to\mp\infty$. Some compactness arguments are also needed.

To construct such a decomposition or projections one looks at the scalar product $x\cdot\xi$, x and ξ being the position and momentum (operators), respectively. The main idea is that if this is positive, the particle will be outgoing (to infinity), and if negative, incoming (from infinity). But since we work in the framework of quantum mechanics, this classical-mechanical intuition should be properly modified.

Besides the completeness of wave operators it can also be shown through the above approach that the singular continuous spectrum of H is absent. For details → [4, XI.17] and V. Enss, *Comm. Math. Phys.*, 61 (1978).

E. Partial Wave Expansion

In this section we assume that the potential $V(x)$ is central, i.e., $V(x)$ is a function of $|x|$ alone. Then the scattering operator S turns out to commute not only with H_0 but also with the †angular momentum operator $L=x\times i^{-1}\nabla$ (vector product). The eigenvalues of $L^2=L\cdot L$ (scalar product) are $l(l+1)$ $(l=0,1,2,\dots)$, and those of L_z, the third component of L, are $m=-l,-(l-1),\dots,l-1,l$ if L^2 has eigenvalue $l(l+1)$, while simultaneous eigenfunctions are given by suitably normalized †spherical harmonics $Y_{lm}(\omega)$, $\omega\in S^2$. Let E_l be the projection onto the subspace spanned by functions of the form $\sum_{m=-l}^{l}u_{lm}(r)Y_{lm}(\omega)$ $(r=|x|>0,\omega=x/r)$. $\mathscr{H}$ becomes an orthogonal sum of $\mathscr{H}_l=E_l\mathscr{H}$. The aforementioned commutation property claims that $E_lS=SE_l\equiv S_l$, and the operator $\hat{S}_l$ reduces to multiplication by a scalar function $e^{2i\delta_l(\lambda)}$ in $\hat{\mathscr{H}}_l$ (→ Section C). $\delta_l(\lambda)$ is called the **phase shift**. Defining the **partial wave scattering amplitude** $f_l(\lambda)=(2i\lambda)^{-1}(e^{2i\delta_l(\lambda)}-1)$, one obtains the **partial wave expansion** of the scattering amplitude:

$$f(\lambda;\omega,\omega')=\sum_{l=0}^{\infty}(2l+1)f_l(\lambda)P_l(\cos\theta),$$

where θ is the angle between ω and ω', and P_l is a †Legendre polynomial. The total cross section is $\sigma(\lambda)=4\pi\lambda^{-2}\sum_{l=0}^{\infty}(2l+1)|f_l(\lambda)|^2$ [1–5, 9].

F. Many-Body Problem (Multichannel Scattering)

We consider only the 3-body case, which is complicated enough compared with the 2-(essentially 1-) body case. The complications are both kinematical and dynamical. The configuration of a 3-body system is given by a point in $\mathbf{R}^9$. Once we choose the center-of-mass coordinates, there is no kinematically natural way to choose the remaining six coordinates. In the 2-body case a freely moving particle in the remote past will be freely moving in the distant future. But in the 3-body case there come into play various other dynamical processes, such as capture, breakup, rearrangement, and excitation.

The 3-body Hamiltonian is a self-adjoint operator in $L_2(\mathbf{R}^9)$ of the form

$$\tilde{H}=-\sum_{i=1}^{3}\frac{1}{2m_i}\Delta_i+\sum_{i<j}V_{ij}(x_i-x_j),$$

where Δ_i is the 3-dimensional Laplacian associated with particle i, m_i is the mass of particle i, and each $V_{ij}(x)$ $(=V_{ji}(x)=V_{ij}(-x))$ is a real-valued function decaying at ∞ in $\mathbf{R}^3$ (not in $\mathbf{R}^9$). If we remove the center-of-mass motion, $\tilde{H}$ can be written in the form

$$\tilde{H}=H_0\otimes I+I\otimes H \quad (\otimes=\text{†tensor product}),$$

where H_0 is the center-of-mass Hamiltonian in $L_2(\mathbf{R}^3)$ representing the uniform free motion of the center of mass, and H is the Hamiltonian of relative motion in $L_2(\mathbf{R}^6)$. One should note as mentioned above that there is no unique natural way of choosing coordinates in $\mathbf{R}^6$ and representing H, but there are many equivalent representations. Suppose, for instance, that particles 1 and 2 and particles 1 and 3 form †bound states (12) and (13) and that there are

no bound states between 2 and 3. We partition the whole system into **clusters**: (1) (2) (3), (12) (3), and (13) (2) (( ) represents a cluster and figures in ( ) are the particles forming the cluster). A **channel** is a partition into clusters together with a specified bound-state eigenfunction. Take, for instance, channel (12) (3), and suppose $\psi \in L_2(\mathbf{R}^3)$ is the eigenfunction in question. If we take $x = x_2 - x_1$ and $y = x_3 - (m_1 + m_2)^{-1}(m_1 x_1 + m_2 x_2)$, then

$$H = -\frac{1}{2m}\Delta_y - \frac{1}{2n}\Delta_x + V_{12}(x)$$

$$+ V_{23}\left(\frac{m_1 x}{m_1 m_2} - y\right) + V_{31}\left(\frac{m_2 x}{m_1 + m_2} + y\right),$$

where $m^{-1} = m_3^{-1} + (m_1 + m_2)^{-1}$, $n^{-1} = m_1^{-1} + m_2^{-1}$. Let us neglect the interactions between (12) and (3), i.e., set $V_{23} = V_{31} = 0$ to define the **cluster decomposition Hamiltonian**

$$H_{(12)(3)} = -\frac{1}{2m}\Delta_y - \frac{1}{2n}\Delta_x + V_{12}(x).$$

Let $\mathscr{H}_{(12)(3)} = L_2(\mathbf{R}^3)$ (called the **channel Hilbert space** consisting of functions of y), and define a mapping $\tau : \mathscr{H}_{(12)(3)} \to \mathscr{H} = L_2(\mathbf{R}^6)$ (functions of x and y) by $(\tau f)(x, y) = \psi(x) f(y)$. The **channel wave operators** $W_{(12)(3)\pm}$ are defined by

$$W_{(12)(3)\pm} = \operatorname*{s-lim}_{t \to \pm\infty} e^{itH} e^{-itH_{(12)(3)}} \tau$$

as isometries from $\mathscr{H}_{(12)(3)}$ into $\mathscr{H}$. Their ranges are not expected to coincide as in the 2-body case. $W_{(1)(2)(3)\pm}$ and $W_{(13)(2)\pm}$ are similarly defined. Note that $\mathscr{H}_{(1)(2)(3)} = \mathscr{H}$. If α and β are distinct channels, we have $\operatorname{Ran} W_{\alpha+} \perp \operatorname{Ran} W_{\beta+}$ and $\operatorname{Ran} W_{\alpha-} \perp \operatorname{Ran} W_{\beta-}$, but no such relations exist between $\operatorname{Ran} W_{\alpha+}$ and $\operatorname{Ran} W_{\beta-}$ or $\operatorname{Ran} W_{\alpha-}$ and $\operatorname{Ran} W_{\beta+}$. Define, for channels α and β, $S_{\alpha\beta} : \mathscr{H}_\beta \to \mathscr{H}_\alpha$ by $S_{\alpha\beta} = (W_{\alpha+})^* W_{\beta-}$. Now the **scattering operator** S for the 3-body system is defined as the †direct sum of $S_{\alpha\beta}$ acting in the Hilbert space $\Sigma_\alpha \oplus \mathscr{H}_\alpha : S = \Sigma_{\alpha,\beta} \oplus S_{\alpha\beta}$. Naturally the question arises: Is S unitary? The first affirmative answer was made by L. D. Faddeev (*Israel Program for Scientific Translations*, 1965 (in English; original in Russian, 1963)), and later the work of J. Ginibre and M. Moulin (*Ann. Inst. H Poincaré*, A21 (1974)) and L. Thomas (*Ann. Phys.*, 90 (1975)) came out. The method of these authors is stationary. There have also been some attempts using time-dependent methods.

G. Inverse Problem

The **inverse problem** in potential scattering may be formulated at least mathematically as follows: Given the scattering operator or scattering amplitude, determine the potential(s) giving rise to the operator or amplitude. We consider here only the 2-body case (as to the many-body case, almost nothing is known). The central-potential case can be reduced to 1-dimensional problems on $(0, \infty)$. In the 1-dimensional case the celebrated Gel'fand-Levitan theory (→ 112 Differential Operators O) has long been known and has been successfully applied even to nonlinear problems such as the †Korteweg–de Vries equation. In the 3-dimensional case, however, the problem becomes difficult; so far there has not been any satisfactory theory comparable to that for the 1-dimensional case. The potential $V(x)$ is a function $\mathbf{R}^3 \to \mathbf{R}$. The scattering amplitude $f(\lambda; \omega, \omega')$ is a function $f : R \times S^2 \times S^2 \to C$. Let M be the mapping that takes V into f. The inverse problem deals with M^{-1}. Several questions may be posed (in order of increasing difficulty): (1) Is M one-to-one? (2) When it is known that M is one-to-one and f is in the image of M, how does one (re-)construct the V that yields f? (3) What conditions characterize the image of M? Question (1) has been rather satisfactorily answered insofar as short-range potentials are concerned. Concerning questions (2) and (3), attempts have been and are being made to generalize the Gel'fand-Levitan theory, but it may be said that we are still at the beginning stage. References are [2, 9, 10] and R. G. Newton (*J. Math. Phys.*, 20 (1980); 21 (1981); 22 (1982)).

H. Scattering for the Wave Equation

Consider the †wave equation $u_{tt} - \Delta u = 0$ in $\mathbf{R}^3$. The solution $u(t) = u(t, x)$ is uniquely determined by the initial data $\{f, g\} = \{u(0), u_t(0)\}$, and $U_0(t)\{f, g\} = \{u(t), u_t(t)\}$ defines the solution operator $U_0(t)$. The set of data $\{f, g\}$ with finite energy: $\int (|\nabla f|^2 + |g|^2) dx < \infty$ forms a Hilbert space $\mathscr{H}_0$. $U_0(t)$ is a unitary group on $\mathscr{H}_0$. A similar description is possible for solutions of the wave equation in an exterior domain Ω outside an obstacle with zero boundary condition. Denote the resulting Hilbert space and solution operator by $\mathscr{H}$ and $U(t)$, respectively. Let $J : \mathscr{H}_0 \to \mathscr{H}$ be the identification operator defined by $(J\{f, g\})(x) = \{f, g\}(x)$, $x \in \Omega$. The **wave operators** are defined by $W_\pm = \operatorname{s-lim}_{t \to \pm\infty} U(-t) J U_0(t)$. The existence of $W_\pm$ is shown rather easily by using †Huygens's principle. As in Section B, we define the **scattering operator** $S = W_+^* W_-$ and say that $W_\pm$ is **complete** if $\operatorname{Ran} W_\pm = \mathscr{H}_{ac}(H)$, where H is the self-adjoint †infinitesimal generator of $U(t) : U(t) = e^{-itH}$. The completeness of $W_\pm$ and the unitarity of S are proved on the basis of the abstract **translation representation**

theorem: Let $U(t)$ be a unitary group on a Hilbert space $\mathscr{H}$. Suppose there exist subspaces D_+ and D_-, called **outgoing** and **incoming subspaces**, such that $U(t)D_{\pm} \subset D_{\pm}$ for $\pm t \geqslant 0$, $\bigcap_{t\in R} U(t)D_{\pm} = \{0\}$, and $\bigcup_{t\in R} U(t)D_{\pm}$ is dense in $\mathscr{H}$. Then we have two unitary operators $\mathscr{R}_{\pm}: \mathscr{H} \to L_2(R; N)$, where N is an auxiliary Hilbert space, such that $\mathscr{R}_{\pm} U(t) \cdot \mathscr{R}_{\pm}^{-1}$ is right translation by t, and D_+ (D_-) is mapped onto $L_2(0, \infty; N)$ $(L_2(-\infty, 0; N))$ by $\mathscr{R}_+$ $(\mathscr{R}_-)$.

Turning to the concrete situation, one can study the detailed properties of S. The uniqueness theorem in the inverse problem is also obtained, to the effect that S determines the obstacle uniquely. The foregoing treatment of scattering is known as the Lax-Phillips theory (P. D. Lax and R. S. Phillips, in [6, 8, 11].

References

[1] W. O. Amrein, J. M. Jauch, and K. B. Sinha, Scattering theory in quantum mechanics, Benjamin, 1977.
[2] R. G. Newton, Scattering theory of waves and particles, McGraw-Hill, 1966.
[3] E. Prugovečki, Quantum mechanics in Hilbert space, Academic Press, 1971.
[4] M. Reed and B. Simon, Methods of modern mathematical physics III, Scattering theory, Academic Press, 1979.
[5] J. R. Taylor, Scattering theory, Wiley, 1972.
[6] Rocky Mountain J. Math., 1 (1) (1971) (special issue on scattering theory).
[7] T. Kato, Perturbation theory for linear operators, Springer, 1976.
[8] J. A. Lavita and J.-P. Marchand (eds.), Scattering theory in mathematical physics, Proc. NATO Adv. Study Inst., Denver, Colo., 1973, Reidel, 1974.
[9] V. De Alfaro and T. Regge, Potential scattering, North-Holland, 1965.
[10] K. Chadan and P. C. Sabatier, Inverse problems in quantum scattering theory, Springer, 1977.
[11] P. D. Lax and R. S. Phillips, Scattering theory, Academic Press, 1967.

376 (XIX.15) Scheduling and Production Planning

Production planning emerges in many situations. Models of economic planning can be classified as (1) fiscal policy-oriented, (2) final demand-oriented, (3) structure-oriented, (4) expenditure-oriented, and (5) industrialization-oriented. Types (2), (3), and (5) belong to production planning in a broad sense, as emphasis is placed on production in these models.

A typical **production planning** theory of primary importance is **activity analysis**, which has made remarkable progress since its initiation by T. C. Koopmans [1]. Its principal theoretical content consists of †linear programming. Most applications of linear programming are more or less concerned with production activities. Because of the additivity and divisibility of production as well as the limitation of production intensities, problems of production planning can be formulated as problems of linear programming. The methods of †linear algebra are used to obtain an optimal production plan and are very important in modern economic analysis, because these methods not only provide practical algorithms but also clarify the role of price, especially in †dual linear programming problems.

The originators of general equilibrium theory (→ 128 Econometrics) failed to give an analytical demonstration of the existence of solutions of certain systems of equations of economic relevance. The existence of a determinate equilibrium was established first by A. Wald for a system of equations of the Walras-Cassel type. On the other hand, J. von Neumann [3] proved the existence of nonnegative solutions α, β, x_i, y_j for a system of inequalities

$$\alpha \sum_{i=1}^{m} a_{ij} x_i \leqslant \sum_{i=1}^{m} b_{ij} x_i, \qquad j = 1, 2, \ldots, n,$$

$$\beta \sum_{j=1}^{n} a_{ij} y_j \geqslant \sum_{j=1}^{n} b_{ij} y_j, \qquad i = 1, 2, \ldots, m,$$

by reducing the problem to the proof of the existence of a †saddle point (→ 292 Nonlinear Programming A) of the function

$$\Phi(X, Y) \equiv \sum_{i=1}^{m} \sum_{j=1}^{n} b_{ij} x_i y_j \Big/ \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_i y_j$$

by means of Brouwer's fixed-point theorem. In this result an equilibrium is defined in the broad sense that demand for goods does not exceed their supply, rather than requiring exact balance.

A second important kind of production planning is related to both †inventory control (→ 227 Inventory Control) and sales planning. An example is the minimization of

$$\int_0^T (z(t) + \beta \max(dz/dt, 0))\, dt$$

subject to the condition $z(t) \geqslant r(t)$, where $z(t)$ and $r(t)$ are the output and demand, respectively, at time t. Stabilization of employment

and production can also be classified as a production planning problem of this kind, in which inventory holding is considered as a means for lessening the change of employment level. This is related to the problem of smoothing production by inventory control. Dynamic programming (→ 127 Dynamic Programming) is very useful in dealing with problems of smoothing production.

Production planning as a production management tool is often embodied in **scheduling**. Consider a project that consists of n indivisible tasks (jobs or activities) J_i, $i=1,2,\ldots,n$, each requiring p_i units of time for processing, where p_i is given either deterministically or probabilistically. A precedence constraint (generally a †partial ordering) partially specifying the order in which these tasks are to be processed is also imposed by technical considerations: One attempts to find a schedule (i.e., a specification of the time to process the tasks J_i) consistent with the given precedence constraint.

Well-known techniques developed for this purpose are **PERT** (program evaluation and review technique) [4] and **CPM** (critical path method) [5], in which the precedence constraint is represented by an acyclic †directed graph, called an arrow diagram, a project network, or a PERT network, such that each task J_i corresponds to an arc of length equal to p_i. An arrow diagram is illustrated in Fig. 1. The longest path from the start node to the end node in the network is called the **critical path**, and gives the minimum time necessary to complete the project. Following computation of the critical path by means of dynamic programming, computations are also made for the earliest (latest) start time, the earliest (latest) finish time, and the floats (i.e., the allowances for such start and finish times) of each task to be satisfied in order to complete the project within the indicated minimum time. These are used to review and control the progress of the project. In PERT the processing time of each task is probabilistically treated on the basis of three estimates: most likely, optimistic, and pessimistic. From these data other parameters,

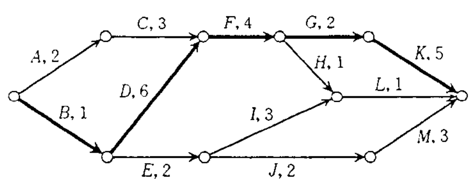


Fig. 1
$A, B, \ldots, M$ denote tasks, while the associated integers are their processing times. Bold arrows indicate the critical path. The start node is on the left, the end node is on the right.

such as the probability of completing the project before the specified due date, are computed. In CPM, on the other hand, a minimum cost schedule to attain the given due date is obtained by utilizing †network flow algorithms (J. E. Kelley [6], D. R. Fulkerson [7]), in which the processing time of a task is determined by linear interpolation between the normal time (achieved with low cost) and the crash time (high cost).

PERT and CPM are used in various areas of application, e.g., civil engineering and the construction industry, shipbuilding, production of automobiles, machines, and electric apparatus, and management of research and development programs. PERT was originally developed by the US Navy to monitor and control the development of the Polaris fleet ballistic missile program, while CPM was developed by the RAND Corporation and the Du Pont Corporation, both in the late 1950s. Computers have been essential from the beginning, to handle the large amount of associated data. A number of application program packages, each with some additional features, are currently available, e.g., PERT/TIME, PERT/COST, CPM, and RAMPS.

The **machine sequencing (scheduling) problem** arises when the resources, instruments, workers, and so forth, required to process a task are abstractly formulated as machines and if the restriction on the number of available machines is taken into consideration (i.e., the conflict between tasks competing for the same machine at the same time must be resolved). Usually one machine is assigned to each task. Such a machine is either (a) uniquely determined for each task or (b) chosen from a given set of machines; in the latter case, there might be (i) parallel machines with the same capability or (ii) machines with different capabilities. The precedence constraints are also ramified into independent (i.e., no constraint), in-tree, out-tree, series-parallel, and general partial ordering constraints. Each task has a ready time (release time) r_i such that J_i cannot be processed before it, and a due time d_i. One is asked to find a schedule satisfying the above machine constraints, precedence constraints, and ready time constraints, while considering an optimality criterion that is a function of the completion time C_i of J_i ($i=1,2,\ldots,n$). Typical criteria for minimization are: (1) maximum completion time (makespan) $C_{\max}=\max_i C_i$; (2) flowtime (total completion time) $F=\sum C_i$; weighted flowtime $\sum w_i C_i$, where $w_i \geqslant 0$ are weights representing the relative importance of J_i; (3) maximum lateness $L_{\max}=\max_i L_i$, $L_i=C_i-d_i$; (4) total tardiness $T=\sum T_i$, $T_i=\max(0, L_i)$, and weighted total

tardiness $\sum w_i T_i$; (5) number of tardy tasks $U=\sum U_i$, $U_i=1$ (if $C_i>d_i$), 0 (otherwise), and weighted number of tardy tasks $\sum w_i U_i$.

Numerous problems can be defined by combining the above conditions. Typical ones might be: the **job-shop scheduling problem**, in which n tasks are scheduled on m machines of type (a), and where the maximum completion time is minimized; the **flow-shop scheduling problem**, which is the same as the job-shop scheduling problem except that $n=n'm$ tasks are divided into n' groups of m tasks processed on machine 1, machine 2, ..., machine m, respectively, in this order; the **multiprocessor scheduling problem**, in which the maximum completion time of n independent tasks on m parallel machines is minimized; the **one-machine sequencing problem**, assuming only one machine (with various types of precedence constraints and optimality criteria), and others [8, 9].

These machine sequencing problems are examples of the combinatorial optimization problem (→ 281 Network Flow Problems E), as the processing time p_i is usually considered to be a given constant. Their computational complexity (→ 71 Complexity of Computations) has been extensively studied with an emphasis on the classification between those problems solvable in polynomial time and those that are †NP-complete, as summarized in [10]. Table 1 lists representative results for one-machine sequencing problems with $r_i=0$ $(i=1,2,\ldots,n)$. The improvement of the algorithm efficiency is pursued for both polynomially solvable problems and NP-complete problems. †Branch and bound (→ 215 Integral Programming D) is a common approach used to solve NP-complete problems such as the job-shop and flow-shop scheduling problems [8, 9]. Many approximation algorithms to obtain good suboptimal schedules in reasonable computation time are also known, and their worst-case and average accuracies have been analyzed [10], as these are important in practical applications.

In more realistic scheduling situations, other factors, such as the set-up cost, balancing of production lines, frequent modifications and updatings of project data, capacity of factories, manpower planning including the possibility of overtime and part-time employment, should be taken into account. Both deterministic and probabilistic models have been proposed for these cases. Mathematical tools used to compute adequate schedules include †mathematical programming, †queuing theory, and †simulation techniques.

References

[1] T. C. Koopmans, Activity analysis of production and allocation, Wiley, 1951.
[2] O. Morgenstern, Economic activity analysis, Wiley, 1954.
[3] J. von Neumann, Über ein ökonomisches Gleichungssystem und eine Verallgemeinerung des Brouwerischen Fixpunktsatzes, Ergebnisse eines Mathematischen Kolloquiums, 8 (1937), 73–83; English translation, A model of general economic equilibrium, Collected Works, Pergamon, 1963, vol. 6, 29–37.
[4] D. G. Malcolm, J. H. Reseboom, C. E. Clark, and W. Fazar, Application of a technique for research and development program evaluation, Operations Res., 7 (1959), 646–669.

Table 1. One-Machine Sequencing Problems with $r_i=0$

Optimality Criterion	Precedence Constraint	Other Constraints	Complexity
$C_{\max}$	Partial order	None	$O(n^2)$
$\sum C_i$	Partial order	None	NP-complete
$\sum w_i C_i$	Series-parallel	None	$O(n \log n)$
	Partial order	$p_i=1$ $(i=1,2,\ldots,n)$	NP-complete
$L_{\max}$	Partial order	None	$O(n^2)$
$\sum T_i$	Independent	None	Not known[a]
	Partial order	None	NP-complete
$\sum w_i T_i$	Independent	None	NP-complete
$\sum U_i$	Independent	None	$O(n \log n)$
	In-tree, out-tree	$p_i=1$ $(i=1,2,\ldots,n)$	NP-complete
$\sum w_i U_i$	Independent	$p_i<p_j \Rightarrow w_i \geqslant w_j$	$O(n \log n)$
	Independent	None	NP-complete[b]

a. An algorithm with $O(n^5 p_{\max})$ running time is known, where $p_{\max}=\max_i p_i$.
b. An algorithm with $O(n \sum p_i)$ running time is known.

[5] J. E. Kelley, Jr., and M. R. Walker, Critical path planning and scheduling, Proc. Eastern Joint Computer Conference, 1959, 160–173.
[6] J. E. Kelley, Jr., Critical-path planning and scheduling: Mathematical basis, Operations Res., 9 (1961), 296–320.
[7] D. R. Fulkerson, A network flow computation for project cost curves, Management Sci., 7 (1961), 167–178.
[8] R. W. Conway, W. L. Maxwell, and L. W. Miller, Theory of scheduling, Addison-Wesley, 1967.
[9] K. R. Baker, Introduction to sequencing and scheduling, Wiley, 1974.
[10] R. L. Graham, E. L. Lawler, J. K. Lenstra, and A. H. G. Rinnooy Kan, Optimization and approximation in deterministic sequencing and scheduling: A survey, Ann. Discrete Math., 5 (1979), 287–326.

377 (XX.27) Second Quantization

A. Fock Space

For a complex Hilbert space K with $\dim K \geqslant 1$, $K^{\otimes n}$ denotes the n-fold tensor product of K with itself (where the vectors $f_1 \otimes \ldots \otimes f_n$ with $f_j \in K$ are total). Let $E_\pm^{(n)}$ be the projection operators on totally symmetric and antisymmetric parts of $K^{\otimes n}$:

$$E_\pm^{(n)}(f_1 \otimes \ldots \otimes f_n)$$
$$=(n!)^{-1}\sum \varepsilon_\pm(P) f_{P(1)} \otimes \ldots \otimes f_{P(n)},$$

where the sum is over all permutations P, $\varepsilon_+(P)=1$ and $\varepsilon_-(P)$ is the signature of P ($+1$ for even permutations and -1 for odd permutations). The following orthogonal direct sum is called a **Fock space** (**symmetric** for E_+, and **antisymmetric** for E_-):

$$\mathscr{F}_\pm(K)=\sum_{n=0}^{\infty}{}^{\oplus} E_\pm^{(n)} K^{\otimes n}.$$

Here the term for $n=0$ is the 1-dimensional space $\mathbf{C}$, and a vector Ω represented by $1 \in \mathbf{C}$ is called the **vacuum vector** in $\mathscr{F}_\pm(K)$. The subspace $\mathscr{F}_\pm(K)_n = E_\pm^{(n)} K^{\otimes n}$ is called the ***n*-particle subspace**. The operator $N=\sum^{\oplus}{}_{n=0}^{\infty} n$, which takes the value n on $\mathscr{F}_\pm(K)_n$, is called the **number operator**.

On the algebraic sum

$$D_\pm=\bigcup_N \sum_{n=0}^{N}{}^{\oplus} \mathscr{F}_\pm(K)_n \subset \mathscr{F}_\pm(K),$$

the **creation operator** $a^+(f)$ for $f \in K$ is defined as the unique linear operator with domain $D_\pm$ satisfying

$$a^+(f) E_\pm^{(n)} f_1 \otimes \ldots \otimes f_n$$
$$=(n+1)^{1/2} E_\pm^{(n+1)} f \otimes f_1 \otimes \ldots \otimes f_n.$$

For $f \in K$, $\bar{f}$ denotes the element of the dual K^* satisfying $\bar{f}(g)=(g,f)$ for $g \in K$ (the inner product is linear in the first entry). The **annihilation operator** $a(\bar{f})$ is defined by

$$a(\bar{f}) E_\pm^{(n)} f_1 \otimes \ldots \otimes f_n$$
$$=n^{-1/2}\sum_{j=1}^{n} \varepsilon^{j-1}(f_j,f) E_\pm^{(n-1)} f_1 \otimes \ldots \hat{j} \ldots \otimes f_n,$$

where the tensor product of f_k, $k \neq j$, appears in the jth term and $\varepsilon = \pm 1$ depending on which of $\pm$ is taken in $\mathscr{F}_\pm(K)$. For $n=0$, $a(\bar{f})\Omega$ is defined to be 0. The adjoint of $a^+(f)$ coincides with $a(\bar{f})$ on $D_\pm$.

The creation and annihilation operators map $D_\pm$ into itself and satisfy the following commutation relations on $D_\pm$:

$$[a^+(f_1), a^+(f_2)]_\mp = [a(\bar{f}_1), a(\bar{f}_2)]_\mp = 0,$$
$$[a(\bar{f}_1), a^+(f_2)]_\mp = (f_2, f_1),$$

where $[A,B]_\mp = AB \mp BA$ and $\mp$ is used depending on the choice of $\pm$ in $\mathscr{F}_\pm(K)$. These relations are often called **canonical commutation relations** for $[\ ,\ ]_-$ (**CCRs**) and **canonical anticommutation relations** for $[\ ,\ ]_+$ (**CARs**).

On $\mathscr{F}_-(K)$, $a^+(f)$ and $a(\bar{f})$ are bounded with $\|a^+(f)\| = \|a(\bar{f})\| = \|f\|$. On $\mathscr{F}_+(K)$, both $a^+(f)$ and $a(\bar{f})$ are not bounded, though $a^+(f)N^{-1/2}$ and $a(\bar{f})N^{-1/2}$ are bounded.

On $\mathscr{F}_+(K)$, $2^{-1/2}(a^+(f)+a(\bar{f}))$ is essentially self-adjoint. Let $\psi(f)$ be its closure. The operator $W(f)=e^{i\psi(f)}$ is unitary and satisfies the identity

$$W(f_1)W(f_2)=W(f_1+f_2)\exp(-i\operatorname{Im}(f_1,f_2)/2).$$

It also satisfies $(W(f)\Omega,\Omega)=\exp(-4^{-1}\|f\|^2)$.

Let K_φ be a real subspace of K such that the inner product (f,g) in K is real for any f and g in K_φ and $K=K_\varphi+iK_\varphi$. (K_φ is then a real Hilbert space.) The unitary operators $U(f)=W(f)$ and $V(f)=W(if)$ for $f \in K_\varphi$ satisfy the following **Weyl form of the CCRs**:

$$U(f_1)U(f_2)=U(f_1+f_2),$$
$$V(f_1)V(f_2)=V(f_1+f_2),$$
$$U(f_1)V(f_2)=V(f_2)U(f_1)\exp(-i(f_1,f_2)).$$

The infinitesimal generators of the continuous one-parameter groups of unitaries $U(tf)$ and $V(tg)$ ($t \in \mathbf{R}$) are denoted by $\varphi(f)$ and $\pi(g)$ and satisfy the following CCRs:

$$[\varphi(f_1),\varphi(f_2)]\Psi=[\pi(g_1),\pi(g_2)]\Psi=0,$$
$$[\varphi(f),\pi(g)]\Psi=i(f,g)\Psi,$$

where $[A,B]=AB-BA$ and $\Psi \in D_+$.

If Q is a linear operator on K, $\Gamma(Q)$ denotes the linear operator on $\mathscr{F}_\pm(K)$ defined as the closure of $\sum^{\oplus}{}_{n=0}^{\infty} Q^{\otimes n} | D_\pm$. It is bounded if $\|Q\| \leqslant 1$. $\Gamma(Q_1)\Gamma(Q_2) = \Gamma(Q_1 Q_2)$ on $D_\pm$. If H is a self-adjoint operator on K, then $\Gamma(e^{itH}) = \exp it\, d\Gamma(H)$ defines a self-adjoint operator $d\Gamma(H)$ on $\mathscr{F}_\pm(K)$, usually called a **bilinear Hamiltonian** and denoted (a^+, Ha). More explicitly,

$$d\Gamma(H) E_\pm^{(n)} f_1 \otimes \ldots \otimes f_n$$

$$= \sum_{j=1}^{n} E_\pm^{(n)} f_1 \otimes \ldots \otimes Hf_j \otimes \ldots \otimes f_n.$$

If U is a unitary operator on K, then

$$\Gamma(U) W(f) \Gamma(U)^{-1} = W(Uf).$$

B. Second Quantization

A single (scalar) particle in quantum mechanics is described by a wave function $\Psi(x)$, $x \in \mathbf{R}^3$, considered as a unit vector in a Hilbert space $K = L_2(\mathbf{R}^3)$. The system consisting of n such identical particles is described by a totally symmetric function $\Psi(x_1, \ldots, x_n)$, $x_j \in \mathbf{R}^3$, considered as a unit vector in the totally symmetric part $E_+^{(n)} K^{\otimes n}$ of the n-fold tensor product of the one-particle Hilbert space K, where the restriction to totally symmetric wave functions is referred to as **Bose statistics**. In a nonrelativistic system, the Hamiltonian operator on a 1-particle space is

$$T = -\hbar^2 (2m)^{-1} \Delta_{\mathbf{x}},$$

called the kinetic energy ($\Delta_{\mathbf{x}}$ denotes the Laplacian); on an n-particle space it is typically given by

$$H_n = -\hbar^2 (2m)^{-1} \sum_{j=1}^{n} \Delta_{\mathbf{x}_j} + \sum_{i<j}^{n} V(\mathbf{x}_i - \mathbf{x}_j),$$

where V is a 2-body potential.

The totality of multiparticle spaces $E_+^{(n)} K^{\otimes n}$ can be described in terms of the Fock space $\mathscr{F}_+(K)$, the vacuum vector Ω (no-particle state), and the annihilation and creation operators, denoted by

$$a(\bar{f}) = \int \Psi(\mathbf{x}) \overline{f(\mathbf{x})} d^3\mathbf{x},$$

$$a^*(f) = \int \Psi^+(\mathbf{x}) f(\mathbf{x}) d^3\mathbf{x}.$$

Since the CCRs (for operator-valued distributions)

$$[\Psi(\mathbf{x}), \Psi(\mathbf{y})]_- = [\Psi^+(\mathbf{x}), \Psi^+(\mathbf{y})]_- = 0,$$

$$[\Psi(\mathbf{x}), \Psi^+(\mathbf{y})]_- = \delta^3(\mathbf{x} - \mathbf{y})$$

are a continuous generalization of CCRs for canonical variables in quantum mechanics and since $\Psi(\mathbf{x})$ comes from the wave function by way of quantization, the above formalism is called **second quantization**. The Hamiltonians H_n for all n can now be combined into the expression

$$H = \int \Psi^+(\mathbf{x}) T \Psi(\mathbf{x}) d^3\mathbf{x}$$

$$+ \frac{1}{2} \int \Psi^+(\mathbf{x}) \Psi^+(\mathbf{y}) V(\mathbf{x} - \mathbf{y}) \times \Psi(\mathbf{x}) \Psi(\mathbf{y}) d^3\mathbf{x}\, d^3\mathbf{y},$$

where the first term is $d\Gamma(T)$.

For particles such as electrons the system of n identical particles is described by a totally antisymmetric wave function, the total antisymmetry being referred to as **Fermi statistics**. Then the antisymmetric Fock space $\mathscr{F}_-(K)$ can be used in exactly the same manner as $\mathscr{F}_+(K)$ in the preceding case.

The method of second quantization was introduced by P. Dirac [2] for the case of bosons and extended by P. Jordan and E. Wigner [3] to fermions. Electromagnetic waves, when quantized in this way, represent a system of **photons**, and the quantization of electron waves leads to the particle picture of the **electron**. The method of second quantization is intimately connected with the notion of fields, as shown below for free fields, and is the basis of the perturbation approach in field theory (→ 150 Field Theory).

C. Free Fields

Let σ_j $(j = 1, 2, 3)$ be †Pauli spin matrices and $\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Let $\tilde{p} = \sum_{\mu=0}^{3} \sigma_\mu p^\mu$ for a 4-vector $p = (p^0, \mathbf{p})$ with $\mathbf{p} = (p^1, p^2, p^3)$ and $p^0 = (m^2 + \mathbf{p}^2)^{1/2}$. Let $u_j(a, A)$ be the irreducible unitary representation $[m_+, j]$ of $\tilde{\mathscr{P}}_+^\uparrow$ on a Hilbert space $K_j = L_2(\mathbf{R}^3, \mathbf{C}^{2j}, (2p^0)^{-1} (m/\tilde{p})^{\otimes 2j} d^3\mathbf{p})$ (→ 258 Lorentz Group C (3)).

Consider first the symmetric Fock space $\mathscr{F}_+(K_0)$. For any complex-valued rapidly decreasing C^∞-function f ($f \in \mathscr{S}(\mathbf{R}^4)$), let

$$\tilde{f}(\mathbf{p}) = (2\pi)^{-3/2} \int e^{ip \cdot x} f(x) d^4 x$$

$$(p = (p^0, \mathbf{p}), \quad p^0 = (\mathbf{p}^2 + m^2)^{1/2}),$$

$$A(f) = a^+(\tilde{f}) + a(\bar{\tilde{\bar{f}}}) \quad \left(= \int A(x) f(x) d^4 x \right),$$

where $p \cdot x = p^0 x^0 - \sum_{j=1}^{3} p^j x^j$ and the bar denotes the complex conjugate. Then $A(x)$ as an operator-valued distribution satisfies the †Wightman axiom and is called the **free scalar field** of mass m. It satisfies the **Klein-Gordon equation**

$$(\square_x + m^2) A(x) \Psi = 0$$

$$\left(\square_x = (\partial/\partial x^0)^2 - \sum_{j=1}^{3} (\partial/\partial x^j)^2 \right),$$

it has the 4-dimensional scalar commutator

$$i[A(x),A(y)]_-\Psi=\Delta_m(x-y)\Psi,$$

and it has the two-point function

$$i(A(x)A(y)\Omega,\Omega)=\Delta_m^+(x-y).$$

Here Ψ is any vector in D_+; for example, Ω is the vacuum vector, and the **invariant distributions** Δ_m and Δ_m^+ are defined by

$$\Delta_m^+(x)=i(2\pi)^{-3}\int e^{-ip\cdot x}(2p^0)^{-1}d^3\mathbf{p},$$

$$\Delta_m(x)=(2\pi)^{-3}\int(\sin p\cdot x)(p^0)^{-1}d^3\mathbf{p}.$$

If we define $U(a,\Lambda(A))=\Gamma(u_0(a,A))$, then

$$U(a,\Lambda)A(x)U(a,\Lambda)^*=A(\Lambda x+a).$$

If $g\in\mathscr{S}(\mathbf{R}^3)$ and $h\in\mathscr{S}(\mathbf{R}^3)$, $\tilde{g}(\mathbf{p})$ and $\tilde{h}(\mathbf{p})$ obtained by substituting $g(\mathbf{x})\delta(x^0)$ and $-h(\mathbf{x})\delta'(x^0)$ into f in the defining equation of $\tilde{f}$ above are in K_0. We then define

$$\varphi(g)=a^+(\tilde{g})+a(\bar{\tilde{g}})\qquad\left(=\int\varphi(\mathbf{x})g(\mathbf{x})d^3\mathbf{x}\right),$$

$$\pi(h)=a^+(\tilde{h})+a(\bar{\tilde{h}})\qquad\left(=\int\pi(\mathbf{x})h(\mathbf{x})d^3\mathbf{x}\right).$$

The operator-valued distributions $\varphi(\mathbf{x})$ and $\pi(\mathbf{x})$ are the **canonical field** and its **conjugate field** at time 0 for the free scalar field and satisfy the following **canonical commutation relations**:

$$[\varphi(\mathbf{x}),\varphi(\mathbf{y})]_-\Psi=[\pi(\mathbf{x}),\pi(\mathbf{y})]_-\Psi=0,$$

$$[\varphi(\mathbf{x}),\pi(\mathbf{y})]_-\Psi=i\delta^3(\mathbf{x}-\mathbf{y})\Psi.$$

If we set $T(t)\equiv U(te_0,1)$ for $e_0=(1,0,0,0)$ and $(\alpha\otimes g)(x)=\alpha(x^0)g(\mathbf{x})$ for $\alpha\in\mathscr{S}(\mathbf{R})$ and $g\in\mathscr{S}(\mathbf{R}^3)$, then for $\Psi\in D_+$

$$A(\alpha\otimes g)\Psi=\int_{-\infty}^{\infty}T(t)\varphi(g)T(t)^*\Psi\alpha(t)dt,$$

$$-A(\alpha'\otimes g)\Psi=\int_{-\infty}^{\infty}T(t)\pi(g)T(t)^*\Psi\alpha(t)dt,$$

or, equivalently,

$$A(x)=T(x^0)\varphi(\mathbf{x})T(x^0)^*,$$

$$\partial A(x)/\partial x^0=T(x^0)\pi(\mathbf{x})T(x^0)^*.$$

If $A(x)$ is a classical field, then $\varphi(\mathbf{x})$ and $\pi(\mathbf{x})$ are the value of $A(x)$ and its time derivative at $x^0=0$, and they serve as initial data for the Klein-Gordon equation

$$(\Box_x+m^2)A(x)=0.$$

Consider next the antisymmetric Fock space $\mathscr{F}_-(K_{1/2}\otimes K_{1/2})$. For $f_\pm\in\mathscr{S}(\mathbf{R}^4,\mathbf{C}^2)$ ($\mathbf{C}^2$-valued rapidly decreasing C^∞-functions), write $f=(f_+,f_-)$, $f_+=(f_1,f_2)$, $f_-=(f_3,f_4)$, define $\tilde{f}_\pm$ as before, and let

$$\psi(f)=m^{1/2}\{a(\overline{(\tilde{f}_-+(\tilde{p}/m)\tilde{f}_+)}\oplus 0)+a^+(0\oplus(\sigma_2\tilde{f}_+-(\tilde{p}/m)\sigma_2\tilde{f}_-))\}$$

$$=\sum_{\alpha=1}^{4}\int\psi_\alpha(x)f_\alpha(x)d^4x,$$

where $\sigma_2=\left(\begin{smallmatrix}0&-i\\ i&0\end{smallmatrix}\right)=-i\varepsilon$. Then $\psi_\alpha(x)$ as an operator-valued distribution satisfies the †Wightman axiom and is called the **free Dirac field** of mass m. In the present formulation, $\{\psi_1,\psi_2\}$ is a contravariant spinor of rank $(1,0)$ and $\{\psi_3,\psi_4\}$ is a covariant spinor of rank $(0,1)$. This field satisfies the **Dirac equation**

$$\left(\sum_\mu\gamma^\mu(\partial/\partial x^\mu)+im\right)\psi(x)=0,$$

as well as the relations

$$[\psi_\alpha(x),\psi_\beta(y)]_+\Psi=0,$$

$$[\psi_\alpha(x),\psi_\beta(y)^*]_+\Psi$$

$$=\left\{\left(\sum_\mu\gamma^\mu(\partial/\partial x^\mu)-im\right)\gamma^0\right\}_{\alpha\beta}\Delta_m(x-y),$$

$$(\psi_\alpha(x)\psi_\beta(y)^*\Omega,\Omega)$$

$$=\left\{\left(\sum_\mu\gamma^\mu(\partial/\partial x^\mu)-im\right)\gamma^0\right\}_{\alpha\beta}\Delta_m^+(x-y).$$

Here $\Psi\in D_+$, Δ_m and Δ_m^+ are as described above, and the γ's are **Dirac matrices** in the following form (somewhat different from but equivalent to the usual form; → 351 Quantum Mechanics):

$$\gamma^0=\begin{pmatrix}0&1\\1&0\end{pmatrix},\qquad\gamma^j=\begin{pmatrix}0&-\sigma_j\\ \sigma_j&0\end{pmatrix},$$

and the σ's are †Pauli spin matrices.

D. Coherent Vectors and Exponential Hilbert Space

In the symmetric Fock space $\mathscr{F}_+(K)$, a vector of the form

$$\exp f\equiv\Omega\oplus\sum_{n=1}^{\infty}{}^{\oplus}(n!)^{-1/2}f^{\otimes n}\in\mathscr{F}_+(K)$$

for $f\in K$ is called a **coherent vector**. The set of $\exp f$ is linearly independent (in the algebraic sense) and total. The inner product is given by

$$(\exp f_1,\exp f_2)=\exp(f_1,f_2).$$

Conversely, we can define $\mathscr{F}_+(K)$ abstractly by introducing this inner product into the formal linear combinations of $\exp f$, $f\in K$, and by completion. In this sense, $\mathscr{F}_+(K)$ is also denoted as $\exp K$ and is called an **exponential Hilbert space** [5]. Then

$$\exp\sum{}^{\oplus}K_j=\otimes\exp K_j,$$

where $\exp\Sigma^{\oplus} f_j$ is identified with $\otimes \exp f_j$. If the number of indices is infinite, the right-hand side is the incomplete infinite tensor product containing the product of the vacuum vector Ω.

If $K=\int_{\Xi} K_\lambda d\mu(\lambda)$ and the measure μ is nonatomic, then for any measurable set S in Ξ, there corresponds a decomposition $\exp K = (\exp K(S)) \otimes (\exp K(S^c))$, where $K(S)= \int_S K_\lambda d\mu(\lambda)$ and S^c is the complement of S in Ξ, and an associated von Neumann algebra $R(S)=B(K(S))\otimes 1$, where $B(K(S))$ is the set of all bounded linear operators on $K(S)$. The system $\{R(S)\}$ forms a complete Boolean lattice of type I factors on $\exp K$. Coherent vectors are characterized by the property of being a product vector for $\{R(S)\}$ in the sense that for any S, $A\in R(S)$ and $A'\in R(S^c)$, the vector $\Psi=\exp f$ satisfies

$$(AA'\Psi,\Psi)\|\Psi\|^2=(A\Psi,\Psi)(A'\Psi,\Psi).$$

In this sense, we can interpret $\exp K$ as a **continuous tensor product** of $\exp K_\lambda$ and also, if $\Psi=\int \Psi_\lambda d\mu(\lambda)$, $\exp\Psi$ as a continuous tensor product of $\exp \Psi_\lambda$ [5].

References

[1] V. Fock, Konfigrationsraum und zweite Quantelung, Z. Phys., 75 (1932), 622–647.
[2] P. A. M. Dirac, The quantum theory of the emission and absorption of radiation, Proc. Roy. Soc. London, 114 (1927), 243–265.
[3] P. Jordan and E. P. Wigner, Über das Paulische Äquivalenzverbot, Z. Phys., 47 (1928), 631–651.
[4] F. A. Berezin, The method of second quantization, Academic Press, 1966. (Original in Russian, 1965.)
[5] H. Araki and E. J. Woods, Complete Boolean algebras of type I factors, Publ. Res. Inst. Math. Sci., A2 (1966), 157–242, 451–452.
[6] J. Avery, Creation and annihilation operators, McGraw-Hill, 1976.

378 (XII.14) Semigroups of Operators and Evolution Equations

A. Introduction

The analytical theory of semigroups was inaugurated around 1948 in order to define exponential functions in infinite-dimensional function spaces. Then it was generalized to the theory of evolution equations as ordinary differential equations in infinite-dimensional linear spaces.

B. The Hille-Yosida Theorem

Let X be a †locally convex topological linear space, and denote by $L(X)$ the totality of continuous linear operators defined on X with values in X. A family $\{T_t \mid t\geqslant 0\}$ of operators $T_t\in L(X)$ is called a **(one-parameter) semigroup of class** (C^0) or a **strongly continuous semigroup** if it satisfies the following two conditions: (i) $T_tT_s=T_{t+s}$ (the semigroup property), $T_0=I$ (the identity operator); and (ii) $\lim_{t\to t_0} T_t x = T_{t_0}x$ ($\forall x\in X, \forall t_0\geqslant 0$). When X is a Banach space, (ii) is implied by $\text{w-lim}_{t\downarrow 0} T_t x = x$ ($\forall x \in X$), as proved by N. Dunford in 1938. In this case there exist constants $M>0$ and $\beta\geqslant 0$ such that (iii') $\|T_t\|\leqslant Me^{\beta t}$ ($\forall t\geqslant 0$). Hence, considering $e^{-\beta t}T_t$ in place of T_t, we can assume the equicontinuity: (iii) For any continuous seminorm p on X, there exists a continuous seminorm q on X such that $p(T_t x)\leqslant q(x)$ ($\forall x\in X$, $\forall t\geqslant 0$). Such semigroups are called **equicontinuous semigroups of class** (C^0) (abbreviated e.c.s.g. (C^0)).

Example 1. $X=L_p(0,\infty)$ with $\infty>p\geqslant 1$.

$$(T_t x)(s)=x(t+s).$$

Example 2. $X=L_p(-\infty,\infty)$ with $\infty>p\geqslant 1$.

$$(T_t x)(s)=(2\pi t)^{-1/2}\int_{-\infty}^{\infty}\exp\left(\frac{(s-u)^2}{2t}\right)x(u)\,du,\quad t>0,$$
$$=x(s),\quad t=0.$$

Example 3. $X=BC(-\infty,\infty)$.

$$(T_t x)(s)=e^{-\lambda t}\sum_{k=0}^{\infty}\frac{(\lambda t)^k}{k!}x(s-k\mu),\quad t\geqslant 0.$$

Here λ and μ are positive constants. (For these examples, we have $\|T_t\|\leqslant 1$; hence (iii) is satisfied.) For L_p and BC → 168 Function Spaces.

We assume in the remainder of the article that X is sequentially complete, that is, if a sequence $\{x_n\}$ of X satisfies $\lim_{n,m\to\infty} p(x_n-x_m)=0$ for every continuous seminorm p on X, then there exists a unique $x\in X$ such that $\lim_{n\to\infty} p(x-x_n)=0$.

The **infinitesimal generator** A of an e.c.s.g. (C^0) $\{T_t \mid t\geqslant 0\}$ is defined by

$$Ax=\lim_{t\downarrow 0} t^{-1}(T_t-I)x. \tag{1}$$

(This is also called the **generator** of T_t.) Then we have the following results.

(I) Differentiability theorem. For every complex number λ with $\mathrm{Re}\,\lambda>0$, the resolvent $(\lambda I-A)^{-1}\in L(X)$ exists and

$$(\lambda I-A)^{-1}x=\int_0^\infty e^{-\lambda t}T_t x\,dt \qquad (\forall x\in X), \tag{2}$$

where the integration is Riemannian. Hence the domain $D(A)$ of A is dense in X and coincides with the range $R((\lambda I-A)^{-1})$, and A is a closed linear operator such that the family

$$\{(\lambda(\lambda I-A)^{-1})^n \mid \lambda>0,\ n=0,1,2,\dots\} \tag{3}$$

is equicontinuous.

(II) Representation theorem. Let $J_n=(I-n^{-1}A)^{-1}$ and consider the approximations to T_t:

$$T_t^{(n)}x=e^{-nt}\sum_{m=0}^{\infty}(m!)^{-1}(ntJ_n)^m x,$$

$$\hat{T}_t^{(n)}x=(J_{nt^{-1}})^n x.$$

Then

$$T_t x=\lim_{n\to\infty}T_t^{(n)}x=\lim_{n\to\infty}\hat{T}_t^{(n)}x \tag{5}$$

uniformly on every compact set of t.

(III) Converse theorem. Let a linear operator A with both dense domain $D(A)$ and range $R(A)$ in X satisfy the condition $(nI-A)^{-1}\in L(X)$ for $n=1,2,\dots$. Then a necessary and sufficient condition for A to be the infinitesimal generator of an e.c.s.g. (C^0) is that the family of operators

$$\{(I-n^{-1}A)^{-m} \mid n=1,2,\dots;m=0,1,\dots\} \tag{3'}$$

be equicontinuous. Since such a semigroup $\{T_t \mid t\geqslant 0\}$ is uniquely determined by A, we can write $T_t=\exp(tA)$.

These three theorems together are called the **Hille-Yosida theorem** or the Hille-Yosida-Feller-Phillips-Miyadera theorem.

Examples of Infinitesimal Generators. $A=d/ds$ for example 1 above, $A=2^{-1}d^2/ds^2$ for example 2, and $(Ax)(s)=\lambda(x(s-\mu)-x(s))$ for example 3.

C. Groups

An operator A in a Hilbert space X generates a group $\{T_t \mid -\infty<t<\infty\}$ of †unitary operators of class (C^0) satisfying $T_tT_s=T_{t+s}$ for $-\infty<t,s<\infty$ if and only if A is equal to iH for some †self-adjoint operator H (**M. H. Stone's theorem**, 1932). In a locally convex space, a necessary and sufficient condition for a given e.c.s.g. (C^0) $\{T_t \mid t\geqslant 0\}$ to be extended to an **equicontinuous group of class** (C^0) $\{T_t \mid -\infty<t<\infty\}$ is that the family (3′) be equicontinuous also for $n=\pm1,\pm2,\dots$.

D. Holomorphic Semigroups

For an e.c.s.g. (C^0) $\{T_t \mid t\geqslant 0\}$, the following three conditions are equivalent (K. Yosida, 1963; the equivalence between (ii) and (iii) for Banach spaces was proved earlier by E. Hille, 1948): (i) When $t>0$,

$$T_t'x=\lim_{h\to 0}h^{-1}(T_{t+h}-T_t)x$$

exists for all $x\in X$ and $\{(CtT_t')^n \mid n=1,2,\dots$ and $0<t\leqslant 1\}$ is equicontinuous for a certain constant $C>0$. (ii) When $t>0$, T_t admits a convergent expansion T_λ given locally by $T_\lambda x=\sum_{n=0}^{\infty}(\lambda-t)^n T_t^{(n)}x/n!$. The extension exists for $|\arg\lambda|\leqslant\arctan(Ce^{-1})$, and the family of operators $\{e^{-\lambda}T_\lambda\}$ is equicontinuous in λ for $|\arg\lambda|\leqslant\arctan(2^{-k}Ce^{-1})$ with some positive constant k. (iii) For the infinitesimal generator A of T_t, there exists a positive constant C_1 such that $\{(C_1\lambda(\lambda I-A)^{-1})^n\}$ is equicontinuous in $n=1,2,\dots$ and in λ with $\mathrm{Re}(\lambda)\geqslant 1+\varepsilon$, $\varepsilon>0$. An e.c.s.g. (C^0) $\{T_t\}$ satisfying the above conditions is called a **holomorphic semigroup**.

For example, introduce

$$f_{t,\alpha}(\lambda)=(2\pi i)^{-1}\int_{\sigma-i\infty}^{\sigma+i\infty}e^{z\lambda-tz^\alpha}dz,$$
$$\lambda\geqslant 0,\quad t>0,\quad \sigma>0,\quad 0<\alpha<1,$$
$$=0,\quad \lambda<0, \tag{6}$$

where the branch of z^α is taken so that $\mathrm{Re}\,z^\alpha>0$ for $\mathrm{Re}\,z>0$. Following S. Bochner (1949), we define

$$\hat{T}_{t,\alpha}x=\hat{T}_t x=\int_0^\infty f_{t,\alpha}(s)T_s x\,ds, \qquad t>0,$$
$$=x, \quad t=0, \tag{7}$$

from a given e.c.s.g. (C^0) $\{T_t \mid t\geqslant 0\}$. Then $\{\hat{T}_{t,\alpha} \mid t\geqslant 0\}$ is a holomorphic semigroup (Yosida, T. Kato, and A. V. Balakrishnan, 1960). Its infinitesimal generator $\hat{A}_\alpha$ can be considered as the **fractional power** $(-A)^\alpha$ of $-A$, multiplied by -1.

Fractional powers $(-A)^\alpha$, $\alpha\in\mathbf{C}$, of operators have also been defined for operators A satisfying the weaker condition than (3′) that $\{\lambda(\lambda-A)^{-1} \mid \lambda>0\}$ is equicontinuous (Balakrishnan, H. Komatsu). If A is such an operator, $-\sqrt{-A}$ generates a holomorphic semigroup and the unique uniformly bounded solution of the "elliptic" equation

$$x_t''=-Ax_t, \quad t>0, \quad \lim_{t\to 0}x_t=x_0 \tag{8}$$

is the solution of

$$x_t'=-\sqrt{-A}\,x_t, \quad t>0, \quad \lim_{t\to 0}x_t=x_0,$$

and therefore $x_t=\exp(-t\sqrt{-A})x_0$ (Balakrishnan). Equation (8) has also been discussed by

M. Sova and H. O. Fattorini from a different point of view.

E. Convergence of Semigroups

Let a sequence $\{\exp(tA_n) \mid n=1,2,\ldots\}$ of e.c.s.g. (C^0) be equicontinuous as a family of operators $\in L(X)$. Then a necessary and sufficient condition for there to exist an e.c.s.g. (C^0) $\exp(tA)$ such that $\lim_{n\to\infty}(\exp(tA_n))x=(\exp(tA))x$ uniformly on every compact interval of t is that $\lim_{n\to\infty}(\lambda_0 I-A_n)^{-1}x=J_{\lambda_0}x$ exist (for some λ_0 with $\mathrm{Re}\,\lambda_0>0$ and for all $x\in X$) and be such that $R(J_{\lambda_0})$ is dense in X (H. F. Trotter, Kato).

F. Miscellaneous Semigroups

(i) **Distribution semigroups.** The semigroup of translations $(T_t x)(s)=x(t+s)$ in $X=L_\infty(-\infty,\infty)$ is not continuous and hence not measurable in t. However, $T_t x$ is an X-valued distribution. For semigroups $\{T_t\}$ such that $T_t x$ is an X-valued distribution for $x\in X$, an analog of the Hille-Yosida theorem is known (J.-L. Lions, 1960). It has been generalized to ultradistribution semigroups by J. Chazarain and to hyperfunction semigroups by S. Ōuchi.

(ii) **Dual semigroups.** The above semigroup $\{T_t\}$ of translations in $L_\infty(-\infty,\infty)$ is obtained as $T_t=S_t^*$ from the e.c.s.g. (C^0) $\{S_t\}$ defined by $(S_t x)(s)=x(s-t)$ in $L_1(-\infty,\infty)$. Let $B=d/ds$ be the infinitesimal generator of $\{S_t\}$. The restriction of $\{S_t^*\}$ to the space of uniformly continuous functions, which is the closure of the domain $D(B^*)$ of the dual B^* in $L_\infty(-\infty,\infty)$, is an e.c.s.g. (C^0). This fact holds for the semigroup $\{S_t^*\}$ of an e.c.s.g. (C^0) $\{S_t\}$ in a Banach space X in general (R. Phillips, 1955) and also in a locally convex space.

(iii) **Locally equicontinuous semigroups.** The infinitesimal generator A of the semigroup of translations $(T_t x)(s)=x(t+s)$ in $X=C(-\infty,\infty)$ is d/ds. A has no resolvent since all complex numbers are eigenvalues of A. $\{T_t\}$ is not equicontinuous but locally equicontinuous, i.e., $\{T_t \mid 0\leqslant t\leqslant t_1\}$ is equicontinuous for any $t_1>0$. For locally equicontinuous (C^0) semigroups an analog of the Hille-Yosida theorem is obtained by using the notion of generalized resolvents (T. Kômura, 1968; Ōuchi, 1973).

(iv) **Differentiable semigroups.** The notion of the holomorphy of semigroups in Section D is weakened to the differentiability. A characterization of a semigroup $\{T_t\}$ such that $T_t x$ is infinitely differentiable in $t>0$ is given by using the resolvent of the infinitesimal generator (A. Pazy, 1968).

(v) **Nonlinear semigroups.** For a (C^0) semigroup $\{T_t\}$ of contractions (i.e., $\|T_t x-T_t y\|\leqslant\|x-y\|$ for $x,y\in X$) in a Hilbert space X, an analog of the Hille-Yosida theorem is known (Y. Kômura, 1969). This result has been partially extended to Banach spaces (→ 286 Nonlinear Functional Analysis X).

G. The Evolution Equation

Let $T_t=\exp(tA)$ be an e.c.s.g. (C^0). Then for $x\in D(A)$,

$$T_t'x=AT_t x(=T_t Ax). \tag{9}$$

Considered in suitable function spaces, the †equation of heat conduction ($A=\Delta=$ the †Laplacian), the †Schrödinger equation ($A=\sqrt{-1}\,(\Delta-V(x))$), and the †wave equation given in matrix form

$$\begin{pmatrix}u\\v\end{pmatrix}'=\begin{pmatrix}u'\\v'\end{pmatrix}=\begin{pmatrix}0&I\\\Delta&0\end{pmatrix}\begin{pmatrix}u\\v\end{pmatrix}$$

are all of the form (9). For a linear operator A_t in X depending on t, the ordinary differential equation in X

$$x_t'=A_t x_t+f(t),\qquad t\geqslant 0, \tag{10}$$

is called the **evolution equation**. A family of operators $\{V(r,s) \mid r\geqslant s\geqslant 0\}$ in X which gives general solutions to the homogeneous evolution equation

$$x_t'=A_t x_t \tag{11}$$

(i.e., for any $s\geqslant 0$, $a\in D(A_s)$, $x_t=V(t,s)a$ is a solution to (11) for $x_s=a$) is called the **evolution operator** associated with the generators $\{A_t\}$. An evolution operator $\{V(r,s)\}$ satisfies (i) $V(r,r)=I$, (ii) $V(r,s)V(s,t)=V(r,t)$. The solution to (10) is formally expressed by

$$x_t=V(t,0)x_0+\int_0^t V(t,s)f(s)\,ds. \tag{12}$$

H. Integration of the Evolution Equation

For equation (11) we have the following result (Kato, 1953; Yosida, 1966). Assume the following four conditions: (i) $D(A_t)$ is independent of t and dense in the Banach space X, and for all $\alpha>0$, $(I-\alpha A_t)^{-1}\in L(X)$ with the estimate $\|(I-\alpha A_t)^{-1}\|\leqslant 1$; (ii) $B_{t,s}=(I-A_t)(I-A_s)^{-1}$ is uniformly bounded in the norm for $0\leqslant s,t\leqslant l$; (iii) $\sum_{j=0}^{n-1}\|B_{t_{j+1},t_0}-B_{t_j,t_0}\|\leqslant N$, where N is independent of the partition ($0=t_0<t_1<\ldots<t_n=l$); (iv) $B_{t,0}$ is weakly differentiable with respect to t such that the differentiated operator $\partial B_{t,0}/\partial t$ is strongly continuous in t. Under these assumptions, we can prove that for $x_0\in D(A_0)$, the limit $V(t,0)x_0=\text{s-}\lim_{k\to\infty}V_k(t,0)x_0$,

with

$$V_k(t,s)=\left(I-\left(t-\frac{[kt]}{k}\right)A\left(\frac{[kt]}{k}\right)\right)^{-1}$$
$$\times\left(I-\frac{1}{k}A\left(\frac{[kt]-1}{k}\right)\right)^{-1}$$
$$\times\ldots\times\left(I-\frac{1}{k}A\left(\frac{[ks]+1}{k}\right)\right)^{-1}$$
$$\times\left(I-\left(\frac{[ks]+1}{k}-s\right)A\left(\frac{[ks]}{k}\right)\right)^{-1}$$

($0\leqslant s\leqslant t\leqslant l$), exists and gives the unique solution of (11). If $f(t)$ is continuously differentiable, the right-hand side of (12) exists and gives a unique solution to the inhomogeneous equation (10).

I. The Evolution Equation of Parabolic Type

Equation (10), for which every A_t is the infinitesimal generator of a holomorphic semigroup, is said to be of **parabolic type** by analogy to parabolic partial differential equations. Under weaker conditions, especially without the condition that $D(A_t)$ is independent of t, the existence of solutions of an equation of this type is obtained. Moreover, differentiability or analyticity of solutions follow from some natural assumptions.

(i) Existence of weak solutions. Let X be a Hilbert space. For $t\in[0,l]$, let V_t be a subspace and at the same time a Hilbert space with respect to a norm $|||\cdot|||_t$ stronger than $\|\cdot\|$. Since the form (A_tx,y) is †sesquilinear (linear in x and antilinear in y), we get a sesquilinear functional $a(t,\cdot,\cdot)$ on $V_t\times V_t$ such that

$$a(t,x,y)=-(A_tx,y),\qquad x,y\in D(A_t),$$

if $D(A_t)$ is dense in V_t with respect to $|||\cdot|||_t$ and if

$$|(A_tx,y)|\leqslant C|||x|||_t|||y|||_t,\qquad x,y\in D(A_t).$$

V_t, $a(t,\cdot,\cdot)$ should be measurable in a certain sense. A solution x_t of the equation (10) in $[0,l]$ satisfies

$$\int_0^l a(t,x_t,v_t)\,dt-\int_0^l (x_t,v_t')\,dt$$
$$=\int_0^l (f(t),v_t)\,dt+(x_0,v_0)\qquad(13)$$

for any differentiable X-valued function v_t such that $v_t\in V_t$, $v_l=0$, $\int_0^l\|v_t\|^2dt<\infty$, and $\int_0^l\|v_t'\|^2dt<\infty$. A solution x_t of (13) is called a **weak solution** of equation (10), though it does not necessarily satisfy (10). If the relation

$$a(t,x,x)+\lambda\|x\|^2\geqslant\alpha|||x|||_t,\qquad x\in V_t,$$

holds for some λ, $\alpha>0$, a weak solution of (10) in the sense of (13) exists for a given $x_0\in X$ (Lions, 1961). In order to obtain the uniqueness or the differentiability of weak solutions, we need some additional conditions.

(ii) Some properties of strong solutions. Let X be a Banach space. Let every semigroup $\{T_\lambda^{(t)}\}$ generated by A_t be holomorphic in a complex sector $|\arg\lambda|\leqslant\theta$, $\theta>0$, independent of t. Suppose one of the following conditions holds: (1) For some α, $0<\alpha<1$, $D(A_t^\alpha)$ is independent of t and for $1-\alpha<\beta<1$,

$$\|(A_t^\alpha-A_s^\alpha)A_0^{-\alpha}\|\leqslant C'|t-s|^\beta,\qquad t,s\in[0,1]$$

(P. E. Sobolevskiĭ, 1958–1961; Kato, 1961); (2) A_t^{-1} is differentiable in t,

$$\|dA_t^{-1}/dt-dA_s^{-1}/ds\|\leqslant C'|t-s|^\beta$$

for some $C'>0$, $0<\beta\leqslant1$, and

$$\left\|\frac{\partial}{\partial t}(A_t-\lambda)^{-1}\right\|\leqslant\frac{N}{|\lambda|^\alpha}$$

for every $\lambda:|\arg\lambda|\geqslant\pi/2-\theta$ for some N, $0<\alpha<1$ (Kato and H. Tanabe, 1962). Then a differentiable evolution operator $\{V(t,s)\}$ associated with (11) exists.

The most interesting property of evolution equations of parabolic type is the analyticity of solutions. If A_t is holomorphic in t in a certain sense, then the solutions are holomorphic (Tanabe, 1967; first noted by Komatsu, 1961). Furthermore, a characterization of evolution operators $\{V(t,s)\}$ holomorphic in some complex neighborhood of $[0,l]$ (called **holomorphic evolution operators**) is obtained by using the resolvent of A_t (Kato and Tanabe, 1967; K. Masuda, 1972; → [8]).

J. Application to Semilinear Evolution Equations

The evolution equation with a nonlinear additive term $f(t,x_t)$: $x_t'=A_tx_t+f(t,x_t)$ can be written as an inhomogeneous integral equation $x_t=V(t,0)x_0+\int_0^t V(t,s)f(s,x_s)\,ds$ in the Banach space X, by means of the evolution operator $\{V(t,s)\}$ introduced in Section G. The existence, differentiability (Kato, H. Fujita, and Sobolevskiĭ, 1963–1966), and analyticity (Masuda, 1967) of solutions of the Navier-Stokes equation has been obtained by reducing it to an integral equation of this type.

Concerning quasilinear equations in which A_t depends on x_t, the existence, differentiability, and analyticity of their solutions have been discussed.

References

[1] E. Hille and R. S. Phillips. Functional analysis and semi-groups, Amer. Math. Soc. Colloq. Publ., 1957.

[2] N. Dunford and J. T. Schwartz, Linear operators I, Interscience 1958.
[3] J. L. Lions, Equations différentielles opérationnelles et problèmes aux limites, Springer, 1961.
[4] K. Yosida, Functional analysis, Springer, 1980.
[5] S. G. Kreĭn, Linear differential equations in a Banach space, Amer. Math. Soc., 1971. (Original in Russian, 1967.)
[6] P. L. Butzer and H. Berens, Semigroups of operators and applications, Springer, 1967.
[7] R. W. Carroll, Abstract methods in partial differential equations, Harper & Row, 1969.
[8] H. Tanabe, Equations of evolution, Pitman, 1979. (Original in Japanese, 1975.)

379 (X.18) Series

A. Convergence and Divergence of Infinite Series

Let $\{a_n\}$ $(n=1,2,3,\ldots)$ be a sequence of real or complex numbers. Then the formal infinite sum $a_1+a_2+\ldots$ is called an **infinite series** (or **series**) and is denoted by $\sum_{n=1}^{\infty} a_n$ or $\sum a_n$. The number a_n is the nth **term** of the series $\sum a_n$, and $s_n=a_1+a_2+\ldots+a_n$ is the nth **partial sum** of $\sum a_n$. Also, for a finite sequence $a_1, a_2, \ldots, a_n$, the sum $a_1+a_2+\ldots+a_n$ is called a series. To distinguish these two series, the latter is called a **finite series**. In this article, *series* means an infinite series. If the sequence of partial sums $\{s_n\}$ †converges to s, we say that the series $\sum a_n$ **converges** or is **convergent** to the **sum** s and write $\sum_{n=1}^{\infty} a_n=s$ or $\sum a_n=s$. If the sequence $\{s_n\}$ is not convergent, we say that the series **diverges** or is **divergent**. In particular, if $\{s_n\}$ is divergent to $+\infty$ $(-\infty)$ or †oscillating, we say that the series is **properly divergent** to $+\infty$ $(-\infty)$ or **oscillating**, respectively.

The notation $\sum a_n$ is customarily used for both the sum s of the convergent series and the formal series, which may or may not be convergent. When the series is convergent, the sum is sometimes called the **Cauchy sum** in contrast to the "summations" of series, which are not necessarily convergent (→ Sections K ff.).

Applying the †Cauchy criterion for the convergence of a sequence, we see that a necessary and sufficient condition for $\sum a_n$ to be convergent is that for any given $\varepsilon>0$, we can take N sufficiently large so that

$$|s_m-s_n|=|a_{n+1}+\ldots+a_m|<\varepsilon$$

for all m, n such that $m>n>N$. Hence if $\sum a_n$ converges, the $a_n \to 0$ as $n\to\infty$, but the converse is not always true.

Elementary properties of the convergence of series are: (1) If $\sum a_n$ and $\sum b_n$ converge to a, b, respectively, then $\sum(a_n+b_n)$ converges to $a+b$. (2) If $\sum a_n$ converges to a, then $\sum ca_n$ converges to ca for any constant c. (3) Suppose that $\{b_m\}$ is a subsequence of $\{a_n\}$ obtained by deleting a finite number of terms a_n from $\{a_n\}$. Then $\sum b_m$ is convergent if and only if $\sum a_n$ is convergent. (4) When a series $\sum a_n$ converges to a and $\{b_m\}$ is a sequence such that $b_1=a_1+a_2+\ldots+a_{r_1}$, $b_2=a_{r_1+1}+a_{r_1+2}+\ldots+a_{r_2}$, $b_3=a_{r_2+1}+\ldots+a_{r_3}, \ldots$, then $\sum b_m$ also converges to a. The converse, however, is not always true. For example, $1-1+1-1+\ldots$ is oscillating, but $(1-1)+(1-1)+\ldots=0$.

B. Series of Positive Terms

Suppose that $\sum a_n$ is a **series of positive** (or **nonnegative**) **terms**. Since its partial sums s_n form a †monotonically increasing sequence, the series is convergent if and only if $\{s_n\}$ is bounded. For example, the series $\sum_{n=1}^{\infty} n^{-p}$ $(p>0)$ converges if $p>1$ because $s_n<2^{p-1}/(2^{p-1}-1)$, whereas it diverges if $p\leqslant 1$ because $s_{2^{m+1}}>1+(m+1)/2$. The **geometric series** $\sum_{n=1}^{\infty} a^{n-1}$ $(a>0)$ converges for $a<1$ because $s_n=(1-a^n)/(1-a)$, and diverges for $a\geqslant 1$ because $s_n\geqslant n$.

Some criteria for the convergence of a series $\sum a_n$ of nonnegative terms are: (1) If $\{a_n\}$ is monotone decreasing, then the series $\sum a_n$ and $\sum 2^\nu a_{2^\nu}$ have the same convergence behavior (**Cauchy's condensation test**). (2) Suppose that $f(x)$ is a positive monotone decreasing function defined for $x\geqslant 1$ such that $f(n)=a_n$ $(n=1,2,\ldots)$. Then the series $\sum a_n$ and the integral $\int_1^\infty f(x)\,dx$ have the same convergence behavior (**Cauchy's integral test**), for example, $\sum n^{-p}$ $(p>0)$ and $\int_1^\infty x^{-p}dx$. (3) If for any positive constant k we have $a_n\leqslant kb_n$ except for a finite number of n, then the convergence of $\sum b_n$ implies the convergence of $\sum a_n$. If $kb_n\leqslant a_n$ and $\sum b_n$ diverges, then $\sum a_n$ also diverges (**comparison test**). (4) Let $a_n>0$ and $b_n>0$. If $a_{n+1}/a_n\leqslant b_{n+1}/b_n$ except for a finite number of values of n and $\sum b_n$ converges, then $\sum a_n$ also converges; if $a_{n+1}/a_n\geqslant b_{n+1}/b_n$ and $\sum b_n$ diverges, then $\sum a_n$ also diverges (→ Appendix A, Table 10).

C. Absolute Convergence and Conditional Convergence

A series $\sum a_n$ (with real or complex terms a_n) is called **absolutely convergent** if the series $\sum |a_n|$ is convergent. If a convergent series is not

absolutely convergent, then it is called **conditionally convergent**. An absolutely convergent series is convergent. A real series $\sum a_n$ whose terms have alternating signs is called an **alternating series**. An alternating series $\sum a_n$ is convergent if the absolute values of terms $|a_n|$ form a monotone decreasing sequence which converges to zero (**Leibniz's test**). An absolutely convergent series remains absolutely convergent under every rearrangement of terms and retains its sum under the rearrangement (**Dirichlet's theorem**). If a series with real terms is conditionally convergent, then it is possible to rearrange its terms so that the rearranged series converges to any given number, diverges to $+\infty$ (or $-\infty$), or is oscillating (**Riemann's theorem**). A convergent series whose convergence behavior is unaffected by rearrangement and whose sum remains unchanged is called **unconditionally convergent** (or **commutatively convergent**). A real or complex series is unconditionally convergent if and only if it is absolutely convergent. The notion of infinite series can be extended to any complete †normed linear space, and absolute convergence can be defined by replacing the absolute values of the terms by the norm of the terms. However, in general, unconditional convergence is not always equivalent to absolute convergence.

D. Abel's Partial Summation

Let $\{a_0, a_1, a_2, \dots\}$ and $\{b_0, b_1, b_2, \dots\}$ be arbitrary sequences, and put $A_n = a_0 + a_1 + \dots + a_n$ for $n \geqslant 0$. Then the following formula of **Abel's partial summation** holds:

$$\sum_{\nu=n+1}^{n+k} a_\nu b_\nu = \sum_{\nu=n+1}^{n+k} A_\nu (b_\nu - b_{\nu+1}) - A_n b_{n+1} + A_{n+k} b_{n+k-1}$$

for any $n \geqslant 0$ and any $k \geqslant 1$; this formula also holds for $n = -1$ if we put $A_{-1} = 0$.

Abel's partial summation enables us to deduce a number of tests of convergence for series of the form $\sum a_n b_n$. In particular, the following criteria are easy to apply:

(1) $\sum a_n b_n$ is convergent if $\sum a_n$ is convergent and if the sequence $\{b_n\}$ is monotone and bounded (**Abel's test**).

(2) $\sum a_n b_n$ is convergent if the sequence $\{s_n\}$ of partial sums of $\sum a_n$ is bounded and if $\{b_n\}$ is monotone and converges to zero (**Dirichlet's test**).

(3) $\sum a_n b_n$ is convergent if $\sum (b_n - b_{n+1})$ is absolutely convergent and if $\sum a_n$ is (at least conditionally) convergent (**test of du Bois-Reymond and Dedekind**).

For example, criterion (2) implies that if $\{b_n\}$ is monotone and converges to zero, then the power series $\sum b_n z^n$ of a complex variable z is convergent on the unit circle $|z| = 1$ except at most for $z = 1$; the case $z = -1$ gives Leibniz's test for alternating series (→ Section C).

E. Double Series

A sequence with two indices, i.e., a mapping from the Cartesian product $\mathbf{N} \times \mathbf{N}$ of two copies of the set of natural numbers $\mathbf{N}$ to a subset of the real or complex numbers, is called a **double sequence** and is denoted by $\{a_{mn}\}$ or $\{a_{m,n}\}$. If there exists a number l such that for any positive ε there is a natural number $N(\varepsilon)$ satisfying $|a_{mn} - l| < \varepsilon$ for all $m > N(\varepsilon)$ and $n > N(\varepsilon)$, then we say that the sequence $\{a_{mn}\}$ has a limit l and write $\lim_{m\to\infty, n\to\infty} a_{mn} = l$. This limit should not be confused with repeated limits such as $\lim_{n\to\infty}(\lim_{m\to\infty} a_{mn})$. If $\lim_{m\to\infty} a_{mn} = \alpha_n$ uniformly in n and $\lim_{n\to\infty} \alpha_n = l$, then $\lim_{m\to\infty, n\to\infty} a_{mn} = l$. For a given double sequence $\{a_{mn}\}$, the formal series $\sum_{m,n=1}^{\infty} a_{mn}$ is called a **double series** and is sometimes denoted by $\sum a_{mn}$. In contrast with double series, the ordinary series discussed previously is called a **simple series**.

Given a double series $\sum a_{mn}$, when the double sequence of partial sums $s_{mn} = \sum_{k=1}^{m} \sum_{l=1}^{n} a_{kl}$ is convergent to s, then $\sum a_{mn}$ is said to be **convergent** to the **sum** s. On the other hand, if s_{mn} is not convergent, $\sum a_{mn}$ is said to be **divergent**. If $\sum_{n=1}^{\infty} a_{mn}$ converges to b_m for each m, then $\sum_{m=1}^{\infty} b_m = \sum_{m=1}^{\infty}(\sum_{n=1}^{\infty} a_{mn})$ is called the **repeated** (or **iterated**) **series by rows**. If $\sum_{m=1}^{\infty} a_{mn}$ converges to c_n for each n, then $\sum_{n=1}^{\infty} c_n = \sum_{n=1}^{\infty}(\sum_{m=1}^{\infty} a_{mn})$ is called the **repeated** (or **iterated**) **series by columns**. Even if two repeated series by rows and columns are convergent, the two sums are not always identical, and $\sum a_{mn}$ is not always convergent. However, if the double series $\sum a_{mn}$ is convergent and $\sum_n a_{mn}$ is convergent for each m, then the repeated series by rows is convergent to the same sum. A similar statement is valid for the repeated series by columns.

Suppose that we are given a double series $\sum a_{mn}$ of nonnegative terms. If any one of $\sum_{m,n} a_{mn}$, $\sum_m \sum_n a_{mn}$, and $\sum_n \sum_m a_{mn}$ is convergent, the other two converge to the same sum. If the **diagonal partial sum** $s_{mm} = \sum_{k=1}^{m} \sum_{l=1}^{m} a_{kl}$ converges to a, then the double series $\sum a_{mn}$ also converges to a.

If $\sum |a_{mn}|$ converges, the double series $\sum a_{mn}$ is called **absolutely convergent**, whereas if $\sum a_{mn}$ converges but not absolutely, then $\sum a_{mn}$ is called **conditionally convergent**. If $\sum a_{mn}$ is absolutely convergent, then any series obtained from $\sum a_{mn}$ by arranging the terms in an arbitrary order is convergent to the same sum.

F. Multiplication of Series

The series $\sum_{n=1}^{\infty} c_n$, where $c_n = a_1 b_n + a_2 b_{n-1} + \ldots + a_n b_1$, is called the **Cauchy product** of two series $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$. (1) Let $\sum a_n$ and $\sum b_n$ be two convergent series and A, B be the sums of these series. If their Cauchy product $\sum c_n$ is also convergent, then it has the sum $C = AB$ (**Abel's theorem**). (2) If at least one of the two convergent series $\sum a_n$ and $\sum b_n$ with the sums A, B, respectively, is absolutely convergent, then their Cauchy product $\sum c_n$ is also convergent and has the sum $C = AB$ (**Mertens's theorem**). (3) If $\sum a_n$ and $\sum b_n$ are absolutely convergent, then their Cauchy product $\sum c_n$ is absolutely convergent (**Cauchy's theorem**). (4) Let $\sum a_n$ and $\sum b_n$ be two convergent series with the sums A, B, respectively. If $\{na_n\}$ and $\{nb_n\}$ are bounded from below, then $\sum c_n$ is convergent and has the sum $C = AB$ (**Hardy's theorem**).

G. Infinite Product

Let $\{a_n\}$ be a given sequence with terms $a_n \neq 0$ $(n = 1, 2, \ldots)$. The formal **infinite product** $a_1 \cdot a_2 \cdot a_3 \cdot \ldots$ is denoted by $\prod_{n=1}^{\infty} a_n$. We call $p_n = a_1 \cdot a_2 \cdot \ldots \cdot a_n$ its nth **partial product**. If the sequence $\{p_n\}$ is convergent to a nonzero limit p, then this infinite product is said to **converge** to p, and p is called the **value** of the infinite product. We write $\prod a_n = p$. If $\{p_n\}$ is not convergent or is convergent to 0, then the infinite product is called **divergent**. Sometimes we consider the infinite product $\prod a_n$ with $a_n = 0$ for a finite number of n's; and then by convergence or divergence of $\prod a_n$ we mean that of the infinite product $\prod a'_n$, where the sequence $\{a'_n\}$ is obtained by deleting zero terms from $\{a_n\}$. Usually we do not treat an infinite product with $a_n = 0$ for an infinite number of n's.

A necessary and sufficient condition for $\prod_{n=1}^{\infty} a_n$ to be convergent is that for any positive ε there is a number N such that $|p_m/p_n - 1| < \varepsilon$ for all $n, m > N$. If $\prod a_n$ converges, then $a_n \to 1$, but the converse is not always true.

It is often convenient to write an infinite product as $\prod(1 + a_n)$. Then $\prod(1 + a_n)$ and $\sum \log(1 + a_n)$ have the same convergence behavior, where the imaginary part $i\theta$ of the logarithm is assumed to satisfy $0 \leqslant |\theta| < \pi$. If $a_n \geqslant 0$, convergence of $\prod(1 + a_n)$ implies convergence of $\sum a_n$, and vice versa.

If $\prod(1 + |a_n|)$ converges, then $\prod(1 + a_n)$ is said to be **absolutely convergent**. An absolutely convergent infinite product is also convergent, and the value of the infinite product is unchanged by the alteration of the order of terms.

H. Termwise Differentiation of Infinite Series with Function Terms

Uniform convergence of an infinite series $\sum f_n(x)$ is defined by uniform convergence of the sequence of the partial sums $\sum_{k=1}^{n} f_k(x)$ (→ 435 Uniform Convergence). If the infinite series $\sum f_n(x)$ defined on an interval I of the real line is convergent at least at one point of I and $\sum f'_n(x)$ is convergent uniformly in I when the derivatives $f'_n(x)$ exist, then $\sum f_n(x)$ is convergent to $f(x)$ uniformly in I, and $\sum f_n(x)$ is **termwise differentiable**, that is, $f'(x) = \sum f'_n(x)$. If the $\varphi_n(z)$ are holomorphic in a complex domain D and $\sum \varphi_n(z)$ converges to $\varphi(z)$ uniformly on every compact subset of D, then $\sum \varphi'_n(z)$ also converges to $\varphi'(z)$ uniformly on every compact subset of D (**Weierstrass's theorem of double series**). (For termwise integration → 216 Integral Calculus.)

I. Numerical Evaluation of Series

In some special cases, we can express the nth partial sum s_n of a series $\sum a_n$ as a well-known function of n. Specifically, if $\sum a_n$ is an **arithmetic progression** $\sum_{k=1}^{n}(a + (k-1)d)$ or a **geometric progression** $\sum_{k=1}^{n} aq^{k-1}$, we have

$$s_n = \frac{n}{2}(2a + (n-1)d), \qquad s_n = \frac{a(q^n - 1)}{q - 1},$$

respectively. If $|q| < 1$, then $\sum_{n=0}^{\infty} aq^n$ converges to $a/(1-q)$. If $B_{r+1}(x)$ is the $(r+1)$st †Bernoulli polynomial, then $s_n = 1^r + 2^r + \ldots + n^r = [B_{r+1}(x)]_1^{n+1}/(r+1)$. This sum was studied by J. Bernoulli, who gave formulas up to $r = 10$ in his *Ars conjectandi*.

In the series $\sum u_n$, if we can find another sequence $\{v_n\}$ such that $u_n = v_n - v_{n-1}$, then $s_n = u_1 + u_2 + \ldots + u_n = v_n - v_0$. For example, if $u_n = n(n+1)(n+2)$, then $v_n = n(n+1)(n+2)(n+3)/4$ and $s_n = v_n$ because $v_0 = 0$ (→ 104 Difference Equations). Series with trigonometric function terms are calculated analogously.

There are cases where the sum $\sum a_n$ itself can be expressed in a satisfactory form although we cannot find an appropriately simple expression for each partial sum s_n. For example, $\zeta(r) = \sum_{m=1}^{\infty} 1/m^r$ can be represented by †Bernoulli numbers if r is even. In particular, $\zeta(2) = \pi^2/6$, $\zeta(4) = \pi^4/90$ (→ Appendix A, Table 10).

If an infinite series converges rapidly, we can get a good approximation by taking a suitable partial sum. On the other hand, if the series converges less rapidly, an effective means for evaluating series is afforded by transformation of series. If the kth †difference is exactly zero, then

$$s_n = \sum_{i=1}^{k} \binom{n+1}{i} \Delta^{i-1} u_1.$$

Since the absolute value of finite differences often decreases rapidly, it is sometimes convenient to consider the series whose terms are the differences of the original series. One finite difference method is **Euler's transformation** of infinite series. In particular, the formula

$$\sum_{k=0}^{\infty}(-1)^k a_k=\sum_{n=0}^{\infty}\frac{\Delta^n a_0}{2^{n+1}}$$

is useful for numerical calculation of sums of slowly converging alternating series. In numerical calculation of the series, we usually start calculating the numerical values of the first few terms; we then apply such transformations as Euler's to the remainder, and calculate the partial sums of the transformed series.

When we calculate the sum of an infinite series approximately, we must estimate the error, i.e., the remainder that must be added to yield the sum of the series itself. We can estimate the maximum error by derivatives or differences of higher orders. We also have the transformations of Markov and Kummer. In the former, every term of the series is represented by convergent series, and in the latter, the given series is reduced by subtracting another convergent series, which has a known sum and similar terms [1].

J. Infinite Series and Integrals

In numerical calculation of functions, we sometimes use the **Euler-Maclaurin formula** [5]:

$$f(x+\xi\omega)=\frac{1}{\omega}\int_x^{x+\omega}f(z)\,dz+\sum_{r=1}^{m}\left(\frac{\omega^r}{r!}B_r(\xi)\Delta_\omega f^{(r-1)}(x)\right)+R_m,$$

$$R_m=-\omega\int_0^1\frac{\bar{B}_m(\xi-z)}{m!}f^{(m)}(x+\omega z)\,dz,$$

where

$$\bar{B}_m(\xi-z)=\begin{cases}B_m(\xi-z+1), & \xi<z,\\ B_m(\xi-z), & \xi\geqslant z.\end{cases}$$

The speed of the convergence for this formula is greater than that for Taylor's expansion when $\xi\omega$ is large, since the terms of the formula are †Bernoulli polynomials $B_r(\xi)$ in $0\leqslant\xi\leqslant 1$. We also have Boole's formula, with †Euler polynomials as its terms [4]. The formulas discussed in this section are also used to calculated approximately the partial sums of infinite series.

Another method of evaluating sums of infinite series analytically entails transforming infinite series to definite integrals using the †residue theorem. If an analytic function $f(z)$ is holomorphic except at poles a_n $(n=1,2,\ldots,k)$ in a domain bounded by the closed curve C and containing the points $z=m$ $(m=1,2,\ldots,N)$, then

$$\sum_{m=1}^{N}f(m)=\frac{1}{2\pi i}\int_C\pi(\cot\pi z)f(z)\,dz-\sum_{n=1}^{k}\operatorname{Res}[\pi(\cot\pi z)f(z)]_{z=a_n}.$$

When the left-hand side of this equation is replaced by

$$\sum_{m=1}^{N}(-1)^m f(m),$$

we replace $\cot\pi z$ by $\operatorname{cosec}\pi z$. $\operatorname{Res}[F(z)]_{z=a}$ is the †residue of $F(z)$ at $z=a$. The line integral along C is often calculated easily by choosing a suitable deformation of C. Sometimes it can be shown immediately that the integral along C is zero, or its asymptotic value can be evaluated by the †method of steepest descent.

K. History of the Study of Divergent Series

Mathematicians in the 18th century did not concern themselves with the question of whether †series were †convergent or †divergent. This indiscriminateness led to various contradictions. In 1821 an exact definition of the notion of convergence of series (Section A) was given by A. L. Cauchy; since then, mathematicians have mostly concerned themselves with convergent series. However, since divergent series appeared in many problems in analysis, the study of such series could not be neglected, and it became desirable to give a suitable definition of their sum. Although some results were given by L. Euler, N. Abel, and others, it was during the latter part of the 19th century that methods of summation of divergent series were studied systematically. This study constituted a new branch of mathematics.

In the following sections, some important methods of summation of divergent series are mentioned. Cesàro's method (→ Section M) was the forerunner of the theory whose general foundation is now the theory of linear transformations.

L. Linear Transformations

For a sequence $\{s_n\}$ $(n=0,1,2,\ldots)$ of real or complex numbers, assume that $\sigma_n=\sum_{i=0}^{\infty}a_{ni}s_i$ converges for $n=0,1,2,\ldots$, where (a_{ik}) is a given matrix $(i,k=0,1,2,\ldots)$. The mapping $T:\{s_n\}\rightarrow\{\sigma_n\}$ is called a **linear transformation**, and $\{\sigma_n\}$ is called the **transform** of $\{s_n\}$ under T. If the matrix satisfies $a_{ik}=0$ $(k>i)$, then T is defined for any sequence $\{s_n\}$ and T is said to

be **triangular**. If the transform $\{\sigma_n\}$ under T is defined and convergent whenever $\{s_n\}$ converges, then T is called a **semiregular transformation.** If in addition $\{\sigma_n\}$ has the same limit as $\{s_n\}$, then T is called a **regular transformation.** If for any bounded sequence $\{s_n\}$ the transform $\{\sigma_n\}$ is defined and convergent, then T is called a **normal transformation.** If T is triangular and the transform $\{\sigma_n\}$ of $\{s_n\}$ under T is divergent to ∞ whenever $s_n \to \infty$, then T is called a **totally regular transformation.**

Let T be a regular transformation. If for at least one divergent sequence $\{s_n\}$ the transform $\{\sigma_n\}$ of $\{s_n\}$ under T converges, then T is called a **method of summation.** The limit s of $\{\sigma_n\}$ is called the **sum** of $\{s_n\}$ under the method T of summation, and $\{s_n\}$ is said to be **T-summable** to s. For a given method of summation T_1, let $D(T_1)$ be the set of sequences that are T_1-summable. If $D(T_1)=D(T_2)$, then the methods T_1 and T_2 are called **equivalent.** If $D(T_1)\subset D(T_2)$, then we say that T_1 is **weaker** than T_2 and T_2 is **stronger** than T_1. If $D(T_1)\not\subset D(T_2)$ and $D(T_1)\not\supset D(T_2)$, then T_1 and T_2 are called **mutually noncomparable.** The following theorems on linear transformations of sequences are important:

(1) **Kojima-Schur theorem.** In order that T be semiregular it is necessary and sufficient that (i) $\lim_{n\to\infty} a_{nk}$ exist for each k; (ii) $t_n=\sum_{k=0}^{\infty}|a_{nk}|$ exist and $\{t_n\}$ be bounded; and (iii) $\lim_{n\to\infty}\sum_{k=0}^{\infty}a_{nk}$ exist. In that case, we have

$$\sum_{k=0}^{\infty}|\lim_{n\to\infty} a_{nk}|<\infty$$

and

$$\lim_{n\to\infty}\sigma_n=\lim_{n\to\infty}\sum_{k=0}^{\infty}a_{nk}s_k$$

$$=\left(\lim_{n\to\infty}s_n\right)\left(\lim_{n\to\infty}\sum_{k=0}^{\infty}a_{nk}\right)$$

$$+\sum_{k=0}^{\infty}\left(\lim_{n\to\infty}a_{nk}\right)\left(s_k-\lim_{n\to\infty}s_n\right).$$

In particular, in order that $\{\sigma_n\}$ converge whenever $s_n\to 0$ it is necessary and sufficient that conditions (i) and (ii) be satisfied. In that case,

$$\lim_{n\to\infty}\sigma_n=\lim_{n\to\infty}\sum_{k=0}^{\infty}a_{nk}s_k$$

$$=\sum_{k=0}^{\infty}\left(\lim_{n\to\infty}a_{nk}\right)s_k$$

(I. Schur, *J. Reine Angew, Math.*, 151 (1921); T. Kojima, *Tôhoku Math. J.*, 12 (1917)).

(2) **Toeplitz's theorem.** In order that T be regular it is necessary and sufficient that (i′) $\lim_{n\to\infty}a_{nk}=0$ for each k; (ii); and (iii′) $\lim_{n\to\infty}\sum_{k=0}^{\infty}a_{nk}=1$ (O. Toeplitz, *Prace Mat.-Fiz.*, 22 (1914)). In particular, (ii) and (i″) $\lim_{n\to\infty}\sum_{k=K}^{\infty}a_{nk}=1$ for each K imply that T is regular (**Perron's theorem**).

(3) **Schur's theorem.** In order that T be normal it is necessary and sufficient that (i); (ii); (iii); and (iv) for any $\varepsilon>0$ there exist a $K>0$ such that $\sum_{k=K+1}^{\infty}|a_{nk}|<\varepsilon$ for each n.

(4) In order that the regular triangular transformation T be totally regular, it is necessary and sufficient that $a_{nk}\geqslant 0$ except for a finite number of k.

M. Cesàro's Method of Summation

We write

$$(1-x)^{-\alpha-1}=\sum_{n=0}^{\infty}A_n^{\alpha}x^n,$$

$$A_n^{\alpha}=\binom{n+\alpha}{n}\sim\frac{n^{\alpha}}{\Gamma(\alpha+1)},$$

$$(1-x)^{-\alpha-1}\sum_{n=0}^{\infty}u_nx^n=(1-x)^{-\alpha}\sum_{n=0}^{\infty}s_nx^n$$

$$=\sum_{n=0}^{\infty}s_n^{\alpha}x^n,$$

where $s_n=\sum_{i=0}^{n}u_i$. Thus the series $\sum u_i$ is associated with the sequence $\{s_n^{\alpha}\}$. If $\sigma_n^{(\alpha)}=s_n^{\alpha}/A_n^{\alpha}$ converges to s as $n\to\infty$, then we say that $\sum u_n$ is **summable by Cesàro's method of order** α (or simply (C, α)-summable) to s and write $\sum_{n=0}^{\infty}u_n=s$ (C, α). This method of summation is called **Cesàro's method of summation of order** α (or simply **(C, α)-summation**).

It is natural to consider (C, α)-summation for $\alpha>-1$. We say that $\sum u_n$ is (C, -1)-summable if $\sum u_n$ converges and $nu_n=o(1)$. Here $\sum_{n=0}^{\infty}u_n=s$ (C, 0) means $\lim_{n\to\infty}s_n=s$, and $\sum_{n=1}^{\infty}u_n=s$ (C, 1) means $s=\lim_{n\to\infty}(s_0+s_1+\ldots+s_{n-1})/n$. Generally, we have the following results:

(1) A_n^{α} is increasing if $\alpha>0$ and decreasing if $0>\alpha>-1$. $A_n^0=1$ and $A_n^{\alpha}>0$ if $\alpha>-1$.

(2) $s_n^{\alpha+\beta+1}=\sum_{k=0}^{n}A_{n-k}^{\beta}s_k^{\alpha}$, $A_n^{\alpha}-A_{n-1}^{\alpha}=A_n^{\alpha-1}$, $s_n^{\alpha}-s_{n-1}^{\alpha}=s_n^{\alpha-1}$.

(3) (C,α) $(\alpha\geqslant 0)$ is regular, and $D(C,\alpha)\supset D(C,\beta)$ if $\alpha>\beta>-1$.

(4) If $\sum u_n=s$ (C, α), then $u_n=o(n^{\alpha})$. Moreover, if $\sum u_n'=s'$ (C, α), then $\sum(u_n+u_n')=s+s'$ (C, α) and $\sum\lambda u_n=\lambda s$ (C, α) for any number λ.

(5) If $\sum u_n=s$ (C, α) and $\sum u_n'=s'$ (C, β), then their †Cauchy product $\sum v_n=ss'$ (C, $\alpha+\beta+1$) (**Chapman's theorem**). Moreover, if $\sum_{k=0}^{n}A_k^{\alpha}|s_{n-k}^{\beta-\alpha-1}(u_{n-k}')|/A_n^{\beta}=O(1)$, then $\sum v_n=ss'$ (C, β) (T. Kojima). If α', $\beta'>-1$, $s_n^{(\alpha')}(u_n)=O(n^{\alpha'})$, and $s_n^{(\beta')}(u_n')=O(n^{\beta'})$, then $\sum v_n=ss'$ (C, $\alpha'+\beta'+2$) (G. Doetsch).

(6) For any integer $\alpha>0$, in order that $\sum u_n=s$ (C, α) it is necessary and sufficient that there exist $\{v_n\}$ such that $u_n=(n+1)(v_n-v_{n+1})$ and $\sum v_n=s$ (C, $\alpha-1$) (G. H. Hardy, *Proc. London Math. Soc.*, (2) 8 (1910)). This condi-

tion is equivalent to (i)–(iii) together: (i) the series $\sum_{k=n}^{\infty} s_n^{\alpha-2}/(k+1)\ldots(k+\alpha)$ converges to the limit b_n; (ii) $b_n=o(1)$ as $n\to\infty$; and (iii) $(s_n^{\alpha-1}/A_n^{\alpha-1})+(n+\alpha)\Gamma(\alpha)b_{n+1}\to s$ as $n\to\infty$.

(7) If $\sum u_n=s$ (C, α) ($\alpha>0$), one of the following five conditions is sufficient for the convergence of $\sum u_n$ (this is a kind of †Tauberian theorem): (i) $nu_n=o(1)$; (ii) $t_n=\sum_{\nu=1}^{n}\nu u_\nu=o(n)$; (iii) $\sum n^p|u_n|^{p+1}<\infty$ ($p\geqslant 0$); (iv) $nu_n>-K$ (K is independent of n); (v) $\liminf(s_m-s_n)\geqslant 0$ as $m>n\to\infty$, $m/n\to 1$ (**R. Schmidt's condition**).

(8) If $\alpha'>\alpha>-1$ and $\sum u_n=s(C,\alpha')$, then $\sum u_n=s$ (C, $\alpha+\varepsilon$) for any $\varepsilon>0$.

For a given series $\sum u_n$, we write H_n^0 for s_n, H_n^1 for the arithmetic mean of $\{H_n^0\}$, and H_n^2 for the arithmetic mean of $\{H_n^1\}$. Similarly, we can define $\{H_n^p\}$ for any integer p. If $H_n^p\to s$ as $n\to\infty$, then $\sum u_n$ is said to be **summable by Hölder's method of order** p (or **(H, p)-summable**) to s, and we write $\sum u_n=s(\mathrm{H},p)$. For any integer $p\geqslant 0$, (H, p)-summability is equivalent to (C, p)-summability (**Knopp-Schnee theorem**).

N. Abel's Method of Summation

If the radius of convergence of the power series $\sum_{n=0}^{\infty}u_n r^n$ is $\geqslant 1$ and $\sum_{n=0}^{\infty}u_n r^n\to s$ as $r\to 1$, then $\sum u_n$ is said to be **summable by Abel's method** (or **A-summable**) to s, and we write $\sum u_n=s$ (A). The transformation matrix is denoted by A, and the transformation is called **Abel's method of summation**.

(1) If $\sum u_n=s$ (A), then $\limsup_{n\to\infty}|u_n|^{1/n}\leqslant 1$. (2) If $\sum u_n=s$ (A) and $\sum u_n'=s'$ (A), then $\sum(u_n+u_n')s+s'$ (A) and $\sum\lambda u_n=\lambda s$ (A) for any constant λ. Moreover, $\sum_{n=k+1}^{\infty}u_n=s-u_0-u_1-\ldots-u_k$ (A). (3) If $\sum u_n=s$ (A) and $\sum u_n'=s'$ (A), then the Cauchy product is $\sum u_n=ss'$ (A). (4) If $\sum u_n=s$ (A) and one of the following five conditions is satisfied, then $\sum u_n=s$: (i) $nu_n=o(1)$; (ii) $t_n=\sum_{\nu=1}^{n}\nu u_\nu=o(n)$; (iii) $nu_n=O(1)$; (iv) $nu_n>-M$; (v) $\liminf(s_m-s_n)\geqslant 0$ as $m>n\to\infty$ and $m/n\to 1$. These theorems are †Tauberian, in the original form proved by Tauber (*Monatsh. Math.*, 8 (1897)). (5) The matrix A is regular, and $D(C,\alpha)\subset D(A)$ for any $\alpha>-1$. (6) If $\sum u_n=s$ (A) and $s_n\geqslant 0$, then $\sum u_n=s$ (C, 1). Moreover, if $\sigma_n^{(\alpha)}=O(1)$, then $\sum u_n=s$ (C, $\alpha+\varepsilon$) for $\alpha>-1$ and $\varepsilon>0$.

O. Borel's Method of Summation

If for a given series $\sum u_n$,

$$u(x)=\sum_{n=0}^{\infty}\frac{s_n x^n}{n!}$$

is convergent for all x, and $u(x)/e^x\to s$ as $x\to\infty$, then $\sum u_n$ is said to be **summable by Borel's exponential method** to the sum s, and we write $\sum u_n=s$ (B). The transformation thus determined is denoted by B and is called **Borel's method of summation**. If

$$\int_0^\infty u(x)e^{-x}dx=s,$$

then $\sum u_n$ is said to be **summable by Borel's integral method** (or **$\mathfrak{B}$-summable**) to s, and we write $\sum u_n=s$ ($\mathfrak{B}$). Then we have: (1) B is regular and $D(C,\alpha)\subset D(B)$ ($\alpha>-1$), while $D(C,\alpha)$ and $D(\mathfrak{B})$ are noncomparable. (2) If the radius of convergence of $\sum_{n=0}^{\infty}u_n x^n$ is $\geqslant 1$ and $\sum u_n=s$ (B), then $\sum u_n=s$ (A). (3) If $\sum_{n=k+1}^{\infty}u_n=s$ (B) (resp. ($\mathfrak{B}$)), then $\sum u_n=s+u_0+u_1+\ldots+u_k$(B) (resp. ($\mathfrak{B}$)), but the converse is not always true. (4) $\sum u_n=s$ (B) implies $|u_n|^{1/n}=o(n)$. (5) If $\sum u_n=s$ (B) and $\sum u_n'=s'$ (B), then $\sum(u_n+u_n')=s+s'$ (B), and $\sum\lambda u_n=\lambda s$ (B) for any constant λ. The same is true for summation ($\mathfrak{B}$). (6) If $\sum u_n=s$ (B) and if one of the following two conditions is satisfied, then $\sum u_n=s$: (i) $\sqrt{n}\,u_n=o(1)$; (ii) $\liminf(s_m-s_n)\geqslant 0$ as $m>n\to\infty$ and $(m-n)\sqrt{n}\to 0$. (7) if $\sum u_n=s$ (B) and $s_{n-1}^{\alpha-1}=o(n^{\alpha-1/2})$, then $\sum u_n=s$ (C, α) ($\alpha\geqslant 0$). (8) If $\sum u_n=s$ (A) and $u(t)>-Mt^{-1}\exp t$, then $\sum u_n=s$ (B). (9) If the sequences $\{n_k\}$ and $\{n_{k'}\}$ satisfy $n_{k+1}>n_{k'}$, $n_{k'}/n_k>1+\varepsilon$ ($k=1,2,\ldots$; $\varepsilon>0$), $u_\nu=0$ ($n_k<\nu\leqslant n_{k'}$), and $\sum u_n=s$ (B), then $s_{n_k}\to s$ as $k\to\infty$.

If $\sum u_n=s$ ($\mathfrak{B}$) and

$$\int_0^\infty\left|\frac{d^\lambda u(x)}{dx^\lambda}\right|e^{-x}dx$$

converges for all $\lambda=0,1,2,\ldots$, then $\sum u_n$ is said to be **absolute Borel summable** (or **$|\mathfrak{B}|$-summable**). Concerning this we have: (1) If $\sum|u_n|$ converges, then $\sum u_n$ is $|\mathfrak{B}|$-summable, but even if $\sum u_n$ converges, $\sum u_n$ is not always $|\mathfrak{B}|$-summable. If $\sum u_n$ is $|\mathfrak{B}|$-summable, then $\sum u_n$ is $\mathfrak{B}$-summable to s. In this case, we write $\sum u_n=s$ ($|\mathfrak{B}|$). (2) $\sum_{n=0}^{\infty}u_n=s$ ($|\mathfrak{B}|$) implies $\sum_{n=k+1}^{\infty}u_n=s-(u_0+u_i+\ldots+u_k)$ ($|\mathfrak{B}|$). (3) If $\sum u_n=s$ (B), $\sum u_n'=s'$ (B), and if at least one of them is $|\mathfrak{B}|$-summable, then their Cauchy product is $\sum v_n=ss'$ ($|\mathfrak{B}|$).

P. Euler's Method of Summation

In the series $\sum u_n$, if

$$\left\{\binom{k+1}{1}s_0+\binom{k+1}{2}s_1+\ldots+\binom{k+1}{k+1}s_k\right\}2^{-(k+1)}$$

($s_n=\sum_{k=0}^{n}u_k$) converges to s as $k\to\infty$, then $\sum u_n$ is said to be **summable by Euler's method**, and we write $\sum u_n=s$ (E). The transformation thus obtained is called **Euler's method of summation**. A necessary and sufficient condition

for $\sum u_n = s$ (E) is that $\sum v_n = s$, where $v_n = 2^{-(n+1)}\sum_{k=0}^{n}\binom{n}{k}u_k$. This summation method is regular. We have $\sum u_n = s$ if $\sum u_n = s$ (E) and if one of the following two conditions is satisfied: (i) $\sqrt{n}\, u_n = O(1)$; (ii) $\liminf(s_m - s_n) \geqslant 0$ for $m > n \to \infty$, $(m-n)/\sqrt{n} \to 1$. Cesàro and Euler summations are noncomparable. As an extension of Euler's method, the Euler method of summation of pth order is also defined (e.g., → [6]).

Q. Nörlund's Method of Summation

For a positive sequence $\{p_n\}$, let $P_n = \sum_{\nu=0}^{n} p_\nu \to \infty$ as $n \to \infty$ and $p_n/P_n \to 0$ as $n \to \infty$. If

$$\left(\sum_{k=0}^{n} s_k p_{n-k}\right)\Big/ P_n = \left(\sum_{k=0}^{n} u_k P_{n-k}\right)\Big/ P_n$$

converges to s as $n \to \infty$, then $\sum u_n$ is said to be **summable by Nörlund's method** of type $\{p_n\}$, and we write $\sum u_n = s$ (N, $\{p_n\}$). The transformation thus obtained is also regular and is called **Nörlund's method of summation**. If $\sum u_n = s$ (C, 1) and $0 \leqslant p_0 \leqslant p_1 \leqslant \ldots$, then $\sum u_n = s$ (N, $\{p_n\}$). Cesàro's method is actually a special case of this method.

R. M. Riesz's Method of Summation

Let $\{\lambda_n\}$ be a sequence with increasing terms and tending to $+\infty$ as $n \to \infty$. If

$$R(\lambda_n, k, \tau) = \left(\sum_{\lambda_n < \tau} (\tau - \lambda_n)^k u_n\right)\Big/ \tau^k$$

converges to s as $\tau \to \infty$, then $\sum u_n$ is said to be **summable by Riesz's method of order** k and type λ_n, and we write $\sum u_n = s$ (R, λ_n, k). The transformation thus obtained is regular and is called **Riesz's method of summation of the** k**th order**. In particular, if $\lambda_n = n$, then $D(R, \lambda_n, k) = D(C, k)$.

S. Riemann's Method of Summation

If

$$\sum_{n=1}^{\infty} u_n \left(\frac{\sin nh}{nh}\right)^k, \quad u_0 = 0,$$

converges for $h > 0$ and tends to s as $h \to 0$, then $\sum u_n$ is said to be (R, k)**-summable** to s. When $k = 1$, this method, often called **Lebesgue's method of summation**, is not regular. When $k = 2$, it is ordinarily called **Riemann's method of summation** and is regular. Corresponding to these cases, if

$$\frac{2}{\pi}\sum s_n \frac{\sin nh}{n} \to s \quad \text{or} \quad \frac{2}{\pi h}\sum s_n \left(\frac{\sin nh}{n}\right)^2 \to s$$

as $h \to 0$, then $\sum u_n$ is called (R_1)-summable or (R_2)-summable to s, respectively. The summation method (R_1) is not regular, while (R_2) is regular. If $\sum u_n$ is (R_1)-summable, then it is also (R_2)-summable, but (R, 2) and (R_2) are noncomparable.

Other methods of summation were developed by G. H. Hardy and J. E. Littlewood, E. Le Roy, C. J. de La Vallée Poussin, and others (e.g., → [6]).

For related topics → 121 Dirichlet Series, 159 Fourier Series, 339 Power Series.

References

[1] K. Knopp, Theorie und Anwendung der unendlichen Reihen, Springer, 1921; fifth edition, 1964; English translation from the second edition, Theory and application of infinite series, Blackie, 1928.
[2] T. J. I'A. Bromwich, An introduction to the theory of infinite series, Macmillan, second edition, 1926.
For numerical computation of series,
[3] L. B. W. Jolley, Summation of series, Chapman & Hall, 1925.
[4] N. E. Nörlund, Vorlesungen über Differenzenrechnung, Springer, 1924.
For the Euler-Maclaurin formula,
[5] N. Bourbaki, Eléments de mathématique, Fonctions d'une variable réele, ch. 6, Actualités Sci. Ind., 1132a, Hermann, second edition, 1961.
For the summation of divergent series and Tauberian theorems,
[6] G. H. Hardy, Divergent series, Clarendon Press, 1949.
[7] R. G. Cooke, Infinite matrices and sequence spaces, Macmillan, 1950.
[8] K. Zeller, Theorie der Limitierungsverfahren, Erg. Math., Springer, 1958; second edition, 1970.
[9] K. Chandrasekharan and S. Minakshisundaram, Typical means, Oxford Univ. Press, 1952.
[10] H. R. Pitt, Tauberian theorems, Oxford Univ. Press, 1958.

380 (X.15) Set Functions

A. General Remarks

A †function whose domain is a †family of sets is called a **set function**. Usually we consider set functions that take real values or $\pm\infty$. For example, if $f(x)$ is a real-valued function

defined on a set X, and if we assign to each subset A of X values such as $\sup_A f$, $\inf_A f$, or $\sup_A f - \inf_A f$, then we obtain a corresponding set function. In particular, a set function whose domain is the family of left open intervals in $\mathbf{R}^m$ is called an **interval function**. To distinguish between set functions and ordinary functions defined at each point of a set, we call the latter **point functions**. For example, if $f(x)$ is an †integrable (point) function with $\mathbf{R}$ $(=\mathbf{R}^1)$ as its domain and we put $F(I)=\int_a^b f(x)\,dx$ for $I=(a,b]$, then we obtain an interval function F on $\mathbf{R}$.

B. Finitely Additive Set Functions

Let $\Phi(E)$ be a real-valued set function defined on a †finitely additive class $\mathfrak{B}$ in a space X. If Φ satisfies the finite additivity condition:

$$E_1, E_2 \in \mathfrak{B}, \quad E_1 \cap E_2 = \varnothing \quad \text{imply}$$

$$\Phi(E_1 \cup E_2) = \Phi(E_1) + \Phi(E_2),$$

then $\Phi(E)$ is called a **finitely additive set function** on $\mathfrak{B}$. For each $E \in \mathfrak{B}$ we denote $\sup\{\Phi(A) \mid A \subset E, A \in \mathfrak{B}\}$ $(\inf\{\Phi(A) \mid A \subset E, A \in \mathfrak{B}\})$ by $\overline{V}(\Phi; E)$ $(\underline{V}(\Phi; E))$, the **upper (lower) variation** of Φ. Since $\Phi(\varnothing)=0$, we have $\underline{V}(\Phi; E) \leqslant 0 \leqslant \overline{V}(\Phi; E)$. $V(\Phi; E) = \overline{V}(\Phi; E) + |\underline{V}(\Phi; E)|$ is called the **total variation** of Φ on E. When we deal with a fixed Φ, instead of $\overline{V}(\Phi; E)$, $\underline{V}(\Phi; E)$, $V(\Phi; E)$ we write simply $\overline{V}(E)$, $\underline{V}(E)$, $V(E)$. If $V(\Phi; E)$ is bounded, then Φ is said to be **of bounded variation**. If $\Phi(E) \geqslant 0$ $(\leqslant 0)$ for every $E \in \mathfrak{B}$, i.e., $E \subset E'$ implies $\Phi(E) \leqslant \Phi(E')$ $(\Phi(E) \geqslant \Phi(E'))$, then Φ is said to be **monotone increasing (decreasing)**. Every finitely additive set function of bounded variation can be represented as the difference of two monotone increasing finitely additive set functions.

Let I_0 be a fixed interval in $\mathbf{R}^m$ and $F(I)$ be an interval function defined for left open intervals $I \subset I_0$, where $\varnothing$ is considered as a degenerate left open interval. If, for any two left open intervals I_1, I_2 such that $I_1 \cup I_2$ is an interval and $I_1 \cap I_2 = \varnothing$, we have $F(I_1 \cup I_2) = F(I_1) + F(I_2)$, then we call $F(I)$ an **additive interval function** in I_0. Specifically, if $f(x)$ is a real-valued bounded function on $\mathbf{R}$ and D is an interval function determined by $D(I) = f(b) - f(a)$, where $I = [a, b)$ (i.e., $D(I)$ is the increment of $f(x)$), then D is an additive interval function on $\mathbf{R}$, called the **increment function** of f. For a given f the increment function is determined uniquely. Conversely, for a given D, a function f such that D is its increment function is determined uniquely up to an additive constant. In this sense an additive interval function in $\mathbf{R}$ may be identified with the corresponding point function on $\mathbf{R}$.

Let $\mathfrak{R}(I_0)$ be the finitely additive class of all finite unions R of left open intervals in I_0. Then any additive interval function $F(I)$ can be extended to a finitely additive set function $F(R)$ defined on $\mathfrak{R}(I_0)$. For the rest of this article, it is understood that an additive interval function means this extended set function. If for any $\varepsilon > 0$ there exists a $\delta > 0$ such that $|I| < \delta$ implies $F(I) < \varepsilon$ (where $|I|$ is the volume of the interval I), then we say that F is **continuous**.

C. Completely Additive Set Functions

Let $\Phi(E)$ be a real-valued set function defined on a †completely additive class $\mathfrak{B}$ in a space X. If Φ satisfies the complete additivity condition:

$$E_j, E_k \in \mathfrak{B}, \quad E_j \cap E_k = \varnothing \quad (j \neq k)$$

$$\text{imply} \quad \Phi\left(\sum_{j=1}^{\infty} E_j\right) = \sum_{j=1}^{\infty} \Phi(E_j),$$

then $\Phi(E)$ is called a **completely additive set function** (or simply **additive set function**) on $\mathfrak{B}$. In this case the corresponding upper variation $\overline{V}(E)$, lower variation $\underline{V}(E)$, and total variation $V(E)$ are all completely additive set functions, and for every $E \in \mathfrak{B}$ we have $\Phi(E) = \overline{V}(E) + \underline{V}(E)$ (**Jordan decomposition**). Furthermore, $V(E) = \sup \sum_{j=1}^{n} |\Phi(E_j)|$, where the supremum is taken over all decompositions of E such that $E = \bigcup_{j=1}^{n} E_j$ $(E_j \in \mathfrak{B}, E_j \cap E_k = \varnothing, j \neq k)$. The completely additive nonnegative set functions are the same as the finite measures. Hence the Jordan decomposition implies that every completely additive set function is represented as the difference of two finite measures. A completely additive set function is also called a **signed measure**.

Any continuous additive interval function of bounded variation can be extended to a completely additive set function. The notion of additive interval function of bounded variation is a generalization of that of function of bounded variation (→ 166 Functions of Bounded Variation).

Let Φ be a completely additive set function and μ a finite or σ-finite measure, both defined on $\mathfrak{B}$. If $\mu(E)=0$ implies $\Phi(E)=0$, then Φ is said to be **absolutely continuous** with respect to μ or **μ-absolutely continuous**. Then Φ is μ-absolutely continuous if and only if for any $\varepsilon > 0$, there exists a $\delta > 0$ such that $\mu(E) < \delta$ implies $|\Phi(E)| < \varepsilon$. If for given Φ and μ there exists an $E_0 \in \mathfrak{B}$ such that $\mu(E_0) = 0$ and $\Phi(E) = \Phi(E \cap E_0)$ for every $E \in \mathfrak{B}$, then Φ is said to be **singular** with respect to μ or **μ-singular**.

In a †σ-finite measure space $(X, \mathfrak{B}, \mu)$, every completely additive set function $\Phi(E)$ defined on $\mathfrak{B}$ can be represented uniquely as the sum of a μ-absolutely continuous set function and a

μ-singular set function (**Lebesgue decomposition theorem**). Also, $\Phi(E)$ is μ-absolutely continuous if and only if $\Phi(E)$ can be represented as the indefinite integral $\int_E f(x)\,d\mu$ of a function f that is integrable on X with respect to μ (**Radon-Nikodym theorem**). This function $f(x)$ is called the **Radon-Nikodym derivative**, $d\Phi/d\mu$, of Φ with respect to μ (→ 270 Measure Theory L (iii)).

D. Differentiation of Set Functions

Let m be the Lebesgue measure in $\mathbf{R}^m$ and E a Lebesgue measurable set. We denote $\sup(m(E)/m(Q))$ for all cubes Q such that $E\subset Q$ by $r(E)$ and call it the **parameter of regularity** of E. If for a sequence of sets $\{E_n\}$ there exists an α such that $r(E_n)>\alpha>0$, then $\{E_n\}$ is called a **regular sequence**. If all the E_n contain a point P and the †diameter of E_n tends to 0 as $n\to\infty$, then we say that $\{E_n\}$ converges to the point P. Let Φ be a set function in $\mathbf{R}^m$. For a regular sequence $\{E_n\}$ of closed sets converging to a point P, we put $l=\limsup(\Phi(E_n)/m(E_n))$ and define the **general upper derivative** of Φ at P to be the least upper bound of l for all such sequences $\{E_n\}$, denoted by $\bar{D}\Phi(P)$. Similarly, the **general lower derivative** $\underline{D}\Phi(P)$ of Φ at P is defined to be the greatest lower bound of $\liminf(\Phi(E_n)/m(E_n))$ for all regular sequences $\{E_n\}$ of closed sets converging to P. The **ordinary upper (lower) derivative**, denoted by $\bar{\Phi}(E)(\underline{\Phi}(E))$, is defined in the same way by taking regular sequences of closed intervals instead of closed sets. $\bar{D}\Phi$, $\underline{D}\Phi$, $\bar{\Phi}$, $\underline{\Phi}$ are point functions derived from Φ. Clearly, $\underline{D}\Phi(P)\leqslant\underline{\Phi}(P)\leqslant\bar{\Phi}(P)\leqslant\bar{D}\Phi(P)$. If $\bar{D}\Phi(P)=\underline{D}\Phi(P)$, then we write it simply as $D\Phi(P)$. If $D\Phi(P)$ is finite, then we call it the **general derivative** of Φ at P and say that Φ is **derivable in the general sense** at P. If $\bar{\Phi}(P)=\underline{\Phi}(P)$, then we write it as $\Phi(P)$. If $\Phi(P)$ is finite, then we call it the **ordinary derivative** of Φ at P and say that Φ is **derivable in the ordinary sense**. We have the following theorems: (1) A completely additive set function is derivable in the general sense †almost everywhere (Lebesgue). The Radon-Nikodym derivative of a set function absolutely continuous with respect to the Lebesgue measure is equal almost everywhere to the generalized derivative of the set function. (2) An additive interval function of bounded variation is derivable in the ordinary sense almost everywhere (Lebesgue). (3) An additive interval function Φ is derivable in the ordinary sense at almost all points such that $\bar{\Phi}(P)<+\infty$ or $\underline{\Phi}(P)>-\infty$.

For the proof of these theorems, **Vitali's covering theorem** is essential: Let A be a given set and $\mathfrak{F}$ a family of measurable sets in Euclidean space. If for each $x\in A$ there is a regular sequence of sets belonging to $\mathfrak{F}$ that converges to x, then there exists a finite or countable set of disjoint $E_n\in\mathfrak{F}$ such that $m(A\setminus\bigcup_{j=1}^{\infty}E_j)=0$.

References

[1] H. Lebesgue, Leçons sur l'intégration et la recherche des fonctions primitives, Gauthier-Villars, 1904, second edition, 1928.
[2] S. Saks, Theory of the integral, Stechert, 1937 (Dover, 1964).
[3] W. Rudin, Real and complex analysis, McGraw-Hill, second edition, 1974.
[4] H. L. Royden, Real analysis, Macmillan, second edition, 1963.

381 (II.1) Sets

A. Definitions and Symbols

G. Cantor defined a **set** as a collection of objects of our intuition or thought, within a certain realm, taken as a whole. Each object in the collection is called an **element** (or **member**) of the set. The notation $a\in A$ $(A\ni a)$ means that a is an element of the set A. In this case we say that a is a member of A or a **belongs** to A. The negation of $a\in A$ $(A\ni a)$ is written $a\notin A$ or $a\bar{\in}A$ $(A\not\ni a$ or $A\bar{\ni}a)$. The set having no element, namely the set A such that $a\notin A$ for every object a, is called the **empty set** (or **null set**) and is usually denoted by $\varnothing$. Two sets A and B are identical, i.e., $A=B$, if every element of A belongs to B, and vice versa. The set containing $a, b, c, \ldots$ as its elements is said to consist of $a, b, c, \ldots$ and is denoted by $\{a,b,c,\ldots\}$. The symbol $\{x\mid C(x)\}$ (or $\{x; C(x)\}$, sometimes $E_x[C(x)]$) denotes the set of all objects that have the property $C(x)$. Thus $\{a\}$ is the set whose only element is a, and $\{a,b\}$ is the set with two elements a and b, provided that $a\neq b$. A set is called a **finite set** or an **infinite set** according as the number of its elements is finite or infinite.

A set A is a **subset** of a set B if each element of A is an element of B. In this case we also say that A is contained in B or that B **contains** A, and we write $A\subset B$ and $B\supset A$. The negation of $A\subset B$ $(B\supset A)$ is $A\not\subset B$ $(B\not\supset A)$. For every set A, $\varnothing\subset A$. $A\subset B$ and $B\subset C$ imply $A\subset C$. If $A\subset B$ and $B\subset A$, then $A=B$. A is a **proper subset** of B (in symbols: $A\subsetneqq B$, $B\supsetneqq A$) if $A\subset B$ and $A\neq B$. Some authors use $\subseteqq$ $(\supseteqq)$ for $\subset$ $(\supset)$, and $\subset$ $(\supset)$ for $\subsetneqq$ $(\supsetneqq)$.

B. Algebra of Sets

The **union** (**join** or **sum**) of sets A and B, written $A \cup B$, is the set of all elements which belong either to A or to B or to both. The **intersection** (**meet** or **product**) of sets A and B, written $A \cap B$, is the set of all elements which belong to both A and B. In other words, $x \in A \cup B$ if and only if $x \in A$ or $x \in B$ or both, and $x \in A \cap B$ if and only if $x \in A$ and $x \in B$. Given sets A, B, and C, $A \cup B = B \cup A$, $A \cap B = B \cap A$ (**commutative law**); $(A \cup B) \cup C = A \cup (B \cup C)$, $(A \cap B) \cap C = A \cap (B \cap C)$ (**associative law**); $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$, $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ (**distributive law**); $A \cup (A \cap B) = A$, $A \cap (A \cup B) = A$ (**absorption law**).

Two sets A and B are **disjoint** if $A \cap B = \varnothing$. In this case the set $C = A \cup B$ is said to be the **disjoint union** (or **sum**) of A and B, and is written sometimes as $C = A + B$. The set of elements of A which are not members of B is denoted by $A - B$, and is called the **difference** of A and B (or **relative complement** of B in A). If $A \supset B$, $A - B$ is called the **complement** (or **complementary set**) of B with respect to A.

We often consider a theory in which we restrict our attention to elements and subsets of a certain fixed set Ω, and call it the **universal set** of the theory. In geometic terms, Ω is also called the **space** or the **abstract space**, elements of Ω are called **points**, and subsets of Ω **point sets**. If A is a subset of Ω, $\Omega - A$ is simply called the **complement** of A and is denoted A^c. For $A \subset \Omega$ and $B \subset \Omega$, $A \supset B$ and $A^c \subset B^c$ are equivalent. Furthermore, we have $A \cup A^c = \Omega$, $A \cap A^c = \varnothing$, $A^{cc} = A$; and $(A \cap B)^c = A^c \cup B^c$, $(A \cup B)^c = A^c \cap B^c$ (**de Morgan's law**).

The **power set** of a set X, written $\mathfrak{P}(X)$, is the set of all subsets of X. A set whose elements are sets is often called a **family of sets**.

The **pair** consisting of objects a and b is denoted by (a, b). Two pairs (a, b), (c, d) are defined to be equal if and only if $a = c$ and $b = d$. A pair (a, b) is called an **ordered pair**, while the set $\{a, b\}$ is sometimes called an **unordered pair**. Generally the ***n*-tuple** $(a, b, c, \ldots, d)$ of n given objects a, b, c, $\ldots$, d is defined to be $((\ldots((a, b), c), \ldots), d)$, so that $(a, b, c, \ldots, d) = (a', b', c', \ldots, d')$ if and only if $a = a'$, $b = b'$, $\ldots$, $d = d'$. The **Cartesian product** (or **direct product**) of sets A and B, written $A \times B$, is the set of all pairs (a, b) such that $a \in A$ and $b \in B$. $A \times B = \varnothing$ if and only if either $A = \varnothing$ or $B = \varnothing$; $A \times B \subset C \times D$ if and only if $A \subset C$ and $B \subset D$, provided that neither A nor B is empty. Furthermore,

$$(A \times B) \cup (A' \times B) = (A \cup A') \times B,$$

$$(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D).$$

The subset $\{(a, a) \mid a \in A\}$ of $A \times A$ is called the **diagonal** of $A \times A$ and is denoted by Δ_A.

Generally the **Cartesian product** (or **direct product**) of sets A, B, $\ldots$, D, written $A \times B \times \ldots \times D$, is defined as the set $\{(a, b, \ldots, d) \mid a \in A, b \in B, \ldots, d \in D\}$.

C. Mappings

If there exists a rule which assigns to each element of a set A an element of a set B, this rule is said to define a **mapping** (or simply **map**), **function**, or **transformation** from A into B. The term *transformation* is sometimes restricted to the case where $A = B$. Usually letters f, g, φ, ψ, $\ldots$ stand for mappings. The expression $f: A \to B$ ($A \xrightarrow{f} B$) means that f is a function which maps A into B. If $f: A \to B$ and $a \in A$, then $f(a)$ denotes the element of B which is assigned to a by f. We call $f(a)$ the **image** of a under f. The notation $f: a \mapsto b$ (or $f: a \to b$) is often used to mean $f(a) = b$ (but not in the present volumes). The **domain** of a mapping $f: A \to B$ is the set A, and its **range** (or **codomain**), written $f(A)$, is the subset $\{f(a) \mid a \in A\}$ of B. Two functions f and g are equal ($f = g$) if their domains coincide and $f(a) = g(a)$ for each a in the common domain.

For a mapping $f: A \to B$ and a set $C \in \mathfrak{P}(A)$, $f(C)$ is defined to be the set $\{f(x) \mid x \in C\}$. This definition induces the mapping from $\mathfrak{P}(A)$ to $\mathfrak{P}(B)$ which is usually also denoted by f. If $A_i \in \mathfrak{P}(A)$ $(i = 1, 2)$, then $f(A_1 \cup A_2) = f(A_1) \cup f(A_2)$ and $f(A_1 \cap A_2) \subset f(A_1) \cap f(A_2)$. The **inverse image** of $D \in \mathfrak{P}(B)$, denoted by $f^{-1}(D)$, is defined to be the set $\{x \mid x \in A, f(x) \in D\}$; thus the mapping $f^{-1}: \mathfrak{P}(B) \to \mathfrak{P}(A)$ is defined. If $B_i \in \mathfrak{P}(B)$ $(i = 1, 2)$, then $f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2)$; $f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2)$; $f^{-1}(B - B_1) = A - f^{-1}(B_1)$. Furthermore, $A_1 \subset f^{-1} \circ f(A_1)$ and $f \circ f^{-1}(B_1) \subset B_1$.

A mapping g is an **extension** of a mapping f to a set A' if A' is the domain of g and contains the domain A of f, and if $g(a) = f(a)$ for each a in A. In this case f is called a **contraction** (or **restriction**) of g to A or simply a **partial mapping** of g, and is denoted by $g \mid A$. A mapping f is the **constant mapping** (or **constant function**) with the value b_0 if $f(a) = b_0$ for every a in the domain of f. The **identity mapping** (or **identity function**) on A, often denoted by 1_A, is the mapping with the domain A such that $f(a) = a$ for every a in A. Given two mappings $f: A \to B$ and $g: B \to C$, the mapping from A to C which assigns $g(f(a))$ to each $a \in A$ is called the **composite** of f and g and is denoted by $g \circ f$. If $f: A \to B$, $g: B \to C$, and $h: C \to D$, then $(h \circ g) \circ f = h \circ (g \circ f)$ (associative law for composition of mappings).

A mapping $f: A \to B$ is from A **onto** B if $f(A) = B$. In this case f is also called a **surjection** (or a **surjective mapping**). A mapping $f: A \to B$

is **one-to-one** (**1-1**, or **injective**) if $a \neq a'$ implies $f(a) \neq f(a')$ for every pair of elements a and a' in A, that is, if for each b in the range of A, there exists only one element a of A such that $f(a)=b$. Such an f is also called an **injection.** In particular, given a subset B of a set A, the injection $f: B \rightarrow A$ defined by the condition $f(b)=b$ for each $b \in B$ is called the **inclusion mapping** (**inclusion** or **canonical injection**). A necessary and sufficient condition for $f: A \rightarrow B$ to be a surjection is that $g \circ f = h \circ f$ imply $g=h$ for every pair of mappings $g: B \rightarrow C$ and $h: B \rightarrow C$. For $f: A \rightarrow B$ to be an injection it is necessary and sufficient that $f \circ g = f \circ h$ imply $g=h$ for every pair of mappings $g: C \rightarrow A$ and $h: C \rightarrow A$. A mapping which is both a surjection and an injection is called a **bijection** (or **bijective mapping**). If $f: A \rightarrow B$ is a bijection, then the mapping from B to A which assigns to each element b of B the unique element a of A such that $f(a)=b$ is called the **inverse mapping** (**inverse function** or simply **inverse**) of f, and is denoted by f^{-1}. We have $f \circ f^{-1} = 1_B$ and $f^{-1} \circ f = 1_A$ for every bijection $f: A \rightarrow B$.

If the domain A of a mapping $f: A \rightarrow B$ is the Cartesian product of A_1 and A_2, $f(a)=b$ (where $a=(a_1, a_2)$) is written as $f(a_1, a_2)=b$. Given $A = A_1 \times A_2$, $B = B_1 \times B_2$, and $f_i: A_i \rightarrow B_i$ $(i=1,2)$, the mapping $f: A \rightarrow B$ defined by the condition $f(a_1, a_2) = (f_1(a_1), f_2(a_2))$ is called the **Cartesian product** (or **direct product**) of the mappings of f_1 and f_2, and is denoted by $f_1 \times f_2$.

For a mapping $f: A \rightarrow B$, the subset $G = \{(a, f(a)) \mid a \in A\}$ of $A \times B$ is called the **graph** of f. The basic properties of the graph G of f are: (1) For every $a \in A$ there exists a $b \in B$ such that $(a, b) \in G$. (2) $(a, b) \in G$ and $(a, b') \in G$ imply $b=b'$. Conversely, a subset G of $A \times B$ with these two properties determines a mapping $f: A \rightarrow B$ such that $(a, b) \in G$ if and only if $f(a)=b$. All notions concerning a mapping $f: A \rightarrow B$ can be transferred by means of its graph to those concerning a subset of a Cartesian product $A \times B$.

Given sets A and B, we denote by B^A the set of mappings from A to B. If a mapping is identified with its graph, B^A is considered to be a subset of $\mathfrak{P}(A \times B)$. For $X \in \mathfrak{P}(A)$, the mapping $c_X: A \rightarrow \{0,1\}$ such that $c_X(x)=1$ if $x \in X$ and $c_X(x)=0$ if $x \notin X$ is called the **characteristic function** (or **representing function**) of X. By assigning to each $X \in \mathfrak{P}(A)$ its characteristic function $c_X \in \{0,1\}^A$, we obtain a one-to-one correspondence between $\mathfrak{P}(A)$ and $\{0,1\}^A$; hence $\mathfrak{P}(A)$ is sometimes denoted by 2^A.

D. Families of Sets

A mapping from a set Λ to a set A is also called a family of elements of A indexed by Λ. Λ is its **index set** (or **indexing set**). In this case, the image $f(\lambda)$ of $\lambda \in \Lambda$ is denoted by a_λ, and the mapping itself is denoted by $\{a_\lambda\}_{\lambda \in \Lambda}$ ($\{a_\lambda\}$ $(\lambda \in \Lambda)$, or simply $\{a_\lambda\}$). In particular, if the set A is the power set of a set, the family $\{a_\lambda\}_{\lambda \in \Lambda}$ is called a **family of sets indexed by** Λ, or simply a **family of sets.** (Moreover, if Λ is chosen to be a subset of the power set $\mathfrak{P}(X)$ of a set X and f to be the identity mapping on Λ, then the family of sets resulting from f can be identified with the set of subsets Λ itself.)

The union $\bigcup_{\lambda \in \Lambda} A_\lambda$ of a family of sets $\{A_\lambda\}_{\lambda \in \Lambda}$ is the set of all elements a such that $a \in A_\lambda$ for at least one λ in Λ. Their intersection $\bigcap_{\lambda \in \Lambda} A_\lambda$ is the set of all elements a such that $a \in A$ for all λ in Λ. A family of sets $\{A_\lambda\}_{\lambda \in \Lambda}$ is **mutually disjoint** if $\lambda \neq \mu$ implies $A_\lambda \cap A_\mu = \varnothing$. In this case $A = \bigcup_{\lambda \in \Lambda} A_\lambda$ is called the **disjoint union** (or **direct sum**) of the sets of the family, and $\{A_\lambda\}_{\lambda \in \Lambda}$ is called a **partition** (or **decomposition**) of A. For families of sets, the following hold: $\bigcup A_{\lambda\mu} = \bigcup_\lambda (\bigcup_\mu A_{\lambda\mu})$, $\bigcap A_{\lambda\mu} = \bigcap_\lambda (\bigcap_\mu A_{\lambda\mu})$ $(\lambda \in \Lambda, \mu \in M_\lambda)$ (associative law); $(\bigcup_{\lambda \in \Lambda} A_\lambda) \cap (\bigcup_{\mu \in M} B_\mu) = \bigcup_{(\lambda, \mu) \in \Lambda \times M} (A_\lambda \cap B_\mu)$, $(\bigcap_{\lambda \in \Lambda} A_\lambda) \cup (\bigcap_{\mu \in M} B_\mu) = \bigcap_{(\lambda, \mu) \in \Lambda \times M} (A_\lambda \cup B_\mu)$ (distributive law); $(\bigcup_{\lambda \in \Lambda} A_\lambda)^c = \bigcap_{\lambda \in \Lambda} A_\lambda^c$, $(\bigcap_{\lambda \in \Lambda} A_\lambda)^c = \bigcup_{\lambda \in \Lambda} A_\lambda^c$ (de Morgan's law).

A family $\{A_\lambda\}_{\lambda \in \Lambda}$ of sets is a **covering** of a set A, or **covers** A, if $A \subset \bigcup_{\lambda \in \Lambda} A_\lambda$.

Given $f: X \rightarrow Y$, $\{A_\lambda\}_{\lambda \in \Lambda}$ and $\{B_\lambda\}_{\lambda \in \Lambda}$ (where $A_\lambda \subset X$ and $B_\lambda \subset Y$), then $f(\bigcup_{\lambda \in \Lambda} A_\lambda) = \bigcup_{\lambda \in \Lambda} f(A_\lambda)$, $f(\bigcap_{\lambda \in \Lambda} A_\lambda) \subset \bigcap_{\lambda \in \Lambda} f(A_\lambda)$; and $f^{-1}(\bigcup_{\lambda \in \Lambda} B_\lambda) = \bigcup_{\lambda \in \Lambda} f^{-1}(B_\lambda)$, $f^{-1}(\bigcap_{\lambda \in \Lambda} B_\lambda) = \bigcap_{\lambda \in \Lambda} f^{-1}(B_\lambda)$.

E. Direct Sum and Direct Product of Families of Sets

Given a family $\{A_\lambda\}_{\lambda \in \Lambda}$ of sets indexed by Λ, a set S, and a family of injections $\{i_\lambda: A_\lambda \rightarrow S\}_{\lambda \in \Lambda}$, then the pair $(S, \{i_\lambda\}_{\lambda \in \Lambda})$ is called the **direct sum** of $\{A_\lambda\}_{\lambda \in \Lambda}$ if $\{i_\lambda(A_\lambda)\}_{\lambda \in \Lambda}$ is a partition of S. In this case, S is written $\sum_{\lambda \in \Lambda} A_\lambda$ (or $\sum_\lambda A_\lambda$ or $\coprod_\lambda A_\lambda$). Each A_λ is called a **direct summand** of S, and each i_λ is called a **canonical injection.**

The **Cartesian product** (or **direct product**) $\prod_{\lambda \in \Lambda} A_\lambda$ $(\prod_\lambda A_\lambda)$ of $\{A_\lambda\}_{\lambda \in \Lambda}$ (where $A_\lambda \subset X$) is the set of all mappings from Λ to X such that $f(\lambda) \in A_\lambda$ for every $\lambda \in \Lambda$. The sets A_λ are the **direct factors** of $\prod_{\lambda \in \Lambda} A_\lambda$. Each element f of $\prod_{\lambda \in \Lambda} A_\lambda$ is denoted by $\{x_\lambda\}_{\lambda \in \Lambda}$ or $(\ldots, x_\lambda, \ldots)$ (where $x_\lambda = f(\lambda)$). The element x_λ is the λth **component** (or **coordinate**) of f. The mapping $\mathrm{pr}_\lambda: \prod_{\lambda \in \Lambda} A_\lambda \rightarrow A_\lambda$ which assigns x_λ to each $\{x_\lambda\}_{\lambda \in \Lambda} \in \prod_{\lambda \in \Lambda} A_\lambda$ is called the **projection** of $\prod_{\lambda \in \Lambda} A_\lambda$ onto its λth component. If $\Lambda = \{1,2\}$, $\prod_{\lambda \in \Lambda} A_\lambda$ can be identified with $A_1 \times A_2$.

F. Set Theory

It was G. †Cantor who introduced the concept of the set as an object of mathematical study.

Cantor stated: "A set is a collection of definite, well-distinguished objects of our intuition or thought. These objects are called the elements of the set" (G. Cantor, *Math. Ann.*, 46 (1895)). Cantor introduced the notions of †cardinal number and †ordinal number and developed what is now known as **set theory**. He proved that the cardinal number of the set of transcendental numbers is greater than that of algebraic numbers, and that all Euclidean spaces have the same cardinal number regardless of their dimension. He stated the †continuum hypothesis and also conjectured the †well-ordering theorem (G. Cantor, *Math. Ann.*, 21 (1883)), which was proved by E. Zermelo [2]. In this proof Zermelo stated the †axiom of choice explicitly for the first time, and used it in an essential way.

Meanwhile it was pointed out that Cantor's naive set concept leads to various logical †paradoxes (→ 319 Paradoxes). Since the set concept plays a fundamental role in every branch of mathematics, the discovery of the paradoxes had a serious impact upon mathematics, and led to a systematic investigation of the †foundations of mathematics. In the course of attempts to avoid paradoxes, set theory was reconstructed as †axiomatic set theory (→ 33 Axiomatic Set Theory), in which Cantor's theory of cardinal numbers and ordinal numbers was restored. Also, the theory of the algebra of sets, which forms a basis for various branches of mathematics, was reconstructed. Axiomatic set theory is considered to be free from paradoxes.

G. Classes

A set in the naive sense is a collection $\{x \mid C(x)\}$ of all x which satisfy a certain condition $C(x)$. The only principle for generating sets in naive set theory is the **axiom of comprehension**, which asserts the existence of the set $\{x \mid C(x)\}$ for any condition $C(x)$. However, this principle leads to paradoxes if the notion of an **arbitrary set** is considered to be well defined; for example, the †Russell paradox is caused by the set $\{x \mid x \notin x\}$. This situation necessitates some restrictions on the axiom of comprehension. The simplest way to overcome the paradoxes is to adopt Zermelo's **axiom of subsets**: Given a set M and a condition $C(x)$, there exists a set $\{x \mid x \in M, C(x)\}$. But this axiom cannot produce any sets other than subsets of sets whose existence is preassumed. Hence further generating principles of sets had to be introduced. The following axioms are usually chosen as generating principles.

Axiom of pairing: For any two objects (possibly sets) a and b, there exists a set $\{a, b\}$.

Axiom of power set: Given a set A, its power set $\mathfrak{P}(A)$ exists.

Axiom of union: For any family of sets the union exists.

Axiom of substitution (or **replacement**): For any set A and any mapping f from A, there exists a set of all images $f(x)$ with $x \in A$.

In ordinary theories of mathematics the set of natural numbers, the set of real numbers, etc., are assumed to exist, in addition to sets generated by the axioms in this section. In pure set theory the **axiom of infinity** is needed to secure the existence of infinite sets.

The concept of the "set" $\{x \mid x \notin x\}$ does not automatically lead to Russell's paradox. The trouble arises when this "set" is regarded as a member of a collection represented by x. This leads to a narrower concept of sets. Consider a fixed collection V consisting of sets in the naive sense and closed under the set-theoretic operations mentioned in the axioms. Call a member of V a set in the narrow sense. Then set theory becomes free from the known paradoxes if the qualification for being a set is restricted in this narrow sense. When sets in the narrow sense are called simply sets, sets in the naive sense are called **classes**. The object $\{x \mid x \notin x\}$ (where x ranges over sets in the narrow sense) is a class which is not a set. Those classes which are not sets are called **proper classes**; for example, the class V of all sets and the class of all ordinal numbers are both proper classes. For classes, unrestricted use of the comprehension axiom again leads to paradoxes, but other set-theoretic operations are justifiably applicable to classes.

The notion of classes was first introduced in connection with the construction of an axiomatic set theory. The term *class* was used originally to denote certain subclasses of the class V of all sets. In these volumes the term *set* is mostly used to mean a set in the naive sense, and most of the notions defined for sets are applicable to classes.

References

[1] G. Cantor, Gesammelte Abhandlungen, Springer, 1932.
[2] E. Zermelo, Beweis, dass jede Menge wohlgeordnet werden kann, Math. Ann., 59 (1904), 511–516.
[3] A. Schoenflies, Entwicklung der Mengenlehre und ihrer Anwendungen, Teubner, 1913.
[4] F. Hausdorff, Grundzüge der Mengenlehre, Veit, 1914 (Chelsea, 1949).
[5] A. Fraenkel, Einleitung in die Mengenlehre, Springer, third edition, 1928.
[6] F. Hausdorff, Mengenlehre, Teubner, 1927; English translation, Set theory, Chelsea 1962.

[7] A. Fraenkel, Abstract set theory, North-Holland, 1953.
[8] J. L. Kelley, General topology, Van Nostrand, 1955, appendix.
[9] N. Bourbaki, Eléments de mathématique, I. Théorie des ensembles, ch. 2, Actualités Sci. Ind., 1212c, Hermann, second edition, 1960; English translation, Theory of sets, Addison-Wesley, 1968.
[10] P. R. Halmos, Naive set theory, Van Nostrand, 1960.

382 (IX.23) Shape Theory

A. General Remarks

In 1968 K. Borsuk introduced the notion of shape as a modification of the notion of †homotopy type. His idea was to take into account the global properties of topological spaces and neglect the local ones. It is a classification of spaces that is coarser than the homotopy type but that coincides with it on †ANR-spaces.

Let $\mathscr{P}$ be the category whose objects are all †polyhedra and whose morphisms are homotopy classes of continuous mappings between them. For spaces X and Y, denote the set of all †homotopy classes of continuous mappings of X to Y by $[X, Y]$ and the homotopy class of a mapping f by $[f]$. For a space X, let Π_X be the functor from $\mathscr{P}$ to the †category of sets and functions that assigns to a polyhedron P the set $\Pi_X(P)=[X,P]$. A morphism $\varphi: P\rightarrow Q$ of $\mathscr{P}$ induces the function $\varphi_{\#}:[X,P]\rightarrow[X,Q]$ defined by $\varphi_{\#}([f])=\varphi\cdot[f]$ for $[f]: X\rightarrow P$. A †natural transformation from Π_Y to Π_X is a **shape morphism** from X to Y. A continuous mapping $f: X\rightarrow Y$ defines the shape morphism $f^{\#}$ of X to Y as follows: For $[g]: Y\rightarrow P$ in $\Pi_Y(P)$, the composition $[g\cdot f]$ is an element of $\Pi_X(P)$. The correspondence: $[g]\rightarrow[g\cdot f]$ defines a natural transformation from Π_Y to Π_X and hence determines the shape morphism $f^{\#}: X\rightarrow Y$. The identity mapping 1_X on X defines the identity shape morphism $1_X^{\#}$ on X. Given spaces X and Y, X **shape dominates** Y if there are shape morphisms $\xi: Y\rightarrow X$ and $\eta: X\rightarrow Y$ such that $\xi\eta=1_X^{\#}$, and we write $\mathrm{Sh}(X)\leqslant\mathrm{Sh}(Y)$. If, in addition, $\eta\xi=1_Y^{\#}$, then X and Y are of the **same shape**, and we write $\mathrm{Sh}(X)=\mathrm{Sh}(Y)$. A **shape category** $\mathscr{S}$ is the category whose objects are all topological spaces and whose morphisms are shape morphisms between them. If we replace topological spaces by pointed ones, the **pointed shape** category is obtained. In what follows, for simplicity, we assume that all spaces are metrizable and the mappings are continous. General references are Borsuk [1], J. Dydak and J. Segal [3], R. H. Fox [4], S. Mardešić [7].

B. Chapman's Complement Theorem

Let X be a compactum. A closed set A of X is a **Z-set** in X if for any $\varepsilon>0$ there is a mapping $f: X\rightarrow X-A$ such that $d(x,f(x))<\varepsilon$ for $x\in X$, where d is a metric on X. The **Hilbert cube** Q is the countable product $\prod_{i=1}^{\infty} I_i$, where I_i is the closed interval $[0,1]$. The subset $s=\prod_{i=1}^{\infty} I_i^0$ ($I_i^0=(0,1)$) is called the **pseudointerior** of Q. The following facts are known: (1) If a compact metric space X is contained in s or $Q-s$, then X is a Z-set in Q. (2) For any continuous mapping f of a compact metric space X into Q, there exists an embedding g of X into Q such that g is arbitrarily close to f and the image $g(X)$ in Q is a Z-set. The **complement theorem** (T. A. Chapman [2]) states: Let X and Y be Z-sets in Q. Then $\mathrm{Sh}(X)=\mathrm{Sh}(Y)$ iff $Q-X$ and $Q-Y$ are homeomorphic.

C. FAR, FANR, Movability, and Shape Group: Shape Invariants

A closed set A of a compactum X is a **fundamental retract** of X if there is a shape morphism $r: X\rightarrow A$ such that $r\cdot i^{\#}=1_A^{\#}$, where i is the inclusion of A into X. A compactum X is a **fundamental absolute retract (FAR)** (resp. **fundamental absolute neighborhood retract (FANR)**) if for any compactum Y containing X, X is a fundamental retract of Y (resp. of some closed neighborhood of X in Y). A compactum X is **movable** if for any embedding $X\subset Q$ and for any neighborhood U of X in Q there is a neighborhood V of X satisfying the following condition: For any mapping f of a compactum Y to V and for any neighborhood W of X, there is a homotopy $H: Y\times I\rightarrow U$ such that $H(y,0)=f(y)$ and $H(y,1)\in W$ for $y\in Y$. In this definition, if Y is replaced by a compactum with dimension $\leqslant k$, then X is said to be **k-movable**. Pointed FAR, FANR, movability, and k-movability are defined similarly in the pointed shape category. The following facts are known (Borsuk [1], Dydak and Segal [3], J. Keeslings [5], J. Krasinkiewicz [8]). A compactum X is a FAR if and only if X is a pointed FAR iff $\mathrm{Sh}(X)=\mathrm{Sh}(point)$, i.e., X has the same shape as a one-point space. A pointed compactum (X,x) is a pointed FANR iff $\mathrm{Sh}(X,x)\leqslant\mathrm{Sh}(K,k)$ for some pointed polyhedron (K,k). An FANR is movable. A compact connected Abelian topological group is movable if and only if it is locally connected.

A continuous image of a pointed 1-movable compactum is pointed 1-movable. It is unknown whether (i) an FANR is a pointed FANR and (ii) movability means pointed movability. For a pointed compactum (X, x), let $\{(K_i, k_i) | i = 1, 2, \ldots\}$ be a countable †inverse system consisting of pointed finite polyhedra whose limit is (X, x). The limit group $\varprojlim \pi_n(K_i, k_i)$ is the kth **shape group** of (X, x), where $\pi_n(K, k)$ is the kth homotopy group of (K, k). It is known that the shape groups for movable compacta behave like homotopy groups for ANR. A property P of spaces is a **shape invariant** if whenever X has P and $\mathrm{Sh}(X) = \mathrm{Sh}(Y)$, then Y has P. FAR, FANR, movability, k-movability, and shape groups are shape invariants.

D. CE Mappings

A mapping f of a space X onto a space Y is a **cell-like (CE) mapping** if it is proper and $\mathrm{Sh}(f^{-1}(y)) = \mathrm{Sh}(point)$ for each point y of Y. It is known (R. B. Sher [9], Y. Kodama [6]) that if there is a CE mapping of X to Y with finite dimension, then $\mathrm{Sh}(X) = \mathrm{Sh}(Y)$. Here the finite-dimensionality of Y is essential. A **Q-manifold** is a space, each point of which has a closed neighborhood homeomorphic to Q. The following are known (Chapman [2], J. E. West [10]): (1) If f is a CE mapping of a Q-manifold M to an ANR X, then the mapping $g: M \times Q \to X \times Q$ defined by $g(m, x) = (f(m), x)$ for $(m, x) \in M \times Q$ is approximated by homeomorphisms. As a consequence, if X is a locally compact ANR, then $X \times Q$ is a Q-manifold. (2) Every compact ANR is a CE image of a compact Q-manifold. (3) Every compact ANR has the same homotopy type as that of a compact polyhedron. The following problem raised by R. H. Bing is open: Is a CE image of a finite-dimensional compactum finite-dimensional?

References

[1] K. Borsuk, Theory of shape, Monograf. mat. 59, Polish Scientific Publishers, 1975.
[2] T. A. Chapman, Lectures on Hilbert cube manifolds, CBMS 28 regional conf. ser. in math., Amer. Math. Soc., 1976.
[3] J. Dydak and J. Segal, Shape theory, Lecture notes in math. 688, Springer, 1978.
[4] R. H. Fox, On shape, Fund. Math., 74 (1972), 47–71.
[5] J. Keesling, Shape theory and compact connected Abelian topological groups, Trans. Amer. Math. Soc., 194 (1974), 349–358.
[6] Y. Kodama, Decomposition spaces and shape in the sense of Fox, Fund. Math., 97 (1977), 199–208.
[7] S. Mardešić, Shapes for topological spaces, General Topology and Appl., 3 (1973), 265–282.
[8] J. Krasinkiewicz, Continuous images of continua and 1-movability, Fund. Math., 98 (1978), 141–164.
[9] R. B. Sher. Realizing cell-like maps in Euclidean spaces, General Topology and Appl., 2 (1972), 75–89.
[10] J. E. West, Mapping Hilbert cube manifolds to ANR's: A solution of a conjecture of Borsuk, Ann. Math., (2) 106 (1977), 1–8.

383 (II.26) Sheaves

A. Presheaves

Let X be a †topological space. Suppose that the following conditions are satisfied: (i) There exists an (additive) Abelian group $\mathscr{F}(U)$ for each open set U of X, and $\mathscr{F}(\varnothing) = \{0\}$; and (ii) there exists a homomorphism $r_{UV}: \mathscr{F}(V) \to \mathscr{F}(U)$ for each pair $U \subset V$, such that $r_{UU} = 1$ (identity) and $r_{UW} = r_{UV} \circ r_{VW}$ for $U \subset V \subset W$. We call $\mathscr{F}$, consisting of a family $\{\mathscr{F}(U)\}$ of Abelian groups and a family of mappings $\{r_{UV}\}$, a **presheaf** (of Abelian groups) on X. If $a \in \mathscr{F}(V)$ and $U \subset V$, we write $r_{UV}(a) = a | U$ and call it the **restriction** of a to U. A **homomorphism** φ between two presheaves $\mathscr{F}$ and $\mathscr{G}$ on X is a family $\{\varphi(U)\}$ of group homomorphisms $\varphi(U): \mathscr{F}(U) \to \mathscr{G}(U)$ satisfying $r_{UV} \circ \varphi(V) = \varphi(U) \circ r_{UV}$ whenever $U \subset V$. The presheaves on X and their homomorphisms form a †category.

B. Axioms for Sheaves

A presheaf $\mathscr{F}$ is called a **sheaf** (of Abelian groups) if it satisfies the following condition: If U is open in X and $(U_i)_{i \in I}$ is an †open covering of U, and if for each $i \in I$ an element s_i of $\mathscr{F}(U_i)$ is given such that $s_i | U_i \cap U_j = s_j | U_i \cap U_j$ for all i and j, then there exists a unique $s \in \mathscr{F}(U)$ such that $s | U_i = s_i$ for all i. By definition, a **homomorphism** between two sheaves is a homomorphism of the presheaves. The sheaves on X also form a category.

Let $\mathscr{F}$ be a presheaf, x a point of X, and $\mathfrak{N}_x$ the †directed set of open neighborhoods of x, with the order opposite to that of inclusion. Then $\{\mathscr{F}(U) | U \in \mathfrak{N}_x\}$ is an inductive system of groups. The †inductive limit $\varinjlim_{x \in U} \mathscr{F}(U)$ of groups $\{\mathscr{F}(U)\}$ is denoted by $\mathscr{F}_x$ and called the **stalk** of $\mathscr{F}$ over x. The image of $s \in \mathscr{F}(U)$ in $\mathscr{F}_x$ is called the **germ** of s at x and is written s_x.

A homomorphism $\varphi: \mathscr{F} \to \mathscr{G}$ of presheaves induces a homomorphism $\varphi_x: \mathscr{F}_x \to \mathscr{G}_x$ of stalks.

C. Sheaf Spaces

We introduce a topology on the †direct sum $\mathscr{F}' = \coprod_{x \in X} \mathscr{F}_x$ in the following way: For each open set U of X and each $s \in \mathscr{F}(U)$, consider the set $M_{U,s} = \{s_x \mid x \in U\}$ of the germs defined by s at the points of U, and take the set of all such $M_{U,s}$ as a †base of open sets of the topology. If $p: \mathscr{F}' \to X$ is the mapping that maps the points of $\mathscr{F}_x$ to x, then p is continuous, and each $p^{-1}(x)$ $(= \mathscr{F}_x)$ has the structure of an Abelian group. Moreover, the following conditions are satisfied: (i) p is a †local homeomorphism, and (ii) the group operations on $p^{-1}(x)$ are continuous in the sense that $(a,b) \to a+b$ is a continuous mapping from the †fiber product $\mathscr{F}' \times_X \mathscr{F}'$ (i.e., the subspace $\{(a,b) \mid p(a) = p(b)\}$ of the product space $\mathscr{F}' \times \mathscr{F}'$) to $\mathscr{F}'$ and $a \to -a$ is a continuous mapping from $\mathscr{F}'$ to itself. In general, a topological space $\mathscr{F}'$ with a structure satisfying these conditions is called a **sheaf space** over X.

When $\mathscr{F}'$ is a sheaf space, a continuous mapping s from a subspace A of X to $\mathscr{F}'$ such that $p \circ s = 1_A$ is called a **section** of $\mathscr{F}'$ over A. The **set of sections** over A, denoted by $\Gamma(A, \mathscr{F}')$, is an Abelian group in the obvious way. If we associate $\Gamma(U, \mathscr{F}')$ with each open set U and define r_{UV} by the restriction of sections $(r_{UV}(s) = s \mid U)$, then we get a sheaf $\mathscr{F}''$ on X. If we start from a presheaf $\mathscr{F}$ and get $\mathscr{F}''$ via $\mathscr{F}'$, the correspondence $\mathscr{F} \to \mathscr{F}''$ is a †covariant functor from the category of presheaves to the category of sheaves, and $\mathscr{F}''$ is called the **sheaf associated with the presheaf** $\mathscr{F}$. If $\mathscr{F}$ is a sheaf, we can prove $\mathscr{F}'' \cong \mathscr{F}$. Conversely, if we start from a sheaf space $\mathscr{F}'$ and construct the sheaf $\mathscr{F}''$ and then the sheaf space $\mathscr{F}'''$, then $\mathscr{F}'''$ is canonically isomorphic to $\mathscr{F}'$. Since we can identify a sheaf and the corresponding sheaf space, both are usually denoted by the same letter. In particular, when $\mathscr{F}$ is a sheaf, $\mathscr{F}(U)$ is usually written $\Gamma(U, \mathscr{F})$.

Given a section $s \in \Gamma(X, \mathscr{F})$ of a sheaf, the points $x \in X$ for which $s_x \neq 0$ in $\mathscr{F}_x$ from a closed set (the sheaf space $\mathscr{F}$ is not necessarily Hausdorff even if X is so). This set is called the **support** of s and is denoted by $\operatorname{supp} s$.

In the theory of this and the previous two sections, we can replace Abelian groups by groups, rings, etc. Then $\mathscr{F}(U)$ is a group or ring accordingly, and $\mathscr{F}(\varnothing)$ the group consisting of the identity element or the ring consisting of the zero element, respectively. We thus obtain the theories of **sheaves of groups**, **sheaves of rings**, etc. In general, a presheaf $\mathscr{F}$ on X with values in a category $\mathscr{C}$ is a †contravariant functor from the category of open sets of X to $\mathscr{C}$, and a homomorphism between presheaves $\mathscr{F}$ and $\mathscr{G}$ on X is a †natural transformation between the functors $\mathscr{F}$ and $\mathscr{G}$.

The presheaves (sheaves) of Abelian groups on a space X form an †Abelian category, denoted by $\mathscr{P}^X$ ($\mathscr{C}^X$). For a homomorphism $f: \mathscr{F} \to \mathscr{G}$ of presheaves, the **image**, **coimage**, **kernel**, and **cokernel** of f in $\mathscr{P}^X$ are given by

$$(\operatorname{Im} f)(U) = \operatorname{Im} f(U),$$

$$(\operatorname{Coim} f)(U) = \operatorname{Coim} f(U),$$

$$(\operatorname{Ker} f)(U) = \operatorname{Ker} f(U),$$

$$(\operatorname{Coker} f)(U) = \operatorname{Coker} f(U).$$

When $\mathscr{F}$ and $\mathscr{G}$ are sheaves, the kernel of f in $\mathscr{C}^X$ coincides with the kernel in $\mathscr{P}^X$, while the image and cokernel of f in $\mathscr{C}^X$ are the associated sheaves of the image and the cokernel in $\mathscr{P}^X$, respectively. Thus, $f: \mathscr{F} \to \mathscr{G}$ induces $f_x: \mathscr{F}_x \to \mathscr{G}_x$ at each $x \in X$, $(\operatorname{Ker} f)_x = \operatorname{Ker} f_x$, $(\operatorname{Im} f)_x = \operatorname{Im} f_x$, $(\operatorname{Coker} f)_x = \operatorname{Coker} f_x$, and a sequence of sheaves $0 \to \mathscr{F} \xrightarrow{f} \mathscr{G} \xrightarrow{g} \mathscr{K} \to 0$ is **exact** if and only if $0 \to \mathscr{F}_x \xrightarrow{f_x} \mathscr{G}_x \xrightarrow{g_x} \mathscr{K}_x \to 0$ is exact at each $x \in X$.

D. Examples

(1) Let G be an Abelian group (or some other †algebraic system) with †discrete topology. The Cartesian product $X \times G$ gives rise to a sheaf on X, called a **constant sheaf** (or **trivial sheaf**).

(2) Let X be a topological space and Y be a topological Abelian group (e.g., the real or complex numbers). We obtain a sheaf $\mathscr{F}$ on X by putting $\mathscr{F}(U)$ = the set of all continuous mappings $U \to Y$ and r_{UV} = the natural restriction. The stalk over $x \in X$ is the set of germs at x of continuous functions into Y. This sheaf is called the **sheaf of germs of continuous functions** with values in Y.

(3) When X is an †analytic manifold and Y is a commutative †Lie group, we define the **sheaf of germs of analytic mappings** with values in Y in the same way. If Y is the complex number field $\mathbf{C}$, this sheaf is the **sheaf $\mathscr{O}$ of germs of analytic** (or **holomorphic**) **functions**. A †connected component of the sheaf space $\mathscr{O}$ can be identified with the †analytic function determined by the function element corresponding to a point on that component. The **sheaf of germs of functions of class C^r** on a C^s-manifold $(r \leqslant s)$ is similarly defined.

(4) Given a †vector bundle B over a topological space X, we define a sheaf on X by $\mathscr{F}(U) = \Gamma(U)$ (= the module of sections of B over U) and r_{UV} = the natural restriction. Here the stalk over $x \in X$ consists of the germs at x of sections of B, and is called the **sheaf of germs of sec-**

tions of the vector bundle B. We have similar definition for the **sheaf of germs of differentiable (analytic) sections** when X is a †differentiable (complex) manifold. The case where B is a †tensor bundle (e.g., the †cotangent bundle $\mathfrak{T}^*(X)$) is important. The sheaf $\mathfrak{A}^r(X)$ of germs of C^∞-sections of the r-fold †exterior power of $\mathfrak{T}^*(X)$ is called the **sheaf of germs of differential forms of degree** r $(0 \leqslant r \leqslant \dim X)$.

E. Sheaf Cohomology

The category $\mathscr{C}^X$ of sheaves of Abelian groups on X has sufficiently many †injective objects. A sheaf $\mathscr{F}$ with the property that $r_{UX}: \Gamma(X, \mathscr{F}) \to \Gamma(U, \mathscr{F})$ is surjective for any open set U is said to be **flabby** (or **scattered**). An injective sheaf is flabby.

Fix a nonempty family Φ of closed subsets of X satisfying the following two conditions: (i) $A, B \in \Phi \Rightarrow A \cup B \in \Phi$; (ii) any closed set contained in an element of Φ belongs to Φ. Putting $\Gamma_\Phi(\mathscr{F}) = \{s \mid s \in \Gamma(X, \mathscr{F}), \operatorname{supp} s \in \Phi\}$ for each $\mathscr{F} \in \mathscr{C}^X$, we obtain a †left-exact †covariant functor Γ_Φ from $\mathscr{C}^X$ to the category (Ab) of Abelian groups. Therefore, by the general theory of homological algebra, we can define the †right derived functors $R^q\Gamma_\Phi: \mathscr{C}^X \to (\mathrm{Ab})$ $(q = 0, 1, 2, \ldots)$. We put $R^q\Gamma_\Phi(\mathscr{F}) = H^q_\Phi(X, \mathscr{F})$ and call the $H^q_\Phi(X, \mathscr{F})$ $(q = 0, 1, \ldots)$ the **cohomology groups with coefficient sheaf** $\mathscr{F}$ and family of supports Φ (→ 200 Homological Algebra I). When Φ is the family of all closed subsets of X, we write $H^q(X, \mathscr{F})$ instead of $H^q_\Phi(X, \mathscr{F})$.

Thus the cohomology group $H^q_\Phi(X, \mathscr{F})$ is the qth cohomology of the complex $\Gamma_\Phi(\mathfrak{L}^0) \xrightarrow{d^0} \Gamma_\Phi(\mathfrak{L}^1) \xrightarrow{d^1} \Gamma_\Phi(\mathfrak{L}^2) \xrightarrow{d^2} \ldots$ induced by an †injective resolution $0 \to \mathscr{F} \to \mathfrak{L}^0 \to \mathfrak{L}^1 \to \ldots$ of the sheaf $\mathscr{F}: H^q_\Phi(X, \mathscr{F}) = \operatorname{Ker} d^q / \operatorname{Im} d^{q-1}$ $(q = 0, 1, \ldots; d^{-1} = 0)$.

$H^0_\Phi(X, \mathscr{F}) = \Gamma_\Phi(\mathscr{F})$, and from an exact sequence of sheaves $0 \to \mathscr{F} \to \mathscr{G} \to \mathscr{H} \to 0$ we get an exact sequence $0 \to H^0_\Phi(X, \mathscr{F}) \to H^0_\Phi(X, \mathscr{G}) \to H^0_\Phi(X, \mathscr{H}) \to H^1_\Phi(X, \mathscr{F}) \to H^1_\Phi(X, \mathscr{G}) \to H^1_\Phi(X, \mathscr{H}) \to H^2_\Phi(X, \mathscr{F}) \to \ldots$.

Similarly, the cohomology groups $H^q_\Phi(X, \mathscr{F})$ can also be calculated with an exact sequence $0 \to \mathscr{F} \to \mathfrak{L}^0 \to \mathfrak{L}^1 \to \ldots$, where each $\mathfrak{L}^i$ is assumed to be Γ_Φ-acyclic (i.e., $H^q_\Phi(X, \mathfrak{L}^i) = 0$ for $q > 0$) instead of injective. The flabby sheaves, for instance, are Γ_Φ-acyclic, so we can compute $H^q_\Phi(X, \mathscr{F})$ by a flabby resolution of $\mathscr{F}$ (R. Godement). For example, let X be an n-dimensional †paracompact C^∞-manifold and $\mathscr{F} = \mathbf{R}$. Then $0 \to \mathbf{R} \to \mathfrak{A}^0(X) \xrightarrow{d^0} \mathfrak{A}^1(X) \xrightarrow{d^1} \ldots$ is exact, where $\mathfrak{A}^q(X)$ is the †sheaf of germs of C^∞-differential forms of degree q, d^q is †exterior differentiation (**Poincaré's theorem**), and we have $H^p(X, \mathfrak{A}^q) = 0$ for $p > 0$. Therefore $H^q(X, \mathbf{R})$ is the qth cohomology of the complex $0 \to D^0(X) \xrightarrow{d^0} D^1(X) \xrightarrow{d^1} \ldots$, where $D^i(X) = \Gamma(X, \mathfrak{A}^i(X))$ = the group of C^∞-differential forms of degree i on X. This proves the de Rham theorem, which says that the †de Rham cohomology group is isomorphic to the †(singular) cohomology group of X with real coefficients (→ 105 Differentiable Manifolds R). For a sheaf $\mathscr{F}$ of noncommutative groups, we can define the first cohomology $H^1(X, \mathscr{F})$ [2].

F. The Čech Cohomology Group

Let $\mathfrak{U} = \{U_i\}$ be an open covering of X, and write $U_i \cap U_j = U_{ij}$, etc. Put

$$C^p(\mathscr{F}) = \prod_{i_0, \ldots, i_p} \Gamma(U_{i_0 \ldots i_p}, \mathscr{F}), \qquad p = 0, 1, 2, \ldots.$$

An element of $C^p(\mathscr{F})$ is called a cochain of degree p. Define $d: C^p(\mathscr{F}) \to C^{p+1}(\mathscr{F})$ by $(df)_{i_0 \ldots i_{p+1}} = \sum_{r=0}^{p+1} (-1)^r (f_{i_0 \ldots \hat{i}_r \ldots i_{p+1}} \mid U_{i_0 \ldots i_{p+1}})$, and denote the qth cohomology of the complex $(C^p(\mathscr{F}), d)$ thus obtained by $H^q(\mathfrak{U}, \mathscr{F})$. When an open covering $\mathfrak{B}$ is a refinement of $\mathfrak{U}$, there is a canonical homomorphism $H^q(\mathfrak{U}, \mathscr{F}) \to H^q(\mathfrak{B}, \mathscr{F})$. So we can take the inductive limit of the groups $H^q(\mathfrak{U}, \mathscr{F})$ with respect to the refinement of open coverings. This limit group is denoted by $\check{H}^q(X, \mathscr{F})$ and is called the **Čech cohomology group with coefficient sheaf** $\mathscr{F}$. It coincides with $H^q(X, \mathscr{F})$ for $q \leqslant 1$, and if X is paracompact, for all q.

G. Relation to Continuous Mappings

Let X and Y be topological spaces and $f: X \to Y$ be a continuous mapping. If $\mathscr{F}$ is a sheaf on Y, the fiber product $X \times_Y \mathscr{F}$ (where $\mathscr{F}$ is viewed as a sheaf space over Y) is a sheaf on X. It is denoted by $f^*(\mathscr{F})$ or $f^{-1}(\mathscr{F})$ and is called the **inverse image** of $\mathscr{F}$. The correspondence $\mathscr{F} \to f^*(\mathscr{F})$ is an exact functor from $\mathscr{C}^Y$ to $\mathscr{C}^X$. Next, let $\mathscr{G}$ be a sheaf on X. Associating $\Gamma(f^{-1}(U), \mathscr{G})$ with each open set U of Y, we obtain a sheaf on Y, which we denote by $f_*(\mathscr{G})$ and call the **direct image** of $\mathscr{G}$. The correspondence f_* is a left-exact functor $\mathscr{C}^X \to \mathscr{C}^Y$, and we can consider its right derived functors $R^q f_*$. The sheaf $R^q f_*(\mathscr{G})$ is the sheaf associated with the presheaf that associates $H^q(f^{-1}(U), \mathscr{G})$ with each open set U.

A homomorphism ψ from $\mathscr{F}$ to $f_*(\mathscr{G})$ is also called an f-homomorphism from $\mathscr{F}$ to $\mathscr{G}$. To give such a ψ is equivalent to giving a family of homomorphisms of the stalks $\psi_x: \mathscr{F}_{f(x)} \to \mathscr{G}_x$ $(x \in X)$ satisfying the continuity condition: For any open set U of Y and any section $s \in \Gamma(U, F)$ over U, the mapping φ from $f^{-1}(U)$ to $\mathscr{G}$ defined by $\varphi(x) = \psi_x(s(f(x)))$ is continuous.

The functors f^* and f_* are related by

$\mathrm{Hom}(f^*(\mathscr{F}),\mathscr{G})\cong\mathrm{Hom}(\mathscr{F},f_*(\mathscr{G}))$. The Leray †spectral sequence

$$E_2^{p,q}=H^p(Y,R^qf_*(\mathscr{G}))\Rightarrow H^n(X,\mathscr{G})$$

exists and connects the cohomologies of X and of Y.

H. Ringed Spaces

Let X be a topological space and $\mathcal{O}$ be a sheaf on X of commutative rings with unity element such that $\mathcal{O}_x\neq\{0\}$ for any $x\in X$. Then the pair $(X,\mathcal{O})$ is called a **ringed space**, and $\mathcal{O}$ is called its **structure sheaf**. A morphism $(X,\mathcal{O})\to(X',\mathcal{O}')$ is by definition a pair (f,θ) consisting of a continuous mapping $f:X\to X'$ and an f-homomorphism $\theta:\mathcal{O}\to\mathcal{O}'$. When each $\mathcal{O}_x$ is a †local ring, $(X,\mathcal{O})$ is called a **local ringed space**. A morphism of local ringed spaces is defined to be a pair $(f,\theta):(X,\mathcal{O})\to(X',\mathcal{O}')$ as before, satisfying the additional condition that θ is local (i.e., $\theta_x:\mathcal{O}'_{f(x)}\to\mathcal{O}_x$ maps the maximal ideal into the maximal ideal for each $x\in X$). These concepts are important in algebraic geometry and the theory of functions of several complex variables.

I. Direct Products and Tensor Products

Let $\mathscr{F}_\lambda$ $(\lambda\in\Lambda)$ be sheaves of Abelian groups on a topological space X. The sheaf $\mathscr{F}$ on X defined by $\mathscr{F}(U)=\prod_\lambda\mathscr{F}_\lambda(U)$ and $r_{UV}=\prod_\lambda r_{UV}^{\mathscr{F}_\lambda}$ is denoted by $\mathscr{F}=\prod_\lambda\mathscr{F}_\lambda$ and called the **direct product** of sheaves $\{\mathscr{F}_\lambda\}$. For each $x\in X$ there is a natural mapping $\mathscr{F}_x\to\prod_\lambda(\mathscr{F}_\lambda)_x$, which is in general neither injective nor surjective. When Λ is a finite set, $\prod\mathscr{F}_i$ is also written $\mathscr{F}=\mathscr{F}_1+\ldots+\mathscr{F}_n$ and is called the **direct sum** of the sheaves. The **inductive limit** $\mathscr{F}=\mathrm{ind\,lim}\,\mathscr{F}_\lambda$ of an inductive system of sheaves on X also exists, and $\mathscr{F}_x=\mathrm{ind\,lim}\,\mathscr{F}_{\lambda,x}$.

Let $(X,\mathcal{O})$ be a ringed space. A sheaf of Abelian groups $\mathscr{F}$ on X is called a **sheaf of $\mathcal{O}$-modules** (or simply an **$\mathcal{O}$-module**) if $\mathscr{F}(U)$ is an $\mathcal{O}(U)$-module for each U and $r_{UV}:\mathscr{F}(V)\to\mathscr{F}(U)$ is a module homomorphism compatible with $\mathcal{O}(V)\to\mathcal{O}(U)$ for each $U\subset V$. Then $\mathscr{F}_x$ is an $\mathcal{O}_x$-module for each $x\in X$. For a fixed $(X,\mathcal{O})$, the $\mathcal{O}$-modules form an Abelian category. When $\mathscr{F}$ and $\mathscr{G}$ are $\mathcal{O}$-modules, the **tensor product** $\mathscr{H}=\mathscr{F}\otimes_{\mathcal{O}}\mathscr{G}$ of $\mathscr{F}$ and $\mathscr{G}$ as sheaves over $\mathcal{O}$ is defined as follows: Define a presheaf by $U\to\mathscr{F}(U)\otimes_{\mathcal{O}(U)}\mathscr{G}(U)$ and $r_{UV}=r_{UV}^{\mathscr{F}}\otimes r_{UV}^{\mathscr{G}}$, and let $\mathscr{H}$ be the associated sheaf of this presheaf. Then we have $\mathscr{H}_x=\mathscr{F}_x\otimes_{\mathcal{O}_x}\mathscr{G}_x$.

The notion of coherent sheaves is important in the theory of $\mathcal{O}$-modules (→ 16 Algebraic Varieties E).

J. History

About 1945, J. Leray established the theory of sheaf coefficient cohomology groups (in a form slightly different from that in Sections E and F) and the theory of spectral sequences to study the relation between the local properties of a continuous mapping and the global cohomologies. In the theory of functions of several complex variables, K. Oka conceived the idea of "ideals of indefinite domain." These two ideas were unified by H. Cartan into the present form of sheaf theory. As a link between local properties and global properties, sheaf theory has been applied in many branches of mathematics (→ 16 Algebraic Varieties; 21 Analytic Functions of Several Complex Variables; 23 Analytic Spaces; 72 Complex Manifolds).

References

[1] F. Hirzebruch, Neue topologische Methoden in der algebraischen Geometrie, Erg. Math., Springer, second edition, 1962; English translation, Topological methods in algebraic geometry, Springer, third enlarged edition, 1966.
[2] A. Grothendieck, Sur quelques points d'algèbre homologique, Tôhoku Math. J., (2) 9 (1957), 119–221.
[3] R. Godement, Topologie algébrique et théorie des faisceaux, Actualités Sci. Ind., 1252, Hermann, 1958.
[4] R. G. Swan, The theory of sheaves, Univ. of Chicago Press, 1964.
[5] G. E. Bredon, Sheaf theory, McGraw-Hill, 1967.
[6] J. Leray, L'anneau spectral et l'anneau filtré d'homologie d'un espace localement compact et d'une application continue, J. Math. Pure Appl., 29 (1950), 1–139.
[7] H. Cartan, Sém. de topologie algébrique, Paris, 1948–1949; 1950–1951.
[8] A. Borel, Sém. de topologie algébrique de l'Ecole polytechnique fédérale, 1951. (Cohomologie des espaces localement compacts d'après J. Leray, Lecture notes in math. 2, Springer, 1964.)

384 (VII.20) Siegel Domains

A. Siegel Domains

Let D be a bounded domain in $\mathbf{C}^n$ and $G_h(D)$ the full †holomorphic automorphism group of

D, which is a †Lie transformation group with respect to the †compact-open topology. If $G_h(D)$ acts transitively on D, then D is called a †**homogeneous bounded domain**. The study of homogeneous bounded domains was initiated systematically by E. Cartan in 1936, while the notion of **Siegel domains**, which was introduced by I. I. Pyatetskiĭ-Shapiro, has made remarkable contributions to the study of homogeneous bounded domains.

Let V be a convex domain in an n-dimensional real vector space R. V is called a **regular cone** if for every $x \in V$ and $\lambda > 0$, $\lambda x \in V$ and if V contains no entirely straight lines. Let W be a complex vector space. A mapping F: $W \times W \to R^{\mathbf{C}}$ (the †complexification of R) is a **V-Hermitian form** if the following conditions are satisfied: (Fi) $F(u,v)$ is $\mathbf{C}$-linear in u, (Fii) $\overline{F(u,v)} = F(v,u)$, where the bar denotes the †conjugation with respect to R, (Fiii) $F(u,u) \in \bar{V}$ (the closure of V), and (Fiv) $F(u,u)=0$ implies $u=0$. Given a regular cone $V \subset R$ and a V-Hermitian form F on W, one can define a **Siegel domain** $D(V,F)$ **(of the second kind)** by putting $D(V,F) = \{(x+iy,u) \in R^{\mathbf{C}} \times W \mid y - F(u,u) \in V\}$, which is holomorphically equivalent to a bounded domain in $R^{\mathbf{C}} \times W$. When $W=(0)$, $D(V,F)$ is reduced to $D(V) = \{x+iy \in R^{\mathbf{C}} \mid y \in V\}$, which is called a **Siegel domain of the first kind**. A mapping $L: W \times W \to R^{\mathbf{C}}$ is a nondegenerate semi-Hermitian form if L can be written as $L = L_1 + L_2$, where L_1 and L_2 are $R^{\mathbf{C}}$-valued functions satisfying the conditions: (Li) L_1 satisfies (Fi) and (Fii); (Lii) L_2 is a symmetric $\mathbf{C}$-bilinear form; (Liii) $L(u,v)=0$ for all $u \in W$ implies $v=0$. Let B be a bounded domain in a complex vector space X and L_p $(p \in B)$ an $R^{\mathbf{C}}$-valued nondegenerate semi-Hermitian form on W depending differentiably on $p \in B$. Consider a domain $D(V,L,B)$ in $R^{\mathbf{C}} \times W \times X$ defined by putting $D(V,L,B) = \{(x+iy,u,p) \in R^{\mathbf{C}} \times W \times X \mid y - \operatorname{Re} L_p(u,u) \in V, p \in B\}$. The domain $D(V,L,B)$ is called a **Siegel domain of the third kind** over B if it is holomorphically equivalent to a bounded domain. $D(V,L,B)$ is a fiber space over B.

By the affine automorphism group G_a of a Siegel domain $D(V,F) \subset R^{\mathbf{C}} \times W$ we mean the group consisting of all elements in the complex affine transformation group of $R^{\mathbf{C}} \times W$ leaving $D(V,F)$ stable. The full holomorphic automorphism group G_h of $D(V,F)$ contains G_a as a closed subgroup. If G_h acts transitively on $D(V,F)$, then $D(V,F)$ is said to be homogeneous. A homogeneous Siegel domain is necessarily affinely homogeneous, i.e., G_a acts transitively on $D(V,F)$ [2]. The †Bergman metric of $D(V,F)$ which is a G_h-invariant †Kähler metric, is †complete [3], and so $D(V,F)$ is a †domain of holomorphy.

Examples. $H(n,\mathbf{R})$ denotes the vector space of all real symmetric matrices of degree n, and $H^+(n,\mathbf{R})$ the regular cone consisting of all positive definite matrices in $H(n,\mathbf{R})$.

(i) The Siegel domain of the first kind $D(H^+(n,\mathbf{R})) = \{X+iY \mid X \in H(n,\mathbf{R}), Y \in H^+(n,\mathbf{R})\}$ is called the †Siegel upper half-plane, which is holomorphically equivalent to the classical †symmetric domain of type III.

(ii) Let $u, v \in \mathbf{C}$ and $F(u,v)$ be the 2×2 diagonal matrix $\operatorname{diag}(u\bar{v},0)$, which is an $H^+(2,\mathbf{R})$-Hermitian form on $\mathbf{C}$. The resulting Siegel domain is

$$D(H^+(2,\mathbf{R}),F) = \left\{(z_1,z_2,z_3,u) \in \mathbf{C}^4 \;\middle|\; \begin{pmatrix} \operatorname{Im} z_1 - |u|^2 & \operatorname{Im} z_3 \\ \operatorname{Im} z_3 & \operatorname{Im} z_2 \end{pmatrix} \in H^+(2,R) \right\}.$$

(iii) Let $B = \{t \in \mathbf{C} \mid |t| < 1\}$, and let $u, v \in \mathbf{C}$. Put $L_t(u,v) = (1-|t|^2)^{-1}(u\bar{v} + \bar{t}uv)$. Then $L_t(u,v)$ is a nondegenerate semi-Hermitian form, and we have $D(H^+(2,\mathbf{R}),L,B) = \{(z,u,t) \in \mathbf{C}^3 \mid \operatorname{Im} z - (1-|t|^2)^{-1}\operatorname{Re}(|u|^2 + \bar{t}u^2) > 0, |t| < 1\}$, which is a Siegel domain of the third kind and is holomorphically equivalent to the Siegel upper half-plane of dimension 3.

The domains in (i) and (ii) are both affinely homogeneous; the latter was originally found by Pyateteskiĭ-Shapiro in 1959 [1] and provides the least-dimensional example of nonsymmetric homogeneous bounded domains, which answered affirmatively Cartan's conjecture (1936): Are there non-†symmetric homogeneous bounded domains in $\mathbf{C}^n$ $(n \geqslant 4)$?

B. Infinitesimal Automorphisms of Siegel Domains

For a Siegel domain $D(V,F) \subset R^{\mathbf{C}} \times W$, the Lie algebra $\mathfrak{g}_h$ of G_h can be identified with the Lie algebra of infinitesimal automorphisms, i.e., all complete holomorphic vector fields on $D(V,F)$. Let $G(V)$ be the group consisting of all the linear automorphisms of R leaving V stable. Let us fix a base in R, and let $(z_1, z_2, \ldots, z_n)$ be the complex linear coordinate system in $R^{\mathbf{C}}$ corresponding to it. Choose a complex linear coordinate system $(u_1, u_2, \ldots, u_m)$ in W. We write $F(u,v)$ as $F(u,v) = (F_1(u,v), \ldots, F_n(u,v))$. Consider the following two vector fields in the Lie algebra $\mathfrak{g}_a$ of G_a:

$$E = 2\sum_k z_k \frac{\partial}{\partial z_k} + \sum_\alpha u_\alpha \frac{\partial}{\partial u_\alpha}, \qquad I = \sum_\alpha i u_\alpha \frac{\partial}{\partial u_\alpha},$$

and put $\mathfrak{g}_a^\lambda = \{X \in \mathfrak{g}_a \mid [E,X] = \lambda X\}$, $\lambda \in \mathbf{Z}$. Then $\mathfrak{g}_a$ can be written as a †graded Lie algebra in the following way: $\mathfrak{g}_a = \mathfrak{g}_a^{-2} + \mathfrak{g}_a^{-1} + \mathfrak{g}_a^0$. Here

we have

$$\mathfrak{g}_a^{-2}=\left\{\sum_k a_k\frac{\partial}{\partial z_k}\,\middle|\, a_k\in\mathbf{R}\right\},$$

$$\mathfrak{g}_a^{-1}=\left\{2i\sum_k F_k(u,c)\frac{\partial}{\partial z_k}+\sum_\alpha c_\alpha\frac{\partial}{\partial u_\alpha}\,\middle|\, c=(c_\alpha),\quad c_\alpha\in\mathbf{C}\right\},$$

and $\mathfrak{g}_a^0$ consists of all vector fields $X_{(A,B)}$ of the form

$$\sum_{k,j}a_{kj}z_j\frac{\partial}{\partial z_k}+\sum_{\alpha,\beta}b_{\alpha\beta}u_\beta\frac{\partial}{\partial u_\alpha},$$

where the matrices $A=(a_{kj})$, $B=(b_{\alpha\beta})$ satisfy the conditions: $\exp tA\in G(V)$, $t\in\mathbf{R}$; $AF(u,v)=F(Bu,v)+F(u,Bv)$. For $X_{(A,B)}\in\mathfrak{g}_a^0$ we define $\operatorname{tr}X_{(A,B)}$ to be the sum of the trace of A and that of B. Let $\mathfrak{g}^\lambda$ be the λ-eigenspace of ad E in $\mathfrak{g}_h$ ($\lambda\in\mathbf{Z}$). Then $\mathfrak{g}_h$ can be written also in the form of a graded Lie algebra: $\mathfrak{g}_h=\mathfrak{g}^{-2}+\mathfrak{g}^{-1}+\mathfrak{g}^0+\mathfrak{g}^1+\mathfrak{g}^2$, and $\mathfrak{g}^\lambda=\mathfrak{g}_a^\lambda$ is valid for $\lambda=-2,-1,0$. Furthermore $\mathfrak{g}_h$ can be nicely determined by $\mathfrak{g}_a$ in the following manner. $p_{\mu\lambda}$ denotes a polynomial on $R^{\mathbf{C}}\times W$ homogeneous of degree μ in $z_1,\dots,z_n$, and homogeneous of degree λ in $u_1,\dots,u_m$. Let $\hat{\mathfrak{g}}^1$ (resp. $\hat{\mathfrak{g}}^2$) be the set of all polynomial vector fields of the form

$$\sum_k p^k_{1,1}\frac{\partial}{\partial z_k}+\sum_\alpha(p^\alpha_{1,0}+p^\alpha_{0,2})\frac{\partial}{\partial u_\alpha}$$

$$\left(\text{resp. }\sum_k p^k_{2,0}\frac{\partial}{\partial z_k}+\sum_\alpha p^\alpha_{1,1}\frac{\partial}{\partial u_\alpha}\right).$$

Then we have $\mathfrak{g}^1=\{X\in\hat{\mathfrak{g}}^1\mid[X,\mathfrak{g}^{-1}]\subset\mathfrak{g}^0\}$, and $\mathfrak{g}^2=\{X\in\hat{\mathfrak{g}}^2\mid[X,\mathfrak{g}^{-2}]\subset\mathfrak{g}^0,\ [X,\mathfrak{g}^{-1}]\subset\mathfrak{g}^1,\ \operatorname{Im}\operatorname{Tr}[X,Y]=0\text{ for }Y\in\mathfrak{g}^{-2}\}$. Another description of $\mathfrak{g}^1$ and $\mathfrak{g}^2$ has been given in terms of Jordan triple systems [4]. The explicit descriptions of $\mathfrak{g}^1$ and $\mathfrak{g}^2$ have been given for most homogeneous Siegel domains $D(V,F)$ over irreducible self-dual cones V (→ Section D; T. Tsuji, *Nagoya Math. J.*, 55 (1974)). $\mathfrak{g}_h=\mathfrak{g}_a$ is valid for the Siegel domains which are irreducible quasisymmetric but not symmetric (→ Section D). Main references for this section are [2–6].

C. *j*-Algebras and Homogeneous Bounded Domains

The notion of ***j*-algebra** was introduced by Pyatetskiĭ-Shapiro [1], which reduces the study of homogeneous bounded domains to purely algebraic problems. Let $\mathfrak{g}$ be a Lie algebra over $\mathbf{R}$, and $\mathfrak{k}$ a subalgebra of $\mathfrak{g}$, (j) a collection of linear endomorphisms of $\mathfrak{g}$, and ω be a linear form on $\mathfrak{g}$. Then the quadruple $\{\mathfrak{g},\mathfrak{k},(j),\omega\}$ (sometimes abbreviated $\mathfrak{g}$) is called a j-algebra if the following conditions are satisfied: (i) $j\mathfrak{k}\subset\mathfrak{k}$ for $j\in(j)$ and $j\equiv j'(\operatorname{mod}\mathfrak{k})$ for $j,j'\in(j)$; (ii) $j^2\equiv-id(\operatorname{mod}\mathfrak{k})$; (iii) $j[k,x]\equiv[k,jx]$ $(\operatorname{mod}\mathfrak{k})$ for $k\in\mathfrak{k}$, $x\in\mathfrak{g}$; (iv) $[jx,jy]\equiv j[jx,y]+j[x,jy]+[x,y](\operatorname{mod}\mathfrak{k})$ for $x,y\in\mathfrak{g}$; (v) $\omega([k,x])=0$ for $k\in\mathfrak{k}$; (vi) $\omega([jx,jy])=\omega([x,y])$; (vii) $\omega([jx,x])>0$ for $x\notin\mathfrak{k}$. Let $\mathfrak{g}'$ be a subalgebra of $\mathfrak{g}$ such that $j\mathfrak{g}'\subset\mathfrak{g}'+\mathfrak{k}$. Then, putting $\mathfrak{k}'=\mathfrak{g}'\cap\mathfrak{k}$, one can naturally induce a j-algebra structure on the pair $\{\mathfrak{g}',\mathfrak{k}'\}$. The j-algebra thus obtained is called a j-subalgebra of $\{\mathfrak{g},\mathfrak{k},(j),\omega\}$. A j-algebra $\{\mathfrak{g},\mathfrak{k},(j),\omega\}$ is called proper (resp. effective), if, for any j-subalgebra $\{\mathfrak{g}',\mathfrak{k}'\}$ with $\mathfrak{g}'$ compact semisimple, $\mathfrak{g}'$ is contained in $\mathfrak{k}$ (resp. if $\{\mathfrak{g},\mathfrak{k}\}$ is an effective pair).

Now let D be a homogeneous bounded domain in $\mathbf{C}^n$, G a connected †Lie subgroup of $G_h(D)$ acting †transitively on D, and K the †isotropy subgroup at a point in D. The Lie algebras of G and K are denoted by $\mathfrak{g}$ and $\mathfrak{k}$, respectively. Then the pair $\{\mathfrak{g},\mathfrak{k}\}$ becomes an effective proper j-algebra. Conversely, to every effective proper j-algebra there corresponds a homogeneous bounded domain. The identity component of $G_h(D)$ is isomorphic to the identity component of a †real algebraic group via the †adjoint representation. Let $\{\mathfrak{g},\mathfrak{k},(j),\omega\}$ be a j-algebra. Suppose that $\mathfrak{g}$ satisfies the following conditions: (i) $\mathfrak{g}=\mathfrak{g}^{-2}+\mathfrak{g}^{-1}+\mathfrak{g}^0$ as a graded Lie algebra; (ii) $\mathfrak{g}^0=\mathfrak{k}+j\mathfrak{g}^{-2}$; (iii) there exists a $j\in(j)$ such that $j\mathfrak{g}^{-1}=\mathfrak{g}^{-1}$; and (iv) there exists an $r\in\mathfrak{g}^{-2}$ such that $[jx,r]=x$ for $x\in\mathfrak{g}^{-2}$. Such a decomposition is called a Siegel decomposition of $\mathfrak{g}$. To an effective j-algebra admitting a Siegel decomposition there corresponds a unique Siegel domain up to affine equivalence. Vinberg, Gindikin, and Pyatetskiĭ-Shapiro (Appendix in [1] or *Trans. Moscow Math. Soc.*, 12 (1963)) proved that the Lie algebra $\mathfrak{g}_h(D)$ of $G_h(D)$ contains a j-subalgebra admitting a Siegel docomposition and corresponding to the same domain D, and obtained the realization theorem: Every homogeneous bounded domain D is holomorphically equivalent to a Siegel domain. In consequence, D is diffeomorphic to a Euclidean space, and the isotropy subgroup $K_h(D)$ is a maximal compact subgroup of $G_h(D)$. We have the decomposition $G_h(D)=K_h(D)\cdot T$ (semidirect), where T is an $\mathbf{R}$-splittable solvable subgroup of $G_h(D)$ acting simply transitively on D. T is uniquely determined up to conjugacy in $G_h^0(D)$ (= the identity component of $G_h(D)$), and is called the **Iwasawa group** of D.

The j-algebra structure of the Lie algebra $\mathfrak{t}$ of the Iwasawa group T is characterized by the following properties: (i) for every $t\in\mathfrak{t}$, the eigenvalues of ad t are all real; (ii) there exists a †complex structure j such that $[jx,jy]=$

$j[jx,y]+j[x,jy]+[x,y]$ for $x, y \in \mathfrak{t}$; and (iii) there exists a linear form ω on $\mathfrak{t}$ such that $\omega([jx,jy])=\omega([x,y])$ and that $\omega([jx,x])>0$ for $x \neq 0$. A Lie algebra satisfying (i)–(iii) is called a **normal** j**-algebra**. There exists a one-to-one correspondence between the set of holomorphic equivalence classes of homogeneous bounded domains and the set of j-isomorphism classes of normal j-algebras; by a j-isomorphism here we mean an isomorphism which commutes with j. Let $\{\mathfrak{g},j,\omega\}$ be a normal j-algebra and define an inner product $\langle\ ,\ \rangle$ on $\mathfrak{g}$ by $\langle x,y\rangle=\omega([jx,y])$. The orthogonal complement $\mathfrak{h}$ with respect to $\langle\ ,\ \rangle$ of the †commutator subalgebra $\mathfrak{g}_1$ of $\mathfrak{g}$ is an Abelian subalgebra of $\mathfrak{g}$, and the adjoint representation of $\mathfrak{h}$ on $\mathfrak{g}_1$ is fully reducible. One has $\mathfrak{g}=\sum_\alpha \mathfrak{k}_\alpha$, $\mathfrak{h}=\mathfrak{k}_0$, and $\mathfrak{g}_1=\sum_{\alpha\neq 0}\mathfrak{k}_\alpha$, where $\mathfrak{k}_\alpha=\{x\in\mathfrak{g} \mid [h,x]=\alpha(h)x, h\in\mathfrak{h}\}$. The linear form α on $\mathfrak{h}$ is called a root of $\mathfrak{g}$. There exist l roots $\alpha_1,\dots,\alpha_l$ ($l=\dim\mathfrak{h}$) such that $\mathfrak{h}$ can be written in the form $\mathfrak{h}=j\mathfrak{k}_{\alpha_1}+\dots+j\mathfrak{k}_{\alpha_l}$, l being the **rank** of $\mathfrak{g}$. Then, after a suitable change of the numbering of the α_i's, any root α will be seen to be of the form $(\alpha_i+\alpha_k)/2$, $(\alpha_i-\alpha_k)/2$ or $\alpha_i/2$, where $1\leqslant i\leqslant k\leqslant l$. A normal j-algebra admits a unique Siegel decomposition which can be constructed by using root spaces.

D. Equivariant Holomorphic Embedding

We retain the notation of Section B. Let $D(V,F)\subset R^{\mathbf{C}}\times W$ be a Siegel domain and $\mathfrak{g}_{\mathbf{C}}$ be the †complexification of the Lie algebra $\mathfrak{g}_h$. $\mathfrak{g}^{-1}$ has the complex structure defined by the endomorphism ad I. Let $\mathfrak{g}_{\pm}^{-1}$ be the $\pm i$-eigenspaces in the complexification $g_{\mathbf{C}}^{-1}$ of g^{-1} under ad I. Let us consider the complex subalgebras $\mathfrak{b}=\mathfrak{g}_{-}^{-1}+\mathfrak{g}_{\mathbf{C}}^{0}+\mathfrak{g}_{\mathbf{C}}^{1}+\mathfrak{g}_{\mathbf{C}}^{2}$ and $\mathfrak{n}=\mathfrak{g}_{\mathbf{C}}^{-2}+\mathfrak{g}_{+}^{-1}$ of $\mathfrak{g}_{\mathbf{C}}$, where the subscripts $\mathbf{C}$ denote the complexification of the respective space. Let $G_{\mathbf{C}}$ be the connected †complex Lie group generated by $\mathfrak{g}_{\mathbf{C}}$ and containing G_h^0 (=the identity component of G_h) as a subgroup. The Lie algebra of the normalizer B of $\mathfrak{b}$ in $G_{\mathbf{C}}$ coincides with $\mathfrak{b}$. Identifying $R^{\mathbf{C}}\times W$ with $\mathfrak{n}$ as a complex vector space, and denoting by π the natural projection of $G_{\mathbf{C}}$ onto the complex coset space $G_{\mathbf{C}}/B$, the composite mapping $\tau=\pi\exp$ is a holomorphic embedding of $\mathfrak{n}$ into $G_{\mathbf{C}}/B$, which induces a holomorphic G_h^0-†equivariant embedding of $D(V,F)$ into $G_{\mathbf{C}}/B$ as an open submanifold. This embedding is called the **Tanaka embedding**. By the **(Shilov) boundary** S of $D(V,F)$ we mean the real submanifold $S=\{(x+iy,u)\in R^{\mathbf{C}}\times W \mid y=F(u,u)\}$ of $R^{\mathbf{C}}\times W$, which is a subset of the boundary of $D(V,F)$. S has the natural †CR-structure induced from the complex structure of $R^{\mathbf{C}}\times W$. Every element of $\mathfrak{g}_h$ can be extended to a unique holomorphic vector field on $R^{\mathbf{C}}\times W$ which is tangent to S, and its restriction to S is an infinitesimal †CR-automorphism on S, i.e., a complete vector field generating a 1-parameter group of †CR-equivalences of S onto itself. Conversely, every infinitesimal CR-automorphism on S can be extended uniquely to a holomorphic vector field on $R^{\mathbf{C}}\times W$, and an element of $\mathfrak{g}_h$ is characterized as an infinitesimal CR-automorphism on S whose extension leaves the †Bergman kernel form of $D(V,F)$ invariant [6]. Let $n=\dim_{\mathbf{C}}\mathfrak{g}_{\mathbf{C}}$, $m=\dim_{\mathbf{C}}\mathfrak{b}$, and let $k=\binom{n}{m}-1$. Then $G_{\mathbf{C}}/B$, and consequently $D(V,F)$, is embedded holomorphically into the complex †Grassmann manifold of m-dimensional subspaces in $\mathfrak{g}_{\mathbf{C}}$ and so into the †complex projective space $P_k(\mathbf{C})$. Any element of G_h^0 is induced from a projective transformation and hence is a birational transformation on $D(V,F)$.

Let D be a homogeneous bounded domain in $\mathbf{C}^n$, $\mathfrak{g}_h$ the Lie algebra of $G_h(D)$, and $\mathfrak{k}_h$ the isotropy subalgebra of $\mathfrak{g}_h$; and let $\mathfrak{g}_{\mathbf{C}}$ be the complexification of $\mathfrak{g}_h$. $\mathfrak{g}_h$ is a j-algebra. Let us define the complex subalgebra $\mathfrak{g}_-$ of $\mathfrak{g}_{\mathbf{C}}$ by putting $\mathfrak{g}_-=\{x+ijx \mid x\in\mathfrak{g}_h, j\in(j)\}$. Then we have $\mathfrak{g}_{\mathbf{C}}=\mathfrak{g}_h+\mathfrak{g}_-$, $\mathfrak{g}_h\cap\mathfrak{g}_-=\mathfrak{k}_h$. Let $G_{\mathbf{C}}$ be the connected Lie group generated by $\mathfrak{g}_{\mathbf{C}}$ and containing $G_h^0(D)$ as a subgroup. Let G_- be the connected (closed) Lie subgroup of $G_{\mathbf{C}}$ generated by $\mathfrak{g}_-$. Then D can be holomorphically embedded in $G_{\mathbf{C}}/G_-$ as the open $G_h^0(D)$-orbit of the origin of $G_{\mathbf{C}}/G_-$ [8]. This embedding is called the **generalized Borel embedding**. $G_{\mathbf{C}}/G_-$ is compact if and only if D is symmetric, and in this case $G_{\mathbf{C}}/G_-$ coincides with the compact dual [9].

Let $\{\mathfrak{t}_0,j,\omega\}$ be a normal j-algebra of rank l corresponding to a homogeneous bounded domain D_0, and define the Hermitian inner product h by $h(x,y)=\omega([jx,y])+i\omega([x,y])$ for $x,y\in\mathfrak{t}_0$. $\mathfrak{t}_0$ has $l-1$ (normal) nontrivial j-ideals (i.e., j-invariant ideals) up to j-isomorphisms. Take a j-ideal $\mathfrak{t}_1$ of $\mathfrak{t}_0$. Then we have $\mathfrak{t}_0=\mathfrak{t}_1+\mathfrak{t}_2$, $\mathfrak{t}_2$ being a (normal) j-subalgebra of $\mathfrak{t}_0$ defined as the orthogonal complement of $\mathfrak{t}_1$ in $\mathfrak{t}_0$ with respect to h. The geometric version of this is that D_0 is represented as a holomorphic fiber space over the homogeneous bounded domain D_2 corresponding to $\mathfrak{t}_2$, with fibers holomorphically equivalent to the homogeneous bounded domain D_1 corresponding to $\mathfrak{t}_1$. For this fibering there exists a universal fiber space D over the product $\hat{D}_2$ of certain classical symmetric domains, with the same fibers, which plays the same role as that of a †universal fiber bundle in topology. Here, D is again a homogeneous bounded domain. The fiber space $D_0\rightarrow D_2$ is induced from the fiber space $D\rightarrow\hat{D}_2$ by the classifying mapping λ of D_2 to $\hat{D}_2$. Let β be the generalized Borel embedding of D into $G_{\mathbf{C}}/G_-$. Then there exists a

complex Abelian subalgebra $\mathfrak{m}$ of $\mathfrak{g}_{\mathbf{C}}$ satisfying $\mathfrak{g}_{\mathbf{C}}=\mathfrak{g}_{-}+\mathfrak{m}$ (semidirect), and one can construct a biholomorphic mapping f of $\beta(D)$ onto a certain Siegel domain of the third kind over D_2 in the vector space $\mathfrak{m}$ [8]. The fiber space $D_0 \rightarrow D_2$ coincides with the one induced from the aforementioned Siegel domain of the third kind by the composite mapping of λ and $f \cdot \beta$. Every realization of a homogeneous bounded domain as a Siegel domain of the third kind is obtained by this method.

E. Classification of Homogeneous Bounded Domains

The main concern is to classify all homogeneous bounded domains in $\mathbf{C}^n$ up to holomorphic equivalence. Since the realization as Siegel domains has been set up, the second step is to get the uniqueness theorem: The holomorphic equivalence of two homogeneous Siegel domains implies that they are linearly equivalent, that is, there exists a (complex) linear isomorphism between the ambient vector spaces which carries the one domain to the other. The uniqueness theorem was first stated in 1963 (Appendix in [1]), rigorously proved in 1967 [10], and in 1970 the homogeneity assumption was removed [2]. A homogeneous Siegel domain is called **irreducible** if it is not holomorphically equivalent to a product of two homogeneous Siegel domains. Every homogeneous Siegel domain is linearly equivalent to a product of irreducible homogeneous Siegel domains [3, 10]. A homogeneous Siegel domain $D(V,F)$ is irreducible if and only if the regular cone V is irreducible, i.e., if V cannot be written as a direct sum of two regular cones. So the problem is to classify irreducible homogeneous Siegel domains up to linear equivalence. This reduces to classifying two kinds of nonassociative algebras with bigradation, called T-algebras and S-algebras [11]. Nonsymmetric homogeneous Siegel domains appear in dimension 4. The numbers of such domains are finite up to dimension 6, but in every dimension $\geqslant 7$ there is at least one continuous family of nonsymmetric irreducible homogeneous Siegel domains, which are not mutually holomorphically equivalent.

There is a remarkable class of homogeneous Siegel domains, called **quasisymmetric** [12], which contains the class of symmetric bounded domains. A regular cone $V \subset R$ is called **self-dual** if there exists an inner product $(\ ,\ )$ on R such that $V=\{x \in R \mid (x,y)>0$ for $y \in \bar{V}-(0)\}$, $\bar{V}$ denoting the closure of V. V is called homogeneous if the group $G(V)$ is transitive on V. Suppose that V is homogeneous self-dual. Then the group $G(V)$ is †self-adjoint with respect to the inner product $(\ ,\ )$ on R. Let $\mathfrak{g}(V)$ be the Lie algebra of $G(V)$. Then the totality $\mathfrak{k}(V)$ of skew-symmetric operators in $\mathfrak{g}(V)$ with respect to $(\ ,\ )$ is a †maximal compact subalgebra of $\mathfrak{g}(V)$ and is the isotropy subalgebra of $\mathfrak{g}(V)$ at a point $e \in V$. Consider the associated †Cartan decomposition $\mathfrak{g}(V)=\mathfrak{k}(V)+\mathfrak{p}(V)$. For each $x \in R$ there exists a unique element $T(x) \in \mathfrak{p}(V)$ such that $T(x)e=x$. Let F be a V-Hermitian form on a complex vector space W. Define a Hermitian inner product $\langle\ ,\ \rangle$ on W by $\langle u,v\rangle=(e,F(u,v))$ for $u, v \in W$, and let $H(W)$ be the set of Hermitian operators on W with respect to $\langle\ ,\ \rangle$. A (homogeneous) Siegel domain $D(V,F) \subset R^{\mathbf{C}} \times W$ is called quasisymmetric if V is homogeneous self-dual and if for each $x \in R$ there exists $R(x) \in H(W)$ such that $F(R(x)u,v)+F(u,R(x)v)=T(x)F(u,v)$ for $u, v \in W$. The normal j-algebra $\mathfrak{t}$ of an irreducible quasisymmetric Siegel domain is characterized by the following conditions: $\dim \mathfrak{k}_{(\alpha_i+\alpha_k)/2}=a$ $(1 \leqslant i \leqslant k \leqslant l)$; and $\dim k_{\alpha_i/2}=b$ $(1 \leqslant i \leqslant l)$, where a, b are some constants and l is the rank of $\mathfrak{t}$ (D'Atri and de Miatello). Quasisymmetric Siegel domains have been completely classified (M. Takeuchi, *Nagoya Math. J.*, 59 (1975), also [12]).

F. Generalized Siegel Domains and Further Results

Let Ω be a domain in $\mathbf{C}^n \times \mathbf{C}^m$ $(n, m \geqslant 0)$ which is holomorphically equivalent to a bounded domain and contains a point of the form $(z,0)$, $z \in C^n$. Ω is called a **generalized Siegel domain** with exponent c $(c \in \mathbf{R})$, if Ω is invariant under holomorphic transformations of $\mathbf{C}^n \times \mathbf{C}^m$ of the types

$$(z,u) \mapsto (z+a,u) \quad \text{for all} \quad a \in \mathbf{R}^n,$$

$$(z,u) \mapsto (z,e^{it}u) \quad \text{for all} \quad t \in \mathbf{R},$$

$$(z,u) \mapsto (e^t z, e^{ct}u) \quad \text{for all} \quad t \in \mathbf{R}.$$

Let D be a bounded domain in $\mathbf{C}^n$, and Γ a subgroup of $G_h(D)$. Γ is said to **sweep** D if there exists a compact set $K \subset D$ such that $\Gamma K=D$. Γ is said to **divide** D if Γ, provided with discrete topology, acts properly on D and sweeps D. D is called **sweepable** (resp. **divisible**) if there exists a subgroup Γ of $G_h(D)$ which sweeps (resp. divides) D. A divisible generalized Siegel domain is symmetric. A sweepable generalized Siegel domain with exponent $c \neq 0$ (resp. $c=0$) is a Siegel domain (resp. a product of a Siegel domain of the first kind and of a homogeneous bounded circular domain) ([13]; also A. Kodama, *J. Math. Soc. Japan*, 33 (1981)).

Some results have been obtained concerning geometry of bounded domains, homogeneous bounded domains, and Siegel domains in

complex Banach spaces [14], and also concerning the unitary representations of the generalized Heisenberg group on the square-integrable cohomology spaces of $\bar{\partial}_b$-complexes on the Shilov boundary of a Siegel domain [15].

References

[1] I. I. Pyatetskiĭ-Shapiro, Automorphic functions and the geometry of classical domains, Gordon & Breach, 1969.
[2] W. Kaup, Y. Matsushima, and T. Ochiai, On the automorphisms and equivalences of generalized Siegel domains, Amer. J. Math., 92 (1970), 267–290.
[3] K. Nakajima, Some studies on Siegel domains, J. Math. Soc. Japan, 27 (1975), 54–75.
[4] J. Dorfmeister, Homogene Siegel-Gebiete, Habilitation, Univ. Münster, 1978.
[5] S. Murakami, On automorphisms of Siegel domains, Lecture notes in math. 286, Springer, 1972.
[6] N. Tanaka, On infinitesimal automorphisms of Siegel domains, J. Math. Soc. Japan, 22 (1970), 180–212.
[7] S. G. Gindikin, E. B. Vinberg, and I. I. Pyatetskiĭ-Shapiro, Homogeneous Kähler manifolds, Geometry of Homogeneous Bounded Domains, C.I.M.E. 1967; Cremonese, 1968, 3–87.
[8] S. Kaneyuki, Homogeneous bounded domains and Siegel domains, Lecture notes in math. 241, Springer, 1971.
[9] K. Nakajima, On Tanaka's embeddings of Siegel domains, J. Math. Kyoto Univ., 14 (1974), 533–548.
[10] S. Kaneyuki, On the automorphism groups of homogeneous bounded domains, J. Fac. Sci. Univ. Tokyo, 14 (1967), 89–130.
[11] M. Takeuchi, Homogeneous Siegel domains, Publ. study group geometry 7, Kyoto, 1973.
[12] I. Satake, Algebraic structure of symmetric domains, Iwanami and Princeton Univ. Press, 1980.
[13] J. Vey, Sur la division des domaines de Siegel, Ann. Sci. Ecole Norm. Sup., 3 (1970), 479–506.
[14] J. P. Vigué, Le groupe des automorphismes analytiques d'un domaine borné d'un espace de Banach complexe, Ann. Sci. Ecole Norm. Sup., 9 (1976), 203–282.
[15] H. Rossi and M. Vergne, Group representations on Hilbert spaces defined in terms of $\bar{\partial}_b$-cohomology on the Shilov boundary of a Siegel domains, Pacific J. Math., 65 (1976), 193–207.

385 (XVI.6) Simulation

A. General Remarks

Simulation, in its widest sense, is a method of utilizing models to study the nature of certain phenomena. This method is employed when experimentation with the actual phenomena in question is difficult because of high cost in time or money. Also, it is sometimes almost impossible to carry out observations when the behavior of the objects can be influenced by their surroundings.

We can classify simulation techniques into the following four types, although simulations in practical use are usually a mixture of them.

The first type is **model experimentation**, which includes model basins and wind tunnels in hydrodynamics and pilot plants in the chemical industry. In advance of construction in a real situation, we perform experiments on a small scale and verify or modify those theories upon which the construction is based.

The second type is **analog simulation** or **experimental analysis**. We investigate the properties of real objects by experiments on alternative phenomena satisfying the same differential equations as those known for or assumed to be satisfied by the real objects. For example, we use an equivalent electric network to study dynamic vibration, and dynamic systems to study heat conduction problems. When theoretical analysis of the actual phenomenon is difficult, we look for other phenomena with similar properties and study them in order to construct mathematical models for them. This type of simulation has come into practical use mainly in engineering problems, but recently it has been utilized for the study of economic phenomena, nervous systems, the circulating system of an artificial heart, etc. Analog simulation was in the past often performed by means of †analog computers. Nowadays, analog simulation is more frequently performed by digital computers than by analog ones. And with the progress of electronics it has become easier to make special-purpose simulators.

The third method of simulation, **simulation in the narrow sense**, has become more important as †digital computers have been developed. In general this method is applied to problems that are more complicated and of larger scale than problems treated by analog simulation. When the mathematical expressions of the phenomenon and the algorithms of its dynamic structure are known, it is easy to simulate it by means of a computer program. In

particular, when these techniques are used to study systems such as sets of machines, equipment at factories, or management organization, we call them **system simulations**. Major fields where system simulation techniques have been used are traffic control on highways or at airports, arrangement or operation of machines at factories, balancing problems in chemical processes, production scheduling in connection with demands and stocks, overall management problems, and design of information systems. The method has also been applied in designing plants and highways and in the study of social or biological phenomena. Also, when we investigate instruction systems of computers that are yet to be completed or develop programming systems for such computers, existing computers can be used to simulate the new ones. †Random numbers play an important role where the simulation must include random fluctuations (→ 354 Random Numbers). In such instances, the method is often called the Monte Carlo method (→ Section C).

The fourth method of simulation deals with systems containing human beings. Among them are war games for training in military-operation planning, business games for training in business enterprises, and simulators for training pilots and operators of atomic power plants. The contribution of human decision to simulation processes is characteristic of these cases. For example, the participants in a business game are divided into several enterprise groups. Each group discusses and decides how to invest in plants, equipment, research, and advertising and how to schedule production for each quarter. On the basis of the decisions, a computer outputs the records of sales, stocks, and cash for each quarter, according to hidden rules. From the results, each group decides on the next steps. In this way, the groups compete for development. This type of simulation is important not only for training but also for investigating the mechanism of human decision. Slightly different from this type of simulation, the "perceptron" and EPAM (Elementary Perceiver and Memorizer) are related to artificial intelligence and have been used extensively in the cognitive sciences and in research into the structure and function of the human brain.

The third type of simulation has attracted attention in particular and has been used both in theoretical problems, such as the explication of various phenomena, and in practical problems, such as design or optimum operation of systems or prediction of their behavior. For school education and training of technicians, this is put to use together with simulations of the fourth type. But their use has also induced heated discussions and controversies on the validity of results.

B. Programming Languages for Simulation

We usually describe models by using general purpose language: such as FORTRAN, or list processing languages such as LISP to simulate situations on computers. For system simulation, a number of programming languages have been developed and put to practical use. They can be divided roughly into two categories: those for which systems change continuously and those describing discrete changes. CSMP (Continuous System Modeling Program), CSSI (Continuous System Simulation Language), and DDS (Digital Dynamics Simulator) belong to the former, and hence all involve integration mechanisms; but each has a different way of describing a model. DYNAMO, which has been implemented, or J. W. Forrester's Industrial Dynamics and World Dynamics, are used extensively. To control simulation time, one may use GPSS (General Purpose Simulation System), SIMULA (Simulation Language), or SIMSCRIPT (Simulation Scriptor), each employing a different method to describe state transitions.

C. The Monte Carlo Method

The **Monte Carlo method** was introduced by J. von Neumann and S. M. Ulam around 1945. They defined this as a method of solving deterministic mathematical problems using †random numbers. L. de Buffon's needle experiment, in which the approximate value of π is obtained by dropping needles at random many times, is a classical example of this method.

Another example is the problem of evaluating a definite integral $I=\int_a^b f(x)\,dx$ $(B\geqslant f(x)\geqslant A\geqslant 0)$. First we generate many pairs of (uniform) random numbers (x, y), where $a\leqslant x\leqslant b$ and $A\leqslant y\leqslant B$. The proportion (p) of pairs satisfying $y\leqslant f(x)$ gives an estimate of the integral, i.e., $I\fallingdotseq p(B-A)(b-a)$. The techniques for inverting matrices, solving †boundary value problems of partial differential equations, and so on, are also examples of the Monte Carlo method in this sense. However, direct numerical calculation seems to be more useful in dealing with this sort of problem. At present, *Monte Carlo methods* are usually used when it is difficult to construct (or solve) mathematical equations describing the phenomena in question, for example, when the phenomena involve †stochastic processes such as †random

walks. Some methods have been devised so as to get precise results efficiently.

References

[1] B. P. Zeigler, Theory of modeling and simulation, Wiley, 1976.
[2] R. S. Lehman, Computer simulation and modeling, Wiley, 1977.
[3] R. E. Shannon, Systems simulations: The art and science, Prentice-Hall, 1975.
[4] J. M. Hammersley and D. C. Handscomb, Monte Carlo methods, Wiley, 1964.
[5] A. Newell and H. A. Simon, Human problem solving, Prentice-Hall, 1972.
[6] M. Minsky and S. Papert, Perceptrons, MIT Press, 1969; second printing with corrections by the authors, June 1972.
[7] J. W. Forrester, World dynamics, Wright-Allen Press, 1972.

386 (XX.29) S-Matrices

A. Basic Notion

It is often useful to focus attention on the relation between a physical system's input and output, without worrying about intermediate processes (the black box), which may be insufficiently understood or too complicated to analyze. For the scattering of particles, this leads to the notion of an **S-matrix** that directly relates the state of incoming particles (before scattering processes take place) to that of outgoing (scattered) particles.

In typical cases, the incoming and outgoing particles are described as mutually noninteracting (these are called free particles). This implies particle motions along straight lines at constant speeds (asymptotic to the actual motion at infinite past for incoming particles and at infinite future for outgoing particles) in classical mechanics, and wave functions (or vectors in a Hilbert space) obeying the Schrödinger equation with a free Hamiltonian H_0 in quantum mechanics and more or less the same in quantum field theory.

A wave function for n particles is an L_2-function of their momenta $p_1, \ldots, p_n$ (each p_j being a 3-dimensional vector) with respect to an appropriate measure (normally the Lebesgue measure $d\mu(p)=\prod d^3p_j$ in quantum mechanics and the Lorentz-invariant measure $d\mu(p)=\prod\{(m^2+p_j^2)^{-1/2}d^3p_j\}$ in quantum field theory, where m is the mass of the particle), and the S-matrix S is described in terms of the S-matrix elements $(p_1, \ldots, p_n|S|p'_1, \ldots, p'_{n'})$ (which is a distribution) as

$$(\Phi, S\Psi)=\int\overline{\Phi(p)}(p|S|p')\Psi(p')d\mu(p)d\mu(p'),$$

where Φ and Ψ are wave functions for n and n' particles, $p=(p_1, \ldots, p_n)$, and similarly for p'. $(p|S|p')$ gives quantities measured in scattering experiments, as will be explained in Section B (2).

In quantum mechanics, the free and actual (interacting) motion of particles is described in the same Hilbert space with free and interacting Hamiltonians H_0 and H. A vector Φ in interacting motion behaves like a vector φ in free motion at infinite past if

$$\|(\exp[-iHt])\Phi-(\exp[-iH_0t])\varphi\|\to 0 \quad \text{as} \quad t\to-\infty$$

and hence

$$\Phi=W_-(H;H_0)\varphi,$$

$$W_-(H;H_0)=\lim_{t\to-\infty}e^{iHt}e^{-iH_0t}.$$

Such a Φ is often written as

$$\Phi=\Phi^{\mathrm{in}}(\varphi)=\int\Phi^{\mathrm{in}}(p)\varphi(p)d\mu(p),$$

and is called an **in-state**. A definition for an **out-state** Φ^{out} is obtained by changing $t\to-\infty$ to $t\to+\infty$ and W_- to W_+. The S-matrix element is defined (as a distribution) by

$$(p|S|p')=(\Phi^{\mathrm{out}}(p), \Phi^{\mathrm{in}}(p')).$$

The existence and properties of $W_\pm$ (called †wave operators) are central subjects in scattering theory (→ 375 Scattering Theory).

In quantum field theory, the asymptotic description is given in terms of vectors in †Fock space, and the in- and out-states are constructed in terms of †asymptotic fields (→ 150 Field Theory).

The foregoing description actually applies only to a system of one-component particles of the same kind. More generally, additional (discrete) variables, say α, are needed to distinguish different kinds of particles and different spin components of each kind of particle, and the p's appearing in the above formulas should be replaced by (p, α)'s, along with related changes in the measure.

Even in a quantum mechanics of many identical particles a bound state, if it exists, is to be treated as another particle (different from the original one) and should be distinguished by α's in the asymptotic description φ.

If the interaction is of long range (e.g.,

Coulomb interaction), the classical path of a particle does not have an asymptote in general, and correspondingly the wave operators $W_\pm$ do not exist for the usual free Hamiltonian. Still, an asymptotic description of scattering is possible in some cases.

In the presence of massless particles, such as photons in quantum field theory, another difficulty, called the infrared problem, can arise in the asymptotic description of scattering because the scattered particle may be accompanied by an infinite number of massless particles (with very small energy). In such a situation, a representation of a free massless field not equivalent to the standard Fock representation is believed to be a possible candidate for the asymptotic description of scattering.

B. Basic Properties

(1) Invariance. Let $\mathscr{H}_0$ be the Hilbert space for the asymptotic description of scattering, such as the space of L_2-functions $\varphi(p_1, \ldots, p_n)$ relative to the measure $d\mu(p)$. The **S-matrix** is an operator on $\mathscr{H}_0$ whose matrix element is as described above. (The corresponding operator **S** in the Hilbert space $\mathscr{H}$ describing the interacting states is defined by $\mathbf{S}\Phi^{\mathrm{out}}(\varphi)=\Phi^{\mathrm{in}}(\varphi)$ and is sometimes called an **S-operator**.)

S is said to be **invariant** under a group G of transformations of the p's (and possibly the α's) if $(U(g)\varphi)(p)=\varphi(g^{-1}p)$ defines a continuous unitary representation $U(g)$ and if $U(g)S=SU(g)$ for all $g\in G$. First, S is usually invariant under time translation. In the quantum mechanics of a particle scattered by a rotationally invariant potential, S is invariant under the group of rotations of 3-dimensional vectors p; in the quantum mechanics of many particles (mutually interacting through central potentials), S is invariant under the 3-dimensional Euclidean group of transformations $p\rightarrow Rp+a$ with rotation R; in relativistic field theory, S is assumed to be invariant under the †inhomogeneous Lorentz group of transformations $p\rightarrow\Lambda p+a$ with $p=(p^0,\mathbf{p})$ and homogeneous Lorentz transformation Λ.

In all these examples, G is of type I, i.e., there is a direct integral decomposition

$$\mathscr{H}_0=\int \mathscr{H}(k)\otimes\mathscr{L}(k)\,dv(k),$$

$$U(g)=\int U_k(g)\otimes 1_{\mathscr{L}(k)}\,dv(k),$$

into irreducible representations U_k on $\mathscr{H}(k)$, which are mutually inequivalent, where $\mathscr{L}(k)$ is some Hilbert space for each k. The S-matrix is invariant under G if and only if

$$S=\int 1_{\mathscr{H}(k)}\otimes S(k)\,dv(k).$$

When a (scalar) particle is scattered by a central potential, irreducible representations are labeled by the energy $E(p)=p^2/(2m)$ (for time translation) and the angular momentum l (for rotations) with $\dim\mathscr{L}(E(p),l)=1$. Therefore each $S(k)=S_l(|p|)$ is a number. For any given energy, $l=0,1,2,\ldots$ are referred to as the S-wave, P-wave, D-wave, ... or generally as **partial waves.**

For relativistic scalar particles, irreducible representations (with positive energy and nonzero real mass) are labeled by the center-of-mass energy squared s $(=(\Sigma(m^2+p_i^2)^{1/2})^2-(\Sigma\,p_i)^2)$ and the total angular momentum l. If s is below the threshold $(3m)^2$ of 3-particle scattering, $\dim\mathscr{L}(k)=1$ and $S(k)=S_l(s)$ is a number.

(2) Unitarity. $S\varphi$ is supposed to represent the $t=+\infty$ asymptotic (free) behavior of the state that initially (i.e., at $t=-\infty$) behaves like a free state φ. If there is no loss of probability in the description of scattering, the S-matrix is isometric. If all asymptotic configurations are realized as a result of scattering, the S-matrix must also be unitary. The mappings $\Phi^{\overset{\mathrm{out}}{\mathrm{in}}}:\varphi\in\mathscr{H}_0\rightarrow\Phi^{\overset{\mathrm{out}}{\mathrm{in}}}(\varphi)\in\mathscr{H}$ ($W_\pm$ in quantum mechanics) are proved to be isometric under a general assumption. The unitarity is then proved in potential scattering (under some conditions on the potential) by showing that the two wave operators $W_\pm$ have the same range. In fact, a somewhat stronger result—that this range is the same as the absolutely continuous spectral subspace for the interacting Hamiltonian H—is usually proved and is called **completeness (of scattering states** in the absolutely continuous spectral subspace of H). If the mappings Φ^{out} and Φ^{in} (or $W_\pm$) are isometric and have the same range, then $\Phi^{\mathrm{in}}(\varphi)=\Phi^{\mathrm{out}}(S\varphi)$, which shows that the state behaving like φ at $t=-\infty$ behaves like $S\varphi$ at $t=+\infty$.

In the simple multiplicity cases such as $S_l(|p|)$ and $S_l(s)$ above (which correspond to the physical situation of purely elastic scattering without any production or change of particles), these numbers must be of the form $e^{2i\delta_l}$ due to unitarity, where the real number δ_l is called the **phase shift.** In terms of the phase shift, the **differential cross section** $d\sigma/d\Omega$, which is the average number of particles scattered per unit time per unit solid angle around the direction forming an angle θ with the incident uniform parallel beam of unit intensity (one particle per unit time per unit area) when

viewed in the center-of-mass system, and the **total elastic cross section** $\sigma_{el}=\int(d\sigma/d\Omega)d\Omega$ are given by the following formulas, called the **partial wave expansion**:

$$d\sigma/d\Omega=|f(s,\theta)|^2,$$

$$f(s,\theta)=k^{-1}\sum_{l=0}^{\infty}(2l+1)f_l P_l(\cos\theta), \qquad f_l=\sin\delta_l e^{i\delta_l},$$

$$\sigma_{el}=4\pi k^{-2}\sum_l(2l+1)\sin^2\delta_l,$$

where $d\Omega=\sin\theta\, d\theta\, d\varphi$ (invariant measure on a 2-dimensional sphere S^2) and k is the wave number of the particle in the center-of-mass system ($k=\hbar^{-1}|p|=\hbar^{-1}[(s/4)-m^2]^{1/2}$). The function f is called the **scattering amplitude**. The forward scattering amplitude $f(s,0)$ is related to the total cross section σ_{tot} by the **optical theorem**:

$$\sigma_{tot}=4\pi k^{-1}\operatorname{Im}f(s,0),$$

which follows from the unitarity of the S-matrix. Here the **total cross section** is expressed as the area of the transverse cross section of a classical (impenetrable) scatterer that would scatter the same amount of particles. In a purely elastic region, $\sigma_{tot}=\sigma_{el}$, and the optical theorem is the same as the assertion $\operatorname{Im}f_l=|f_l|^2$.

Even for values of s above the threshold of inelastic scattering, the restriction of S_0 to the subspace of two particles is again described by numbers $e^{2i\delta_l}$, where δ_l is now complex, and the same formulas for the differential and total elastic cross sections hold, except that unitarity of S_0 now implies $\operatorname{Im}f_l\geq|f_l|^2$.

(3) TCP Symmetry. In the framework of either †axiomatic quantum field theory or the †theory of local observables, the TCP theorem (or PCT theorem) shows [1] that to every particle there corresponds another particle with the same mass and spin, which is called the **antiparticle** and can be the original particle itself, such that any particle is the antiparticle of its antiparticle, and the following relation, called **TCP invariance** (or **PCT invariance** [2]) holds for S-matrix elements:

$$(p,\alpha|S|p',\alpha')$$

$$=\eta(\alpha)\eta(\alpha')i^{F(\alpha)-F(\alpha')}(p',\theta_0\alpha'|S|p,\theta_0\alpha).$$

Here $F(\alpha)$ and $F(\alpha')$ are the number of particles with half odd integer spin amongst the incoming and outgoing particles (i.e., in α and α'), $\eta(\alpha)$ and $\eta(\alpha')$ are the product of ±1 assigned to each particle in α and α', respectively, with the assignment to a particle and its antiparticle the same for bosons and opposite for fermions, (p,α) is an abbreviation for an ordered n-tuple of energy-momentum 4-vectors p_i and other indices α_i (for spin components and particle species), $i=1,\ldots,n$, and similarly for (p',α'), and θ_0 is the mapping from particles to their antiparticles (without changing spin indices). We may view the mapping $|p,\alpha)\rightarrow\eta(\alpha)i^{F(\alpha)}|p,\theta_0\alpha)$ as an antiunitary operator Θ_0 on $\mathscr{H}_0$ satisfying $\Theta_0 S\Theta_0^{-1}=S^*$ so that

$$(\varphi_1,S_0\varphi_2)=(\Theta_0\varphi_2,S_0\Theta_0\varphi_1).$$

This is related to the †TCP operator Θ in quantum field theory by $\Theta\Psi^{out}(\varphi)=\Psi^{in}(\Theta_0\varphi)$ and $\Theta\Psi^{in}(\varphi)=\Psi^{out}(\Theta_0\varphi)$. The name TCP comes from the combination of T for time reversal (incoming $\rightleftarrows$ outgoing, $v\equiv p/m\rightarrow -v$), C for charge conjugation (particle $\rightleftarrows$ antiparticle), and P for parity, which is a quantum number for space inversion ($p\rightarrow -p$).

TCP symmetry was suggested to G. Lüders (*Dansk. Mat. Fys. Medd.*, 1954) by B. Zumino in the form that P-invariance implies TC symmetry. W. Pauli (*Niels Bohr and the development of physics*, Pergamon 1955, 30) realized that TCP is a symmetry. R. Jost (*Helv. Phys. Acta*, 1957) gave its proof in the framework of axiomatic quantum field theory.

(4) Crossing Symmetry. In the framework of either axiomatic quantum field theory or the theory of local observables, it has been shown by J. Bros, H. Epstein, and V. Glaser (*Comm. Math. Phys.*, 1 (1965); also [3]) that there exist analytic functions $H(k,\alpha)$ of complex variables $k=(k_1,\ldots,k_4)$ in a certain domain D such that each k_j is a complex 4-vector, the variables are on the mass-shell manifold defined by $k_j^2=m_j^2$ (k_j^2 is in the Minkowski metric) and $\sum k_j=0$, and the boundary value of $H(k,\alpha)$ as k_j approaches $\varepsilon_j p_j$ from $\operatorname{Im}s>0$ (s being the square of the sum of two k's for incoming particles) in D is the scattering amplitude for the following processes involving the particles A_j and their antiparticles $\bar{A}_j$ ($j=1,\ldots,4$) with 4-momenta p_j, $\varepsilon_j=1$ for A_j and -1 for $\bar{A}_j$ (some of the A's and $\bar{A}$'s may coincide):

(i) $A_1+A_2\rightarrow\bar{A}_3+\bar{A}_4$,

(i′) $A_3+A_4\rightarrow\bar{A}_1+\bar{A}_2$,

(ii) $A_1+A_3\rightarrow\bar{A}_2+\bar{A}_4$,

(ii′) $A_2+A_4\rightarrow\bar{A}_1+\bar{A}_3$,

(iii) $A_1+A_4\rightarrow\bar{A}_2+\bar{A}_3$,

(iii′) $A_2+A_3\rightarrow\bar{A}_1+\bar{A}_4$.

Any pair of relations taken from (i)–(iii) constitutes a **crossing symmetry**.

(5) High Energy Theorems. M. Froissart (*Phys. Rev.*, 123 (1961)) obtained from the Mandelstam representation the following bound for the forward scattering amplitude (called the

Froissart bound) at large s:

$$|F(s,0)|<(\text{const})s(\log s)^2,$$

$$F(s,t)=kf(s,\theta), \quad t=2k^2(\cos\theta-1).$$

A. Martin has obtained such a bound in the axiomatic framework. As a consequence we have the **Froissart-Martin bound**:

$$\sigma_{\text{tot}}<4\pi R^{-1}(1+\varepsilon)^2(\log(s/s_0))^2,$$

where $\varepsilon>0$ is arbitrary, R can be taken to be $4m_\pi^2$ (m_π is the mass of a pion) for many cases, such as $\pi\pi$, πK, πN, and $\pi\Lambda$ scattering, and s_0 is an unknown constant.

Many other upper and lower bounds for scattering amplitudes have been obtained under other assumptions [4, 5].

I. Ya. Pomeranchuk (*Sov. Phys. JETP*, 7 (1958)) suggested the asymptotic coincidence of total cross sections at high energy for scattering of AB and $A\bar{B}$, where $\bar{B}$ is the antiparticle of B. This **Pomeranchuk theorem** has been shown to hold by Martin (*Nuovo Cimento*, 39 (1965)) by using the analyticity derived in the axiomatic framework under the following sufficient condition: The existence of $\lim_{s\to\infty}[\sigma(AB)-\sigma(A\bar{B})]$ and $\lim_{s\to\infty}(s\log s)^{-1}f(s,0)=0$.

C. S-Matrix Approach

(1) History. All the information needed to understand the experimental elementary-particle scattering data seems to be expressible by S-matrix elements. It was therefore natural to try to develop a foundation (and practical methods of calculation) for the theory of elementary particles on the basis of some general properties of the S-matrix, especially when other approaches, such as quantum field theory, faced difficulties. W. Heisenberg (*Z. Phys.*, 120 (1943)) first pointed out the possibility of such an approach soon after the introduction of the S-matrix by J. A. Wheeler (*Phys. Rev.*, 52 (1937)). Unfortunately, in the early 1940s not much dynamical content could be given to such an S-matrix-theoretic approach because only unitarity and some invariance properties, such as †Lorentz invariance, were used to characterize the S-matrix. In the late 1950s, through the study of the analyticity of the S-matrix in connection with dispersion relations in quantum field theory, it became evident that causality is another important determinant of S-matrix structure. In practice, causality in position space is used in the form of the analyticity of the S-matrix elements in the energy-momentum space (variables dual to space-time positions in the Fourier transform). Subsequently it was realized that analyticity combined with unitarity gives surprisingly strong control over the structure of the S-matrix (G. F. Chew [6]; H. P. Stapp, *Phys. Rev.*, 125 (1962); J. C. Polkinghorne, *Nuovo Cimento*, 23 and 25 (1962); J. Gunson, *J. Math. Phys.*, 6 (1965); D. I. Olive, *Phys. Rev.*, 135B (1964); Chew [7]; R. J. Eden et al. [8]). The study of the S-matrix based on its general properties, such as invariance, unitarity, and analyticity, is called **S-matrix theory**. In the present form, it is adapted only to massive particles with short-range interactions, and its applicability is believed to be limited to strongly interacting systems.

(2) Landau-Nakanishi Variety. C. Chandler and Stapp (*J. Math. Phys.*, 10 (1969)) and D. Iagolnitzer and Stapp (*Comm. Math. Phys.*, 14 (1969)) clarified the analytic structure of the S-matrix in terms of Landau equations (→ 146 Feynman Integrals) based on the important physical idea of **macroscopic causality**, which gives much more precise information in the physical region than a superficial application of †locality (also called microcausality) in axiomatic quantum field theory, though it is possible that a detailed study starting from microcausality and incorporating †asymptotic completeness (the so-called nonlinear program in axiomatic quantum field theory) might eventually entail the macroscopic causality condition (e.g., J. Bros, in [9]; Iagolnitzer [10, ch. IV]; also K. Symanzik, *J. Math. Phys.*, 1 (1960)).

(3) Microlocal Analysis. An important fact is that the **normal analytic structure** of the S-matrix discussed by Iagolnitzer and Stapp essentially coincides with the notion of analyticity in microlocal analysis, i.e., with microanalyticity (→ 274 Microlocal Analysis; F. Pham and Iagolnitzer, *Lecture notes in math. 449*, Springer, 1975; M. Sato, *Lecture notes in phys. 39*, Springer, 1975)). In a word, the †Landau equations have acquired a new interpretation in the description of the microanalytic structure of the S-matrix. In the new developments, the Landau equations define a variety in the cotangent bundle (over the mass-shell manifold in momentum space), and the †singularity spectrum of S-matrix elements is assumed to be confined to $\bigcup_G \mathscr{L}^+(G)$ (except for the so-called $\mathscr{M}_0$-points), where G ranges over all possible **Feynman graphs** and $\mathscr{L}^+(G)$ denotes the positive-α Landau-Nakanishi variety associated with G (→ 146 Feynman Integrals). The union $\bigcup_G \mathscr{L}^+(G)$ is known to be locally finite and hence makes sense (Stapp, *J. Math. Phys.*, 8 (1967)). The old interpretation of Landau equations, as defining a variety in energy-momentum space, corresponds now to considering the variety $L^+(G)$

obtained by projecting $\mathscr{L}^+(G)$ from the cotangent bundle to the base manifold (i.e., the mass shell manifold). The new interpretation of Landau equations led Sato (*Lecture notes in phys. 39*, Springer, 1975) to make a further intriguing conjecture that the S-matrix would satisfy a special †overdetermined system (a †holonomic system) of †(micro-) differential equations whose †characteristic variety is given by the complexification of Landau-Nakanishi varieties. This conjecture is closely related to the monodromy-theoretic approach by T. Regge (*Publ. Res. Inst. Math. Sci.*, 12 suppl. (1977) and the references cited therein) and his co-workers.

(4) Discontinuity Formula. It has turned out that the above approach is closely related to the so-called **discontinuity formula** obtained by combining the unitarity and the analyticity of the S-matrix. Actually T. Kawai and Stapp (*Publ. Res. Inst. Math. Sci.*, 12 suppl. (1977)) have shown that Sato's conjecture can be verified at several physically important points on the basis of the discontinuity formula. The discontinuity formula was first found by R. E. Cutkosky (*J. Math. Phys.*, 1 (1960)) for Feynman integrals. It describes the ramification property of the integral around its singularities (→ 146 Feynman Integrals). An analogous formula has been shown to be valid also for the S-matrix, and it demonstrates how strict are the constraints derived from unitarity and analyticity (Eden et al. [8, ch. 4]; M. J. D. Bloxham et al., *J. Math. Phys.*, 10 (1969); J. Coster and Stapp, *J. Math. Phys.*, 10 (1969); also Stapp in [11] and Iagolnitzer [10]). Note, however, that the derivation of the hitherto-known discontinuity uses either some *ad hoc* assumptions or some heuristic reasoning which is not rigorous or sometimes is even erroneous from the mathematical viewpoint. Efforts to give a rigorous proof are still being made, and these present several mathematically interesting problems (e.g., Iagolnitzer in [9] and M. Kashiwara and Kawai in [9]).

(5) Regge Poles. The results stated so far concerning the analyticity of the S-matrix have been primarily derived in the low-energy region. It is commonly hoped that these results can be related to its high-energy behavior through the inner consistency of S-matrix theory, even though it is still unclear to what extent such a relationship can be developed. Such a hope was advocated by Chew, who had been inspired by the results of Regge (*Nuovo Cimento*, 14 (1959); 18 (1960)) for potential scattering. After being adapted to the relativistic case, Regge's idea took the following form: Consider the scattering of two incoming and two outgoing scalar particles with equal mass $m>0$. Let $f_l(s)$ be the partial scattering amplitude defined earlier. Regge introduced the idea of extending the function $f_l(s)$ to an analytic function $f(l,s)$ ($l\in\mathbf{C}$) and of applying the Sommerfeld-Watson transformation in order to replace the partial wave expansion by the integral

$$\sqrt{-1}/2\int_C (2l+1)f(l,s)P_l(-\cos\theta)\,dl/\sin\pi l$$

$$=F(s,t)$$

for a certain contour C in the complex l-plane which encircles $\{0,1,2,\dots\}$. If $f(l,s)$ is meromorphic in $\operatorname{Re} l>-1/2$ and if it tends to zero sufficiently rapidly at infinity, then one can change the contour C so that, with the help of Cauchy's integral formula, $F\sim\text{constant}\cdot t^{\alpha(s)}$, $\alpha(s)=\max\operatorname{Re} l(s)$, where the maximum is taken over all the poles of $f(l,s)$. Thus the poles of the extended function $f(l,s)$ determine the asymptotic behavior of F as $t\to\infty$ (**Regge behavior**) under the assumption that $f(l,s)$ can be chosen to satisfy suitable analyticity and growth order conditions. These poles are called **Regge poles**. Even though meromorphy conditions are found to be satisfied for scattering by a (Yukawa) potential, they do not seem to be satisfied for the full S-matrix in the relativistic case. More general cases than those discussed here, i.e., the cases where more variables are considered, are also being studied but without full success at the moment. For details and references → [7, 12, 13].

(6) Veneziano Model. In connection with Regge-pole theory, we note an interesting observation by G. Veneziano (*Nuovo Cimento*, 57A (1968)) to the effect that $\Gamma(1-\alpha(s))\Gamma(1-\alpha(t))/\Gamma(1-\alpha(s)-\alpha(t))$, with $\alpha(s)$ being linear in s, satisfies a crossing symmetry (in s and t) and shows an exact Regge-pole behavior. Although the many results that have been obtained give rise to a hope of constructing a realistic model of the S-matrix starting from the aforementioned function, no one has yet succeeded [14]. A more promising approach is the topological expansion procedure in which the first term of the expansion apparently shares with the potential-scattering functions the property of having only poles in the complex l-plane, along with several other physically important properties of Veneziano's function.

References

[1] H. Epstein, CPT invariance of the S-matrix in a theory of local observables, J. Math. Phys., 8 (1967), 750–767.

[2] R. F. Streater and A. S. Wightman, PCT, spin and statistics, and all that, Benjamin, 1964.
[3] H. Epstein, in Axiomatic field theory, M. Chretien and S. Deser (eds.), Gordon & Breach, 1966.
[4] A. Martin and F. Cheung, Analyticity property and bounds of the scattering amplitudes, Gordon & Breach, 1970.
[5] A. Martin, Scattering theory in unitarity, analyticity and crossing, Lecture notes in phys. 3, Springer, 1969.
[6] G. F. Chew, S-matrix theory of strong interaction physics, Benjamin, 1962.
[7] G. F. Chew, The analytic S-matrix: A basis for nuclear democracy, Benjamin, 1966.
[8] R. J. Eden, P. V. Landshoff, D. I. Olive, and J. C. Polkinghorne, The analytic S-matrix, Cambridge Univ. Press, 1966.
[9] D. Iagolnitzer (ed.), Proc. Les Houches, Lecture notes in phys. 126, Springer, 1980.
[10] D. Iagolnitzer, The S-Matrix, North-Holland, 1978.
[11] R. Balian, and D. Iagolnitzer (ed.), Structural analysis of collision amplitudes, North-Holland, 1976.
[12] S. Frautschi, Regge poles and S-matrix theory, Benjamin, 1963.
[13] P. O. B. Collins, An introduction to Regge theory and high energy physics, Cambridge Univ. Press, 1977.
[14] M. Jacob, Dual theory, North-Holland, 1974.

387 (XX.34) Solitons

A. General Remarks

Solitons are nonlinear waves that preserve their shape under interaction. Mathematically, the theory of solitons continues to develop as a theory of completely integrable mechanical systems. Typical examples are the **Korteweg–de Vries equation** (→ Section B), the **Toda lattice** (→ 287 Nonlinear Lattice Dynamics), and the **Sine-Gordon equation**

$$u_{tt}-u_{xx}=\sin u,$$

studied classically in connection with transformations of surfaces of constant negative curvature.

B. The KdV Equation

In the late 19th century J. Boussinesq and then D. Korteweg and G. de Vries obtained equations describing water waves having traveling-wave solutions. The equation

$$u_t-6uu_x+u_{xx}=0, \qquad u=u(x,t), \tag{1}$$

derived by de Vries is called the **KdV equation** for short. Putting $u(x,t)=s(x-ct-\delta)$, we find that s is an †elliptic function, and we obtain

$$s=-\frac{c}{2}\operatorname{sech}^2\left(\frac{\sqrt{c}}{2}x\right)$$

as its degenerate form. This solution is called a **solitary wave**.

Around 1965, M. Kruskal and N. Zabusky solved the KdV equation numerically, taking several separated solitary waves as the initial data. They found that the waves interact in a complicated way but that eventually the initial solitary waves reappear. Noting the particle-like character of the waves, they called each of these waves a **soliton**. Subsequently, the KdV equation was found to have an infinite number of constants of motion.

C. Gardner, J. Greene, Kruskal, and R. Miura associated the 1-dimensional †Schrödinger operator $-d^2/dx^2+u(x,t)$ to each solution $u(x,t)$ of the KdV equation and showed that its †eigenvalues are preserved in time. Moreover, they applied inverse scattering theory (→ Section D) and obtained explicit formulas for the solutions.

C. Lax Representation

Let

$$L=-D^2+u(x), \qquad D=d/dx.$$

For

$$M=-4D^3+6uD+3u_x,$$

the commutator $[M,L]=ML-LM$ is the operator of multiplication by the function $6uu_x-u_{xxx}$. So $u=u(x,t)$ is a solution of the KdV equation if and only if

$$L_t=[M,L]. \tag{2}$$

Equation (2) is also the condition that all $L=L(t)$ are †unitarily equivalent to each other and the †spectrum of L is preserved through time.

Most equations having soliton solutions can be represented in the form of (2) for a suitable pair of L and M. This representation is called the **Lax representation**. One sometimes says that an **isospectral deformation** of L is given by (2).

On the other hand, isomonodromic deformations (→ 253 Linear Ordinary Differential Equations (Global Theory)) have been studied extensively by M. Sato and his co-workers, and relations to soliton theory have been

discovered (Sato, T. Miwa, and M. Jimbo, *Publ. Res. Inst. Math. Sci.*, 14 (1978), 15 (1979)).

In the present case, the requirement that the commutator $[M, L]$ be a multiplication by a function determines an essentially unique $(2n+1)$th-order ordinary differential operator $M = A_n$, the differential operator part of the †fractional power $L^{n+1/2}$. $[A_n, L]$ is a polynomial in u and its derivatives, denoted by $K_n[u]$. The equation $u_t = K_n[u]$ is called the nth KdV equation. The transformation taking the initial data $u(x)$ to the solution $u_n(x,t)$ of the nth KdV equation is denoted by $T_n(t)$. Then $T_n(t)T_m(s) = T_m(s)T_n(t)$, i.e., the flows defined by these higher-order KdV equations commute. This property and the existence of infinite number of invariant integrals are consequences of the complete integrability of the higher-order KdV equations considered as infinite-dimensional Hamiltonian systems. The KdV equations can be studied group-theoretically as Hamiltonian systems on a certain coadjoint orbit in the †dual space of the †Lie algebra of a certain class of †pseudo-differential operators (M. Adler, *Inventiones Math.*, 50 (1979); also → B. Kostant, *Advances in Math.*, 34 (1979) for the analogous facts for the Toda lattice).

The ordinary differential equation $K_n[u] = 0$ is called a stationary KdV equation. By the commutativity of the flows $T_n(t)$, each KdV flow leaves invariant the space of solutions of $K_n[u] = 0$. The flows restricted to this space form a completely integrable Hamiltonian system with finitely many degrees of freedom (S. P. Novikov, *Functional Anal. Appl.*, 8 (1974)).

$K_n[u] = 0$ is also the condition that there exist an ordinary differential operator M which commutes with L. J. Burchnal and T. Chaundy (*Proc. London Math. Soc.*, (2) 21 (1922)) studied this problem and showed that such L and M are connected by the relation $M^2 = P(L)$, where P is a certain polynomial of degree $2n+1$ and that the potential u is expressed by the †theta function associated with the †hyperelliptic curve $w^2 = P(z)$.

D. Inverse Scattering Method

Let $u(x)$ be a potential such that $u(x) \to 0$ as $x \to \pm\infty$. The equation $Lf = \zeta^2 f$ $(\operatorname{Im}\zeta \geqslant 0)$ has solutions $f_\pm(x,\zeta)$ that can be represented as

$$f_\pm(x,\zeta) = e^{\pm i\zeta x}\left(1 + \int_0^\infty K_\pm(x,t)e^{\pm 2i\zeta t}\,dt\right).$$

Putting $\zeta = \xi + i\eta$ and noting that $f_+(x,\xi)$ and $f_+(x,-\xi)$ are independent solutions of $Lf = \xi^2 f$, one can express f_- as

$$f_-(x,\xi) = a(\xi)f_+(x,-\xi) + b(\xi)f_+(x,\xi).$$

The coefficient $a(\xi)$ can be continued analytically to the upper half-plane, where it has only a finite number of zeros, all of which are simple and lie on the imaginary axis. Denote them by $i\eta_j$ $(j = 1, \ldots, n)$. The †point spectrum of the operator L consists of the numbers $-\eta_j^2$, and the associated eigenfunctions are $f_+(x, i\eta_j)$, which are real-valued. Put

$$c_j = \left(\int f_+(x, i\eta_j)^2\,dx\right)^{-1},$$

and call $t(\xi) = a(\xi)^{-1}$ and $r(\xi) = b(\xi)/a(\xi)$ the **transmission coefficient** and the **reflection coefficient**, respectively. The triplet $r(\xi)$, η_j, c_j $(j = 1, \ldots, n)$ is called the **scattering data**. It is connected with the kernel $K = K_+$ by the **Gel'fand-Levitan-Marchenko equation**

$$K(x,t) + F(x+t) + \int_0^\infty F(x+t+s)K(x,s)\,ds \qquad (t>0),$$

$$F(x) = \pi^{-1}\int_{-\infty}^{\infty} r(\xi)e^{2i\xi x}\,dx + 2\sum_{j=1}^{n} c_j e^{-2\eta_j x}.$$

The potential is given by

$$u(x) = -(\partial K/\partial x)(x, 0).$$

When the reflection coefficient $r(\xi)$ vanishes identically, the potential is called a **reflectionless potential**. The kernel K then becomes a †degenerate kernel and the potential is expressed by

$$u(x) = -2\frac{d^2}{dx^2}\log D(x),$$

where $D(x)$ is the determinant of the $n \times n$ matrix whose j, k entry is $\delta_{jk} + c_j\exp\{-(\eta_j + \eta_k)x\}/(\eta_j + \eta_k)$.

The authors of [1] showed that if $u(x,t)$ is a solution of the KdV equation and if $u(x,t) \to 0$ $(x \to \pm\infty)$, then the time development of the scattering data of the potential $u(x,t)$ is as follows: n and η_j do not depend on t, and

$$c_j(t) = c_j e^{8\eta_j^3 t}, \qquad r(\xi, t) = r(\xi)e^{8i\xi^3 t}.$$

The solution associated with the reflectionless potentials are obtained by replacing c_j by $c_j(t)$ in the formula for $D(x)$. These are soliton solutions of the KdV equation.

R. Hirota developed a method of treating functions like $D(x)$ directly for most of the equations in the soliton theory (*Lecture notes in math.* 515, Springer, 1976).

A certain geometric method that enables one to obtain solutions of the Sine-Gordon equation from a known solution has been studied in the transformation theory of surfaces of constant negative curvature (G. Darboux, *Leçons sur la théorie générale des surfaces*, Chelsea, 1972, vol. 3, ch. 12), and its generalizations are called Bäcklund transfor-

mations. Soliton solutions of the KdV equation can also be constructed by this method. Relations to the inverse scattering method, differential systems, and transformation groups have also been studied (*Lecture notes in math.* *515*, Springer, 1976).

E. Periodic Problem

Let the potential $u(x)$ be of period l, and consider Hill's equation (→ 268 Mathieu Functions E) $Lf=\lambda f$. The real (λ-) axis is divided into intervals of unstable solutions (u-intervals) alternating with intervals of stable solutions. One of the u-intervals is of the form $(-\infty,\lambda_0]$ and the others are finite, possibly degenerating to points. The special potentials that have only a finite number of nondegenerate u-intervals are the periodic analog of the reflectionless potentials, and are called the **finite-gap** (or **-band**) **potentials**.

Consider the eigenvalue problem $Lf=\lambda f$, $f(\tau)=f(\tau+l)=0$ for a fixed real parameter τ. Exactly one of its eigenvalues belongs to each finite u-interval. These eigenvalues move with τ, but those in degenerate intervals cannot move.

Let u be a finite-gap potential, $I_j=[\lambda_{2j-1},\lambda_{2j}]$ $(j=1,\dots,g)$ be the nondegenerate (finite) u-intervals, and $\mu_j(\tau)$ be the associated eigenvalues in I_j. The potential is recovered by the formula

$$u(x)=\sum_{j=0}^{2g}\lambda_j-2\sum_{j=1}^{g}\mu_j(x).$$

Put $P(\lambda)=\prod_{j=0}^{2g}(\lambda-\lambda_j)$, realize the †Riemann surface S of the hyperelliptic curve $w^2=P(z)$ as a two-sheeted cover of the †Riemann sphere, and consider the $\mu_j(\tau)$ as points on S. Let ω_j $(j=1,\dots,g)$ be a basis for the space of the †differentials of the first kind on S. Fix a point P_0 in S and put

$$w_j(\tau)=\sum_{k=1}^{g}\int_{P_0}^{\mu_k(\tau)}\omega_j\qquad(j=1,\dots,g)$$

(→ 3 Abelian Varieties L). Then the locus $(w_1(\tau),\dots,w_g(\tau))(-\infty<\tau<\infty)$ on the †Jacobian variety turns out to be a straight line, the direction v_x being determined by the †periods of certain †differentials of the second kind on S. Employing the solution of †Jacobi's inverse problem, we can write the potential in terms of the †Riemann theta function as

$$u(x)=-2\frac{d^2}{dx^2}\log\theta(xv_x+c)+C,\qquad(3)$$

where c is a certain constant vector and C is a constant.

Suppose now that $u(x,t)$ is a solution of the KdV equation which is a finite-gap potential for each t. Then g and λ_j are preserved in time. The determination of the direction v_t on the Jacobian variety is similar to that of v_x, and the solutions of the KdV equation are obtained by replacing the vector c by tv_t+c' in (3). The case $g=1$ is the elliptic traveling wave solution (→ Section B). Most of these results have been extended to the general periodic problem (H. P. McKean and E. Trubowitz, *Comm. Pure Appl. Math.*, 29 (1976)).

F. Two-Dimensional KdV Equation

Let S be a compact †Riemann surface of †genus g, and let p_∞ be a fixed point on S. Put $F(\kappa)=a_n\kappa^n+\dots+a_0$ and $G(\kappa)=b_m\kappa^m+\dots+b_0$. A function $\psi(x,y,t,p)$ of $p\in S$ and of x, y, t is uniquely determined by the following conditions: (a) it is meromorphic on $S-\{p_\infty\}$, and its pole †divisor is a general divisor of degree g and does not depend on x, y, t; (b) for a †local parameter z at p_∞ $(z(p_\infty)=0)$ and $\kappa=z^{-1}$, $\psi_0=\psi\exp(-\kappa x-F(\kappa)y-G(\kappa)t)$ is holomorphic near p_∞, and $\psi_0(p_\infty)=1$.

Moreover, there is a differential operator $L=a_nD^n+a_{n-1}D^{n-1}+\sum_{j=0}^{n-2}u_j(x,y,t)D^j$ such that $\psi_y=L\psi$. Expanding ψ_0 at p_∞ as $1+\sum_{j=1}\xi_j(x,y,t)z^j$, one can express the coefficients u_j by $\xi_1,\xi_2,\dots$. Analogously, there is an $M=b_mD^m+b_{m-1}D^{m-1}+\sum_{j=0}^{m-2}v_jD^j$ such that $\psi_t=M\psi$. The operators L and M satisfy the relation

$$L_t-M_y=[L,M],\qquad(4)$$

which is a generalization of the Lax representation. The coefficients of L and M satisfy a certain system of nonlinear differential equations (V. E. Zakharov and A. B. Shabat, *Functional Anal. Appl.*, 8 (1974); I. M. Krichever, *Functional Anal. Appl.*, 11 (1977)).

Example. Let $F(\kappa)=\kappa^2$ and $G(\kappa)=\kappa^3+c\kappa$. Then one finds

$$L=D^2+u,\qquad M=D^3+(3u/2+c)D+v,$$

where $u=-2\xi_1'$ and $v=3\xi_1\xi_1'-3\xi_1''-3\xi_2'$. Eliminating v from (4), one has

$$3u_{yy}+(-4u_t+4cu_x+u_{xxx}+6uu_x)_x=0,$$

the so-called **two-dimensional KdV equation (Kadomtsev-Petvyashvili equation)**. If $u(x,y,t)$ does not depend on y, the equation reduces to the KdV equation, and if u does not depend on t, to the **Boussinesq equation**.

The condition for reduction to these special cases can be described in terms of the meromorphic functions admitted by S. Suppose that there is a meromorphic function $E_F(p)$ holomorphic for $p\neq p_\infty$ and of †principal part $F(\kappa)$ at $p=p_\infty$. Then ψ is written as $\varphi(x,t,p)\exp\{E_F(p)y\}$, and the coeffi-

cients of L and M do not depend on y. φ satisfies $L\varphi=E\varphi$, and (4) reduces to the Lax representation.

If such an E_F exists, S is hyperelliptic, and p_∞ is one of its †branch points over the Riemann sphere. Thus the result of the previous section is recovered.

G. Solvable Models in Field Theory

The Sine-Gordon equation has been studied extensively as a solvable model in †field theory. It is a special case of a field in two space-time dimensions with values in a †symmetric space; this can also be treated by a variant of the inverse scattering method (V. E. Zakharov and A. V. Mikhailov, *Sov. Phys. JETP*, 47 (1978)).

Much work has been done on the semiclassical †quantization of equations encountered in soliton theory. Recently, a method of exact quantization (called quantum inverse scattering) was developed (see, for example, L. A. Takhtadzhyan and L. D. Faddeev, *Russian Math. Surveys*, 34 (5) (1979)).

References

[1] C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Method for solving the Korteweg-de Vries equation, Phys. Rev. Lett., 19 (1967), 1095–1097.
[2] P. D. Lax, Integrals of nonlinear equations of evolution and solitary waves, Comm. Pure Appl. Math., 21 (1968), 467–490.
[3] B. A. Dubrovin, V. B. Matveev, and S. P. Novikov, Nonlinear equations of Korteweg-de Vries type, finite-band operators and Abelian varieties, Russian Math. Surveys 31 (1) (1976), 59–146. (Original in Russian, 1976.)
[4] H. P. McKean, Integrable systems and algebraic curves, Lecture notes in math. 755, Springer, 1979, 83–200.
[5] Yu. I. Manin, Algebraic aspects of nonlinear differential equations, J. Sov. Math., 11 (1) (1979), 1–122. (Original in Russian, 1978.)

388 (XIII.32) Special Functional Equations

A. General Remarks

The term **special functional equations** usually means functional equations that do not involve limit operations. Such functional equations appear in various fields, but there is no systematic method for solving them. Usually they are solved by reduction to functional equations of some standard type. In this article functions are all real-valued functions of real variables unless otherwise specified.

B. Additive Functional Equations and Related Equations

Suppose that we are given an equation

$$f(x+y)=f(x)+f(y). \tag{1}$$

Clearly, $f(x)=cx$ (c a constant) is a solution. If $f(x)$ is continuous, (1) has no other solution (Cauchy). The same conclusion holds under any one of the following weaker conditions: (i) $f(x)$ is continuous at a point; (ii) $f(x)$ is bounded in a neighborhood of a point; (iii) $f(x)$ is †measurable in a neighborhood of a point. However, it was shown by G. Hamel and H. Lebesgue by means of †transfinite induction that equation (1) has infinitely many nonmeasurable solutions. On the other hand, it was proved by A. Ostrowski [2] that if a solution $f(x)$ of equation (1) does not take any value between two distinct numbers for x on a set of positive measure, then $f(x)$ is continuous. This result can be extended to the case where x is a point $(x_1,\dots,x_n)$ of an n-dimensional Euclidean space. In this case, any continuous solution is of the form

$$f(x)=\sum_{j=1}^{n} C_j x_j$$

(C_j constant).

Next, we consider the equation

$$f\left(\frac{x+y}{2}\right)=\frac{f(x)+f(y)}{2}. \tag{2}$$

Any solution of equation (1) satisfies this equation. When x is a point $(x_1,\dots,x_n)$ of an n-dimensional Euclidean space, any continuous solution of (2) is of the form $f(x)=\sum_{j=1}^{n} C_j x_j+C_0$ (C_j constant). If a solution $f(x)$ of (2) defined on a †convex set K does not take any value between two distinct numbers for x on a set of positive measure, then $f(x)$ is continuous (M. Hukuhara).

Consider the equation

$$g(x+y)=g(x)g(y). \tag{3}$$

If a solution $g(x)$ vanishes at some point ξ, then $g(x)\equiv 0$. Excluding this trivial case, we assume that $g(x)$ never vanishes. Then, putting $y=x$, we see that $g(x)>0$ for all x. The substitution $f(x)=\log g(x)$ then reduces equation (3) to equation (1). Thus we see that any continuous solution of (3) is of the form $g(x)=\exp(cx)$.

Next, we consider the equation

$$g(uv)=g(u)g(v). \tag{4}$$

If a solution $g(u)$ vanishes at some $\xi \neq 0$, then $g(u) \equiv 0$. Excluding this case, we assume that $g(u) \neq 0$ for $u \neq 0$. For $u, v > 0$, by the substitution $x = \log u$, $y = \log v$, we have an equation of the form (3). On the other hand, putting $v = -1$, we have $g(-u) = g(-1)g(u)$. Since $g^2(-1) = g(1) = 1$, we see that any continuous solution of (4) is of the form $|u|^c$ or $(\operatorname{sgn} u)|u|^c$ according as $g(-1) = 1$ or -1.

C. The General Addition Theorem and Related Functional Equations

The **general addition theorem** is: If the equation

$$f(x+y) = F(f(x), f(y)) \tag{5}$$

has a continuous nonconstant solution $f(x)$ on $-\infty < x < +\infty$, then $f(x)$ is strictly monotone, and $F(u, v)$ is strictly monotone increasing and continuous with respect to u and v for $\alpha < u$, $v < \beta$ and satisfies $\alpha < F(u, v) < \beta$ for $\alpha < u$, $v < \beta$. There is also a constant c satisfying $F(c, c) = c$, and the identity $F(F(u, v), w) = F(u, F(v, w))$ holds for any u, v, w in the interval (α, β). Conversely, if $F(u, v)$ is such a function, then (5) has a continuous nonconstant solution on $-\infty < x < +\infty$. Let $f(x)$ be such a solution. Then any other continuous solution is given by $f(cx)$. When $F(u, v)$ is continuously differentiable, a continuous solution $f(x)$ of (5) can be obtained as a solution of the differential equation

$$f'(x) = F_v(f(x), a)c, \qquad c = f'(0),$$

satisfying the initial condition $f(0) = a$.

Consider the equation

$$F(f(x), f(y), f(x+y)) = 0.$$

Suppose that $F(u, v, w)$ is a polynomial in u, v, and w. If this equation has a †meromorphic solution $f(x)$, then $f(x)$ must be a rational function, a rational function of $\exp cx$, or an elliptic function [1, p. 64].

Next, consider the equation

$$f(x+y) + f(x-y) = 2(f(x) + f(y)). \tag{6}$$

Any solution continuous on $-\infty < x < \infty$ is of the form $f(x) = cx^2$. When x is a point $(x_1, \ldots, x_n)$ of an n-dimensional Euclidean space, any continuous solution of (6) is given by a quadratic form $f(x) = \sum_{i,j}^n c_{ij} x_i x_j$.

Consider the equation

$$f(x+y) + f(x-y) = 2f(x)f(y). \tag{7}$$

Any solution continuous on $-\infty < x < \infty$ is of the form $f(x) = \cosh cx = (e^{cx} + e^{-cx})/2$ or $f(x) = \cos cx$. If $f(x)$ is allowed to take complex values, then any continuous solution can be written in the form $f(x) = (e^{bx} + e^{-bx})/2$ in terms of a complex number b. A special case is $\cos x$, since b may take purely imaginary values.

D. Schröder's Functional Equation

Schröder's functional equation is

$$f(\theta(x)) = cf(x), \tag{8}$$

where $\theta(x)$ is a given function and c is a constant. A general solution of (8) can be written as $f(x) = f_1(x)\varphi(x)$, where $f_1(x) \neq 0$ is a particular solution of (8) and $\varphi(x)$ is a general solution of the equation $\varphi(\theta(x)) = \varphi(x)$. Suppose that there is a point a such that $\theta(a) = a$, and $\theta(x)$ and $f(x)$ are both differentiable in a neighborhood of $x = a$. Then we have $f'(a) = 0$ or $\theta'(a) = c$. Consider the case where $\theta'(a) = c$, and suppose that $\theta(x)$ is twice differentiable at $x = a$. When $|c| = |\theta'(a)| < 1$, define $\theta_n(x)$ ($n = 0, 1, 2, \ldots$) by $\theta_0(x) = x$ and $\theta_n(x) = \theta(\theta_{n-1}(x))$. Then the sequence $\{(\theta_n(x-a)c^{-n}\}$ ($n = 0, 1, 2, \ldots$) converges uniformly in a neighborhood of $x = a$, and its limit function $f(x)$ is a solution of (8). When $|c| = |\theta'(a)| > 1$, put $\theta(x) = u$. Then we have the equation $f(\theta^{-1}(u)) = c^{-1}f(u)$, and the problem reduces to the previous case. The results obtained for equation (8) can be extended to the following system of equations:

$$f_j(\theta_1(x), \theta_2(x), \ldots, \theta_n(x)) = \lambda_j f_j(x) + \delta_j f_{j-1}(x) + g_j(x), \quad j = 1, 2, \ldots, n, \tag{9}$$

where the $\theta_j(x)$ are given functions holomorphic in a neighborhood of $x = 0$, $\theta_j(0) = 0$, the coefficients of f in the right-hand side of (9) are numbers such that the matrix $A = (a_{ij})$ ($a_{ij} = 0$ except for $a_{jj} = \lambda_j$, $a_{j-1,j} = \delta_j$) is the †Jordan canonical form of the matrix formed by $\{(\partial\theta_j/\partial x_k)_{x=0}\}$, and the $g_j(x)$ are polynomials consisting of terms of the form constant $\times$ $x_1^{m_1} x_2^{m_2} \ldots x_n^{m_n}$ with exponents $m_1, m_2, \ldots, m_n$ for which $\lambda_j = \lambda_1^{m_1} \ldots \lambda_n^{m_n}$. If $0 < |\lambda_j| < 1$ ($j = 1, 2, \ldots, n$), then we can always choose the coefficients of the polynomials $g_j(x)$ so that equation (9) has a solution $\{f_j(x)\}$ holomorphic in a neighborhood of $x = 0$. The same conclusion can be obtained for $|\lambda_j| > 1$ ($j = 1, \ldots, n$).

Consider **Abel's functional equation**

$$f(\theta(x)) = f(x) + 1. \tag{10}$$

If we put $\exp f(x) = \varphi(x)$, then we have Schröder's functional equation

$$\varphi(\theta(x)) = e\varphi(x).$$

Consider the equation

$$f(x+1) = A(x)f(x). \tag{11}$$

If we put $\varphi(x) = \log f(x)$, then we have a †linear difference equation of the form

$$\varphi(x+1) - \varphi(x) = \log A(x).$$

References

[1] E. C. Picard, Leçons sur quelques équations fonctionnelles, Gauthier-Villars, 1928.
[2] A. Ostrowski, Mathematische Miszellen XIV, Uber die Funktionalgleichung der Exponentialfunktion und verwandte Funktionalgleichungenm, Jber. Deutsch. Math. Verein., 38 (1929), 54–62.
[3] F. Hausdorff, Mengenlehre, de Gruyter, second edition, 1927, 175; English translation, Set theory, Chelsea, 1962.
[4] J. Aczél, Lectures on functional equations and their applications, Academic Press, 1966. (Original in German, 1961.)

389 (XIV.1) Special Functions

A. Special Functions

The term **special functions** usually refers to the classes of functions listed in (1)–(4) (other terms, such as **higher transcendental functions**, are sometimes used). (1) The †gamma function and related functions (→ 174 Gamma Function); (2) †Fresnel's integral, the †error function, the †logarithmic integral, and other functions that can be expressed as indefinite integrals of elementary functions (→ 167 Functions of Confluent Type D); (3) †elliptic functions (→ 134 Elliptic Functions); (4) solutions of †linear ordinary differential equations of the second order derived by the method of separation of variables in certain partial differential equations, e.g., †Laplace's equation, in various †curvilinear coordinates. Recently, new types of special functions, such as †Painlevé's, have been introduced as the solutions of special differential equations.

In this article we discuss class (4); for the other classes, see the articles quoted. Class (4) is further divided into the following three types, according to the character of the †singular points of the differential equations of which the functions are solutions. Equations with a smaller number of singular points than those indicated in (1)–(3) below can be integrated in terms of elementary functions.

(1) **Special functions of hypergeometric type** are solutions of differential equations with three †regular singular points on the Riemann sphere. Examples are the †hypergeometric function and the †Legendre function. Any function of this type reduces to a hypergeometric function through a simple transformation (→ 206 Hypergeometric Functions; 393 Spherical Functions).

(2) **Special functions of confluent type** are solutions of differential equations that are derived from †hypergeometric differential equations by the confluence of two regular singular points, that is, by making one of the regular singular points tend to the other one so that the resulting singularity is an †irregular singular point of class 1 (→ 167 Functions of Confluent Type). Any function of this type can be expressed by means of †Whittaker functions, of which many important special functions, such as †Bessel functions, are special cases (→ 39 Bessel Functions). Also, one can reduce to this type the †parabolic cylindrical functions, that is, the solutions of differential equations with only one singular point which is at infinity and is irregular of class 2.

(3) **Special functions of ellipsoidal type** are solutions of differential equations with four or five regular singular points, some of which may be confluent to become irregular singular points. Examples are †Lamé functions, †Mathieu functions, and †spheroidal wave functions (→ 133 Ellipsoidal Harmonics; 268 Mathieu Functions). In contrast to types (1) and (2), functions of type (3) are difficult to characterize by means of †difference-differential equations and have not been fully explored. Sometimes the term *special function* in the strict sense is not applied to them. To specify the special functions of types (1) and (2), the term *classical special functions* has been proposed.

B. Unified Theories of Special Functions

Though many special functions were introduced separately to solve practical problems, several unified theories have been proposed. The classification in Section A based on differential equations may be regarded also as a kind of unified theory. Other examples are:

(1) Expression by †Barnes's extended hypergeometric function or its extension to the case of several variables by means of a definite integral of the form

$$\int(\zeta-a_1)^{b_1}(\zeta-a_2)^{b_2}\ldots(\zeta-a_m)^{b_m}(\zeta-z)^c\,d\zeta$$

(→ 206 Hypergeometric Functions).

(2) A unified theory [14] that includes the gamma function and is based upon Truesdell's difference-differential equation

$$\partial F(z,\alpha)/\partial z=F(z,\alpha+1).$$

(3) Unification from the standpoint of expansions in terms of †zonal spherical functions of a differential operator (the Laplacian) invariant under a transitive group of motions on a †symmetric Riemannian manifold (→ 437

Unitary Representations). With this approach a great variety of formulas can be derived in a unified way [3, 4].

References

[1] J. Dieudonné, Special functions and linear representations of Lie groups, Expository lectures from the CBMS regional conference held at East Carolina Univ., March 1979, Amer. Math. Soc., 1980.
[2] B. C. Carlson, Special functions of applied mathematics, Academic Press, 1977.
[3] J. D. Talman, Special functions: A group-theoretic approach, Benjamin, 1968.
[4] N. Ya. Vilenkin, Special functions and the theory of group representations, Amer. Math. Soc. Transl. of Math. Monographs 22, 1968. (Original in Russian, 1965.)
[5] W. Miller, Jr., Lie theory and special functions, Academic Press, 1968.
[6] Y. L. Luke, The special functions and their approximations I, II, Academic Press, 1969.
[7] S. Moriguti, K. Udagawa, and S. Hitotumatu, Mathematical formulas, Special functions (in Japanese), Iwanami, 1960.
[8] E. T. Whittaker and G. N. Watson, A course of modern analysis, Cambridge Univ. Press, 1902; fourth edition, 1958.
[9] E. Jahnke and F. Emde, Funktionentafeln mit Formalen und Kurven, Teubner, 1933; bilingual edition, Tables of functions with formulas and curves, Dover, 1945.
[10] W. Magnus and F. Oberhettinger, Formulas and theorems for the special functions of mathematical physics, Springer, third enlarged edition, 1966.
[11] R. Courant and D. Hilbert, Methods of mathematical physics I, II, Interscience, 1953, 1962.
[12] A. Erdélyi (ed.), Higher transcendental functions I–III, McGraw-Hill, 1953–1955.
[13] J. Meixner, Spezielle Funktionen der mathematischen Physik, Handbuch der Physik II, Springer, second edition, 1933.
[14] C. A. Truesdell, An essay toward a unified theory of special functions based upon the functional equation $\partial F(z,\alpha)/\partial z = F(z,\alpha+1)$, Ann. Math. Studies, Princeton Univ. Press, 1948.
[15] I. N. Sneddon, Special functions of mathematical physics and chemistry, Oliver & Boyd, 1956.
[16] F. W. Schäfke, Einführung in die Theorie der spezielle Funktionen der mathematischen Physik, Springer, 1963.
[17] G. Szegö, Orthogonal polynomials, Amer. Math. Soc. Colloq. Publ., revised edition, 1959.
[18] E. D. Rainville, Special functions, Macmillan, 1960; Chelsea, second edition, 1971.

390 (XII.12) Spectral Analysis of Operators

A. General Remarks

Throughout this article, X stands for a †complex linear space and A for a †linear operator in X. Except when X is finite-dimensional, A need not be defined over all X. A linear operator A in X is by definition a linear mapping whose †domain $D(A)$ and †range $R(A)$ are linear subspaces of X. A complex number λ is said to be an **eigenvalue (proper value** or **characteristic value)** of A if there exists an $x \in D(A)$ such that $Ax = \lambda x$, $x \neq 0$. Any such x is called an **eigenvector (eigenelement, proper vector, characteristic vector)** associated with λ. When X is a †function space, the word **eigenfunction** is also used. For an eigenvalue λ of A, the subspace $M(\lambda)$ of X given by

$$M(\lambda) = M(\lambda; A) = \{x \mid Ax = \lambda x\},$$

i.e., the subspace consisting of 0 and all eigenvectors associated with λ, is called the **eigenspace** associated with λ, and the number $m(\lambda) = \dim M(\lambda)$ is called the **geometric multiplicity** of λ. The eigenvalue λ is said to be **(geometrically) simple** or **degenerate** according as $m(\lambda) = 1$ or $m(\lambda) \geqslant 2$. The problem of seeking eigenvalues and eigenvectors is referred to as the **eigenvalue problem**.

When X is a †topological linear space, the notion of eigenvalues leads to a more general object called the spectrum of A. Let λ be a complex number and put $A_\lambda = \lambda I - A$, where I is the †identity operator in X. Furthermore, put $R_\lambda = (A_\lambda)^{-1} = (\lambda I - A)^{-1}$, if the inverse exists. Then the **resolvent set** $\rho(A)$ of A is defined to be the set of all λ such that R_λ exists, has domain †dense in X, and is continuous. The **spectrum** $\sigma(A)$ of A is, by definition, the complement of $\rho(A)$ in the complex plane, and it is divided into three mutually disjoint sets: the **point spectrum** $\sigma_P(A)$, the **continuous spectrum** $\sigma_C(A)$, and the **residual spectrum** $\sigma_R(A)$. These are defined as follows: $\sigma_P(A) = \{\lambda \mid R_\lambda$ does not exist$\} = \{\lambda \mid \lambda$ is an eigenvalue of $A\}$; $\sigma_C(A) = \{\lambda \mid R_\lambda$ exists and has domain dense in X, but is not continuous$\}$; $\sigma_R(A) = \{\lambda \mid R_\lambda$ exists, but its domain is not dense in $X\}$.

Let X be a †Banach space and $\mathbf{B}(X)$ the set of all †bounded linear operators with domain X. If A is a †closed operator in X, then $\lambda \in \rho(A)$ if and only if $R_\lambda \in \mathbf{B}(X)$. Moreover, $\sigma(A)$ is a closed set. In particular, if $A \in \mathbf{B}(X)$, then $\sigma(A)$ is a nonempty compact set. In this case the **spectral radius** $r_\sigma(A)$ is defined as $r_\sigma(A) = \sup_{\lambda \in \sigma(A)} |\lambda|$. Then $r_\sigma(A) \leqslant \|A^n\|^{1/n}$, $n = 1, 2, \ldots$, and $\|A^n\|^{1/n} \to r_\sigma(A)$, $n \to \infty$.

In many problems of analysis crucial roles have been played by methods involving the spectrum and other related concepts. This branch of analysis is called **spectral analysis.** For an infinite-dimensional X the theory is well developed when X is a †Hilbert space and A is †self-adjoint or †normal.

B. Eigenvalue Problems for Matrices

Throughout this section let X be an N-dimensional complex linear space ($N<\infty$) and A a linear operator in X. (We assume that A is defined over all X.) With respect to a fixed basis $(\psi_1, \ldots, \psi_N)$ of X, the operator A is represented by an $N \times N$ matrix, also denoted by A. Then the eigenvalues of A coincide with the roots of the †characteristic equation $\det(\lambda I - A)=0$. There are no points of the spectrum other than eigenvalues, that is $\sigma(A)=\sigma_P(A)$. Let $\lambda\in\sigma(A)$. The multiplicity $\tilde{m}(\lambda)$ of λ as a root of the characteristic equation is called the **(algebraic) multiplicity** of the eigenvalue λ. The sum of $\tilde{m}(\lambda)$ over all the eigenvalues of A is equal to N. The eigenvalue λ is said to be **(algebraically) simple** or **degenerate** according as $\tilde{m}(\lambda)=1$ or $\tilde{m}(\lambda)\geqslant 2$. Let $\nu=1,2,\ldots$, and $N_\nu(\lambda)=\{x \mid (\lambda I - A)^\nu x=0\}$. Then $\{N_\nu(\lambda)\}$ forms a nondecreasing chain of subspaces $M(\lambda)=N_1(\lambda)\subset N_2(\lambda)\subset \ldots$, which ceases to increase after a finite number of steps. When $\nu\geqslant\tilde{m}(\lambda)$, the space $N_\nu(\lambda)$ is equal to a fixed subspace $\tilde{M}(\lambda)$, sometimes called the **root subspace** (or **generalized eigenspace** or **principal subspace**) of A associated with λ. A vector in the root subspace is called a **root vector** (or a **generalized eigenvector**). Then $\dim \tilde{M}(\lambda)=\tilde{m}(\lambda)$ and hence $m(\lambda)\leqslant\tilde{m}(\lambda)$. When A is a †normal matrix, $M(\lambda)=\tilde{M}(\lambda)$ and $m(\lambda)=\tilde{m}(\lambda)$.

If two matrices A and B are †similar, i.e., if there exists an invertible matrix P such that $B=P^{-1}AP$, then A and B have the same eigenvalues with the same algebraic (and geometric) multiplicities. The same conclusion holds for A and A', where A' is the †transpose of A. For the adjoint matrix A^* we have $\sigma(A^*)=\overline{\sigma(A)}\equiv\{\bar{\lambda} \mid \lambda\in\sigma(A)\}$. For an arbitrary polynomial f the relation $\sigma(f(A))=f(\sigma(A))\equiv\{f(\lambda) \mid \lambda\in\sigma(A)\}$ holds **(Frobenius's theorem)**. These relations can be extended to operators in a Banach space. In particular, $\sigma(f(A))=f(\sigma(A))$ if A is a bounded operator and f is a function holomorphic in a neighborhood of $\sigma(A)$ (for the †spectral mapping theorem → 251 Linear Operators).

In the next four paragraphs, in which the spectral properties of †normal or †Hermitian matrices is discussed, we introduce into X the Euclidean †inner product (,), regarding X as a space of N-tuples of scalars. Let A be an $N \times N$ normal matrix. Then the eigenspaces associated with different eigenvalues of A are mutually orthogonal. Moreover, the eigenspaces of A as a whole span the entire space X. One can therefore choose a †basis of X formed by a †complete orthonormal set of eigenvectors of A. Specifically, there exists a basis $\{\varphi_j \mid j=1,\ldots,N\}$ of X such that $A\varphi_j=\mu_j\varphi_j$ and $(\varphi_i,\varphi_j)=\delta_{ij}$, where δ_{ij} is the †Kronecker delta. Moreover, $\mu_1,\ldots,\mu_N$ exhaust all the eigenvalues of A. In terms of the basis $\{\varphi_j\}$, an arbitrary $x\in X$ can be expanded as

$$Ax=\sum_{j=1}^{N}\mu_j(x,\varphi_j)\varphi_j=\sum_{\lambda\in\sigma(A)}\lambda P_\lambda x, \tag{1}$$

where P_λ is the orthogonal †projection on the eigenspace associated with the eigenvalue λ.

Of particular importance among normal matrices are Hermitian matrices and †unitary matrices. The eigenvalues of a Hermitian matrix are real, and those of a unitary matrix have the absolute value 1.

Solving the eigenvalue problem of a normal matrix A leads immediately to the diagonalization of A. For instance, let U be the $N \times N$ matrix whose jth column is equal to φ_j. Here the basis $\{\varphi_j\}$ is as before, and each φ_j is regarded as a column vector. Then U is unitary, and the transform U^*AU of A by U is the diagonal matrix whose diagonal entries are the μ_j. The problem of transforming a †Hermitian form to its canonical form can also be solved by means of U. In fact, a Hermitian form $Q=Q(x)$ on X is expressed as $Q(x)=(Ax,x)$ with a Hermitian matrix A. For this A, construct U as before. Then by the transformation $x=Uy$ of the coordinates of X, the form Q is converted to its canonical form $Q=\mu_1|y_1|^2+\ldots+\mu_N|y_N|^2$. When X is a real linear space and A is a real symmetric matrix, U is an orthogonal matrix. By means of the orthogonal transformation $x=Uy$, the surface of the second order $Q(x)=1$ in $\mathbf{R}^N$ is converted to the form $\mu_1 y_1^2+\ldots+\mu_N y_N^2=1$. The orthogonal transformation $x=Uy$ is called the **transformation to principal axes** of the surface $Q(x)=1$.

When A is not normal, it can be transformed into †Jordan's canonical form by a basis φ_j taken from the root subspaces $\tilde{M}(\lambda)$. However, φ_j need not be orthonormal even when A is diagonalizable.

C. Spectral Analysis in Hilbert Spaces

Throughout the rest of this article except for the last section, X is assumed to be a Hilbert space with inner product (,). Furthermore, the most complete discussions will be confined to †normal or †self-adjoint operators. A funda-

mental theorem in spectral analysis for such operators is the spectral theorem, which asserts that a representation such as (1) holds in a generalized form. When the operator is †compact, we have only to replace the sum by an infinite sum (→ 68 Compact and Nuclear Operators). In the general case we need a kind of integral. This is discussed in detail in Sections D and E. The general theory of spectral analysis for nonnormal operators, however, is rather involved even in Hilbert spaces, but two important developments can be noted. One is the theory of Volterra operators, and the other is the theory of essentially normal operators. The former is discussed in Section H and the latter and its related results in Sections I and J.

D. Spectral Measure

Let $\mathscr{B}$ be a †completely additive class of subsets of a set Ω, that is, $(\Omega, \mathscr{B})$ is a †measurable space. An operator-valued set function $E=E(\cdot)$ defined on $\mathscr{B}$ is said to be a (self-adjoint) **spectral measure** if (i) $E(M)$, $M\in\mathscr{B}$, is an †orthogonal projection in X; (ii) $E(\Omega)=I$; and (iii) E is †countably additive, that is,

$$E\left(\bigcup_{n=1}^{\infty} M_n\right)=\sum_{n=1}^{\infty} E(M_n)$$

(†strong convergence) for a disjoint sequence $\{M_n\}$ of subsets in $\mathscr{B}$. A spectral measure E satisfies $E(M\cap N)=E(M)E(N)=E(N)E(M)$, $M, N\in\mathscr{B}$. Spectral measures which are frequently used in spectral analysis are those defined on the family $\mathscr{B}_r$ ($\mathscr{B}_c$) of all †Borel sets in the field of real (complex) numbers $\mathbf{R}$ ($\mathbf{C}$). A spectral measure on $\mathscr{B}_r$ ($\mathscr{B}_c$) is sometimes referred to as a **real (complex) spectral measure**. For such a spectral measure E the **support** (or the **spectrum**) of E, denoted by $\Lambda(E)$, is defined to be the complement of the largest open set G for which $E(G)=0$. A complex spectral measure such that $\Lambda(E)\subset\mathbf{R}$ can be identified with a real spectral measure.

Let E be a spectral measure on $\mathscr{B}_r$, and put

$$E_\lambda=E((-\infty,\lambda]), \qquad -\infty<\lambda<\infty. \tag{2}$$

Then E_λ satisfies the relations

$$E_\lambda E_\mu=E_{\min(\lambda,\mu)}, \qquad \operatorname*{s-lim}_{\lambda\to\mu+0} E_\lambda=E_\mu,$$

$$\operatorname*{s-lim}_{\lambda\to-\infty} E_\lambda=0, \qquad \operatorname*{s-lim}_{\lambda\to+\infty} E_\lambda=I, \tag{3}$$

where s-lim stands for strong convergence. A family $\{E_\lambda\}_{\lambda\in\mathbf{R}}$ of orthogonal projections satisfying the relation (3) is called a **resolution of the identity**. Relation (2) gives a one-to-one correspondence between the resolutions of the identity and the spectral measures on $\mathscr{B}_r$.

Let E be a spectral measure on $\mathscr{B}_r$, and let $x, y\in X$. Then the set function $M\to(E(M)x,x)=\|E(M)x\|^2$ is a bounded regular †measure in the ordinary sense, and the set function $M\to(E(M)x,y)$ is a complex-valued regular †completely additive set function. For every complex Borel †measurable function f on $\mathbf{R}$, the operator $S(f)$ in X is defined by the relations

$$D(S(f))=\left\{x\,\middle|\,\int_{-\infty}^{\infty}|f(\lambda)|^2(E(d\lambda)x,x)<\infty\right\},$$

$$(S(f)x,y)=\int_{-\infty}^{\infty}f(\lambda)(E(d\lambda)x,y), \tag{4}$$

$$x\in D(S(f)), \qquad y\in X.$$

$S(f)$ is a densely defined closed operator and is denoted by $S(f)=\int_{-\infty}^{\infty}f(\lambda)E(d\lambda)$. The correspondence $f\mapsto S(f)$ satisfies formulas of the so-called operational calculus (→ 251 Linear Operators). In particular, $S(\bar{f})=S(f)^*$, and hence $S(f)$ is self-adjoint if f is real-valued. If f is bounded on the support of E, then $S(f)$ is everywhere defined in X and is bounded. $S(f)$ is sometimes called the **spectral integral** of f with respect to E. The operator $S(f)$ can be defined in a similar way for a spectral measure on $\mathscr{B}_c$ (and for a more general spectral measure).

E. Spectral Theorems

For every self-adjoint operator H in a Hilbert space X, there exists a unique real spectral measure E such that

$$H=\int_{-\infty}^{\infty}\lambda E(d\lambda). \tag{5}$$

In other words, H and E correspond to each other by the relations

$$D(H)=\left\{x\,\middle|\,\int_{-\infty}^{\infty}\lambda^2(E(d\lambda)x,x)<+\infty\right\},$$

$$(Hx,y)=\int_{-\infty}^{\infty}\lambda(E(d\lambda)x,y), \tag{6}$$

$$x\in D(H), \qquad y\in X.$$

This is the **spectral theorem** for self-adjoint operators. The support of E is equal to the spectrum $\sigma(H)$, so that we can write

$$H=\int_{\sigma(H)}\lambda E(d\lambda)=\int_{-\infty}^{\infty}\lambda\chi_{\sigma(H)}(\lambda)E(d\lambda),$$

where χ_M stands for the †characteristic function of M. Formulas (5) and (6) are called the **spectral resolution** (or **spectral representation**) of the **self-adjoint operator** H. We call E the spectral measure for H, and the $\{E_\lambda\}$ corresponding to E by formula (2) (or sometimes E itself) the resolution of the identity for H. Let λ be a real number. Then $\lambda\in\sigma_P(H)$ if and only if $E(\{\lambda\})\neq0$. Also, $\lambda\in\sigma_C(H)$ if and only if $E(\{\lambda\})=0$ and $E(V)\neq0$ for any neighborhood

V of λ. The spectral measure E can be represented in terms of the resolvent $R(\alpha; H)=(\alpha I-H)^{-1}$ of H by the formula

$$E((a,b))=\lim_{\delta\downarrow 0}\lim_{\varepsilon\downarrow 0}\frac{1}{2\pi i}\int_{a+\delta}^{b-\delta}\{R(\mu-\varepsilon i; H)-R(\mu+\varepsilon i; H)\}\,d\mu$$

(strong convergence).

For every normal operator A in X, there exists a unique spectral measure E on the family of all complex Borel sets $\mathscr{B}_c$ such that

$$A=\int_{\mathbf{C}} zE(dz).$$

This is called the **complex spectral resolution** (or **complex spectral representation**) of the **normal operator** A. The support $\Lambda(E)$ is equal to $\sigma(A)$. There are characterizations of point and continuous spectra similar to the case of self-adjoint operators. Normal operators have no residual spectra. For a unitary operator U, the support of the associated spectral measure is contained in the unit circle Γ, so that U can be represented as

$$U=\int_{\Gamma} e^{i\theta}F(d\theta) \tag{7}$$

with a spectral measure F defined on Γ. Formula (7) is the **spectral resolution** of the **unitary operator** U.

For a self-adjoint operator $H=\int_{-\infty}^{\infty}\lambda E(d\lambda)$ the following two types of classification of $\sigma(H)$ are often useful.

(i) The **essential spectrum** $\sigma_e(H)$ is by definition the set $\sigma(H)$ minus all the isolated eigenvalues of H with finite multiplicity. When H is bounded this definition of the essential spectrum coincides with that to be given in Section I for a general $A\in B(X)$. $\sigma(H)\setminus\sigma_e(H)$ is called the **discrete spectrum** of H.

(ii) The set $X_{ac}(H)$ (resp. $X_s(H)$) (called the **space of absolute continuity** (resp. **singularity**) with respect to H) of all $u\in X$ such that the measure $(E(d\lambda)u,u)$ is absolutely continuous (resp. singular) with respect to the †Lebesgue measure is a closed subspace of X that reduces H. The restriction of H to X_{ac} (resp. $X_s(H)$) is called the absolutely continuous (resp. singular) part of H, and its spectrum, denoted by $\sigma_{ac}(H)$ (resp. $\sigma_s(H)$), is called the **absolutely continuous** (resp. **singular**) **spectrum** of H. Note that $\sigma_{ac}(H)$ and $\sigma_s(H)$ may not be disjoint.

F. Functions of a Self-Adjoint Operator

Let $H=\int_{-\infty}^{\infty}\lambda E(d\lambda)$ be a self-adjoint operator in X. For a complex-valued Borel measurable function f on $\mathbf{R}$, we define $f(H)$ to be the operator $S(f)$ determined by (4) in reference to the resolution of the identity E associated with H:

$$f(H)=\int_{-\infty}^{\infty} f(\lambda)E(d\lambda).$$

For an arbitrary $a\in X$, let $L_2(a)$ be the L_2-space over the measure $\mu_a=\mu_a(\cdot)=(E(\cdot)a,a)$. In other words, $f\in L_2(a)$ if and only if $a\in D(f(H))$. The correspondence $f\leftrightarrow f(H)a$ gives an isometric isomorphism between $L_2(a)$ and the subspace $M(a)=\{f(H)a \mid f\in L_2(a)\}$ of X. (In particular, $M(a)$ is closed.) H is reduced by $M(a)$, and the part of H in $M(a)$ corresponds to the multiplication $f(\lambda)\to\lambda f(\lambda)$ in $L_2(a)$.

For a given self-adjoint operator H there exists a (not necessarily countable) family $\{a_\theta\}_{\theta\in\Theta}$ of elements a_θ of X such that

$$X=\sum_{\theta\in\Theta} M(a_\theta), \tag{8}$$

where $\sum$ stands for the †direct sum of mutually orthogonal closed subspaces. Consequently, X is represented by the direct sum $\sum_{\theta\in\Theta}L_2(a_\theta)$ of L_2-spaces. If $x\in D(H)$ is represented by $\{f_\theta\}_{\theta\in\Theta}$ in this representation, Hx is represented by $\{\lambda f_\theta\}_{\theta\in\Theta}$.

G. Unitary Equivalence and Spectral Multiplicity

In this section X is assumed to be a †separable Hilbert space. Then (8) can be made more precise. Namely, for a self-adjoint operator H, we can find a countable family $\{a_n\}_{n=1}^{\infty}$ of elements of X such that

$$X=\sum_{n=1}^{\infty} M(a_n)\cong\sum_{n=1}^{\infty} L_2(a_n), \tag{9}$$

$$\mu_{a_{n+1}} \text{ is †absolutely continuous with respect to } \mu_{a_n}, \quad n=1,2,\ldots. \tag{10}$$

Furthermore, if $\{a_n'\}$ is another family satisfying (9) and (10), then μ_{a_n} and $\mu_{a_n'}$ are absolutely continuous with respect to each other (**Hellinger-Hahn theorem**). μ_{a_1} is said to be the **maximum spectral measure**.

Two operators H_1 and H_2 are said to be **unitarily equivalent** if there exists a unitary operator U such that $H_2=U^*H_1U$. A criterion for unitary equivalence of self-adjoint operators can be given in terms of the spectral representation given previously. Namely, let $\{a_n^{(i)}\}$, $i=1,2$, be a sequence satisfying (9) and (10) with respect to H_i. Then H_1 and H_2 are unitarily equivalent if and only if $\mu_{a_n^{(1)}}$ and $\mu_{a_n^{(2)}}$ are absolutely continuous with respect to each other for all $n=1,2,\ldots$.

A self-adjoint operator H is said to have a **simple spectrum** if there exists an $a\in X$ such that $M(a)=X$. Such an $a\in X$ is called a **generating element** of X with respect to H.

Self-adjoint operators with simple spectra are closely related to Jacobi matrices. Let H be such an operator with a generating element $a \in X$. Take a complete orthonormal set $\{G_n\}_{n=1}^{\infty}$ in $L_2(a)$ such that $G_n = G_n(\lambda)$ is a polynomial of degree $n-1$ and $\lambda G_n(\lambda) \in L_2(a)$. Then $\{g_n\}_{n=1}^{\infty}$, $g_n = G_n(H)a$, is a complete orthonormal set in X. The matrix representation $\{a_{mn}\}$, $a_{mn} = (Hg_n, g_m)$ of H with respect to the basis $\{g_n\}$ has the following properties: (i) $a_{mn} = 0$ if $|m-n| \geqslant 2$; (ii) $a_{n,n+1} = \overline{a_{n+1,n}} \neq 0$; (iii) a_{nn} is real. Any infinite matrix $\{a_{mn}\}$ satisfying (i), (ii), and (iii) is called a **Jacobi matrix**. A Jacobi matrix determines a †symmetric operator whose †deficiency index is either (0, 0) or (1, 1). Any self-adjoint extension has a simple spectrum. (For more details about Jacobi matrices and their applications → [8].)

H. Triangular Representation of Volterra Operators

A linear operator A in a Hilbert space X is called a †Volterra operator if it is †compact and †quasinilpotent (i.e., 0 is the only spectrum). The name is justified because under very general assumptions such an operator is unitarily equivalent to the integral operator of Volterra type in the vector-valued L_2 space on [0, 1]. Let $\mathfrak{P}$ be a maximal †totally ordered family of orthogonal projections in X such that PX is an †invariant subspace of a Volterra operator A for every $P \in \mathfrak{P}$. Such a family $\mathfrak{P}$ always exists and is called a maximal **eigenchain** of A. Then generalizing the triangular representation of nilpotent matrices, we have the integral representation

$$A = 2i \int_{\mathfrak{P}} PA_I \, dP,$$

where $A_I = (A - A^*)/(2i)$ is the imaginary part of A and the integral is the limit in norm of approximating sums of the form $\sum Q_i A_I (P_i - P_{i-1})$ for finite partitions $0 = P_0 < P_1 < \ldots < P_n = I$ of $\mathfrak{P}$ in which Q_i is an arbitrary projection in $\mathfrak{P}$ such that $P_{i-1} \leqslant Q_i \leqslant P_i$ (M. S. Brodskiĭ). Conversely, let $\mathfrak{P}$ be a totally ordered family of orthogonal projections that contains 0 and the identity. If the integral $A = \int_{\mathfrak{P}} PBdP$ converges in norm for a compact linear operator B, then A is a Volterra operator and $\mathfrak{P}$ is an eigenchain of A. If, moreover, B is self-adjoint, we have $B = A_I$ (I. C. Gokhberg and M. G. Kreĭn; Brodskiĭ). Furthermore, assume for simplicity that $\mathfrak{P}$ is continuous in the sense that for every $P_1 < P_2$ in $\mathfrak{P}$ there exists an element P in $\mathfrak{P}$ such that $P_1 < P < P_2$. If B is a †Hilbert-Schmidt operator, then the integral $A = \int_{\mathfrak{P}} PBdP$ converges in the †Hilbert-Schmidt norm and the mapping $B \mapsto A$ is an orthogonal projection to the set of all Volterra operators of Hilbert-Schmidt class possessing $\mathfrak{P}$ as an eigenchain (Gokhberg and Kreĭn). Volterra operators with the imaginary part of the †trace class are especially important for applications. In this case we have the following fundamental theorem on the density of the spectrum of the real part A_R of the Volterra operator A: Let $n_+(r; A_R)$ and $n_-(r; A_R)$ be the numbers of eigenvalues of A_R in the intervals $[1/r, \infty)$ and $(-\infty, -1/r]$, respectively.

Then $\lim_{r\to\infty} n_+(r; A_R)/r = \lim_{r\to\infty} n_-(r; A_R) = h/\pi$.

The number h is given by

$$h = \int_{\mathfrak{P}} \|dP \, A_I \, dP\|_1$$

$$= \inf \sum \|(P_i - P_{i-1}) A_I (P_i - P_{i-1})\|_1,$$

where the norm is the †trace norm and the infimum is taken over all finite partitions P_i of a maximal eigenchain $\mathfrak{P}$ for A. We refer for further details and applications to the books by Gokhberg and Kreĭn [9, 10].

I. Fredholm Operators and Essential Spectra of Operators

Throughout Sections I and J we assume that X is a separable infinite-dimensional complex Hilbert space, and we consider only bounded linear operators in X. The set $\mathbf{B}^{(c)}(X)$ of all †compact linear operators in X is a †maximal two-sided ideal of the †C^*-algebra $\mathbf{B}(X)$ of all bounded linear operators. The simple quotient C^*-algebra $\mathbf{A}(X) = \mathbf{B}(X)/\mathbf{B}^{(c)}(X)$ is called the **Calkin algebra**. We denote the quotient mapping by $\pi: \mathbf{B}(X) \to \mathbf{A}(X)$. Then an operator $A \in \mathbf{B}(X)$ is a †Fredholm operator if and only if its image $\pi(A)$ is an invertible element of $\mathbf{A}(X)$. Let $\mathbf{F}(X)$ be the set of all Fredholm operators in X, and let $\mathbf{F}_n(X)$, $n \in \mathbf{Z}$, be its subset of all operators of †index n. $\mathbf{F}_n(X)$ is a connected set in $\mathbf{B}(X)$, and in particular, $\mathbf{F}_0(X)$ is the inverse image of the connected component of the identity in the multiplicative topological group $\pi(\mathbf{F}(X))$ of all invertible elements in $\mathbf{A}(X)$. The index gives the group isomorphisms $\mathbf{F}(X)/\mathbf{F}_0(X) \cong \pi(\mathbf{F}(X))/\pi(\mathbf{F}_0(X)) \cong \mathbf{Z}$. More generally, we have for any compact topological space Y the group isomorphisms $[Y, \mathbf{F}(X)] \cong [Y, \pi(\mathbf{F}(X))] \cong K(Y)$ of the groups of †homotopy classes of continuous mappings and the K-group in the †K-theory (M. F. Atiyah [12]).

If N is a †normal operator, its **essential spectrum** $\sigma_e(N)$ is defined to be the set of all $\lambda \in \sigma(N)$ that is not an isolated eigenvalue of finite multiplicity. Let H_1 and H_2 be self-adjoint operators. Then $\sigma_e(H_1) = \sigma_e(H_2)$ if and

only if $U^*H_1U=H_2+K$ for a unitary operator U and a compact operator K (Weyl-von Neumann theorem).

I. D. Berg and W. Sikonia (1971) extended this result to normal operators N_1 and N_2. Moreover, for any compact subset Y of $\mathbf{C}$ there exists a normal operator N such that $\sigma_e(N)=Y$. Hence it follows that the essential spectrum $\sigma_e(N)$ of a normal operator N coincides with the †spectrum $\sigma(\pi(N))$ of the image in $\mathbf{A}(X)$. Thus we define the **essential spectrum** $\sigma_e(A)$ of an arbitrary operator A to be the spectrum $\sigma(\pi(A))$ in $\mathbf{A}(X)$. An operator $A\in\mathbf{B}(X)$ is said to be **essentially normal** (resp. **essentially self-adjoint, essentially unitary**) if $\pi(A)$ is normal (resp. self-adjoint, unitary) in $\mathbf{A}(X)$. (Note that this definition of essentially self-adjoint operators is completely different from that in 251 Linear Operators E.) Since an essentially self-adjoint operator is the sum of a self-adjoint operator and a compact operator, the Weyl-von Neumann theorem classifies essentially self-adjoint operators up to †unitary equivalence modulo $\mathbf{B}^{(c)}(X)$. An operator A is essentially normal if and only if the commutator $[A,A^*]$ is compact, but it need not be the sum of a normal operator and a compact operator. For example, let V be the **unilateral shift operator** that maps the orthonormal basis $\mathbf{e}_i$ of X into $\mathbf{e}_{i+1}$ for every $i=1,2,\ldots$. Then V is essentially unitary, but it cannot be written as the sum of a normal operator and a compact operator. The essential spectrum $\sigma_e(V)$ is the unit circle, whereas the spectrum $\sigma(V)$ is the unit disk and $\operatorname{ind}(V-\lambda)=-1$ for $|\lambda|<1$.

J. The Brown-Douglas-Fillmore (BDF) Theory

The following is the main theorem for essentially normal operators, due to L. G. Brown, R. G. Douglas, and P. A. Fillmore [14]. Let A_1 and A_2 be essentially normal operators. There are a unitary operator U and a compact operator K such that $U^*A_1U=A_2+K$ if and only if $\sigma_e(A_1)=\sigma_e(A_2)$ and $\operatorname{ind}(A_1-\lambda)=\operatorname{ind}(A_2-\lambda)$ for every λ in the complement of the essential spectrum. An essentially normal operator A is the sum of a normal operator and a compact operator if and only if $\operatorname{ind}(A-\lambda)=0$ for every λ in the complement of $\sigma_e(A)$.

To prove this and many other facts, they developed the theory of extension of $\mathbf{B}^{(c)}(X)$ by the C^*-algebra $C(Y)$ of continuous complex-valued functions on a compact metrizable space Y [14–16]. This revealed deep relations between the theory of operator algebras on Hilbert spaces (→ 36 Banach Algebras, 308 Operator Algebras) and algebraic topology (in particular, K-theory; → 237 K-Theory). Extension theory also gives a natural setting for the †index theory of elliptic differential operators due to Atiyah and I. M. Singer [13, 16].

An **extension** of $\mathbf{B}^{(c)}(X)$ by $C(Y)$ is a †short exact sequence

$$0\to\mathbf{B}^{(c)}(X)\xrightarrow{\varepsilon}E\xrightarrow{\varphi}C(Y)\to 0 \qquad (11)$$

of a C^*-subalgebra E of $\mathbf{B}(X)$ and †$*$-homomorphisms, i.e., E is a C^*-subalgebra of $\mathbf{B}(X)$ containing the identity I and including $\mathbf{B}^{(c)}(X)$ as a C^*-subalgebra, and φ is a $*$-homomorphism onto $C(Y)$ whose kernel is equal to $\mathbf{B}^{(c)}(X)$. Or equivalently, an extension is a unital (identity preserving) $*$-monomorphism τ: $C(Y)\to\mathbf{A}(X)$ defined by $\tau=\pi\circ\varphi^{-1}$. (For general extension of C^*-algebras → 36 Banach Algebras.) Two extensions (E_1,φ_1) and (E_2,φ_2) (or τ_1 and τ_2) are said to be equivalent if there exists a $*$-isomorphism $\psi:E_1\to E_2$ such that $\varphi_2\circ\psi=\varphi_1$ (or equivalently there exists a unitary operator U such that $\pi(U^*)\tau_1(f)\pi(U)=\tau_2(f)$ for every $f\in C(Y)$). We denote by $\operatorname{Ext}(Y)$ the set of all equivalence classes of extensions of $\mathbf{B}^{(c)}(X)$ by $C(Y)$.

Let A be an essentially normal operator in X with the essential spectrum $\sigma_e(A)=Y$. Then the C^*-algebra E_A generated by $\mathbf{B}^{(c)}(X)$, A and the identity I, and the $*$-homomorphism φ_A of E_A onto $C(Y)$ which sends A to the function $\chi(z)=z$ define an extension (E_A,φ_A). It is easy to see that two essentially normal operators A_1 and A_2 are unitarily equivalent modulo $\mathbf{B}^{(c)}(X)$ if and only if $\sigma_e(A_1)=\sigma_e(A_2)$ and the extensions (E_{A_1},φ_{A_1}) and (E_{A_2},φ_{A_2}) are equivalent. Conversely, if Y is a compact subset of $\mathbf{C}$ and (11) is an extension, then (E,φ) is equivalent to (E_A,φ_A), where A is an essentially normal operator in E such that $\varphi(A)=\chi$.

Extensions of $\mathbf{B}^{(c)}(X)$ by $C(Y)$ appear also in different parts of analysis. Let X be the Hilbert space $L_2(M)$ on a compact differentiable manifold M relative to a fixed smooth measure and let E be the C^*-subalgebra of $\mathbf{B}(X)$ generated by all zeroth-order †pseudodifferential operators together with $\mathbf{B}^{(c)}(X)$. Then E and the †symbol mapping $\varphi:E\to C(S^*(M))$ define an extension of $\mathbf{B}^{(c)}(X)$ by $C(S^*(M))$, where $S^*(M)$ is the †cosphere bundle of M. This extension is closely related to the †Atiyah-Singer index theorem. Let Ω be a †strongly pseudoconvex domain in $\mathbf{C}^n$. Then the C^*-algebra generated by †Toeplitz operators with continuous symbol gives rise to an extension of $\mathbf{B}^{(c)}(H_2(\partial\Omega))$ by $C(\partial\Omega)$.

Let τ_1 and τ_2 be $*$-monomorphisms from $C(Y)$ into $\mathbf{A}(X)$ and $a_1=[\tau_1]$ and $a_2=[\tau_2]$ be corresponding elements in $\operatorname{Ext}(Y)$. Then the sum $a_1+a_2\in\operatorname{Ext}(Y)$ is defined to be the equivalence class of $\tau:C(Y)\to\mathbf{A}(X)$ which sends f to $\tau_1(f)\oplus\tau_2(f)\in\mathbf{A}(X)\oplus\mathbf{A}(X)\subset\mathbf{A}(X\oplus X)\cong\mathbf{A}(X)$. It does not depend on the

choice of τ_1, τ_2 and the unitary $X \oplus X \cong X$. An extension $\tau: C(Y) \to \mathbf{A}(X)$ is said to be **trivial** if there exists a unital $*$-monomorphism $\sigma: C(Y) \to \mathbf{B}(X)$ such that $\tau = \pi \circ \sigma$. For each compact metrizable space Y there exists a unique equivalence class of trivial extensions in $\mathrm{Ext}(Y)$. $\mathrm{Ext}(Y)$ is an Abelian group in which the class of trivial extensions is the identity element. The extension (E_A, φ_A) for an essentially normal operator A is trivial if and only if $A = N + K$ with N normal and K compact. Hence follows the Weyl-von Neumann-Berg-Sikonia theorem. The BDF theorem for essentially normal operators is proved by the pairing $\mathrm{Ext}(Y) \times K^1(Y) \to \mathbf{Z}$ defined by the index, where $K^1(Y) = \tilde{K}(SY^+) = \lim_{n\to\infty}[Y, GL(n, \mathbf{C})]$ (→ 237 K-Theory; [13]). The induced homomorphism $\gamma_\infty: \mathrm{Ext}(Y) \to \mathrm{Hom}(K^1(Y), \mathbf{Z})$ is always surjective, and it is an isomorphism for $Y \subset \mathbf{R}^3$ or $Y = S^n$ but not for $Y \subset \mathbf{R}^4$.

Ext is a †covariant functor from the category of compact metrizable spaces to the category of Abelian groups. It is †homotopy invariant. Define for $n = 0, 1, \ldots$ the group $\mathrm{Ext}_{1-n}(Y)$ by $\mathrm{Ext}(S^n Y)$, where $S^n Y$ is the n-fold †suspension. Then we have the **periodicity** $\mathrm{Ext}_{n+2}(Y) \cong \mathrm{Ext}_n(Y)$. Moreover, for each pair of compact metrizable spaces $Y \supset Z$ we have the long exact sequence

$$\underset{n}{\mathrm{Ext}}(Z) \xrightarrow{i_*} \underset{n}{\mathrm{Ext}}(Y) \xrightarrow{p_*} \underset{n}{\mathrm{Ext}}(Y/Z) \xrightarrow{\partial} \underset{n-1}{\mathrm{Ext}}(Z) \xrightarrow{i_*} \ldots,$$

where Y/Z is the space obtained from Y by collapsing Z to a point and ∂ is $q_* r_*^{-1}$: $\mathrm{Ext}_n(Y/Z) \to \mathrm{Ext}_n(SZ)$ defined by $q: Y \cup CZ \to (Y \cup CZ)/Y = SZ$ and $r: Y \cup CZ \to (Y \cup CZ)/CZ = Y/Z$, CZ being the †cone over Z. Ext_* satisfies the †Eilenberg-Steenrod axioms for homology theory except for the dimension axiom, which is replaced by $\mathrm{Ext}_n(S^0) = \mathrm{Ext}(S^{n-1}) = \mathbf{Z}$ for n even and $= 0$ for n odd.

The von Neumann algebra $\mathbf{B}(X)$ is classified as a †factor of type I_∞. In the case of a factor M of type II_∞ another index theory has been developed by H. Breuer [17] and others replacing $\mathbf{B}^{(c)}(X)$ by the closed ideal of M generated by finite projections and using the †semifinite trace on M for the dimensions of kernels and cokernels of operators in M.

K. Spectral Analysis in Banach Spaces

Spectral analysis becomes rather involved for general operators in a Banach space as well as for nonnormal operators in Hilbert space.

For a †compact operator A, the nature of the spectrum $\sigma(A)$ and the structure of A in the root subspace associated with a nonzero eigenvalue are well known (→ 68 Compact and Nuclear Operators). However, a full spectral analysis may not be possible without further assumptions.

For a †closed operator with nonempty resolvent set $\rho(A)$, an †operational calculus can be developed by means of a function-theoretic method based on the fact that the resolvent $R_\lambda = (\lambda I - A)^{-1}$ is a $\mathbf{B}(X)$-valued holomorphic function of λ in $\rho(A)$. In particular, the †spectral mapping theorem holds (→ 251 Linear Operators G).

A general class of operators having associated spectral resolution was introduced by N. Dunford. Let X be a Banach space. An operator $E \in \mathbf{B}(X)$ is called a projection if $E^2 = E$. As before we can define a (projection-valued countably additive) **spectral measure** on $\mathscr{B}_c$. An operator $A \in \mathbf{B}(X)$ is said to be a **spectral operator** if there exists a **spectral measure** E on $\mathscr{B}_c$ satisfying the following properties: (i) $E(M)A = AE(M)$, $M \in \mathscr{B}_c$; (ii) $\sigma(A|_{E(M)X}) \subset \overline{M}$, $M \in \mathscr{B}_c$, where $A|_Y$ is the restriction of A to Y and $\overline{M}$ is the closure of M; (iii) there exists a $k \geqslant 0$ such that $\|E(M)\| \leqslant k$ for all $M \in \mathscr{B}_c$. E is unique. A spectral operator A is expressed as $A = S + N$, where $S = \int_{\mathbf{C}} zE(dz)$ and N is †quasinilpotent. A is said to be a **scalar operator** if $N = 0$. Unbounded spectral operators are defined similarly, with (i′) $E(M)A \subset AE(M)$ in place of (i). However, for unbounded spectral operators A we no longer have the decomposition $A = S + N$. (For more details about spectral operators → [3]. For other topics related to the material discussed in this section → 68 Compact and Nuclear Operators; 251 Linear Operators; and 287 Numerical Computation of Eigenvalues.)

References

[1] N. I. Akhiezer and I. M. Glazman, Theory of linear operators in Hilbert space I, II, Ungar, 1961, 1963. (Original in Russian, 1950.)
[2] F. Riesz and B. Sz.-Nagy, Functional analysis, Ungar, 1955. (Original in French, 1952.)
[3] N. Dunford and J. T. Schwartz, Linear operators, I, II, III, Interscience, 1958, 1963, 1971.
[4] K. Yosida, Functional analysis, Springer, 1965.
[5] T. Kato, Perturbation theory for linear operators, Springer, 1966.
[6] M. Reed and B. Simon, Methods of modern mathematical physics, Academic Press, I, 1972; II, 1975; III, 1979; IV, 1978.
[7] P. R. Halmos, Introduction to Hilbert space and the theory of spectral multiplicity, Chelsea, second edition, 1957.

[8] M. H. Stone, Linear transformations in Hilbert space and their applications to analysis, Amer. Math. Soc. Colloq. Publ., 1932.
[9] I. C. Gokhberg (Gohberg) and M. G. Kreĭn, Introduction to the theory of linear non-self-adjoint operators, Amer. Math. Soc., 1969. (Original in Russian, 1965.)
[10] I. C. Gokhberg (Gohberg) and M. G. Kreĭn, Theory and applications of Volterra operators in Hilbert space, Amer. Math. Soc., 1970. (Original in Russian, 1967.)
[11] R. G. Douglas, Banach algebra techniques in operator theory, Academic Press, 1972.
[12] M. F. Atiyah, K-theory, Benjamin, 1967.
[13] M. F. Atiyah, Global theory of elliptic operators, Proc. Intern. Conf. on Functional Analysis and Related Topics, Univ. Tokyo Press, 1970, 21–30.
[14] L. G. Brown, R. G. Douglas, and P. A. Fillmore, Unitary equivalence modulo the compact operators and extensions of C^*-algebras, Lecture notes in math. 345, Springer, 1973, 58–128.
[15] L. G. Brown, R. G. Douglas, and P. A. Fillmore, Extensions of C^*-algebras and K-homology, Ann. Math., 105 (1977), 265–324.
[16] R. G. Douglas, C^*-algebra extensions and K-homology, Ann. math. studies 95, Princeton Univ. Press, 1980.
[17] M. Breuer, Fredholm theories on von Neumann algebras, Math. Ann., I, 178 (1968), 243–254; II, 180 (1969), 313–325.

391 (VII.21) Spectral Geometry

A. General Remarks

Let E^3 be a Euclidean 3-space with a standard coordinate system (x_1, x_2, x_3) and E^2 be the x_1x_2-plane. Consider a domain D in E^2 as a vibrating membrane with the fixed boundary ∂D. Then the height $x_3 = F(x_1, x_2, t)$ obeys the differential equation of hyperbolic type, $\partial^2 F(x_1, x_2, t)/\partial t^2 = c^2 \Delta F(x_1, x_2, t)$, where $\Delta = \partial^2/\partial x_1^2 + \partial^2/\partial x_2^2$ denotes the †Laplacian in E^2 and c is a constant (we put $c=1$ in the following). For solutions of the form $F(x_1, x_2, t) = U(x_1, x_2) V(t)$, U is a solution of the †Dirichlet problem in $D \subset E^2$; $\Delta U + \lambda U = 0$, where λ is a positive constant called an †eigenvalue of Δ. Solutions of the form $F(x_1, x_2, t) = U(x_1, x_2) \sin\sqrt{\lambda}t$ represent the pure tones that the membrane produces as normal modes. That is, the shape of D is related to the possible sounds or vibrations (i.e., to the eigenvalues of Δ) through the Dirichlet problem. The set of eigenvalues of Δ is called the **spectrum of** D and is denoted by Spec(D). There arises the question of how much information Spec(D) can impart about the geometric properties (i.e., the shape, extent, and connectedness) of D. Generally, **spectral geometry** is the study of the relations between the spectrum of domains D of †Riemannian manifolds or compact Riemannian manifolds (M, g) and the geometric properties of D or (M, g).

B. Spectra

Let $\mathscr{D}^p(M)$ denote the space of smooth †p-forms on a compact m-dimensional C^∞-Riemannian manifold (M, g). Then eigenvalues of the †Laplacian (Laplace-Beltrami operator) Δ acting on $\mathscr{D}^p(M)$ are discretely distributed in $[0, \infty)$, and each multiplicity is finite (→ 68 Compact and Nuclear Operators, 323 Partial Differential Equations of Elliptic Type). The **spectrum for** p**-forms** Spec$^p(M, g)$ is $\{\lambda_{p,1} \leqslant \lambda_{p,2} \leqslant \ldots\}$, where each eigenvalue is repeated as many times as its multiplicity indicates. If 0 is an eigenvalue, its multiplicity is equal to the pth †Betti number of M.

In the following, mainly the case $p=0$ is explained. Spec$^0(M, g)$ is denoted by Spec(M, g). 0 is always in Spec(M, g) and its multiplicity is 1. So we put $\lambda_0 = 0$, and λ_1 is the first nonzero eigenvalue. A geometric meaning of Δf at x for a function f is as follows: If $\{\gamma_h\}_{h=1}^m$ are m geodesics mutually orthogonal at x and parametrized by arc length, then $(\Delta f)(x) = \sum_h (f \circ \gamma_h)''(0)$.

Let $\{\varphi_i\}_{i=0}^\infty$ be an orthonormal basis of $\mathscr{D}^0(M)$ consisting of eigenfunctions: $\Delta\varphi_i + \lambda_i \varphi_i = 0$, $\langle \varphi_i, \varphi_j \rangle = \int_M \varphi_i \varphi_j = \delta_{ij}$. Then the †fundamental solution $E(x, y, t)$ of the heat equation $\Delta - \partial/t = 0$ is given by $E(x, y, t) = \sum_i e^{-\lambda_i t} \varphi_i(x) \otimes \varphi_i(y)$ as a function on $M \times M \times (0, \infty)$. We put $Z(t) = \int_M E(x, x, t) = \sum_i e^{-\lambda_i t}$. $Z(t)$ and Spec(M, g) are equivalent. The Minakshisundaram-Pleijel **asymptotic expansion of** $Z(t)$,

$$Z(t) \sim (1/4\pi t)^{m/2}(a_0 + a_1 t + a_2 t^2 + \ldots), \qquad t \downarrow 0,$$

is the bridge connecting Spec(M, g) and geometric properties of (M, g), because $a_0, a_1, \ldots$ can be expressed as the integrals of functions over M defined by $g = (g_{ij})$, the components R^i_{jkl} of the †Riemannian curvature tensor, and their derivatives of finite order [1]. a_0 is the volume of (M, g) and $a_1 = (1/6)\int_M S$, where S is the †scalar curvature. a_2 was calculated by H. P. Mckean, I. M. Singer, and M. Berger, and a_3 by T. Sakai.

Let D be a bounded domain in E^2 or, more generally, a bounded domain in a Riemannian manifold, and assume that the boundary ∂D is piecewise smooth. For smooth functions which

take the value 0 on ∂D, eigenvalues of the Laplacian Δ are discretely distributed in $(0, \infty)$, and each multiplicity is finite. We denote the spectrum of D by $\mathrm{Spec}(D)=\{\lambda_1<\lambda_2\leqslant\ldots\}$. The multiplicity of λ_1 is 1, and an eigenfunction f corresponding to λ_1 takes the same sign in D. The behavior of $Z(t)=\sum_i e^{-\lambda_i t}$ for D is different from that for (M,g) since $Z(t)$ reflects the geometric situation of ∂D in this case.

(M,g) and (N,h), or D_1 and D_2 are called **isospectral** if they have the same spectra.

Examples for which the spectrum is explicitly calculable are as follows: spheres $(S^m, g_0=$ canonical), real projective spaces (RP^m, g_0), complex projective spaces (CP^n, J_0, g_0), $(S^{2n+1}, g_s=$ suitably deformed from $g_0)$, flat tori, (and for domains D) unit disks, rectangles, equilateral triangles, etc.

C. Congruence and Characterization

Let D_1 and D_2 be bounded domains in E^2. An open question is whether isospectral D_1, D_2 are congruent. Concerning this, there is M. Kac's paper with the famous title "Can one hear the shape of a drum?" Let D be a bounded domain in E^2 with smooth boundary ∂D. If D has r holes, then

$$Z(t)\sim A(D)/4\pi t-L(\partial D)/4\sqrt{4\pi t}+(1-r)/12, \qquad t\downarrow 0,$$

holds (A. Pleijel, Kac, P. Mckean, I. M. Singer; → [7]), where $A(D)$ denotes the area of D and $L(\partial D)$ denotes the length of ∂D. This theorem implies that the area, the length of ∂D, and the number of holes are determined by $\mathrm{Spec}(D)$. In particular, if D_2 is a unit disk, $\mathrm{Spec}(D_1)=\mathrm{Spec}(D_2)$ implies that D_1 and D_2 are congruent. There are some other results on $Z(t)$ for domains D of surfaces in E^m or for domains D in Riemannian manifolds (M,g).

Two isometric (M,g), (N,h) are isospectral. Concerning the question of whether isospectral (M,g), (N,h) are isometric, there are some counterexamples. The first is the case of two flat tori T^{16}, given by J. Milnor. Examples with nonflat metrics were given by N. Ejiri using warped products and by M. F. Vigneras for surfaces of constant negative curvature. In those examples, M and N are homeomorphic. A. Ikeda showed that there are †lens spaces that are isospectral but not homotopy equivalent.

Examples of affirmative cases are as follows: $\mathrm{Spec}(M,g)=\mathrm{Spec}(S^m,g_0)$, $m\leqslant 6$, implies that (M,g) is isometric to (S^m,g_0) (M. Berger, S. Tanno). The result is the same for (RP^m,g_0), $m\leqslant 6$. For $n\leqslant 6$, (CP^n,J_0,g_0) is characterized by a spectrum among †Kählerian manifolds (M,J,g).

The number of nonisometric flat tori (or more generally, compact flat Riemannian manifolds) with the same spectrum is finite (M. Kneser, T. Sunada).

If one considers spectra for two types of forms, then the situation turns out to be simpler. For example, if $\mathrm{Spec}^p(M,g)=\mathrm{Spec}^p(N,h)$ for $p=0$, 1, then (M,g) is of constant curvature K if and only if (N,h) is also and $K'=K$ (V. K. Patodi [9]).

D. The First Eigenvalue

The first eigenvalue λ_1 for (M,g) or for a domain D in (M,g) reflects the geometric situation of (M,g) or D. A lower bound of λ_1 given by J. Cheeger is $\lambda_1\geqslant h(M)^2/4$, where $h(M)$ is the **isoperimetric constant**, defined by

$$h(M)=\inf\{\mathrm{vol}(S)/\min[\mathrm{vol}(M_1),\mathrm{vol}(M_2)]\},$$

where the inf is taken over all smoothly embedded hypersurfaces S dividing M into two open submanifolds M_1, M_2, $\partial M_1=\partial M_2=S$, and vol means the volume. 1/4 is the best possible estimate.

Let ρ denote the maximum radius of a disk included in a simply connected $D\subset E^2$. Then $\lambda_1\geqslant 1/(4\rho^2)$ (W. K. Hayman, R. Osserman).

If (M,g) has †Ricci curvature $\geqslant k>0$, then $\lambda_1\geqslant mk/(m-1)$ holds, and the equality holds if and only if (M,g) is isometric to $(S^m,(1/k)g_0)$ (A. Lichnerowicz, M. Obata). λ_1 can also be estimated from k and the diameter $d(M)$ of M (P. Li, S. T. Yau).

To obtain upper bounds of λ_1 the **minimum principle of** λ_1 is useful. We state it only for (M,g):

$$\lambda_1=\inf\left\{\int_M(\nabla f,\nabla f)\Big/\int_M f^2\right\},$$

where inf is taken over all piecewise smooth functions f satisfying $\int_M f=0$, $f\neq 0$, and $(\ ,\ )$ denotes the local inner product with respect to g. An upper bound of λ_1 for (M,g) of nonnegative curvature is $\lambda_1\leqslant c(m)/d(M)^2$, where $c(m)$ is some constant depending on m (Cheeger).

For (M,g) or D a submanifold of another Riemannian manifold, there exist some estimates of λ_1 in terms of second fundamental forms, etc.

$\lambda_1\geqslant j^2/(A(D))$ holds for $D\subset E^2$, and the equality holds if and only if ∂D is a circle (C. Faber, E. Krahn), where j denotes the first zero of the †Bessel function J_0. This estimate is generalized in many directions; for example, for $D\subset(M^2,g)$ in terms of the integral of the Gauss curvature, etc. It is very useful to note that the estimate of λ_1 for D is deeply related to the isoperimetric inequality (→ 228 Isoperimetric Problems).

E. Hersch Type Theorem

With respect to the first eigenvalue $\lambda_1(g)$ and the volume $\mathrm{Vol}(M,g)$ of (M,g), $\lambda_1(g)\cdot\mathrm{vol}(M,g)^{2/m}$ is invariant under a change of metric $g\to c^2 g$ (c is a constant). **Hersch's problem** is stated as follows: Is there a constant $k(M)$ depending on M so that for any Riemannian metric g on M, $\lambda_1(g)\cdot\mathrm{vol}(M,g)^{2/m}\leqslant k(M)$? J. Hersch proved this for a 2-sphere $M=S^2$ with $k(S^2)=8\pi$, and in this case the equality holds if and only if g is proportional to the canonical metric g_0.

The Hersch type theorem holds for an oriented Riemann surface M of genus q with $k(M)=8\pi(q+1)$ (P. C. Yang, S. T. Yau). There is no such constant $k(S^m)$ for an m-sphere S^m, $m\geqslant 3$ (H. Urakawa, H. Muto, S. Tanno).

F. The Multiplicity of λ_i

By a theorem of K. Uhlenbeck each eigenvalue for a Riemannian manifold (M,g) is simple. However, for (S^m, g_0) the first eigenvalue λ_1 is m and its multiplicity is $m+1$. Furthermore, for some g_s deformed from g_0, $\lambda_1(g_s)$ of (S^{2n+1}, g_s) has multiplicity n^2+4n+2, which is larger than $m+1$ $(=2n+2)$.

The multiplicity $m(\lambda_i)$ of the ith eigenvalue λ_i for a Riemann surface of genus q satisfies $m(\lambda_i)\leqslant 4q+2i+1$ (S. Y. Cheng, G. Besson).

G. kth Eigenvalue

The **minimum principle for** λ_k of $\mathrm{Spec}(M,g)$ is stated as follows: Let f_i be an eigenfunction corresponding to λ_i, $0\leqslant i\leqslant k-1$. Define H_{k-1} to be the set of piecewise smooth functions $f\neq 0$ orthogonal to each f_i, i.e., $\int_M ff_i=0$. Then

$$\lambda_k=\inf\left\{\int_M(\nabla f,\nabla f)\Big/\int_M f^2\right\},$$

where inf is taken over $f\in H_{k-1}$. We have the **minimax principle for** λ_k of the first and second type. We state the second type only:

$$\lambda_k=\inf_{L_{k+1}}\sup_{0\neq f\in L_{k+1}}\left\{\int_M(\nabla f,\nabla f)\Big/\int_M f^2\right\},$$

where L_k denotes a k-dimensional linear subspace of $\mathscr{D}^0(M)$. From this, for 1-parameter metrics g_u $(a<u<b)$ on M, the continuity of $\lambda_k(g_u)$ with respect to u follows.

H. Courant-Cheng Nodal Domain Theorem

Let f be an eigenfunction on (M,g) or D. The set of all zero points of f is called the **nodal set** of f (or the **nodal curve** of f if $m=2$). Each connected component of the complement of the nodal set in (M,g) or D is called a **nodal domain** of f. The nodal set of f is a smooth submanifold of (M,g) or D except for a set of lower dimension. The number of nodal domains of an eigenfunction corresponding to the ith eigenvalue is $\leqslant i+1$ for (M,g) and $\leqslant i$ for D (**Courant-Cheng nodal domain theorem**).

I. Estimate of $N(\lambda)$

$N(\lambda)$ is defined as the number of eigenvalues of (M,g) or D which are less than or equal to λ. For $D\subset E^2$, H. A. Lorentz conjectured that the behavior of $N(\lambda)$ for $\lambda\to\infty$ does not depend on the shape of D but only on the area $A(D)$ of D, i.e., $\lim_{\lambda\to\infty}N(\lambda)/\lambda=A(D)/4\pi$. This was proved by H. Weyl. Generally, for D or (M,g), the behavior of $N(\lambda)$ for $\lambda\to\infty$ is $\mathrm{vol}(D)\lambda^{m/2}/(4\pi)^{m/2}\Gamma(m/2+1)$, and this is related to the first term $\mathrm{vol}(D)/(4\pi t)^{m/2}$ of the asymptotic expansion of $Z(t)$ by †Tauberian and †Abelian theorems.

J. Spec(M,g) and Geodesics

Let $T^m=E^m/\Gamma$ be a flat torus, Γ being the lattice for T^m. Let Γ^* be the lattice dual to Γ. Then **Poisson's formula**,

$$\sum_{y\in\Gamma^*}e^{-4\pi^2|y|^2t}=(\mathrm{vol}(T^m)/(4\pi t)^{m/2})\sum_{x\in\Gamma}e^{-|x|^2/4t},$$

gives a clear relation between $\mathrm{Spec}(T^m)=\{4\pi^2|y|^2, y\in\Gamma^*\}$ and the set $\{|x|, x\in\Gamma\}$ of lengths of closed geodesics on T^m.

If (M,g) satisfies some conditions, then $\mathrm{Spec}(M,g)$ determines the set of lengths of periodic geodesics (Y. Colin de Verdiere), and the spectrum characterizes those Riemannian manifolds whose geodesics are all periodic (J. J. Duistermaat, V. W. Guillemin).

K. Spec$^p(M,g)$ and the Euler-Poincaré Characteristic

Let (M,g) be oriented and even dimensional. Let $E^p(x,y,t)$ be the †fundamental solution of the †heat equation for p-forms. Corresponding to $Z(t)$ for $\mathrm{Spec}(M,g)$, we get $Z^p(t)=\int_M E^p=\sum_i e^{-\lambda_{p,i}t}$. Then

$$\sum_{p=0}^{m}(-1)^pZ^p(t)=\sum_{p=0}^{m}(-1)^p\int_M\mathrm{tr}\,E^p=\chi(M),$$

where $\chi(M)$ denotes the †Euler-Poincaré characteristic of M (Mckean, Singer). On the other hand, the †Gauss-Bonnet theorem is $\chi(M)=\int_M C$, where C is a function on M expressed as a homogeneous polynomial of components of the Riemannian curvature tensor. Then

Patodi proved

$$\sum_{p=0}^{m}(-1)^p \operatorname{tr} E^p = C + O(t), \qquad t \downarrow 0.$$

L. η-Function

Let (X,g) be a compact oriented $4k$-dimensional Riemannian manifold with boundary $\partial X = Y$ and assume that some neighborhood of Y in (X,g) is isometric to a Riemannian product $Y \times [0,\varepsilon)$. Define an operator B acting on forms of even degree on Y by

$$Bw = (-1)^{k+r+1}(*d - d*)w, \qquad w \in \mathscr{D}^{2r}(Y),$$

where $*$ denotes the †Hodge star operator and d denotes exterior differentiation on Y. Then $B^2 = \Delta$ holds. Using the spectrum $\{\mu\}$ of B, we define the **η-function** by

$$\eta(s) = \sum_{\mu \neq 0} (\operatorname{sgn}\mu)|\mu|^{-s}.$$

$\eta(s)$ is a spectral invariant, and

$$\operatorname{sgn}(X) = \int_X L_k(p_1, \ldots, p_k) - \eta(0)$$

holds (Atiyah, Patodi, and Singer [2]), where $\operatorname{sgn}(X)$ is the †signature of the quadratic form defined by the †cup product on the image of $H^{2k}(X, Y)$ in $H^{2k}(X)$, L_k is the kth †Hirzebruch L-polynomial, and $p_1, \ldots, p_k$ are the Pontryagin forms of (X,g).

M. Analytic Torsion

Let χ be a representation of the fundamental group $\pi_1(M)$ of (M,g) by the orthogonal group and E_χ be the associated vector bundle. Let Δ^χ be the Laplacian acting on E_χ-valued p-forms on M and $\{\lambda^\chi_{p,i}\}$ be its spectrum. Then

$$\log T(M,\chi) = \sum_{p=0}^{m}(-1)^p p \log \frac{d}{ds}\left(\sum_i (\lambda^\chi_{p,i})^{-s}\right)\Bigg|_{s=0}$$

is independent of the choice of g. $T(M,\chi)$ is called the **analytic torsion** of M. $T(M,\chi)$ is equal to the †R-torsion $\tau(M,\chi)$ (W. Müller, Cheeger).

N. Concluding Remarks

An †isometry ψ of (M,g) commutes with the Laplacian and induces a linear transformation $\psi^\#_\lambda$ of each eigenspace V_λ. Using the asymptotic expansion of $\sum \operatorname{tr}(\psi^\#_\lambda)e^{-\lambda t}$, the Atiyah-Singer †$G$-signature theorem has been proved (H. Donnelly, Patodi).

The Atiyah-Singer †index theorem has been proved by using Gilkey's theory and the heat equation (Atiyah, R. Bott, Patodi).

Let (N,h) be a complete Riemannian manifold of negative curvature. Then Δ is extended to an unbounded self-adjoint operator for $L_2(N)$. Generally Δ has a continuous spectrum. Some conditions for (N,h) to have pure point spectrum were given in terms of curvature (Donnelly, P. Li).

If D is a minimal submanifold of another Riemannian manifold, estimates of λ_1 are related to the stability of D (→ 275 Minimal Submanifolds).

On the nonexistence of the 1-parameter isospectral deformation $(M,g_0) \to (M,g_t)$, there are results for (i) $m=2$ and g_0 of negative curvature, $m \geqslant 3$ and negatively pinched g_0 (V. Guillemin, D. Kazhdan); (ii) flat metrics g_0 (R. Kuwabara); and (iii) g_0 of constant positive curvature (Tanno).

As for spectral geometry for complex Laplacian on Hermitian manifolds, there are results by P. Gilkey, Tsukada, and others.

References

[1] M. F. Atiyah, R. Bott, and V. K. Patodi, On the heat equation and the index theorem, Inventiones Math., 19 (1973), 279–330.

[2] M. F. Atiyah, V. K. Patodi, and I. M. Singer, Spectral asymmetry and Riemannian geometry I, II, III, Proc. Cambridge Philos. Soc., 77 (1975), 43–69, 405–432; 79 (1976), 71–99.

[3] M. Berger, P. Gauduchon and E. Mazet, Le spectre d'une variété riemannienne, Lecture notes in math. 194, Springer, 1971.

[4] J. Cheeger, A lower bound for the smallest eigenvalue of the Laplacian, Problems in Analysis: Symposium in Honor of Salomon Bochner, R. C. Gunning (ed.), Princeton Univ. Press, 1970, 175–199.

[5] P. B. Gilkey, The index theorem and the heat equation, Publish or Perish, 1974.

[6] M. Kac, Can one hear the shape of a drum ? Amer. Math. Monthly, 73 (April 1966), 1–23.

[7] H. P. Mckean and I. M. Singer, Curvature and eigenvalues of the Laplacian, J. Differential Geometry, 1 (1967), 43–69.

[8] S. Minakshisundarum and A. Pleijel, Some properties of the eigenfunctions of the Laplace operator on Riemannian manifolds, Canad. J. Math., 1 (1949), 242–256.

[9] V. K. Patodi, Curvature and the fundamental solution of the heat operator, J. Indian Math. Soc., 34 (1970), 269–285.

392 (XX.5) Spherical Astronomy

Spherical astronomy is concerned with the apparent positions of celestial bodies and their

motions on a celestial sphere with center at an observer on the Earth, while †celestial mechanics is concerned with computing heliocentric true positions of planets and comets and geocentric true positions of satellites. The purpose of spherical astronomy is to find all possible causes of displacement of the apparent positions of celestial bodies on the celestial sphere from their geocentric positions and to study their effects. Atmospheric refraction, geocentric parallax, aberration, annual parallax, precession, nutation, and proper motion are examples of these causes.

When light from a celestial body passes through the Earth's atmosphere, it is refracted since air densities at different heights are different. This phenomenon is called **atmospheric refraction**. The effect of refraction on the apparent direction of the celestial body is a minimum when the body is at its culmination, and vanishes when this coincides with the observer's zenith, while the maximum refraction of 34′.5 occurs when the body is at the horizon.

Topocentric positions differ appreciably from geocentric positions for the Moon and planets, since their geocentric distances are not large compared with the radius of the Earth. The difference is largest when the observer is on the equator and the celestial body is at the horizon, and this largest value is called the **geocentric parallax**. The geocentric parallax of the Moon is between 53′.9 and 60′.2; those of the Sun, Mercury, Venus, Mars, Jupiter, and Saturn are, respectively, 8″.64–8″.94, 6″–16″.5, 5″–32″, 3″.5–23″.5, 1″.4–2″.1, and 0″.8–1″.1. For fixed stars geocentric parallaxes can be regarded as zero since the stars are far from the Earth.

The Earth moves in an orbit around the Sun with period of one year (365.2564 days) and rotates around the polar axis, which is inclined at 66°.5 to the orbital plane (the **ecliptic**), with period of one day (23 hours, 56 minutes, 4.091 seconds). Therefore the observer on the Earth moves with a speed depending on the latitude (0.465 km/sec on the equator) due to the rotation and moves with an average speed of 29.785 km/sec on the ecliptic. Due to these motions of the observer, apparent directions of celestial bodies are displaced from their geometric directions. Displacement due to the rotation is called **diurnal aberration**, and that due to the orbital motion, **annual aberration**. The effect of diurnal aberration is between 0″ and 0″.32 and varies with a period of one day, while that of the annual aberration is between 0″ and 20″.496 and varies with a period of one year. Moreover, to compute the positions of celestial bodies, the travel time of light to the observer should be taken into account.

Annual parallax for a fixed star is half the difference of its apparent directions, which are measured at the ends of a diameter perpendicular to the direction of the star from the orbit of the Earth. The effect of the annual parallax varies with a period of one year. However, except for nearby stars, it is not necessary to take this effect into account when computing apparent positions.

The pole of the Earth on the celestial sphere moves on a circle around the pole of the ecliptic due to the gravitational attraction of the Moon, Sun, and planets, and therefore the equinox moves clockwise on the ecliptic. Because the resultant of the attractive force of the Moon, Sun, and planets changes periodically, the motion of the equinox is not uniform. Therefore the motion is expressed as the sum of secular motion, called **precession**, and periodic motion, called **nutation**. Since the positons of fixed stars on the celestial sphere are measured with respect to the equator and the equinox, their right ascensions and declinations are continuously changing because of precession and nutation.

Since the stars are not fixed in space but themselves have **proper motions**, their positions on the celestial sphere are continuously changing.

Spherical astronomy also includes predictions of solar and lunar eclipses, the theory of †orbit determination to compute apparent positions of celestial bodies in the solar system by use of orbital elements, and the computation of ephemerides for the Sun, Moon, planets, and fixed stars. Practical astronomy, which develops theories and methods of observation by use of meridian circles, transit instruments, zenith telescopes, sextants, theodolites, telescopes with equatorial mountings, and astronomical clocks, and navigational astronomy, which deals with methods for determining the positions of ships and aircraft, are closely connected to spherical astronomy.

It should be noted that recently radar has been used to measure distances to the Moon and planets accurately, a contribution to determining the size of the solar system with precision.

References

[1] W. Chauvenet, A manual of spherical and practical astronomy I, II, Dover, 1960.
[2] S. Newcomb, A compendium of spherical astronomy, Dover, 1960.
[3] W. M. Smart, Textbook on spherical astronomy, Cambridge Univ. Press, fifth edition, 1965.

[4] E. W. Woolard and G. M. Clemence, Spherical astronomy, Academic Press, 1966.

393 (XIV.6) Spherical Functions

A. Spherical Functions

The term *spherical functions* in modern terminology means a certain family of functions on †symmetric Riemannian spaces obtained as simultaneous †eigenfunctions of certain integral operations (→ 437 Unitary Representations). In this article, however, we explain only the classical theory of **Laplace's spherical functions** with respect to the rotation group in 3-dimensional space.

Solutions of †Laplace's equation $\Delta V=0$ that are homogeneous polynomials of degree n with respect to the orthogonal coordinates x, y, z are called **solid harmonics** of degree n. If n is a positive integer, there are $2n+1$ linearly independent solid harmonics of degree n. In †polar coordinates (r, θ, φ) they are of the form $r^n Y_n(\theta, \varphi)$, where $Y_n(\theta, \varphi)$, the **surface harmonic** of degree n, satisfies the differential equation

$$\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial Y_n}{\partial\theta}\right)+\frac{1}{\sin^2\theta}\frac{\partial^2 Y_n}{\partial\varphi^2}+n(n+1)Y_n=0.$$

Here, if we apply †separation of variables to θ and φ and put $z=\cos\theta$, then the component in φ is represented by trigonometric functions, and the other component in θ reduces to a solution of **Legendre's associated differential equation**

$$(1-z^2)\frac{d^2w}{dz^2}-2z\frac{dw}{dz}+\left(n(n+1)-\frac{m^2}{1-z^2}\right)w=0. \quad (1)$$

B. Legendre Functions

With $m=0$ in (1) and n replaced by an arbitrary complex number ν, the equation is reduced to **Legendre's differential equation**

$$(1-z^2)\frac{d^2w}{dz^2}-2z\frac{dw}{dz}+\nu(\nu+1)w=0, \quad (2)$$

whose fundamental solutions are represented by

$$P_\nu(z)=\frac{1}{2\pi i}\oint^{(1+,z+)}\frac{(\zeta^2-1)^\nu}{2^\nu(z-\zeta)^{\nu+1}}d\zeta, \quad (3)$$

$$Q_\nu(z)=\frac{1}{4i\sin\nu\pi}\oint^{(1+,-1+)}\frac{(\zeta^2-1)^\nu}{2^\nu(z-\zeta)^{\nu+1}}d\zeta, \quad (4)$$

where the contour of integration in (3) is a closed curve with positive direction on the ζ-plane, avoiding the half-line $(-\infty, -1)$, and admitting 1 and z as inner points of the domain it bounds, whereas the contour in (4) is a closed ∞-shaped curve encircling the point 1 once in the negative direction and the point -1 once in the positive direction. The functions $P_\nu(z)$ and $Q_\nu(z)$ are called **Legendre functions of the first and second kind,** respectively. The integral representation (3) is **Schläfli's integral representation.** If $\mathrm{Re}(\nu+1)>0$, we can deform the contour of integration and obtain

$$Q_\nu(z)=\frac{1}{2^{\nu+1}}\int_{-1}^{1}\frac{(1-\zeta^2)^\nu}{(z-\zeta)^{\nu+1}}d\zeta. \quad (5)$$

If ν is an integer, it is convenient to use the representation (5).

From (3)–(5), we can obtain the recurrence formulas for Legendre functions of distinct degrees. The recurrence formulas for $P_\nu(z)$ and $Q_\nu(z)$ have exactly the same form (→ Appendix A, Table 18.II). Expanding the integrand in (3) and (4) with respect to $z-1$ and ζ/z, the following identities are obtained:

$$P_\nu(z)=F\left(\nu+1,-\nu,1,\frac{1-z}{2}\right), \quad |1-z|<2,$$

$$Q_\nu(z)=\frac{\sqrt{\pi}\,\Gamma(\nu+1)}{(2z)^{\nu+1}\Gamma(\nu+3/2)}\times F\left(\frac{\nu+1}{2},\frac{\nu+2}{2},\nu+\frac{3}{2},\frac{1}{z^2}\right),$$

$$|z|>1,\ |\arg z|<\pi,$$

where $F(\alpha, \beta, \gamma, z)$ is the †hypergeometric function. These expansions are the solutions in series of Legendre's differential equation in the neighborhood of the †regular singular points $z=1$ and ∞, respectively (→ Appendix A, Table 18.II).

If ν is a positive integer, since $\zeta=1$ is not a †branch point in (3), $P_n(z)$ is represented by **Rodrigues's formula**

$$P_n(z)=\frac{1}{2\pi i}\oint\frac{(\zeta^2-1)^n}{2^n(z-\zeta)^{n+1}}d\zeta=\frac{1}{2^n n!}\frac{d^n}{dz^n}(z^2-1)^n. \quad (6)$$

In this case, $P_n(z)$ is a polynomial of degree n such that

$$P_n(z)=\sum_{r=0}^{[n/2]}(-1)^r\frac{(2n-2r)!}{2^n r!(n-r)!(n-2r)!}z^{n-2r},$$

$$P_0(z)=1,$$

which is called the **Legendre polynomial** (Legendre, 1784). The †generating function for the

Legendre polynomials is $(1-2\rho\cos\theta+\rho^2)^{-1/2}$, whose expansion with respect to ρ is of the form $\sum_{n=0}^{\infty}P_n(z)\rho^n$, $z=\cos\theta$. Here the †generating function $(1-2\rho\cos\theta+\rho^2)^{-1/2}$ is the inverse of the distance between two points (ρ,θ) and $(1,0)$ in polar coordinates. Hence $P_n(z)$ is also called the **Legendre coefficient**. If z is real, $\{((2n+1)/2)^{1/2}P_n(z)\}_{n=0}^{\infty}$ constitutes an orthonormal system on $[-1,1]$ (→ 317 Orthogonal Functions). The n zeros of $P_n(z)$ are all real, simple, and lie in $(-1,1)$. For sufficiently large n, we have

$$P_n(\cos\theta)$$
$$=\sqrt{\frac{2}{n\pi\sin\theta}}\sin\left(\left(n+\frac{1}{2}\right)\theta+\frac{\pi}{4}\right)+O(1/n^{3/2}),$$
$$Q_n(\cos\theta)$$
$$=\sqrt{\frac{\pi}{2n\sin\theta}}\cos\left(\left(n+\frac{1}{2}\right)\theta+\frac{\pi}{4}\right)+O(1/n^{3/2})$$

as $n\to\infty$.

C. Associated Legendre Functions

For any positive integer m, the functions

$$P_\nu^m(z)=(1-z^2)^{m/2}d^mP_\nu(z)/dz^m,$$

$$Q_\nu^m(z)=(1-z^2)^{m/2}d^mQ_\nu(z)/dz^m$$

are called the **associated Legendre functions of the first and second kind**, respectively. This definition, due to N. M. Ferrers, is convenient for the case $-1<z<1$. For arbitrary complex z in a domain G obtained by deleting the segment $[-1,1]$ from the complex plane, the following definition, due to H. E. Heine and E. W. Hobson, is used:

$$P_\nu^m(z)=(z^2-1)^{m/2}d^mP_\nu(z)/dz^m,$$

$$Q_\nu^m(z)=(z^2-1)^{m/2}d^mQ_\nu(z)/dz^m.$$

The associated Legendre functions satisfy the associated Legendre differential equation (1). In particular, for $\nu=n$ (a positive integer) and $z=x$ (real),

$$\{(2n+1)(n-m)!/2(n+m)!\}^{1/2}P_n^m(z),$$
$$n=0,1,\ldots\ m=\text{constant},$$

constitute an orthonormal system on $[-1,1]$.

The **addition theorem** for the Legendre functions is

$$P_n\left(z_1z_2\pm\sqrt{\pm(1-z_1^2)}\sqrt{\pm(1-z_2^2)}\cos\varphi\right)$$
$$=P_n(z_1)P_n(z_2)+2\sum_{m=1}^{n}\frac{(n-m)!}{(n+m)!}P_n^m(z_1)$$
$$\times P_n^m(z_2)\cos m\varphi,$$

where the equality with the plus sign was obtained by Ferrers and that with the minus sign by Heine and Hobson.

D. Surface Harmonics

From the considerations so far for the surface harmonics $Y_n(\theta,\varphi)$, $2n+1$ independent solutions

$$P_n(\cos\theta),\quad P_n^m(\cos\theta)\sin m\varphi,$$

$$qP_n^m(\cos\theta)\cos m\varphi,\quad 1\leqslant m\leqslant n,$$

are obtained. Since $P_n(\cos\theta)$ vanishes on n latitudes of the unit sphere, and $P_n^m(\cos\theta)\cdot\cos m\varphi$ and $P_n^m(\cos\theta)\sin m\varphi$ vanish on $n-m$ latitudes and m longitudes of the unit sphere, respectively, the former functions are called **zonal harmonics** and the latter, **tesseral harmonics**. The general form of surface harmonics Y_n of order n is given by a linear combination of zonal and tesseral harmonics:

$$Y_n(\theta,\varphi)=A_{n,0}P_n(\cos\theta)$$
$$+\sum_{m=1}^{n}(A_{n,m}\cos m\varphi+B_{n,m}\sin m\varphi)P_n^m(\cos\theta). \qquad (7)$$

Expressing two surface harmonics $Y_n^{(1)}$ and $Y_l^{(2)}$ in linear combinations such as (7), the following orthogonality relations hold:

$$\int_{-\pi}^{\pi}\int_0^{\pi}Y_n^{(1)}(\theta,\varphi)Y_l^{(2)}(\theta,\varphi)\sin\theta\,d\theta\,d\varphi$$
$$=\delta_{n,l}\frac{4\pi}{2n+1}\left(A_{n,0}^{(1)}A_{n,0}^{(2)}\right.$$
$$\left.+\frac{1}{2}\sum_{m=1}^{n}\frac{(n+m)!}{(n-m)!}(A_{n,m}^{(1)}A_{n,m}^{(2)}+B_{n,m}^{(1)}B_{n,m}^{(2)})\right).$$

Since the family of all zonal and tesseral harmonics constitutes a †complete orthogonal system, it is possible to expand a function $f(\theta,\varphi)$ on the sphere into an orthogonal series:

$$f(\theta,\varphi)=\sum_{n=0}^{\infty}Y_n(\theta,\varphi)$$
$$=\sum_{n=0}^{\infty}\left(A_{n,0}P_n(\cos\theta)+\sum_{m=1}^{n}(A_{n,m}\cos m\varphi\right.$$
$$\left.+B_{n,m}\sin m\varphi)P_n^m(\cos\theta)\right).$$

To obtain surface harmonics, the following method is effective. Let ν be a direction proportional to the direction cosines l, m, n. Then a function

$$\left(l\frac{\partial}{\partial x}+m\frac{\partial}{\partial y}+n\frac{\partial}{\partial z}\right)\frac{1}{r}=\alpha\frac{\partial}{\partial\nu}\left(\frac{1}{r}\right),$$

$$\alpha=\sqrt{l^2+m^2+n^2}$$

is a solution of Laplace's equation. Physically, this corresponds to a †potential of double pole with moment α and direction ν. A more gen-

eral multipole potential

$$V=c\left(\prod_{i=1}^{n}\alpha_i\frac{\partial}{\partial v_i}\right)\left(\frac{1}{r}\right)$$

also satisfies Laplace's equation. If we put $V=U_n(x,y,z)r^{-2n-1}$, U_n is a spherical function of order n (**Maxwell's theorem**). Various spherical functions correspond to particular directions v_i. For example, if every v_i is equal to z, we have zonal harmonics; and if $n-m$ of the v_i are equal to z and m of the v_i are symmetric on the xy-plane, we obtain tesseral harmonics. Let γ be an angle between two segments connecting the origin to the points (r,θ,φ) and (r',θ',φ') in polar coordinates. Then $\cos\gamma=\cos\theta\cos\theta'+\sin\theta\sin\theta'\cos(\varphi-\varphi')$, and if we choose the line connecting the origin to a point (r',θ',φ') as the axis defining P_n, we have

$$P_n(\cos\gamma)=(-1)^n\frac{r^{n+1}}{n!}\left(\frac{x'}{r'}\frac{\partial}{\partial x}+\frac{y'}{r'}\frac{\partial}{\partial y}+\frac{z'}{r'}\frac{\partial}{\partial z}\right)^n\left(\frac{1}{r}\right).$$

These are called **biaxial spherical harmonics**, which can also be represented (by the addition theorem) by means of spherical harmonics with respect to each axis.

E. Extension of the Legendre Functions

We extend the associated functions with positive integer m to any number m. First, if m is a negative integer, we put

$$P_\nu^{-m}(z)=(1-z^2)^{-m/2}\int_1^z d\zeta_m\int_1^{\zeta_m}d\zeta_{m-1}\cdots\times\int_1^{\zeta_3}d\zeta_2\int_1^{\zeta_2}P_\nu(\zeta_1)d\zeta_1,$$

$$Q_\nu^{-m}(z)=(1-z^2)^{-m/2}\int_\infty^z d\zeta_m\int_\infty^{\zeta_m}d\zeta_{m-1}\cdots\times\int_\infty^{\zeta_3}d\zeta_2\int_\infty^{\zeta_2}Q_\nu(\zeta_1)d\zeta_1,$$

a definition due to Ferrers. Then

$$P_\nu^{-m}(z)=(-1)^m\frac{\Gamma(\nu-m+1)}{\Gamma(\nu+m+1)}P_\nu^m(z),$$

$$Q_\nu^{-m}(z)=(-1)^m\frac{\Gamma(\nu-m+1)}{\Gamma(\nu+m+1)}Q_\nu^m(z).$$

When we use the definition due to Heine and Hobson, the factor $(-1)^m$ in these formulas is excluded. Two fundamental solutions, called **hypergeometric functions** of the **hyperspherical differential equation**

$$(1-z^2)d^2w/dz^2-2(\mu+1)z\,dw/dz+(\nu-\mu)(\nu+\mu+1)w=0,$$

are

$$P_{\nu-\mu}^{(\mu,\mu)}(z)=\frac{e^{-\nu\pi i}}{2^{\nu-\mu}4\pi\sin\nu\pi}\times\oint^{(z+,1+,z-,1-)}\frac{(\zeta^2-1)^\nu}{(\zeta-z)^{\nu+\mu+1}}d\zeta,$$

$$Q_{\nu-\mu}^{(\mu,\mu)}(z)=\frac{e^{(1+\nu)\pi i}}{2^{\nu-\mu}4i\sin\nu\pi}\oint^{(-1+,1-)}\frac{(\zeta^2-1)^\nu}{(\zeta-z)^{\nu+\mu+1}}d\zeta,$$

where the contour of integration for the latter integral is a curve encircling the point -1 once in the positive direction and the point 1 once in the negative direction. Then the associated Legendre functions for an arbitrary number μ are defined as follows:

$$P_\nu^\mu(z)=\frac{\Gamma(\nu+\mu+1)}{2^\mu\Gamma(\nu+1)}(z^2-1)^{\mu/2}P_{\nu-\mu}^{(\mu,\mu)}(z),$$

$$Q_\nu^\mu(z)=\frac{\Gamma(\nu+\mu+1)}{2^\mu\Gamma(\nu+1)}(z^2-1)^{\mu/2}Q_{\nu-\mu}^{(\mu,\mu)}(z).$$

If $\nu-\mu$ is a positive integer, $P_{\nu-\mu}^{(\mu,\mu)}$ is called the **Gegenbauer polynomial**, also denoted by $C_{\nu-\mu}(z)$. The $C_{\nu-\mu}(z)$ are obtained as coefficients of the expansion of the generating function $(1-2hz+z^2)^{-(2\mu+1)/2}$ (→ Appendix A, Table 20.I).

For spherical functions of several variables there is an investigation by P. Appell and J. Kempé de Fériet [2] (→ 206 Hypergeometric Functions D).

References

[1] E. W. Hobson, The theory of spherical and ellipsoidal harmonics, Cambridge Univ. Press, 1931 (Chelsea, 1955).
[2] P. Appell and J. Kempé de Fériet, Fonctions hypergéométriques et hypersphériques, polynomes d'Hermite, Gauthier-Villars, 1926.
[3] C. Müller, Spherical harmonics, Lecture notes in math. 7, Springer, 1967.
[4] T. MacRobert, Spherical harmonics, Snedden, third revised edition, 1967.
Also → references to 206 Hypergeometric Functions, 389 Special Functions.

394 (XIII.13) Stability

A. General Remarks

Stability was originally a concept concerned with stationary physical states. When a state is affected by a small disturbance, this state is said to be **stable** if the disturbance subsequently remains small, and **unstable** if the dis-

turbance gradually increases. For instance, consider a rod placed in the Earth's gravitational field with one end fixed at a point around which the rod can rotate freely. When the rod is placed vertically, this state is stationary. It is stable if the rod is hanging down from the fixed end, and unstable if it is standing on the fixed end. In physical systems only the stable state is practically realizable, so this distinction is important.

The concept of stability is used not only in relation to physical states but also in many other fields of science. We shall restrict ourselves to stability of solutions of differential equations. There, the term *stability* is used in the sense that a small change in the initial values results in a small change in the solution. As long as the solution is considered within a finite interval of the independent variable, this stability is naturally guaranteed by the continuity of the solution with respect to its initial values (→ 316 Ordinary Differential Equations (Initial Value Problems)). The problem arises when an independent variable moves over an unbounded interval.

Let $(x_1, \ldots, x_n) = \mathbf{x}$, $(x_1(t), \ldots, x_n(t)) = \mathbf{x}(t)$, $(x_1'(t), \ldots, x_n'(t)) = \mathbf{x}'(t)$ (the symbol $'$ means differentiation by t), and $|\mathbf{x}| = \sum_{j=1}^n |x_j|$. Consider the differential equation

$$\mathbf{x}' = \mathbf{f}(t, \mathbf{x}), \tag{1}$$

for which the existence and uniqueness of the solution of the initial value problem is assumed for $|t| < \infty$, $|\mathbf{x}| < \infty$. Let $\mathbf{x} = \varphi(t)$ be a solution of (1). If for any $\varepsilon > 0$ and t_0, a $\delta > 0$ can be chosen so that $|\mathbf{x}(t_0) - \varphi(t_0)| < \delta$ implies $|\mathbf{x}(t) - \varphi(t)| < \varepsilon$ for $t_0 \leqslant t < \infty$ $(-\infty < t \leqslant t_0)$, where $\mathbf{x}(t)$ is any solution of (1), then $\mathbf{x} = \varphi(t)$ is said to be **(Lyapunov) stable in the positive (negative) direction**. If it is stable both in the positive and negative directions, it is said to be **stable in both directions**. In the remainder of this article we will consider stability in the positive direction only. Corresponding assertions for stability in the negative direction can be obtained by reversing the sign of t.

B. Classification

We denote by $\mathbf{x} = \mathbf{x}(t, t_0, \mathbf{x}_0)$ a solution of (1) such that $\mathbf{x} = \mathbf{x}_0$ at $t = t_0$.

Suppose a solution $\mathbf{x} = \varphi(t)$ is stable. If for any t_0 there exists a $\zeta > 0$ such that $|\mathbf{x}(t, t_0, \mathbf{x}_0) - \varphi(t)| \to 0$ as $t \to \infty$ for any $x(t, t_0, x_0)$ with $|x_0 - \varphi(t_0)| < \zeta$, $\varphi(t)$ is said to be **asymptotically stable**.

If a constant δ in the definition of stability can be chosen independently of t_0, $\varphi(t)$ is said to be **uniformly stable**. When equation (1) is †autonomous, stability implies uniform stability.

If (1) $\varphi(t)$ is uniformly stable and (2) for any t_0 and $\eta > 0$, there exist a $\zeta > 0$ independent of t_0 and a $T > 0$ independent of t_0 such that $|\mathbf{x}_0 - \varphi(t_0)| < \zeta$ and $t > t_0 + T$ imply $|\mathbf{x}(t, t_0, \mathbf{x}_0) - \varphi(t)| < \eta$, then $\varphi(t)$ is said to be **uniformly asymptotically stable**.

Suppose that there exists a positive number λ with the following property: For any t_0 and $\varepsilon > 0$ one can take a $\delta(\varepsilon) > 0$ such that $|\mathbf{x}_0 - \phi(t_0)| < \delta(\varepsilon)$ implies $|\mathbf{x}(t, t_0, \mathbf{x}_0) - \phi(t)| < \varepsilon e^{-\lambda(t - t_0)}$ for $t \geqslant t_0$. Then $\phi(t)$ is said to be **exponentially stable**. Exponential stability implies uniform asymptotic stability.

C. Criteria

To deal with the stability of $\mathbf{x} = \varphi(t)$, we need consider only the case $\varphi(t) \equiv 0$, since the transformation $\mathbf{x} = \mathbf{y} + \varphi(t)$ reduces equation (1) to

$$\mathbf{y}' = \mathbf{f}(t, \mathbf{y} + \varphi(t)) - \mathbf{f}(t, \varphi(t)) = \mathbf{F}(t, \mathbf{y}),$$

$$\mathbf{F}(t, \mathbf{0}) \equiv \mathbf{0}, \tag{2}$$

and thus $\mathbf{x} = \varphi(t)$ is transformed into $\mathbf{y} \equiv \mathbf{0}$. If $\mathbf{F}$ is continuously differentiable with respect to y, (2) can be written in the form

$$\mathbf{y}' = \mathbf{F}_{\mathbf{y}}(t, \mathbf{0})\mathbf{y} + \mathbf{g}(t, \mathbf{y}), \qquad \mathbf{g}(t, \mathbf{y}) = o(|\mathbf{y}|).$$

The linear part of this equation,

$$\mathbf{y}' = \mathbf{F}_{\mathbf{y}}(t, \mathbf{0})\mathbf{y},$$

is called the **variational equation** for (1). So, in this section, we can state several criteria for stability of the null solution $\mathbf{y} \equiv \mathbf{0}$ of the equation

$$\mathbf{y}' = P(t)\mathbf{y} + \mathbf{g}(t, \mathbf{y}), \qquad |\mathbf{g}(t, \mathbf{y})| = o(|\mathbf{y}|). \tag{3}$$

(I) If (3) is linear (i.e., $\mathbf{g}(t, \mathbf{y}) \equiv \mathbf{0}$), then $\mathbf{y} \equiv \mathbf{0}$ is stable if and only if every solution of (3) is bounded as $t \to \infty$.

(II) If (3) is linear, uniform asymptotic stability implies exponential stability.

Let $f(t, \mathbf{y})$ be a function defined for $|\mathbf{y}| < \rho$, $t > \alpha$. If there exists a continuous function $w(\mathbf{y})$ such that $w(\mathbf{0}) = 0$, $w(\mathbf{y}) > 0$ $(\mathbf{y} \neq \mathbf{0})$, $f(t, \mathbf{y}) \geqslant w(\mathbf{y})$ $(|\mathbf{y}| < \rho, t > \alpha)$, then $f(t, \mathbf{y})$ is said to be **positive definite**. If $-f(t, \mathbf{y})$ is positive definite, then $f(t, \mathbf{y})$ is said to be **negative definite**.

(III) The existence of a **Lyapunov function** $V(t, \mathbf{y})$ with the following properties implies the stability of $\mathbf{y} \equiv \mathbf{0}$: (i) $V(t, \mathbf{y})$ is positive definite and differentiable, (ii) $V(t, \mathbf{0}) = 0$, (iii) $\dot{V}(t, \mathbf{y}) = V_t + V_{\mathbf{y}} \cdot (P(t)\mathbf{y} + \mathbf{g}(t, \mathbf{y})) \leqslant 0$.

The existence of $V(t, \mathbf{y})$ with the following properties implies the uniform asymptotic stability of $\mathbf{y} \equiv \mathbf{0}$: (i) same as (i) above, (ii) there exists a continuous function $v(\mathbf{y})$ such that

$v(\mathbf{0})=0$, $v(\mathbf{y})>0$ ($\mathbf{y}\neq\mathbf{0}$), $V(t,\mathbf{y})\leqslant v(\mathbf{y})$, (iii) $\dot{V}(t,\mathbf{y})$ is negative definite.

Hereafter we shall assume that $|\mathbf{g}(t,\mathbf{y})|=o(|\mathbf{y}|)$ as $\mathbf{y}\to\mathbf{0}$ uniformly with respect to t.

(IV) If $P(t)$ is a constant matrix all of whose †eigenvalues have negative real parts, then $\mathbf{y}\equiv\mathbf{0}$ is asymptotically stable [3,4].

(V) Let $P(t)$ be continuous and periodic with period T and Z be a †fundamental system of solutions of the variational equation

$$\mathbf{z}'=P(t)\mathbf{z}. \tag{4}$$

Then there exists a constant matrix C such that $Z(t+T)=Z(t)C$. Let $\lambda_1,\ldots,\lambda_n$ be the eigenvalues of C. Then the numbers $\mu_k=(\log\lambda_k)/T$ ($k=1,\ldots,n$) are called the **characteristic exponents** of (4). Obviously they are determined up to integral multiples of $2\pi i/T$. If the real parts of the characteristic exponents are all negative, then $\mathbf{y}\equiv\mathbf{0}$ is asymptotically stable [3,4].

(VI) If $\mathbf{f}(t,\mathbf{x})$ in (1) is periodic in t with period T and (1) admits a periodic solution $\mathbf{x}=\varphi(t)$ with period T, then (1) can be reduced to (3) by putting $\mathbf{x}=\mathbf{y}+\varphi(t)$, and $P(t)$, $\mathbf{g}(t,\mathbf{y})$ are both periodic in t with period T. Thus criterion (V) can be applied as a stability criterion for the periodic solution of (1). There are many other criteria for various particular forms of the equation (→ 290 Nonlinear Oscillation). For the †autonomous case where $\mathbf{f}(t,\mathbf{x})$ is of the form $\mathbf{p}(\mathbf{x})$ or $\mathbf{p}(\mathbf{x})+\mathbf{q}(t)$ with $\mathbf{q}(t+T)=\mathbf{q}(t)$, many results have been found.

(VII) If the solution $\mathbf{z}\equiv\mathbf{0}$ of

$$\mathbf{z}'=P(t)\mathbf{z}$$

is uniformly asymptotically stable, then the solution $\mathbf{y}\equiv\mathbf{0}$ of (3) is also uniformly asymptotically stable [4].

D. Conditional Stability

Let $\varphi(t)$ be a solution and $\mathfrak{F}$ a family of solutions of (1). If for any $\varepsilon>0$, a $\delta>0$ can be determined so that $|\mathbf{x}(t_0)-\varphi(t_0)|<\delta$ implies $|\mathbf{x}(t)-\varphi(t)|<\varepsilon$ for $t_0\leqslant t<\infty$ for any solution $\mathbf{x}(t)$ in $\mathfrak{F}$, then $\varphi(t)$ is said to be stable with respect to the family $\mathfrak{F}$. If a family $\mathfrak{F}$ can be found so that a solution is stable with respect to $\mathfrak{F}$, the solution is said to be **conditionally stable**. For instance, in equation (3), if $P(t)$ is a constant matrix some of whose eigenvalues have negative real parts, $\mathbf{g}(t,\mathbf{y})$ is differentiable with respect to $\mathbf{y}$, and $\mathbf{g}_y(t,\mathbf{y})=o(1)$ uniformly in t as $\mathbf{y}\to 0$, then $\mathbf{y}\equiv 0$ is conditionally stable.

We now mention a weaker kind of stability called orbital stability. Let $\varphi(t)$ be a solution and ε any positive number. If there can be found a positive number δ such that for any solution $\mathbf{x}(t)$ with $|\mathbf{x}(t_1)-\varphi(t_0)|<\delta$ for some t_0 and t_1, $\bigcup_{t_1\leqslant t<\infty}\mathbf{x}(t)$ belongs to the ε-neighborhood of $\bigcup_{t_0\leqslant t<\infty}\varphi(t)$, then $\varphi(t)$ is said to have **orbital stability**.

When $\mathbf{f}(t,\mathbf{x})$ in equation (1) is independent of t, (1) is often called a †dynamical system. In the theory of dynamical systems, not merely the stability of a solution itself but also the stability of a closed invariant set is of importance (→ 126 Dynamical Systems).

It is also of importance to investigate the change in solution caused by a small change in the right-hand member of the equation. Suppose, for instance, that the right-hand member of the equation depends continuously on a parameter ε. Then the question arises as to how the solution changes if ε changes. The theory of such problems is called †perturbation theory. Suppose that the equation

$$\mathbf{x}'=\mathbf{f}(t,\mathbf{x},\varepsilon)$$

admits a periodic solution $\varphi(t)$ for $\varepsilon=0$. Then $\varphi(t)$ is said to be stable under perturbation if for $\varepsilon\neq 0$ the same equation admits a periodic solution lying near $\varphi(t)$. In †celestial mechanics and †nonlinear oscillation theory this concept plays an important role.

References

[1] R. E. Bellman, Stability theory of differential equations, McGraw-Hill, 1953.
[2] L. Cesari, Asymptotic behavior and stability problems in ordinary differential equations, Springer, third edition, 1971.
[3] E. A. Coddington and N. Levinson, Theory of ordinary differential equations, McGraw-Hill, 1955.
[4] S. Lefschetz, Differential equations, geometric theory, Interscience, 1957.
[5] A. Lyapunov (Ljapunov), Problème général de la stabilité du mouvement, Ann. Math. Studies, Princeton Univ. Press, 1947.
[6] V. V. Nemytskiĭ (Nemyckiĭ) and V. V. Stepanov, Qualitative theory of differential equations, Princeton Univ. Press, 1960. (Original in Russian, 1947.)
[7] O. Perron, Über eine Matrixtransformation, Math. Z., 32 (1930), 465–473.
[8] O. Perron, Die Stabilitätsfrage bei Differentialgleichungen, Math. Z., 32 (1930), 703–728.
[9] A. Halanay, Differential equations, stability, oscillations, time lags, Academic Press, 1966.
[10] T. Yoshizawa, Stability theory by Ljapunov's second method, Publ. Math. Soc. Japan, 1966.

395 (XVII.12) Stationary Processes

A. Definitions

Stationary process is a general name given to all †stochastic processes (→ 407 Stochastic Processes) that have the property of being stationary (to be defined in the next paragraph) under a shift of a time parameter t that extends over T, which is either the set of all real numbers $\mathbf{R}$ (a continuous parameter) or the set of all integers $\mathbf{Z}$ (a discrete parameter).

Let $(\Omega, \mathfrak{B}, P)$ be a †probability space and $\{X_t(\omega)\}$ $(t \in T, \omega \in \Omega)$ a complex-valued †stochastic process. If for every n, every $t_1, t_2, \ldots, t_n \in T$, and every †Borel subset E_n of complex n-dimensional space $\mathbf{C}^n$, the equality

$$P((X_{t_1+t}, \ldots, X_{t_n+t}) \in E_n) = P((X_{t_1}, \ldots, X_{t_n}) \in E_n) \qquad (1)$$

holds, then $\{X_t\}$ is called a **strongly** (or **strictly**) **stationary process**; while if $E(|X_t|^2)$ is finite for every t, and if the †moments up to the second order are stationary, i.e., if

$$E(X_{t+h}) = E(X_t),$$
$$E(X_{t+h}\bar{X}_{s+h}) = E(X_t\bar{X}_s), \qquad (2)$$

then $\{X_t\}$ is called a **weakly stationary process** or a **stationary process in the wider sense**. The "stationary" in the latter sense obviously includes the former if $E(|X_t|^2) < \infty$. Condition (2) is equivalent to

$$E(X_t) = m \quad \text{(a constant independent of } t\text{)},$$
$$E((X_t - m)\overline{(X_s - m)}) = \rho(t-s) \quad \text{(a function of } t-s\text{)}. \qquad (3)$$

We call m and $\rho(t)$ the **mean** and the **covariance function** of $\{X_t\}$.

In the continuous parameter case, we assume †continuity in probability,

$$\lim_{h \to 0} P(|X_{t+h} - X_t| > \varepsilon) = 0, \quad \varepsilon > 0,$$

for a strongly stationary process, and continuity in mean square,

$$\lim_{h \to 0} E(|X_{t+h} - X_t|^2) = 0,$$

for a weakly stationary process. The latter assumption is equivalent to continuity of the covariance function $\rho(t)$.

A †Gaussian process is strongly stationary if and only if it is weakly stationary; and so it is simply called a †stationary Gaussian process. Such processes constitute a typical class of stationary processes (→ 176 Gaussian Processes C).

B. Spectral Decomposition of Weakly Stationary Processes

The covariance function $\rho(t)$ is obviously †positive definite and continuous. Therefore, by †Bochner's theorem, we have the †spectral decomposition of $\rho(t)$:

$$\rho(t) = \int_{T'} e^{i\lambda t} F(d\lambda), \qquad (4)$$

where T' is either $\mathbf{R}$ (when $T = \mathbf{R}$) or $[-\pi, \pi]$ (when $T = \mathbf{Z}$) and F is a bounded measure on T'. The decomposition (4) is called the **Khinchin decomposition** of $\rho(t)$, and $F(d\lambda)$ is called the **spectral measure**. If the process $\{X_t\}$ is real-valued, then the spectral measure $F(d\lambda)$ is symmetric with respect to the origin.

To obtain the spectral decomposition of a weakly stationary process X_t itself, we introduce the †Hilbert space $L_2(\Omega)$ (where $\Omega = \Omega(\mathfrak{B}, P)$ is the basic probability space on which each X_t is regarded as a †square integrable function). Let $\mathfrak{M}(X)$ be the subspace of $L_2(\Omega)$ spanned by the X_t $(t \in T)$ and the constant function 1. Since $\{X_t\}$ is weakly stationary, we can define a one-parameter group of †unitary operators U_t $(t \in T)$ determined by $U_t X_s = X_{t+s}$ and $U_t 1 = 1$. By †Stone's theorem we have the spectral decomposition of U_t:

$$U_t = \int_{T'} e^{i\lambda t} E(d\lambda). \qquad (5)$$

Setting $M(\Lambda) = E(\Lambda)(X_0 - m)$, we obtain the **spectral decomposition** of X_t:

$$X_t = U_t X_0 = m + \int_{T'} e^{i\lambda t} M(d\lambda). \qquad (6)$$

We also have

$$(M(\Lambda), 1) = 0,$$
$$(M(\Lambda_1), M(\Lambda_2)) = F(\Lambda_1 \cap \Lambda_2). \qquad (7)$$

The study of weakly stationary processes is based on the decomposition (6). For example, the **weak law of large numbers** for $\{X_t\}$,

$$\operatorname*{l.i.m.}_{B-A \to \infty} \frac{1}{B-A} \int_A^B X_t \, dt = m + M(\{0\}), \qquad (8)$$

is an immediate consequence of (6). In the discrete parameter case a similar result is obtained by replacing the integral sign in expression (8) by the summation sign. In particular, if F is continuous at the origin, we have $M(\{0\}) = 0$, and only the constant m remains in the right-hand side of (8) [1, 2].

C. Weakly Stationary Random Distributions

Just as we introduce †distributions as generalizations of ordinary functions, we define

weakly stationary random distributions as generalizations of weakly stationary processes. Let $\mathscr{D}$ be the space of all functions of class C^∞ on $T=\mathbf{R}$ with compact support. We introduce the same topology on $\mathscr{D}$ as in the theory of distributions. If a random variable $X_\varphi \in L_2(\Omega)$ is defined for every $\varphi \in \mathscr{D}$ and the mapping $\varphi \to X_\varphi$ is continuous in the L_2-sense and linear, then the family $\{X_\varphi\}$ of random variables is called a **random distribution in the wider sense** (→ 407 Stochastic Processes). Furthermore, if

$$(X_{\tau_h\varphi}, 1)=(X_\varphi, 1),$$

$$(X_{\tau_h\varphi}, X_{\tau_h\psi})=(X_\varphi, X_\psi) \tag{9}$$

for every $h\in\mathbf{R}^1$, where (,) stands for the inner product in $L_2(\Omega)$ and

$$\tau_h\varphi(t)=\varphi(t-h),$$

then $\{X_\varphi\}$ is said to be a **weakly stationary random distribution**. With a weakly stationary process we can associate a weakly stationary random distribution by the relation

$$X_\varphi=\int_T X_t\varphi(t)\,dt. \tag{10}$$

This correspondence is one-to-one, and therefore we can identify $\{X_t\}$ with $\{X_\varphi\}$ as we identify an ordinary function with a distribution. From equations (9) it follows that there exist a constant m and a distribution ρ such that $E(X_\varphi)=m\int\varphi(t)\,dt$ and $E(X_\varphi - E(X_\varphi))\overline{(X_\psi-E(X_\psi))}=\rho(\varphi*\check{\psi})$, where $*$ denotes convolution, and $\check{\psi}(t)=\overline{\psi(-t)}$. We call m and ρ the **mean value** and **covariance distribution** of $\{X_\varphi\}$, respectively. By the generalized Bochner theorem ρ can be expressed in the form

$$\rho(\varphi)=\int\hat{\varphi}(\lambda)F(d\lambda), \quad \hat{\varphi}(\lambda)=\int e^{i\lambda t}\varphi(t)\,dt,$$

where $F(d\lambda)$ is a slowly increasing measure, i.e.,

$$\int(1+\lambda^2)^{-k}F(d\lambda)<\infty \tag{11}$$

for some positive integer k. $F(d\lambda)$ is called the **spectral measure**. This expression is the generalization of the Khinchin decomposition. The spectral decomposition corresponding to (6) and the †law of large numbers for X_φ can be discussed in a manner similar to that for weakly stationary processes (K. Itô [3]).

D. Prediction Theory

Let $\{X_t\}$ be a weakly stationary process. Suppose that its values X_s $(s\leqslant t)$ up to time t are observed. **Prediction theory** deals with the problem of forecasting the future value $X_{t+\tau}$ $(\tau>0)$ from the known values X_s $(s\leqslant t)$. If the domain of the admissible predictors is limited to linear functions of X_s $(s\leqslant t)$, the theory is called **linear prediction theory**. We can assume without loss of generality that the mean value m of X_t is zero and that the spectral measure $F(d\lambda)$ of $\{X_t\}$ is not a zero measure.

Let $\mathscr{M}_t(X)$ be the subspace of $L_2(\Omega)$ spanned by the X_s $(s\leqslant t)$, then $\mathscr{M}(X)=\bigvee_t\mathscr{M}_t(X)$. A **linear predictor** for $X_{t+\tau}$ from X_s $(s\leqslant t)$ is an element Y of $\mathscr{M}_t(X)$. If a linear predictor minimizes the prediction error $\sigma^2(\tau)=E(|X_{t+\tau}-Y|^2)$ in $\mathscr{M}_t(X)$, it is called an **optimum linear predictor**, which turns out to be the †(orthogonal) projection of $X_{t+\tau}$ on $\mathscr{M}_t(X)$ and which is denoted by $\hat{X}_{t,\tau}$. Since $\{X_t\}$ is stationary, the error $\sigma^2(\tau)$ does not depend on t for such a predictor. Corresponding to the spectral decomposition (6) of X_t, the optimum linear predictor is expressed in the form

$$\hat{X}_{t,\tau}=\int_{T'}e^{i\lambda t}\hat{\varphi}_\tau(\lambda)M(d\lambda), \tag{12}$$

where $\hat{\varphi}_\tau(\cdot)$ is square integrable with respect to the spectral measure $F(d\lambda)$.

The subspace $\mathscr{M}_t(X)$ is nondecreasing in t. If $\mathscr{M}_t(X)$ is independent of t, i.e., $\mathscr{M}_t(X)=\mathscr{M}(X)$ for every t, then $\{X_t\}$ is said to be **deterministic**. In this case we have $\hat{X}_{t,\tau}=X_{t+\tau}$ for every t and $\tau>0$, since $X_{t+\tau}\in\mathscr{M}_t(X)$. This means that the linear predictor enables us to determine the unknown quantities without error, and therefore such a process is of no probabilistic interest. On the other hand, if $\bigcap_t\mathscr{M}_t(X)=\{0\}$, then $\{X_t\}$ is said to be **purely nondeterministic**. A general $\{X_t\}$ is expressed as the sum of the deterministic part $\{X_t^d\}$ and the purely nondeterministic part $\{X_t^n\}$ (**Wold decomposition**). Furthermore, we have $\mathscr{M}(X^d)=\bigcap_t\mathscr{M}_t(X)$, and $\mathscr{M}_t(X)=\mathscr{M}(X^d)+\mathscr{M}_t(X^n)$ (direct sum). Thus $\{X_t^d\}$ and $\{X_t^n\}$ can be dealt with separately.

A weakly stationary process $\{X_t\}$ is purely nondeterministic if and only if the spectral measure $F(d\lambda)$ is absolutely continuous with respect to the Lebesgue measure, and the density $f(\lambda)$ is positive almost everywhere and satisfies

$$\int_{-\pi}^{\pi}\log f(\lambda)\,d\lambda>-\infty$$

(discrete parameter case),

$$\int_{-\infty}^{\infty}\frac{\log f(\lambda)}{1+\lambda^2}\,d\lambda>-\infty$$

(continuous parameter case).

By using $f(\lambda)$ the optimum linear predictor can be obtained.

First, we explain the discrete parameter case. There exists a function $\gamma(z)=\sum_{t=0}^{\infty}a_t z^t$ in the †Hardy class H_2 relative to the unit disk

such that its boundary value satisfies

$f(\lambda)=(1/2\pi)|\gamma(e^{-i\lambda})|^2.$

Then we can find a sequence of mutually orthogonal random variables $\{\xi_t\}$ $(t\in\mathbf{Z})$ such that $\{X_t\}$ admits a **backward moving average representation**

$$X_t=\sum_{s=-\infty}^{t} a_{t-s}\xi_s. \tag{13}$$

There are many pairs $\{a_t\}$ and $\{\xi_t\}$ which give the representation (13), but if $\gamma(z)$ is maximal (optimal), namely, if $\gamma(z)$ is expressed as

$$\gamma(z)=\sqrt{2\pi}\exp\left(\frac{1}{4\pi}\int_{-\pi}^{\pi}\log f(\lambda)\frac{e^{-i\lambda}+z}{e^{-i\lambda}-z}d\lambda\right), \tag{14}$$

then the representation (13) is **canonical** in the sense that $\mathscr{M}_t(X)=\mathscr{M}_t(\xi)$ for every t. Hence the optimum predictor $\hat{X}_{t,\tau}$ for $X_{t+\tau}$ is given by

$$\hat{X}_{t,\tau}=\sum_{s=-\infty}^{t} a_{t+\tau-s}\xi_s=\int_{-\pi}^{\pi} e^{i\lambda t}\varphi_\tau(\lambda)M(d\lambda), \tag{15}$$

where

$$\varphi_\tau(\lambda)=\frac{e^{i\lambda\tau}}{\gamma(e^{-i\lambda})}\left(\gamma(e^{-i\lambda})-\sum_{s=0}^{\tau-1}a_s e^{-i\lambda s}\right).$$

The prediction error $\sigma^2(\tau)$ of this predictor is given by

$$\sigma^2(\tau)=\sum_{s=0}^{\tau-1}|a_s|^2.$$

Example. Let the covariance function of a weakly stationary process $\{X_t\}$ be $e^{-\alpha|t|}$ $(\alpha>0)$. Then we have

$f(\lambda)=(1/2\pi)(1-\beta^2)|1-\beta e^{-i\lambda}|^{-2},$

$\beta=e^{-\alpha}.$

The maximal $\gamma(z)$ is expressed as $\sqrt{1-\beta^2}(1-\beta z)^{-1}$, and

$$\hat{X}_{t,\tau}=\int_{-\pi}^{\pi} e^{i\lambda t}\frac{e^{i\lambda\tau}}{1-\beta e^{-i\lambda}}\sum_{s=\tau}^{\infty}\beta^s e^{-i\lambda s}M(d\lambda)$$
$$=\beta^\tau X_t,$$

$$\sigma^2(\tau)=1-2\beta^\tau e^{-\alpha\tau}+\beta^{2\tau}=1-e^{-2\alpha\tau}.$$

We now come to the continuous parameter case. By replacing the holomorphic function $\gamma(z)$ on the unit disk with the one on the half-plane, we see that almost all results obtained in the discrete parameter case hold similarly in this case. The maximal $\gamma(z)$ is expressed as

$$\gamma(z)=\sqrt{\pi}\exp\left(\frac{1}{2\pi i}\int_{-\infty}^{\infty}\log f(\lambda)\frac{1+\lambda z}{z-\lambda}\frac{d\lambda}{1+\lambda^2}\right).$$

Using the †Fourier transform a_t of the boundary function of $\gamma(z)$ and a process $\{\xi_t\}$ with orthogonal increments, we have the **canonical backward moving average representation** for the process $\{X_t\}$:

$$X_t=\int_{-\infty}^{t} a_{t-s}d\xi_s,$$

which enables us to obtain the optimum predictor and the prediction error in a manner similar to the discrete parameter case [4]. In particular, if the optimal $\gamma(z)$ is of the form $\gamma(z)=c/P(iz)$, where c is a constant and $P(z)$ is a polynomial of degree p, then X_t is $p-1$ times differentiable and $P(d/dt)X_t=(d/dt)\xi_t$ up to a multiplicative constant; therefore $\hat{X}_{t,\tau}$ is obtained explicitly. To obtain the optimum linear predictor for $Y\in\mathscr{M}(X)$, we first establish the expression

$$Y=\sum_{s=-\infty}^{\infty} f(s)\xi_s \quad\text{or}\quad \int_{-\infty}^{\infty} f(s)d\xi_s,$$

and then take $\sum_{s=-\infty}^{0} f(s)\xi_s$ or $\int_{-\infty}^{0} f(s)d\xi_s$ for the optimum linear predictor.

The results stated above can be generalized to multivariate (n-dimensional) stationary processes [6, 7] and to the case where the parameter space T is multidimensional.

N. Wiener observed the individual †sample process $X(t,\omega)$ and discussed a method of finding the optimum predictor for $X(t+\tau,\omega)$ by using a linear functional

$$\int_0^{\infty} X(t-s,\omega)dK(s)$$

(K is of †bounded variation) of the values X_s $(s\leqslant t)$ [8]. The spectral measure played an important role in his observation. Calculations in this case are analogous to those mentioned above.

For a weakly stationary random distribution $\{X(\varphi)\}$ $(\varphi\in\mathscr{D})$, the prediction theory is reduced to that for ordinary stationary processes. Assume that the spectral measure $F(d\lambda)$ of $\{X(\varphi)\}$ satisfies (11). Set $e(t)=\exp(t)$ $(t\leqslant 0)$, $=0$ $(t>0)$, and let $e_k(t)$ be the k-times convolution of $e(t)$ with itself. Set $Y(\varphi)=X(e_k*\varphi)$. Then $\{Y(\varphi)\}$ is equivalent to a weakly stationary process. It is obvious that $\mathscr{M}_t(X)=\mathscr{M}_t(Y)$ for every t, where $\mathscr{M}_t(X)$ is the linear subspace spanned by $\{X(\varphi)\,|\,\text{support of }\varphi\subset(-\infty,t]\}$. This consideration allows us to reduce the prediction problem for $\{X(\varphi)\}$ to that for the stationary process corresponding to $\{Y(\varphi)\}$.

Nonlinear prediction theory is formulated as follows. Let $\mathfrak{B}_t$ be the smallest σ-algebra with respect to which every X_s $(s\leqslant t)$ is measurable and $H_t(X)$ be the subspace of $L_2(\Omega)$ consisting of all $\mathfrak{B}_t$-measurable elements. The problem is to forecast $X_{t+\tau}$ $(\tau>0)$ by using an element of $H_t(X)$. The optimum predictor is obviously equal to $E(X_{t+\tau}|\mathfrak{B}_t)$. For a stationary Gaussian process it has been proved that

the optimum predictor found in $H_t(X)$ belongs to $\mathcal{M}_t(X)$. Therefore the optimum nonlinear predictor coincides with the optimum linear predictor. However, except for stationary Gaussian processes, no systematic approach for nonlinear prediction theory has been established so far. (For a typical case that arises from a stationary Gaussian process → 176 Gaussian Processes H.)

E. Interpolation and Filtering

Interpolation and **filtering** of stationary processes have many similarities with prediction theory, both in the formulation of the problems and in their method of solution.

Let $\{X_t\}$ be a weakly stationary process, all of whose values $\{X_t \mid t \notin T_1\}$. T_1 some interval, are known with the exception of those at $t \in T_1$. The problem of linear interpolation of the unknown value X_t $(t \in T_1)$ is to find the best approximation of this random variable by the limit of linear combinations of the known values. The following example illustrates the problem in the discrete parameter case.

Example. Let $T_1 = \{t_0\}$ and $f(\lambda)d\lambda$ be the spectral measure of $\{X_t\}$. The interpolation of X_{t_0} has an error if and only if

$$\int_{-\pi}^{\pi} \frac{1}{f(\lambda)} d\lambda < \infty.$$

Expressing X_t in the form (6) with $m=0$, the best (linear) interpolation $\hat{X}_{t_0}$ of X_{t_0} is given by

$$\int_{-\pi}^{\pi} e^{i\lambda t_0}\left(1 - 2\pi\left(f(\lambda)\int_{-\pi}^{\pi}\frac{1}{f(\mu)}d\mu\right)^{-1}\right)M(d\lambda),$$

and the error of the interpolation is expressed by

$$E(|X_{t_0} - \hat{X}_{t_0}|^2) = 4\pi^2\left(\int_{-\pi}^{\pi}\frac{1}{f(\lambda)}d\lambda\right)^{-1}.$$

The problem of interpolation for multivariate stationary processes has also been discussed [7].

The filtering problem originated in communication theory as a technique to extract the relevant component from a received signal with noise [8, 13]. Suppose that a complex-valued stationary process $\{X_t\}$ with continuous parameter is expressed in the form

$$X_t = \int_{-\infty}^{\infty} e^{i\lambda t} M(d\lambda) = S_t + N_t,$$

where (S_t, N_t) is a (2-dimensional) weakly stationary process with mean vector 0. Here, S_t and N_t indicate the signal and noise, respectively. The filtering problem is to find the element of $\mathcal{M}_t(X)$ that approximates $S_{t+\tau}$ as closely (relative to the $\mathcal{M}(X)$-norm) as possible. The best approximation is the projection of $S_{t+\tau}$ on $\mathcal{M}_t(X)$, but its expression in terms of the spectral measure becomes extremely complicated [8]. This problem is usually discussed under the assumption that S_t and N_t are orthogonal. Let us further assume that their spectral measures are absolutely continuous. The density functions are denoted by $f_S(\lambda)$ and $f_N(\lambda)$, respectively. If $\{X_t \mid t \in T\}$ is observed, then the best (linear) approximation $\hat{S}_t$ of S_t is given by

$$\hat{S}_t = \int_{-\infty}^{\infty} e^{i\lambda t}\varphi_0(\lambda)M(d\lambda),$$

where $\varphi_0(\lambda) = f_S(\lambda)/(f_S(\lambda) + f_N(\lambda))$. The mean square error $E(|S_t - \hat{S}_t|^2)$ of this filtering is

$$\int_{-\infty}^{\infty} f_S(\lambda) f_N(\lambda)/(f_S(\lambda) + f_N(\lambda))\, d\lambda.$$

F. Strongly Stationary Processes and Flows

Let $\{X_t(\omega)\}$ $(t \in T,\ \omega \in \Omega(\mathfrak{B}, P))$ be a strongly stationary process. To study it we take the coordinate representation of $\{X_t\}$ as follows. Let Ω be the complex vector space $\mathbf{C}^T$, $\mathfrak{B}$ the σ-algebra generated by the Borel †cylinder sets, $X_t(\omega)$ the tth coordinate of the function $\omega \in \mathbf{C}^T$, and P the probability distribution of the process $\{X_t\}$ defined on $(\Omega, \mathfrak{B})$. Define the shift transformation S_t of Ω onto itself by $(S_t\omega)(s) = \omega(s+t)$. Then $\{S_t\}$ forms a group of †measure-preserving transformations on $\Omega(\mathfrak{B}, P)$ (→ 136 Ergodic Theory) since $\{X_t\}$ is strongly stationary. Thus we are given a (measure-preserving) †flow $\{S_t\}$ $(t \in T)$ on $\Omega(\mathfrak{B}, P)$. Conversely, if $\{S_t\}$ $(t \in T)$ is a (measure-preserving) flow on a probability space $\Omega(\mathfrak{B}, P)$, then $\{X_t\}$, given by $X_t(\omega) = f(S_t\omega)$, is a strongly stationary process, provided that f is measurable. Many properties of a strongly stationary process are closely related to those of the corresponding flow. For example, the †strong law of large numbers for a strongly stationary process follows from †Birkhoff's individual ergodic theorem for flows. †Ergodicity, several kinds of †mixing properties, and the spectral properties of a strongly stationary process are defined in accordance with the respective notions for the corresponding flow. Now we give some examples of flows corresponding to strongly stationary processes.

(1) If X_t $(t \in \mathbf{Z})$ are mutually independent and have the same probability distribution, then the process $\{X_t\}$ $(t \in \mathbf{Z})$ is strongly stationary and $\bigcap_t \mathfrak{B}_t$ is trivial (the definition of $\mathfrak{B}_t$ → D). Hence the corresponding flow is a †Kolmogorov flow.

(2) Similarly to (1), the flow corresponding to †Gaussian white noise (→ 176 Gaussian Process) is also a Kolmogorov flow.

(3) The mixing properties of the flow corresponding to a stationary Gaussian process is determined by the smoothness of its spectral measure $F(d\lambda)$. The flow is ergodic if and only if F is continuous (i.e., F has no point mass). In this case, the flow is also †weakly mixing. For the flow to be †strongly mixing, it is necessary and sufficient that the covariance function $\rho(t)$ of the process tend to zero as $|t|\to\infty$. In this case the flow is †mixing of all orders (→ 136 Ergodic Theory) [14–16].

G. Analytic Properties of Sample Functions of Stationary Processes

In the continuous parameter case, we always assume that †continuity in probability holds for strongly stationary processes and †mean square continuity holds for weakly stationary processes. Hence the processes discussed here are all continuous in probability, and without loss of generality we can assume that the stationary processes are †separable and †measurable (→ 407 Stochastic Processes).

Let $\{X_t\}$ be a weakly stationary process. Assume that the moments up to order $2n$ of the spectral measure $F(d\lambda)$ are all finite. Then almost all †sample functions of $\{X_t\}$ are $n-1$ times continuously differentiable, and almost all sample functions of $\{X_t^{(n-1)}\}$ are absolutely continuous. Define the spectral distribution function $F(\lambda)=F((-\infty,\lambda])$ for the spectral measure $F(d\lambda)$ of $\{X_t\}$. If F satisfies

$$\sum_{n=-\infty}^{\infty} |n|^r(F(n+1)-F(n))^{1/2}<\infty \tag{16}$$

for a nonnegative integer r, then almost all sample functions of $\{X_t\}$ have continuous rth derivatives. In particular, if the condition (16) is satisfied for $r=0$, then almost all sample functions are continuous. Conditions for Hölder continuity of almost all sample functions of a weakly stationary process have also been obtained [17, 18]. (For sample functions of stationary Gaussian processes → 176 Gaussian Processes F.)

For a strongly stationary process $\{X_t\}$ with $E(X_0)=0$ and finite $E(X_0^2)$, the **sample covariance function**

$$R(t)=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} X_{t+s}(\omega)\overline{X_s(\omega)}\,ds$$

is determined with probability 1. We can therefore apply the theory of †generalized harmonic analysis, due to Wiener [9]. (Further results on sample function properties are found in [19].)

H. Strongly Stationary Random Distributions

Let $\mathscr{D}$ be the space of all C^∞-functions with compact support and $\mathscr{D}'$ be the space of †distributions. If $X_\varphi(\omega)$ is defined for $\omega\in\Omega(\mathfrak{B},P)$ and $\varphi\in\mathscr{D}$, and if for almost all ω, $X_\varphi(\omega)$ belongs to $\mathscr{D}'$ as a linear functional of φ, then $\{X_\varphi\}$ is called a **random distribution**. Suppose that the joint distribution of the random variables $X_{\tau_h\varphi_1}, X_{\tau_h\varphi_2},\dots,X_{\tau_h\varphi_n}$ ($\tau_h\varphi(t)=\varphi(t-h)$) is independent of h. Then we call $\{X_\varphi(\omega)\}$ a **strongly** (or **strictly**) **stationary random distribution**. If we identify random distributions that have the same probability law, then $\{X_\varphi\}$ is determined by the characteristic functional

$$C(\varphi)=E(e^{iX_\varphi}).$$

For $\{X_\varphi\}$ to be strictly stationary it is necessary and sufficient that the equality $C(\tau_h\varphi)=C(\varphi)$ hold. The simplest example of a strictly stationary random distribution is the Gaussian white noise (→ 176 Gaussian Processes, 341 Probability Measures) [20].

I. Generalizations of Stationary Processes

The concept of stationary processes is generalized in many directions.

(1) Let T be a set different from $\mathbf{R}$ or $\mathbf{Z}$, and suppose that there is given a group G of transformations that map T onto itself. If a family $\{X_t\}$ of random variables with parameter $t\in T$ has the property that for every choice of random variables $X_{t_1}, X_{t_2},\dots,X_{t_n}$, the joint distribution of $(X_{gt_1},\dots,X_{gt_n})$ is always independent of $g\in G$, then $\{X_t\}$ $(t\in T)$ is said to be a **strictly G-stationary** system of random variables. Similarly, a **weakly G-stationary** system of random variables can be defined [21, 22].

(2) Let T be a Riemannian space, and let G be the group of all isometric transformations on T or one of its subgroups. Suppose that a †tensor field $\mathbf{X}_t(\omega)$ of constant rank is associated with any $\omega\in\Omega(\mathfrak{B},P)$ at every point t. Then $\mathbf{X}(\omega)\equiv\{\mathbf{X}_t(\omega)\mid t\in T\}$ is called a **random tensor field** over the Riemannian space T. Any $g\in G$ induces an isometric transformation of the tangent vector space at t to that at gt. Hence g maps a tensor field $\mathbf{X}(\omega)$ to another tensor field $g\mathbf{X}(\omega)$ for every ω. If $\mathbf{X}(\omega)$ and $g\mathbf{X}(\omega)$ have the same probability law, then $\mathbf{X}(\omega)$ is said to be **strictly G-stationary**. $X(\omega)$ is defined to be **weakly G-stationary** in a similar way [21, 22].

(3) In the same way as we extended stochastic processes to random distributions, we can generalize random tensor fields to **random currents** and discuss stationary random currents [21].

(4) **Stochastic process with stationary incre-**

ments of order n. Assume that $\{X_t\}$ ($t \in \mathbf{R}$) is not necessarily a stationary process but that the nth-order increment of X_t is stationary. Then by taking the nth derivative $D^n X_t$ in the sense of random distributions, we obtain a stationary random distribution. From the properties of $D^n X_t$ we can investigate the original process itself. Brownian motion is an example of a stochastic process with stationary increments of order 1.

(5) **Weakly stationary processes of degree** k. A weakly stationary process is a process whose moments up to order 2 are stationary. Generalizing this, we can define a weakly stationary process of degree k by requiring the moments up to order k to be stationary. We can obtain more detailed properties of such processes than those of weakly stationary processes [23].

References

[1] J. L. Doob, Stochastic processes, Wiley, 1953.
[2] U. Grenander and M. Rosenblatt, Statistical analysis of stationary time series, Wiley, 1957.
[3] K. Itô, Stationary random distributions, Mem. Coll. Sci. Univ. Kyôto, (A) 28 (1954), 209–223.
[4] K. Karhunen, Über die Struktur stationärer zufälliger Funktionen, Ark. Mat., 1 (1950), 141–160.
[5] H. Cramér, On the linear prediction problem for certain stochastic processes, Ark. Mat., 4 (1963), 45–53.
[6] N. Wiener and P. Masani, The prediction theory of multivariate stochastic processes I, II, Acta Math., 98 (1957), 111–150; 99 (1958), 93–137.
[7] Yu. A. Rozanov, Stationary random processes, Holden-Day, 1967. (Original in Russian, 1963.)
[8] N. Wiener, Extrapolation, interpolation, and smoothing of stationary time series, MIT Press, 1949.
[9] N. Wiener, Generalized harmonic analysis, Acta Math., 55 (1930), 117–258.
[10] N. Wiener, The homogeneous chaos, Amer. J. Math., 60 (1938), 897–936.
[11] N. Wiener, Nonlinear problems in random theory, MIT Press, 1958.
[12] P. Masani and N. Wiener, Non-linear prediction, in Probability and Statistics: The H. Cramér Volume, Wiley, 1959, 190–212.
[13] A. Blanc-Lapierre and R. Fortet, Théorie des fonctions aléatoires, Masson, 1953.
[14] G. Maruyama, The harmonic analysis of stationary stochastic processes, Mem. Fac. Sci. Kyushu Univ., (A) 4 (1949), 45–106.
[15] S. Kakutani, Spectral analysis of stationary Gaussian processes, Proc. 4th Berkeley Symp. Math. Stat. Prob. II, Univ. of California Press (1961), 239–247.
[16] H. Totoki, The mixing property of Gaussian flows, Mem. Fac. Sci. Kyushu Univ., (A) 18 (1964), 136–139.
[17] I. Kubo, On a necessary condition for the sample path continuity of weakly stationary processes, Nagoya Math. J., 38 (1970), 103–111.
[18] T. Kawata and I. Kubo, Sample properties of weakly stationary processes, Nagoya Math. J., 39 (1970), 7–21.
[19] H. Cramér and M. R. Leadbetter, Stationary and related processes, Wiley, 1967.
[20] I. M. Gel'fand and N. Ya. Vilenkin, Generalized functions IV, Academic Press, 1964. (Original in Russian, 1961.)
[21] K. Itô, Isotropic random current, Proc. 3rd Berkeley Symp. Math. Stat. Prob. II, Univ. of California Press (1956), 125–132.
[22] A. M. Yaglom, Second-order homogeneous random fields, Proc. 4th Berkeley Symp. Math. Stat. Prob. II, Univ. of California Press (1961), 593–622.
[23] A. N. Shiryaev, Some problems in the spectral theory of higher-order moments I, Theory Prob. Appl., 5 (1960), 265–284. (Original in Russian, 1960.)

396 (XVIII.3) Statistic

A. General Remarks

A statistic is a function of a value (i.e., a sample value) observed in the process of statistical inference (→ 401 Statistical Inference). A statistic is used for two purposes: (a) to characterize the set of observed values or sample values, and (b) to summarize the information contained in the sample about the unknown parameters of the population from which it is assumed to have been drawn.

B. Samples and Statistics

The basic concepts in statistical inference are †population and †sample. Let $(\Omega, \mathscr{B}, P)$ be a †probability space, where P is a †probability measure on $\mathscr{B}$. A †random variable X defines a 1-dimensional probability distribution $\Phi(A) = P\{\omega \mid X(\omega) \in A\}$, where A is a 1-dimensional †Borel set, which gives rise to a 1-dimensional probability space $(\mathbf{R}, \mathscr{B}^1, \Phi)$. Here $\mathscr{B}^1$ is the

family of all 1-dimensional †Borel sets. Let $X_1, X_2, \ldots, X_n$ be †independent random variables with identical 1-dimensional distributions. The †n-dimensional random variable $X=(X_1, X_2, \ldots, X_n)$ is called a **random sample** of **size** n from the population $(\Omega, \mathscr{B}, P)$. In particular, when each of $X_1, \ldots, X_n$ takes only two values (usually 0 and 1), the sample is called a **Bernoulli sample** or a **sequence of Bernoulli trials**. Generally, if Φ_n is the †n-dimensional probability distribution determined by X (i.e., the direct product of n copies of the 1-dimensional probability distribution Φ), then the n-dimensional probability space $(\mathbf{R}^n, \mathscr{B}^n, \Phi_n)$, where $\mathscr{B}^n$ is the family of n-dimensional Borel sets, is called an **n-dimensional sample space**. A point belonging to the set of actually observed values of the sample X, which is a random variable by definition, is called a **sample value** and is denoted by x. Thus the sample value can be expressed as $x=X(\omega)$ $(\omega \in \Omega)$ and regarded as a point in the sample space (**sample point**). The basic underlying structure which determines the probability distribution is the set Ω, which we can view as describing the physical structure of the observed phenomena, but statistical procedures are always carried out through the observations of samples, and Ω itself is often disregarded. The 1-dimensional probability distribution Φ (the n-dimensional probability distribution Φ_n determined by X) is called the **population distribution** in the 1-dimensional (n-dimensional) sample space, since it is induced from the probability measure on $(\Omega, \mathscr{B})$.

A **statistic** Y is a random variable expressed as $Y=f(X)$, where f is a †measurable function from the sample space $(\mathbf{R}^n, \mathscr{B}^n, \Phi_n)$ into a measurable space $(\mathbf{R}, \mathscr{B}^1)$. The value of the statistic Y corresponding to a sample value x of the sample X is denoted by $y=f(x)$.

When we deal with a statistical problem we often have no exact knowledge of the population distribution $\Phi(\Phi_n)$ except that it belongs to a family $\mathscr{P}=\{P_\theta \mid \theta \in \Theta\}$ of probability measures on $\mathscr{B}^1(\mathscr{B}^n)$. We call θ the **parameter** of the probability distribution and Θ the **parameter space**. The typical cases described in this section can be extended as follows: (1) The distribution Φ may be an r-dimensional probability distribution. In this case a sample of size n induces an nr-dimensional sample space. (2) Random variables $X_1, \ldots, X_n$, being mutually independent, may not have identical distributions. (3) Random variables $X_1, \ldots, X_n$ may not be mutually independent. In both cases (2) and (3) the sample space is of the form $(\mathbf{R}^n, \mathscr{B}^n, \Phi_n)$, but n may not be the sample size itself, nor Φ_n be the direct product of n copies of identical 1-dimensional components. (4) The most general sample space is expressed as a certain measurable space $(\mathscr{X}, \mathscr{A})$ and a family $\mathscr{P}=\{P_\theta \mid \theta \in \Theta\}$ of probability measures on $\mathscr{A}$.

A **statistic**, in general, is a random variable expressed as $Y=f(X)$ by a measurable function f defined on a sample space $(\mathscr{X}, \mathscr{A})$ taking values in another meaurable space $(\mathscr{Y}, \mathscr{C})$. When $(\mathscr{Y}, \mathscr{C})$ is $(\mathbf{R}, \mathscr{B}^1)$ or $(\mathbf{R}^n, \mathscr{B}^n)$, $Y=f(X)$ is accordingly called a **1-dimensional** or **n-dimensional statistic**.

C. Population and Sample Characteristics in the 1-Dimensional Case

In a 1-dimensional probability space $(\mathbf{R}, \mathscr{B}^1, P_0)$ the following quantities, called **population characteristics**, are used to characterize the population distribution P_0: Letting $F(z)=P_0((-\infty, z])$ be the †distribution function of P_0, we use the **population mean** $\mu=\int z\,dF(z)$; the **population variance** $\sigma^2=\int(z-\mu)^2\,dF(z)$; the **population standard deviation** $\sigma(\geqslant 0)$; the **population moment of order** k $\mu'_k=\int z^k\,dF(z)$ $(\mu'_1=\mu)$; the **kurtosis** μ_4/σ^4; the **coefficient of excess** $(\mu_4/\sigma^4)-3$; the **skewness** μ_3/σ^3; the **α-quantile** or **$100\alpha\%$-point** m satisfying $F(m-0) \leqslant \alpha \leqslant F(m+0)$; the **median**, which is the 50%-point; the **first** and **third quartiles**, which are the 25%-point and 75%-point, respectively; the **range**, which is the third quartile minus the first quartile; and the **mode**, which is the value or values of z for which $dF(z)/dz$ attains its maximum.

Sometimes the kurtosis and others are called **population kurtosis**, etc. Here the word "population" is used when it is desirable to distinguish population characteristics from the sample characteristics defined in 341 Probability Measures.

Let $x=(x_1, \ldots, x_n)$ be a point of an n-dimensional sample space (a sample value). Corresponding to each 1-dimensional Borel set A, the number of components of x that belong to A is called the **frequency** of A in the sample value $x=(x_1, \ldots, x_n)$, and (frequency)/n is called the **relative frequency** of A. If we take $A=(-\infty, z]$ and regard its relative frequency $F_x(z)$ as a function of z, it becomes a †distribution function for every $x \in \mathbf{R}^n$, called the **empirical distribution function** based on x.

Various characteristics can be defined from the empirical distribution function in exactly the same way as population characteristics are derived from a population distribution function. These are called **sample characteristic values** and can be expressed as functions of $x_1, \ldots, x_n$.

Assuming that $x=(x_1, \ldots, x_n)$ is a sample value of a sample $X=(X_1, \ldots, X_n)$, the statistic obtained by substituting X for x in the function denoting a sample characteristic value is

called a **sample characteristic** and given the same name as the corresponding population characteristic, except that the word "population" is replaced by the word "sample." A sample characteristic is a function of random variables. Hence it is also a random variable, and the problem of deriving its probability distribution from the assumed population distribution is called that of sampling distribution (→ 374 Sampling Distributions). Thus we define the **sample mean** $\bar{X}=\sum_{i=1}^{n} X_i/n$; the **sample variance** $\sum_{i=1}^{n}(X_i-\bar{X})^2/n$ (sometimes $\sum_{i=1}^{n}(X_i-\bar{X})^2/(n-1)$ is taken as the sample variance); the **sample standard deviation**

$$\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2/n},$$

which is the positive square root of the sample variance; the **sample mode**, which is the value taken by the largest number of X_i; and the **sample moment of order** k $\sum_{i=1}^{n}(X_i-\bar{X})^k/n$.

Among other statistics of frequent use are the **order statistic**, i.e., the set of values of $X_1, \dots, X_n$ arranged in order of magnitude and usually denoted by $X_{(1)}<X_{(2)}<\dots<X_{(n)}$. Various other statistics are defined in terms of order statistics: the **sample median** $X_{\mathrm{med}}=X_{((n+1)/2)}$ for odd n and $=(X_{(n/2)}+X_{((n/2)+1)})/2$ for even n, the **sample range** $R=\max X_i-\min X_i=X_{(n)}-X_{(1)}$, and so on. The empirical distribution function $F_X(z)$ or its standardized form $S_x(z)=\sqrt{n}\{F_x(z)-F(z)\}$ can also be considered to be a function of the order statistics, and hence is a statistic taking values in the space of functions of a real variable. So is the **empirical characteristic function**

$$\phi_X(t)=\int \exp(itz)\,dF_x(z)=\sum_j \exp(itx_{(j)})/n.$$

In a sequence of Bernoulli trials, a set of successive components with an identical value is called a **run**. For example, (01100010) has a run of 0 of the length 3 and a run of 1 of the length 2.

Among the statistics listed in the previous paragraphs, the order statistic is an n-dimensional statistic and all others are 1-dimensional.

D. Other Cases

Let $(\mathbf{R}^2, \mathscr{B}^2, P_0)$ be a 2-dimensional probability space with a 2-dimensional population distribution P_0, and let $(X_1, \dots, X_n)$ $(X_i=(U_i, V_i))$ be a random sample of size n from P_0. In this case also, the population characteristics for the †marginal distributions of U_i and V_i and sample characteristics for $(U_1, \dots, U_n)$ and $(V_1, \dots, V_n)$ are defined as in Section C.

As an index for association between U_i and V_i the **population covariance** $\int(u-\mu_{(1)})(v-\mu_{(2)})\,dF(u,v)$ of U_i and V_i and the **population correlation coefficient**, which is equal to (population covariance)$/\sigma_{(1)}\sigma_{(2)}$, are defined. Here $F(u,v)$ is the joint distribution function of U_i and V_i, $\mu_{(1)}$ and $\mu_{(2)}$ are the respective population means of U_i and V_i, and $\sigma_{(1)}$ and $\sigma_{(2)}$ are the respective standard deviations. As corresponding sample characteristics, we have the **sample covariance** $\sum_{i=1}^{n}(U_i-\bar{U})(V_i-\bar{V})/n$ of $(U_1, \dots, U_n)$ and $(V_1, \dots, V_n)$ and the **sample correlation coefficient**

$$\frac{\sum_{i=1}^{n}(U_i-\bar{U})(V_i-\bar{V})}{\left(\sum_{i=1}^{n}(U_i-\bar{U})^2\right)^{1/2}\left(\sum_{i=1}^{n}(V_i-\bar{V})^2\right)^{1/2}},$$

where $\bar{U}=\sum_{i=1}^{n} U_i/n$ and $\bar{V}=\sum_{i=1}^{n} V_i/n$.

Similarly, statistics of the samples from a population of k-dimensional distribution $(k\geqslant 3)$ can be defined (→ 280 Multivariate Analysis). More generally, in statistical inference we encounter samples where observed values may not be mutually independent or identically distributed, but have more complicated probability structures. Statistics as functions of such samples are also considered.

E. General Properties of Statistics

The general theory of statistics has been studied in a measure-theoretic framework. $(\mathscr{X}, \mathscr{A}, \mathscr{P})$ is called a **statistical structure**, where $(\mathscr{X}, \mathscr{A})$ is a measurable space and $\mathscr{P}$ is a family of probability measures on $(\mathscr{X}, \mathscr{A})$. A σ-subfield $\mathscr{B}$ of $\mathscr{A}$ (hereafter abbreviated σ-field) is called **sufficient** for $\mathscr{P}$ if for any $A\in\mathscr{A}$ there exists a $\mathscr{B}$-measurable conditional probability of A independent of $P_\theta\in\mathscr{P}$, that is, a $\mathscr{B}$-measurable function $\varphi_A(x)$ satisfying

$$P_\theta(A\cap B)=\int_B \varphi_A(x)\,dP_\theta(x)$$

for all $P_\theta\in\mathscr{P}$ and $B\in\mathscr{B}$.

For any two σ-fields $\mathscr{B}_1$ and $\mathscr{B}_2$, the notation $\mathscr{B}_1\subset\mathscr{B}_2[\mathscr{P}]$ means that to each set A_1 in $\mathscr{B}_1$ there corresponds an A_2 in $\mathscr{B}_2$ satisfying $P_\theta((A_1-A_2)\cup(A_2-A_1))=0$ for all $P_\theta\in\mathscr{P}$. When the reverse relation $\mathscr{B}_2\subset\mathscr{B}_1[\mathscr{P}]$ also holds, we write $\mathscr{B}_1=\mathscr{B}_2[\mathscr{P}]$.

For a statistic t which is a measurable function from $(\mathscr{X}, \mathscr{A})$ to $(\mathscr{Y}, \mathscr{C})$, $\mathscr{B}(t)=\{B\mid B\in\mathscr{A}, t(B)\in\mathscr{C}\}$ is a σ-field and is called the σ-field induced by t. If $\mathscr{B}(t)$ is sufficient for $\mathscr{P}$, t is said to be sufficient for $\mathscr{P}$. Since sufficiency of a statistic means that of a σ-field, we consider only sufficiency of a σ-field.

$\mathscr{B}$ is called **necessary** if for any sufficient $\mathscr{B}_0$ we have $\mathscr{B}\subset\mathscr{B}_0[\mathscr{P}]$. A necessary and sufficient σ-field is called a **minimal sufficient σ-field**. A

necessary and sufficient statistic is also called a **minimal sufficient statistic**. Such a statistic does not always exist; $\mathscr{B}_2$ containing a sufficient $\mathscr{B}_1$ is not always sufficient.

$\mathscr{B}$ is said to be **complete** if for every $\mathscr{B}$-measurable integrable function φ, $\int_{\mathscr{X}} \varphi(x) dP_\theta(x) = 0$ for all $P_\theta \in \mathscr{P}$ implies $P_\theta(\{x \mid \varphi(x) \neq 0\}) = 0$ for all $P_\theta \in \mathscr{P}$. $\mathscr{B}$ is said to be **boundedly complete** if for every bounded $\mathscr{B}$-measurable φ, $\int_{\mathscr{X}} \varphi(x) dP_\theta(x) = 0$ for all $P_\theta \in \mathscr{P}$ implies $P_\theta(\{x \mid \varphi(x) \neq 0\}) = 0$ for all $P_\theta \in \mathscr{P}$. When $\mathscr{B}(t)$ is (boundedly) complete, t is called (boundedly) complete. If $\mathscr{B}_1 \subset \mathscr{B}_2$ and $\mathscr{B}_2$ is (boundedly) complete, $\mathscr{B}_1$ is also (boundedly) complete. If $\mathscr{B}_1$ is (boundedly) complete and sufficient and $\mathscr{B}_2$ is minimal sufficient, we have $\mathscr{B}_1 = \mathscr{B}_2 [\mathscr{P}]$.

F. Dominated Statistical Structure

When all $P_\theta \in \mathscr{P}$ are absolutely continuous with respect to a σ-finite measure λ on $\mathscr{A}$, then $(\mathscr{X}, \mathscr{A}, \mathscr{P})$ is said to be a **dominated** statistical structure and $\mathscr{P}$ is said to be a dominated family of probability distributions. In this case, P_θ has the density $f_\theta(x) = dP_\theta / d\lambda$ with respect to λ by the Radon-Nikodym theorem. If $\mathscr{A}$ is separable, $\mathscr{P}$ is a separable metric space with respect to the metric $\rho(P_{\theta_1}, P_{\theta_2}) = \sup_{B \in \mathscr{A}} |P_{\theta_1}(B) - P_{\theta_2}(B)|$. There exists a countable subset $\mathscr{P}' = \{P_{\theta_1}, P_{\theta_2}, \dots\}$ of $\mathscr{P}$ such that $P_\theta(N) = 0$ for all $P_\theta \in \mathscr{P}'$ implies $P_\theta(N) = 0$ for all $P_\theta \in \mathscr{P}$. If we put $\lambda_0 = \sum_i c_i P_{\theta_i}$, $c_i > 0$, $\sum_i c_i = 1$, λ_0 dominates $\mathscr{P}$, and if $\mathscr{B}$ is sufficient for $\mathscr{P}$ we can choose a $\mathscr{B}$-measurable version of $dP_\theta / d\lambda_0$. Conversely, if there exists a σ-finite measure λ such that we can choose a $\mathscr{B}$-measurable version of $dP_\theta / d\lambda$ for all $P_\theta \in \mathscr{P}$, then $\mathscr{B}$ is sufficient.

If $\mathscr{P}$ is dominated by a σ-finite λ, $\mathscr{B}$ is sufficient if and only if there exist a $\mathscr{B}$-measurable g_θ and an $\mathscr{A}$-measurable h independent of θ satisfying

$$\frac{dP_\theta}{d\lambda} = g_\theta h \text{ a.e. } (\mathscr{A}, \lambda) \quad \text{for all} \quad P_\theta \in \mathscr{P}.$$

This is called **Neyman's factorization theorem**.

With a dominated statistical structure, there exists a minimal sufficient σ-field, and a σ-field containing a sufficient σ-field is also sufficient.

We say that a σ-field $\mathscr{B}$ is **pairwise sufficient** for $\mathscr{P}$ if it is sufficient for every pair $\{P_{\theta_1}, P_{\theta_2}\}$ of measures in $\mathscr{P}$. A necessary and sufficient condition for $\mathscr{B}$ to be sufficient for a dominated set $\mathscr{P}$ is that $\mathscr{B}$ be pairwise sufficient for $\mathscr{P}$.

Recently, a more general statistical structure has been studied. Put $\mathscr{A}_e(\mu) = \{A \mid A \in \mathscr{A}, \mu(A) < \infty\}$. A measure μ on $\mathscr{A}$ is said to be a localizable measure if there exists ess-sup $\mathscr{F}(\mu)$ for any subfamily $\mathscr{F} \subset \mathscr{A}_e(\mu)$, that is, if there exists a set $E \in \mathscr{A}$ such that $\mu(A - E) = 0$ holds for all $A \in \mathscr{F}$, and $\mu(A - S) = 0$ for all $A \in \mathscr{F}$ implies $\mu(E - S) = 0$. A σ-finite measure is localizable. A measure μ is said to have the **finite subset property** if for any A satisfying $0 < \mu(A)$, there exists a $B \subset A$ satisfying $0 < \mu(B) < \infty$. A statistical structure $(\mathscr{X}, \mathscr{A}, \mathscr{P})$ is said to be **weakly dominated** if $\mathscr{P}$ is dominated by a localizable measure μ with the finite subset property and a density $dP_\theta / d\mu$ exists for all $P_\theta \in \mathscr{P}$. In this case a minimal sufficient σ-field exists, and a pairwise sufficient σ-field is sufficient. For example, let $\mathscr{A} = 2^{\mathscr{X}}$ and $\mathscr{P}$ be the set of all discrete probability measures on $\mathscr{A}$. $\mathscr{P}$ is weakly dominated by the counting measure μ which is localizable on $\mathscr{A}$.

The order statistic is sufficient if $\mathscr{P}$ is dominated, $X_1, \dots, X_n$ are mutually independent and identically distributed random variables with $\mathscr{X} = \mathbf{R}^n$, and each $P_\theta \in \mathscr{P}$ is invariant under every permutation of the components of the points $x = (x_1, \dots, x_n)$ in $\mathscr{X}$. Moreover, the order statistic is complete if $\mathscr{P}$ is large enough. For example, we have the following theorem: The order statistic is complete if every P_θ^0 (the component of P_θ on $\mathscr{B}_0 \subset \mathscr{B}^1$, $\theta \in \Theta$) is absolutely continuous with respect to the Lebesgue measure l on $\mathbf{R}$ and $\{P_\theta^0 \mid \theta \in \Theta\}$ contains all P_θ^0 for which $g(z) = dP_\theta^0 / dl$ is constant on some finite disjoint intervals in $\mathscr{X}_0 \subset \mathbf{R}$. A similar result holds for discrete distributions.

We call θ a **selection parameter** when $f_\theta(x) = c(\theta) \chi_{E_\theta}(x) h(x)$, where $h(x)$ is a positive $\mathscr{B}$-measurable function, $\chi_{E_\theta}(x)$ is the indicator function of a set $E_\theta \in \mathscr{B}$, and $c(\theta)$ is a constant depending on θ. Here θ determines the carrier E_θ of $f_\theta(x)$ but does not essentially affect the functional form of $f_\theta(x)$. A necessary and sufficient statistic is given by $t^*(x) = \bigcap \{E_\theta \mid E_\theta \ni x, P_\theta \in \mathscr{P}\}$. Here the class of sets of the form given in the right-hand side of this expression is taken as $\mathscr{Y}$, and we set $\mathscr{C} = \{C \mid C \subset \mathscr{Y}, t^{*-1}(C) \in \mathscr{B}\}$. We call $t^*(x)$ the **selection statistic**. Two examples follow.

(I) †Uniform distributions. Let $\Theta = \{(\alpha, \beta) \mid -\infty < \alpha < \beta < \infty\}$, $\mathscr{X}_0 = \mathbf{R}$, P_θ^0 be the uniform distribution on (α, β), and $X = (X_1, \dots, X_n)$ be a random sample of size n having P_θ^0 as its population distribution. Then $E_\theta = \{x \mid \alpha \leqslant \min_i x_i \leqslant \max_i x_i \leqslant \beta\}$ and $f_\theta(x) = (\beta - \alpha)^{-n} \chi_{E_\theta}(x)$. If we put $t(x) = (\min_i x_i, \max_i x_i)$, $\mathscr{Y} = \mathbf{R}^2$, and $\mathscr{C} =$ the set of all Borel sets of $\mathbf{R}^2$, it follows that $t(x) = t^*(x) [\mathscr{P}]$, where $t^*(x)$ is the selection statistic. Hence $t(x)$ itself is necessary and sufficient.

(II) †Exponential distributions. Put $\Theta = (-\infty, \infty)$, $\mathscr{X}_0 = \mathbf{R}$, and

$$g_\theta(z) = \begin{cases} \alpha e^{-\alpha(z-\theta)}, & z \geqslant \theta, \\ 0, & z < \theta, \end{cases}$$

where α is a known constant, and let $X = (X_1, \dots, X_n)$ be a random sample of size n

having a population distribution with density function $g_\theta(z)$. Then

$$E_\theta=\left\{x \mid \theta\leqslant\min_i x_i\right\},$$

$$f_\theta(x)=\alpha^n e^{n\alpha\theta}\chi_{E_\theta}(x)\exp\left(-\alpha\sum_{i=1}^n x_i\right).$$

If we put $t(x)=\min_i x_i$ and let $t^*(x)$ be the selection statistic, it follows that $t(x)=t^*(x)[\mathscr{P}]$, and $t(x)$ is necessary and sufficient. If $\Theta=\{(\alpha,\theta) \mid 0<\alpha<\infty, -\infty<\theta<\infty\}$, then $t(x)=\{\min_i x_i, \sum_i x_i\}$ is a necessary and sufficient statistic.

G. Exponential Families of Distributions

A dominated $\mathscr{P}$ is called an **exponential family of distributions** if and only if $f_\theta(x)=dP_\theta/d\lambda$ can be expressed in the form

$$f_\theta(x)=\exp\left(\sum_{j=1}^k s_j(x)\alpha_j(\theta)+\alpha_0(\theta)+s_0(x)\right),\qquad x\in\mathscr{X},\quad \theta\in\Theta, \quad (1)$$

where the $s_j(x)$ $(j=0,1,\ldots,k)$ are real-valued $\mathscr{B}$-measurable functions and the $\alpha_j(\theta)$ $(j=0,1,\ldots,k)$ are constants depending on θ. If there exists a sufficient statistic for $\mathscr{P}$ that is not equivalent to but is in a certain sense simpler than the sample itself or the order statistics, then it can be shown under some regularity conditions that $\mathscr{P}$ must be an exponential family. The following theorem provides an instance of the hypotheses that guarantee such a conclusion: Let X be a sample from a 1-dimensional probability space $(\mathscr{X}_0,\mathscr{B}_0,P_\theta^0)$ with P_θ^0 the population distribution, where $\mathscr{X}_0$ is a finite or infinite interval in $\mathbf{R}$ and $\mathscr{B}_0$ is the class of all Borel sets. Let l denote the Lebesgue measure. Assume that $\{P_\theta^0\}$ is dominated by l and $g_\theta(z)=dP_\theta^0/dl$ is greater than a positive constant and continuously differentiable in z on $\mathscr{X}_0$. Assume further that there exists a sufficient statistic $t(x)$ with the property that for each open subset B of $\mathscr{X}$ $(\subset\mathbf{R}^n)$ and λ-null set N there are two points $x\neq x'$ in $B-N$ such that $t(x)\neq t(x')$. Then $\mathscr{P}$ is an exponential family, and the k given in (1) is less than n. Similar results are known also for cases where $X_1,\ldots X_n$ are not identically distributed.

It is evident from the construction of a necessary and sufficient statistic that the statistic $t(x)=(s_1(x),\ldots,s_k(x))$, where the $s_j(x)$ are those appearing in (1), is sufficient for an exponential family and necessary if $\alpha_1(\theta),\ldots,\alpha_k(\theta)$ are linearly independent. If $\{(\alpha_1(\theta),\ldots,\alpha_k(\theta)) \mid \theta\in\Theta\}$ contains a k-dimensional interval, $t(x)$ is complete. The distribution of $t(x)$ is of exponential type. When $X_1,\ldots,X_n$ are mutually independent with a common distribution of exponential type, the distribution of $X=(X_1,\ldots,X_n)$ is of exponential type, and vice versa. The family (1) of distributions is a special form of †Pólya-type distributions, and various distributions given in (III)–(VII) below are written in this form. In the following examples, for a sample of size n from the specified distribution, $f_\theta(x)$ is the density with respect to Lebesgue measure in (III), (IV), and (V) and to counting measure in (VI) and (VII) (→ Appendix A, Table 22).

(III) †Normal distributions $N(\mu,\sigma^2)$, $\mathscr{X}_0=(-\infty,\infty)$.

$$f_\theta(x)=\exp\left(-\left(\sum_{i=1}^n x_i^2\right)\frac{1}{2\sigma^2}+\left(\sum_{i=1}^n x_i\right)\frac{\mu}{\sigma^2}-\frac{n\mu^2}{2}-n\log\sigma-\frac{n}{2}\log 2\pi\right).$$

(IV) †Γ-distributions $\Gamma(p,\sigma)$, $\mathscr{X}_0=(0,\infty)$.

$$f_\theta(x)=\exp\left((p-1)\sum_{i=1}^n \log x_i-\frac{1}{\sigma}\sum_{i=1}^n x_i-np\log\sigma-n\log\Gamma(p)\right).$$

(V) †Exponential distributions $e(\mu,\sigma)$, $\mathscr{X}_0=(\mu,\infty)$.

$$f_\theta(x)=\exp\left(-\left(\sum_{i=1}^n x_i\right)\frac{1}{\sigma}-n\log\sigma+n\frac{\mu}{\sigma}\right).$$

(VI) †Binomial distributions $Bin(N,p)$, $\mathscr{X}_0=\{0,1,\ldots,N\}$.

$$f_\theta(x)=\exp\left(\left(\sum_{i=1}^n x_i\right)\log\frac{p}{1-p}+nN\log(1-p)-\sum_{i=1}^n\log\binom{N}{x_i}\right).$$

(VII) †Poisson distributions $P(\lambda)$, $\mathscr{X}_0=\{0,1,2,3,\ldots\}$.

$$f_\theta(x)=\exp\left(\left(\sum_{i=1}^n x_i\right)\log\lambda-\sum_{i=1}^n\log(x_i!)-n\lambda\right).$$

H. Ancillary Statistics

A statistic $t(x)$ is called an **ancillary statistic** when for every element A in $\mathscr{A}(t)$, $P_\theta(A)$ is independent of θ, or in other words, when the distribution of $t(x)$ is independent of θ. A sufficient condition for a statistic to be ancillary is that it is independent of some sufficient statistic. Conversely, an ancillary statistic is independent of all boundedly complete sufficient statistics.

I. Invariant Statistics

Suppose that we are given groups of one-to-one measurable transformations G and $\bar{G}$ on

$\mathscr{X}$ and Θ, respectively. Suppose also that we are given a †homomorphism $g \to \bar{g}$ from G to $\bar{G}$ satisfying $P_\theta(g^{-1}B) = P_{\bar{g}\theta}(B)$. In this case $\mathscr{P} = \{P_\theta \mid \theta \in \Theta\}$ is called **G-invariant**. If $\bar{G}$ is transitive, there exists a fixed element θ_0 of Θ that is sent to an arbitrary θ by an element g of G. In this case, θ is called a **transformation parameter**. In particular, if $\Theta = \mathbf{R}$, X is a random sample from a population distribution, and $P_\theta(B) = P_{\theta_0}(B - \theta)$, where $B - \theta = \{x \mid (x_1 + \theta, \ldots, x_n + \theta) \in B\}$, then θ is called a **location parameter**. When $\Theta = (0, \infty)$ and $P_\theta(B) = P_1(B/\theta)$, where $B/\theta = \{x \mid (\theta x_1, \ldots, \theta x_n) \in B\}$, θ is called a **scale parameter**. Now assume that θ is a combination of these two kinds of parameters such that $\theta = (\alpha, \beta)$ $(-\infty < \alpha < \infty, 0 < \beta < \infty)$ and $P_\theta(B) = P_{\theta_0}((B - \alpha)/\beta)$, where $\theta_0 = (0, 1)$. Then if $\mathscr{P}$ is an exponential family, (1) of Section G can be written as

$$g_e(z) = \frac{dP_\theta^0}{dl} = \frac{1}{\beta}\exp\left(\sum_{j=0}^{m} k_j \left(\frac{z-\alpha}{\beta}\right)^j\right),$$

where the k_j $(j = 0, 1, \ldots, m)$ are constants.

We call $t(x)$ an **invariant statistic** with respect to a general transformation group G when $t(gx) = t(x)$ for all $g \in G$ and $x \in \mathscr{X}$. An invariant statistic is said to be **maximal invariant** with respect to G if, for $t(x) = t(x')$, there exists a $g \in G$ such that $x = gx'$. If t_0 is maximal invariant with respect to G, a statistic t is invariant under G if and only if $t_0(x) = t_0(x')$ implies $t(x) = t(x')$.

When $\mathscr{P}$ is G-invariant, a set A $(\in \mathscr{A})$ is called G-invariant if $gA = A$ for all $g \in G$. We denote by $\mathscr{A}^0$ the set of all G-invariant sets in $\mathscr{A}$. $\mathscr{A}^0$ is clearly a σ-field. A set A $(\in \mathscr{A})$ is called **almost G-invariant** if $gA = A$ $(\mathscr{A}, \mathscr{P})$ for all $g \in G$. We denote by $\mathscr{A}^*$ the σ-field consisting of all almost G-invariant sets. If $\mathscr{P}$ is G-invariant, $\mathscr{B}$ is sufficient for $\mathscr{P}$, $g\mathscr{B} = \mathscr{B}$ for all $g \in G$, and moreover, $\mathscr{B}^0 = \mathscr{B}^*$ $(\mathscr{A}, \mathscr{P})$, then $\mathscr{B}^0$ is a sufficient σ-subfield of $\mathscr{A}^0$, where $\mathscr{B}^0 = \mathscr{B} \cap \mathscr{A}^0$ and $\mathscr{B}^* = \mathscr{B} \cap \mathscr{A}^*$.

J. Various Definitions of Sufficiency

There are many different definitions of sufficiency, and the relations among them have been investigated. A σ-field $\mathscr{B}$ is called **decision-theoretically sufficient** or **D-sufficient** if for a given $\mathscr{A}$-measurable decision function δ there exists a $\mathscr{B}$-measurable decision function δ' such that

$$\int_{\mathscr{X}} \delta(x, E)\, dP_\theta(x) = \int_{\mathscr{X}} \delta'(x, E)\, dP_\theta(x)$$

for all $E \in \mathscr{D}$, $P_\theta \in \mathscr{P}$, where a decision space $(D, \mathscr{D})$ is quite arbitrary. $\mathscr{B}$ is called **test sufficient** if for any given $\mathscr{A}$-measurable test function φ, there exists a $\mathscr{B}$-measurable test function φ' satisfying $E_\theta(\varphi) = E_\theta(\varphi')$ for all $P_\theta \in \mathscr{P}$. Let $(\Theta, \mathscr{C})$ be a measurable space of parameter θ and $\hat{\Theta}$ be the set of all probability measures on $\mathscr{C}$. Moreover we assume that $P_\theta(B)$ is $\mathscr{C}$-measurable as a function of θ for any fixed $B \in \mathscr{A}$. For any $\xi \in \hat{\Theta}$, we define λ_ξ by

$$\lambda_\xi(A \times C) = \int_C P_\theta(A)\, d\xi(\theta), \qquad A \in \mathscr{A},\ C \in \mathscr{C}.$$

We denote by $\tilde{\lambda}_\xi$ the extension of λ_ξ to $\mathscr{A} \times \mathscr{C}$. $\mathscr{B}$ is said to be **Bayes sufficient** if

$$E_{\tilde{\lambda}_\xi}(I_{\mathscr{X} \times C} \mid \mathscr{A} \times \mathscr{C}) = E_{\tilde{\lambda}_\xi}(I_{\mathscr{X} \times C} \mid \mathscr{B} \times \mathscr{C})$$

for all $\xi \in \hat{\Theta}$, $C \in \mathscr{C}$, that is, the a posteriori distribution on Θ given $\mathscr{A}$ coincides with that given $\mathscr{B}$ for any a priori ξ. When $\mathscr{P}$ is dominated, these definitions coincide with the classical definition of sufficiency. Generally, a D-sufficient σ-field contains at least one sufficient σ-field. A σ-field containing a sufficient σ-field is Bayes sufficient. Hence Bayes sufficiency follows from D-sufficiency and from classical sufficiency. If a D-sufficient σ-field is separable it is sufficient.

The notion of **prediction sufficiency** or **adequacy** was defined by Skibinsky [10]. Let (X, Y) be a pair of random variables defined over the probability space $(\mathscr{X} \times \mathscr{Y}, \mathscr{A} \times \mathscr{B}, \mathscr{P}_\theta)$. We suppose that X is the sample to be observed, and Y is (are) the value(s) of future observation(s) about which we are to make prediction(s) based on X. X and Y have joint probability distribution with an unknown parameter. A statistic $T = T(X)$ or a subfield $\mathscr{C}$ of $\mathscr{A}$ is said to be prediction sufficient or adequate if (a) given T, X and Y are conditionally independent (or given $\mathscr{C}$, $\mathscr{A}$ and $\mathscr{B}$ are conditionally independent or Markov) and (b) T is sufficient for X ($\mathscr{C}$ is sufficient for $\mathscr{A}$). It was proved that in any form of prediction on Y, we may restrict ourselves to the class of procedures that are functions of T (or are $\mathscr{C}$-measurable).

References

[1] D. A. S. Fraser, Nonparametric methods in statistics, Wiley, 1957.
[2] E. L. Lehmann, Testing statistical hypotheses, Wiley, 1959.
[3] S. S. Wilks, Mathematical statistics, Wiley, 1962.
[4] R. R. Bahadur, Sufficiency and statistical decision functions, Ann. Math. Statist., 25 (1954), 423–462.
[5] T. S. Pitcher, A more general property than domination for sets of probability measures, Pacific J. Math., 15 (1965), 597–611.
[6] L. Brown, Sufficient statistics in the case of independent random variables, Ann. Math. Statist., 35 (1964), 1456–1474.

[7] E. W. Barankin and A. P. Maitra, Generalization of the Fisher-Darmois-Koopman-Pitman theorem on sufficient statistics, Sankhyā, ser. A, 25 (1963), 217–244.
[8] W. J. Hall, R. A. Wijsman, and J. K. Ghosh, The relationship between sufficiency and invariance with applications in sequential analysis, Ann. Math. Statist., 36 (1965), 575–614.
[9] K. K. Roy and R. V. Ramamoorthi, Relationship between Bayes, classical and decision theoretic sufficiency, Sankhyā, ser. A, 41 (1979), 48–58.
[10] M. Skibinsky, Adequate subfields and sufficiency, Ann. Math. Statist., 38 (1967), 155–161.

397 (XVIII.1) Statistical Data Analysis

A. Statistical Data

Statistical data analysis is comprised of a collection of mathematical methods whereby we can deal with numerical data obtained through observations, measurements, surveys, or experiments on the "objective" world. The purpose of statistical analysis is to extract the relevant information from that numerical data pertinent to the subject under consideration. The nature and the properties of the subject and also the purpose of the analysis may vary greatly. The subject may be physical, biological, chemical, sociological, psychological, economic, etc. in nature, and the purpose of the analysis can be purely scientific, as well as technological, medical, or managerial. Because of the great diversity of statistical data, the methods of statistical data analysis and the manner of application should differ greatly from situation to situation; we cannot expect a single unified system of methods to be applicable to all cases. Nevertheless, we have several formal methods of statistical analysis that are more or less mutually related and have been successfully applied to most, if not all, statistical data.

Statistical data can be classified into several types according to a few criteria: according to the property of each observation or measurement, they can be either **quantitative** or **qualitative**; according to whether only one observation is made on each object under investigation or many observations on the same object, they can be either **univariate** or **multivariate**; and according to whether the observations are made at one time or consecutively in the course of time, they may be either **cross sectional** or **time series**. Each different type of statistical data requires a different type of procedure (→ 280 Multivariate Analysis, 421 Time Series Analysis).

B. Frequency Distributions and Histograms

Statistical data have the simplest structure when they consist of a collection of observations made on an aggregate of objects supposedly of the same kind. Such an aggregate is usually called a **population**, and the number of its members (its size) is denoted by N. When the data are qualitative or **categorical**, each member of the population is classified into several types according to some criteria, the data consist of the numbers of the members of the population classified into each of the categories. Such numbers are usually called **frequencies**, and the set of frequencies is called the **frequency distribution**.

When the data are quantitative and univariate, one quantitative attribute of each member of the population is observed, and the results are given as a set of N real numbers $(x_1, x_2, \dots, x_N)$. When N is large, as is usually expected, it is necessary to summarize these results in some manner. One common method is to tabulate the frequency distribution: We define a certain number of intervals $(a_{i-1}, a_i]$, $i=1, \dots, K$, $a_0 < a_1 < \dots < a_K$, $a_0 \leqslant \min x_i$, $\max x_i \leqslant a_K$; and we count the numbers of those x's falling within each of the intervals and tabulate those numbers or frequencies f_i, $i=1,2,\dots,K$. Frequency distribution is often represented in the form of a **histogram**, where the endpoints of the intervals are marked on the horizontal axis, and above each interval a rectangle of area proportional to the frequency for the interval is drawn. It is usually recommended that the widths of the intervals in the frequency distribution be equal, especially when it is to be represented by a histogram. It is, however, often impossible or impractical to do so, and sometimes a logarithmic or other functional scale is used in the abscissa of the histogram; then it is desirable that the intervals of the transformed values are approximately of equal lengths. The number K of the intervals should also be of an appropriate magnitude, neither too large nor too small; K is often constrained by the size N of the population, the shape of the distribution, or other factors. Usually, K is chosen to be between 6 and 20.

From the frequency distribution, we obtain the **cumulative distribution** by associating with each endpoint a_i of the intervals the number F_i of x's not greater than a_i, namely, $F_i = \sum_{j \leqslant i} f_j$. The curve obtained by connecting $K+1$ points

of coordinates (a_i, F_i), $i=0, 1, \ldots, K$, by linear segments is called the **cumulative distribution curve** (or **polygon**).

C. Characteristics of the Distribution

In order to summarize univariate quatitative data, various values are calculated from the values $x_1, \ldots, x_N$. Such values are called **statistics** (singular, †statistic) and are used to characterize the distribution of the values. Various types of statistics characterize different aspects of the distribution:

(a) Representative value or **measure of location**: a value which is supposed to give the "representative," "typical," or "most common" value in the population. By far the most commonly used measure is the **mean** $\bar{x}=\sum_{i=1}^{N} x_i/N$. $\bar{x}$ is sometimes called the **arithmetic mean**, and some other "means" are also calculated: especially when all the values are positive, the **geometrical mean** $\bar{x}_G=(\prod_i x_i^{1/N})$ or **harmonic mean** $\bar{x}_H=(\sum_i(1/x_i)/N)^{-1}$ may be calculated; more generally, for some monotone function $f(x)$ we can calculate the f-mean by $\bar{x}_f=f^{-1}(\sum_i f(x_i)/N)$, of which the geometric and the harmonic means are special cases. Another measure of location is the **median**, which is the value in the population located exactly in the middle of the ordering of the magnitudes; more precisely, if $x_{(1)}<x_{(2)}<\ldots<x_{(N)}$ are the values in the population arranged according to their magnitudes, the median $x_{\text{med}}=x_{((N+1)/2)}$ for odd N, and $=\frac{1}{2}(x_{(N/2)}+x_{((N/2)+1)})$ for even N. The **mode** is also sometimes used; this is defined as the value (usually the center) corresponding to the highest frequency.

(b) The **measure of variability** or **dispersion** shows how widely the values in the population vary. The most common measure is the **standard deviation**, which is defined by $s_x=\sqrt{\sum_i(x_i-\bar{x})^2/N}$, and its square is called the **variance** V_x^2. A similar measure is the **mean absolute deviation** $D_x=\sum_i|x_i-\bar{x}|/N$. Another type of a measure of dispersion is the **range** $R_x=\max x_i-\min x_i$, and the **interquartile range** $Q_x=x_{(3N/4)}-x_{(N/4)}$ and more generally the interquantile range $x_{(\alpha N)}-x_{((1-\alpha)N)}$ for some α. The ratio of the standard deviation to the mean is called the **coefficient of variation** (C. V. for short) and is used as a measure of relative variability when all the values in the population are positive.

(c) Characteristics often used to characterize the "shape" of distributions are the **moments** (around the origin) $m_k=N^{-1}\sum_i x_i^k$ and the **central moments** (moments around the mean) $M_k=N^{-1}\sum_i(x_i-\bar{x})^k$ for a positive integer k; for a specific k these are called the kth moments. Central kth moments with odd k are equal to zero when the distribution is symmetric, i.e., when the histogram is symmetrically shaped. Hence the third moment or its ratio to the third power of the standard deviation s_x^3 is used as a measure of the asymmetry of a distribution; this is called the **skewness**. The fourth moment is large if there are some values which are far off from others and small when all values are concentrated; hence it tends to be large when the histogram has a rather sharp peak in the center and has a long tail in either direction or both, and tends to be small when the histogram is flat in the center and drops off sharply at both ends. Accordingly, the ratio M_4/V_x^2 is used as a measure of long-tailedness of the distribution; this ratio minus 3 is called the **kurtosis**.

D. Theoretical Frequency Distribution

When the observed values can be any real number (sometimes in an interval), the size of the population N is increased indefinitely, and the widths of the intervals are decreased to 0, the histogram is expected to approach a smooth curve. And in the limit when N is infinity, we can assume that the distribution is represented by a mathematically well-behaved function $f(x)$ and that the ratio of the numbers of those values in the population within the interval (a, b) to the size of the population approaches $\int_a^b f(x)\,dx$. Such a function $f(x)$ is called the **frequency function** or **density function**. Various types of functions have been proposed and used as "theoretical" frequency functions to approximate the actually observed frequency distributions. The most important is the **normal density function**

$$\varphi(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}.$$

The following density functions most commonly appear in applications: the **gamma density**, $f(x)=x^{p-1}\exp(-x/a)/a^p\Gamma(p)$ for $x>0$ and $=0$ for $x\leqslant 0$; the **beta density**, $f(x)=x^{p-1}(1-x)^{q-1}/B(p,q)$ for $0<x<1$ and $=0$ otherwise.

We can conceive of a population of infinite size with some density function; the term **theoretical distribution** is used to mean such a population with its density function, and more specifically the †normal distribution, etc. Such a population and associated density is often called a continuous distribution. For a theoretical distribution, the mean, variance, and moments are naturally defined by

$$\mu=\int xf(x)\,dx,\qquad \sigma^2=\int(x-\mu)^2 f(x)\,dx,$$

$$\mu_k=\int(x-\mu)^k f(x)\,dx,\qquad \mu_k'=\int x^k f(x)\,dx.$$

It should, however, be noted that the mean, the variance, or the moments may not exist for particular distributions.

K. Pearson introduced a system of density functions defined as solutions of the differential equation

$$d\ln f(x)/dx=(A+Bx)/(C+Dx+Ex^2),$$

where $A, B, \ldots, E$ are constants. A distribution thus obtained is called a **Pearson distribution**. The normal, gamma, and beta distributions together with some other commonly used distributions, such as the t- and F-distributions, are Pearson distributions.

E. Measure of Concentration

When all the observed values are nonnegative in nature, we may sometimes require some measure of inequality or concentration of the distribution. For such a purpose we order the observed values according to their magnitudes and obtain $x_{(1)}\leqslant x_{(2)}\leqslant\ldots\leqslant x_{(N)}$; we define $S_i=\sum_{j\leqslant i}x_{(j)}$ for $i=1,\ldots,N$, and $S_0=0$, plot $N+1$ points $(S_i/S_N, i/N)$, $i=0,1,\ldots,N$, and connect them by line segments. The graph thus obtained is called the **Lorentz curve** or the curve of concentration, and it connects the origin and the point (1,1). It lies below the 45° line, and if all the values are nearly equal the curve comes close to the 45° line, but if values are widely unequal, the curve comes close to the horizontal axis and suddenly jumps to the point (1,1). The area between the curve and the 45° line is called the **area of concentration**, and it is equal to one-fourth of the mean difference δ divided by the mean, where δ is defined by

$$\delta=\frac{1}{N^2}\sum_i\sum_j|x_i-x_j|.$$

$G=\delta/\bar{x}$ is called the **Gini coefficient of concentration** and is used as a measure of concentration or inequality of distribution. Other measures, including the coefficient of variation, are also used to represent the concentration.

F. Discrete Distributions

There are cases where the observed values are taken only from the nonnegative integers, e.g., the number of individual animals of a specific species in an area, of accidents during a specified time, etc. In such cases, when we increase the number of observations, the distribution does not approach one with a continuous density function but rather one with a certain theoretical discrete distribution. Among theoretical discrete distributions, the most commonly used are the **binomial distribution**: $f_j={}_nC_jp^j(1-p)^{n-j}$ for $0\leqslant j\leqslant n$, and the **Poisson distribution**: $f_j=e^{-\lambda}\lambda^j/j!$ for $j=0,1,\ldots$. The **hypergeometric distribution**: $f_j={}_MC_j\cdot{}_{N-M}C_{n-j}/{}_NC_n$, for $\max(0,M+n-N)\leqslant j\leqslant\min(n,M)$, and the **negative binomial distribution**: $f_j={}_{j+r-1}C_{r-1}p^r(1-p)^j, j=0,1,\ldots$, are also often used.

For discrete distributions we can define moments by $\mu=\sum_j jf_j$ and $\mu_k=\sum_j(j-\mu)^kf_j$, and $\mu'_k=\sum_j j^kf_j$, $k=2,3,\ldots$.

G. Generating Functions and Cumulants

For a theoretical distribution with the density function $f(x)$, the **moment generating function** $M(\theta)$ is defined by $M(\theta)=\int e^{\theta x}f(x)dx$. When $M(\theta)$ is well defined in an open interval including the origin, the distribution has all kth moments, and it can be expanded as

$$M(\theta)=1+\mu'_1\theta+\frac{1}{2!}\mu'_2\theta^2+\ldots+\frac{1}{k!}\mu'_k\theta^k,$$

from which the term "moment generating function" is derived. When θ is replaced by it with real t, we have the **characteristic function** $\varphi(t)=M(it)$, which can be expanded as

$$\phi(t)=1+\mu'_1(it)+\frac{1}{2!}\mu'_2(it)^2+\ldots+\frac{1}{k!}\mu'_k(it)^k$$
$$+o(|t|^k)$$

if the distribution has moments up to the kth. The function $K(\theta)=\ln M(\theta)$ is called the **cumulant generating function**, and the coefficients κ_j in the expansion

$$K(\theta)=\kappa_1\theta+\frac{\kappa_2}{2}\theta^2+\ldots$$

are called the **cumulants**. The kth cumulant κ_k is expressed as a polynomial of the moments of order not exceeding k; thus

$$\kappa_1=\mu'_1,\quad \kappa_2=\mu'_2-\mu'^2_1,\quad \kappa_3=\mu'_3-3\mu'_2\mu'_1+2\mu'^3_1,$$

etc.

For the normal distribution,

$$M(\theta)=\exp\left\{\mu\theta+\frac{\sigma^2\theta^2}{2}\right\};$$

hence

$$K(\theta)=\mu\theta+\frac{\sigma^2\theta^2}{2},$$

and the kth cumulant for $k\geqslant 3$ is equal to 0. Cumulants are used as measures indicating whether the distribution is close to or different from the normal.

For discrete distributions, the moment generating function is defined as $M(\theta)=\sum e^{\theta j}f_j$, but the **probability generating func-**

tion $P(t)=\sum f_j t^j = M(\ln t)$, the **factorial moment generating function** $\tilde{M}(\theta)=P(\theta+1)$, and the **factorial cumulant generating functions** $\tilde{K}(t)=\ln \tilde{M}(t)$ are also of use. The coefficient of t^j in the Maclaurin expansion of $\tilde{M}(t)$ is expressed as $\mu_{[k]}$ and called the kth **factorial moment**; it is equal to $\mu_{[k]}=\sum_j j(j-1)\ldots(j-k+1)f_j$. That in the expansion of $\tilde{K}(t)$ is the **factorial cumulant**. The factorial cumulant $\kappa_{[k]}$ is expressed by the same polynomial in $\mu_{[k]}$ as κ_k is expressed in μ'_k. For the Poisson distribution, $M(\theta)=\exp\lambda(e^\theta-1)$; hence $\tilde{K}(t)=\lambda t$ and it follows that the factorial cumulants $\kappa_{[k]}$ for $k\geqslant 2$ are all equal to zero if and only if the distribution is Poisson.

H. Bivariate Distribution

When two quantitative observations are obtained for each member of a population of size N, the results are given as N pairs of real numbers (x_i, y_i), $i=1,2,\ldots,N$. Such data are called **bivariate data** and the distribution, **bivariate distribution**. Those data can be illustrated as N points in a plane with coordinates (x_i, y_i), and such an illustration is called a **scatter diagram**. In order to characterize a bivariate distribution, we often use **bivariate moments**

$$M_{k,l}=\frac{1}{N}\sum(x_i-\bar{x})^k(y_i-\bar{y})^l,$$

where $\bar{x}$ and $\bar{y}$ are the means of x and y, respectively; especially, the (1,1) moment $M_{1,1}$ is called the **covariance** and is denoted as $\mathrm{Cov}(x,y)$. The most often used measure of the strength of the relation between x and y values is the **correlation coefficient** $r_{x,y}=\mathrm{Cov}(x,y)/s_x s_y$, where s_x and s_y are the standard deviations of x and y. It is easily shown that $-1\leqslant r_{x,y}\leqslant 1$, and when there exists a nearly linear relationship between x and y values, $r_{x,y}$ is close to either $+1$ or -1 according to whether the x and y values change in the same direction or in opposite directions. When there is no clear relationship between x and y, the correlation coefficient is close to zero, but it may not be a good measure of the relationship when x and y values are related nonlinearly.

A linear function $y=a+bx$ is called the **linear regression function** of y on x, for which the sum of the square distances $\sum_i(y_i-a-bx_i)^2$ is minimized. For the linear regression function the coefficients a and b are determined by $b=\mathrm{Cov}(x,y)/s_x^2$ and $a=\bar{y}-b\bar{x}$, and b is called the **regression coefficient**. We have that $\sum_i(y_i-a-bx_i)^2/\sum(y_i-\bar{y})^2=1-r_{xy}^2$, i.e., that the square of the correlation coefficient is equal to 1 minus the ratio of the variance of the residual y_i-a-bx_i to that of y; hence it is sometimes called the **coefficient of determination**. Similarly, the linear regression function of x on y is defined by $x=c+dy$, where $d=\mathrm{Cov}(x,y)/s_y^2$ and $c=\bar{x}-d\bar{y}$. We have $bd=r_{xy}^2$, and $|1/d|=|b/r_{xy}^2|\geqslant|b|$.

We can tabulate the bivariate frequency distribution by splitting the range of x values into K intervals $(a_{i-1},a_i]$, $i=1,\ldots,K$, and the range of y values into L intervals $(b_{j-1},b_j]$, $j=1,\ldots,L$, and counting the number f_{ij} of cases for which $a_{i-1}<x\leqslant a_i$ and $b_{j-1}<y\leqslant b_j$. In contrast to the bivariate frequency distribution, the distributions of x and y values are called the **marginal distributions**.

I. Bivariate Density Function

As we did for the univariate distribution, we can consider the limiting shape of the bivariate frequency distribution when the size N of the population tends to infinity and define a continuous bivariate distribution with density function $f(x,y)$, with which the ratio of those members in the population with values (x,y) in a set S in a plane is given by $\iint_S f(x,y)\,dx\,dy$. The bivariate density function is also called the **joint density**, and then the density functions of x and y are called the **marginal density functions** and are given by $f_1(x)=\int f(x,y)\,dy$ and $f_2(y)=\int f(x,y)\,dx$. The joint moments of a continuous bivariate distribution are defined by $\mu_{k,l}=\iint(x-\mu_1)^k(y-\mu_2)^l f(x,y)\,dx\,dy$, where μ_1 and μ_2 are the means of x and y, respectively. The **joint moment generating function** is defined by $M(t_1,t_2)=\iint e^{t_1x+t_2y}f(x,y)\,dx\,dy$, and the cumulant generating function by $K(t_1,t_2)=\log M(t_1,t_2)$, from which the **joint cumulants**

$$\kappa_{k,l}=\frac{\partial^{k+l}}{\partial t_1^k\partial t_2^l}K(t_1,t_2)|_{t_1=0,t_2=0}$$

are derived.

The **conditional density** of y given x is defined by $f(y|x)=f(x,y)/f_1(x)$, and the distribution with this density function is the **conditional distribution** of y given x; this latter can be interpreted as the distribution of y of those members in the population with x values in the interval $(x,x+dx]$, where dx is small. The conditional density and the conditional distribution of x given y are similarly defined. The mean and the moments of the conditional distribution are called the **conditional mean** and the **conditional moments**. The conditional mean of y given x, considered as a function of x, is called the **regression function** of y on x.

By far the most important theoretical bivariate density is the **bivariate normal density**,

which is given by

$$f(x,y)=(2\pi\sigma_1\sigma_2\sqrt{1-\rho^2})^{-1}$$
$$\times\exp\left\{\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}+\frac{(y-\mu_2)^2}{\sigma_2^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right]\right\},$$

for which the mean of x is μ_1 and that of y is μ_2, the variances of x and y are σ_1^2 and σ_2^2, respectively, and the covariance of x and y is equal to $\rho\sigma_1\sigma_2$. The bivariate normal distribution has several remarkable properties: All the k, l joint cumulants are equal to zero for $k+l\geqslant 3$; the marginal distributions of x and y are normal; the regression functions of y on x and x on y are both linear; the conditional distribution of y given x (and x given y) is normal and the conditional variance is constant; and the contours $f(x,y)=c$ for different values of c are equicentric ellipsoids.

J. Higher-Dimensional Data

When the data are of more than two dimensions, i.e., more than two observations are made on each of the objects, we designate the data by Nk-tuples of real numbers $(x_{i1},x_{i2},\ldots,x_{ik})$, $i=1,\ldots,N$ $(k\geqslant 3)$. Then we can calculate the moments of each of the variates and the joint moments, which are defined by

$$M_{k_1,k_2,\ldots,k_k}=\frac{1}{N}\sum(x_{i1}-\bar{x}_1)^{k_1}(x_{i2}-\bar{x}_2)^{k_2}\ldots(x_{ik}-\bar{x}_k)^{k_k}.$$

Also, we can arrange the variances and covariances in a symmetric matrix of order k, and we call it the **(variance-) covariance matrix**. A covariance matrix is easily shown to be nonnegative definite. The determinant of the covariance matrix is called the **generalized variance**. The matrix with the (i,j) element equal to the correlation coefficient of the ith and the jth variates r_{ij} (r_{ii} is set equal to 1) is called the **correlation matrix** and is denoted by R. R is also nonnegative definite. If we denote the (i,j) cofactor of R by R_{ij}, the quantity defined by

$$R_{i|1,\ldots,(i),\ldots,k}=\sqrt{1-|R|/R_{ii}}$$

is called the **multiple correlation coefficient** of the ith variate and all other variates; and

$$r_{ij|1,\ldots,(1),\ldots,(j),\ldots,k}=-R_{ij}/\sqrt{R_{ii}R_{jj}}$$

is called the **partial correlation coefficient** of the ith and the jth variates given all other variates. The meaning of these coefficients will be elucidated below. A linear function $a_0+a_1x_1+a_2x_2+\ldots+a_{k-1}x_{k-1}$ is called the **linear regression function** of x_k on $x_1,\ldots,x_{k-1}$, when the coefficients $a_1,\ldots,a_{k-1}$ are so determined that the sum $Q=\sum_i(x_{ik}-a_0-a_1x_{i1}-\ldots-a_{k-1}x_{ik-1})^2$ is minimized. They are determined from the equation

$$C_{i1}a_1+C_{i2}a_2+\ldots+C_{ik-1}a_{k-1}=C_{ik},\qquad (*)$$
$$i=1,\ldots,k-1,$$

with $a_0=\bar{x}_k-a_1\bar{x}_1-\ldots-a_{k-1}\bar{x}_{k-1}$, where C_{ij} are the covariances. $a_1,\ldots,a_{k-1}$ thus determined are called the **regression coefficients** of x_k on $x_1,\ldots,x_{k-1}$, and such a procedure is called the **method of least squares**. The equation (*) is called the **normal equation**. If we write $\hat{x}_{ik}=a_0+a_1x_{i1}+\ldots+a_{k-1}x_{ik-1}$, we have $Q=\sum(x_{ik}-\hat{x}_{ik})^2=\sum(x_{ik}-\bar{x}_k)^2-\sum(\hat{x}_{ik}-\bar{x}_k)^2=\sum(x_{ik}-\bar{x}_k)^2\times(1-R_{k|1,\ldots,k-1}{}^2)$, where $R_{k|1,\ldots,k-1}$ is the multiple correlation coefficient of x_k and $x_1,\ldots,x_{k-1}$, which is also equal to the correlation coefficient of x_k and $\hat{x}_k$. The square of the multiple correlation coefficient is also called the **coefficient of determination**. The quantities $x_{ik}-\hat{x}_{ik}$ are called the residuals. Let $\hat{x}_{ik-1}$ and $\hat{x}_{ik}$ be the values of regression functions of x_{k-1} and x_k, respectively, on $x_1,\ldots,x_{k-2}$, and let $y_{ik-1}=x_{ik-1}-\hat{x}_{ik-1}$ and $y_{ik}=x_{ik}-\hat{x}_{ik}$ be the residuals; then the correlation coefficient of y_{k-1} and y_k is equal to the partial correlation coefficient of x_{k-1} and x_k given $x_1,\ldots,x_{k-2}$. We have the following relation between the multiple and the partial correlation coefficients:

$$1-R^2_{k|1,\ldots,k-1}$$
$$=(1-R^2_{k|1,\ldots,k-2})(1-r^2_{k-1,k|1,\ldots,k-2}).$$

Multiple and partial correlation coefficients are also expressed in terms of the correlation coefficients of the variates. For example, it can be shown that

$$R^2_{3|12}=(r_{13}^2+r_{23}^2-2r_{12}r_{13}r_{23})/(1-r_{12}^2)$$

and that

$$r_{23|1}=(r_{23}-r_{12}r_{13})/\sqrt{(1-r_{12}^2)(1-r_{13}^2)}.$$

For higher dimensions, we can also define the **(joint) density function** $f(x_1,x_2,\ldots,x_k)$ and the **(joint) moment generating function**

$$M(t_1,t_2,\ldots,t_k)$$
$$=\int\ldots\int\exp(t_1x_1+t_2x_2+\ldots+t_kx_k)$$
$$\times f(x_1,x_2,\ldots,x_k)\,dx_1\ldots dx_k.$$

The most important multivariate joint density is that of the **multivariate normal distribution**, which is expressed by

$$f(x_1,\ldots,x_k)=(2\pi)^{-R/2}|\Sigma|\exp\left\{\frac{1}{2}\sum\sum\sigma^{ij}(x_i-\mu_i)(x_j-\mu_j)\right\},$$

where Σ is the covariance matrix with elements σ_{ij}, $\Sigma^{-1}=(\sigma^{ij})$, and μ_i is the mean of the ith variate. For the multivariate normal distribution the moment generating function is given by

$$M(t_1,\ldots,t_k)=\exp\left(\sum\mu_i t_i+\frac{1}{2}\sum\sigma_{ij}t_it_j\right).$$

K. Contingency Tables

When several qualitative observations are made on N objects, each object is classified according to the combination of the categories, and the data are summarized by the numbers $N(i_1,i_2,\ldots,i_k)$ of the objects that fall in the i_1th category according to the first observation, i_2th category in the second observation, etc. A table that shows the results of such observations is called a (k**-way**) **contingency table**. If there are m_1 categories in the first criterion, m_2 categories in the second, etc., the contingency table is also called an m_1 by m_2 by ... by m_k table. The numbers $\tilde{N}(j,i_j)$ of the objects which are classified into the i_jth category according to the jth criterion are called marginal frequencies. If we have

$$N(i_1,i_2,\ldots,i_k)/N=\tilde{N}(1,i_1)\tilde{N}(2,i_2)\ldots\tilde{N}(k,i_k)/N^k \quad \text{for all } i_1,i_2,\ldots,i_k,$$

then the k observations or criteria are independent.

The simplest contingency table is a 2 by 2 table, where several measures for the relation of two observations or criteria have been proposed, among which the most commonly used are the **measure of association** defined by

$$Q=\frac{N(1,1)N(2,2)-N(1,2)N(2,1)}{N(1,1)N(2,2)+N(1,2)N(2,1)}$$

and the **odds ratio**

$$\delta=\frac{N(1,1)N(2,2)}{N(1,2)N(2,1)}$$

and also

$$V=\frac{N(1,1)N(2,2)-N(1,2)N(2,1)}{\sqrt{\tilde{N}(1,1)\tilde{N}(1,2)\tilde{N}(2,1)\tilde{N}(2,2)}},$$

where $N(i,j_i)$ are marginal frequencies. The two observations are independent if and only if $Q=0$ or $\sigma=1$ and $V=0$. V is equal to the correlation coefficient of the variables x_1 and x_2, for which $x_i=0$ if the object is classified into the first category according to the ith criterion and $x_i=1$ if it is classified into the second category. In a two-way m_1 by m_2 table a measure of association is defined by

$$\frac{X^2}{N}=\sum\sum\left(\frac{N(i_1,i_2)N}{\tilde{N}(1,i_1)\tilde{N}(2,i_2)}-1\right)^2;$$

it can be shown that $X^2/N=0$ if and only if the two criteria are independent, and that $0\leqslant X^2/N\leqslant\min(m_1-1,m_2-1)$. When we take $x_i=1$ if the object is classified into the ith category according to the first criterion and $x_i=0$ otherwise and take $y_j=1$ if it belongs to the jth category in the second criterion and $y_j=0$ otherwise, it is shown that the sum of squares of the multiple correlation coefficients of x_i and $y_1,\ldots,y_{m_2-1}$ is equal to X^2/N.

L. Decomposition of the Variance

When one observation is qualitative while another is quantitative the objects are classified into several categories according to the first observation, while for each object the value of the second observation is also given. Let x_{ij} be the observed value of the jth object in the group of the ith category; then for each i we can obtain frequency distributions of x_{ij}, and compare these distributions. Let N_i be the number of objects in the ith category, and $\bar{x}_i=\sum x_{ij}/N_i$ be the mean in the ith category. Then the weighted variance of the $\bar{x}_i$ defined by $v_{\mathrm{B}}=\sum_i N_i(\bar{x}_i-\bar{x})^2/N$, where $\bar{x}$ is the mean of all the observations, i.e., $\bar{x}=\sum N_i\bar{x}_i/N$, is called the **between-group variance**, and the weighted mean of the variances of each of the groups defined by $v_{\mathrm{W}}=\sum_i\sum_j(x_{ij}-\bar{x}_i)^2/N$ is called the **within-group variance**. It can be shown that $V_{\mathrm{B}}+V_{\mathrm{W}}=V=\sum\sum(x_{ij}-\bar{x})^2/N$, i.e., the variance of all the observations is decomposed as a sum of the between-group and the within-group variances. The ratio V_{B}/V is called the **correlation ratio**, which is equal to the square of the multiple correlation coefficient of x and $y_1,\ldots,y_m$, where $y_i=1$ if the object is in the ith category and $y_i=0$ otherwise.

M. Ordinal Data

When the observation is not quantitative but there exists a natural ordering among the categories into which the objects are classified, the observation is said to be in an **ordinal scale**, or simply ordinal.

When two ordinal observations are made on the same set of N objects, we can define several measures of association between the two ordinal scales. For each pair of objects we define a variable c_{ij}, $i,j=1,\ldots,N$, $i\neq j$, as $c_{ij}=1$ if the ith object is classified as "better" (or "superior") than the jth object according to both of the measurements, $c_{ij}=-1$ if the orderings are different in the two scales, or $c_{ij}=0$ if they are in the same category according to either or both of the scales. A measure of association is then given by $S=\sum\sum c_{ij}/N(N-1)$,

which takes a value between -1 and $+1$ but usually cannot attain ± 1. Other ways of normalizing the sum $\sum\sum c_{ij}$ have been proposed.

Another method of calculating the association is the **scoring method**, i.e., giving a set of ordered real numbers to the categories of each of the scales and calculating the correlation coefficient between the scores. The simplest scores are $0, 1, \ldots, m-1$ when there are m categories, but other methods of scoring are also used. Scores that give the largest possible correlation are called **canonical scores**, which are obtained as the characteristic vectors of the matrices NN' and $N'N$, where N is the matrix of the contingency table.

N. Time Series Data

Time series data can be recorded in a continuous time scale, but usually measurements are made at discrete times, which are most commonly equally spaced. Hence we here denote them as $x(t)$, $t=1,2,\ldots,T$. First we consider the quantitative univariate case. The intertemporal change of $x(t)$ is often decomposed into three parts:

$$x(t)=m(t)+c(t)+e(t),$$

where $m(t)$ is called the **trend**, and represents the secular, systematic change of x; $c(t)$ is called the **cycle**, and represents the recurrent pattern of the change; and $e(t)$ is called the error or random fluctuation, and represents the irregular changes. Such a decomposition cannot be defined rigorously without assuming some probabilistic or stochastic model for $x(t)$, but it is intuitively clear and practically useful in many applications.

There are two ways to estimate the trend. One is to calculate the **moving average** $\hat{x}(t)=(x(t-k)+x(t-k+1)+\ldots+x(t)+\ldots+x(t+k))/(2k+1)$ and use it as an estimate of the trend of $x(t)$; here k should be chosen to substantially eliminate the cyclic and random parts. More generally we can use the weighted moving average defined as $x(t)=\sum_{j=-k}^{k} w(j)x(t+j)$, where $w(-j)=w(j)$ and $\sum w(j)=1$. The second method is to assume some functional form, usually a polynomial in t, for the trend: $m(t)=a_0+a_1t+\ldots+a_kt^k$, and to determine the coefficient by least squares, i.e., to calculate the values of $a_0, a_1, \ldots, a_k$ which minimize $\sum_t(x(t)-a_0-a_1t-\ldots-a_kt^k)^2$.

There are two cases of cyclical changes. One is the case when there is a clearly defined relevant external time period, such as the seasons of the year or the days of the week. In such cases the effects of such external periodical cycles must be eliminated, and the process which does that is called **seasonal adjustment** of the time series data. Various methods of seasonal adjustment have been proposed and applied, but none is definitive. The other case is where the cyclical changes are produced from the observed process itself; here, the length of the period and the pattern of the cyclical change must be estimated.

Now assume that the data do not contain any trend, or that the trend has been effectively eliminated. First we calculate the correlation coefficient between $x(t)$ and $x(t+s)$ by

$$r(s)=\frac{\sum(x(t)-\bar{x}'_s)(x(t+s)-\bar{x}''_s)}{\sqrt{\sum(x(t)-\bar{x}'_s)^2\sum(x(t+s)-\bar{x}''_s)^2}},$$

where

$$\bar{x}'_s=\sum_{t=1}^{T-s}x(t)/(T-s), \qquad \bar{x}''_s=\sum_{t=s+1}^{T}x(t)/(T-s),$$

or more simply by

$$r(s)=T\sum(x(t)-\bar{x})(x(t+s)-\bar{x})/\left((T-s)\sqrt{\sum(x(t)-\bar{x})^2}\right),$$

where $\bar{x}=\sum x(t)/T$. $r(s)$ is called the **serial correlation coefficient** or the **autocorrelation coefficient** of lag s. When there exists a clear and definite cyclical change of period s in the data, $r(s)$ is close to 1. The diagram in which the serial correlation coefficient $r(s)$ is plotted against s is called the **correlogram**. In order to see the cyclical properties of the data more clearly, we calculate the **power spectral density**

$$w(\lambda)=1+2\sum_s r(s)\cos\lambda s.$$

The graph of $w(\lambda)$ is called the **power spectrum**, or simply the **spectrum**. $w(\lambda)$ represents the square of the width of the sine curve of frequency $\lambda/2\pi$ or of period $2\pi/\lambda$ contained in the data. The spectral density is closely related to the intensity, defined by

$$I(\lambda)=\frac{1}{T}\left(\left(\sum_t x(t)\cos\lambda t\right)^2+\left(\sum_t x(t)\sin\lambda t\right)^2\right),$$

which is proportional to the square of the multiple correlation coefficient of $x(t)$ and the functions $\cos\lambda t$ and $\sin\lambda t$ and is large if the data contains a sine curve of frequency $\lambda/2\pi$. It can be shown that $I(\lambda)$ is approximately equal to $w(\lambda)V(x)/\pi$. The spectral density thus obtained usually oscillates irregularly and far from smoothly; hence smoothing by use of a "spectral window" is often applied (→ 421 Times Series Analysis).

When several observations are made in time series data, we speak of multivariate time series. Let $x_i(t)$ be the ith observation in the tth period. The correlation coefficient between $x_i(t)$ and $x_j(t+s)$ is called the **serial cross-correlation coefficient** and is denoted by $r_{ij}(s)$ $(s=0,\pm 1,\pm 2,\ldots)$. Analogously to the univari-

ate case we define

$$w_{ij}(\lambda)=\sum_s r_{ij}(s)\cos\lambda s+i\sum_s r_{ij}(s)\sin\lambda s$$
$$=p_{ij}(\lambda)+iq_{ij}(\lambda),$$

and we call $p_{ij}(\lambda)$ the **cospectral density** between the ith and the jth variables, and $q_{ij}(\lambda)$ the **quadrature spectral density**. $p_{ij}^2+q_{ij}^2$ is called the **amplitude**, and

$$C_{ij}(\lambda)=\frac{p_{ij}^2(\lambda)+q_{ij}^2(\lambda)}{\sqrt{w_i(\lambda)w_j(\lambda)}},$$

where w_i and w_j are the spectral densities of the ith and the jth variables, is called the **coherence**.

O. Events in Time Scale

Some data give us the time points at which a specific event occurs. Let T_i be the time when the event occurs for the ith time. Then usually the most important information we want to obtain is about the time intervals $d_i=T_{i+1}-T_i$. If there is a periodicity in the occurrences of the event, the d_i will be approximately equal. On the other hand, if the event tends to occur repeatedly after its first appearance, some d_i will be small while others will be large. When there is no periodicity, no tendency to repetition, and no increasing or decreasing trend in the occurrences, we can suppose that the event occurs simply by chance, and this is good reason to suppose that the density function of the distribution of the intervals is exponential, i.e., it can be expressed by $f(d)=(\exp(-d/\alpha))/\alpha$ for $d>0$. Also in such a case the number of occurrences in fixed time intervals are distributed according to the Poisson distribution. Such a sequence of occurrences of an event is called a †Poisson process. More generally, let $f(d)$ be the density function of the time intervals; then

$$h(d)=\frac{f(d)}{1-\int_0^d f(c)\,dc}$$

is called the **hazard rate** or **hazard function**. The hazard function is constant if and only if the process is Poisson.

P. Probabilistic Models

In many applications of statistical data analysis, the data exhibit variabilities and fluctuations that are due to fortuitous or hazardous causes or chance effects and that obscure the information contained in the data. In such cases we assume that the chance variabilities and fluctuations are random variables distributed according to some probability distribution, and the information we require is represented by a set of unknown constants that characterize the probability distribution of the data as unknown parameters. Suppose, to take the simplest case, that we have repeated observations of results of some experiment under a fixed condition and that we obtain the values $x_1, x_2, \ldots, x_n$. Since the experimental condition is fixed, the variations among the x_i values can be considered to be due to chance causes, such as variations in materials, uncontrolled small fluctuations in experimental conditions or instruments, and various other variations usually called the errors. Whatever the true causes of the variations, we can consider them to be random, and we can regard the values $x_1, x_2, \ldots, x_n$ as the results of random experiments or the realizations of random variables $X_1, X_2, \ldots, X_n$ independently and identically distributed according to some probability distribution. Or we may think of a **hypothetical infinite population** of the results of supposedly infinite replications of the experiment under the same fixed condition, and regard the actual observations as n values chosen from this population at random. We may also consider that in this hypothetical infinite population, the frequency distribution is represented by a density function f, which in turn determines the probability distribution of each observation. We may be interested in the "average" values of the result of the experiment as well as the magnitude of the variability; then those values are represented by the mean and the variance of the population distribution. If the form of the population distribution is assumed to be completely specified except for the mean μ and the variance σ^2, the density function f is determined without these two parameters, and is expressed as $f(x;\mu,\sigma)$. The joint density for n repeated observations is $\prod_i f(x_i;\mu,\sigma)$. The set of assumptions that determines the probability distributions of the observations in terms of the unknown parameters is called the **probabilistic model**, and its determination is called the **problem of specification**.

Once the probabilistic model is given, the purpose of statistical data analysis can be formulated as making judgments on the values of the parameters, which may sometimes go wrong but can be relied on with some margin of probability of error that can be mathematically rigorously ascertained. The formal procedure of making such judgments is called statistical inference, and its mathematical theory has been well established over the last hundred years (→ 401 Statistical Inference).

In most cases of statistical inference, the joint density function of the data plays an important role, and when it is regarded as a function of the unknown parameters for given

values of observations it is called the †likelihood function.

Q. Exploratory Procedures

In many applications of statistical data analysis, we are not quite sure of the validity of the probabilistic model assumed, or we admit that the models are, at best, approximations to reality and hence cannot be exactly correct; the approximation may not be precise enough for the conclusions drawn from the assumptions to be practically reliable. Therefore we have to check whether the model assumed is at least approximately valid for the data, and if not, we have to look for a better model that reflects more accurately the structure of the actual data. Thus in many practical applications of data analysis, we have to scrutinize the structure of the data and try various models before settling on a model and drawing final conclusions (which are still susceptible to further revisions when more data are obtained). Methods used in such a process are called **exploratory procedures**, which depend partly on the formal procedure of testing hypotheses and partly on intuitive reasoning sometimes combined with graphical presentations of the data, and also on scientific and empirical understanding of the subject matter.

Suppose in the simplest case that n observations $X_1, \ldots, X_n$ are assumed to be independently and identically distributed. Under the condition that all those values are observed under the same well-controlled situation, this assumption is reasonable. But in reality some of the observations may be subject to some unexpected effect due to either a fortuitous outside cause or some "gross error" in the measurement procedure, the process of reporting, etc., and may show much greater variation than others. Such observations can be detected by certain **outlier tests** or simply by looking at the data carefully, and if it is established that some observations are definitely outside of the possible random variability or are subject to some hazardous external effect, those data could be omitted from consideration. Further, the assumed probability density $f(x, \theta)$ may not well approximate the distribution of the actual data even after the "outliers" are omitted. Some test for **goodness of fit** should be applied, and if the hypothesis is rejected, we have to modify our model. Also, if we are provided with several candidates for the model to be adopted, we have to apply some procedure of model selection (→ 400 Statistical Hypothesis Testing, 403 Statistical Models). It could also happen that the supposedly uniform conditions of observation, measurement, or experimentation did not actually prevail but that there has been some heterogeneity among the observations. Bimodality or multimodality of the histogram, i.e., existence of two or more peaks in the histogram, usually strongly suggests such heterogeneity. In such cases, grouping or stratification of the observations is required to make the conditions of observation within each group nearly uniform.

References

[1] R. A. Fisher, Statistical methods for research workers, Oliver & Boyd, first edition, 1925.
[2] M. G. Kendall and A. Stuart, The advanced theory of statistics, 3 vols., Griffin, 1958–1966.
[3] J. W. Tukey, Exploratory data analysis, Addison-Wesley, 1977.

398 (XVIII.6) Statistical Decision Functions

A. General Remarks

The theory of statistical decision functions was established by A. Wald as a mathematically unified theory of statistics (→ 401 Statistical Inference). In this theory, the problems of mathematical statistics, for example, statistical hypothesis testing and †statistical estimation, are formulated in a unified way [1].

A †measurable space $(\mathscr{X}, \mathfrak{B})$ with a fixed †probability measure is called a **sample space**, and an element $x \in \mathscr{X}$ is called a **sample point**. Suppose that we are given a family $\mathscr{P} = \{P_\theta \mid \theta \in \Omega\}$ of probability measures on $(\mathscr{X}, \mathfrak{B})$, where Ω is called the **parameter space**, and a †random variable X takes values in $\mathscr{X}$ according to a true probability distribution P assumed to belong to $\mathscr{P}$. This article deals with the problems involved in making a decision about the parameter θ, called determining the **true value of the parameter**, such that $P = P_\theta$. To describe the procedure for such a decision based on the observation of the behavior of X, we need a triple $(\mathscr{A}, \mathfrak{C}; D)$ consisting of a set $\mathscr{A}$, a †σ-algebra $\mathfrak{C}$ of subsets of $\mathscr{A}$, and a set D of mappings δ from $\mathscr{X}$ into the set of probability measures on $(\mathscr{A}, \mathfrak{C})$, $\delta: x \to \delta(\cdot \mid x)$, such that for a fixed $C \in \mathfrak{C}$, the function $\delta(C \mid x)$ is $\mathfrak{B}$-measurable. We call $\mathscr{A}$ an **action space** or **decision space**, δ a **statistical decision function** (or simply a **decision function**) or **statistical deci-**

sion procedure, and D a **space of decision functions**. In actual decision procedures, $\delta(C|x)$ is the probability that an action belonging to C is taken, based on the observation of sample point x. We further consider a nonnegative function $w:\Omega\times\mathscr{A}\to\mathbf{R}$, called a **loss function**, such that for a fixed θ, $w(\theta,a)$ is $\mathfrak{C}$-measurable. By averaging the loss, we obtain the **risk function**:

$$r(\theta,\delta)=\int_{\mathscr{X}}\int_{\mathscr{A}}w(\theta,a)\delta(da|x)P_\theta(dx).$$

Two decision functions δ and δ' are identified if $\delta(C|x)=\delta'(C|x)$ for almost every x with respect to P_θ, for all $\theta\in\Omega$ and all $C\in\mathfrak{C}$. When to each $x\in\mathscr{X}$ there corresponds a unique action a_x such that $\delta(\{a_x\}|x)=1$, the decision function δ is said to be **nonrandomized**; otherwise, **randomized**. The system $(\mathscr{X},\mathfrak{B},\mathscr{P},\Omega,\mathscr{A},\mathfrak{C},W,D)$ is called a **statistical decision problem**.

From the point of view of this theory, †point estimation, †interval estimation, and statistical hypothesis testing (→ 400 Statistical Hypothesis Testing) are described as follows.

(1) In point estimation we assume that the action space $\mathscr{A}$ is a subset of $\mathbf{R}$ and that we are given functions $\varphi:\mathscr{X}\to\mathscr{A}$ and $l:\mathscr{P}\to\mathbf{R}$. The problem is to estimate the value of $l(P)$ by using the real value $\varphi(x)$ at an observed sample point $x\in\mathscr{X}$. As the loss function, we often set $w(\theta,a)=C(\theta)(a-l(P_\theta))^2$, where $C(\theta)$ is a function of θ, and call it a **quadratic loss function** (→ 399 Statistical Estimation).

(2) In interval estimation we assume that each action is represented as an interval in $\mathbf{R}$. Each interval $[u,v]$ can be represented by a point (u,v) of the half-space $\mathbf{R}^{2\prime}=\{z=(z_1,z_2)|z_1\leqslant z_2\}$, which may be taken as the action space. A weighted sum $\alpha w_1(\theta,z)+\beta w_2(\theta,z)$ $(\alpha,\beta\geqslant0)$ of two functions,

$$w_1(\theta,z)=\begin{cases}1 & (\theta\notin[z_1,z_2]),\\ 0 & (\theta\in[z_1,z_2]),\end{cases}$$

$$w_2(\theta,z)=z_2-z_1,$$

often supplies the loss function (→ 399 Statistical Estimation).

(3) In testing a hypothesis $\mathbf{H}:\theta\in\omega_0$ versus an †alternative $\mathbf{A}:\theta\in\omega_1$ $(\omega_0\cap\omega_1=\varnothing,\ \omega_0\cup\omega_1=\Omega)$, the action space can be expressed by the set consisting of two points a_1,a_2, where a_1 denotes the decision to reject $\mathbf{H}$ and a_2 the decision to accept $\mathbf{H}$. The loss function is defined as follows:

$$w(\theta,a_1)=\begin{cases}1 & (\theta\in\omega_0),\\ 0 & (\theta\in\omega_1),\end{cases}$$

$$w(\theta,a_2)=\begin{cases}0 & (\theta\in\omega_0),\\ 1 & (\theta\in\omega_1).\end{cases}$$

This is called a **simple loss function**. Whatever testing procedure δ is adopted, the probabilities of the †errors of the first or second kind coincide with the values of $r(\theta,\delta)$ for $\theta\in\omega_0$ or $\theta\in\omega_1$, respectively (→ 400 Statistical Hypothesis Testing).

When Ω is the union of mutually disjoint nonempty sets $\omega_1,\omega_2,\ldots,\omega_n$, $\mathscr{A}=\{a_1,a_2,\ldots,a_n\}$, and $w(\theta,a_j)=c_{ij}$ $(\theta\in\omega_i)$ with $c_{ij}\geqslant0$ $(i\neq j)$, $c_{ii}=0$, the decision problem is called an **n-decision problem**.

B. Optimality of Statistical Decision Functions

Consider the problem of choosing the best decision function δ. When $r(\theta,\delta_1)\leqslant r(\theta,\delta_2)$ for all θ and there exists at least one θ_0 such that $r(\theta_0,\delta_1)<r(\theta_0,\delta_2)$, the decision function δ_1 is said to be **uniformly better** than δ_2. If there exists a decision function δ_0 in D that is uniformly better than any other δ in D, it is the **best** decision function. However, such a function δ_0 does not always exist. A decision function $\tilde\delta$ in D is said to be **admissible** if there exists no other decision function δ in D that is uniformly better than $\tilde\delta$. In other words, $\tilde\delta$ is admissible if and only if the validity of the inequality $r(\theta,\delta)\leqslant r(\theta,\tilde\delta)$ for some $\delta\in D$ and all $\theta\in\Omega$ implies $r(\theta,\delta)=r(\theta,\tilde\delta)$ for all $\theta\in\Omega$. When there is no information about P except that it is a member of $\mathscr{P}$, we follow the **minimax principle** and choose a function δ^* for which we have $\inf_{\delta\in D}\sup_{\theta\in\Omega}r(\theta,\delta)=\sup_{\theta\in\Omega}r(\theta,\delta^*)$. This decision function δ^* is called a **minimax decision function** or **minimax solution**.

Let $\mathfrak{F}$ be a σ-algebra of subsets of Ω, and suppose that $r(\theta,\delta)$ is $\mathfrak{F}$-measurable for any fixed δ. If, furthermore, we are given a probability measure ξ, called an **a priori distribution**, on $(\Omega,\mathfrak{F})$, we choose a $\hat\delta$ that satisfies

$$\inf_{\delta\in D}\int_\Omega r(\theta,\delta)d\xi(\theta)=\int_\Omega r(\theta,\hat\delta)d\xi(\theta).$$

Such a $\hat\delta$ and the integral of $r(\theta,\hat\delta)$ are called a **Bayes solution** and the **Bayes risk** relative to ξ, respectively. Let F be a family of a priori distributions on $(\Omega,\mathfrak{F})$. If $\hat\delta$ satisfies

$$\inf_{\xi\in F}\left(\int_\Omega r(\theta,\hat\delta)d\xi(\theta)-\inf_{\delta\in D}\int_\Omega r(\theta,\delta)d\xi(\theta)\right)=0,$$

$\hat\delta$ is called a **Bayes solution in the wider sense** relative to F. If $\mathscr{P}$ is †dominated by λ with a $\mathfrak{B}\times\mathfrak{F}$-measurable $f(x,\theta)=dP_\theta/d\lambda$, $w(\theta,a)$ is $\mathfrak{F}\times\mathfrak{C}$-measurable, and

$$A_x=\left\{\tilde a\in\mathscr{A}\,\middle|\,\int_\Omega w(\theta,\tilde a)f(x,\theta)d\xi(\theta)=\inf_{a\in\mathscr{A}}\int_\Omega w(\theta,a)f(x,\theta)d\xi(\theta)\right\}$$

is nonempty and $\mathfrak{C}$-measurable for λ-almost

every x, then a Bayes solution δ with respect to ξ satisfies $\delta(A_x|x)=1$ for λ-almost every x. For a sample point x with $\int_\Omega f(x,\theta)d\xi(\theta)\neq 0$, a probability measure $\eta(\cdot|x,\xi)$ on $(\Omega,\mathfrak{F})$ defined by

$$\eta(B|x,\xi)=\frac{\int_B f(x,\theta)d\xi(\theta)}{\int_\Omega f(x,\theta)d\xi(\theta)}$$

is called an **a posteriori distribution**. To get a Bayes solution it is enough to minimize the value of $\int_\Omega w(\theta,a)d\eta(\theta|x,\xi)$ for every observed x.

If ξ in the definition of A_x defined above is a σ-finite measure on $(\Omega,\mathfrak{F})$, a decision function δ satisfying $\delta(A_x|x)=1$ for λ-almost every x is called a **generalized Bayes solution** with respect to ξ.

Let D' be a subset of the space D. If for any $\delta\in D-D'$ there exists a $\delta'\in D'$ that is uniformly better than δ, then D' is called a **complete class**. If for any $\delta\in D$ there exists a $\delta'\in D'$ that is either uniformly better than δ or has the same risk function as δ, then D' is called an **essentially complete class**. If D' is complete and any proper subset of D' is not complete, then D' is called a **minimal complete class**. If a minimal complete class exists, it is unique and coincides with the set of all admissible decision functions.

C. n-Decision Problems

In an n-decision problem where $\mathscr{A}=\{a_1,\ldots,a_n\}$, we set $\delta_i(x)=\delta(a_i|x)$ for a decision function δ, where $\delta_i(x)$ is $\mathfrak{B}$-measurable and satisfies $\delta_i(x)\geqslant 0$, $\delta_1(x)+\ldots+\delta_n(x)=1$. We consider the set $\mathscr{D}$ of vector-valued functions $\Delta(x)=(\delta_1(x),\ldots,\delta_n(x))$ whose components $\delta_i(x)$ satisfy the conditions just given. Such a vector-valued function $\Delta(x)$ can be identified with $\delta(x)$; we write $\delta(x)$ instead of $\Delta(x)$ also. If in addition the parameter space Ω is a finite set $\{\theta_1,\theta_2,\ldots,\theta_k\}$, we can consider a mapping $\psi:\mathscr{D}\to\mathbf{R}^k$ defined by $\psi(\delta)=(r(\theta_1,\delta),\ldots,r(\theta_k,\delta))$, and then $S=\psi(\mathscr{D})=\{\psi(\delta)|\delta\in\mathscr{D}\}$ is convex and closed in $\mathbf{R}^k$. If δ is nonrandomized, then for each x, one and only one of the $\delta_i(x)$ is 1, and all others are 0. Hence, in this case, $\mathscr{X}$ is the disjoint union of $\mathfrak{B}$-measurable subsets B_i $(i=1,\ldots,n)$ such that $\delta(a_i|x)=1$ if and only if $x\in B_i$. A probability measure m on $(\mathscr{X},\mathfrak{B})$ is said to be **atomless** if for any set $A\in\mathfrak{B}$ with $m(A)>0$ and any b with $0<b<m(A)$, there exists a subset $B\in\mathfrak{B}$ of A such that $m(B)=b$. If Ω is finite and every member of $\mathscr{P}$ is atomless, the image $\psi(\mathscr{D}^0)$ of the set $\mathscr{D}^0$ of all nonrandomized decision functions coincides with $\psi(\mathscr{D})$. This shows that $\mathscr{D}^0$ is an essentially complete class in $\mathscr{D}$. However, when some members of $\mathscr{P}$ are not atomless, $\psi(\mathscr{D})$ is not always equal to $\psi(\mathscr{D}^0)$, but $\psi(\mathscr{D}^0)$ is a closed subset of $\mathbf{R}^k$. In particular, when $n=2$ for given probability measures $P_1,\ldots,P_k$ over $(\mathscr{X},\mathfrak{B})$, the set S of all points $(P_1(B),\ldots,P_k(B))$ $(B\in\mathfrak{B})$ in $\mathbf{R}^k$ is a closed set. If in addition $P_1,\ldots,P_k$ are all atomless, S is convex. These results are known as the **Lyapunov theorem**.

A two-decision problem with $\omega_1=\{1\}$, $\omega_2=\{2\}$, and $c_{ij}=1$ $(i\neq j)$, $=0$ $(i=j)$ is called a **dichotomy**. We discuss this problem in some detail in order to explain the concept of optimality of decision functions. For a dichotomy, $S=\psi(\mathscr{D})$ is a set in $\mathbf{R}^2$ that (i) is convex, (ii) is closed, (iii) is symmetric about $(1/2,1/2)$, (iv) contains the points $(0,1)$, $(1,0)$, and (v) is a subset of interval $[0,1]\times[0,1]$ (Fig. 1). The set of decision functions δ that is mapped under ψ onto the curve $ACDB$ in Fig. 1 constitutes the minimal complete class, and a decision function δ mapped onto the point D is a minimax solution. Let ξ be an a priori distribution such that

$$\xi(1)=\alpha,\quad \xi(2)=\beta\quad(\alpha+\beta=1,\alpha\geqslant 0,\beta\geqslant 0).$$

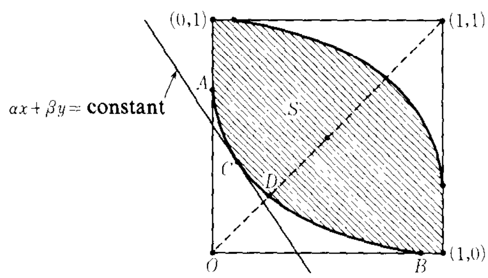


Fig. 1

Then a Bayes solution relative to ξ is mapped under ψ onto a supporting point c with direction ratio $-\alpha/\beta$ and is obtained as the characteristic function of the set $E=\{x|\alpha f(x)<\beta g(x)\}$, where f and g are †Radon-Nikodym derivatives $dP_1/d\lambda$ and $dP_2/d\lambda$ with respect to $\lambda=P_1+P_2$, that is, f, g are measurable functions on $\mathscr{X}$ such that for any measurable set E, we have

$$P_1(E)=\int_E fd\lambda,\qquad P_2(E)=\int_E gd\lambda.$$

This fact implies that the most powerful test constructed in the Neyman-Pearson fundamental lemma (→ 400 Statistical Hypothesis Testing B) is precisely a Bayes solution.

D. Complete Class Theorems

Suppose that $\mathscr{P}$ is †dominated by a σ-finite measure λ, and let $f(x,\theta)$ be the Radon-Nikodym derivative of P_θ with respect to λ.

Consider a subspace L of $L_1(\mathscr{X}, \lambda)$ containing $\{f(x,\theta) | \theta \in \Omega\}$, and the following equivalence relation between bounded $\mathfrak{B}$-measurable functions on $\mathscr{X}$: φ_1 and φ_2 are defined to be equivalent if and only if $\int \varphi_1 f d\lambda = \int \varphi_2 f d\lambda$ holds for every $f \in L$. We further assume that the dual space of L is the linear space $\mathfrak{M}$ of the equivalence classes of bounded measurable functions just defined. Let $C_0(\mathscr{A})$ be the set of all continuous functions on $\mathscr{A}$ with compact support, $\pi \circ \alpha$ the integral of $\alpha \in C_0(\mathscr{A})$ with respect to a probability measure π over $\mathscr{A}$, and $g \circ \delta$ the integral

$$\int \delta(\cdot | x) g(x) d\lambda$$

for $g \in L$, $\delta \in D$. As a base for neighborhoods of the space of decision functions around $\delta_0 \in D$, we consider $V(\delta_0 : \alpha_1, \ldots, \alpha_n, g_1, \ldots, g_n, \varepsilon) = \{\delta \mid |g_i \circ \delta_0 \circ \alpha_i - g_i \circ \delta \circ \alpha_i| < \varepsilon, i = 1, \ldots, n\}$, where $\alpha_i \in C_0(\mathscr{A})$, $g_i \in L$, $\varepsilon > 0$.

Let F be the set of all a priori distributions ξ on $(\Omega, \mathfrak{F})$ each of which assigns the total mass 1 to a finite subset of Ω, B the set of all Bayes solutions relative to some $\xi (\in F)$, and W the set of all Bayes solutions in the wide sense relative to F. Suppose that $\mathscr{A}$ is a †locally compact, separable metric space and $w(\theta, a)$ is lower semicontinuous with respect to a for every fixed θ. Then if D is compact and convex, the intersection of W and the closure $\bar{B}$ of B constitutes an essentially complete class [2]. Moreover, if Ω is compact with respect to the metric $d(\theta_1, \theta_2) = \sup_{B \in \mathfrak{B}} |P_{\theta_1}(B) - P_{\theta_2}(B)|$ and $\{w(\cdot, a) | a \in \mathscr{A}\}$ is a uniformly bounded and †equicontinuous family on Ω, then the class of Bayes solutions relative to some a priori distribution is a complete class [1]. These propositions are called **complete class theorems** (for complete classes in specified problems and admissibility of individual procedures → 399 Statistical Estimation; 400 Statistical Hypothesis Testing; and Appendix A, Table 23).

Two measurable spaces $(S, \mathscr{S})$ and $(R, \mathscr{R})$ are said to be **isomorphic** if there exists a correspondence ρ of S and R ($E \in \mathscr{S}$ implies $\rho(E) \in \mathscr{R}$ and, conversely, $\rho(E) \in \mathscr{R}$ implies $E \in \mathscr{S}$) where ρ is one-to-one onto. Let $\mathscr{D}$ be the set of all decision functions associated with a sample space $(\mathscr{X}, \mathfrak{B})$ and an action space $(\mathscr{A}, \mathfrak{C})$, T be a †statistic on $(\mathscr{X}, \mathfrak{B})$ taking values in a measurable space $(\mathscr{Y}, \mathfrak{E})$, and $\mathscr{D}^*$ be the set of all decision functions having sample space $(\mathscr{Y}, \mathfrak{E})$ and action space $(\mathscr{A}, \mathfrak{C})$. The set of all $\delta \in \mathscr{D}$ for which there exists a $\delta^* \in \mathscr{D}^*$ satisfying $\delta(C | x) = \delta^*(C | T(x))$ $(C \in \mathfrak{C})$ is denoted by $\mathscr{D}_T$. If (i) T is a †sufficient statistic and if (ii) $(\mathscr{A}, \mathfrak{C})$ is isomorphic to the measurable space $\mathbf{R}^k$ associated with the σ-algebra of its Borel subsets, then $\mathscr{D}_T$ is essentially complete in $\mathscr{D}$. Conversely, if (i) $\mathscr{P}$ is a dominated family, (ii) $f(x, \theta) > 0$ always holds on $\mathscr{X}$, and (iii) for any pair $\theta_1, \theta_2 (\in \Omega)$, there exists a $\theta_3 (\in \Omega)$ such that no element a in $\mathscr{A}$ makes two of the $w(\theta_i, a)$ $(i = 1, 2, 3)$ attain their minimum simultaneously, then essential completeness of $\mathscr{D}_T$ in $\mathscr{D}$ implies that T is sufficient [3] (→ 399 Statistical Estimation).

E. Invariance

Suppose that there exist one-to-one transformation groups G, $\bar{G}$, and $\tilde{G}$ of X, Ω, and $\mathscr{A}$, respectively, onto themselves (transformations belonging to G, $\bar{G}$, and $\tilde{G}$ are †measurable with respect to $\mathfrak{B}$, $\mathfrak{F}$, and $\mathfrak{C}$, respectively) and that there exist homomorphisms $g \to \bar{g}$, $g \to \tilde{g}$ of G to $\bar{G}$ and $\tilde{G}$, respectively, such that for $B \in \mathfrak{B}$ we have $P_\theta(g^{-1}B) = P_{\bar{g}\theta}(B)$ and $w(\bar{g}\theta, \tilde{g}a) = w(\theta, a)$. Then a decision problem $(\mathscr{X}, \mathfrak{B}, \mathscr{P}, \Omega, \mathscr{A}, \mathfrak{C}, w, D)$ is said to be **invariant** under $(G, \bar{G}, \tilde{G})$. In a decision problem invariant under $(G, \bar{G}, \tilde{G})$, a decision function satisfying $\delta(\tilde{g}C | gx) = \delta(C | x)$ is called an **invariant decision function**.

Suppose that the transformation group G is locally compact and is the union of a countable family $\{K_n\}$ of compact subsets. Let Γ be the σ-algebra of Borel subsets of G such that the mapping $(g, x) \to (g, gx)$ is measurable in the sense that the inverse image of any set in $\Gamma \times \mathfrak{B}$ is also a set in $\Gamma \times \mathfrak{B}$. For such a Γ the †orbit $G(x)$ of G through x is Γ-measurable. We assume that following conditions: (1) $\mathscr{P}$ is dominated; (2) G operates †effectively on $\mathscr{X}$; (3) $\bar{G}$ operates †transitively on Ω; (4) for any compact subset J of G,

$$\lim_{n \to \infty} \{\mu^r(K_n) / \mu^r(K_n \cdot J^{-1})\} = 1,$$

where μ^r is a right-invariant Haar measure (→ 225 Invariant Measures C) on G and $K_n \cdot J^{-1} = \{gh^{-1} | g \in K_n, h \in J\}$; (5) there is a †conditional probability distribution $P_\theta(\cdot : z)$, given $z \in G(x)$, on $\mathscr{X}$; (6) the integral $\int_{G(x)} w(\theta, \tilde{g}a) P_\theta(dgx : z)$ attains its minimum value $b(\theta, z)$ for any θ and $z \in G(x)$; and (7)

$$\lim_{n \to \infty} \left\{ \int_{K_n x} w(\theta, \tilde{g}a) P_\theta(dgx : z) - b(\theta, z) \right\} \geqslant 0$$

uniformly in a, where $K_n x = \{gx | g \in K_n\}$. Then the best invariant decision function exists which is also minimax in $\mathscr{D}$ [4] (→ 400 Statistical Hypothesis Testing). It is shown in [9] that some invariant minimax decision functions are not admissible.

F. Sequential Decision Problems

In the general framework of statistical decision theory, not only the decisions to be taken but the number of samples to be observed

may be determined based on the previous observations.

A simple formulation is illustrated for the sequential decision problem given below. Suppose that $X_1, X_2, \ldots, X_n$ are independently and identically distributed random variables with probability measure P_θ. We assume that the X's are to be observed one by one, and at the ith stage, when we have observed $X_1, \ldots, X_i$, we are to decide whether to continue sampling and observe X_{i+1} or to stop observation and choose an action or decision in $\mathscr{D}$, utilizing all the observations thus far obtained. Then a decision rule is defined by a sequence of pairs (δ_n, s_n), $n=0,1,\ldots$, where δ_n is a mapping from the space X^n to $\mathscr{D}$ (for the sake of simplicity we here exclude randomized decisions), and $s_n = s_n(x_1, \ldots, x_n)$ is a measurable function from X^n to the interval $[0,1]$. s_n gives the probability of stopping the sampling when the first n of the X's are observed, and δ_n defines the decision taken when the observations are stopped. We include s_0 and δ_0 to denote the probability of taking a decision without making any observation and the decision to be taken then. We call δ_n the **terminal decision rule** and s_n the **stopping rule**. Then for such a decision rule $\tilde{\delta}$ the total expected loss or the risk is given by

$$r(\theta,\tilde{\delta}) = \sum_{n=0}^{\infty} \int \prod_{j=0}^{n-1} (1 - s_j(x_1, \ldots, x_j)) \times \{s_n(x_1, \ldots, x_n) w(\theta, \delta_n(x_1, \ldots, x_n)) + c_n(x_n | x_1, \ldots, x_{n-1})\} \prod_{j=1}^{n} dP_\theta(x_j),$$

where $c_n(x_n | x_1, \ldots, x_{n-1})$ is the **cost of observation** of $X_n = x_n$ when $X_1 = x_1, \ldots, X_{n-1} = x_{n-1}$. The rule $\tilde{\delta}$ is called a **sequential decision rule** or a **sequential decision function**, and the whole setup a **sequential decision problem**. In most of the sequential decision problems the cost of observation is assumed to be equal to a constant c per observation, and then the Bayes risk $r^*(\xi_\theta)$ for the prior distribution ξ_θ satisfies the relation

$$r^*(\xi_\theta) = \min\left\{ \inf_d \int w(\theta, d)\, d\xi_\theta, \iint r^*(\pi_\theta(x_1, \xi_\theta))\, dP_\theta(x_1)\, d\xi_\theta + c \right\},$$

where $\pi_\theta(x_1, \xi_\theta)$ denotes the posterior distribution when $X_1 = x_1$ is observed under the prior distribution ξ_θ; and the Bayes decision rule satisfies the condition

$$s_n(x_1, \ldots, x_n) = 1 \quad \text{if } r^*(\pi_\theta(x_1, \ldots, x_n; \xi_\theta)) = \inf \int w(\theta, d)\, d\pi_\theta(x_1, \ldots, x_n; \xi_\theta),$$
$$= 0 \quad \text{otherwise},$$

and $\delta_n(x_1, \ldots, x_n) = d^*$ if $\int w(\theta, d^*)\, d\pi_\theta(x_1, \ldots, x_n; \xi_\theta) = \inf_d \int w(\theta, d)\, d\pi_\theta(x_1, \ldots, x_n; \xi_\theta)$, where $\pi_\theta(x_1, \ldots, x_n; \xi_\theta)$ is the posterior distribution given $X_1 = x_1, \ldots, X_n = x_n$ under the prior ξ_θ.

For the dichotomy problem with the simple loss function discussed above, the Bayes decision rule has the form

$$s_n(x_1, \ldots, x_n) = 1 \quad \text{if } \pi_0(x_1, \ldots, x_n) \leqslant \pi_1 \text{ or } \geqslant \pi_2,$$
$$= 0 \quad \text{otherwise},$$

and

$$\delta_n(x_1, \ldots, x_n) = d_1 \quad \text{if } \pi_0(x_1, \ldots, x_n) \leqslant \pi_1,$$
$$= d_0 \quad \text{if } \pi_0(x_1, \ldots, x_n) \geqslant \pi_2,$$

where $\pi_0(x_1, \ldots, x_n)$ is the posterior probability that $\theta = \theta_0$, given $X_1 = x_1, X_n = x_n$. This amounts to the following rule: Continue sampling as long as

$$\gamma_1 < \lambda(x_1, \ldots, x_n) = \prod_i P_{\theta_0}(x_i) / \prod P_{\theta_1}(x_i) < \gamma_2,$$

and decide on d_0 as soon as $\lambda(x_1, \ldots, x_n) \geqslant \gamma_2$ and on d_1 as soon as $\lambda(x_1, \ldots, x_n) \leqslant \gamma_1$, which is actually equivalent to a †sequential probability ratio test (→ 400 Statistical Hypothesis Testing).

G. Information in Statistical Experiments

The part $\mathscr{E} = (\mathscr{X}, \mathfrak{B}, \mathscr{P}, \Omega)$ of a decision problem $\Delta = (\mathscr{X}, \mathfrak{B}, \mathscr{P}, \Omega, \mathscr{A}, \mathfrak{C}, w, D)$ is called a **statistical experiment**. In this section, we consider n-decision problems, i.e., those wherein Ω consists of a finite number $\{1, 2, \ldots, n\}$ of states, and we denote the set $S = \{r(1,\delta), r(2,\delta), \ldots, r(n,\delta) | \delta \in \mathscr{D}\}$ by $L(\Delta)$. Let $\mathscr{E}_1$ and $\mathscr{E}_2$ be two experiments having a common parameter space Ω, and let Δ_1 and Δ_2 be two decision problems composed of $\mathscr{E}_1$ and $\mathscr{E}_2$, respectively, and a common $(\mathscr{A}, \mathfrak{C}, w)$. We say that the experiment $\mathscr{E}_1$ is **more informative** than the experiment $\mathscr{E}_2$ if $L(\Delta_1) \supset L(\Delta_2)$ for any action space $(\mathscr{A}, \mathfrak{C})$ and any loss function w [6]; that is, whatever the actions proposed and the loss incurred, the experiment $\mathscr{E}_1$ can offer a decision procedure at least as efficient as the experiment $\mathscr{E}_2$. Thus the set $L(\Delta)$ with $\Delta = (\mathscr{E}, \mathscr{A}, \mathfrak{C}, w, D)$ represents some feature of information that $\mathscr{E}$ can provide about the states Ω. However, comparison of $L(\Delta)$ is not easy to carry out. S. Kullback and R. A. Leibler defined the concept of information for the case of a dichotomy Δ [5]. If the probability distribution induced by a random variable X has a Radon-Nikodym derivative $f_1(x)$ (>0) or $f_2(x)$ (>0) with respect to λ, we define

$$I(X:1,2) = I(f_1, f_2) = \int_{\mathscr{X}} \log \frac{f_1(x)}{f_2(x)} f_1(x)\, d\lambda,$$

calling this the **Kullback-Leibler information number** (or **K-L information number**). This number is uniquely determined by the set $S = L(\Delta)$ in Fig. 1, and the larger S becomes, the larger $I(f_1, f_2)$ becomes. If α, β $(\alpha+\beta=1)$ are a priori probabilities and $\eta(1|x)$ and $\eta(2|x)$ are a posteriori probabilities of 1 and 2, respectively, we have, from the Bayes theorem,

$$\log\frac{f_1(x)}{f_2(x)} = \log\frac{\eta(1|x)}{\eta(2|x)} - \log\frac{\alpha}{\beta}.$$

Here the right-hand side stands for the change in the probability of the occurrence of the state after an observation of x, and the †expectation of the left-hand side under f_1 is $I(f_1, f_2)$. The K-L information number has the following properties: (i) $I(X:1,2) \geqslant 0$, and $I(X:1,2)=0 \Leftrightarrow f=g$; (ii) †independence of X and Y implies $I(X:1,2)+I(Y:1,2)=I(X,Y:1,2)$; (iii) for a statistic $T=t(x)$, $I(X:1,2) \leqslant I(T:1,2)$, where the equality holds if and only if T is sufficient for $\mathscr{P}=\{P_1, P_2\}$, in which $dP_1/d\lambda = f_1$ and $dP_2/d\lambda = f_2$; and, as a result of (ii), (iv) if $X_1, \ldots, X_n$ are distributed independently with the same distribution, $I(X_1, \ldots, X_n:1,2) = nI(X_1:1,2)$.

Suppose next that Ω is the real line and that the Radon-Nikodym derivative $g(t:\theta)$ of the distribution of a statistic T with respect to a measure λ has the following properties: (i) the set of all t at which $g(t:\theta)>0$ is independent of θ; (ii) $g(t:\theta)$ is continuously twice differentiable; and (iii) the order of differentiation with respect to θ and integration with respect to t can be interchanged. Then $I(T:\theta, \theta+d\theta) = I(T:\theta)d\theta^2$ for an infinitesimal displacement $d\theta$ of θ, where

$$I(T:\theta) = \int_{\mathscr{X}} \left(\frac{\partial \log g(t:\theta)}{\partial\theta}\right)^2 g(t:\theta)\,d\lambda.$$

Here $I(T:\theta)$ coincides with the Fisher information (→ 399 Statistical Estimation).

Suppose that we are given a sequence $(X_1, X_2, \ldots)$ of independent random variables whose distributions have as their density either $(f_1, f_2, \ldots)$ or $(g_1, g_2, \ldots)$, that is, f_i and g_i are candidates for the density of distribution of X_i $(i=1,2,\ldots)$. A method to determine which of the sequences (f_i), (g_i) actually corresponds to (X_i) is given by the **Kakutani theorem**. Let F be the distribution of $(X_1, X_2, \ldots)$ when each X_i is distributed according to f_i, and let G be that of $(X_1, X_2, \ldots)$ when each X_i is distributed according to g_i. To see how $X_1, X_2, \ldots$ are actually distributed, we assume that the loss incurred by an incorrect decision is 1 and the loss incurred by a correct decision is 0. Denote such a decision problem (dichotomy) by Δ. Then we generally have $L(\Delta) \subset I$, where $I=\{(x,y) | 0 \leqslant x, y \leqslant 1\}$. A necessary and sufficient condition for $L(\Delta)=I$ is that $\prod_{n=1}^{\infty} \rho(f_n, g_n)=0$, where

$$\rho(f_n, g_n) = \int_x \sqrt{f_n(x) g_n(x)}\,dx.$$

Consequently, if the X_n $(n=1,2,\ldots)$ have the same distribution, we have $L(\Delta)=I$, and the correct decision can be made with no error based on infinitely many independent observations of $X_1, X_2, \ldots$.

H. Relation to Game Theory

The theory of statistical decision functions is closely related to game theory. From the game-theoretic viewpoint, a statistical decision problem is considered to be a zero-sum two-person game played by the statistician against nature. A strategy of nature is the true distribution P of the variable X or the true value of θ, and a strategy of the statistician is a decision δ. In this setup, the risk function $r(\theta, \delta)$ can be regarded as a †payoff function paid by the statistician to nature. An a priori distribution ξ is a †mixed strategy of nature. A randomized decision function is a mixed strategy of the statistician. A minimax decision function corresponds to a minimax strategy of the statistician. A minimax strategy of nature is called a **least favorable a priori distribution**. If a decision problem is †strictly determined as a game, a minimax solution is a Bayes solution in the wide sense.

If δ_0 is a Bayes solution with respect to ξ_0 and $r(\theta, \delta_0) \leqslant R(\xi_0, \delta_0)$ for all θ, the decision problem is strictly determined, and δ_0 is minimax and ξ_0 is a least favorable a priori distribution, where $R(\xi_0, \delta_0) = \int_\Omega r(\theta, \delta_0) d\xi_0(\theta)$. If δ_0 is a Bayes solution in the wide sense and $r(\theta, \delta_0)$ is constant as a function of θ, the decision problem is strictly determined, and δ_0 is minimax. If δ_0 is admissible and $r(\theta, \delta_0) \equiv c < \infty$, δ_0 is minimax. If Ω is a finite set and $\inf_\delta \sup_\theta r(\theta, \delta) < \infty$, the decision problem is strictly determined and there exists a least favorable a priori distribution. We have few general results about generalized Bayes solutions (→ 173 Game Theory; [6]).

References

[1] A. Wald, Statistical decision functions, Wiley, 1950 (Chelsea, 1971).
[2] L. LeCam, An extension of Wald's theory of statistical decision functions, Ann. Math. Statist., 26 (1955), 69–81.
[3] R. R. Bahadur, Sufficiency and statistical decision functions, Ann. Math. Statist., 25 (1954), 423–462.

[4] H. Kudō, On minimax invariant estimates of the transformation parameter, Nat. Sci. Rep. Ochanomizu Univ., 6 (1955), 31–73.
[5] S. Kullback, Information theory and statistics, Wiley, 1959.
[6] D. H. Blackwell and M. A. Girshick, Theory of games and statistical decisions, Wiley, 1954.
[7] L. Weiss, Statistical decision theory, McGraw-Hill, 1961.
[8] H. Raiffa and R. Schlaifer, Applied statistical decision theory, Harvard Univ. Graduate School of Business Administration, 1961.
[9] W. James and C. Stein, Estimation with quadratic loss, Proc. 4th Berkeley Symp. Math. Stat. Prob. I, Univ. of California Press (1961), 361–379.
[10] J. O. Berger, Statistical decision theory, foundations, concepts, and methods, Springer, 1980.
[11] R. H. Farrell, Weak limits of sequences of Bayes procedures in estimation theory, Appendix, Decision theory, Proc. 5th Berkeley Symp. Math. Stat. Prob. I, Univ. of California Press, 1967, 83–111.

399 (XVIII.7) Statistical Estimation

A. General Remarks

Statistical estimation is one of the most important methods of statistical inference (→ 401 Statistical Inference). Its purpose is to estimate the values of †parameters (or their functions) involved in a distribution of a statistical †population by using observations on the population (→ 396 Statistic). Let $\mathscr{P}=\{P_\theta \mid \theta\in\Theta\}$ be a family of †probability distributions, indexed by a parameter θ and defined over a †measurable space (i.e., sample space) $(\mathscr{X},\mathfrak{B})$. Let X be a †random variable taking values in $\mathscr{X}$ and distributed according to a probability distribution P that is a member of $\mathscr{P}$. **Statistical estimation** is a method of estimating the †true value of the parameter θ (i.e., the parameter θ such that $P=P_\theta$) or the (true) value $g(\theta)$ of a given **parametric function** g (i.e., a function defined over Θ) or both, at θ, based on the observed value x of the random variable X. The function g maps the parameter values into $\mathbf{R}$, $\mathbf{R}^k$, or some function space. Statistical estimation methods are classified into two types: point estimation, which deals with individual values of $g(\theta)$, and interval (or region) estimation, by means of which regions that may contain the value $g(\theta)$ are considered. We can also include in statistical estimation the problem of predicting the tolerance region in which the value of a yet unobserved random variable may come out.

B. Point Estimation

In the method of **point estimation** for a given parametric function g, we choose a measurable mapping φ from the sample space $(\mathscr{X},\mathfrak{B})$ into a measurable space $(\mathscr{A},\mathfrak{C})$ and state that "the value of $g(\theta)$ is $\varphi(x)$" for an observed value x, where $\mathscr{A}$ is a set containing the range of g and $\mathfrak{C}$ is a †complete additive class of subsets in $\mathscr{A}$. The mapping φ, or the random variable $\varphi(X)$ taking values in the space $\mathscr{A}$, is called an **estimator** of $g(\theta)$, while the value $\varphi(x)$ determined by the observed value x is called an **estimate** of $g(\theta)$. This estimate is sometimes termed a **nonrandomized estimate** in contrast to the following generalized notion of estimator. A mapping from $\mathscr{X}$ to a set of probability distributions defined over $(\mathscr{A},\mathfrak{C})$ is called a **randomized estimator**, which reduces to a nonrandomized estimator when each image distribution degenerates to a single point. We assume that $\mathscr{A}=\mathbf{R}$ and $\mathfrak{C}=$ the class $\mathfrak{B}$ of all †Borel sets in $\mathbf{R}$, unless stated otherwise. We denote the †expectation and †variance with respect to P_θ by E_θ and V_θ, respectively.

C. Unbiasedness

An estimator $\varphi(X)$ of $g(\theta)$ may not be exactly equal to $g(\theta)$ for any $\theta\in\Theta$ except for trivial cases, but could instead be stochastically distributed around it. An estimator $\varphi(X)$ is said to have **unbiasedness** if it is stochastically balanced around $g(\theta)$ in some sense, such as mean, median, or mode. A statistic $\varphi(X)$ is called a (mean) **unbiased estimator** of $g(\theta)$ if

$$E_\theta(\varphi(X))=g(\theta)$$

for any $\theta\in\Theta$. A parametric function g is said to be **estimable** if it has an unbiased estimator. For example, the sample mean is unbiased for the population mean: $E_\theta(\bar{X})=E_\theta(X)$ for any $\theta\in\Theta$. Unbiasedness usually implies mean unbiasedness, and we assume this unless stated otherwise. The function

$$b(\theta)=E_\theta(\varphi(X))-g(\theta)$$

is called the **bias** of the estimator $\varphi(X)$. If we restrict ourselves to unbiased estimators $\varphi(X)$ only, it is best to choose, if possible, a $\varphi(X)$ whose variance $V_\theta(\varphi(X))$ is minimum uniformly for every $\theta\in\Theta$.

Theorem (Rao-Blackwell). If $T=t(X)$ is a †sufficient statistic, then for any unbiased

estimator $\varphi(X)$ of g the †conditional expectation $\psi(t)=E(\varphi(X)\mid T=t)$ yields another unbiased estimator $\varphi^*(X)=\psi(t(X))$ of g, which satisfies $V_\theta(\varphi^*)\leqslant V_\theta(\varphi)$ for all $\theta\in\Theta$, with the equality holding if and only if $\varphi(x)=\varphi^*(x)$ (a.e.$\mathscr{P}$). The notation a.e.$\mathscr{P}$ means that the statement concerned holds with probability 1 with respect to P_θ for each $\theta\in\Theta$. An estimator φ of $g(\theta)$ is called a **uniformly minimum variance** (or **UMV**) **unbiased estimator** if φ is unbiased for $g(\theta)$ and has a minimum variance uniformly in Θ among the class of unbiased estimators for $g(\theta)$.

Theorem (Lehmann-Scheffé). If T is a sufficient and †complete statistic, then for any estimable parametric function $g(\theta)$, there exists a unique UMV unbiased estimator of $g(\theta)$ that is a function of T. For example, suppose that $X=(X_1,X_2,\ldots,X_n)$ is a random sample from a population with exponential type distribution P_θ with density $p_\theta(x)=\beta(\theta)u(x)\exp(\sum_{i=1}^k\alpha_i(\theta)t_i(x))$ with respect to Lebesgue measure and that the set $\{(\alpha_1(\theta),\ldots,\alpha_k(\theta))\mid\theta\in\Theta\}$ contains some open set of $\mathbf{R}^k$. In this case, $T=(t_1(X),\ldots,t_k(X))$ is a sufficient and complete statistic, and hence every real-valued measurable function $\psi(T)$ is the unique UMV unbiased estimator of the parametric function $E_\theta(\psi(T))$. If for any $\theta\in\Theta$ the †median of the distribution of an estimator $\varphi(X)$ equals a real parametric function $g(\theta)$ when X is distributed as P_θ, i.e., if

$$P_\theta\{\varphi(X)<g(\theta)\}\leqslant\tfrac{1}{2}\leqslant P_\theta\{\varphi(X)\leqslant g(\theta)\},$$

then $\varphi(X)$ is called a **median unbiased estimator**. For example, a sample median (suitably defined for the case of an even number of samples) is median unbiased for the population median. If $\varphi(X)$ is a median unbiased estimator of g, then for any real-valued monotone function h, an estimator $h(\varphi(X))$ for $h(g(\theta))$ is median unbiased, that is, median unbiasedness is preserved under monotone transformations, which is not the case with mean unbiasedness.

Restricting our consideration to the class of all median unbiased estimators, we can use the function

$$a(u,\theta,\varphi)=\begin{cases}P_\theta\{\varphi(X)\leqslant u\} & \text{for } u<\theta\\ P_\theta\{\varphi(X)\geqslant u\} & \text{for } u>\theta\end{cases}$$

as an indicator of the behavior pattern of an estimator φ. The estimator φ that minimizes $a(u,\theta,\varphi)$ for all values of u and θ $(u\neq\theta)$ is said to be **uniformly best**. This property is also preserved under monotone transformations.

Theorem (Birnbaum). If a family of distributions $\{P_\theta\mid\theta\in\Theta\subset\mathbf{R}\}$ has a monotone †likelihood ratio with respect to a statistic $t(x)$ and the distribution function $F(t,\theta)$ of $T=t(X)$ is continuous both in t for any θ and in θ for any t, then there exists a uniformly best median unbiased estimator of θ. Actually, if $\theta=\hat\theta(t)$ is a solution of $F(t,\theta)=1/2$, then $\varphi(X)=\hat\theta(t(X))$ is such an estimator.

If for any $\theta\in\Theta$ the †mode of the density function or the probability mass function of an estimator $\varphi(X)$ is equal to $g(\theta)$, then $\varphi(X)$ is called a **modal unbiased estimator** of $g(\theta)$.

D. Lower Bounded of the Variance of an Unbiased Estimator

When there does not exist a sufficient and complete statistic, we can still seek to minimize the variance of the (mean) unbiased estimator at every fixed point $\theta=\theta_0$. In the remainder of this section, $\mathscr{P}$ is assumed to be dominated by a measure μ, and $p_\theta(x)$ denotes the density function of P_θ with respect to μ and $\pi_\theta(x)=p_\theta(x)/p_{\theta_0}(x)$. The following theorem guarantees the existence of the locally best unbiased estimator.

Theorem (Barankin). Let $\mathscr{M}$ be the set of all unbiased estimators of a parametric function $g(\theta)$ with finite variance at $\theta=\theta_0$. Assume that $\mathscr{M}$ is not empty and $E_{\theta_0}((\pi_\theta(X))^2)<\infty$. Then there exists an estimator φ_0 in $\mathscr{M}$ that minimizes the variance at θ_0 within $\mathscr{M}$. Actually, $\{\varphi_0\}=\mathscr{M}\cap\mathscr{P}_0$, where $\mathscr{P}_0$ is the linear space generated by $\{\pi_\theta(x)\mid\theta\in\Theta\}$. The minimum variance is given as follows:

$$\begin{aligned}&V_{\theta_0}(\varphi_0)\\&=\inf\{V_{\theta_0}(\varphi(X));\varphi\in\mathscr{M}\}\\&=\sup\nolimits_{\mathrm{I}}\left\{\left(\sum_{i=1}^n a_ih(\theta_i)\right)^2\Big/E_{\theta_0}\left(\sum_{i=1}^n a_i\pi_{\theta_i}(X)\right)^2\right\}\\&=\sup\nolimits_{\mathrm{II}}\left\{\sum_{i=1}^n\sum_{j=1}^n h(\theta_i)h(\theta_j)\lambda^{ij}\right\},\end{aligned}$$

where $h(\theta)=g(\theta)-g(\theta_0)$ and λ^{ij} is the (i,j)-component of the inverse of the $n\times n$ matrix (λ_{ij}) with $\lambda_{ij}=E_{\theta_0}(\pi_{\theta_i}(X)\pi_{\theta_j}(X))$ and where the supremum $\sup_{\mathrm{I}}$ is taken over all positive integers n, $\theta_1,\ldots,\theta_n\in\Theta$ and $a_1,\ldots,a_n\in R$, and becomes $\sup_{\mathrm{II}}$ when the supremum is taken over n and the θ_j. This theorem leads to the following three theorems with respect to the lower bound of the variance of an unbiased estimator. The first is immediate and the last two are obtained by replacing some order differences of π_θ with the corresponding differentials.

Theorem. For any unbiased estimator $\varphi(X)$ of g and $\theta_0\in\Theta$, we have

$$V_{\theta_0}(\varphi(X))\geqslant\sup_{\theta\in\Theta}\{(g(\theta)-g(\theta_0))^2/E_{\theta_0}(\pi_\theta(X)-\pi_{\theta_0}(X))^2\}$$

(Chapman-Robbins-Kiefer inequality).

Theorem. Suppose $\Theta \subset \mathbf{R}$. For any unbiased estimator $\varphi(X)$ of g and under certain regularity conditions, we have

$$V_{\theta_0}(\varphi(X)) \geqslant \frac{(g'(\theta_0))^2}{E_{\theta_0}((\partial \log p_\theta(X)/\partial\theta|_{\theta=\theta_0})^2)}$$

(Cramér-Rao inequality), where the equality holds only for the exponential type distribution $p_\theta(x)=\beta(\theta)u(x)\exp(\alpha(\theta)\varphi(x))$. An example of such regularity conditions is (i)–(iii): (i) $E_{\theta_0}((\pi_\theta(X))^2)<\infty$ for all $\theta\in\Theta$; (ii) $p_\theta(x)$ has a partial derivative $p'_\theta(x)$ at $\theta=\theta_0$ (a.e. P_{θ_0}); and (iii)

$$\lim_{\Delta\theta\to 0} E_{\theta_0}\left[\left(\frac{p_{\theta_0+\Delta\theta}(X)-p_{\theta_0}(X)}{p_{\theta_0}(X)\Delta\theta}-\frac{p'_{\theta_0}(X)}{p_{\theta_0}(X)}\right)^2\right]=0.$$

Corollary. Let $X=(X_1,\ldots,X_n)$ be a random sample from a distribution with density $f(x,\theta)$ and let

$$I(\theta)=E_\theta((\partial\log f(X_1,\theta)/\partial\theta)^2).$$

Since $E_\theta((\partial\log p_\theta(X)/\partial\theta)^2)=nI(\theta)$, the Cramér-Rao inequality implies

$$V_{\theta_0}(\varphi(X))\geqslant(g'(\theta_0))^2/nI(\theta_0).$$

The number $I(\theta_0)$ is called the **Fisher information** of the distribution $f(x,\theta)$. When the equality holds for an unbiased estimator $\varphi(X)$, $\varphi(X)$ is called an **efficient estimator** of $g(\theta)$. In general, the **efficiency** of an unbiased estimator φ at $\theta=\theta_0$ is defined by

$$\mathrm{Eff}(\varphi)=(g'(\theta_0))^2/(nI(\theta_0)V_{\theta_0}(\varphi)).$$

Theorem. For any unbiased estimator $\varphi(X)$ of $g(\theta)$ and under certain regularity conditions, we have

$$V_{\theta_0}(\varphi(X))\geqslant\sum_{i=1}^{k}\sum_{j=1}^{k}g^{(i)}(\theta_0)g^{(j)}(\theta_0)K^{ij}$$

(Bhattacharyya inequality), where $g^{(i)}(\theta_0)=d^ig(\theta)/d\theta^i|_{\theta=\theta_0}$ and K^{ij} is the (i,j)-component of the inverse of the matrix (K_{ij}) with

$$K_{ij}=E_{\theta_0}\left(\frac{p^{(i)}_{\theta_0}(X)\,p^{(j)}_{\theta_0}(X)}{p_{\theta_0}(X)\,p_{\theta_0}(X)}\right),\qquad i,j=1,\ldots,k.$$

An example of such regularity conditions is (i)–(iii): (i) $E_{\theta_0}((\pi_\theta(X))^2)<\infty$; (ii) $p_\theta(x)$ is k-times differentiable with respect to θ at $\theta=\theta_0$; (iii) the ith partial derivative $p^{(i)}_{\theta_0}(x)$, $i\leqslant k$, satisfies

$$\lim_{\Delta\theta\to 0}E_{\theta_0}\left[\left(\frac{\Delta^i p_\theta(X)|_{\theta=\theta_0}}{p_{\theta_0}(X)(\Delta\theta)^i}-\frac{p^{(i)}_{\theta_0}(X)}{p_{\theta_0}(X)}\right)^2\right]=0,$$

where $\Delta P_\theta(X)|_{\theta=\theta_0}=P_{\theta_0+\Delta\theta}(X)-P_{\theta_0}(X)$ and $\Delta^iP_\theta(X)=\Delta(\Delta^{i-1}P_\theta(X))$ for $i\geqslant 2$. For $k=1$ the Bhattacharyya lower bound is the same as the Cramér-Rao lower bound. In general, the former gives a sharper lower bound than the latter.

If the parameter is multidimensional, $\theta=(\theta_1,\ldots,\theta_k)'$, then for any unbiased estimator $\varphi(X)$ of $g(\theta)$ and under similar conditions to those for the 1-dimensional case, we have

$$V_{\theta_0}(\varphi(X))\geqslant\sum_{i=1}^{k}\sum_{j=1}^{k}g'_i(\theta_0)g'_j(\theta_0)J^{ij},$$

where $g'_i(\theta_0)=\partial g(\theta)/\partial\theta_i|_{\theta=\theta_0}$ and J^{ij} is the (i,j)-component of the inverse of the matrix (J_{ij}) with

$$J_{ij}=E_{\theta_0}(\partial\log p_\theta(X)/\partial\theta_i|_{\theta=\theta_0}\partial\log p_\theta(X)/\partial\theta_j|_{\theta=\theta_0}).$$

If $\theta^*(X)=(\theta^*_1(X),\ldots,\theta^*_k(X))'$ is an unbiased estimator of $\theta=(\theta_1,\ldots,\theta_k)'$ (i.e., $E_\theta(\theta^*_i(X))=\theta_i$ for $i=1,2,\ldots,k$), then the †covariance matrix $V_{\theta_0}(\theta^*(X))$ of $\theta^*(X)$ at θ_0 satisfies $V_{\theta_0}(\theta^*(X))\geqslant J^{-1}$, that is, the difference $V_\theta(\theta^*(X))-J^{-1}$ is a nonnegative definite matrix. If $X=(X_1,\ldots,X_n)$ is a random sample from a distribution having density $f(x_1,\theta)$, then by setting

$$I_{ij}=E_{\theta_0}(\partial\log f(X_1,\theta)/\partial\theta_i|_{\theta=\theta_0}\times\partial\log f(X_1,\theta)/\partial\theta_j|_{\theta=\theta_0}),$$

we have $J_{ij}=nI_{ij}$. The matrix $I=(I_{ij})$ is called the **Fisher information matrix** of the distribution $f(x_1,\theta)$.

E. Decision-Theoretic Formulation

(→ 398 Statistical Decision Functions)

Let $W(\theta,a)$ ($\geqslant 0$) be the loss incurred from an estimate (or action) a of the parameter when the true value of the parameter is θ. The risk function of an estimator $\varphi(X)$ of the parametric function $g(\theta)$ is then defined as

$$r(\theta,\varphi)=E_\theta(W(\theta,\varphi(X))).$$

Statistical decision theory deals with the problem of minimizing, in an appropriate manner, the risk function by a suitable choice of φ. The notions of complete class, Bayes estimator, admissibility, minimax estimator, and invariant estimator, explained here and in Sections F–I, are the most important of the theory. The unbiased estimator explained in Section C may also be considered an important concept of the theory.

A class C of estimators is said to be essentially complete if for any estimator φ there exists an estimator φ_0 in C such that

$$r(\theta,\varphi_0)\leqslant r(\theta,\varphi)$$

for any $\theta\in\Theta$. The following two theorems hold, provided that the action space $\mathscr{A}$ is $\mathbf{R}$ and the loss function $W(\theta,a)$ is convex with respect to $a\in\mathscr{A}$ for any $\theta\in\Theta$.

Theorem (Hodges-Lehmann). If $W(\theta,a)\to\infty$ as $|a|\to\infty$, then the class of all nonrandomized estimators is essentially complete.

Theorem. If $T=t(X)$ is a sufficient statistic, then the class of all functions of T is essentially

complete. Actually, given any estimator $\varphi(X)$ of $g(\theta)$, the conditional expectation $\psi(t)=E(\varphi(X)|T=t)$ yields an estimator $\varphi_0(X)=\psi(t(X))$ satisfying $r(\theta,\varphi_0)\leqslant r(\theta,\varphi)$ for any θ, where the equality holds when and only when $\varphi\equiv\varphi_0$, provided that W is convex in the strict sense.

A loss function of the form

$$W(\theta,a)=\lambda(\theta)(a-g(\theta))^2, \qquad \lambda(\theta)>0,$$

is called a **quadratic loss functions**. If, in particular, $\lambda(\theta)\equiv 1$, then

$$r(\theta,\varphi)=E_\theta((\varphi(X)-g(\theta))^2)$$

is called the **mean square error** of the estimator $\varphi(X)$ of $g(\theta)$. This error coincides with the variance when the estimator is unbiased.

F. Bayes Estimators

Let ξ be an a priori distribution over the parameter space Θ associated with a certain †σ-algebra $\mathfrak{F}$, and assume that $r(\theta,\varphi)$ is $\mathfrak{F}$-measurable for every φ. Denote by E^ξ the average operator relative to ξ. The infimum of the average risk $r(\xi,\varphi)=E^\xi(r(\theta,\varphi))=E^\xi E_\theta(W(\theta,\varphi(X)))$ for φ running over its range is called the †Bayes risk relative to ξ, while an estimator $\varphi(X)$ of $g(\theta)$ at which the average risk $r(\xi,\varphi)$ attains the infimum is called a **Bayes estimator** relative to ξ. A Bayes estimator is obtained as follows: Assume that $\mathscr{P}$ is dominated by a measure μ with $\mathfrak{B}\times\mathfrak{F}$-measurable density $p_\theta(x)$, the loss function $W(\theta,a)$ is $\mathfrak{F}\times\mathfrak{C}$-measurable, and

$$\int_\Theta p_\theta(x)\,d\xi(\theta)<\infty.$$

For each observed value of x, the Bayes estimator $\varphi(x)$ takes the value a that minimizes

$$r(a|x)=E^\xi(W(\theta,a)|x)=\int_\Theta W(\theta,a)p(\theta|x)\,d\xi(\theta),$$

where $p(\theta|x)$ is the probability density, with given x, of θ. We call $r(a|x)$ the **posterior risk**.

Theorem (Girshick-Savage). Suppose that the loss function is quadratic. For any x the value of the posterior risk is either ∞ (for every value of a) or finite (for all or only one value of a). If the Bayes risk relative to ξ is finite, then a Bayes estimator $\varphi^*(X)$ relative to ξ is determined uniquely as follows: $\varphi^*(x)=a_0$ if $r(a_0|x)<\infty$ for only one value a_0, whereas $\varphi^*(x)=E^\xi(g(\theta)\lambda(\theta)|x)/E^\xi(\lambda(\theta)|x)$ if $r(a|x)<\infty$ for every a. If $E^\xi(\lambda(\theta))<\infty$, then a Bayes estimator is either biased or has average risk zero.

G. Admissibility of Estimators

An estimator $\varphi_0(X)$ of a parametric function $g(\theta)$ is said to be **admissible** if and only if for any estimator $\varphi(X)$ of $g(\theta)$ the inequality $r(\theta,\varphi)\leqslant r(\theta,\varphi_0)$ for all $\theta\in\Theta$ implies that

$$r(\theta,\varphi)=r(\theta,\varphi_0)$$

for all $\theta\in\Theta$. If an estimator is the unique Bayes estimator relative to some a priori distribution, then it is admissible. For example, let $X_1,\ldots,X_n$ be a random sample from $N(\theta,1)$, and let $W(\theta,a)=(a-\theta)^2$. Then $\varphi(X)=(c+n\sigma^2\bar{X})/(1+n\sigma^2)$ is the unique Bayes estimator relative to the prior distribution $N(c,\sigma^2)$ of θ, where $\bar{X}$ is the sample mean. Hence any estimator of the form $\varphi(X)=a\bar{X}+b$ is admissible when $0<a<1$ and $-\infty<b<\infty$.

In the rest of this section and the next section, we restrict ourselves to quadratic loss functions. If a statistic of the form $c\varphi(X)$ with real c is an admissible estimator of $g(\theta)$, then

$$\inf_\theta\left(g(\theta)\frac{E_\theta(\varphi)}{E_\theta(\varphi^2)}\right)\leqslant c\leqslant\sup_\theta\left(g(\theta)\frac{E_\theta(\varphi)}{E_\theta(\varphi^2)}\right).$$

Theorem (Karlin). Let X be a random variable having a 1-dimensional exponential type distribution $dP_\theta(x)=\beta(\theta)e^{\theta x}d\mu(x)$ with a parameter space $\Theta=I(\underline{\theta},\bar{\theta})=\{\theta\mid\int_{-\infty}^{\infty}e^{\theta x}d\mu(x)<\infty\}$ a closed or open interval, and let $g(\theta)=E_\theta(X)=-\beta'(\theta)/\beta(\theta)$ be a parametric function to be estimated. Then the estimator $\varphi_\lambda(X)=X/(\lambda+1)$ for real λ is admissible provided that $\int_c^b(\beta(\theta))^{-\lambda}d\theta\to\infty$ as $b\to\bar{\theta}$ and $\int_a^c(\beta(\theta))^{-\lambda}d\theta\to\infty$ as $a\to\underline{\theta}$ for any $c\in(\underline{\theta},\bar{\theta})$.

Corollary. When $\Theta=(-\infty,\infty)$ the estimator $\varphi_0(X)=X$ is admissible.

Corollary. Let $\Theta=(-\infty,\infty)$, and assume that both intervals $(-\infty,0]$ and $[0,\infty)$ have positive measure with respect to μ. Then every estimator of the form $\varphi(X)=aX$ with $0<a\leqslant 1$ is admissible. This theorem can be applied to a random sample $X_1,\ldots,X_n$ drawn from an exponential type distribution because the sufficient statistic $\bar{X}=\sum X_i/n$ has an exponential type distribution and $E_\theta(\bar{X})=E_\theta(X_1)$.

Theorem (Karlin). Let X be a random variable having a distribution $dP_\theta(x)=q(\theta)\cdot r(x)dx$ for $0\leqslant x\leqslant\theta$ and $=0$ otherwise, with the parameter space $\Theta=(0,\infty)$, where $\int_0^\theta r(x)dx<+\infty$ and $\int_0^\infty r(x)dx=\infty$. Among the estimators of the type $\varphi_c(X)=c(q(X))^{-\alpha}$ for $g(\theta)=(q(\theta))^{-\alpha}$ with $\alpha>0$, only one estimator with the value $c=(2\alpha+1)/(\alpha+1)$ is admissible. This theorem is applicable also when the size of the random sample is larger than 1.

Theorem (Stein). Let $X_1,\ldots,X_n$ be a random sample from a univariate distribution $dP_\theta(x)=f(x-\theta)dx$ with a location parameter θ. Define

$$A_k=A_k(x_1,\ldots,x_n)=\int_{-\infty}^{\infty}\theta^k\left(\prod_{i=1}^n f(x_i-\theta)\right)d\theta$$

for $k=0, 1, 2$. If

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left(\frac{A_2}{A_0}-\left(\frac{A_1}{A_0}\right)^2\right)^{3/2}\prod_{i=1}^{n}f(x_i)\times\prod_{i=1}^{n}dx_i<\infty,$$

then the **Pitman estimator** $\varphi_0(X_1,\dots,X_n)=A_1(X_1,\dots,X_n)/A_0(X_1,\dots,X_n)$ of the parameter θ is admissible.

Inadmissibility of the Usual Estimator for Three or More Location Parameters. Let $X=(X_1,\dots,X_k)'$ be a k-variate normal random variable with mean $\theta=(\theta_1,\dots,\theta_k)'$ and covariance matrix I, the identity. Then the Pitman estimator of θ is $\hat{\theta}=X$. However, Stein showed that X is inadmissible. It is strictly dominated by the estimator $\theta^*(X)=(1-(k-2)/|X|^2)X$, where $|\cdot|$ denotes the Euclidean norm $|X|^2=\sum_{i=1}^{k}X_i^2$. That is, if $k\geqslant 3$, $E|\theta^*(X)-\theta|^2<E|X-\theta|^2$ for any θ.

An estimator such as $\theta^*(X)$ is called **Stein's shrinkage estimator** (James and Stein, *Proc. 4th Berkeley Symp.*, 1 (1960)).

H. Minimax Estimation

An estimator $\varphi^*(X)$ is said to be **minimax** if and only if

$$\sup_\theta r(\theta,\varphi^*)=\inf_\varphi\sup_\theta r(\theta,\varphi).$$

If an estimator φ^* is admissible and the risk $r(\theta,\varphi^*)$ is constant with respect to θ, then φ^* is minimax.

Theorem (Hodges-Lehmann). A Bayes estimator φ^* relative to an a priori distribution ξ is minimax if ξ assigns the whole probability to a subset $\omega\subset\Theta$, $r(\theta,\varphi^*)$ is constant (say, c) for $\theta\in\omega$, and $r(\theta,\varphi^*)\leqslant c$ for $\theta\in\Theta$. Let X have a binomial distribution $B(n,\theta)$, $0<\theta<1$; θ is unknown. If the prior distribution of θ is a beta distribution $\beta(\sqrt{n}/2,\sqrt{n}/2)$, then the Bayes estimator is $T^*(X)=(X+\sqrt{n}/2)/(n+\sqrt{n})$, which has constant risk $E_\theta(T^*(X)-\theta)^2=(2(1+\sqrt{n}))^{-2}$ for all θ, $0<\theta<1$. Thus, according to this theorem, $T^*(X)$ is minimax. It is interesting to compare the mean square error of $T^*(X)$ to that of minimum variance unbiased estimator $\hat{\theta}=X/n$, $\theta(1-\theta)/n$.

Theorem (Wald). If there exists a sequence of prior distributions $\{\xi_n\}$ such that

$$\liminf_{n\to\infty}\left(r(\theta,\varphi^*)-\inf_\varphi\int_\Theta r(\theta,\varphi)\,d\xi_n(\theta)\right)\leqslant 0$$

for any $\theta\in\Theta$, then φ^* is minimax. For example, the last theorem led to the proof of the fact that the Pitman estimator of a location parameter is minimax [4]. In the discussion of robust estimation, Huber (*Ann. Math. Statist.*, 35 (1964)) proved that Huber's minimax robust location estimator minimizes the maximum asymptotic variance over some family of symmetric distributions in a neighborhood of the normal distribution (→ 371 Robust and Nonparametric Methods).

I. Invariant Estimator

For simplicity, assume that the range A of a parametric function $g(\theta)$ coincides with the parameter space, and consider the point estimation of $g(\theta)\equiv\theta$ (θ is not necessarily real). Suppose that there exist two groups $G=\{\tau\}$ and $\bar{G}=\{\bar{\tau}\}$ of one-to-one measurable transformations of $\mathscr{X}$ and Θ, respectively, onto themselves such that (i) there exists a homeomorphic mapping $\tau\to\bar{\tau}$ from G to $\bar{G}$; (ii) if X has a distribution P_θ, then τX has the distribution $P_{\bar{\tau}\theta}$; and (iii) $W(\bar{\tau}\theta,\bar{\tau}a)=W(\theta,a)$ for any θ, a, and τ. An estimator φ is said to be **invariant** if it satisfies $\varphi(\tau x)=\bar{\tau}\varphi(x)$ for any $\tau\in G$, a.e. $\mathscr{P}$. An estimator φ is called a **best invariant estimator** if the risk function $r(\theta,\varphi')$, where φ' is an invariant estimator, takes its minimum value when $\varphi'=\varphi$.

If the group $\bar{G}$ is †transitive on Θ, then the risk function of any invariant estimator is independent of the value of θ, and hence any admissible invariant estimator is best invariant. For example, in the point estimation of a location parameter with a quadratic loss function, the Pitman estimator is best invariant.

Theorem. If Θ is a compact topological space, $\bar{G}$ is a group of homeomorphisms of Θ onto itself, and $\bar{G}$ is homeomorphic to Θ under $g\in\bar{G}\to\bar{g}\theta_0\in\Theta$ with a fixed $\theta_0\in\Theta$, then any Bayes estimator relative to the †right-invariant Haar measure over $\bar{G}$ is best invariant. This result can be generalized to a locally compact Θ (→ 398 Statistical Decision Functions).

J. Sequential Estimation

Estimation methods based on sequential sampling are not as popular as †sequential tests, because their efficiency is not very large compared to that of nonsequential estimation. A generalization of the Cramér-Rao inequality to any sequential unbiased estimator $\varphi(X)$ of a parametric function $g(\theta)$ is the **Wolfowitz inequality,**

$$V_\theta(\varphi(X))\geqslant(g'(\theta))^2/(E_\theta(N)I(\theta))$$

for every $\theta\in\Theta$, under regularity conditions similar to those for the fixed-size sample problem, where N is the sample size and $I(\theta)$ the Fisher information.

K. Asymptotic Theory

In practical problems of statistical inference the sample size n is often large enough to give sharp estimates of the parameters involved; then the sample distributions of estimators can be approximated closely by their asymptotic distributions, which are of a simpler nature. Assume that $X=(X_1, X_2, \ldots)$ is a sequence of independent and identically distributed (i.i.d.) random variables with the common distribution P_θ, $\theta \in \Theta$. For each n, let $\varphi_n = \varphi_n(X_1, \ldots, X_n)$ be an estimator of $g(\theta)$ that is a function of θ to $\mathscr{A}$ $(\subset \mathbf{R}^p)$. Thus φ_n is a measurable mapping from $(\mathscr{X}^n, \mathfrak{B}^n)$ to $(\mathscr{A}, \mathscr{C})$. Let us denote the distribution of φ_n by $\mathscr{L}(\varphi_n)$, $\mathscr{L}(\varphi_n \mid \theta)$ or $\mathscr{L}(\varphi_n \mid P_\theta)$, the last two emphasizing that the underlying probability distribution is P_θ. For example, if the mean vector $E_\theta(\varphi_n) = m_n(\theta)$ and the covariance matrix $v_n(\theta) = V_\theta(\varphi_n)$ exist for every n, and if $\mathscr{L}[v_n(\theta)^{-1/2}(\varphi_n(X) - m_n(\theta)) \mid \theta] \rightarrow N_p(0, I)$ as $n \rightarrow \infty$, then $\mathscr{L}(\varphi_n \mid \theta)$ is approximated by a p-variate normal distribution $N_p(m_n(\theta), v_n(\theta))$ (→ 341 Probability Measures D). φ_n is said to be **asymptotically (mean) unbiased** for $g(\theta)$ if $m_n(\theta) \rightarrow g(\theta)$ for any $\theta \in \Theta$ as $n \rightarrow \infty$. But we often calculate the asymptotic distribution without obtaining the exact mean and covariance matrix of the estimator φ_n for each n. In the asymptotic theory it may be reasonable to regard the sequence of estimators $\{\varphi_n\}$ rather than each estimator φ_n as an "estimator," but we do not bother with the difference between these definitions of an estimator.

Consistency. $\{\varphi_n\}$ is called a **consistent estimator** of $g(\theta)$ if φ_n converges to $g(\theta)$ in probability as $n \rightarrow \infty$:

$$\lim P_\theta\{|\varphi_n - g(\theta)| > \varepsilon\} = 0 \quad \text{for any } \varepsilon > 0 \text{ and every } \theta \in \Theta.$$

If the convergence is almost sure, it is called **a.s. consistent**. For example, if φ_n is asymptotically unbiased with the covariance matrix $v_n(\theta)$ such that $|v_n(\theta)| \rightarrow 0$ as $n \rightarrow \infty$, then φ_n is a consistent estimator of $g(\theta)$. A sufficient condition for existence of a consistent estimator is given by the following result.

Theorem (LeCam). Let $\mathscr{X}$ be a Euclidean space and $\mathfrak{B}$ the σ-algebra of all Borel sets in $\mathscr{X}$. If the parameter space Θ is a locally compact subset of $\mathbf{R}^k$, $P_\theta \neq P_{\theta'}$ for any $\theta \neq \theta'$ (identifiability condition), and $P_{\theta_n} \rightarrow P_\theta$ whenever $\theta_n \rightarrow \theta$, then there exists a consistent estimator of θ. $\varphi_n = g(T_n)$ is a consistent estimator of $g(\theta)$ if $\{T_n\}$ is a consistent estimator of θ and $g(\theta)$ is a continuous function of θ.

Asymptotic Normality. The class of estimators is restricted to what are called consistent estimators. An estimator $\{\varphi_n\}$ is said to be **asymptotically normally distributed** if the asymptotic distribution of $n^{1/2}(\varphi_n - g(\theta))$ is normal:

$$\mathscr{L}[n^{1/2}(\varphi_n - g(\theta)) \mid \theta] \rightarrow N_p(\mu(\theta), v(\theta)) \quad \text{as} \quad n \rightarrow \infty.$$

$\mu(\theta)$ and $v(\theta)$ are called the **asymptotic bias** and **asymptotic covariance matrix**, respectively. They are not always equal to the limits, if any, of the mean and covariance matrix of $n^{1/2}(\varphi_n - g(\theta))$. $\{\varphi_n\}$ is usually called a **consistent and asymptotically normal (CAN) estimator** if the asymptotic distribution of $n^{1/2}(\varphi_n - g(\theta))$ is normal with the asymptotic bias zero. Then the distribution $\mathscr{L}(\varphi_n \mid \theta)$ is approximated by $N_p(0, v(\theta)/n)$. For example, the †moment method estimator is a CAN estimator.

Theorem. Let $\{\varphi_n\}$ be a CAN estimator of $g(\theta) \in \mathbf{R}^1$ with asymptotic variance $v(\theta)$. Then

$$\liminf_{n \rightarrow \infty} E_\theta\{n |\varphi_n - g(\theta)|^2\} \geqslant v(\theta) \quad \text{for every} \quad \theta \in \Theta.$$

Theorem. Suppose that $g(\theta)$ is a continuously differentiable function from Θ $(\subset \mathbf{R}^k)$ to $\mathbf{R}^p$ $(p \leqslant k)$. Let $G(\theta) = (\partial g_i(\theta)/\partial \theta_j)$, the †Jacobian matrix. If $\{T_n\}$ is a CAN estimator of θ with asymptotic covariance matrix $v(\theta)$, then $\{g(T_n)\}$ is a CAN estimator of $g(\theta)$:

$$\mathscr{L}[n^{1/2}(g(T_n) - g(\theta)) \mid \theta] \rightarrow N_P(0, G(\theta)v(\theta)G(\theta)').$$

An estimator $\{T_n\}$ is said to be a **best asymptotically normal (BAN) estimator of** θ if $\{T_n\}$ is a CAN estimator of θ with asymptotic variance $I^{-1}(\theta)$, where $I(\theta)$ is the †Fisher information matrix on θ in a single observation. We can see that the maximum likelihood (ML) estimator (→ Section M) is a BAN estimator.

Functional on Distribution Functions. Let $\varphi(F)$ be a functional on distribution functions to $\mathbf{R}^1$. Let us consider the class of estimators that are defined by $\varphi_n = \varphi(\hat{F}_n)$, where $\hat{F}_n$ is the †empirical distribution function of n samples $X_1, \ldots, X_n$. An estimator $\{\varphi_n\}$ with $\varphi_n = \varphi(\hat{F}_n)$ for each n is said to be **Fisher consistent** for $g(\theta)$ if $\varphi(F_\theta) = g(\theta)$ for every $\theta \in \Theta$ when F_θ is the true distribution function. $\{\varphi_n\}$ is also a.s. consistent for $g(\theta)$ if $\{\varphi_n\}$ is a Fisher consistent estimator of $g(\theta)$ and if φ is a continuous functional. Furthermore, if φ is differentiable, we can see that $\{\varphi_n\}$ is also a CAN estimator by using the fact that $n^{1/2}(\hat{F}_n(F_\theta^{-1}(t)) - t)$, $0 \leqslant t \leqslant 1$, converges weakly to the †Brownian bridge. Let S be a set of distribution functions. S is said to be a star-shaped set of F if $H \in S$ implies $F^{(t)} = (1-t)F + tH \in S$ for any $t \in [0,1]$.

Theorem (Von Mises). Assume that
(F1) there exists a star-shaped set S_θ at F such that $\lim_n P_\theta\{\hat{F}_n \in S_\theta\} = 1$;
(F2) for any $t \in [0,1]$ and $H \in S_\theta$, there exist derivatives $(d^i/dt^i)\varphi[(1-t)F + tH]$, $i = 1, 2$;
(F3) there exists $\psi_\theta(y)$ from $\mathbf{R}^1$ to $\mathbf{R}^1$ such that

$\frac{d}{dt}\varphi[(1-t)F_\theta+tH]|_{t=0}=\int_{-\infty}^{\infty}\psi_\theta(y)d(H(y)-F_\theta(y))$ for all $H\in S_\theta$; and

(F4) $\lim_{n\to\infty} P_\theta\left\{n^{1-\delta}\sup_{t\in[0,1]}\left|\frac{d^2}{dt^2}\varphi[F_{\theta n}^{(t)}]\right|>\varepsilon\right\}=0$ for any δ and $\varepsilon>0$, where $F_{\theta n}^{(t)}=(1-t)F_\theta+t\hat{F}_n$. Then if φ_n with $\varphi_n=\varphi(\hat{F}_n)$ is Fisher consistent for $g(\theta)$, we have

$$\mathscr{L}[n^{1/2}(\varphi_n-g(\theta))\,|\,\theta]\to N(0,v(\theta)),$$

where

$$v(\theta)=\int_{-\infty}^{\infty}\psi_\theta(y)^2\,dF_\theta(y)-\left\{\int_{-\infty}^{\infty}\psi_\theta(y)\,dF_\theta(y)\right\}^2.$$

$\{c_n\}$-Consistency. For a sequence of positive numbers c_n tending to infinity as $n\to\infty$, an estimator T_n is called **consistent for $\theta\in\Theta$ with order c_n** (or **$\{c_n\}$-consistent** for short) if for every $\varepsilon>0$ and every $\theta\in\Theta$ there exist a sufficiently small positive number δ and a sufficiently large number K satisfying

$$\limsup_{n\to\infty}\sup[P_\tau\{c_n|T_n-\tau|\geqslant K\};|\tau-\theta|<\delta]<\varepsilon.$$

Let $\{c_n\}$ be a maximal order of consistency. This notion was introduced by Takeuchi and Akahira. They studied consistent estimators of location parameters with various orders. Let $\mathscr{X}=\Theta=\mathbf{R}^1$. Suppose that for every $\theta\in\Theta$, P_θ has a density function $f(x-\theta)$ with respect to the Lebesgue measure.

Theorem. Assume that
(OC1) $f(x)>0$ if $a<x<b$ and $f(x)=0$ if $x\leqslant a$ or $x\leqslant b$;
(OC2) there exist positive numbers $0<\alpha\leqslant\beta<\infty$ and $0<A'$, $B'<\infty$ such that

$$\lim_{x\to a+0}(x-a)^{1-\alpha}f(x)=A',$$

$$\lim_{x\to b-0}(b-x)^{1-\beta}f(x)=B';$$

(OC3) $f(x)$ is twice continuously differentiable in the interval (a,b) and there exist positive numbers $0<A''$, $B''<\infty$ such that

$$\lim_{x\to a+0}(x-a)^{2-\alpha}|f'(x)|=A'',$$

$$\lim_{x\to b-0}(b-x)^{2-\beta}|f'(x)|=B''.$$

Assume further that $f''(x)$ is bounded if $\alpha\geqslant 2$. Then for each α there exists a consistent estimator with the order given in Table 1.

Table 1

α	order c_n	$\{c_n\}$-consistent estimator
$0<\alpha<2$	$n^{1/\alpha}$	$\{\min X_i+\max X_i-(a+b)\}/2$
$\alpha=2$	$(n\log n)^{1/2}$	ML estimator
$\alpha>2$	$n^{1/2}$	ML estimator

L. Moment Method

The moment method is also utilized to obtain estimators. Suppose that $\mathscr{X}\subset\mathbf{R}^1$ and $\Theta\subset\mathbf{R}^k$. Denote the †population distribution function by F_θ and the †empirical distribution function of n samples $X_1,\ldots,X_n$ by $\hat{F}_n$. The following system of simultaneous equations is derived from letting the jth †population moment

$$\mu_j(\theta)=E_\theta(X^j)=\int x^j F_\theta(dx)$$

be equal to the jth †sample moment

$$m_{nj}=n^{-1}\sum_{i=1}^n X_i^j=\int x^j\hat{F}_n(dx).$$

For example, for $j=1,\ldots,k$,

$$\mu(\theta)=(\mu_1(\theta),\ldots,\mu_k(\theta))'=(m_{n1},\ldots,m_{nk})'=m_n.$$

A **moment method estimator** is determined as a solution $\theta=\tilde{\theta}_n(x)\in\Theta$ of k numbers of simultaneous equations.

Theorem. Assume that the function $\mu(\theta)$ from Θ to $\mathbf{R}^k$ is continuously differentiable and that the Jacobian matrix $M(\theta)=\partial\mu(\theta)/\partial\theta=(\partial\mu_i(\theta)/\partial\theta_j)$, $i,j=1,\ldots,k$, is nonsingular in a neighborhood of the true parameter. Then the moment method estimator exists and is a CAN estimator:

$$\mathscr{L}[n^{1/2}(\tilde{\theta}_n-\theta)\,|\,\theta]\to N_k(0,M(\theta)^{-1}v(\theta)M(\theta)'^{-1}),$$

where $v(\theta)=(\mathrm{cov}_\theta(X^i,X^j))$, $i,j=1,\ldots,k$. In general, a moment method estimator is not a BAN estimator. However, in view of its simple form, a moment method estimator is important and often utilized as a first-step estimator in order to determine the maximum likelihood estimator by the iteration method.

M. Maximum Likelihood Method

Suppose that a distribution P_θ has the density function $f(x,\theta)$, $\theta\in\Theta\subset\mathbf{R}^k$, with respect to a †$\sigma$-finite measure μ, and let $x_1,\ldots,x_n$ be observed values of random samples $X_1,\ldots,X_n$ from the population $f(x,\theta)$. Then the function L_n of θ defined by

$$L_n(\theta;x_1,\ldots,x_n)=\prod_{i=1}^n f(x_i,\theta)$$

is called the **likelihood function.** If $\theta=\hat{\theta}_n(x_1,\ldots,x_n)$ maximizes the value of $L_n(\theta)$ for fixed $x_1,\ldots,x_n$ and if it is a measurable mapping from $(\mathscr{X}^n,\mathfrak{B}^n)$ to $(\Theta,\mathscr{C})$ with $\mathscr{C}$ a †σ-algebra of subsets of Θ, then $\hat{\theta}_n(X)=\hat{\theta}_n(X_1,\ldots,X_n)$ is called the **maximum likelihood (ML) estimator of θ.** This method of finding estimators is called the **maximum likelihood method.** If the parameter is transformed into a new parameter $\eta=h(\theta)$ by means of a known one-to-

one bimeasurable transformation h and if there exists a unique ML estimator $\hat{\theta}_n$ of θ, then $\hat{\eta}_n(X)=h(\hat{\theta}_n(X))$ is a unique ML estimator of η. In other words, the ML estimator is invariant for every one-to-one transformation.

Many statisticians have investigated and improved known adequate regularity conditions under which the ML estimator exists and is a BAN estimator.

Theorem (Wald). Assume that
(C1) Θ is a closed subset of $\mathbf{R}^k$ with nonempty interior Θ^0;
(C2) for any $x\in\mathcal{X}$, $f(x,\theta)$ is continuous with respect to θ and $\lim_{|\theta|\to\infty}f(x,\theta)=0$ if Θ is not bounded;
(C3) if $\theta_1\neq\theta_2$, then $P_{\theta_1}\neq P_{\theta_2}$ and $\int|f(x,\theta_1)-f(x,\theta_2)d\mu(x)>0$;
(C4) $E_{\theta_0}(|\log f(X,\theta_0)|)<\infty$; and
(C5) $E_{\theta_0}(\log^+ f(X,\theta,\rho))<\infty$ and $E_{\theta_0}(\log^+\varphi(X,r))<\infty$, where $f(x,\theta,\rho)=\sup\{f(x,\theta');|\theta'-\theta|\leqslant\rho\}$ and $\varphi(x,r)=\sup\{f(x,\theta);|\theta|>r\}$. (The last two functions are measurable according to assumption (C2).)
Then if a sequence of measurable functions, $\{\overline{\theta}_n(x_1,\dots,x_n)\}$, satisfies

$$\liminf_{n\to\infty}\left\{\frac{L_n(\overline{\theta}_n;x_1,\dots,x_n)}{L_n(\theta_0;x_1,\dots,x_n)}\right\}\geqslant C>0\qquad(\text{a.s. }P_{\theta_0}),$$

then as $n\to\infty$, $\overline{\theta}_n(X_1,\dots,X_n)$ converges a.s. to the true value θ_0 of the parameter. Hence if the ML estimator exists, it is a.s. consistent. Pfanzagl (*Metrika*, 14 (1969)) and Fu and Gleser (*Ann. Inst. Statist. Math.*, 27 (1975)) gave rigorous proofs for the existence of the ML estimator.

Theorem. Under assumptions (C1) and (C2), there exists a maximum likelihood estimator $\hat{\theta}_n$ for any positive integer n. That is, $\hat{\theta}_n=\hat{\theta}_n(x_1,\dots,x_n)$ is a measurable function from $(\mathcal{X}^n,\mathfrak{B}^n)$ to $(\Theta,\mathscr{C})$ and satisfies $L(\hat{\theta}_n;x_1,\dots,x_n)=\sup_\theta L(\theta;x_1,\dots,x_n)$.

In the remainder of this section we suppose that assumptions (C1)–(C5) are satisfied. We use the notation

$$\frac{\partial f}{\partial\theta}=\left(\frac{\partial}{\partial\theta_i}f(x,\theta)\right),\quad\text{a }k\text{-column vector,}$$

$$\frac{\partial^2 f}{\partial\theta^2}=\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}f(x,\theta)\right),\quad\text{a }k\times k\text{ matrix,}$$

$$\frac{\partial^3\log f}{\partial\theta^3}=\left(\frac{\partial^3\log f(x,\theta)}{\partial\theta_h\partial\theta_i\partial\theta_j}\right),\qquad h,i,j=1,\dots,k.$$

Theorem (Cramér). Assume that
(AN1) for a.s. $[\mu]x$, $f(x,\theta)$ is three-times continuously differentiable with respect to each component of $\theta=(\theta_1,\dots,\theta_k)'\in\Theta^0$;
(AN2) for $\theta\in\Theta^0$, $\displaystyle\int_{\mathcal{X}}\frac{\partial f(x,\theta)}{\partial\theta}d\mu=0$ and $\displaystyle\int_{\mathcal{X}}\frac{\partial^2 f(x,\theta)}{\partial\theta^2}d\mu=0$;
(AN3) $E_\theta\left(\left|\frac{\partial}{\partial\theta}\log f(x,\theta)\right|^3\right)<\infty$ for $\theta\in\Theta^0$;
(AN4) the Fisher information matrix $I(\theta)=E_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(x,\theta)\right)\left(\frac{\partial}{\partial\theta}\log f(x,\theta)\right)'\right]$ exists and is positive definite for $\theta\in\Theta^0$; and
(AN5) there exists an $H(x)$ such that $\left|\frac{\partial^3}{\partial\theta^3}\log f(x,\theta)\right|<H(x)$ and $E_\theta(H(X))<C$, a constant for $\theta\in\Theta^0$.
Then the maximum likelihood estimator is a BAN estimator: $\mathscr{L}[n^{1/2}(\hat{\theta}_n-\theta)|\theta]\to N(0,I(\theta)^{-1})$ as $n\to\infty$ for $\theta\in\Theta^0$. Note that under assumption (AN1) the likelihood function attains the maximum in Θ^0 with probability tending to 1 as $n\to\infty$ if the true value of the parameter exists in Θ^0. Hence the ML estimator is determined as a root $\theta=\hat{\theta}_n(x_1,\dots,x_n)$ of the **likelihood equation** with the same probability as above:

$$\frac{\partial}{\partial\theta}\log L_n(\theta)=\left(\sum_{i=1}^n\frac{\partial}{\partial\theta_j}\log f(x_i,\theta)\right)_{j=1,\dots,k}=0.$$

We also call $n^{-1}\partial\log L_n(\theta)/\partial\theta$ the **likelihood estimating function**. The essential fact used in the proof of the above theorem is the asymptotic equivalence of the ML estimator and the likelihood estimating function:

$$\Delta_n(\theta)-I(\theta)n^{1/2}(\hat{\theta}_n-\theta)\to 0\ \text{ in }\ P_\theta\ \text{ as }\ n\to\infty,$$

where $\Delta_n(\theta)=n^{-1/2}\partial\log L_n(\theta)/\partial\theta$. Note the fact that

$$\mathscr{L}[\Delta_n(\theta)|\theta]\to N_k(0,I(\theta))\ \text{ as }\ n\to\infty$$

holds according to the central limit theorem (→ 250 Limit Theorems in Probability Theory B (1)).

Contiguity. We now turn to the situation where we need asymptotic distributions under the alternative distribution $P_{\theta+n^{-1/2}h}$ with $\theta+n^{-1/2}h\in\Theta$ in estimation as we do in testing hypotheses (→ 400 Statistical Hypothesis Testing). The notion of contiguity, due to LeCam (1960), is basic for the asymptotic methods of estimation theory. We consider sequences $\{P_n\}$ and $\{P_n'\}$ of probability measures on $(\mathcal{X}^n,\mathfrak{B}^n)$ with the †Radon-Nikodym derivatives p_n and p_n' with respect to a σ-finite measure, such as P_n+P_n'. Denote by $\chi_n=\Lambda[P_n';P_n]$ a generalized log-likelihood ratio that is defined by $\log p_n'/p_n$ on the set $\{p_np_n'>0\}$ and is an arbitrary measurable function on the set $\{p_np_n'=0\}$. Let $\{B_n\}$ be any sequence of $\{\mathfrak{B}^n\}$-measurable sets, and let $\{T_n\}$ be any sequence of $\{\mathfrak{B}^n\}$-measurable functions. A sequence of distributions $\{\mathscr{L}[T_n|P_n]\}$ is said to be **relatively compact** if every subsequence $\{n'\}\subset\{n\}$ contains a further subsequence $\{m\}\subset\{n'\}$ along which it converges to a prob-

ability distribution. In the Euclidean space relative compactness is equivalent to **tightness**: that is, for every $\varepsilon>0$ there is a $b(\varepsilon)$ such that $P_n\{|T_n|>b(\varepsilon)\}<\varepsilon$ for every n.

Theorem. The following statements are all equivalent.

(1) For any $\{T_n\}$, $T_n\to 0$ in P_n if and only if $T_n\to 0$ in P_n'.

(2) For any $\{T_n\}$, $\{\mathscr{L}[T_n|P_n]\}$ is relatively compact if and only if $\{\mathscr{L}[T_n|P_n']\}$ is relatively compact.

(3) For any $\{B_n\}$, $P_n\{B_n\}\to 0$ if and only if $P_n'\{B_n\}\to 0$.

(4) Whatever the choice of χ_n, $\{\mathscr{L}[\chi_n|P_n]\}$ and $\{\mathscr{L}[\chi_n|P_n']\}$ are relatively compact.

(5) Whatever the choice of χ_n, $\{\mathscr{L}[\chi_n|P_n]\}$ is relatively compact. Furthermore, if $\{m\}\subset\{n\}$ is a subsequence of $\{n\}$ such that $\mathscr{L}[\chi_m|P_m]$ converges to $\mathscr{L}[\chi]$, then $E\{e^{\chi}\}=1$.

Two sequences $\{P_n\}$ and $\{P_n'\}$ satisfying requirements (1)–(5) of the above theorem are said to be **contiguous**.

Theorem. Suppose that $\{P_n\}$ and $\{P_n'\}$ are contiguous. Let $\{m\}\subset\{n\}$ be a subsequence such that $\mathscr{L}[\chi_m, T_m|P_m]$ converges to a limit $\mathscr{L}[\chi, T]$. Then $\mathscr{L}[\chi_m, T_m|P_m']$ converges to $e^{\chi}\mathscr{L}[\chi, T]$, where $v=e^{\chi}\mathscr{L}[\chi, T]$ is given by $v(A)=\int_A e^{\chi}d\mathscr{L}[\chi, T]$.

Now, set $P_n=P_\theta^n$ and $P_n'=P_{\theta+n^{-1/2}h}^n$ for each n. Under suitable regularity conditions, such as assumptions (C1)–(C5) and (AN1)–(AN5), it is easy to see that $\{P_\theta^n\}$ and $\{P_{\theta+n^{-1/2}h}^n\}$ are contiguous. At the same time, we can see that the asymptotic linearity of $\Lambda_n(\theta+n^{-1/2}h;\theta)=\Lambda[P_{\theta+n^{-1/2}h}^n; P_\theta^n]$ holds (say) in the vicinity of the true parameter as follows:

$$\Lambda_n(\theta+n^{-1/2}h;\theta)-h'\Delta_n(\theta)+\frac{1}{2}h'I(\theta)h\to 0 \text{ in } P_\theta.$$

The asymptotic equivalence of the ML estimator and $\Delta_n(\theta)$, and the asymptotic linearity of the log-likelihood function and $\Delta_n(\theta)$, leads to the regularity (→ Section N) of the ML estimator.

Theorem. Under suitable regularity conditions as above, the ML estimator is regular:

$$\mathscr{L}[n^{1/2}(\hat{\theta}_n-\theta)-h|\theta+n^{-1/2}h]\to N_k(0, I(\theta)^{-1}),$$

for any $h\in\mathbf{R}^k$ with $\theta+n^{-1/2}h\in\Theta$.

N. Asymptotic Efficiency

In Section D we discussed lower bounds of variances of unbiased estimators for finite sample size and defined the efficiency of an unbiased estimator with variance $v_n(\theta)$ by $(v_n(\theta)nI(\theta))^{-1}$. In this section we first discuss the asymptotic efficiency of a CAN estimator in the same vein as in the case of finite sample size. Second, we see a specific approach to the large-sample theory of estimation. Throughout this section we assume (C1)–(C5) and (AN1)–(AN5) stated in Section M.

BAN Estimators. We suppose that the parameter space Θ is a subset of $\mathbf{R}^1$ in this paragraph. We restrict our attention to the class of CAN estimators $\{T_n\}$ of the real-valued parameter θ for which $\mathscr{L}[n^{1/2}(T_n-\theta)|\theta]\to N(0, v(\theta))$ as $n\to\infty$. Fisher's conjecture concerning the lower bound to asymptotic variance $v(\theta)$ of any CAN estimator is

$$v(\theta)\geqslant I(\theta)^{-1},$$

where $I(\theta)$ is the Fisher information on θ in a single observation. The **asymptotic efficiency** of a CAN estimator with asymptotic variance $v(\theta)$ is defined by $(v(\theta)I(\theta))^{-1}$. A CAN estimator with asymptotic variance $I(\theta)^{-1}$ is called a **BAN estimator** or an **asymptotically efficient estimator**. Note that under suitable regularity conditions there always exists an asymptotic efficient estimator, for example a ML estimator, although for a sample of finite size there exists an efficient estimator if and only if the family of density functions is of the exponential type.

Unfortunately, Fisher's conjecture is not true without further conditions on the competing estimators. A counterexample was provided by Hodges and reported by LeCam (1953). Let $\{T_n\}$ be any CAN estimator with the asymptotic variance $v(\theta)$. Consider the estimator

$$T_n'=\begin{cases}\alpha T_n & \text{if } |T_n|<n^{-1/4},\\ T_n & \text{if } |T_n|\geqslant n^{-1/4},\end{cases}$$

where $0<\alpha<1$ is a constant. $\{T_n'\}$ is also a CAN estimator with asymptotic variance $v'(\theta)$ such that

$$v'(\theta)=\begin{cases}\alpha^2 v(0) & \text{if } \theta=0,\\ v(\theta) & \text{otherwise.}\end{cases}$$

Let $\{T_n\}$ be a BAN estimator; then T_n' is an estimator with asymptotic variance less than $I(\theta)^{-1}$ and is called a **superefficient estimator**.

Theorem (LeCam). The set of points θ for which the inequality due to Fisher fails is of Lebesgue measure zero. A condition due to Bahadur leads to the continuity of asymptotic variance which implies the validity of the above inequality.

Theorem (Bahadur). Suppose that $\{T_n\}$ is a CAN estimator with asymptotic variance $v(\theta)$ satisfying the condition

$$\liminf_{n\to\infty} P_{\theta_0+n^{-1/2}}\{T_n<\theta_0+n^{-1/2}\}\leqslant\frac{1}{2}$$

or

$$\liminf_{n\to\infty} P_{\theta_0+n^{-1/2}}\{T_n>\theta_0+n^{-1/2}\}\leqslant\frac{1}{2}.$$

Then the following inequality, due to Fisher, holds at $\theta=\theta_0$:

$$v(\theta_0)\geqslant I(\theta_0)^{-1}.$$

Regular Estimators. Wolfowitz and Kaufman considered an operationally more justifiable restriction on competing estimators, called the **uniformity property**, stating that, for an estimator $\{T_n\}$ of θ, $\mathscr{L}[n^{1/2}(T_n-\theta)\,|\,\theta](y)$ converges to any limit $L_\theta(y)$ uniformly in $(y,\theta)\in \mathbf{R}^k\times C$, where C is any compact subset of the interior Θ^0 of $\Theta\subset\mathbf{R}^k$. The ML estimator $\{\hat{\theta}_n\}$ also has this uniformity property under suitable regularity conditions, such as (C1)–(C5) and (AN1)–(AN5) to which some uniformity properties are added [29]. We note that asymptotic variance is not a good measurement of asymptotic efficiency unless an estimator is a CAN estimator, and that **asymptotic concentration** is in general a more pertinent measurement.

Theorem. For an estimator $\{T_n\}$ with uniformity property above, it holds that the limit $L_\theta(y)$ is a probability distribution function and continuous for either one of the variables y or θ if the other is fixed and furthermore that the probability measure L_θ is absolutely continuous with respect to the Lebesgue measure on $\mathbf{R}^k$.

Theorem. The asymptotic concentration of the ML estimator $\{\hat{\theta}_n\}$ about θ is not less than that of any estimator $\{T_n\}$ with uniformity property: For any convex and symmetric set $S\subset\mathbf{R}^k$ about the origin,

$$\lim_{n\to\infty} P_\theta\{n^{1/2}(\hat{\theta}_n-\theta)\in S\}\geqslant \lim_{n\to\infty} P_\theta\{n^{1/2}(T_n-\theta)\in S\}.$$

Schmetterer (*Research papers in statistics*, F. N. David (ed.), 1966) provided the notion of the continuous convergence of distributions of estimators of θ which is weaker than uniform convergence. An estimator $\{T_n\}$ is said to be **regular** if

$$\mathscr{L}[n^{1/2}(T_n-\theta)-h\,|\,\theta+n^{-1/2}h]\to L_\theta \quad\text{as}\quad n\to\infty,$$

where L_θ is a probability distribution independent of h with $\theta+n^{-1/2}h\in\Theta$. It was shown that the ML estimator $\{\hat{\theta}_n\}$ is regular under ordinary regularity conditions. Hájek obtained the following characterization of the asymptotic distribution of any regular estimator and, independently Inagaki (*Ann. Inst. Statist. Math.*, 22 (1970); 25 (1973)) obtained a similar result.

Theorem. The asymptotic distribution L_θ of any regular estimator $\{T_n\}$ is represented as the †convolution of a normal distribution N_θ and some residual distribution G_θ:

$$L_\theta=N_\theta * G_\theta,$$

where $N_\theta=N_k(0,I(\theta)^{-1})$, the asymptotic distribution of the ML estimator $\{\hat{\theta}_n\}$ in the ordinary regular case. It follows from the characterization $L_\theta=N_\theta * G_\theta$, that the first two theorems in this section hold also for regular estimators. (→ LeCam, *Proc. 6th Berkeley Symp.*, 1972, and Roussas and Soms, *Ann. Inst. Statist. Math.*, 25 (1973).)

O. Higher-Order Asymptotic Efficiency

In Section N it was shown that the ML estimator is a BAN estimator. In general, however, there exist many BAN estimators. For example, consider the case of a †multinomial distribution where probabilities of events are parametrized. That is, let $X=(n_1,\dots,n_m)'$, $n=n_1+\dots+n_m$, be distributed as a multinomial distribution $M(n;\pi_1,\dots,\pi_m)$, $\pi_1+\dots+\pi_m=1$, and let Π_Θ be a subset of m-vectors, $\Pi_\Theta=\{\pi(\theta)=(\pi_1(\theta),\dots,\pi_m(\theta))';\theta\in\Theta\}$, $\Theta\subset\mathbf{R}^1$. Define $\hat{p}=(\hat{p}_1,\dots,\hat{p}_m)'=(n_1/n,\dots,n_m/n)'$. Then we consider (i) the ML estimator, (ii) the minimum-chi-square estimator, (iii) the minimum modified chi-square estimator, (iv) the minimum Haldane discrepancy (D_k) estimator, (v) the minimum Hellinger distance (HD) estimator, and (vi) the minimum Kullback-Leibler (K-L) information estimator. These are defined as the values of the parameter θ that minimize the following quantities, respectively: (i) $\mathrm{ML}=-\log L_n=-n\sum_{i=1}^m \hat{p}_i\log\pi_i(\theta)$; (ii) $\chi^2=\sum_{i=1}^m(n\hat{p}_i-n\pi_i(\theta))^2/(n\pi_i(\theta))$; (iii) $\mathrm{mod}\,\chi^2=\sum_{i=1}^m(n\hat{p}_i-n\pi_i(\theta))^2/(n\hat{p}_i)$; (iv) $D_k=\sum_{i=1}^m\pi_i(\theta)^{k+1}/\hat{p}_i^k$; (v) $\mathrm{HD}=\cos^{-1}\sum_{i=1}^m(\hat{p}_i\pi_i(\theta))^{1/2}$; and (vi) $\text{K-L}=\sum_{i=1}^m\pi_i(\theta)\log(\pi_i(\theta)/\hat{p}_i)$. Rao [32] showed that under suitable regularity conditions these estimators are Fisher consistent and BAN estimators.

Fisher-Rao Approach to Second-Order Asymptotic Efficiency. For $\theta\in\Theta\subset\mathbf{R}^1$, let $p_{\theta n}$ be the density for n i.i.d. observations $x=(x_1,\dots,x_n)$, and let $q_{\theta n}$ be the density of estimator T_n. The †Fisher information contained in X and in T_n are defined by $nI(\theta)=E(d\log p_{\theta n}/d\theta)^2$ and $nI_{T_n}(\theta)=E(d\log q_{\theta n}/d\theta)^2$, respectively. Rao defined the **first-order (asymptotic) efficient estimator** T_n satisfying one of the following conditions: (1) $n^{1/2}|d\log p_{\theta n}/d\theta-d\log q_{\theta n}/d\theta|\to0$ in probability; (2) $I(\theta)-I_{T_n}(\theta)\to0$ as $n\to\infty$; (3) the asymptotic correlation between $n^{1/2}(T_n-\theta)$ and $n^{1/2}d\log p_{\theta n}/d\theta$ is unity; (4) $|n^{1/2}d\log p_{\theta n}/d\theta-\alpha-\beta n^{1/2}(T_n-\theta)|\to0$ in probability. We note that the larger the condition number (j), the stronger the condition. A first-order efficient estimator is a BAN estimator. Fisher proposed

$$E_2'=\lim_{n\to\infty}(nI(\theta)-nI_{T_n}(\theta))=\lim_{n\to\infty}V_\theta(d\log p_{\theta n}/d\theta-d\log q_{\theta n}/d\theta)$$

as a measure of **second-order (asymptotic) efficiency** to distinguish different BAN estimators. Fisher stated without any sort of proof that the maximum likelihood estimator minimizes E_2', i.e., maximizes second-order efficiency. Rao proposed

$$E_2=\min_\lambda V_\theta(d\log p_{\theta n}/d\theta-\alpha n^{1/2}-\beta n(T_n-\theta)-\lambda n(T_n-\theta)^2)$$

as a measure of second-order efficiency for first-order efficient estimators $\{T_n\}$ satisfying condition (4). He showed that the estimators mentioned above for multinomial distribution are first-order efficient estimators satisfying condition (4) and furthermore calculated second-order efficiencies of these estimators measured in terms of E_2: (i) $E_2(\mathrm{ML})=\gamma^2(\theta)I(\theta)$; (ii) $E_2(\chi^2)=\Delta(\theta)+E_2(\mathrm{ML})$; (iii) $E_2(\mathrm{mod}\,\chi^2)=4\Delta(\theta)+E_2(\mathrm{ML})$; (iv) $E_2(D_k)=(k+1)^2\Delta(\theta)+E_2(\mathrm{ML})$; (v) $E_2(\mathrm{HD})=\Delta(\theta)/4+E_2(\mathrm{ML})$; (vi) $E_2(\text{K-L})=\Delta(\theta)+E_2(\mathrm{ML})$ with

$$\gamma^2(\theta)=(\mu_{02}-2\mu_{21}+\mu_{40})/I^2(\theta)-1-(\mu_{11}^2+\mu_{30}^2-2\mu_{11}\mu_{30})/I^3(\theta),$$

$$\Delta(\theta)=\sum_{i=1}^m(\pi_i'(\theta)/\pi_i(\theta))^2/2-\mu_{40}/I(\theta)+\mu_{30}^2/(2I^2(\theta)),$$

where

$$\mu_{rs}=\mu_{rs}(\theta)=E_\theta\{((df(X,\theta)/d\theta)/f(X,\theta))^r\times((d^2f(X,\theta)/d\theta^2)/f(X,\theta))^s\}$$
$$=\sum_{i=1}^m\pi_i(\theta)(\pi_i'(\theta)/\pi_i(\theta))^r(\pi_i''(\theta)/\pi_i(\theta))^s.$$

Rao [33] gave another definition of second-order efficiency based on the expansion of the variance of T_n after correcting for bias: $V_\theta(T_n)=(nI(\theta))^{-1}+\psi(\theta)n^{-2}+o(n^{-2})$. The quantity $\psi(\theta)$ is considered to be a measure of second-order efficiency. The results of Fisher and Rao were confined to multinomial distributions. Efron (*Ann. Statist.*, 3 (1975)) and Ghosh and Subramanyan (*Saṇkhyā*, sec. A, 36 (1974)) extended the results to the so-called curved exponential family of distributions. Efron gave a geometric interpretation to the effect that second-order efficiency is related to the curvature of a statistical problem corresponding to $\gamma(\theta)$ above, and S. Amari recently extended this differential-geometric approach to the discussion of higher-order efficiency of estimators. Rao suggested that E_2' is equal to E_2. Ghosh and Subramanyan gave a sufficient condition for the equality to hold, whereas Efron provided a counterexample to show that $E_2'\neq E_2$ in general.

Pfanzagl and Takeuchi-Akahira Approaches to Higher-Order Asymptotic Efficiency. For each $k=1,2,\ldots$, a $^\dagger\{c_n\}$-consistent estimator $\{T_n\}$ is said to be the **kth-order asymptotically median unbiased (AMU) estimator** if for any $\theta\in\Theta$ $(\subset\mathbf{R}^1)$ there exists a positive number δ such that

$$\lim_{n\to\infty}\sup_{|\tau-\theta|<\delta}c_n^{k-1}\left|P_\tau\{T_n\leqslant\tau\}-\frac{1}{2}\right|=0$$

or

$$\lim_{n\to\infty}\sup_{|\tau-\theta|<\delta}c_n^{k-1}\left|P_\tau\{T_n\geqslant\tau\}-\frac{1}{2}\right|=0.$$

This notion, which is an extension of the condition due to Bahadur for the asymptotic efficiency, was introduced by Takeuchi and Akahira. For a kth-order AMU estimator $\{T_n\}$, $G_0(t,\theta)+c_n^{-1}G_1(t,\theta)+\ldots+c_n^{-k+1}G_{k-1}(t,\theta)$ is called the **kth-order asymptotic distribution** of $c_n(T_n-\theta)$ if

$$\lim_{n\to\infty}c_n^{k-1}|P_\theta\{c_n(T_n-\theta)\leqslant t\}-G_0(t,\theta)-c_n^{-1}G_1(t,\theta)-\ldots-c_n^{-k+1}G_{k+1}(t,\theta)|=0.$$

Pfanzagl and Takeuchi and Akahira obtained the concrete form of the second- or third-order asymptotic distribution of the ML estimator. A kth-order AMU estimator is said to be **kth-order asymptotic efficient** if the kth-order asymptotic distribution of it attains uniformly the bound for the kth-order asymptotic distributions of the kth-order AMU estimators. Takeuchi and Akahira showed that under suitable regularity conditions, T_n is second-order asymptotic efficient if

$$P_\theta\{(nI(\theta))^{1/2}(T_n-\theta)\leqslant t\}=\Phi(t)+n^{-1/2}(3\mu_{11}(\theta)+2\mu_{30}(\theta))/(6I(\theta)^{3/2})\times t^2\varphi(t)+o(n^{-1/2}),$$

where $\Phi(t)$ is the standard normal distribution function and $\varphi(t)$ is its density function, and further that the modified ML estimator

$$\hat\theta_n^*=\hat\theta_n+n^{-1}\mu_{30}(\hat\theta_n)/(6I(\hat\theta_n)^2)$$

for the ML estimator $\hat\theta_n$ is second-order asymptotic efficient. Pfanzagl (*Ann. Statist.*, 1 (1973)) obtained a similar result. The formulation due to Takeuchi and Akahira is more extensive since it can be applied to the so-called nonregular cases.

P. Estimating Equations

We often determine an estimator as a solution $\theta=T_n(x_1,\ldots,x_n)$ of an equation $\Psi_n(x_1,\ldots,x_n;\theta)=0$; for example, the ML estimator as a solution of the likelihood equation. In such case, such an equation is called an **estimating equation** and $\Psi_n(\theta)=\Psi_n(X_1,\ldots,X_n;\theta)$, a random function, is called an **estimating function** [3].

Call T_n an **estimator based on an estimating function** $\Psi_n(\theta)$. The following result is a modification of a theorem due to Hodges and Lehmann (*Ann. Math. Statist.*, 34 (1963)).

Theorem. Let Θ be an interval of $\mathbf{R}^1$. Suppose that a real-valued estimating function $\Psi_n(\theta)$ satisfies the following three conditions:
(M1) $\Psi_n(\theta)$ is a nonincreasing function of the real parameter θ;
(M2) for any real number h, $n^{1/2}\Psi_n(\theta_0+n^{-1/2}h)-n^{1/2}\Psi_n(\theta_0)+\gamma h\to 0$ in probability, where γ is a positive constant; and
(M3) $\mathscr{L}[n^{1/2}\Psi_n(\theta_0)](y)\to\Phi(y/\sigma)$, where Φ is a continuous distribution function with zero mean and unit variance.
Define an estimator based on Ψ_n by $T_n=(\theta_n^*+\theta_n^{**})/2$, where $\theta_n^*=\sup\{\theta\mid\Psi_n(\theta)>0\}$ and $\theta_n^{**}=\inf\{\theta\mid\Psi_n(\theta)<0\}$, Then we have $\mathscr{L}[n^{1/2}(T_n-\theta_0)](y)\to\Phi(y\gamma/\sigma)$ as $n\to\infty$. Huber considered a formulation that guarantees the asymptotic normality of an ***M*-estimator**. An M-estimator T_n is defined by a minimum problem of the form $\sum_{i=1}^n\rho(X_i,T_n)=\min_\theta\sum_{i=1}^n\rho(X_i,\theta)$ or by an implicit equation $\sum_{i=1}^n\psi(X_i,T_n)=0$, where ρ is an arbitrary function and $\psi(x,\theta)=\partial\rho(x,\theta)/\partial\theta$. Note that $\rho(x,\theta)=-\log f(x,\theta)$ gives the ordinary ML estimator. Let Θ be a closed subset of $\mathbf{R}^k$, let $(\mathscr{X},\mathfrak{B},P)$ be a probability space, and let $\psi(x,\theta)$ be some function on $\mathscr{X}\times\Theta$ with value in $\mathbf{R}^k$. Assume that X_1, $X_2,\dots$ are independent random variables with values in $\mathscr{X}$ having the common probability distribution P that need not be a member of the parametric family. Consider an estimating function

$$\Psi_n(\theta)=\frac{1}{n}\sum_{i=1}^n\psi(X_i,\theta).$$

Assume that
(N1) for each fixed $\theta\in\Theta$, $\psi(x,\theta)$ is $\mathfrak{B}$-measurable and $\psi(x,\theta)$ is †separable;
(N2) the expected value $\lambda(\theta)=E\{\psi(X,\theta)\}$ exists for all $\theta\in\Theta$, and has a unique zero at $\theta=\theta_0\in\Theta^0$;
(N3) there exists a continuous function that is bounded away from zero, $b(\theta)\geqslant b_0>0$, such that $E(\sup_\theta\{|\psi(X,\theta)|/b(\theta)\})<\infty$, $\liminf_{|\theta|\to\infty}\{|\lambda(\theta)|/b(\theta)\}\geqslant 1$, and $E(\limsup_{|\theta|\to\infty}\{|\psi(X_i,\theta)-\lambda(\theta)|/b(\theta)\})<1$;
(N4) for $u(x,\theta,d)=\sup\{|\psi(x,\tau)-\psi(x,\theta)|\mid |\tau-\theta|\leqslant d\}$,
(i) as $d\to 0$, $E(u(X,\theta,d))\to 0$,
(ii) there exist positive numbers d_0, b_1, and b_2 such that $E(u(X,\theta,d))\leqslant b_1 d$ and $E(u(X,\theta,d)^2)\leqslant b_2 d$ for θ and d satisfying $0<|\theta-\theta_0|+d\leqslant d_0$;
(N5) in some neighborhood of θ_0, $\lambda(\theta)$ is continuously differentiable and the Jacobian matrix at $\theta=\theta_0$, $\Lambda=(\partial\lambda_i(\theta_0)/\partial\theta_j)$, is nonsingular; and
(N6) the covariance matrix $\Sigma=E\{\psi(X,\theta_0)\cdot\psi(X,\theta_0)'\}$ exists and is positive definite.

Theorem. Suppose that an estimator $\{T_n\}$ satisfies $\Psi_n(T_n)\to 0$ a.s. (or in probability) as $n\to\infty$. Then, under (N1)–(N4) (i), $T_n\to\theta_0$ a.s. (or in probability) as $n\to\infty$.

Lemma. Under (N1)–(N5),

$$\sup\{|\Psi_n(\tau)-\Psi_n(\theta_0)-\lambda(\tau)|/(n^{-1/2}+|\lambda(\tau)|);\ |\tau-\theta_0|\leqslant d_0\}\to 0$$

in probability as $n\to\infty$.

Suppose that $\{\mathscr{L}[n^{1/2}\Psi_n(T_n)]\}$ is †tight. It implies that $\Psi_n(T_n)\to 0$ in probability, and hence from the above theorem $T_n\to\theta_0$ in probability. Thus letting $\tau=T_n$ we have

$$|\Psi_n(T_n)-\Psi_n(\theta_0)-\lambda(T_n)|/(n^{-1/2}+|\lambda(T_n)|)\to 0$$

in probability. That is, for any $\varepsilon>0$ there exists an n_0 such that for $n\geqslant n_0$, $P\{n^{1/2}|\lambda(T_n)|<\varepsilon+|n^{1/2}\Psi_n(T_n)-n^{1/2}\Psi_n(\theta_0)|\}>1-\varepsilon$. This and the tightness of $\{\mathscr{L}[n^{1/2}\Psi_n(T_n)]\}$ and $\{\mathscr{L}[n^{1/2}\Psi_n(\theta_0)]\}$ lead to the tightness of $\{\mathscr{L}[n^{1/2}(T_n-\theta_0)]\}$, and the converse is also true. At the same time we have $n^{1/2}\Psi_n(T_n)-n^{1/2}\Psi_n(\theta_0)-\Lambda n^{1/2}(T_n-\theta_0)\to 0$ in probability. The following theorem is a straightforward consequence.

Theorem. Suppose that an estimator $\hat\theta_n$ satisfies $n^{1/2}\Psi_n(\hat\theta_n)\to 0$ as $n\to\infty$. Then, under (N1)–(N6), $\mathscr{L}[n^{1/2}(\hat\theta_n-\theta_0)\mid P]\to N_k(0,\Lambda^{-1}\Sigma\Lambda'^{-1})$ as $n\to\infty$.

Q. Interval Estimation

Interval estimation or **region estimation** is a method of statistical inference utilized to estimate the true value $g(\theta)$ of the given parametric function by stating that $g(\theta)$ belongs to a subset $S(x)$ of $\mathscr{A}$, based on the observed value x of the random variable X. If

$$P_\theta\{g(\theta)\in S(X)\}\geqslant 1-\alpha \qquad \text{for any } \theta\in\Theta$$

for a constant α $(0<\alpha<1)$, then the random region $S(X)$ is called a **confidence region** of $g(\theta)$ of **confidence level** $1-\alpha$, and the infimum of the left-hand side with respect to $\theta\in\Theta$ is called the **confidence coefficient**. In particular, if $\Theta\subset\mathbf{R}$ and a confidence region is an interval, as is often the case, then the region is called a **confidence interval**, and two boundaries of the interval are called **confidence limits**. If a particular subset $S(X)$ among the set of confidence regions of $g(\theta)$ of confidence level $1-\alpha$ minimizes $P_\theta\{g(\theta')\in S(X)\}$ for all pairs θ and θ' $(\neq\theta)$, $S(X)$ is said to be **uniformly most powerful**. If a confidence region $S(X)$ of $g(\theta)$ of confidence level $1-\alpha$ satisfies $P_\theta\{g(\theta')\in S(X)\}\leqslant 1-\alpha$ for all pairs θ and θ' $(\neq\theta)$, then it is said to be **unbiased**. The notion of **invariance of a confidence region** can be defined similarly, and the definition for $S(X)$ being **uniformly most powerful unbiased** (UMPU) or **uniformly most**

powerful invariant (**UMPI**) can be formulated in an obvious manner.

For each $\theta_0 \in \Theta$ let $A(\theta_0)$ be an †acceptance region of a †test of level α of the †hypothesis $\theta = \theta_0$. For each $x \in \mathscr{X}$ let $S(x) = \{\theta \mid x \in A(\theta), \theta \in \Theta\}$. Then $S(X)$ is a confidence region of θ of confidence level $1-\alpha$. If $A(\theta_0)$ is an acceptance region of a UMP test of the hypothesis $\theta = \theta_0$ for each θ_0, then $S(X)$ is a UMP confidence region of θ of confidence level $1-\alpha$. Furthermore, corresponding to an acceptance region $A(\theta_0)$ of a UMP unbiased (invariant) test, a UMP unbiased (invariant) confidence region can be constructed in a similar manner.

R. Tolerance Regions

Let X and Y be distributed according to probability distributions P_θ^X and P_θ^Y labeled by a common $\theta \in \Theta$ over measurable spaces $(\mathscr{X}, \mathfrak{B})$ and $(\mathscr{Y}, \mathfrak{C})$, respectively, and consider the problem of predicting a future value of the random variable Y using the observed value x of the random variable X. If a mapping $S(x)$ sending a point x to a set belonging to $\mathfrak{C}$ is used to predict that the value of Y will lie in the set $S(x)$, then the random region is called a **tolerance region** of Y. In particular, if a tolerance region of a real random variable is an interval, it is called a **tolerance interval** and its boundaries **tolerance limits**.

For simplicity, suppose that X and Y are independent. There are several kinds of tolerance regions. First, if $P_\theta^X(P_\theta^Y\{Y \in S(X)\} \geqslant \beta) \geqslant \gamma$ for any $\theta \in \Theta$, then $S(X)$ is called a tolerance region of Y of **content** β and **level** γ. Second, if $E_\theta^X(P_\theta^Y\{Y \in S(X)\}) \geqslant \beta$ for any $\theta \in \Theta$, then $S(X)$ is called a tolerance region of Y of **mean content** β. Suppose that the random variable $X = (X_1, \ldots, X_n)'$ is a †random sample and that both X_1 and Y obey the same distribution. If further the set $\{P_\theta^Y \mid \theta \in \Theta\}$ forms the totality of 1-dimensional †continuous distributions and the distribution of $P_\theta^Y\{Y \in S(X)\}$ does not depend on the choice of θ, then $S(X)$ is called a distribution-free tolerance region. For example, if $X_{(1)} \leqslant X_{(2)} \leqslant \ldots \leqslant X_{(n)}$ is an †order statistic, then the interval $[X_{(r)}, X_{(s)}]$ (for $r < s$) is a distribution-free tolerance interval for a random variable Y, independent of $X_1, \ldots, X_n$, which has the same distribution as X_1.

References

[1] H. Cramér, Mathematical methods of statistics, Princeton Univ. Press, 1946.
[2] C. R. Rao, Linear statistical inference and its applications, Wiley, second edition, 1973.
[3] S. S. Wilks, Mathematical statistics, Wiley, 1962.
[4] S. Zacks, The theory of statistical inference, Wiley, 1971.
[5] A. Birnbaum, On the foundation of statistical inference, J. Amer. Statist. Assoc., 57 (1962), 269–326.
[6] E. L. Lehmann and H. Scheffé, Completeness, similar regions and unbiased estimation I, II, Sankhyā, sec. A, 10 (1950), 305–340; 15 (1955), 219–236.
[7] H. Morimoto, N. Ikeda, and Y. Washio, Unbiased estimates based on sufficient statistics, Bull. Math. Statist., 6 (1956).
[8] E. W. Barankin, Locally best unbiased estimates, Ann. Math. Statist., 20 (1949), 477–501.
[9] A. Bhattacharya, On some analogues to the amount of information and their uses in statistical estimation, Sankhyā, sec. A, 8 (1946), 1–14.
[10] K. Ishii, Inequalities of the types of Chebyshev and Cramer-Rao and mathematical programming, Ann. Inst. Statist. Math., 16 (1964), 277–293.
[11] C. R. Rao, Information and the accuracy attainable in the estimation of statistical parameters, Bull. Calcutta Math. Soc., 37 (1945), 81–91.
[12] M. A. Girshick and L. G. Savage, Bayes and minimax estimates for quadratic loss function, Proc. 2nd Berkeley Symp. Math. Prob., 1 (1951), 53–74.
[13] S. Karlin, Admissibility for estimation with quadratic loss, Ann. Math. Statist., 29 (1958), 406–436.
[14] E. J. Pitman, The estimation of the location and scale parameters of a continuous population of any given form, Biometrika, 30 (1939), 391–421.
[15] C. Stein, Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, Proc. 3rd Berkeley Symp. Math. Prob., 1 (1956), 197–206.
[16] C. Stein, The admissibility of the Pitman's estimator for a single location parameter, Ann. Math. Statist., 30 (1959), 970–979.
[17] J. L. Hodges and E. L. Lehmann, Some problems in minimax estimation, Ann. Math. Statist., 21 (1950), 182–197.
[18] H. Kudo, On minimax invariant estimates of the transformation parameter, Nat. Sci. Rep. Ochanomizu Univ., 6 (1955), 31–73.
[19] J. Wolfowitz, The efficiency of sequential estimates and Wald's equation for sequential processes, Ann. Math. Statist., 18 (1947), 215–230.
[20] R. A. Fisher, On the mathematical foundations of theoretical statistics, Philos. Trans. Roy. Soc. London, 222 (1922), 309–368.
[21] R. A. Fisher, Theory of statistical estimation, Proc. Cambridge Philos. Soc., 22 (1925), 700–725.

[22] R. von Mises, On the asymptotic distributions of differentiable statistical functions, Ann. Math. Statist., 18 (1947), 309–348.
[23] L. LeCam, On some asymptotic properties of maximum likelihood estimates and related Bayes estimates, Univ. California Publ. Statist., 1 (1953), 277–330.
[24] L. LeCam, Locally asymptotically normal families of distributions, Univ. California Publ. Statist., 3 (1960), 37–98.
[25] G. G. Roussas, Contiguity of probability measures, Cambridge Univ. Press, 1972.
[26] A. Wald, Note on the consistency of the maximum likelihood estimate, Ann. Math. Statist., 20 (1949), 595–601.
[27] R. R. Bahadur, On Fisher's bound for asymptotic variances, Ann. Math. Statist., 35 (1964), 1545–1552.
[28] J. Hájek, A characterization of limiting distributions of regular estimates, Z. Wahrsheinlichkeitstheorie und Verw. Gebiete, 14 (1970), 323–330.
[29] S. Kaufman, Asymptotic efficiency of the maximum likelihood estimator, Ann. Inst. Statist. Math., 18 (1966), 155–178.
[30] J. Wolfowitz, Asymptotic efficiency of the maximum likelihood estimator, Theory Prob. Appl., 10 (1965), 247–260.
[31] M. Akahira and K. Takeuchi, Asymptotic efficiency of statistical estimators: Concepts and higher order asymptotic efficiency, Lecture notes in statist. 7, Springer, 1981.
[32] C. R. Rao, Asymptotic efficiency and limiting information, Proc. 4th Berkeley Symp. Math. Prob., 1 (1961), 531–546.
[33] C. R. Rao, Criteria of estimation in large samples, Sankhyā, sec. A, 25 (1963), 189–206.
[34] P. J. Huber, The behavior of maximum likelihood estimates under nonstandard conditions, Proc. 5th Berkeley Symp. Math. Statist. Prob., 1 (1967), 221–233.

400 (XVIII.8) Statistical Hypothesis Testing

A. General Remarks

A **statistical hypothesis** is a proposition about the †probability distribution of a †sample X. If it is known that the †distribution of X belongs to a family of distributions $\mathscr{P}=\{P_\theta \mid \theta\in\Omega\}$ with a parameter space Ω, the hypothesis can be stated as follows: The value of the parameter θ belongs to ω_H, where ω_H is a nonempty subset of the parameter space Ω. This hypothesis is also written simply as $\mathbf{H}:\theta\in\omega_H$. When ω_H consists of one point, it is called a **simple hypothesis**, otherwise a **composite hypothesis**.

Let $\mathscr{X}$ be a †sample space associated with a †σ-algebra $\mathfrak{B}$ of subsets of $\mathscr{X}$. To test a hypothesis $\mathbf{H}$ is to decide whether $\mathbf{H}$ is false, based on the observation of a sample $X(\in\mathscr{X})$. The assertion that $\mathbf{H}$ is not false does not necessarily imply the validity of $\mathbf{H}$. Such an assertion is called the **acceptance** of $\mathbf{H}$, while the opposite assertion, that $\mathbf{H}$ is false, is called the **rejection** of $\mathbf{H}$. In this framework for the testing problem, $\mathbf{H}$ is often called a **null hypothesis** (→ 401 Statistical Inference).

Consider a testing procedure in which $\mathbf{H}$ is rejected with probability $\varphi(x)$ $(0\leqslant\varphi(x)\leqslant 1)$ and accepted with probability $1-\varphi(x)$, when $x\in\mathscr{X}$ is observed. This testing procedure is characterized by the function φ on $\mathscr{X}$ with range in $[0,1]$. Here $\varphi(x)$ is taken as $\mathfrak{B}$-measurable on $\mathscr{X}$, and is called a **test function** or **test**. If $\varphi(x)$ is the indicator function $\chi_B(x)$ of a set B $(\in\mathfrak{B})$, then the test is rejecting $\mathbf{H}$ when x belongs to B and accepting $\mathbf{H}$ otherwise. The set B is called a **critical region**, and its complementary set an **acceptance region**. A test is called a **nonrandomized test** if it is the indicator function of a set. Other tests are called **randomized tests**.

Suppose that the †distribution of the sample X is a probability measure P_θ on $(\mathscr{X},\mathfrak{B})$. The probability of rejecting $\mathbf{H}$ when θ is the true value of the parameter is calculated from

$$E_\theta(\varphi)=\int_{\mathscr{X}}\varphi(x)P_\theta(dx).$$

Let α be a given constant in $(0,1)$. If a test $\varphi(x)$ satisfies $E_\theta(\varphi)\leqslant\alpha$ for all $\theta\in\omega_H$, or, in other words, if the probability of rejecting $\mathbf{H}$ when $\mathbf{H}$ is true is not greater than α, α is called the **level** of φ and such a test is called a **level** α **test**. We denote the set of all level α tests by $\Phi(\alpha)$, and $\sup_{\theta\in\omega_H}E_\theta(\varphi)$ is called the **size** of φ. To judge the merit of tests, we introduce a different hypothesis $\mathbf{A}$: The true value of θ belongs to $\omega_A\subset\Omega-\omega_H$. This is called an **alternative hypothesis**, or, for simplicity, an **alternative**. The errors of a test are divided into two kinds: errors owing to the rejection of the hypothesis $\mathbf{H}$ when it is true, and errors owing to the acceptance of $\mathbf{H}$ when it is false. The former are called **errors of the first kind**, and the latter, **errors of the second kind**. The probability $E_\theta(\varphi)$ of rejecting $\mathbf{H}$ when $\theta\in\omega_A$, that is, the probability of the correct decision being made for $\theta\in\omega_A$, is called the **power** of a test or the **power function**. The probability of committing an error of the second kind is $1-E_\theta(\varphi)$ for $\theta\in\omega_A$. A testing problem is indicated by the notation $(\mathscr{X},\mathfrak{B},\mathscr{P},\omega_H,\omega_A)$. A test φ in a class $\Phi(\alpha)$ of tests is said to be **uniformly most powerful** in $\Phi(\alpha)$ (or **UMP** in $\Phi(\alpha)$) if for any

$\psi \in \Phi(\alpha)$, $E_\theta(\varphi) \geqslant E_\theta(\psi)$ for all $\theta \in \omega_A$. When ω_A consist of a single point, it is said to be **most powerful**.

B. The Neyman-Pearson Fundamental Lemma

Let μ be a †σ-finite measure over $(\mathscr{X}, \mathfrak{B})$ and $f_1, \ldots, f_{n+1}$ be μ-integrable real-valued functions. If $c_1, \ldots, c_n$ are constants such that the set $\Phi(c_1, \ldots, c_n)$ of test functions φ satisfying

$$\int \varphi f_i \, d\mu \leqslant c_i, \qquad i = 1, 2, \ldots, n,$$

is not empty, then there exists at least one test φ_0 in $\Phi(c_1, \ldots, c_n)$ that maximizes $\int \varphi f_{n+1} \, d\mu$ among all φ in $\Phi(c_1, \ldots, c_n)$. A test $\tilde{\varphi}$ is one of these tests if it satisfies the following two conditions:

(1) For appropriate constants $k_1, \ldots, k_n \geqslant 0$,

$$\tilde{\varphi}(x) = \begin{cases} 1 & \text{when } f_{n+1}(x) > \sum_{i=1}^{n} k_i f_i(x), \\ 0 & \text{when } f_{n+1}(x) < \sum_{i=1}^{n} k_i f_i(x) \end{cases}$$

almost everywhere with respect to μ, and

(2) the equation

$$\int \tilde{\varphi} f_i \, d\mu = c_i, \qquad i = 1, 2, \ldots, n,$$

holds.

If $(c_1, \ldots, c_n)$ is an interior point of the subset

$$\mathbf{M} = \left\{ \left(\int \varphi f_1 \, d\mu, \ldots, \int \varphi f_n \, d\mu \right) \middle| \varphi \text{ is a test} \right\}$$

of the n-space $\mathbf{R}^n$ and $\tilde{\varphi}$ satisfies (2) and maximizes $\int \varphi f_{n+1} \, d\mu$ among all φ in $\Phi(c_1, \ldots, c_n)$, then $\tilde{\varphi}$ satisfies (1). These statements are called the **Neyman-Pearson lemma**.

As an illustration, suppose that $\Omega = (1, 2, \ldots, n, n+1)$ and that $\{P_\theta \mid \theta \in \Omega)$ is dominated by a σ-finite measure μ. Let $f_i(x)$ be the density of P_i with respect to μ. When ω_H is a finite set $\{1, \ldots, n\}$ and ω_A consists of a single point $n+1$, then $\tilde{\varphi}$ satisfying (1) and (2) with $c_1 = \ldots = c_n = \alpha$ is a uniformly most powerful level α test.

If $\mathfrak{B}$ is generated by a countable number of sets, then there exists a most powerful level α test for any hypothesis against a simple alternative $\mathbf{A}: \theta = \tilde{\theta}$. A method of obtaining such a most powerful test is given in the **Lehman-Stein theorem**: Denote by f_θ the density function of P_θ with respect to a σ-finite measure μ, and define

$$h_\lambda(x) = \int_{\omega_H} f_\theta(x) \, d\lambda(\theta)$$

for a probability measure λ on ω_H. Consider testing the simple hypotheses $\mathbf{H}_\lambda$: The density of distribution of the sample is h_λ, against the alternative $\mathbf{A}: \theta = \tilde{\theta}$, and let φ_λ be a most powerful level α test for this problem $(\mathbf{H}_\lambda : \mathbf{A})$. If $\sup_{\theta \in \omega_H} E_\theta(\varphi_\lambda) \leqslant \alpha$, φ_λ is a most powerful level α test for testing $\mathbf{H}: \theta \in \omega_H$ against $\mathbf{A}: \theta = \tilde{\theta}$. The measure λ satisfies $E_\theta(\varphi_{\lambda'}) \geqslant E_\theta(\varphi_\lambda)$ for any probability measure λ' on ω_H and is called a **least favorable distribution**.

When the alternative hypothesis ω_A consists of more than one point, a uniformly most powerful test does not generally exist. However, if $\Omega = \mathbf{R}$, $\omega_H = (-\infty, \theta_0]$, $\omega_A = (\theta_0, \infty)$, and $f_\theta(x)$ is a density function with †monotone likelihood ratio with respect to a statistic $T(x)$, then a UMP level α test $\varphi(x)$ exists and is defined by $\varphi(x) = 1$ if $T(x) > c$; $=$ a constant if $T(x) = c$; and $= 0$ if $T(x) < c$. For a one-parameter exponential family of distributions, there exists a UMP level α test for testing $\mathbf{H}: \omega_H = (-\infty, \theta_1] \cup [\theta_2, \infty)$ against $\mathbf{A}: \omega_A = (\theta_1, \theta_2)$ $(\theta_1 < \theta_2)$. However, a UMP test does not exist for the problem obtained by interchanging the positions of ω_H and ω_A.

Since hypothesis-testing problems admitting UMP tests are rather rare, alternative ways for judging the merit of tests are needed, and two have been devised. The first is to restrict the class Φ of tests and to find a UMP test in this restricted class. The second is to introduce an alternative criterion of optimality and to select a test accordingly. The first is discussed in detail in Sections C, D, and E, and the second in Section F.

C. Unbiased Tests

The unbiasedness criterion is based on the idea that the probability of rejecting the hypothesis $\mathbf{H}$ when it is true (the probability of an error of the first kind) should preferably be no larger than that of rejecting $\mathbf{H}$ when it is false (the power). If a level α test φ satisfies $E_\theta(\varphi) \geqslant \alpha$ for $\theta \in \omega_A$, then φ is called an **unbiased level** α **test**. Let $\Phi_u(\alpha)$ be the set of all unbiased level α tests. A UMP test in $\Phi_u(\alpha)$ is called a **uniformly most powerful** (or **UMP**) **unbiased level** α **test**.

If P_θ is of the †exponential family whose parameter space Ω is a finite or an infinite open interval of $\mathbf{R}^k$, then there exists a UMP unbiased level α test for the following two problems: (1) $\mathbf{H}: \theta_1 \leqslant a$, $\mathbf{A}: \theta_1 > a$, where θ_1 is the first coordinate of $\theta = (\theta_1, \ldots, \theta_k)$; and (2) $\mathbf{H}: \theta_1 = a$, $\mathbf{A}: \theta_1 \neq a$. For example, when the sample is normally distributed with unknown mean μ and unknown variance σ^2, the Student test (defined in Section G) for a hypothesis $\mathbf{H}: \mu = \mu_0$ against an alternative $\mathbf{A}: \mu \neq \mu_0$ is a UMP unbiased test.

D. Similar Tests and Neyman Structure

If $E_\theta(\varphi)$ is constant for all $\theta\in\omega$ ($\subset\Omega$), φ is called a **similar test** with respect to ω. If $E_\theta(\varphi)$ is a continuous function of θ, an unbiased test φ is similar with respect to the common boundary $\tilde{\omega}$ of ω_H and ω_A, provided that Ω is a topological space and the density function is continuous in ω. Therefore, in this case, unbiased tests are found in the class of all tests similar with respect to $\tilde{\omega}$. Let a statistic $y=T(x)$ be †sufficient for $\mathscr{P}_\omega=\{P_\theta \mid \theta\in\omega\}$. A test φ is said to have **Neyman structure** with respect to T if the conditional expectation $E(\varphi \mid T(x)=y)$ of φ equals a constant $[\mathscr{P}_\omega]$. (For the notation $[\mathscr{P}_\omega]$ → 396 Statistic.) A test φ having Neyman structure with respect to the statistic $T(x)$ is similar with respect to ω. A test φ similar with respect to ω has Neyman structure with respect to $T(x)$ if and only if the family $\mathscr{Q}=\{Q_\theta=P_\theta T^{-1} \mid \theta\in\omega\}$ of †marginal distributions of T is †boundedly complete.

E. Invariant Tests

Consider groups G and $\bar{G}$ of one-to-one transformations on $\mathscr{X}$ and Ω, respectively. Suppose that each element of G is a measurable transformation of $\mathscr{X}$ onto itself (i.e., $gB\in\mathfrak{B}$ for any $B\in\mathfrak{B}$) and that a homomorphism $g\rightarrow\bar{g}$ of G into $\bar{G}$ is defined so that $P_\theta(g^{-1}B)=P_{\bar{g}\theta}(B)$. The hypothesis $\mathbf{H}:\theta\in\omega_H$ and the alternative $\mathbf{A}$: $\theta\in\omega_A$ are said to be **invariant** under G if $\bar{g}\omega_H=\omega_H$ and $\bar{g}\omega_A=\omega_A$ for all $g\in G$, and in this case the testing problem $(\mathscr{X},\mathfrak{B},\mathscr{P},\omega_H,\omega_A)$ is said to be **invariant** under G. A test is called an **invariant test** if $\varphi(gx)=\varphi(x)$ for every $g\in G$, and $E_{\bar{g}\theta}(\varphi)=E_\theta(\varphi)$ holds for any invariant φ. Accordingly, if $\bar{G}$ is †transitive on ω_H, and invariant test is similar with respect to ω_H. If the sample space $\mathscr{X}$ is a subset of $\mathbf{R}^n$ that is invariant under the translation $(x_1,\ldots,x_n)\rightarrow(x_1+a,\ldots,x_n+a)$ with a real a and if there exists a $\theta'=\bar{a}\cdot\theta\in\Omega$ such that $P_{\theta'}(B)=P_\theta(\{(x_1-a,\ldots,x_n-a) \mid (x_1,\ldots,x_n)\in B\})$ for any $\theta\in\Omega$, then $\bar{a}$ is a transformation on Ω. In this case the real number a is called a **location parameter**. Furthermore, if the sample space $\mathscr{X}$ is a subset invariant under the similarity transformation $(x_1,\ldots,x_n)\rightarrow(ax_1,\ldots,ax_n)$ $(a>0)$ and if there exists a $\theta'=\bar{a}\cdot\theta\in\Omega$ such that $P_{\theta'}(B)=P_\theta(\{(x_1/a,\ldots,x_n/a) \mid (x_1,\ldots,x_n)\in B\})$ for any $\theta\in\Omega$, then the real number a is called a **scale parameter**. The **invariance principle** states that a test for a testing problem invariant under G should preferably be invariant under G. A test $\varphi(x)$ is called an **almost invariant test** if $\varphi(gx)=\varphi(x)$ $[\mathscr{P}]$ for all $g\in G$.

Suppose that the testing problem of a hypothesis under consideration is invariant under a transformation group G on the sample space $\mathscr{X}$. Denote the set of all invariant level α tests by $\mathbf{\Phi}_I(\alpha)$. A test that is uniformly most powerful in $\mathbf{\Phi}_I(\alpha)$ is called a **uniformly most powerful** (in short, **UMP**) **invariant level** α **test**. If there exists a unique UMP unbiased level α test φ^*, then a UMP invariant level α test (if it exists) coincides with $\varphi^*[\mathscr{P}]$. When $T(x)$ is †maximal invariant under G, a necessary and sufficient condition for $\varphi(x)$ to be invariant is that φ be a function of $T(x)$.

For example, suppose that the sample $X=(X_1,\ldots,X_n)$ is taken from $N(\mu,\sigma^2)$ with unknown μ and σ^2. In this situation, $Y=(\bar{X},S)$ is a sufficient statistic, where $\bar{X}=\sum_i X_i/n$ and $S=\sqrt{\sum(X_i-\bar{X})^2}$. Let G be the group of transformations $(\bar{x},s)\rightarrow(c\bar{x},cs)$ $(c>0)$ on the range $\mathscr{Y}$ of Y and $\bar{G}$ be the group of transformations $(\mu,\sigma^2)\rightarrow(c\mu,c\sigma^2)$ $(c>0)$ on the parameter space. Both the hypotheses $\mathbf{H}_1:\mu/\sigma^2\leqslant 0$ and $\mathbf{H}_2$: $\mu/\sigma^2=0$ are invariant under G. Since $t=\sqrt{n}\cdot\bar{x}/(s/\sqrt{n-1})$ is maximal invariant, any invariant level α test is in the class of functions of t. The Student test, defined in Section G, is UMP invariant under G.

F. Minimax Tests and Most Stringent Tests

Minimax tests and most stringent tests are sometimes used as alternatives to UMP tests. Suppose that $\mathscr{P}=\{P_\theta \mid \theta\in\Omega\}$ is a †dominated family and $\mathfrak{B}$ is generated by a countable number of sets. A level α test φ^* is called a **minimax level** α **test** if for any level α test φ,

$$\inf_{\theta\in\omega_A} E_\theta(\varphi^*)\geqslant\inf_{\theta\in\omega_A} E_\theta(\varphi).$$

Such a test exists for any $\alpha\in(0,1)$. If a group G of measurable transformations on $\mathscr{X}$ leaves a testing problem invariant, then an intimate relation exists between the minimax property and invariance. Concerning this relation, we have the following theorem: For each $\alpha\in(0,1)$ there is an almost invariant level α test that is minimax if there exists a σ-field $\mathfrak{A}$ of subsets of G and a sequence $\{v_n\}$ of probability measures on $(G,\mathfrak{A})$ such that (i) $B\in\mathfrak{B}$ implies $\{(x,g) \mid gx\in B\}\in\mathfrak{B}\times\mathfrak{A}$; (ii) $A\in\mathfrak{A}$, $g\in G$ implies $Ag\in\mathscr{A}$; and (iii) $\lim_{n\rightarrow\infty}|v_n(Ag)-v_n(A)|=0$ for any $A\in\mathfrak{A}$ and $g\in G$. Fundamental in the invariant testing problem is the **Hunt-Stein lemma**: Under the condition just stated, for any φ there exists an almost invariant test ψ such that

$$\inf_{\bar{g}\in\bar{G}} E_{\bar{g}\theta}(\varphi)\leqslant E_\theta(\psi)\leqslant\sup_{\bar{g}\in\bar{G}} E_{\bar{g}\theta}(\varphi).$$

The following six types of transformation groups satisfy the condition of the theorem: (1) the group of translations on $\mathbf{R}^n$, (2) the group of similarity transformations on $\mathbf{R}^n$, (3) the

group of transformations $g=(a,b):(x_1,\ldots,x_n)\in \mathbf{R}^n\to(ax_1+b,\ldots,ax_n+b)\in\mathbf{R}^n$ $(0<a<\infty,\ -\infty<b<\infty)$, (4) finite groups, (5) the group of orthogonal transformations on $\mathbf{R}^n$, and (6) the direct product of a finite number of the groups mentioned in (1)–(5).

We call $\beta_\alpha^*(\theta)=\sup_{\varphi\in\Phi(\alpha)}E_\theta(\varphi)$ an **envelope power function**, and φ^* $(\in\Phi(\alpha))$ is called a **most stringent level α test** if

$$\sup_{\theta\in\omega_A}(\beta_\alpha^*(\theta)-E_\theta(\varphi^*))\leqslant\sup_{\theta\in\omega_A}(\beta_\alpha^*(\theta)-E_\theta(\varphi))$$

for any $\varphi\in\Phi(\alpha)$. There exists a most stringent level α test for each $\alpha\in(0,1)$. If a testing problem is invariant under a transformation group G on X and G satisfies the condition in the Hunt-Stein lemma, then a uniformly most powerful invariant level α test is most stringent among the level α tests (→ 398 Statistical Decision Functions).

Admissibility of a test and completeness of a class of tests are defined with respect to the probability of an error of the second kind (→ 398 Statistical Decision Functions). The uniformly most powerful level α test and the uniformly most powerful unbiased level α test are admissible.

G. Useful Tests Concerning Normal Distributions (→ Appendix A, Table 23)

In this section, we treat the rejection regions S that are commonly used in testing problems related to normal distributions. Let α be the level of S, and let $c(\alpha)$ and $d(\alpha)$ be positive numbers determined by α. In (1)–(5) below, the sample consists of n mutually independent random variables $X_1,\ldots,X_n$ each of which is assumed to be normally distributed with mean μ and variance σ^2. For any sample point $x=(x_1,\ldots,x_n)$, denote $\sum_{i=1}^n x_i/n$ by $\bar{x}$ and $\sum_{i=1}^n(x_i-\bar{x})^2$ by s^2. (1) To test the hypothesis $\mu\leqslant\mu_0$ against an alternative $\mu>\mu_0$, we can use as a critical region $S=\{x\,|\,t(x)>c(\alpha)\}$, where the test statistic $t(x)$ is given by $\sqrt{n}(\bar{x}-\mu_0)/\sqrt{s^2/(n-1)}$. (2) To test the hypothesis $\mu=\mu_0$ against an alternative $\mu\neq\mu_0$, we can use $S=\{x\,|\,|t(x)|\geqslant c(\alpha)\}$ with the same test statistic $t(x)$ as in (1). These tests based on the statistic $t(x)$ are generally called **Student tests** or ***t*-tests**. (3) To test the hypothesis $\sigma^2=\sigma_0^2$ against the alternative $\sigma^2>\sigma_0^2$ with $\sigma_0^2>0$, we can use $S=\{x\,|\,\chi^2(x)\geqslant c(\alpha)\}$, where $\chi^2=s^2/\sigma_0^2$. (4) To test the hypothesis $\sigma^2>\sigma_0^2$ against the alternative $\sigma^2<\sigma_0^2$, we can use $S=\{x\,|\,\chi^2(x)\leqslant c(\alpha)\}$, where χ^2 is the same as in (3). (5) To test the hypothesis $\sigma^2=\sigma_0^2$ against $\sigma^2\neq\sigma_0^2$, we can use $S=\{x\,|\,\chi^2(x)\leqslant c(\alpha)$ or $\geqslant d(\alpha)\}$, where χ^2 is the same as in (3). Each of these tests based on the statistic χ^2 is called a **chi-square test**. Among these tests, (3) is UMP and (1) is UMP when $\alpha\geqslant 1/2$. All tests (1)–(5) are UMP unbiased, and (3)–(5) are UMP invariant under the translations $(x_1,\ldots,x_n)\to(x_1+a,\ldots,x_n+a)$ $(-\infty<a<\infty)$. Since (1) and (2) are also UMP invariant under the transformations $(x_1,\ldots,x_n)\to(ax_1,\ldots,ax_n)$ $(0<a<\infty)$, they are most stringent tests.

Suppose that $X_1,\ldots,X_m$ are independently distributed according to $N(\mu_1,\sigma_1^2)$ and that $Y_1,\ldots,Y_n$ are independently distributed according to $N(\mu_2,\sigma_2^2)$, where μ_1, μ_2, σ_1, and σ_2 are assumed unknown unless otherwise stated. Here we give the important tests for μ_1, μ_2, σ_1^2, σ_2^2. Let $x=(x_1,\ldots,x_m)$ and $y=(y_1,\ldots,y_n)$ be sample points in $\mathbf{R}^m$ and $\mathbf{R}^n$, respectively, and denote $\sum_{i=1}^m x_i/m$, $\sum_{i=1}^n y_i/n$, $\sum_{i=1}^m(x_i-\bar{x})^2$, and $\sum_{i=1}^n(y_i-\bar{y})^2$ by $\bar{x}$, $\bar{y}$, s_x^2, and s_y^2, respectively. (6) Assume that σ_1 and σ_2 are known, and consider a hypothesis $\mu_1=\mu_2$. When an alternative $\mu_1>\mu_2$ $(\mu_1\neq\mu_2)$ is taken, we can use as a critical region $S=\{(x,y)\,|\,T(x,y)\geqslant c(\alpha)\}$ $(S=\{(x,y)\,|\,|T(x,y)|\geqslant c(\alpha)\})$, where $T(x,y)=(\bar{x}-\bar{y})/\sqrt{\sigma_1^2/m+\sigma_2^2/n}$. These tests are both UMP unbiased and invariant under the translations $(x_1,\ldots,x_m,y_1,\ldots,y_n)\to(x_1+a,\ldots,x_m+a,y_1+a,\ldots,y_n+a)$. (7) Assume that $\sigma_1=\sigma_2$, and consider a hypothesis $\mu_1=\mu_2$. When $\mu_1\neq\mu_2$ $(\mu_1>\mu_2)$ is the alternative, $S=\{(x,y)\,|\,|T(x,y)|\geqslant c(\alpha)\}$ $(S=\{(x,y)\,|\,T(x,y)>c(\alpha)\})$ can be used as a critical region, where $T(x,y)=(\bar{x}-\bar{y})\sqrt{m+n-2}/(\sqrt{1/m+1/n}\sqrt{s_x^2+s_y^2})$. Both tests are UMP unbiased and are invariant under $(x_1,\ldots,x_m,y_1,\ldots,y_n)\to(ax_1+b,\ldots,ax_m+b,ay_1+b,\ldots,ay_n+b)$ $(-\infty<b<\infty,0<a<\infty)$. (8) Testing the hypothesis $\mathbf{H}$: $\mu_1=\mu_2$ is called the **Behrens-Fisher problem**. Note that nothing is assumed about the relation of the variances σ_1^2 and σ_2^2 of the two samples X and Y, in contrast to (7). It is difficult to construct a statistic whose distribution is independent of σ_1^2 and σ_2^2 when $\mathbf{H}$ is true. Compare this with (1)–(7), where the proposed statistics have this property and are used to construct similar critical regions S. The critical region

$$\{(x,y)\,|\,(\bar{x}-\bar{y})/\sqrt{s_x^2/m(m-1)+s_y^2/n(n-1)}\geqslant f(s_y^2/s_x^2)\},$$

with an appropriately chosen f, is similar to such a region S. This test is called **Welch's test**. (9) For a hypothesis $\sigma_1=\sigma_2$, we can use as a critical region $S=\{(x,y)\,|\,F(x,y)\leqslant c(\alpha)\}$, $S=\{(x,y)\,|\,F(x,y)\geqslant c(\alpha)\}$, or $S=\{(x,y)\,|\,F(x,y)\geqslant d(\alpha)$ or $<c(\alpha)\}$, when $\sigma_1<\sigma_2$, $\sigma_2<\sigma_1$, or $\sigma_1\neq\sigma_2$, respectively, is taken as alternative, where $F(x,y)=(n-1)s_x^2/(m-1)s_y^2$. All these tests are UMP unbiased and are invariant under the transformations $(x_1,\ldots,x_m,y_1,\ldots,y_n)\to(ax_1+b,\ldots,ax_m+b,ay_1+b,\ldots,ay_n+b)$ $(-\infty<$

$b<\infty, 0<a<\infty$). A test based on $F(x, y)$ is called an **F-test**.

H. Linear Hypotheses

Let $X_1, \ldots, X_n$ be independent and distributed according to $N(\mu_i, \sigma^2)$ $(i=1,2,\ldots,n)$, where $\mu_1, \mu_2, \ldots, \mu_s$ $(s<n)$, σ are assumed to be unknown and $\mu_i=0$ $(s<i\leqslant n)$. The hypothesis is $\mathbf{H}: \mu_1=\mu_2=\ldots=\mu_r=0$ $(r\leqslant s)$, and the alternative hypothesis is that at least one μ_i, $1\leqslant i\leqslant r$, does not vanish. The critical region $S=\{(x_1, \ldots, x_n) \mid F(x_1, \ldots, x_r; x_{s+1}, \ldots, x_n)=\sum_{i=1}^r x_i^2/\sum_{i=s+1}^n x_i^2 \geqslant c(\alpha)\}$ is a UMP unbiased test for this problem, and S is invariant under the group g_1 of translations

$$(x_1, \ldots, x_r, x_{r+1}, \ldots, x_s, x_{s+1}, \ldots, x_n)$$
$$\rightarrow (x_1, \ldots, x_r, x_{r+1}+a, \ldots, x_s+a, x_{s+1}, \ldots, x_n),$$

the group g_2 of similarity transformations $(x_1, \ldots, x_n)\rightarrow(cx_1, \ldots, cx_n)$, the group $g_3=O(r)$ of orthogonal transformations in $\mathbf{R}^r=\{(x_1, \ldots, x_r)\}$, the group $g_4=O(n-s)$ of orthogonal transformations in $\mathbf{R}^{n-s}=\{(x_{s+1}, \ldots, x_n)\}$, and finite products of elements of the groups g_1, g_2, g_3, and g_4. This test is also a kind of F-test. More generally, let us denote $\mathbf{X}=(X_1, X_2, \ldots, X_n)'$ and assume that it is expressed as

$$\mathbf{X}=A\xi+\mathbf{W}, \qquad \mathbf{W}=(W_1, W_2, \ldots, W_n)', \tag{1}$$

where $\xi=(\xi_1, \xi_2, \ldots, \xi_s)'$, $s\leqslant n$ is a vector of unknown parameters and A a matrix of known constants, and $W_1, W_2, \ldots, W_n$ are distributed independently according to the normal distribution with mean 0 and variance σ^2. Then a **general linear hypothesis** is a hypothesis stating that the vector ξ lies within a linear subspace M of $\mathbf{R}^s$. The set of points $A\xi$ with ξ satisfying a linear hypothesis $\mathbf{H}$ is the linear subspace $L(B)$ of $L(A)$ spanned by the column vectors of an $n\times k_1$ matrix B. Assume, for example, that the dimension of $L(B)$ (= the rank of B) is $s-r$. Let C be an $n\times k_2$ matrix whose column vectors span the orthocomplement $L_A^\perp(B)$ of the space $L(B)$ with respect to the space $L(A)$. Then the model (1) can be written as

$$\mathbf{X}=B\eta+C\zeta+\mathbf{W}, \qquad E(\mathbf{W})=\mathbf{0}, \tag{2}$$

with a k_1-vector η and a k_2-vector ζ, and hence the hypothesis $\mathbf{H}$ is represented by $\zeta=\mathbf{0}'$. We denote by $\mathbf{Y}=P_A\mathbf{X}$ the projection of $\mathbf{X}$ onto the space $L(A)$ and by $\mathbf{Z}$ the projection $P_B\mathbf{X}$ of $\mathbf{X}$ onto the space $L(B)$. The quantity $Q_H=\mathbf{X}'(P_A-P_B)\mathbf{X}$ equals the square of the length of the vector $\mathbf{Y}-\mathbf{Z}$ and represents the sum of squares of residuals for the hypothesis $\mathbf{H}$. The error mean square $\hat{\sigma}^2=\mathbf{X}'(I-P_A)\mathbf{X}/(n-s)=Q_e/(n-s)$ and also $\hat{\sigma}_H^2=Q_H/r$ under $\mathbf{H}$ are unbiased estimators of σ^2.

Assume that $B'C=0$. Let U be an orthogonal transformation in $\mathbf{R}^n$ such that the first, $\ldots, r$th, $(s+1)$st, $\ldots, n$th rows of UB and the $(s+1)$st, $\ldots, n$th rows of UC are all equal to zero vectors. Using the notation $\tilde{\mathbf{X}}=U\mathbf{X}$, $\tilde{\eta}=UB\eta$, $\tilde{\zeta}=UC\zeta$, and $\tilde{\mathbf{W}}=U\mathbf{W}$, we obtain the **canonical form** $\tilde{\mathbf{X}}=\tilde{\eta}+\tilde{\zeta}+\tilde{\mathbf{W}}$ of the model (2). $\tilde{\mathbf{W}}$ is also a vector of independently and identically distributed normal random variables. The hypothesis $\mathbf{H}$ is expressed as $\tilde{\zeta}=\mathbf{0}$. In this model, we have $E(\tilde{X}_i)=0$ for $i=s+1, \ldots, n$, and moreover, $E(\tilde{X}_i)=0$ for $i=1, \ldots, r, s+1, \ldots, n$ if and only if $\mathbf{H}$ is true.

The least squares estimator $\mathbf{Y}$ of $A\xi$ is the †maximum likelihood estimator, and $\mathbf{X}-\mathbf{Y}$, $\mathbf{Y}-\mathbf{Z}$, $\mathbf{Z}$ are distributed independently according to the n-variate normal distributions $N(\mathbf{0}, \sigma^2(I-P_A))$, $N((P_A-P_B)C\zeta, \sigma^2(P_A-P_B))$, and $N(B\eta+P_BC\zeta, \sigma^2P_B)$, respectively. Hence Q/σ^2, Q_H/σ^2, and $\mathbf{X}'P_B\mathbf{X}/\sigma^2$ are distributed independently according to the †noncentral χ^2-distributions with $n-s$, r, and $s-r$ degrees of freedom and †noncentrality parameters 0, $\zeta'C'(P_A-P_B)C\zeta/\sigma^2$, and $(B\eta+P_BC\zeta)'(B\eta+P_BC\zeta)/\sigma^2$, respectively. The †likelihood ratio test of the hypothesis $\mathbf{H}$ has a critical region $\hat{\sigma}_H^2/\hat{\sigma}^2>c$, is a †uniformly most powerful invariant test with respect to the group of linear transformations leaving the hypothesis $\mathbf{H}$ invariant, and is the †most stringent test. This test is also uniformly most powerful among the tests whose †power function has a single variable $\zeta'C'(P_A-P_B)C\zeta/\sigma^2$. Furthermore, for $s-r=1$, this test is a †uniformly most powerful unbiased test. In the decomposition $\mathbf{X}'\mathbf{X}=\mathbf{X}'(P_A-P_B)\mathbf{X}+\mathbf{X}'P_B\mathbf{X}+\mathbf{X}'(I-P_A)\mathbf{X}=Q_H+Q_B+Q_e$ the terms Q_H and Q_e are called the sum of squares due to the hypothesis and due to the error, respectively. Such a process of decomposition is called the analysis of variance and its result is summarized in the analysis of variance table (Table 1).

I. The Likelihood Ratio Test

The likelihood ratio test is comparatively easy to construct. Let $L(x_1, \ldots, x_n; \theta)$ be the †likelihood function. Then

$$\Lambda(x_1, \ldots, x_n)=\frac{\sup_{\theta\in\omega_H} L(x_1, \ldots, x_n; \theta)}{\sup_{\theta\in\omega_H\cup\omega_A} L(x_1, \ldots, x_n; \theta)}$$

is called the **likelihood ratio**, and the test corresponding to the critical region $S=\{(x_1, \ldots, x_n) \mid \Lambda(x_1, \ldots, x_n)\leqslant c_\alpha\}$ is called the **likelihood ratio test**, where c_α is a positive constant determined by the level α. Let $\hat{\theta}_H(x_1, \ldots, x_n)$ and $\hat{\theta}_{H\vee A}(x_1, \ldots, x_n)$ be the †maximum likelihood estimators for θ in ω_H and in $\omega_H\cup\omega_A$, respectively; that is, $L(x; \hat{\theta}_H(x))=\sup_{\theta\in\omega_H} L(x; \theta)$ and $L(x; \hat{\theta}_{H\vee A}(x))=\sup_{\theta\in\omega_H\cup\omega_A} L(x; \theta)$. Then

Table 1

Factor	Sum of Squares	Degrees of Freedom	Mean Square	Ratio of Variances
H	$Q_H = \mathbf{X}'(P_A - P_B)\mathbf{X}$	r	$\hat{\sigma}_H^2 = Q_H/r$	$\hat{\sigma}_H^2/\hat{\sigma}^2$
B	$Q_B = \mathbf{X}'P_B\mathbf{X}$	$s-r$	$\hat{\sigma}_B^2/(s-r)$	
Error	$Q_e = \mathbf{X}'(I - P_A)\mathbf{X}$	$n-s$	$\hat{\sigma}^2 = Q_e/(n-s)$	
Total	$\mathbf{X}'\mathbf{X}$	n		

$$\Lambda(x) = \frac{L(x; \hat{\theta}_H)}{L(x; \hat{\theta}_{H \vee A})}.$$

The F-test for a linear hypothesis is a likelihood ratio test, and other examples of the likelihood ratio test are shown in Appendix A, Table 23. However, the likelihood ratio test does not necessarily have the desirable properties stated in the preceding sections.

J. Complete Classes

The set of critical regions of the type $\{x \mid T(x) > c\}$ for the problems $\mathbf{H}: \theta \leqslant \theta_0$ and $\mathbf{A}: \theta > \theta_0$, where the distribution family $\mathscr{P}$ of the statistic T is of †Pólya type 2 in the strict sense, and the set of regions of the type $\{x \mid c < T(x) < d\}$ for the problem $\mathbf{H}: \theta_1 \leqslant \theta \leqslant \theta_2$ and $\mathbf{A}: \theta < \theta_1$ or $\theta_2 < \theta$, where the distribution family $\mathscr{P}$ of the statistic T is of †Pólya type 3 in the strict sense, are examples of minimal complete classes [6]. It has been proved under a mild condition that the set of all tests with convex critical regions is essentially complete when the underlying distributions are of exponential type and the null hypothesis is simple.

Let $P_{\theta_i} \in \mathscr{P}$ $(i = 0, 1)$ and let $\mathfrak{B}_0$ be a σ-subalgebra of $\mathfrak{B}$. $\mathfrak{B}_0$ is sufficient for $\mathfrak{B}$ w.r.t. $\{P_{\theta_0}, P_{\theta_1}\}$ if and only if the class of all $\mathfrak{B}_0$-measurable test functions is essentially complete, i.e. iff for every critical region $B \in \mathfrak{B}$ there exists a $\mathfrak{B}_0$-measurable test function φ_0 such that $E_{\theta_0}(\varphi_0) \leqslant E_{\theta_0}(\chi_B)$ and $E_{\theta_1}(\varphi_0) \geqslant E_{\theta_1}(\chi_B)$. Assume that $\mathscr{P} = \{P_\theta \mid \theta \in \Theta\}$ is dominated; if for every $\mathfrak{B}$-measurable test function φ there exists a $\mathfrak{B}_0$-measurable test function ψ such that $E_\theta(\psi) = E_\theta(\varphi)$ for all $\theta \in \Theta$, then $\mathfrak{B}_0$ is sufficient for $\mathfrak{B}$ w.r.t. $\mathscr{P}$ [9, 10] (→ 398 Statistical Decision Functions D, 399 Statistical Estimation E).

K. Asymptotic Theory

Let $(\mathscr{X}_\nu, \mathfrak{B}_\nu, P_{\nu\theta})$ $(\nu = 1, 2, \ldots, n)$ be a sequence of probability spaces, where the parameter space Ω is common to all ν. Let $(\mathscr{X}^{(n)}, \mathfrak{B}^{(n)}, P_\theta^{(n)})$ be the direct product probability space of $(\mathscr{X}_\nu, \mathfrak{B}_\nu, P_{\nu\theta})$ for $\nu = 1, 2, \ldots, n$. For each sample space $(\mathscr{X}^{(n)}, \mathfrak{B}^{(n)}, P_\theta^{(n)})$, denote a test function for $\mathbf{H}: \theta \in \omega_H (\subset \Omega)$ and $\mathbf{A}: \theta \in \omega_A (\subset \Omega - \omega_H)$ by $\varphi_n(x_1, \ldots, x_n)$. A sequence $\{\varphi_n\}$ $(n = 1, 2, \ldots)$ is often called a test. For example, a likelihood ratio test is frequently understood as a sequence of tests $S_n = \{(x_1, \ldots, x_n) \mid \Lambda_n(x_1, \ldots, x_n) \leqslant \lambda_n\}$, where $\{\lambda_n\}$ is a sequence of constants and $\Lambda_n(x_1, \ldots, x_n)$ is the likelihood ratio defined by $(\mathscr{X}^{(n)}, \mathfrak{B}^{(n)}, P_\theta^{(n)})$ and ω_H. If a test $\{\varphi_n\}$ satisfies $E_\theta(\varphi_n\} \to 0$ $(\theta \in \omega_H)$ and $E_\theta(\varphi_n) \to 1$ $(\theta \in \omega_A)$ as $n \to \infty$, $\{\varphi_n\}$ is said to be a **consistent test**. If these convergences are uniform with respect to θ, $\{\varphi_n\}$ is said to be a **uniformly consistent test**. When a uniformly consistent test exists, ω_H and ω_A are said to be **finitely distinguishable**. Suppose that the observed values are identically distributed (that is, $(\mathscr{X}_\nu, \mathfrak{B}_\nu, P_{\nu\theta})$ is a copy of a probability space $(\mathscr{X}, \mathfrak{B}, P_\theta))$ and ω_H and ω_A are both compact with respect to the metric $\rho(\theta, \theta') = \sup_{B \in \mathfrak{B}} |P_\theta(B) - P_{\theta'}(B)|$. In this case, ω_H and ω_A are finitely distinguishable if $E_\theta(\varphi)$ is a continuous function of θ for any φ [4]. Kakutani's theorem (→ 398 Statistical Decision Functions) is regarded as a proposition concerning distinguishability when the null hypothesis and the alternative are both simple.

The following result about the †limit distribution of a likelihood ratio is due to H. Chernoff [3]: Let $\mathscr{X}^{(n)}$ be an n-space and Ω be an open subset of $\mathbf{R}^k$ containing the origin 0. Suppose that the observed random variables are independent and distributed according to a density $f(x, \theta)$; that is, the likelihood function $L(x; \theta)$ is $\prod_{i=1}^n f(x_i, \theta)$. Moreover, assume the following regularity conditions:
(1) $\log f(x, \theta)$ is three-times differentiable with respect to θ at every point of the closure of some neighborhood N of $\theta = 0$.
(2) There exist an integrable function F and a measurable function H such that (i) $|\partial f/\partial \theta_i| < F(x)$ for every $\theta \in N$; (ii) $|\partial^2 f/\partial \theta_i \partial \theta_j| < F(x)$ for every $\theta \in N$; (iii) $|\partial^3 \log f/\partial \theta_i \partial \theta_j \partial \theta_m| < H(x)$; and (iv) $\sup_\theta E_\theta(H(x)) < \infty$.
(3) For every $i, j = 1, 2, \ldots, k$, we have $J_\theta^{ij} = E_\theta[(\partial \log f/\partial \theta_i)(\partial \log f/\partial \theta_j)] < \infty$, and the matrix $J_\theta = (J_\theta^{ij})$ is positive definite for all $\theta \in N$.
Let $P(x_1, \ldots, x_n; \omega) = \sup_{\theta \in \omega} L(x_1, \ldots, x_n; \theta)$ for a subset ω of Ω. Consider testing a hypothesis $\theta \in \omega_H$ against an alternative $\theta \in \omega_A$, where 0 is an accumulation point of ω_H. If $\lambda^*(x_1, \ldots, x_n) = P(x_1, \ldots, x_n; \omega_H)/P(x_1, \ldots, x_n; \omega_A)$, λ^* plays essentially the same role as the likelihood ratio λ and hence can be used in its place. We call a subset C of Ω a cone if $\theta \in C$ implies $a\theta \in C$ for

any $a>0$. A subset ω of Ω is said to be approximated by C if ω satisfies $\inf_{x\in C}\|x-y\|=o(\|y\|)$ for all $y\in\omega$ and $\inf_{y\in\omega}\|x-y\|=o(\|x\|)$ for all $x\in C$ around the origin, where $\|x\|^2=\sum_{i=1}^n x_i^2$. Suppose that ω_H and ω_A are approximated by two cones C_H and C_A, respectively. Then, setting $z_n=\sqrt{n}\,J_0^{-1}A(x)$ with

$$A(x)=\left[\frac{1}{n}\sum_{\alpha=1}^n\frac{\partial\log f(x_\alpha,0)}{\partial\theta_1},\ \ldots,\frac{1}{n}\sum_{\alpha=1}^n\frac{\partial\log f(x_\alpha,0)}{\partial\theta_k}\right]',$$

the limit distribution of $-2\log\lambda^*$, when $\theta=0$, coincides with that of $\inf_{\theta\in C_H}(z_n-\theta)'J_0(z_n-\theta)-\inf_{\theta\in C_A}(z_n-\theta)'J_0(z_n-\theta)$. In particular, when $\Omega=\mathbf{R}^k$ and $\omega_H=\{(\theta_1^0,\ldots,\theta_s^0,\theta_{s+1},\ldots,\theta_k)\mid -\infty<\theta_i<\infty,\ i=s+1,\ldots,k\}$ and some regularity conditions are assumed, the limit distribution of $-2\log\lambda^*$ is the †χ^2-distribution with s degrees of freedom if the hypothesis is true.

The asymptotic behavior of the chi-square test of goodness of fit is also very important. Suppose that $(X_1,\ldots,X_k)$ has a multinomial distribution $n!(x_1!\ldots x_k!)^{-1}p_1^{x_1}\ldots p_k^{x_k}(\sum_{i=1}^k x_i=n,\ x_i\geqslant 0)$, and consider testing $\mathbf{H}:p_1=p_1^0,\ldots,p_k=p_k^0$. The **chi-square test of goodness of fit** has a critical region of the type $\{x\mid\chi^2(x_1,\ldots,x_k)\geqslant c\}$, where $\chi^2(x_1,\ldots,x_k)$ is $\sum_{i=1}^k((x_i-np_i^0)^2/np_i^0)$, i.e., the weighted sum of the squares of the differences between the value p_i^0 of p_i and the maximum likelihood estimator x_i/n of p_i. The limit distribution of $\chi^2(x_1,\ldots,x_k)$ when $p_i=p_i^0$, $i=1,2,\ldots,k$, is the chi-square distribution with $k-1$ degrees of freedom.

Suppose moreover that k functions $p_i(\theta)$ $(i=1,\ldots,k;s<k)$ of θ $(\in\mathbf{R}^s)$ are given and that the hypothesis to be tested is that the sample has been drawn from a population having a distribution determined by $\mathbf{H}:p_i=p_i(\theta)$ $(i=1,\ldots,k;s<k)$ for some value of θ. In this case the chi-square test of goodness of fit could be applied after replacing the parameter θ in $p_i=p_i(\theta)$ $(i=1,2,\ldots,n)$ by the solutions $\tilde{\theta}_n(x_1,\ldots,x_k)$ of the system of equations of the **modified minimum chi-square method**,

$$\sum_{i=1}^k\frac{x_i-np_i}{p_i}\frac{\partial p_i}{\partial\theta_j}=0$$

$(j=1,\ldots,s)$. Suppose that (1) $p_i(\theta)>c^2>0$ $(i=1,\ldots,k)$ and $\sum_{i=1}^k p_i(\theta)=1$; (2) $p_i(\theta)$ is twice continuously differentiable with respect to the coordinates of θ; and (3) the rank of the matrix $(\partial p_i/\partial\theta_j)$ is k. Then the system of equations above has a unique solution $\theta=\tilde{\theta}_n(x_1,\ldots,x_k)$, and $\tilde{\theta}_n$ converges in probability to θ_0 when $\theta=\theta_0$. The asymptotic distribution of $\chi^2(x)=\sum_{i=1}^k((x_i-np_i(\hat{\theta}_n))^2/np_i(\hat{\theta}_n))$ is the chi-square distribution with $n-s-1$ degrees of freedom [5]. For the test of goodness of fit, the empirical distribution function may also be used (→ 371 Robust and Nonparametric Methods).

A test of independence by contingency tables is one application of the chi-square test of goodness of fit. We suppose that n individuals are classified according to two categories A and B, where A has r ranks $A_1,A_2,\ldots,A_r$ and B has s ranks $B_1,B_2,\ldots,B_s$. Let $p_{i\cdot},p_{\cdot j},p_{ij}$ be the probabilities that the observed value of an individual belongs to A_i, B_j, $A_i\cap B_j$, respectively. Let $x_{i\cdot},x_{\cdot j},x_{ij}$ be the numbers of individuals belonging to A_i, B_j, and $A_i\cap B_j$, respectively. Table 2 is called a **contingency table**. To test the null hypothesis $\mathbf{H}$ that the divisions of A and B into their ranks are independent, that is, $\mathbf{H}:p_{ij}=p_{i\cdot}p_{\cdot j}$, the statistic $\chi^2=\sum_{i=1}^r\sum_{j=1}^s(x_{ij}-x_{i\cdot}x_{\cdot j}/n)^2/(x_{i\cdot}x_{\cdot j}/n)$ is applied. When $\mathbf{H}$ is true, χ^2 is asymptotically distributed according to the chi-square distribution with $(r-1)(s-1)$ degrees of freedom as $n\to\infty$.

Likelihood ratio tests and chi-square tests of goodness of fit are consistent tests under conditions stated in their respective descriptions. In general, there are many consistent tests for a problem. Therefore it is necessary to consider another criterion that has to be satisfied by the best test among consistent tests. Pitman's asymptotic relative efficiency is such a criterion. Other notions of efficiency have also been introduced.

A completely specified form of distribution is rather exceptional in applications. More often we encounter cases where distribution of the sample belongs in a large domain. Various tests independent of the functional form of distribution have been proposed, and the asymptotic theory plays an important role in those cases (→ 371 Robust and Nonparametric Methods).

The following concept of asymptotic efficiency is due to R. R. Bahadur [11]: Let $\{T_n\}$ be a sequence of real-valued statistics defined on $\mathscr{X}^{(n)}$. $\{T_n\}$ is said to be a standard sequence (for testing **H**) if the following three conditions are satisfied.

(I) There exists a continuous probability distribution function F such that for each $\theta\in\omega_H$, $\lim_{n\to\infty}P_\theta^{(n)}\{T_n<t\}=F(t)$ for every $t\in\mathbf{R}^1$.

Table 2. Contingency Table

	B_1	B_2	$\cdots$	B_s	Total
A_1	x_{11}	x_{12}	$\cdots$	x_{1s}	$x_{1\cdot}$
A_2	x_{21}	x_{22}	$\cdots$	x_{2s}	$x_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
A_r	x_{r1}	x_{r2}	$\cdots$	x_{rs}	$x_{r\cdot}$
Total	$x_{\cdot 1}$	$x_{\cdot 2}$	$\cdots$	$x_{\cdot s}$	n

(II) There exists a constant a, $0<a<\infty$, such that $\log\{1-F(t)\}=-(at^2/2)\{1+o(1)\}$ as $t\to\infty$.

(III) There exists a function $b(\theta)$ on $\Omega-\omega_H$ with $0<b(\theta)<\infty$ such that for each $\theta\in\Omega-\omega_H$,

$$\lim_{n\to\infty} P_\theta^{(n)}\{|(T_n/n^{1/2})-b(\theta)|>t\}=0$$

for every $t>0$.

Suppose that $\{T_n\}$ is a standard sequence. Then T_n has the asymptotic distribution F if **H** is satisfied, but otherwise $T_n\to\infty$ in probability. Consequently, large values of T_n are significant when T_n is regarded as a test statistic for **H**. Accordingly, for any given $x\in\mathscr{X}^{(n)}$, $1-F(T_n(x))$ is called the critical level in terms of T_n, and is regarded as a random variable defined on $\mathscr{X}^{(n)}$ [1]. It is convenient to describe the behavior of this random variable as $n\to\infty$ in terms of K_n, where $K_n(x)=-2\log[1-F(T_n(x))]$. Then for each $\theta\in\omega_H$, K_n is asymptotically distributed as a chi-square variable χ_2^2 with 2 degrees of freedom and for $\theta\in\Omega-\omega_H$, $K_n/n\to ab^2(\theta)$ in probability as $n\to\infty$. The asymptotic slope of the test based on $\{T_n\}$ (or simply the slope of $\{T_n\}$) is defined to be $c(\theta)=ab^2(\theta)$. Note that the statistic $K_n^{1/2}$ is equivalent to T_n in the following technical sense: (i) $\{K_n^{1/2}\}$ is a standard sequence; (ii) for each $\theta\in\Omega$, the slope of $\{K_n^{1/2}\}$ equals that of $\{T_n\}$; and (iii) for any given n and x, the critical level in terms of $K_n^{1/2}$ equals the critical level in terms of T_n. Since the critical level of $K_n^{1/2}$ is found by substituting $K_n^{1/2}$ into the function representing the upper tail of a fixed distribution independent of F, $\{K_n^{1/2}\}$ is a normalized version of $\{T_n\}$. Suppose that $\{T_n^{(1)}\}$ and $\{T_n^{(2)}\}$ are two standard sequences defined on $\mathscr{X}^{(n)}$, and let $F^{(i)}(x)$, a_i, and $b_i(\theta)$ be the functions and constants prescribed by conditions (I)–(III) for $i=1, 2$. Consider an arbitrary but fixed θ in $\Omega-\omega_H$, and suppose that x is distributed according to P_θ. The asymptotic efficiency of $\{T_n^{(1)}\}$ relative to $\{T_n^{(2)}\}$ is defined to be $\varphi_{12}(\theta)=c_1(\theta)/c_2(\theta)$, where $c_i(\theta)=a_ib_i^2(\theta)$ is the slope of $\{T_n^{(i)}\}$, $i=1, 2$. The asymptotic efficiency is called **Bahadur efficiency**.

Several comparisons of standard sequences are given in [11]. The relationship between Bahadur efficiency and Pitman efficiency for hypothesis-testing problems has also been studied. Under suitable conditions the two efficiencies coincide.

L. Sequential Tests

Let $X_1, X_2, \ldots$ be a given sequence of random variables. To test a hypothesis concerning the distributions of these variables (sample sizes are not predetermined), we observe first X_1, then X_2, etc. At each stage a decision is made on the basis of the previously obtained data whether the observation should be stopped and a judgment made on the acceptability of the hypothesis. Such a test is called a **sequential test**. Let $X_1, X_2, \ldots$ be independent and identically distributed by $f_\theta(x)$. For testing a simple hypothesis $\mathbf{H}:\theta=0$ against a simple alternative $\mathbf{A}:\theta=1$, we have the **sequential probability ratio test**: Let $G_n(x_1, x_2, \ldots, x_n)=\prod_{i=1}^n f_1(x_i)/\prod_{i=1}^n f_0(x_i)$, and preassign two constants $a_0<a_1$. After the observations of $X_1, \ldots, X_n$ are performed, the next random variable X_{n+1} is observed if $a_0<G_n(x_1, \ldots, x_n)<a_1$. Otherwise the experiment is stopped, and we accept **H** when $G_n\leqslant a_0$ or accept **A** when $a_1\leqslant G_n$. The constants a_0 and a_1 are determined by the desired probabilities α_1 and α_2 of errors of the first and second kind, respectively. It is known that, among the class of sequential tests in which the probability of error of the first (second) kind is not greater than α_0 (α_1), the sequential probability ratio test minimizes the expected number of observations when either **H** or **A** is true (→ 398 Statistical Decision Functions; 404 Statistical Quality Control).

References

[1] E. L. Lehmann, Testing statistical hypotheses, Wiley, 1959.
[2] A. Wald, Sequential analysis, Wiley, 1947.
[3] H. Chernoff, On the distribution of the likelihood ratio, Ann. Math. Statist., 25 (1954), 573–578.
[4] W. Hoeffding and J. Wolfowitz, Distinguishability of sets of distributions, Ann. Math. Statist., 29 (1958), 700–718.
[5] M. G. Kendall and A. Stuart, The advanced theory of statistics II, Griffin, 1961.
[6] S. Karlin, Decision theory for Pólya type distributions, Case of two actions I, Proc. 3rd Berkeley Symp. Math. Stat. Prob. I, Univ. of California Press (1956), 115–128.
[7] A. Birnbaum, Characterizations of complete classes of tests of some multiparametric hypotheses, with applications to likelihood ratio tests, Ann. Math. Statist., 26 (1955), 21–36.
[8] H. Cramér, Mathematical methods of statistics, Princeton Univ. Press, 1946.
[9] K. Takeuchi and M. Akahira, Characterizations of prediction sufficiency (adequacy) in terms of risk functions, Ann. Statist., 3 (1975), 1018–1024.
[10] J. Pfanzagl, A characterization of sufficiency by power functions, Metrika, 21 (1974), 197–199.

[11] R. R. Bahadur, Stochastic comparison of tests, Ann. Math. Statist., 31 (1960), 276–295.

401 (XVIII.2) Statistical Inference

A. The Statistical Model

Broadly and loosely speaking, the term "statistical inference" may imply any procedure for drawing conclusions from statistical data. But now it is usually understood more rigorously to mean those procedures based upon a †probabilistic model of the data to obtain conclusions concerning the unknown parameters of the population that represents the probabilistic model by viewing the observed data as a †random sample extracted from the population.

As the simplest example, suppose that, for some system, we have a number of observations from repeated measurements or experiments under a supposedly uniform condition. If we can assume that there are no systematic trends or tendencies involved, we can suppose that the variations among repeated observations are due to random causes and assume that the observed values $X_1, X_2, \ldots$ are independently and identically distributed random variables. Our purpose in making observations is to draw some information from the data, that is, to make some judgment on an unknown system quantity θ, which together with some other quantity (quantities) η characterizing the measurement or the experiment, determines the distribution of the X_i. We assume that the distribution has a density function $f(x;\theta,\eta)$. This amounts to assuming that the observed values $X_1, X_2, \ldots$ are a sample randomly drawn from a hypothetical population of the results of the measurements or experiments supposedly continued indefinitely. Then the problem of statistical inference is one of making some judgment based on the random sample. The set of hypotheses postulating the distribution of the observed values is called the probabilistic model of the observations, and the problem of determining a model in a specific situation is called that of **specification**.

B. Bayesian and Non-Bayesian Approaches

There are two different ways to make inferences on the population parameters: the **Bayesian approach** and the **non-Bayesian approach**.

In the Bayesian approach it is assumed that we have some probability density $\pi(\theta,\eta)$ for the parameters θ and η. Then, given the observations $X_1=x_1$, $X_2=x_2, \ldots$, the conditional probability density for θ and η is given by

$$\tilde{p}(\theta,\eta \mid x_1, x_2, \ldots) = \frac{\prod f(x_i;\theta,\eta)\pi(\theta,\eta)}{\int\int \prod f(x_i;\theta,\eta)\pi(\theta,\eta)\,d\theta\,d\eta}.$$

π is called the **prior density** and $\tilde{p}$ the **posterior density** for the parameters. Then all the information obtained from the sample is considered to be contained in the **posterior distribution** with the density $\tilde{p}(\theta,\eta)$, and conclusions on the parameters can be drawn from it.

The prior density $\pi(\theta,\eta)$ does not necessarily represent a frequency function of a population of which the parameters are a random sample, but in most cases treated by the Bayesian approach it is considered to be a summary of the statistician's judgment over relative possibilities of the different values of the parameters based on all the information obtained before the observations are made. Bayesians claim that it is always possible to determine such a **prior distribution** in a coherent way, specifying the **subjective probability**, representing a person's judgment under uncertainty, as opposed to the **objective probability**, representing the relative frequencies in a population. L. J. Savage [7] succeeded in developing a formal mathematical theory of the subjective probability from a set of postulates about the consistent behavior of a person under uncertainty.

The non-Bayesian statisticians, however, do not accept the Bayesians' viewpoint and insist that statistical inference should be free from any subjective judgment and be based solely on the objective properties of the sample derived from the assumed model. The theory developed by R. A. Fisher, J. Neyman, and E. S. Pearson, and others is based on the non-Bayesian approach.

C. Problems of Non-Bayesian Inference

The most commonly used forms of statistical inference are **point estimation**, used when we want to get a value as the estimate for the parameter; **interval estimation**, when we want to get an interval that contains the true value of the parameter with a probability not smaller than the preassigned level; and **hypothesis testing**, when it is required to determine whether or not some hypothesis about the parameter values is wrong (→ 399 Statistical Estimation, 400 Statistical Hypothesis Testing).

In any type of statistical inference, the problem can be abstractly formulated by determining a procedure that defines a rule, based on the sample observed, for choosing an element from the set of possible conclusions.

Such a procedure is evaluated by the probabilistic properties derived under different values of the parameter from the distribution of the sample, and it is usually required to satisfy some type of validity criteria (such as unbiasedness of an estimator, size of a test, etc.), and among those satisfying them, one which is considered to be best according to some optimality criterion (such as minimum variance or most-powerfulness) is looked for. But in the sense of objective probability, the probabilistic property of a procedure is relevant only for the frequencies in repeated trials when the same procedure is applied to a sequence of samples obtained from the population and has no direct implications for the conclusion obtained by applying the procedure to a specific sample we have in hand. For this reason Neyman argued that in statistical inference there is really no such thing as inductive inference but only inductive behavior. Fisher disagreed strongly with this argument and emphasized that statistical analysis is induction and that its purpose is to allow us to draw the proper conclusions from a particular sample and that the probabilistic properties of the procedure should and could have relevance for a particular conclusion obtained from a specific sample, provided that all the information contained in the sample is used. The arguments between Fisher and Neyman led to a heated controversy between their followers that is still not completely settled. Fisher's arguments lead to the **principle of sufficiency** and the **principle of conditionality**. The principle of sufficiency dictates that all inferences should be based on a sufficient statistic if there is one, and the principle of conditionality requires that any inference should be based on the conditional distribution given the **ancillary statistic**, i.e., a statistic whose distribution is independent of the parameter, if there is such a statistic. These two principles are accepted by many statisticians who do not follow all of Fisher's arguments, though the principle of conditionality sometimes leads to difficulties due to nonuniqueness of the ancillary.

D. Specification Problem

It is often difficult and sometimes impossible to have an exactly correct model for the data, and we must be satisfied with a model that gives a sufficiently close approximation and is mathematically tractable as well. It may also happen, however, that a model first specified may be far from reality and could lead to erroneous conclusions if relied on blindly. Here, the problem of **model selection** arises (→ 403 Statistical Models), i.e., choosing the best of various possible models.

We may also seek procedures that are little affected by the departure of the distribution of the data from the assumed or some other model that satisfies the condition of validity without any assumption about the exact shape of the distributions (→ 371 Robust and Nonparametric Methods). Generally, the problem of determining the model or specification should not be dealt with by mathematical methods alone, and it should be considered by taking into account the properties and nature of the subject under consideration and also the process of measurement or experimentation.

E. History

The first appearance of statistical inference as a method of grasping numerical characteristics of a collective was seen in the study by J. Graunt (1662) of the number of people who died in London. W. Petty applied Graunt's method further to the comparison of communities in his *Political arithmetic* (1690). J. P. Süssmilch, a member of the Graunt school, perceived the regularity in mass observations and stressed the statistical importance of this regularity. The development of the theory of probability inevitably affected the theory of statistical inference. The method of T. Bayes was the first procedure of statistical inference in the current meaning of this expression. We now have a theorem bearing his name (the Bayes theorem), which is stated in current language as follows: If we know the probability $P_C(E)$ that a cause C produces an effect E and if the prior (or †a priori) probability $P(C)$ of the existence of the cause C is also known, then the posterior (or †a posteriori) probability of C, given an effect E, is equal to

$$P_E(C)=\frac{P(C)P_C(E)}{\sum_C P(C)P_C(E)}$$

(→ 342 Probability Theory F). This theorem, easily extendable to the continuous case, suggests the following inference procedure: If we are informed that an effect E has taken place, then we calculate the probabilities $P_E(C)$ for every cause C, compare them, and infer that the C^* with $P_E(C^*)=\max_C P_E(C)$ is the **most probable cause** of E.

Both P. S. Laplace and C. F. Gauss discussed the theory of estimation of parameters (→ 399 Statistical Estimation) as an application of the Bayes theorem. In his research, Laplace considered a monotone function $W(|t-\theta|)$, and $W=|t-\theta|$ in particular, of the distance $|t-\theta|$ between a parameter value θ and its

estimate t as a measure of significance of the error of the estimate t. Gauss, following Laplace, used this weight function $W(|t-\theta|)$ of error, and going beyond Laplace, realized that it would be mathematically fruitful to put $W(|t-\theta|)=(t-\theta)^2$. Such considerations led him to the study of the †least squares method, in which the terminology and notation he devised are still in use. He also developed the theory of errors and recognized the importance of the normal distribution and found that the least square estimate is equal to the most probable value if the errors are normally distributed.

F. Galton, a biologist, revealed the usefulness of statistical methods in biological research and explored what we call †regression analysis (→ 403 Statistical Models) by introducing the concepts of regression line and †correlation coefficient. His research on regression analysis originated from the study of the correlation between characteristics of parents and children, but he failed to realize the difference between †population characteristics and †sample characteristics.

Following Galton, K. Pearson developed the theory of regression and correlation, with which he succeeded in establishing the basis of biometrics (→ 40 Biometrics). He arrived at the concept of population in statistics: A statistical **population** is a collective consisting of observable individuals, while a **sample** is a set of individuals drawn out of the population and containing something telling us about characteristics of the population. Thus statistical research is regarded as investigation that focuses not on a sample as such but on a population from which the sample has been drawn. This consideration raised the problem of the **goodness-of-fit test** (→ 400 Statistical Hypothesis Testing), that is, the problem of knowing whether a sample is likely to have been drawn from a population whose distribution was determined by theoretical considerations. K. Pearson characterized some population distributions occurring in practice by a differential equation, and classified them into several types. Using this classification, he discussed goodness-of-fit tests and developed the χ^2-distribution (†chi-square distribution) in relation to the problem of testing hypotheses.

Statisticians in the time of K. Pearson thought of a population as a collective having infinitely many individuals (i.e., an **infinite population**), which led to the idea that the larger the **size** of a sample (i.e., the number of individuals in the sample), the more precisely could the sample give information about the population. They carried out inferences, including the testing of hypotheses (→ 400 Statistical Hypothesis Testing), by approximate methods, which later came to be termed **large sample theory**. Suppose, for instance, that $\{X_1, \ldots, X_n\}$ is an †independent sample of size n from a normal population $N(\mu, \sigma^2)$. The random variable $Z=\sqrt{n}(\bar{X}-\mu_0)/\sigma$ with $\bar{X}=\sum_i X_i/n$ is distributed according to $N(0,1)$ when $\mu=\mu_0$. Therefore, if the size n is sufficiently large ($n\to\infty$), we estimate σ by $\hat{\sigma}=\{\sum_i(X_i-\bar{X})^2/(n-1)\}^{1/2}$ and deal with the random variable $T=\sqrt{n}(\bar{X}-\mu_0)/\hat{\sigma}$ obtained by inserting $\hat{\sigma}$ in place of σ in the expression for Z, as if T itself were distributed according to $N(0,1)$.

F. Development in the 20th Century

W. S. Gosset, writing under the pen name "Student," reported in 1908 the discovery of the exact distribution of T and thereby opened the new epoch of **exact sampling theory** (→ 374 Sampling Distributions). This work of Student made it possible to perform statistical inference by means of **small samples** and consequently changed statistical research from the study of collectives to that of uncertain phenomena; in other words, the concept of population was once again related to a †probability space with a †probability distribution (i.e., a **population distribution**) containing unknown **parameters**. Thus it began to be emphasized that a sample has to be drawn **at random** (i.e., a **random sample**) from the population if we are to make an inference about a parameter based on the sample.

Fisher presented a complete derivation, using the multiple integration method, of the †t-distribution (the sampling distribution of T). In addition, Fisher introduced the concepts of †null hypothesis and significance test, which were the starting points for later progress in the theory of hypothesis testing. He also added the concepts of †consistency, †efficiency, and †sufficiency to the list of possible properties of †estimators, and he studied the connection between the information contained in a sample and the accuracy of an estimator, which led to the idea of amount of information. Fisher also proposed the †maximum likelihood estimator, which is formally equivalent to the most probable value, but he renamed it and gave it a foundation completely independent of any prior information and showed that it leads to the at least asymptotically efficient estimator.

Fisher made efforts to obtain a distribution of the parameter directly from the sample observation, hence independently of the concept of prior probability. He sought in this way to be released from the weakness of the Bayes method. For this purpose he introduced

the concept of **fiducial distribution**, which was the subject of bitter controversy in the period that followed. As an example of a fiducial distribution, we consider here the †Behrens-Fisher problem: Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be samples drawn independently from the populations $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively, where the parameters μ_1, μ_2, σ_1, and σ_2 are all unknown. The problem raised is to test the hypothesis $\mu_1 = \mu_2$ or to estimate $\delta = \mu_1 - \mu_2$ by an interval. To solve this problem, we put

$$\bar{X} = \sum_i X_i/m, \quad \bar{Y} = \sum_j Y_j/n,$$

$$S_1^2 = \sum_i (X_i - \bar{X})^2/(m-1),$$

$$S_2^2 = \sum_j (Y_j - \bar{Y})^2/(n-1),$$

and learn that $T_1 = \sqrt{m}(\bar{X} - \mu_1)/S_1$ and $T_2 = \sqrt{n}(\bar{Y} - \mu_2)/S_2$ are mutually independent and distributed according to the t-distribution with degrees of freedom $m-1$ and $n-1$, respectively. From this fact Fisher reasoned as follows: Given observed values $\bar{x}, \bar{y}, s_1, s_2$ of the variables $\bar{X}, \bar{Y}, S_1, S_2$, the distributions of the parameters μ_1 and μ_2 are induced from the distributions of T_1 and T_2 by means of transformations

$$\mu_1 = \bar{x} - \frac{T_1 s_1}{\sqrt{m}}, \quad \mu_2 = \bar{y} - \frac{T_2 s_2}{\sqrt{n}}.$$

Consequently the distribution of $\delta = \bar{x} - \bar{y} - (T_1 s_1/\sqrt{m} - T_2 s_2/\sqrt{n})$ is obtained. These distributions are called the fiducial distributions of the parameters μ_1, μ_2, and δ. The interval $|\delta - (\bar{x} - \bar{y})| < c$ of δ deduced from the fiducial distribution of δ is called a **fiducial interval** of δ.

Neyman and E. S. Pearson developed a mathematical theory of testing hypotheses, in which they deliberately defined a family of population distributions admissible for formal treatment and considered alternative hypotheses within the family. They proposed to relate a test to its †power function, on the basis of which the test would be judged. Their ideas brought mathematical clarity to the theory of inference. Furthermore, concerning interval estimation, Neyman devised an alternative to the fiducial interval, the †confidence interval, which has full mathematical justification. Unfortunately it was later found that the confidence interval, fiducial interval, and the Bayes posterior interval based on the posterior distribution often gave distinctly different results to the same problem, which became a source of controversy among different schools of thought.

Since the publication of A. Wald's theory of statistical decision functions (→ 398 Statistical Decision Functions) in 1939, there has been a steady increase in its importance. In this theory the totality $\mathscr{D}$ of available statistical procedures, which is considered implicitly in the Neyman-Pearson theory, is put forth explicitly as a set and defined as the space of decision functions. Wald also defined the †risk function of a statistical decision procedure and used it as a basis for judging procedures. In addition, he employed the concept of prior probability and the Bayes procedure for the purpose of proving the †complete class theorem. Wald's idea of bringing the concept of prior probability back into statistical theory carried a great deal of weight, and much literature has now been accumulated on this subject. Prior probability as a technique in statistics was abandoned after Fisher's introduction of the maximum likelihood method independent of prior probability and Neyman's assertion that a probability distribution on the †parameter space made no sense. In addition, Wald linked statistical inference to games (→ 173 Game Theory) and introduced the †minimax principle into statistics. The decision-theoretic setup also enabled him to develop a theory of sequential analysis by comparing the cost of sampling with the risk of erroneous decisions (→ 400 Statistical Hypothesis Testing).

After the publication of Savage's book in 1954, there was a revival of the Bayesian approach, i.e., one based on the concept of subjective probability, and now the group of those statisticians who accept the Bayesian approach are called Bayesians or neo-Bayesians.

G. Applications

Methods of statistical inference are applied in many fields where statistical data are used for scientific, engineering, medical, or managerial purposes. Methods of producing data that are appropriate for statistical inference have also been developed. R. A. Fisher developed the method of statistical †design of experiments (→ 102 Design of Experiments) that when it is impossible or impractical to eliminate completely experimental errors or variabilities, provides the procedures to obtain such data. These data, though subject to random errors, are susceptible to rigorous statistical inference. For this purpose Fisher introduced the principles of †randomization, †local control, and †replication in the design of experiments. W. A. Shewhart defined the †state of statistical control in mass-production processes, where the variabilities of the products can be considered to be due to chance causes alone and hence are statistically analyzable.

Applying the idea of statistical inference to this situation, Shewhart established the method of statistical quality control (→ 404 Statistical Quality Control). Neyman introduced the method of †random sampling into statistical surveys and developed the theory of estimation and allocation based on the theory of statistical inference (→ 373 Sample Survey).

In many applied fields there exist systems of statistical methods which have been developed specifically for the respective fields, and although all of them are based essentially on the same general principles of statistical inference, each has its own special techniques and procedures. Specific names have been invented, such as biometrics (→ 40 Biometrics) econometrics (→ 128 Econometrics), psychometrics (→ 346 Psychometrics), technometrics, sociometrics, etc.

References

[1] R. A. Fisher, Statistical methods for research workers, Oliver & Boyd, 1925; eleventh edition, 1950.
[2] R. A. Fisher, The design of experiments, Oliver & Boyd, 1935; fourth edition, 1947.
[3] R. A. Fisher, Contributions to mathematical statistics, Wiley, 1950.
[4] R. A. Fisher Statistical methods and scientific inference, Oliver & Boyd, 1956.
[5] J. Neyman and E. S. Pearson, Contributions to the theory of testing statistical hypotheses I, II, Statist. Res. Mem. London Univ., 1 (1936), 1–37; 2 (1938), 25–57.
[6] A. Wald, Statistical decision functions, Wiley, 1950 (Chelsea, 1971).
[7] L. J. Savage, The foundation of statistical inference, Methuen, 1962.
[8] C. R. Rao, Linear statistical inference and its applications, Wiley, 1965.
[9] D. A. Fraser, The structure of inference, Wiley, 1968.
[10] D. V. Lindley, Introduction to probability and statistics from a Bayesian viewpoint I, II, Cambridge Univ. Press, 1965.

402 (XX.19) Statistical Mechanics

A. General Remarks

One cubic centimeter of water contains about 3×10^{22} water molecules. A macroscopic system of matter thus consists of an enormous number of particles incessantly moving in accordance with the laws of dynamics (→ 271 Mechanics; 351 Quantum Mechanics). Dynamical description of such microscopic motion in full detail is impossible and even meaningless. A physical process in thermodynamics or †hydrodynamics is described in terms of a relatively small number of macroscopic variables, such as temperature, pressure, and a velocity field. Such a process shows a remarkable simplicity which is a statistical result of the molecular chaos. This is the reason why statistical mechanics is needed as a theoretical model to unify microscopic dynamics and †probability theory. Thus **statistical mechanics** aims at deriving physical laws in the macroscopic world from the atomistic structures of the microscopic world on the basis of microscopic dynamical laws and probabilistic laws. Its function is twofold. First, statistical mechanics should give microscopic proofs of the macroscopic laws of physics, such as those of thermodynamics or the laws of macroscopic electromagnetism. Second, it should also provide us with detailed knowledge of physical properties of a given material system once its microscopic structure is known. In this sense, statistical mechanics is an essential basis of the modern science of materials.

Strictly speaking, the dynamics of the microscopic world obeys †quantum mechanics. However, even before the birth of quantum mechanics, statistical mechanics had progressed on the basis of classical mechanics. This stage of statistical mechanics is often called **classical statistical mechanics**, in contrast to **quantum statistical mechanics** based on quantum mechanics. Statistical mechanics has a fully developed formalism to apply to physical systems in thermal equilibrium. This is sometimes called **statistical thermodynamics** or **equilibrium statistical mechanics**. Until the 1950s the term "statistical mechanics" had often been used in this narrow sense. In a wider sense it is concerned with systems in more general states, for instance, in nonequilibrium states. In the modern literature, a general statistical mechanical theory of nonequilibrium systems is often referred to as the statistical mechanics of **irreversible processes**.

B. History

The early stage of statistical mechanics can be traced back to the **kinetic theory of gases**, which started in the 18th century. In dilute gases, gas molecules fly freely through the whole volume of the vessel and collide only from time to time. In thermal equilibrium, the average energy of each molecule is determined by the temperature of the gas; namely

$m\overline{v_x^2}/2 = m\overline{v_y^2}/2 = m\overline{v_z^2}/2 = kT/2,$

where (v_x, v_y, v_z) is the velocity, an overbar means the average, m is the mass of a molecule, T is the absolute temperature, and k is the **Boltzmann constant** ($=1.38\times 10^{-16}$ erg $\cdot$ deg^{-1}). The velocity of each molecule is only probabilistic, and a †distribution function $f(v_x, v_y, v_z)$ is defined as the †probability density that the velocity of a given molecule is found to be in the neighborhood of (v_x, v_y, v_z). In a dilute gas, this is given by

$$f(v_x, v_y, v_z) = C\exp\left[-\frac{m}{2}(v_x^2+v_y^2+v_z^2)/kT\right], \quad (1)$$

the **Maxwell-Boltzmann distribution law**.

L. Boltzmann viewed the velocity distribution function as changing in time as a result of molecular collisions and gave an equation of the form

$$\frac{\partial f}{\partial t} = A[f] + \Gamma[f], \quad (2)$$

where $A[f]$ is the change of the distribution function f by acceleration due to the presence of external forces and $\Gamma[f]$ is the change caused by molecular collisions. $\Gamma[f]$ is an integral which is nonlinear in f. This type of equation is called a **Boltzmann equation** [1, 2].

Boltzmann introduced the **H-function** by the definition

$$H = \iiint f\log f\, dv_x dv_y dv_z \quad (3)$$

and proved on the basis of equation (2) that $dH/dt \leqslant 0$. This theorem is known as the **H-theorem** [1–4]. The equilibrium distribution (1) is therefore obtained from equation (2) as the solution that makes H a minimum. In fact the H-function is related to the **entropy** S by

$$S = -kH. \quad (4)$$

Boltzmann further showed (1877) that the distribution function of a system in thermal equilibrium can be obtained on more general grounds without relying on a kinetic equation of the type (2) and that the statistical mechanics of systems in equilibrium can thus be constructed on a basis much more general than that given by a kinetic theory. It was W. Gibbs, however, who clearly established (1902) the complete framework of statistical thermodynamics, although he had to confine himself to classical statistical mechanics [5].

C. The Ergodic Hypothesis

For a given dynamical system with n †degrees of freedom, the **phase space** is defined as a $2n$-dimensional space with †generalized coordinates $q_1, \ldots, q_n$ and †generalized momenta $p_1, \ldots, p_n$. Dynamical states of the system constitute a set of points in this space. At a given time, the state of the system is represented by a point P in the phase space, and hence the motion of the system is represented by the motion of P. If the system is conservative, its energy function is constant. Let $\mathscr{H}$ be the †Hamiltonian function. Then the motion of P is confined to an **energy surface** defined by the condition $\mathscr{H} = E =$ constant. Measure on an energy surface is defined as the limit of the volume element lying between two neighboring energy surfaces corresponding to the energies E and $E + dE$. The motions of P form a †topological group that makes this measure invariant (†Liouville's theorem).

A dynamical quantity $A(p, q)$ of the system changes its value as the phase point P moves on the energy surface. The time average $\bar{A}$ of A is identified with the value of A observed in the equilibrium state of the system, namely, the average of A with respect to the invariant measure. Boltzmann justified this assumption by the following reasoning. If the energy surface has a finite measure and the trajectory of P does not make a closed curve on the energy surface, it can be assumed that the trajectory will move around practically everywhere on the surface. Mathematically formulated, the only measurable subset of the surface that has a nonzero measure and is invariant under the motion is the whole surface. This assumption is the **ergodic hypothesis** [6–9]. The long-time average of A will then equal the average of A over the entire energy surface with weight function equal to the measure previously introduced. The latter average is called the **phase average** and is denoted by $\langle A\rangle$. Boltzmann thus asserted that

$$\bar{A} = \langle A\rangle. \quad (5)$$

Efforts of mathematicians to study the ergodic hypothesis created an important branch of mathematics called ergodic theory (→ 136 Ergodic Theory).

D. Ensembles in Classical Statistical Mechanics

Once we admit the ergodic hypothesis, or more specifically the assumption (5), the calculation of the observed value of a physical quantity A for a system in equilibrium is reduced to finding the phase average of A on an energy surface. The task of statistical mechanics of systems in equilibrium is thus reduced essentially to calculating phase averages and establishing relationships between them [10–13].

For a set (called an **ensemble** in this case) of

identical systems with the same energy, we consider the phase average for the †probability space with the measure mentioned in Section C on the energy surface corresponding to the given energy value. Gibbs called a probability space of this kind a **microcanonical ensemble.** An average in this probability space is defined by

$$\langle A\rangle=\int_{\mathscr{H}=E}\frac{A\,dS}{|\mathrm{grad}\,\mathscr{H}|}\Big/\int_{\mathscr{H}=E}\frac{dS}{|\mathrm{grad}\,\mathscr{H}|},\qquad(6)$$

where $\mathscr{H}$ is the Hamiltonian function, grad $\mathscr{H}$ is its gradient in the $2n$-dimensional phase space, and the integration is carried over the energy surface with dS as surface element.

When the observed system is in mechanical contact with a heat reservoir, the composite system consisting of the system and the heat reservoir is regarded as an isolated system with constant energy. Then an ensemble of the composite systems is treated as a microcanonical ensemble. It is more convenient and more physical, however, to consider the heat reservoir simply as providing an environment characterized by its temperature T, and to concentrate only on the system in which we are interested. Then the system is no longer isolated and exchanges energy with its environment. Since the energy of the system is no longer constant, the system will be found in any part of the phase space with a certain probability. To find the probability distribution for an ensemble of this system is a problem of asymptotic evaluation which is solved on the basis of the ergodic hypothesis and the fact that a heat reservoir has an extremely large number of degrees of freedom. This asymptotic evaluation is traditionally done with the help of †Stirling's formula or by using the Fowler-Darwin method [10], but it is essentially based on the †central limit theorem [11].

The probability space of this kind of ensemble of systems in contact with heat reservoirs was called a **canonical ensemble** by Gibbs [5]. If $d\Gamma$ is a volume element of the phase space of the system, the probability of finding a system arbitrarily chosen from the ensemble in a volume element $d\Gamma$ is given by

$$\Pr(d\Gamma)=C\exp(-\mathscr{H}/kT)\,d\Gamma.\qquad(7)$$

Accordingly, the average of a dynamical quantity A is given by

$$\langle A\rangle=\int Ae^{-\mathscr{H}/kT}\,d\Gamma\Big/\int e^{-\mathscr{H}/kT}\,d\Gamma.\qquad(8)$$

For example, the average energy is

$$\langle\mathscr{H}\rangle\equiv E=\int\mathscr{H}e^{-\mathscr{H}/kT}\,d\Gamma\Big/\int e^{-\mathscr{H}/kT}\,d\Gamma.\qquad(9)$$

By a traditional convention we introduce the parameter

$$\beta=1/kT$$

and write (9) as

$$E=-\partial\log Z(\beta)/\partial\beta,$$

where $Z(\beta)$ is called the **partition function** or the **sum over states** and is given for a system composed of N identical particles by

$$Z(\beta)=\int e^{-\beta\mathscr{H}}\,d\Gamma/N!.\qquad(10)$$

If an exchange of particles with the environment takes place in addition to an exchange of energy, the probability of finding a system with particle number N in the volume element $d\Gamma$ is given by

$$C\exp(-\beta\mathscr{H}+\beta\mu N)\,d\Gamma/N!,$$

where μ is a real parameter called the **chemical potential**; this characterizes the environment with regard to the exchange of particles. This ensemble is called the **grand canonical ensemble**. The average of a dynamical quantity A is then given by

$$\langle A\rangle=\sum_N N!^{-1}\int A_N e^{-\beta\mathscr{H}(N)+\beta\mu N}\,d\Gamma/\Xi(\beta,\mu),\qquad(11)$$

where the dependence of A and $\mathscr{H}$ on N is now explicitly written, and where

$$\Xi(\beta,\mu)=\sum_N N!^{-1}\int e^{-\beta\mathscr{H}(N)+\beta\mu N}\,d\Gamma\qquad(12)$$

is called the **grand partition function.**

E. Ensembles in Quantum Statistical Mechanics

The quantum counterpart of the classical ergodic hypothesis is that to each of these quantum states an equal probability weight should be assigned [10]. A **microcanonical ensemble** is then defined by this **principle of equal weight**, which yields in turn

$$\langle A\rangle=\sum_l A_l\Big/\sum_l 1\qquad(13)$$

instead of (6). Here the index l refers to the quantum states lying in the interval ΔE, and A_l is the quantum-mechanical expectation of a dynamical variable A in the quantum state l. A canonical ensemble is now defined by assigning

$$P_j=e^{-\beta E_j}\Big/\sum_j e^{-\beta E_j}\qquad(14)$$

to the jth quantum state as the probability that the system will be found in that state. The

expectation value of A must be given by

$$\langle A\rangle=\sum_j A_j e^{-\beta E_j}/\sum_j e^{-\beta E_j}=\operatorname{tr} A e^{-\beta H}/\operatorname{tr} e^{-\beta H}, \tag{15}$$

where H is the Hamiltonian. The partition function is defined by

$$Z=\sum_j e^{-\beta E_j}=\operatorname{tr} e^{-\beta H}, \tag{16}$$

corresponding to (10).

For a system consisting of identical particles, quantum mechanics requires a particular symmetry of its †wave function; namely, the wave function must be even or odd with respect to permutation of any two particles according as the particles are bosons or fermions. This symmetry requirement is peculiar to quantum mechanics. Thus, even for an ideal gas consisting of noninteracting particles, quantum statistics leads to results characteristically different from those of classical statistical mechanics. This difference becomes more significant when the particle mass is smaller, the density is larger, and the temperature is lower. Quantum effects of this kind are seen in metallic electrons, in liquid helium, in an assembly of photons or phonons, and in high-density stars. The statistical laws obeyed by bosons are called **Bose statistics**, and those obeyed by fermions, **Fermi statistics**.

The expectation value of A in the **grand canonical ensemble** is given in quantum statistics by

$$\langle A\rangle=\Xi(\beta,\mu)^{-1}\operatorname{tr}(A e^{-\beta H+\beta\mu N}), \tag{17}$$

where H is the †second-quantized Hamiltonian, N is the number operator, the trace tr is taken on the (nonrelativistic) †Fock space (symmetric or antisymmetric according to Bose or Fermi statistics), and $\Xi(\beta,\mu)$ is the **grand partition function** given by

$$\Xi(\beta,\mu)=\operatorname{tr} e^{-\beta H+\beta\mu N}. \tag{18}$$

F. Many-Body Problems in Statistical Mechanics

Since statistical mechanics is primarily concerned with systems with large numbers of particles, problems in statistical mechanics are essentially many-body problems. In practice, however, there are some cases where extreme idealization is possible, as in ideal gases, where the interaction between gas molecules is ignored. In some cases we can proceed by successive approximation, taking the particle interactions as perturbations. Such perturbational treatments are, however, entirely useless for some problems, such as phase transitions, of which an example is the condensation of gases into liquid states, where the interaction of particles plays a critical role. Such problems are clearly many-body problems. There are a number of important and interesting problems in this category, for example, transitions between ferromagnetic and paramagnetic states and those between the superconducting and normal states of metals. Transition from a high-temperature phase to a low-temperature phase is generally regarded as a consequence of the appearance of a certain type of order in thermal motion. This kind of phase change is called an **order-disorder transition** [14–16].

G. Thermodynamic Limit and Characterization of Equilibrium States

Although an actual system is finitely extended, the enormous sizes of the usual macroscopic systems in comparison to the sizes of their constituent particles justifies the idealization to infinitely extended systems. At the same time, there are several mathematical advantages in considering infinitely extended systems, such as the absence of walls (replaced by the boundary condition at infinity, should it be relevant), appearance of phase transitions as mathematical discontinuities rather than mathematically smooth though quantitatively sudden changes, and mathematically clear-cut occurrence of broken symmetries.

Equilibrium states of infinitely extended systems are usually obtained by taking the limit of the equilibrium states of systems in a finite volume V as both V and the number of particles N tend to ∞ with the density $\rho=N/V$ fixed; this is called the **thermodynamic limit**.

It is sometimes possible to formulate the dynamics of infinitely extended systems directly and to characterize their equilibrium states, which more or less coincide with the thermodynamic limit of equilibrium states of finitely extended systems [17–21]. The simplest and most fully investigated case of **lattice spin systems** is explained below in detail [17]. Since classical systems can be viewed as special cases of quantum systems, we start with the latter. To be definite, we take a ν-dimensional cubic lattice $\mathbf{Z}^\nu$ with a lattice site $n=(n_1,\ldots,n_\nu)$ specified by its integer coordinates n_j. (In the lattice case, the thermodynamic limit is simply the limit as $V\to\infty$.)

The C^*-algebra $\mathfrak{A}$ of observables is generated by the subalgebra $\mathfrak{A}_n$ at each lattice site n, which is assumed to be the algebra of all $d\times d$ matrices (for example, linear combinations of †Pauli spin matrices $\boldsymbol{\sigma}^{(n)}=(\sigma_x^{(n)},\sigma_y^{(n)},\sigma_z^{(n)})$ and the identity for $d=2$) and to commute with operators at other lattice sites. The group of lattice translations $n\to n+a$ is represented by

automorphisms γ_a of $\mathfrak{A}$, satisfying $\gamma_a\mathfrak{A}_n=\mathfrak{A}_{n+a}$ $(\gamma_a\boldsymbol{\sigma}^{(n)}=\boldsymbol{\sigma}^{(n+a)})$. For any subset Λ of $\mathbf{Z}^\nu$, $\mathfrak{A}(\Lambda)$ denotes the C^*-subalgebra of $\mathfrak{A}$ generated by $\mathfrak{A}_n$, $n\in\Lambda$.

A model is specified by giving a **potential** Φ which assigns to each finite nonempty subset I of $\mathbf{Z}^\nu$ an operator $\Phi(I)=\Phi(I)^*\in\mathfrak{A}(I)$. The Hamiltonian for a finite subset Λ of $\mathbf{Z}^\nu$ is given by $U(\Lambda)=\sum_{I\subset\Lambda}\Phi(I)$. In order to control long-range interactions, various assumptions are introduced. Examples are finiteness of either of the following:

$$\|\Phi\|=\sup_n\sum_{I\ni n}N(I)^{-1}\|\Phi(I)\|, \qquad (19)$$

$$|||\Phi|||=\sup_n\sum_{I\ni n}\|\Phi(I)\|. \qquad (20)$$

Here $N(I)$ is the number of points in I.

Let $\mathfrak{A}_n^{\mathrm{cl}}$ be a maximal Abelian *-subalgebra of $\mathfrak{A}_n$ (such as $\{c_1+c_2\sigma_z^{(n)}\}$) satisfying $\gamma_a\mathfrak{A}_n^{\mathrm{cl}}=\mathfrak{A}_{n+a}^{\mathrm{cl}}$ and $\mathfrak{A}^{\mathrm{cl}}$ be the Abelian C^*-subalgebra of $\mathfrak{A}$ generated by $\mathfrak{A}_n^{\mathrm{cl}}$, $n\in\mathbf{Z}^\nu$. If $\Phi(I)$ is in $\mathfrak{A}^{\mathrm{cl}}$ for all I, we call the potential Φ **Abelian** or **classical**. There exists a conditional expectation π^{cl} which is a positive mapping of norm 1 from $\mathfrak{A}$ onto $\mathfrak{A}^{\mathrm{cl}}$ satisfying $\pi^{\mathrm{cl}}(ABC)=A\pi^{\mathrm{cl}}(B)C$ for A and C in $\mathfrak{A}^{\mathrm{cl}}$ and $\pi^{\mathrm{cl}}(1)=1$. If a state φ on $\mathfrak{A}$ satisfies $\varphi(A)=\varphi(\pi^{\mathrm{cl}}(A))$, we call the state φ **classical**. Classical states are in one-to-one correspondence with the restriction on $\mathfrak{A}^{\mathrm{cl}}$, which can be viewed as a probability measure on the spectrum (also called **configuration space**) of the C^*-algebra $\mathfrak{A}^{\mathrm{cl}}$ of observables for classical spin lattice systems. This correspondence makes it possible to view classical spin lattice systems as quantum spin lattice systems with Abelian interactions.

For a given potential Φ, the time evolution of the infinitely extended system is described by the one-parameter group α_t, $t\in\mathbf{R}$, of *-automorphisms of $\mathfrak{A}$ defined as the following limit:

$$\alpha_t(A)=\lim_{\Lambda\nearrow\mathbf{Z}^\nu}e^{itU(\Lambda)}Ae^{-itU(\Lambda)} \quad (A\in\mathfrak{A}). \qquad (21)$$

The limit exists if $\Phi(I)=0$ for $N(I)>N$ and $|||\Phi|||<\infty$, or if for some $\lambda>0$ $\sum_n e^{\lambda n}(\sup_x\sum_I\{\|\Phi(I)\|\mid I\ni x,\ N(I)=n\})<\infty$, or if $\nu=1$ (1-dimensional lattice) and $\sup_x\sum_I\{\|\Phi(I)\|\mid I\cap(-\infty,x)\neq\varnothing,\ I\cap[x,\infty)\neq\varnothing\}<\infty$. An alternative way is first to define $\delta_\Phi(A)=\sum_I i[\Phi(I),A]$ for $A\in\bigcup_\Lambda\mathfrak{A}(\Lambda)$ (Λ is a finite subset of $\mathbf{Z}^\nu$), which exists if $|||\Phi|||<\infty$, and to prove that the closure $\bar\delta_\Phi$ of δ_Φ is a generator of a one-parameter subgroup α_t $(=\exp t\bar\delta_\Phi)$. In the above cases, $\bar\delta_\Phi$ is a generator.

A general canonical ensemble for a system in a finite subset Λ of the lattice $\mathbf{Z}^\nu$, with some boundary condition in the outside $\Lambda^c=\mathbf{Z}^\nu\setminus\Lambda$, is given by $\varphi_\Lambda(A)=(\tau_\Lambda\otimes\psi)(e^{-\beta H(\Lambda)}\times Ae^{-\beta H(\Lambda)/2})$, where $\tau_\Lambda\otimes\psi$ is the product of the unique tracial state τ_Λ on $\mathfrak{A}(\Lambda)$ and a state ψ on $\mathfrak{A}(\Lambda^c)$ (the boundary condition), $H(\Lambda)=U(\Lambda)+W(\Lambda)$ and $W(\Lambda)=\sum_I\{\Phi(I)\ I\cap\Lambda\neq\varnothing,\ I\cap\Lambda^c\neq\varnothing\}$ (the **surface energy**). The following conditions on a state φ of $\mathfrak{A}$ are mutually equivalent under the condition that $\bar\delta_\Phi$ is a generator (which holds under any one of the conditions described above) and is satisfied by any limit state of the above φ_Λ as $\Lambda\nearrow\mathbf{Z}^\nu$ (i.e., a state in $\bigcap_{\Lambda'}\overline{\{\varphi_\Lambda\mid\Lambda'\subset\Lambda,\psi\}}$, with the bar denoting weak closure).

1. **KMS condition**: $\varphi(A\alpha_{i\beta}(B))=\varphi(BA)$ for any A, $B\in\mathfrak{A}$ such that $\alpha_t(B)$ is an entire function of t. (φ is called a **β-KMS state**.)

2. **Local thermodynamic stability**: For any finite subset Λ of $\mathbf{Z}^\nu$ and for any state ψ having the same restriction to $\mathfrak{A}(\Lambda)$ as the state φ under consideration, $\tilde F_{\Lambda,\beta}(\varphi)\leqslant\tilde F_{\Lambda,\beta}(\psi)$ (the minimality of the **free energy** multiplied by β), where $\tilde F_{\Lambda,\beta}(\varphi)=\beta\varphi(H(\Lambda))-S_\Lambda(\varphi)$, $S_\Lambda(\varphi)=\lim\{S_{\Lambda'}(\varphi)-S_{\Lambda'\setminus\Lambda}(\varphi)\}$ as $\Lambda'\nearrow\mathbf{Z}^\nu$ (the **open system entropy**), $S_\Lambda(\varphi)=-\varphi(\log\rho_\Lambda(\varphi))$ (the **closed system entropy**) and the density matrix $\rho_\Lambda(\varphi)\in\mathfrak{A}(\Lambda)^+$ is defined by $\varphi(A)=\tau_\Lambda(\rho_\Lambda(\varphi)A)$ for all $A\in\mathfrak{A}(\Lambda)$.

3. **Gibbs condition**: For every finite subset Λ of $\mathbf{Z}^\nu$, the perturbed state $\varphi^{\beta W(\Lambda)}$ (not necessarily normalized) is the product $\varphi_\Lambda^G\times\psi$ of the Gibbs state $\varphi_\Lambda^G(A)=\mathrm{tr}(e^{-\beta U(\Lambda)}A)/\mathrm{tr}\,e^{-\beta U(\Lambda)}$ on $A\in\mathfrak{A}(\Lambda)$ and some (unknown) state ψ on $\mathfrak{A}(\Lambda^c)$, where the representative vector Φ for φ for the GNS representation π_φ is assumed to be separating for $\pi_\varphi(\mathfrak{A})''$, and then $\varphi^{\beta W(\Lambda)}(A)=(\Omega,\pi_\varphi(A)\Omega)$ for

$$\Omega=\sum_{n=0}^{\infty}\beta^n\int_0^{1/2}ds_1\int_0^{s_1}ds_2\dots\int_0^{s_{n-1}}ds_n\Delta_\Phi^{s_n}\pi_\varphi \times(W(\Lambda))\Delta_\Phi^{s_{n-1}-s_n}\dots\Delta_\Phi^{s_1-s_2}\pi_\varphi(W(\Lambda))\Phi, \qquad (22)$$

where Δ_Φ is the †modular operator for Φ and the series converges.

For a classical potential, this condition reduces to the conditions that φ is classical and that the restriction of φ to $\mathfrak{A}^{\mathrm{cl}}$ as a measure on the configuration space $\{1\dots d\}^{\mathbf{Z}^\nu}$ satisfies the following **DLR equation** due to R. L. Dobrushin, O. E. Lanford, and D. Ruelle: The conditional probability for $\xi(\Lambda)\in\{1\dots d\}^\Lambda$ knowing $\xi(\Lambda^c)\in\{1\dots d\}^{\mathbf{Z}^\nu\setminus\Lambda}$ is proportional to $\exp(-\beta H(\Lambda))$, where $H(\Lambda)=U(\Lambda)+W(\Lambda)$ is a function of $\xi(U(\Lambda)$ depending only on $\xi(\Lambda))$.

4. **Roepstorff-Araki-Sewell inequality**: For any $A\in\bigcup_\Lambda\mathfrak{A}(\Lambda)$, $i\varphi(A^*\delta_\Phi(A))$ is real and

$$-i\beta\varphi(A^*\delta_\Phi(A))\geqslant S(\varphi(A^*A),\varphi(AA^*)), \qquad (23)$$

where $S(u,v)=u\log(u/v)$ if $u>0$, $v>0$, $S(0,v)=0$ for $v\geqslant0$ and $S(u,0)=+\infty$ for $u>0$.

5. **Roepstorff-Fannes-Verbeure inequality**:

For α-entire $A \in \mathfrak{A}$,

$$\beta^{-1}\int_0^\beta \varphi(A^*\alpha_{i\lambda}(A))\,d\lambda \leqslant F(\varphi(A^*A), \varphi(AA^*)), \tag{24}$$

where $F(u,v)=(u-v)/\log(u/v)$ for $u>0$, $v>0$, $u \neq v$, $F(u,u)=u$ for $u>0$, $F(u,0)=F(0,v)=0$.

If the interaction is translationally invariant (i.e., $\gamma_a\Phi(I)=\Phi(I+a)$ for all $a\in\mathbf{Z}^\nu$ and I) and if we restrict our attention to translationally invariant states (i.e., $\varphi(\gamma_a(A))=\varphi(A)$ for all $A\in\mathfrak{A}$ and $a\in\mathbf{Z}^\nu$), then the following conditions are also equivalent to the above.

6. **Variational principle**: $\beta e(\varphi)-s(\varphi)\leqslant\beta e(\psi)-s(\psi)$ for all translationally invariant ψ (the minimality of the **mean free energy**), where $e(\varphi)=\lim N(\Lambda)^{-1}\varphi(U(\Lambda))=\lim N(\Lambda)^{-1}\varphi(H(\Lambda))$ (the **mean energy**), $s(\varphi)=\lim N(\Lambda)^{-1}S_\Lambda(\varphi)$ (the **mean entropy**), the infimum value $\beta e(\varphi)-s(\varphi)$ is $-P(\beta\Phi)$ with $P(\beta\Phi)=\lim N(\Lambda)^{-1}\log\tau_\Lambda(e^{-\beta U(\Lambda)})$ (the **pressure**), and the limits exist if $\Lambda\nearrow\mathbf{Z}^\nu$ is taken in the following **van Hove sense**: For any given cube C of lattice points, the minimal number $n_\Lambda^+(C)$ of translations of C that cover Λ and the maximal number $n_\Lambda^-(C)$ of mutually disjoint translations of C in Λ satisfy $n_\Lambda^+(C)/n_\Lambda^-(C)\to 1$ as $\Lambda\nearrow\mathbf{Z}^\nu$.

7. **Tangent to the pressure function**: $P(\Phi)$ is a continuous convex function on the Banach space of translationally invariant Φ with $\|\Phi\|<\infty$. A continuous linear functional α on this Banach space is a **tangent** to P at Φ if $P(\Phi+\Psi)\geqslant P(\Phi)+\alpha(\Psi)$ for all Ψ. For a translationally invariant state ψ, we define $\alpha_\psi(\Psi)=\psi(\sum_{I\ni 0}N(I)^{-1}\Psi(I))$. The condition is that $-\alpha_\varphi$ is a tangent to P at $\beta\Phi$. (Conversely, any tangent α to P at $\beta\Phi$ arises in this manner.)

The set K_β of all (normalized) β-KMS states is nonempty, compact, and convex. A β-KMS state φ is an extremal point of K_β if and only if it is factorial (i.e., the associated von Neumann algebra $\pi_\varphi(\mathfrak{A})''$ has a trivial center). It then has the **clustering property** $\lim_{a\to\infty}\{\varphi(A\gamma_a(B))-\varphi(A)\varphi(\gamma_a(B))\}=0$ and is interpreted as a **pure phase**. Any β-KMS state has a unique integral decomposition into extremal β-KMS states.

For any Φ, K_β is a one-point set for sufficiently small $|\beta|$. For a 1-dimensional system ($\nu=1$), K_β consists of only one point (uniqueness of equilibrium states usually interpreted as indication of no phase transition) if the surface energy $W([-N,N])$ is uniformly bounded (H. Araki, *Comm. Math. Phys.*, 44 (1975); A. Kishimoto, *Comm. Math. Phys.*, 47 (1976)). For the two-body interaction $\Phi(\{m,n\})=-J|m-n|^{-\alpha}\sigma_z^{(m)}\sigma_z^{(n)}$, this condition is satisfied if $\alpha>2$ while $\|\Phi\|<\infty$ and α_t defined if $\alpha>1$. There is more than one KMS state (with spontaneous magnetization) for $2\geqslant\alpha>1$ and large $\beta J>0$, and hence a phase transition exists (F. J. Dyson, *Comm. Math. Phys.*, 12 (1969); J. Fröhlich and T. Spencer, *Comm. Math. Phys.*, 83 (1982)). If a 1-dimensional interaction has a finite range (i.e., $\Phi(I)=0$ if the diameter of I exceeds some number r_0) or if it is classical and $\sum_{I\ni 0}N(I)^{-1}(\operatorname{diam} I+1)\|\Phi(I)\|<\infty$ for $d=2$, then $\varphi(A)$ for $\varphi\in K_\beta$ and $A\in\mathfrak{A}(\Lambda)$ for a finite Λ is real analytic in β and any other analytic parameter in the potential (Araki, *Comm. Math. Phys.*, 14 (1969); [22]; M. Cassandro and E. Olivieri, *Comm. Math. Phys.*, 80 (1981)).

For a 2-dimensional **Ising model** with the nearest-neighbor ferromagnetic interaction [23], K_β consists of only one point for $0\leqslant\beta\leqslant\beta_c$ while K_β for $\beta>\beta_c$ has exactly two extremal points corresponding to positive and negative magnetizations (M. Aizenman, *Comm. Math. Phys.*, 73 (1980); Y. Higuchi, *Colloquia Math. Soc. János Bolyai*, 27 (1979)). In this case, all KMS states are translationally invariant, while there exist (infinitely many) translationally noninvariant KMS states for sufficiently large β if $\nu=3$ (Dobrushin, *Theory Prob. Appl.*, 17 (1972); H. van Beijeren, *Comm. Math. Phys.*, 40 (1975)).

The accumulation points of β-KMS states as $\beta\to+\infty$ (or $-\infty$) provide examples of **ground** (or **ceiling**) **states** defined by any one of the following mutually equivalent conditions 1_+, 2_+ (or 1_-, 2_-) (O. Bratteli, A. Kishimoto, and D. W. Robinson, *Comm. Math. Phys.*, 64 (1978)):

1_+ (1_-). Positivity (negativity) of energy: For any $A\in\bigcup_\Lambda\mathfrak{A}(\Lambda)$, $-i\varphi(A^*\delta_\Phi(A))$ is real and positive (negative).

2_+ (2_-). Local minimality (maximality) of energy: For any finite subset Λ of $\mathbf{Z}^\nu$ and for any state ψ with the same restriction to $\mathfrak{A}(\Lambda^c)$ as the state φ under consideration, $\varphi(H(\Lambda))\leqslant\psi(H(\Lambda))$ ($\varphi(H(\Lambda))\geqslant\psi(H(\Lambda))$).

For translationally invariant potentials and states, the following condition is also equivalent to the above:

3_+ (3_-). Global minimality (maximality) of energy: $e(\varphi)\leqslant e(\psi)$ ($e(\varphi)\geqslant e(\psi)$) for all translationally invariant states ψ.

The totality of KMS, ground, and ceiling states can be characterized by the following formulation of the impossibility of **perpetual motion**: Let $P_t=P_t^*\in\mathfrak{A}$ be a norm-differentiable function of the time $t\in\mathbf{R}$ with a compact support, representing (external) time-dependent perturbations. Then there exists a unique perturbed time evolution α_t^P as a one-parameter family of *-automorphisms of $\mathfrak{A}$ satisfying $(d/dt)\alpha_t^P(A)=\alpha_t^P(\delta_\Phi(A)+i[P_t,A])$ for all $A\in\mathfrak{A}$ in the domain of δ_Φ. A state φ changes with time t as $\varphi_t(A)=\varphi(\tau_t^P(A))$ under the perturbed dynamics α_t^P, and the total

energy given to the system (mechanical work performed by the external forces) is given by $L^P(\varphi)=\int_{-\infty}^{\infty}\varphi_t(dP_t/dt)dt$. For KMS states at any β, as well as ground and ceiling states, $L^P(\varphi)\geqslant 0$ for any P_t. If φ is a factor state, the converse holds, i.e., $L^P(\varphi)\geqslant 0$ for all P_t implies that φ is either a KMS, ground, or ceiling state. The condition $L^P(\varphi)\geqslant 0$ for all P_t is equivalent to $-i\varphi(U^*\delta_\Phi(U))\geqslant 0$ for all unitary U in the domain of δ_Φ and in the identity component of the group of all unitaries of $\mathfrak{A}$. A state φ satisfying this condition is called **passive**, and a state φ whose n-fold product with itself as a state on $\mathfrak{A}^{\otimes n}$ is passive relative to $\alpha_t^{\otimes n}$ for all n is called **completely passive**. The last property holds if and only if φ is a KMS, ground, or ceiling state (W. Pusz and S. L. Woronowicz, *Comm. Math. Phys.*, 16 (1970)).

The totality of KMS, ground, and ceiling states can be characterized by a certain stability under perturbations (P_t considered above) under some additional condition on α_t (R. Haag, D. Kastler, and E. B. Trych-Pohlmeyer, *Comm. Math. Phys.*, 38 (1974); O. Bratteli, A. Kishimoto, and D. W. Robinson, *Comm. Math. Phys.*, 61 (1978)).

When a lattice spin system is interpreted as a lattice gas, an operator $N_n\in\mathfrak{A}_n^{\mathrm{cl}}$ (such as $(\sigma_z^{(n)}+1)/2$) is interpreted as the particle number at the lattice site n and $N(\Lambda)=\sum_{n\in\Lambda}N_n$ is the particle number in Λ. It defines a representation of a unit circle T by automorphisms τ_θ of $\mathfrak{A}$ defined as $\tau_\theta(A)=\lim e^{iN(\Lambda)\theta}Ae^{-iN(\Lambda)\theta}$ ($\Lambda\nearrow\mathbf{Z}^\nu$), called **gauge transformations** (of the first kind). The **grand canonical ensemble** can be formulated as a β-KMS state with respect to $\alpha_t\tau_{\mu t}$ (instead of α_t), where the real constant μ is called the **chemical potential**. It can be interpreted as an equilibrium state when the gauge-invariant elements $\{A\in\mathfrak{A}\mid\tau_\theta(A)=A\}$ instead of $\mathfrak{A}$ are taken to be the algebra of observables or as a state stable under those perturbations that do not change the particle number.

H. The Boltzmann Equation

Statistical mechanics of irreversible processes originated from the kinetic theory of gases. Long ago, Maxwell and Boltzmann tried to calculate viscosity and other physical quantities characterizing gaseous flow in nonequilibrium. The †Boltzmann equation is generally a nonlinear †integrodifferential equation. On the basis of this equation mathematical theories were developed by D. Enskog, S. Chapman, and D. Hilbert [2].

Free electrons in a metal can be regarded as forming an electron gas, in which electron scattering by lattice vibrations or by impurities is more important than electron-electron scattering. Following the example of gas theories H. A. Lorentz set forth a simple theory of irreversible processes of metallic electrons. His theory was, however, not quite correct, since metallic electrons are highly quantum-mechanical and classical theories cannot be applied to them. Quantum-mechanical theories of metal electrons were developed by A. Sommerfeld and F. Bloch.

I. Master Equations

The Boltzmann equation gives the velocity distribution function of a single particle in the system. This line of approach can be extended in two directions. The first is the so-called master equation. For example, consider a gaseous system consisting of N particles, and ask for the probability distribution of all the momenta, namely, the distribution function $f_N(p_1,\ldots,p_N;t)$, where $p_1,\ldots,p_N$ are the momenta of the N particles. The equations of motion are deterministic with respect to the complete set of dynamical variables $(x_1,p_1,\ldots,x_N,p_N)$. The equation for $f(p_1,\ldots,p_N,t)$ may not be deterministic, but it may be stochastic because we are concerned only with the variables $p_1,\ldots,p_N$, with all information about the space coordinates $x_1,\ldots,x_N$ disregarded. This situation is essentially the same in both classical and quantum statistical mechanics. If the duration of the observation process is limited to a finite length of time and the precision of the observation to a certain degree of crudeness, the time evolution of the momentum distribution function f_N can be regarded as a †Markov process. In general, an equation describing a Markov process of a certain distribution function is called a **master equation**. Typically it takes the following form for a suitable choice of variables x:

$$(\partial/\partial t)f(x,t)=\int dx'(W(x',x)f(x',t)-W(x,x')f(x,t)), \qquad (25)$$

where $W(x,x')$ is the transition probability from x to x'. By expanding the first integrand into a power series in $x-x'$, with x' fixed and by retaining the first few terms, we obtain the **Fokker-Planck equation**:

$$(\partial/\partial t)f(x,t)=-(\partial/\partial x)(\alpha_1(x)f(x,t))+(\partial^2/\partial x^2)(\alpha_2(x)f(x,t))/2, \qquad (26)$$

$$a_n(x)=\int W(x,x+r)r^n\,dr. \qquad (27)$$

J. The Hierarchy of Particle Distribution Functions

Another way of extending the Boltzmann equation is to consider a set of distribution functions of one particle, of two particles, and generally of n ($<N$) particles selected from the whole system of N particles. For example, a two-particle distribution function is the function $f_2(x_1, v_1, x_2, v_2, t)$ for positions and velocities of two particles at time t. The complete dynamics of the entire system of particles can be projected to the time evolution of this hierarchy of distribution functions. The equation of motion for f_1 then contains the function f_2 if the interaction of particles if pairwise, the equation for f_2 contains f_3, and so on. Thus the equations of motion for the set of distribution functions make a chain of equations. The whole chain is equivalent to the deterministic equations of motion for the dynamics of N particles. However, if the particle number N is made indefinitely large, with the time scale of observation always finite, the chain of equations for the distribution functions asymptotically approaches a stochastic process if certain conditions are satisfied. Approximate methods of solving the hierarchy equations in classical cases have been developed by J. Yvon, J. G. Kirkwood, M. Born, H. S. Green, and others.

In quantum statistics, similar hierarchy equations can be considered. A typical example is the so-called **Green's function method** [27].

K. Irreversible Processes and Stochastic Processes

The statistical mechanics of physical processes evolving in time is a hybrid of dynamics and the mathematical theory of stochastic processes. A typical example is the theory of Brownian motion. A colloidal particle floating in a liquid moves incessantly and irregularly because of thermal agitation from surrounding liquid molecules. For simplicity, an example of 1-dimensional Brownian motion is considered here. Phenomenologically we assume that a colloid particle follows an equation of motion of the form

$$m\dot{u} = -m\gamma u + f(t), \tag{28}$$

called the **Langevin equation**, where m is the mass of the colloid particle and u is the velocity. The first term on the right-hand side is the friction force due to viscous resistance, and the second term represents a random force acting on the particle from surrounding molecules.

If (28) describes the Brownian motion in thermal equilibrium, the friction constant $m\gamma$ and the random force cannot be independent, but are related by a theorem asserting that

$$m\gamma = \int_0^\infty \langle f(t_1) f(t_1 + t) \rangle \, dt. \tag{29}$$

In an electric conductor, the thermal motion of charge carriers necessarily induces irregularities of charge distribution, and so an electromotive force that varies in time in a random manner is created. This random electromotive force is similar to the random motion of a Brownian particle and is called the **thermal noise**. For such a thermal noise there exists a relation similar to (29) between the resistance and the random electromotive force. This relation is known as the **Nyquist theorem**. These theorems are contained in a more general theorem called the **fluctuation-dissipation theorem**.

When an external force is applied to a system in thermal equilibrium, some kind of irreversible flow, an electric current, for example, is induced in the system. The relationship between the flow and the external force is generally represented by an admittance. If the external force is periodic, the admittance is a function of frequency ω and is given by

$$\chi(\omega) = \int_0^\infty e^{-i\omega t} \psi(t) \, dt, \tag{30}$$

where $\psi(t)$ is equal to the correlation function of the flow that appears spontaneously as the fluctuation in thermal equilibrium when no external force is applied. This general expression for an admittance, often called the **Kubo formula**, gives a unified viewpoint from which responses of physical systems to weak external disturbances can be treated without recourse to the traditional kinetic approach.

The static limit ($\omega \to 0$) of the admittance is the **transport coefficient**. The reversibility of dynamics leads to relations among transport coefficients, called **Onsager's reciprocity relations** in the thermodynamics of irreversible processes.

When external disturbances are so large that the system deviates considerably from thermal equilibrium, the responses may show characteristic nonlinearities. Such nonlinear phenomena are important from both experimental and theoretical points of view, and constitute a central subject of modern research (→ 433 Turbulence and Chaos).

References

[1] L. Boltzmann, Lectures on gas theory, Univ. of California Press, 1964. (Original in German, 1912.)

[2] S. Chapman and T. G. Cowling, The mathematical theory of non-uniform gases, Cambridge Univ. Press, third edition, 1970.
[3] R. C. Tolman, The principles of statistical mechanics, Oxford Univ. Press, 1938.
[4] D. ter Haar, Elements of statistical mechanics, Holt, Rinehart and Winston, 1961.
[5] J. W. Gibbs, Elementary principles in statistical mechanics, Yale Univ. Press, 1902 (Dover, 1960).
[6] P. and T. Ehrenfest, The conceptual foundations of the statistical approach in mechanics, Cornell Univ. Press, 1959. (Original in German, 1911.)
[7] I. E. Farquhar, Ergodic theory in statistical mechanics, Interscience, 1964.
[8] V. I. Arnol'd and A. Avez, Ergodic problems of classical mechanics, Benjamin, 1968.
[9] R. Jancel, Foundations of classical and statistical mechanics, Pergamon, 1963.
[10] R. H. Fowler, Statistical mechanics, Cambridge Univ. Press, second edition, 1936.
[11] A. Ya. Khinchin, Mathematical foundation of statistical mechanics, Dover, 1949. (Original in Russian, 1943.)
[12] L. D. Landau and E. M. Lifshits, Statistical physics, Pergamon, 1969. (Original in Russian, 1964.)
[13] R. Kubo, Statistical mechanics, North-Holland, 1965.
[14] H. S. Green and C. A. Hurst, Order-disorder phenomena, Interscience, 1964.
[15] H. E. Stanley, Introduction to critical phenomena, Oxford Univ. Press, 1971.
[16] C. Domb and M. S. Green, Phase transitions and critical phenomena, Academic Press, 1972.
[17] O. Bratteli and D. W. Robinson, Operator algebras and quantum statistical mechanics II, Springer, 1981.
[18] D. Ruelle, Statistical mechanics: Rigorous results, Benjamin, 1969.
[19] D. Ruelle, Thermodynamic formalism, Addison-Wesley, 1978.
[20] D. W. Robinson, The thermodynamic pressure in quantum statistical mechanics, Springer, 1971.
[21] R. B. Israel, Convexity in the theory of lattice gas, Princeton Univ. Press, 1979.
[22] D. H. Mayer, The Ruelle-Araki transfer operator in classical statistical mechanics, Springer, 1980.
[23] B. M. McCoy and T. T. Wu, The two-dimensional Ising model, Harvard Univ. Press, 1973.
[24] N. S. Krylov, Works on the foundation of statistical physics, Princeton Univ. Press, 1979.
[25] K. Huang, Statistical mechanics, Wiley, 1963.
[26] F. Reif, Fundamentals of statistical and thermal physics, McGraw-Hill, 1965.
[27] A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskiĭ, Methods of quantum field theory in statistical physics, Prentice-Hall, 1963. (Original in Russian, 1962.)
[28]E. H. Lieb and D. C. Mattis, Mathematical physics in one dimension, Academic Press, 1966.

403 (XVIII.5) Statistical Models

A. General Remarks

A statistical model is defined by specifying the structure of the probability distributions of the relevant quantities. When a statistical model is used for the analysis of a set of data, its role is to measure the characteristics of a certain configuration of the data points. R. A. Fisher [1] advanced a systematic procedure for the application of statistical models. The process of statistical inference contemplated by Fisher may be characterized by the following three phases: (1) specification of the model, (2) estimation of the unknown parameters, and (3) testing the goodness of fit. The last phase is followed by the first when the result of the testing is negative. Thus the statistical inference contemplated by Fisher is realized through the process of introduction and selection of statistical models.

We always assume that the true distribution of an observation exists in each particular application of statistical inference, even though it may not be precisely known to us. Our partial knowledge of the generating mechanism of the observation suggests various possible constraints on the form of the true distribution. The basic problem of statistical inference is then to generate an approximation to the true distribution by using the available observational data and a model defined by a set of probability distributions satisfying the constraints.

B. The Criterion of Fit

The use of statistical models can best be explained by adopting the predictive point of view, which defines the objective of statistical inference as the determination of the predictive distribution, the probability distribution of a future observation defined as a function of the information available at present. The performance of a statistical inference procedure is then evaluated in terms of the expected discrepancy of the predictive distribution from the true distribution of the future observation.

The probabilistic interpretation of thermodynamic entropy developed by L. Boltzmann [2] provides a natural measure of the discrepancy between two probability distributions. The **entropy of a distribution** specified by the density $f(y)$ with respect to the distribution specified by $g(y)$ is defined by

$$B(f;g) = -\int \frac{f(y)}{g(y)} \log\left[\frac{f(y)}{g(y)}\right] g(y)\,dy,$$

where, as in what follows, the integral is taken with respect to some appropriate measure dy. This definition of entropy is a faithful reproduction of the original probabilistic interpretation of the thermodynamic entropy by Boltzmann and allows the interpretation that the entropy $B(f;g)$ is proportional to the logarithm of the probability of getting a statistical distribution of observations closely approximating $f(y)$ by taking a large number of independent observations from the distribution $g(y)$. (For a detailed discussion → [3].)

Obviously we have

$$B(f;g) = \int f(y)\log g(y)\,dy - \int f(y)\log f(y)\,dy.$$

The second term on the right-hand side is a constant depending only on $f(y)$. The first term is the expected log likelihood of the distribution $g(y)$ with respect to the true distribution $f(y)$. Thus a distribution with a larger value of the expected log likelihood provides a better approximation to the true distribution. Even when $f(y)$ is unknown, $\log g(y)$ provides an unbiased estimate of the expected log likelihood. This fact constitutes the basis of the objectivity of the †likelihood as a criterion for judging the goodness of a distribution as an approximation to the true distribution.

C. Parametrization of Probability Distributions

When we construct a statistical model it is a common practice to represent the uncertain aspect of the true distribution by a family of probability distributions with unknown parameters. This type of family is called a parametric family; the model is called a parametric model. The parameters in a parametric model are the keys to the realization of the information extraction from data by statistical methods. Accordingly, the introduction of mathematically manageable parametric models forms the basis for the advance of statistical methods.

(1) Pearson's System of Distributions. A wide family of distributions can be generated by assuming a rational-function representation of the sensitivity of the density function $f(y)$ given by

$$\frac{d}{dy}\log f(y) = \frac{a_0 + a_1 y + \ldots + a_p y^p}{b_0 + b_1 y + \ldots + b_q y^q}.$$

Pearson's system of distributions is defined by putting $p=1$ and $q=2$ and by assuming various constraints on the parameters a_i and b_j and the support of $f(y)$ [4].

E. Wong [5] discussed the construction of continuous-time stationary Markov processes with the distributions of Pearson's system as their stationary distributions. This allows a structural interpretation of the parameters of a distribution of the system.

(2) Maximum Entropy Principle and the Exponential Family. To develop a formal theory of statistical mechanics E. T. Janes [6] introduced the concept of the maximum entropy estimate of a probability distribution. This concept leads to a natural introduction of the exponential family. Following Kullback [7], we start with a distribution $g(y)$ and try to find $f(y)$ with prescribed expectations of statistics $T_1(y), \ldots, T_k(y)$ and with maximum entropy $B(f;g)$. Such a distribution $f(y)$ is given by the relation

$$f(y) \propto \exp[\tau_1 T_1(y) + \ldots + \tau_k T_k(y)]\,g(y),$$

where it is assumed that the right-hand side is integrable. By varying the parameters $\tau_1, \ldots, \tau_k$ over the allowable range we get the exponential family of distributions. I. J. Good [8] considered the Janes procedure as a principle for the generation of statistical hypotheses and called it the maximum entropy principle.

(3) Parametric Models of Normal Distributions. Of particular interest within the exponential family is the family of normal distributions. This is obtained by assuming the knowledge of the first- and second-order moments of a distribution and applying the maximum entropy principle [9]. Obviously, the parametrization of a normal distribution is concerned only with the mean vector and the variance-covariance matrix.

Let $\mathbf{X} = (X_1, \ldots, X_n)'$ be an n-dimensional normal random variable with mean $E\mathbf{X} = (m_1, \ldots, m_n)'$ and variance-covariance matrix $\Sigma = (\sigma_{ij})$, where $\sigma_{ij} = E(X_i - m_i)(X_j - m_j)$. ($E$ denotes expectation and $'$ denotes the transpose.) A nonrestrictive family of n-dimensional normal distributions is characterized by $n + n(n+1)/2$ parameters, m_i $(i = 1, \ldots, n)$, and σ_{ij} $(i = 1, \ldots, n; j = 1, \ldots, n)$. The prior information on the generating mechanism of $\mathbf{X}$ introduces constraints on these parameters and reduces the number of free parameters.

Reduction of the dimensionality of the

parameter vector $\mathbf{m}=(m_1, \ldots, m_n)'$ is realized by assuming that $\mathbf{m}$ is an element of a k-dimensional subspace of $\mathbf{R}^n$ spanned by the vectors $\mathbf{a}_1=(a_{11}, \ldots, a_{1n})', \ldots, \mathbf{a}_k=(a_{k1}, \ldots, a_{kn})'$, i.e., by assuming the relation

$$\mathbf{m}=A\mathbf{c},$$

where $A=(\mathbf{a}_1, \ldots, \mathbf{a}_k)$ is an $n \times k$ matrix and $\mathbf{c}=(c_1, \ldots, c_k)'$ is a k-dimensional vector with $k<n$. This parametrization is obtained when for each X_i the observation $(a_{i1}, \ldots, a_{ik})$ is made on a set of k factors and the analysis of the linear effect of these factors on the mean of X_i is required. We have the representation $\mathbf{X}=A\mathbf{c}+\mathbf{W}$, where $\mathbf{W}$ is an n-dimensional normal random vector with $E\mathbf{W}=\mathbf{0}$ and variance-covariance matrix Σ (→ Section D).

To complete the model we have to specify the variance-covariance matrix Σ. One of the simplest possible specifications is obtained by assuming that the X_i are mutually independent and of the same variance σ^2. This reduces Σ to $\sigma^2 I$, where I denotes an $n \times n$ identity matrix. With this assumption the number of necessary parameters to represent the variance-covariance matrix reduces from $n(n+1)/2$ to 1. The model obtained with the assumptions $\Sigma=\sigma^2 I$ and $\mathbf{m}=A\mathbf{c}$ is called the general linear model (or regression model) with normal error, or simply the **normal linear model**. The model is called a regression model also when the a_i are random variables (→ e.g., [10, 11]).

A typical example of nontrivial parametrization of the covariance structure of $\mathbf{X}$ is obtained by assuming the representation

$$\mathbf{X}=\mathbf{m}+A\mathbf{F}+\mathbf{W},$$

where $\mathbf{F}=(f_1, \ldots, f_g)'$ denotes the vector of random effects and $\mathbf{W}$ the vector of measurement errors. It is assumed that $\mathbf{F}$ and $\mathbf{W}$ are mutually independent and normally distributed with $E\mathbf{F}=\mathbf{0}$ and $E\mathbf{W}=\mathbf{0}$. Also the components of $\mathbf{W}$ are assumed to be mutually independent. The variance-covariance matrix Σ of $\mathbf{X}$ is then given by

$$\Sigma=A\Phi A'+\Delta,$$

where $\Phi=E\mathbf{F}\mathbf{F}'$ and $\Delta=E\mathbf{W}\mathbf{W}'$, which is diagonal.

When A is a design matrix, the parametrization provides a **components-of-variance model** (or **random-effects model**) of which the main use is the measurement of the variance-covariance matrix Φ of the random effects $f_1, \ldots, f_g$ rather than the measurement of $\mathbf{F}$ itself. If we consider $\mathbf{F}$ to be representing the effects of some latent factors for which A is not uniquely specified, the above representation of Σ gives merely a formal, or noncausal, parametrization of Σ. In this case the model is called the **factor analysis model** and the dimension g of $\mathbf{F}$ is called the number of factors. By keeping the number of factors sufficiently smaller than the dimension of $\mathbf{X}$, we get a parametrization of Σ with a smaller number of free parameters than the unconstrained model. Starting with $g=1$ and successively increasing the number of factors, we can get a hierarchy of models with successively increasing numbers of parameters. (→ [12] for a very general modeling of the variance-covariance matrix.)

(4) Parametrization of Discrete Distribution. Consider the situation where the observation produces one of the events represented by $r=0, 1, 2, \ldots, k$ with probability $p(r)$, where k may be infinite. Represent by $\mathbf{X}=(X_1, \ldots, X_n)$ the result of n independent observations. The probability $p(\mathbf{X})$ of getting such a result is given by the relation

$$\log p(\mathbf{X})=\sum_{r=0}^{k} \theta_r n_r(\mathbf{X}),$$

where $\theta_r=\log p(r)$ and $n_r(\mathbf{X})$ denotes the number of X_i's which are equal to r. (The term not depending on the θ_r's is omitted in the above and subsequent formulas, since it is immaterial for problems of inference.) Thus a nonrestrictive model is obtained by assuming only the relations $\theta_r \leqslant 0$ and $\sum_{r=0}^{k} e^{\theta_r}=1$. Obviously the model defines an exponential family and various useful parametrizations are realized by introducing some constraints on the parameters θ_r.

When the events r are arranged in a 2-dimensional array (i,j) $(i=1, \ldots, m; j=1, \ldots, n)$ we have

$$\log p(\mathbf{X})=\sum_{i=1}^{m} \sum_{j=1}^{n} \theta_{ij} n_{ij}(\mathbf{X}).$$

One simple parametrization is given by

$$\theta_{ij}=\mu+\alpha_i+\beta_j+\gamma_{ij},$$

where it is assumed that $\sum_{i=1}^{m} \alpha_i=\sum_{j=1}^{n} \beta_j=\sum_{i=1}^{m} \gamma_{ij}=\sum_{j=1}^{n} \gamma_{ij}=0$. Obviously this is a parametrization of θ_{ij} as a linear function of the parameters α_i, β_j, and γ_{ij}, and the model thus obtained is called the **log linear model**. The model shows a formal similarity to the analysis-of-variance model (→ Section D). By introducing successively more restrictive assumptions on the parameters, we can get a hierarchy of models for the analysis of a two-way contingency table. Extension to cases when more than two factors are involved is obvious (→ e.g., [13]).

Here we consider that X_i is a dichotomous variable, i.e., $k=1$, and that the probability of $X_i=1$ may depend on i, i.e., we have

$$\log p(X_i)=n\theta_{0i}+(\theta_{1i}-\theta_{0i})n_1(X_i),$$

where $\theta_{0i} = \log \text{Prob}(X_i = 0)$ and $\theta_{1i} = \log \text{Prob}(X_i = 1)$. We assume that a vector of observations $\mathbf{a}_i = (a_{i1}, \ldots, a_{im})'$ is available simultaneously with X_i and that we are interested in analyzing the relation between $\mathbf{a}_i$ and the probability distribution of X_i. The analysis is realized by exploring the functional relation between $\tau_i = \theta_{1i} - \theta_{0i}$ and $\mathbf{a}_i$. We can assume a linear relation

$$\boldsymbol{\tau} = A\mathbf{c},$$

where $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_n)'$, $A = (\mathbf{a}_1, \ldots, \mathbf{a}_n)'$, and $\mathbf{c} = (c_1, \ldots, c_m)'$. The parameter $\tau_i = \log\{p(X_i = 1)/(1 - p(X_i = 1))\}$ is the log odds ratio or logit of the event $X_i = 1$, and the model is called the **linear logistic model** [14]. A hierarchy of models can be generated by assuming a successively more restrictive linear relations among the components of $\mathbf{C}$.

D. General Linear Models

Another class of models often used in practical applications is composed of general linear models or linear regression models, where the observed value is considered to be the sum of the effects of some fixed causes and the error.

Let $\mathbf{X} = (X_1, \ldots, X_n)'$ be an n-dimensional †random variable, and denote the expectation of $\mathbf{X}$ by $E(\mathbf{X}) = (\mu_1, \ldots, \mu_n)'$. If $E(\mathbf{X})$ is of the form $A\xi$ with an unknown parameter $\xi = (\xi_1, \ldots, \xi_k)'$, and a given $n \times k$ matrix A, then we can express $\mathbf{X}$ as

$$\mathbf{X} = A\xi + \mathbf{W}, \qquad E(\mathbf{W}) = (0, \ldots, 0)', \tag{1}$$

with the **error term** $\mathbf{W} = (W_1, \ldots, W_n)'$. We frequently assume a set of conditions on the distribution of $\mathbf{X}$; for example, (i) $X_1, \ldots, X_n$ are mutually †independent, (ii) $X_1, \ldots, X_n$ have a common unknown †variance σ^2, (iii) $(X_1, \ldots, X_n)$ is distributed according to an n-dimensional †normal distribution. The equations (1) together with conditions on the distribution are called a **linear model**.

Among the methods of statistical analysis of linear models are regression analysis, analysis of variance, and analysis of covariance as explained below, but these are not clearly distinguished from each other. (I) In **design-of-experiment analysis**, i.e., **analysis of variance**, the matrix A and the vector ξ in (1) are called a **design matrix** and an **effect**, respectively. In this case entries of A are assumed to be either 1 or 0. (II) In **regression analysis**, we are first given a linear form $x = \sum_{j=1}^{k} a_j \xi_j$ of a vector $\mathbf{a} = (a_1, \ldots, a_k)'$ with coefficient vector $\xi = (\xi_1, \ldots, \xi_k)'$. Let $X_1, \ldots, X_n$ be the observed values of x at n points $\mathbf{a}_1 = (a_{11}, \ldots, a_{1k})', \ldots, \mathbf{a}_n = (a_{n1}, \ldots, a_{nk})'$, respectively, where $n > k$. If the observations are unbiased, that is, if $E(X_i) = \sum_{j=1}^{k} a_{ij}\xi_j$, $i = 1, \ldots, n$, then the model (1) is obtained with $A = (a_{ij})$. Usually one of the components of the vector $\mathbf{a}$ is taken as unity. In this framework the form $x = \sum_j a_j \xi_j$ is called the **linear regression function** or **regression hyperplane**, and for $k = 2$, the graph of the linear function $x = a_1\xi_1 + \xi_2$ and its coefficient ξ_1 are called the **regression line** and **regression coefficient**, respectively. The components of the vector $\mathbf{a}$ are called **fixed variates** or **explanatory variables**. Frequently we encounter the case where the vector $\mathbf{a}$, and consequently the matrix A, are random variables. When this is the case, a discussion like that above can be carried out for given A by regarding $A\xi$ in (1) as the conditional expectation of the vector $\mathbf{X}$.

E. The Method of Least Squares

Consider the subspace $L(A)$ of the †sample space $\mathbf{R}^n$ spanned by the column vectors of A. Then the dimension s of $L(A)$ equals the rank of A, and $L(A)$ and its †orthocomplement $L^{\perp}(A)$ are called the **estimation space** and the **error space**, respectively. The †orthogonal projection $\mathbf{y}$ of a point $\mathbf{x}$ to the space $L(A)$ is expressed as $\mathbf{y} = P_A\mathbf{x}$ with a real †projection matrix P_A. The variable $\mathbf{Y} = P_A\mathbf{X}$ is called the **least squares estimator** of $E(\mathbf{X})$, and the routine of getting such an estimator $\mathbf{Y}$, called the **method of least squares**, minimizes the squared error $(\mathbf{X} - A\xi)'(\mathbf{X} - A\xi)$ for a given $\mathbf{X}$. This method consists of two operations solving the **normal equation** $A'A\xi = A'\mathbf{X}$ with respect to ξ, and setting $\mathbf{Y} = A\hat{\xi}$, where $\hat{\xi}$ is a solution of the equation. For $s = k$, we obtain $\mathbf{Y} = A(A'A)^{-1}A'\mathbf{X}$ directly. Even when $s < k$, where the solution of the normal equation is not unique, $\mathbf{Y}$ is uniquely determined. The quantity $Q = \mathbf{X}'(I - P_A)\mathbf{X}$, where I is the unit $n \times n$ matrix, is the squared distance of the point $\mathbf{X}$ from the space $L(A)$ and is called the **error sum of squares with $n - s$ degrees of freedom**.

A linear function $\boldsymbol{\beta}'\xi$ of the parameter ξ with coefficient vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$ is called a **linearly estimable parameter** (or **estimable parameter**) if there is a linear unbiased estimator, that is, an unbiased estimator of the form $\mathbf{b}'\mathbf{X}$, of $\boldsymbol{\beta}'\xi$. In order that $\boldsymbol{\beta}'\xi$ be estimable it is necessary and sufficient that $\boldsymbol{\beta}'$ be a linear combination $\mathbf{u}'A$ of the row vectors of the matrix A. A linear unbiased estimator that has minimum variance among all linear unbiased estimator uniformly in ξ is called the **best linear unbiased estimator (b.l.u.e.)**. If the conditions (i) and (ii) of Section D are satisfied, then for any given n-vector $\mathbf{u}$ the b.l.u.e. of a parameter $\gamma = \mathbf{u}'A\xi$ is given by $\hat{\gamma} = \mathbf{u}'\mathbf{Y}$ with $\mathbf{Y} = P_A\mathbf{X}$, and its variance equals $(\mathbf{u}'P_A\mathbf{u})\sigma^2$, while the expectation of the quantity Q is

given by $(n-s)\sigma^2$. This proposition is known as the **Gauss-Markov theorem**. Hence the b.l.u.e. $\hat{\gamma}=\mathbf{u}'\mathbf{Y}$ is frequently cited as the least squares estimator of γ. The quantity $\hat{\sigma}^2 = Q/(n-s)$ is an unbiased estimator of σ^2 and is called the **mean square error**. If in addition the condition (iii) of Section D is assumed, then $\hat{\gamma}$ and $\hat{\sigma}^2$ are the †uniformly minimum variance unbiased estimators of γ and σ^2, respectively.

When the error term $\mathbf{W}$ in (1) has covariance matrix $\Sigma=\sigma^2\Sigma_0$ with an unknown real parameter σ^2 and a known matrix Σ_0, the valaue of the parameter ξ minimizing $Q=(\mathbf{X}-A\xi)'\Sigma_0^{-1}(\mathbf{X}-A\xi)$ is called the **generalized least squares estimator** of ξ if it exists. This estimator has properties similar to those of the least squares estimator.

F. Model Selection and the Method of Maximum Likelihood

When a parametric family of distributions $\{f(\cdot\,|\,\theta);\theta\in\Theta\}$ is given and an observation x is made, $\log f(x\,|\,\theta)$ provides an unbiased estimate of the expected log likelihood of the distribution $f(\cdot\,|\,\theta)$ with respect to the true distribution of the observation. The value of θ which maximizes this estimate is the maximum likelihood estimate of the parameter and is denoted by $\theta(x)$ (→ 399 Statistical Estimation).

In a practical application we often have to consider a **multiple model**, defined by a set of component models $\{f_i(\cdot\,|\,\theta_i);\ \theta_i\in\Theta_i\}$ $(i=1,\ldots,k)$. The problem of model selection is concerned with the selection of a **component model** from a multiple model. The difference of the difficulties of handling a **simple model** defined by a one-component model, and a multiple model is quite significant. For a simple model $\{f(\cdot\,|\,\theta);\theta\in\Theta\}$, each member of the family is a probability distribution. In the case of a multiple model, its member is a model which is simply a collection of distributions and does not uniquely specify a probabilistic structure for the generation of an observation. Thus the likelihood of each component model with respect to the observation x cannot be defined and the direct extension of the method of maximum likelihood to the problem of model selection is impossible. This constitutes a serious difficulty for the handling of multiple models. Apparently, Fisher used the procedure of testing to solve this difficulty.

(1) Analysis of log Likelihood Ratios. The procedure of model selection by testing, which is applicable to a wide class of models, is the method of analysis of log likelihood ratios [15]. Consider the situation where a model is to be determined by using a hierarchy of models $\{f(\cdot\,|\,\theta_i);\theta_i\in\Theta_i\}$ $(i=1,\ldots,k)$ such that $\Theta_1\subset\Theta_2\subset\ldots\subset\Theta_k$. The comparison of models is then realized through the comparison of the maximum likelihoods $f(x\,|\,\theta_i(x))$, where $\theta_i(x)$ denotes the maximum likelihood estimate of θ_i based on the data x. For $\Theta_i\subset\Theta_j$ the log likelihood ratio is defined by

$$\Lambda(\Theta_i/\Theta_j;x)=-2\log\{f(x\,|\,\theta_i(x))/f(x\,|\,\theta_j(x))\}.$$

The analysis of log likelihood ratios is realized by the decomposition

$$\Lambda(\Theta_1/\Theta_k;x)=\Lambda(\Theta_1/\Theta_2;x)+\Lambda(\Theta_2/\Theta_3;x)+\ldots+\Lambda(\Theta_{k-1}/\Theta_k;x).$$

The log likelihood ratios $\Lambda(\Theta_{k-1}/\Theta_k;x)$, $\Lambda(\Theta_{k-2}/\Theta_{k-1};x),\ldots,\Lambda(\Theta_1/\Theta_2;x)$ are successively tested by referring to chi-square distributions with the degrees of freedom $d(k)-d(k-1)$, $d(k-1)-d(k-2),\ldots,d(2)-d(1)$, respectively, where $d(i)$ denotes the dimension of the manifold Θ_i. The assumption of the chi-square distributions is only asymptotically valid under the usual regularity conditions (→ 400 Statistical Hypothesis Testing). The model defined with Θ_i for which $\Lambda(\Theta_i/\Theta_{i-1};x)$ first becomes significant is selected and $f(y\,|\,\theta_i(x))$ is accepted as the predictive distribution. The problem of how to choose the levels of significance to make the test procedure a procedure for model selection remains open.

(2) Model selection by AIC. One way out of the difficulty of model selection is to assume a prior distribution over Θ_i for each model $\{f_i(\cdot\,|\,\theta_i);\theta_i\in\Theta_i\}$. This leads to Bayesian modeling, which is discussed in Section G. Another possibility is to replace each component model $\{f_i(\cdot\,|\,\theta_i);\theta_i\in\Theta_i\}$ by a distribution $f_i(\cdot\,|\,\theta_i(x))$ specified by the maximum-likelihood estimate $\theta_i(x)$. The problem here is how to define the likelihood of each distribution $f_i(\cdot\,|\,\theta_i(x))$. An information criterion **AIC** was introduced by H. Akaike [16] for this purpose; it is defined by

$$\mathrm{AIC}=(-2)\log_e(\text{maximum likelihood})+2\,(\text{number of estimated parameters}).$$

We may consider -0.5 AIC to be the log "likelihood" of $f(\cdot\,|\,\theta(x))$ which is corrected for its bias as an estimate of $E_xE_y\log f(y\,|\,\theta(x))$, where E_x denotes the expectation with respect to the true distribution of x, and where it is assumed that x and y are independent and identically distributed. The maximum "likelihood" estimate of the model is then defined by the model with minimum AIC. This realizes a procedure of model selection that avoids the ambiguity of the testing procedure. It is applicable, at least formally, even to the case of a nonhierarchical set of models.

G. Bayesian Models

Consider the situation where an observation x is made and it is desired to produce an estimate $p(y|x)$ of the true distribution of a future observation y. $p(y|x)$ is called a **predictive distribution**. Assume that x and y are sampled from one of the distributions within the family $\{g(y|\theta)f(x|\theta); \theta \in \Theta\}$, i.e., x and y are stochastically independent but share common structural information represented by θ. As a design criterion of $p(y|x)$ we assume a probability distribution $\pi(\theta)$ of θ. The model $\{f(\cdot|\theta); \theta \in \Theta\}$ with $\pi(\theta)$ is called a **Bayesian model** and $\pi(\theta)$ is called the **prior distribution**. From the relation

$$E_\theta E_{x|\theta} E_{y|\theta} \log p(y|x) = E_x E_{y|x} \log p(y|x),$$

where $E_{y|x}$ denotes the expectation of y conditional on x and E_x the expectation with respect to the marginal distribution of x, it can be seen that the optimal choice of $p(y|x)$ which maximizes the expected log likelihood is given by the conditional distribution

$$p(y|x) = \int g(y|\theta) p(\theta|x) d\theta,$$

where $p(\theta|x)$ is the **posterior distribution** of θ defined by

$$p(\theta|x) = f(x|\theta)\pi(\theta)\left(\int f(x|\theta)\pi(\theta) d\theta\right)^{-1}.$$

When a prior distribution $\pi(\theta)$ is specified, the parametric family of distributions $\{f(\cdot|\theta); \theta \in \Theta\}$ is converted into a stochastic structure which specifies a probability distribution of the observations. The likelihood of the structure, or the Bayesian model, with respect to an observation x is defined by

$$\int f(x|\theta)\pi(\theta) d\theta.$$

When there is uncertainty about the choice of the prior distribution we can consider a set of possible prior distributions and apply the method of maximum likelihood. Such a procedure is called the method of type II maximum likelihood by I. J. Good [17]. For a multiple model $\{f_i(\cdot|\theta_i); \theta_i \in \Theta_i\}$ $(i = 1, \ldots, k)$, if prior distributions $\pi_i(\theta_i)$ are defined, a model selection procedure is realized by selecting the Bayesian model with maximum likelihood.

Bayesian modeling has often been considered as not quite suitable for scientific applications unless the prior distribution is objectively defined. However, even the construction of an ordinary statistical model is always heavily dependent on our subjective judgment. Once the objective nature of the likelihood of a Bayesian model is recognized, the selection or determination of a Bayesian model can proceed completely analogously to Fisher's scheme of statistical inference (→ e.g., [18]). The basic underlying idea of both the minimum AIC procedure and Bayesian modeling is the balancing of the complexity of the model against the amount of information available from the data. This unifying view of the construction of statistical models is obtained by the introduction of entropy as the criterion for judging the goodness of fit of a statistical model (→ [19] for more details).

References

[1] R. A. Fisher, Statistical methods for research workers, Oliver & Boyd, 1925; Hafner, fourteenth edition, 1970.
[2] L. Boltzmann, Über die Beziehung zwischen dem Zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung respektive den Sätzen über das Wärmegleichgewicht, Wiener Berichte, 76 (1877), 373–435.
[3] I. N. Sanov, On the probability of large deviations of random variables, IMS and AMS Selected Transl. Math. Statist. Prob., 1 (1961), 213–244. (Original in Russian, 1957.)
[4] K. Pearson, Contributions to the mathematical theory of evolution II, Skew variation in homogeneous material, Philos. Trans. Roy. Soc. London, ser. A, 186 (1895), 343–414. Also included in Karl Pearson's early statistical papers, Cambridge Univ. Press, 1948, 41–112.
[5] E. Wong, The construction of a class of stationary Markoff processes, Proc. Amer. Math. Soc. Symp. Appl. Math., 16 (1963), 264–276. Also included in A. H. Haddad (ed.), Nonlinear systems, Dowden, Hutchinson & Ross, 1975, 33–45.
[6] E. T. Janes, Information theory and statistical mechanics, Phys. Rev., 106 (1957), 620–630.
[7] S. Kullback, Information theory and statistics, Wiley, 1959 (Dover, 1967).
[8] I. J. Good, Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables, Ann. Math. Statist., 34 (1963), 911–934.
[9] C. E. Shannon and W. Weaver, The mathematical theory of communication, Univ. of Illinois Press, 1949.
[10] S. R. Searle, Linear models, Wiley, 1971.
[11] F. A. Graybill, Theory and application of the linear model, Duxbury Press, 1976.
[12] K. G. Jöreskog, A general method for analysis of covariance structures, Biometrika, 57 (1970), 239–251.
[13] Y. M. M. Bishop, S. E. Feinberg, and P. W. Holland, Discrete multivariate analysis: Theory and practice, MIT Press, 1975.

[14] D. R. Cox, The analysis of binary data, Chapman & Hall, 1970.
[15] I. J. Good, Comments on the paper by Professor Anscombe, J. Roy. Statist. Soc., ser. B, 29 (1967), 39–42.
[16] H. Akaike, A new look at the statistical model identification, IEEE Trans. Automatic Control, AC-19 (1974), 716–723.
[17] I. J. Good, The estimation of probabilities, MIT Press, 1965.
[18] G. E. P. Box, Sampling and Bayes' inference in scientific modeling and robustness, J. Roy. Statist. Soc., ser. A, 143 (1980), 383–430.
[19] H. Akaike, A new look at the Bayes procedure, Biometrika, 65 (1978), 53–59.

404 (XVIII.14) Statistical Quality Control

A. General Remarks

According to the Japanese Industrial Standard (JIS) Z 8101, "**Quality Control** (QC) is a system comprising all the methods used in manufacturing products or providing services economically that meet the quality requirements of consumers." To emphasize that modern quality control makes use of statistical methods, it is sometimes referred to as **Statistical Quality Control** (SQC). In order to implement effective QC, statistical concepts and methods must be applied and the "Plan-Do-Check-Action" (PDCA) cycle must be followed in research and development, design, procurement, production, sales, and so on. These QC activities are executed on a company-wide basis from the top management to the production workers. This type of QC is called Company-Wide Quality Control (CWQC) or Total Quality Control (TQC).

The quality Q is an abstract notion of the conformity of a product or service to consumers' requirements; it also refers to the total of the characteristics of a product or service as perceived by consumers. The quality characteristics may include both measurable physical and/or chemical features, such as strength and purity, or features such as color or texture as appreciated by individuals. These latter characteristics could be called "consumer qualities." Furthermore, the concept "quality" has also been used to describe the social impact of a product or service. This might be called "social quality." Examples of social-quality issues are pollution by solid waste or drainage in the production stage; degradation, maintainability, and safety of a product in daily use; and pollution following disposal. For a product or service to conform to this sort of quality it has become necessary to conduct QC activities not only during production but also at early stages of design and development of new products.

The measured characteristics of quality vary from one product to another because of natural variability in the material and production process involved, ability of individual workers, errors in different sorts of measurement, etc. If the variations among the measured values from a process can be attributed to "chance causes" and their distribution expressed by a probability or a probability density function, the process is said to be in a "state of statistical control" according to W. A. Shewhart or in a "stable state" by JIS. In this case the value of a characteristic is deemed to be the realization of a random variable X. Sometimes the variations are attributed to "assignable causes," which must be identified and eliminated.

B. Control Charts

The control chart provides a means of evaluating whether a process is in a stable state.

The control chart is made by plotting points illustrating a statistic of the quality characteristics or manufacturing conditions for an ordered series of samples or subgroups. A sheet of the control chart is provided with a middle line between a pair of lines depicting the upper control limit (UCL) and the lower control limit (LCL). The stable state is assumed to be exhibited by points within the control limits. Points falling outside the control limits suggest some assignable causes, which should be eliminated through corrective measures.

The idea underlying control charts as developed by Shewhart is to apply the statistical principle of significance to the control of production processes. Other types of control charts have also been developed, for example, acceptance control charts and adaptive control charts. These have been successfully applied to many quality control problems.

The foundation of Shewhart's control chart is the division of observations into what are called "rational subgroups." A rational subgroup is the one within which variability is due only to chance causes. Between different subgroups, however, variations due to assignable causes might be detected. In most production processes the rational subgroup comprises the data collected over a short period of time during which essentially the same con-

dition in material, tool setting, environmental factors, etc., prevails.

The limits on the control charts are placed, according to Shewhart, at a 3σ distance from the middle line, where σ is the population standard deviation (or standard error) of the statistic; 3σ expresses limits of variability within the subgroup. Assuming that the population distribution of an observed characteristic is "normal," the range between the limits should include 99.7% of the points plotted so long as the process is "in control" at the middle value. Accordingly, 0.3% of the plotted points from the "in control" process fall outside the limits, and thus give an erroneous "out of control" signal.

To determine that the process is in control for a normally distributed characteristic $N(\mu, \sigma^2)$, we have to investigate the variability between the means μ and the standard deviations σ of different distributions of X for different subgroups. Thus the state of control of a process is determined with control charts for

$$\bar{X} = \sum_{i=1}^{n} X_i/n \quad \text{and} \quad s = \sqrt{\sum_{i=1}^{n} (X_i - \bar{X})/(n-1)},$$

which are the appropriate statistics corresponding to μ and σ. Despite the theoretical drawback of the statistical range $R = \max_i X_i - \min_i X_i$ against s, use of the range is often preferred in QC work because of its simplicity in computation. Hence the $\bar{X}$–R charts are obtained from the previously collected k rational subgroups each of size n as follows:

$$\mathrm{UCL} = \bar{\bar{X}} + A_2 \bar{R}, \qquad \mathrm{UCL} = D_4 \bar{R},$$

$$\mathrm{LCL} = \bar{\bar{X}} - A_2 \bar{R}. \qquad \mathrm{LCL} = D_3 \bar{R},$$

where $\bar{\bar{X}}$ and $\bar{R}$ are the averages of the k values of $\bar{X}$ and R, respectively, and

$$A_2 = \frac{3}{\sqrt{n}d_2}, \qquad D_4 = 1 + 3\frac{d_3}{d_2}, \qquad D_3 = 1 - 3\frac{d_3}{d_2},$$

$$E[R] = d_2\sigma, \qquad V[R] = E[(R - E[R])^2] = {d_3}^2\sigma^2.$$

For $n < 7$, LCL for R cannot be given because D_3 becomes negative.

The other commonly used control charts are the p chart (proportion of nonconformity: binomial distribution) and the c chart (number of defects: Poisson distribution). For those charts the above theory of normal distribution is also used to approximate the binomial and Poisson values.

It is generally sufficient to use the agreed-on decision criterion (3σ limits) and to recognize a relatively small risk ($\alpha = 0.003$) for practical purposes. It should be noted, however, that a shift of the process mean μ by 1σ would not be observed at a ratio of 97.7%, which is the value of the risk β for each plotted point, under the normal distribution of the plotted statistic. One reason why the control chart has been a practical tool in many applications is this lack of sensitivity to a relatively small shift of the level. If greater sensitivity is required, 2σ limits, "warning limits" are used. This results in a greater risk α of erroneously finding a process out of control.

Other decision criteria based on aspects of run theory are also used. Charts using accumulated data from several rational subgroups for each plotted value are sometimes recommended: Moving average and moving range charts and the cusum ("cumulative sum") chart are examples. The statistical theory for these charts is more complicated than that for the simple charts discussed here.

C. Sampling Inspection

Sampling inspection determines whether a lot should be accepted or rejected by drawing a sample from it, observing a quality characteristic of the sample, and comparing the observed value to a prescribed acceptance criterion. Sampling may be conducted in several stages. Definite criteria are required to decide at each stage whether to accept or reject the lot or continue sampling on the basis of sample values observed so far. There also must be some rules to determine the size of the next sample if it is to be taken. These criteria and rules together are called the **sampling inspection plan**. The number of samples eventually drawn and observed and their sizes are generally random variables. In **single sampling inspection** the final decision is always reached after one stage of sampling is completed. **Double sampling inspection** makes the final decision after at most two stages of sampling are completed. **Multiple sampling inspection** makes the final decision after at most N stages of sampling are completed ($N < \infty$). Inspection without a predetermined limit on the number of sampling stages, **sequential sampling inspection**, is usually constructed so that the probability of the indefinite continuation of sampling is 0.

Once a sampling inspection plan is determined, the probability for accepting a lot with given composition can be calculated. This probability as a function of lot composition is called the **operating characteristic** of the plan. In most cases, the quality of a lot is expressed by a real parameter θ (e.g., fraction defective, i.e., percentage of defective products, or the average of some quality characteristic), and we use only inspection plans whose operating

characteristics are expressed as a function of θ. The graph of this function is called an **OC-curve**. We impose certain desirable conditions on the OC-curve and design plans to satisfy them. Tables for this purpose, **sampling inspection tables**, are prepared for practical use. The condition most frequently employed is expressed in the following form in terms of four constants θ_0, θ_1, α, β: The probability of rejection is required to be at most α when $\theta \leqslant \theta_0$ (or $\theta \geqslant \theta_0$), and the probability of acceptance at most β when $\theta \geqslant \theta_1$ (or $\theta \leqslant \theta_1$). Here α is called the **producer's risk**, and β the **consumer's risk**.

If the rejection of a lot is identified with the rejection of a statistical hypothesis $\theta \leqslant \theta_0$, the OC-curve is actually the power curve of the test upside down (i.e., the graph of 1 minus the †power function), and the producer's and consumer's risks are precisely the †errors of the first and second kind. The choice of a plan is actually the choice of a test under certain conditions on its power curve. Commonly used plans are mostly based on well-established tests, some of which have certain optimum properties. A few examples, (1)–(4), are given below. Here **sampling inspection by attribute** is an inspection plan that uses a statistic with a discrete distribution, whereas **sampling inspection by variables** uses a statistic with a continuous distribution (→ 400 Statistical Hypothesis Testing).

(1) Single sampling inspection by attribute concerning the fraction of defective items in a lot: Let the defective fraction be denoted by p and identified with θ in the preceding paragraph. Assign two values of p, say p_0 and p_1 ($0<p_0<p_1<1$), the producer's risk α, and the consumer's risk β. Together they give conditions to be fulfilled by the OC-curve. Draw n items from a lot at random, and suppose that they contain Z defective items. The decision is then made after observing Z, whose distribution is †hypergeometric and approximately †binomial when the size of the lot is large enough. There exists a plan that minimizes n among all plans satisfying the imposed conditions under either of the two assumptions about the distribution of Z. It rejects the lot when Z is greater than a fixed number determined by p_0, p_1, α, and β. This plan is based on the †UMP test of the hypothesis $p \leqslant p_0$ against the alternative $p \geqslant p_1$.

(2) Single sampling inspection by variables concerning the population mean μ in the case where the population distribution is $N(\mu,\sigma^2)$ with known σ^2: Draw a sample $(X_1,\ldots,X_n)$ of size n from a lot. Assume that the X are independently distributed with the same distribution $N(\mu,\sigma^2)$. Suppose that smaller values of the quality characteristic stand for a more desirable quality. If two values of μ, say μ_0 and μ_1, α, and β are assigned, a plan can be established to minimize n. It rejects the lot when the sample mean $\bar{X}=\sum_{i=1}^{n} X_i/n$ exceeds a fixed number determined by μ_0, μ_1, α, and β. This too is based on the UMP test of $\mu \leqslant \mu_0$ against $\mu \geqslant \mu_1$.

(3) Cases where the samples are drawn in more than one stage: As in (2), assign two values of θ, α, and β; there is still liberty to choose $n_1, n_2, \ldots$, which are the sizes of the samples drawn at each stage. Hence there are many possible plans fulfilling the imposed conditions. Among them a plan is sought to minimize the expectation of $n=n_1+n_2+\ldots$ (called the **average sample number**). For example, plans based on the sequential probability ratio tests are in common use (→ 400 Statistical Hypothesis Testing).

(4) Among other special plans, **sampling inspection with screening** and **sampling inspection with adjustment** are worthy of mention. In the first plan, all the units in the rejected lots are inspected and defective units replaced by nondefective ones. In this case, fixing p_1 (the **lot tolerance percent defective**) and β, or the average fraction defective after the inspection (**average outgoing quality level**), we attempt to minimize the **expected amount of inspection**, that is, the expected number of inspected units including those in the rejected lots. In the second plan, acceptance criteria are tightened or loosened according to the quality of the lots just inspected.

References

[1] J. M. Juran, Quality control handbook, McGraw-Hill, 1974.
[2] Japanese Standard Association, Terminology, JIS Z 8101, 1963.
[3] Japanese Standard Association, Sampling inspection, JIS Z 9001–9006, 1957.
[4] Japanese Standard Association, Control charts, JIS Z 9021–9023, 1963.

405 (XVII.16) Stochastic Control and Stochastic Filtering

A. General Description of Stochastic Control

Stochastic control is an optimization method for systems subject to random disturbance. Let Γ be a compact convex subset of $\mathbf{R}^k$, called a control region. Let W_t be an †n-dimensional Brownian motion and $\sigma_t(W)=\sigma(W_s; s\leqslant t)$ (say $\mathscr{F}_t$) be the least †σ-field for which W_s, $s \leqslant t$, are

measurable. An †$\mathcal{F}_t$-progressible measurable Γ-valued process is called an **admissible control**. For an admissible control U_t the system evolves according to an n-dimensional **controlled stochastic differential equation** (CSDE)

$$dX_t = \alpha(X_t, U_t)dW_t + \gamma(X_t, U_t)dt,$$

where a symmetric $n \times n$ matrix $\alpha(x,u)$ and n-vector $\gamma(x,u)$ are continuous in $\mathbf{R}^n \times \Gamma$ and Lipschitz continuous in $x \in \mathbf{R}^n$. Hence the CSDE has a unique solution, called the **response for** U_t. The problem is to maximize (or minimize) the performance J:

$$J(\tau, x, \varphi, U) = E_x\left[\int_0^\tau e^{-\int_0^t c(X_s, U_s)ds} f(X_t, U_t)dt + e^{-\int_0^\tau c(X_s, U_s)ds}\varphi(X_\tau)\right],$$

where X_t is the response for U_t with $X_0 = x$ and τ is a constant time or a †hitting time associated with a target set. We put $V(\tau, x, \varphi) = \sup_{U \in \text{adm. control}} J(\tau, x, \varphi, U)$ the **value function** as a function of x. If the supremum value is attained at an admissible control $\tilde{U}_t$, then $\tilde{U}_t$ is called an **optimal control**.

B. Bellman Principle

In order to get $V(t+s, x, \varphi)$, R. Bellman applied the following two-stage optimization. After using any U up to time t, a controller changes U to an optimal one. Then at time $t+s$ the performance $J(t, x, V(s, \cdot, \varphi), U)$ is obtained. Taking the supremum with respect to U, one gets $V(t+s, x, \varphi)$. This is called the **Bellman principle**. Let C be the †Banach lattice of the totality of bounded and uniformly continuous functions on $\mathbf{R}^n$. Suppose that α, γ, f and $c(\geqslant 0)$ are bounded and smooth; then for constant time t, the value function $V(t, x, \varphi)$ belongs to C whenever $\varphi \in C$. Moreover, the family of operators $V(t)$ defined by $V(t)\varphi(x) = V(t, x, \varphi)$ becomes a †monotone contraction semigroup on C. The semigroup property $V(t+s, x, \varphi) = V(t, x, V(s, \cdot, \varphi))$ is nothing but the Bellman principle. The †generator G is expressed by

$$G\varphi(x) = \sup_{u \in \Gamma}\{L^u\varphi(x) - c(x,u)\varphi(x) + f(x,u)\}$$

for a smooth function φ, where L^u is the generator of †diffusion of the response for constant control $u (\in \Gamma)$, namely,

$$L^u = \frac{1}{2}\sum_{i,j,p=1}^n \alpha_{ip}(x,u)\alpha_{jp}(x,u)\frac{\partial^2}{\partial x_i \partial x_i} + \sum_{i=1}^n \gamma_i(x,u)\frac{\partial}{\partial x_i}.$$

Furthermore, assume that α is uniformly positive definite and φ is smooth. Then $V(t)\varphi(x)$ belongs to †$W^{1,2}_{p,\text{loc}}$ for any p and is the unique solution of the **Bellman equation** (= dynamic programming equation)

$$\frac{\partial W}{\partial t} = \sup_{u \in \Gamma}\{L^u W - c(x,u)W + f(x,u)\}$$

a.e. on $(0, \infty) \times \mathbf{R}^n$, $\quad W(0,x) = \varphi(x) \quad$ on $\mathbf{R}^n$.

In addition if $\inf_{x,u} c(x,u) > 0$, then $W = \lim_{t \to \infty} V(t)\varphi$ exists and is the unique solution of the Bellman equation $\sup_{u \in \Gamma}\{L^u W - c(x,u)W + f(x,u)\} = 0$ a.e. on $\mathbf{R}^n$. When τ is the hitting time, the value function is related to the Bellman equation with a boundary condition.

C. Feedback Control

In practical problems we specify the kind of information on which the decision of the controller can be based at each time. We frequently assume that the data obtained up to that time is available. The following situations are possible: (1) The controller knows the complete state of the system. This is called the case of **complete observation**. (2) The controller has partial knowledge of the state of system. This is called the case of **partial observation**. A feedback control (= **policy**) is a function of available information, namely, a Γ-valued progressible measurable function defined on $[0, \infty) \times C^j[0, \infty)$, where j is the dimension of data and $C^j[0, \infty)$ is a metric space of totality of j-vector valued continuous functions on $[0, \infty)$. A policy U is called a **Markovian policy** if $U(t, \xi)$ is a Borel function of t and the tth coordinate of ξ. When a policy U is applied, the system is governed by the †SDE

$$dX_t = \alpha(X_t, U(t, Y))dW_t + \gamma(X_t, U(t, Y))dt$$

with data process Y. When the SDE has a †weak solution, U is called admissible. For example, when $X = Y$, any Markovian policy is admissible if α is uniformly positive definite. Let X_t be a weak solution for U. Then its performance $J(\tau, x, \varphi, U) \leqslant V(t, x, \varphi)$.

(1) The case of **complete observation**. When α is uniformly positive definite, an optimal Markovian policy can be constructed in the following way. Since Γ is compact, there exists a Borel function $\tilde{U}$ on $[0, \infty) \times \mathbf{R}^n$ which gives the supremum, namely,

$$\sup_{u \in \Gamma}\{L^u V(t)\varphi(x) - c(x,u)V(t)\varphi(x) + f(x,u)\}$$
$$= L^{\tilde{U}(t,x)}V(t)\varphi(x) - c(x, \tilde{U}(t,x))V(t)\varphi(x) + f(x, \tilde{U}(t,x)).$$

This relation implies that $V(t)\varphi(x) = J(t, x, \varphi, \tilde{U}(t, X_t))$ for any weak solution X_t. Hence

$\tilde{U}$ is an optimal Markovian policy. Especially when α, c, and f are independent of u and $\gamma(x,u)=R(x)\cdot u$, where $R(x)$ is an $n\times k$ matrix, an optimal Markovian policy $\tilde{U}$ is obtained by a measurable selection of maximum points of (grad. $V(t)\varphi$, $R(x)\cdot u$). Since the supremum of a linear form is attained at the boundary $\partial\Gamma$, one can suppose that $\tilde{U}(t,x)$ belongs to $\partial\Gamma$. This is called **bang-bang control.**

(2) The case of **partial observation**. One useful method is the **separation principle**. This means that the control problem can be split into two parts. The first is the mean square estimate for the system using a †filtering. The second is a stochastic optimal control with complete observation. But generally speaking, the problem of deciding under what conditions the separation principle is valid is difficult. In the case of the following linear regulator the separation principle holds.

Suppose that the system process X_t and the observation process Y_t obey the following SDEs:

$$dX_t=A(t)dW_t+(B(t)X_t+b(t,U(t,Y))dt,$$

$$dY_t=d\tilde{W}_t+H(t)X_t dt,$$

where $A(t)$, $B(t)$, and $H(t)$ are nonrandom matrix-valued functions and $\tilde{W}_t$ is a j-dimensional Brownian motion independent of W_t. The problem is to search for a feedback control which gives the maximum value. Suppose that a feedback control $U(t,\xi)$ is Lipschitz continuous in $\xi\in C^j[0,\infty)$. Put

$$Q(s,x)$$

$$=\sup_{U,\mathrm{Lip}} E_{sx}\left[\int_s^T L(t,X_t,U(t,Y))dt+\Phi(X_\tau)\right],$$

where (X_t, Y_t) is the unique solution for U with the initial condition $X_s=x$, $Y_s=0$. By the Lipschitz condition of U, $\sigma_t=\sigma(Y_s;s\leqslant t)$ is independent of U, and the †conditional expectation $\hat{X}_t=E(X_t/\sigma_t)$ is governed by the following SDE, by way of the †Kalman-Bucy filter:

$$d\hat{X}_t=P(t)H'(t)dW_t^*+(B(t)\hat{X}_t+b(t,\hat{U}_t))dt$$

with some σ_t-progressible measurable control $\hat{U}_t$ and an n-dimensional Brownian motion W_t^* adapted to σ_t. Moreover, $P(t)$ is the †error matrix satisfying the †Riccati equation, and H' is the transpose of H. Let $g(t,x)$ be the probability density of the normal distribution $N(0,P(t))$, and put $\tilde{L}(t,\hat{x},u)=\int L(t,x,u)\,g(t,x-\hat{x})dx$ and $\hat{\Psi}(\hat{x})=\int\Psi(x)g(T,x-\hat{x})dx$; then the problem turns into

$$Q(s,x)=\sup_{\hat{U}} E_{sx}\left[\int_s^T \tilde{L}(t,\hat{X}_t,\hat{U}_t)dt+\hat{\Psi}(\hat{X}_\tau)\right].$$

Recalling the SDE for $\hat{X}_t$, we can use the Bellman equation for choosing an optimal one.

D. Stochastic Maximum Principle

A stochastic version of †Pontryagin's maximum principle gives a necessary condition for optimality. This means that the instantaneous value of optimal control maximizes the stochastic analog of Pontryagin's Hamiltonian. Suppose that the system evolves according to an n-dimensional CSDE

$$dX_t=\alpha(X_t)dW_t+\gamma(X_t,U_t)dt.$$

The problem is to seek conditions on admissible control U_t such that $E_x[\int_0^T f(X_t,U_t)dt]$ is maximized, where T is a constant time. Assume that α, γ, and f are bounded and smooth. Define a Hamiltonian H on $\mathbf{R}^n\times\Gamma\times\mathbf{R}^n$ by $H(x,u,\Psi)=\gamma(x,u)\cdot\Psi+f(x,u)$. Let $\tilde{U}_t$ be optimal and $\tilde{X}_t$ its response starting at x. Then under some conditions there exist $\lambda\geqslant 0$ and $\mathscr{F}_t$-progressively measurable $q_{t,k}=(q_{t,k_1},q_{t,k_2},\ldots,q_{t,k_n})$ $(k=1,\ldots,n)$ and $\Psi_t=(\Psi_{t,1}\ldots\Psi_{t,n})$ which satisfy the SDE

$$d\Psi_{t,k}=-\left(\frac{\partial\gamma}{\partial x_k}(\tilde{X}_t,\tilde{U}_t)\cdot\Psi_t+\lambda\frac{\partial f}{\partial x_k}(\tilde{X}_t,\tilde{U}_t)\right)dt$$

$$+q_{t,k}dW_t,\quad k=1,\ldots,n,$$

and $H(\tilde{X}_t,\tilde{U}_t,\Psi_t)=\max_{u\in\Gamma}H(\tilde{X}_t,u,\Psi_t)$ a.e.

E. Optimal Stopping and Impulse Control

Suppose that X_t is an n-dimensional diffusion whose generator A is an elliptic differential operator. Let τ be a †stopping time. The optimal stopping problem is to seek a stopping time $\tilde{\tau}$ so that $E_x[g(X_\tau]$ is maximized, where g is nonnegative and continuous. $\tilde{\tau}$ is called optimal. The value function $V(x)=\sup_\tau E_x[g(X_\tau)]$ is characterized as the least †excessive majorant of g. Moreover, under some conditions V belongs to the domain of A and is the unique solution of the †free boundary problem; $V\geqslant g$, $AV\leqslant 0$, and $(V-g)\cdot AV=0$. Therefore, in the Hilbert space framework, the value function is related to the variational inequality. An optimal stopping time is provided by the hitting time for the set $\{x\mid V(x)=g(x)\}$.

Impulse control is a variant of the optimal stopping problem. At some moment (=stopping time) a controller shifts the current state to some other state. But not all shifts are allowed: State x can be shifted to a state of $x+[0,\infty)^n$. Let τ_k, $k=1$, 2, be a sequence of increasing stopping times and ξ_k be a $[0,\infty)^n$-valued $\sigma_{\tau_k}(X)$-measurable random variable. The sequence $U=\{\tau_1,\xi_1,\tau_2,\xi_2,\ldots\}$ is called an impulse control. U transfers the process X_t to

$$Y_t^U=X_0+\int_0^t\alpha(Y_s^U)dw_s+\int_0^t\gamma(Y_s^U)ds+\sum_{\tau_i\leqslant t}\xi_i,$$

if

$$A=\frac{1}{2}\sum_{i,j}\alpha\alpha^*(x)_{ij}\frac{\partial^2}{\partial x_i\partial x_j}+\sum_i\gamma_i(x)\frac{\partial}{\partial x_i},$$

and the problem is to seek U so as to maximize $E_x[\int_0^\tau e^{-\lambda t}f(Y_t^U)dt-\sum_{k=1}^\infty e^{-\lambda\tau_k}K(\xi_k)]$, where λ (>0) is constant and the function K $(\geqslant 0)$ stands for the cost of shifting. The value function is related to a quasivariational inequality.

F. General Description of Stochastic Filtering

The problem of estimating the original signal from data disturbed by noises is called a **stochastic filtering** problem. Let X_t, $t\in[0,T]$, be a continuous stochastic process with values in $\mathbf{R}^n$, called a **signal** (or **system**) process. It is transformed (or coded) to $h(t,X_t)$, where $h(t,x)$ is an m-vector-valued continuous function. Suppose that it is disturbed by a noise $\dot{W}_t$ and we observe $\dot{Y}_t=h(t,X_t)+\dot{W}_t$. Usually $\dot{W}_t$ is assumed to be a †white noise independent of X_t. Since the white noise is a generalized function, the integral of $\dot{Y}_t$, i.e.,

$$Y_t=\int_0^t h(s,X_s)ds+W_t,$$

is called an **observation process**, where W_t is an †m-dimensional Brownian motion independent of X_t. It is assumed for convenience that $|X_t|^2$ and $\int_0^t|h(s,X_s)|^2ds$ are integrable.

Assume that X_t is a 1-dimensional signal process. The least square estimation of X_t by nonlinear functions of observed data Y_s, $s\leqslant t$, is called a **nonlinear filter** of X_t and is denoted by $\hat{X}_t$. Let $\mathscr{F}_t$ or $\sigma(Y_s; s\leqslant t)$ be the least †σ-field for which Y_s, $s\leqslant t$, are measurable. Then the filter $\hat{X}_t$ is equal to an †$\mathscr{F}_t$-measurable random variable such that $E|X_t-\hat{X}_t|^2\leqslant E|X_t-Z|^2$ holds for any $\mathscr{F}_t$-measurable L^2 random variable Z. Hence it coincides with the †conditional expectation $E[X_t|\mathscr{F}_t]$. Now let H_t be the closed linear space spanned by Y_s, $s\leqslant t$. The least square estimation of X_t by elements of H_t, i.e., the orthogonal projection of X_t onto H_t, is called the **linear filter** of X_t and is denoted by $\tilde{X}_t$. Obviously, the mean square error of a nonlinear filter is less than or equal to that of a linear filter, but a linear filter is calculated more easily. If (X_t, Y_t) is a †Gaussian process, both filters coincide.

When X_t is an n-dimensional process $(X_t^1,\ldots,X_t^n)$, the n-vector process $\hat{X}_t=(\hat{X}_t^1,\ldots,\hat{X}_t^n)$ (or $\tilde{X}_t=(\tilde{X}_t^1,\ldots,\tilde{X}_t^n)$) is called the **nonlinear** (or **linear**) **filter** of X_t.

G. Kalman-Bucy Filter

Suppose that the signal process X_t is governed by a †linear stochastic differential equation (LSDE)

$$X_t=X_0+\int_0^t A(s)X_sds+\int_0^t B(s)d\tilde{W}_s,$$

where $A(s)$ (or $B(s)$) is an $n\times n$ (or $n\times r$) matrix-valued continuous function, $\tilde{W}_t$ is an r-dimensional Brownian motion independent of the noise W_t, and the initial data X_0 is a Gaussian random variable independent of W_t and $\tilde{W}_t$. Suppose further that $h(t,x)$ is linear, i.e., $h(t,x)=H(t)x$, where $H(t)$ is an $m\times n$-matrix-valued function. Then the joint process (X_t, Y_t) is Gaussian. Hence the nonlinear filter $\hat{X}_t$ coincides with the linear filter and satisfies

$$\hat{X}_t=E[X_0]+\int_0^t(A(s)-P(s)H(s)'H(s))\hat{X}_sds$$
$$+\int_0^t P(s)H(s)'dY_s,$$

where $H(s)'$ is the transpose of $H(s)$, and $P(t)=(P_{ij}(t))$ is the **error matrix** defined by $P_{ij}(t)=E(X_t^i-\hat{X}_t^i)(X_t^j-\hat{X}_t^j)$. It satisfies the **matrix Riccati equation**

$$\frac{dP(t)}{dt}=A(t)P(t)+P(t)A(t)'$$
$$-P(t)H(t)'H(t)P(t)+B(t)B(t)',$$
$$P(0)=\text{covariance of } X_0.$$

Let $\Phi(t,s)$ be the †fundamental solution of the linear differential equation $dx/dt=(A(t)-P(t)H(t)'H(t))x$. Then the solution $\hat{X}_t$ is represented by

$$\hat{X}_t=\Phi(t,0)E[X_0]+\int_0^t\Phi(t,s)P(s)H(s)'dY_s.$$

This algorithm is called the **Kalman-Bucy filter** [1]. Analogous results for discrete-time models have been obtained by Kalman.

H. Nonlinear Filter

In the study of nonlinear filters, the †conditional distribution $\pi(dx)=P(X_t\in dx|F_t)$ is considered besides $\hat{X}_t$. Suppose that X_t is governed by the SDE

$$X_t=X_0+\int_0^t a(s,X_s)ds+\int_0^t b(s,X_s)d\tilde{W}_s,$$

where $a(s,x)$ (or $b(s,x)$) is an n-vector ($n\times r$-matrix) valued Lipschitz continuous function. Then $\pi_t(f)=\int f(x)\pi_t(dx)$ satisfies the SDE

$$\pi_t(f)=E[f(X_0)]+\int_0^t\pi_s(Lf)ds$$
$$+\int_0^t(\pi_s(h_sf)-\pi_s(h_s)\pi_s(f))(dY_s-\pi_s(h_s)ds),$$

where $h_s(x)=h(s,x)$ and

$$Lf(x)=\sum_i a^i(s,x)\frac{\partial f}{\partial x^i}+\frac{1}{2}\sum_{i,j}\left(\sum_k b_{ik}(s,x)b_{jk}(s,x)\right)\frac{\partial^2 f}{\partial x^i \partial x^j}.$$

Under additional conditions on $a(s,x)$ and $b(s,x)$, $\pi_t(dx)$ has a density function $\pi_t(x)$, and it satisfies

$$\pi_t(x)=\pi_0(x)+\int_0^t L^*\pi_s(x)ds+\int_0^t \pi_s(x)(h_s(x)-\int\pi_s(x)h_s(x)dx)\left(dy_s-\left(\int\pi_s(x)h_s(x)dx\right)ds\right),$$

where L^* is the formal adjoint of L.

The process $I_t\equiv Y_t-\int_0^t\pi_s(h_s)ds$ is a Brownian motion such that $\sigma(I_s; s\leqslant t)\subset\mathscr{F}_t$ holds for any t. If $\sigma(I_s; s\leqslant t)=\mathscr{F}_t$ holds for all t, I_t is called the **innovation** of Y_t. The innovation property is not valid in general. A sufficient condition is that (X_t, Y_t) is a Gaussian process or $h(t,x)$ be a bounded function. However, in any case, †$\mathscr{F}_t$-adapted martingales are always represented as †stochastic integrals of the form $\sum_{i=1}^n\int_0^t f_s^i(\omega)dI_s^i$, where the f_s^i are $\mathscr{F}_s$-adapted processes.

I. Bayes Formula

Let C (or D) be the space of all continuous mappings x (or y) from $[0,T]$ into $\mathbf{R}^n$ (or $\mathbf{R}^m$) equipped with the uniform topology. x_t (y_t) is the value of x (y) at time t. Let $\mathscr{B}_t(C)$ be the least σ-field of C for which x_s, $s\leqslant t$, are measurable. $\mathscr{B}_t(D)$ is defined similarly. We denote by Φ_X, Φ_W, $\Phi_{X,Y}$ the †laws of processes X_t, W_t, and (X_t, Y_t), respectively. These are defined on $\mathscr{B}_T(C)$, $\mathscr{B}_T(D)$, and $\mathscr{B}_T(C)\otimes\mathscr{B}_T(D)$, respectively. Then $\Phi_{X,Y}$ is equivalent (mutually †absolutely continuous) to the product measure $\Phi_X\otimes\Phi_W$ on each $\mathscr{B}_t(C)\otimes\mathscr{B}_t(D)$. The †Radon-Nikodym density α_t of $\Phi_{X,Y}$ with respect to $\Phi_X\otimes\Phi_W$ is written as

$$\alpha_t(x,y)=\exp\left\{\sum_i\int_0^t h^i(s,x_s)dy_s^i-\frac{1}{2}\sum_i\int_0^t h^i(s,x_s)^2ds\right\},$$

where h^i and y^i are the corresponding components of vectors and dy_s^i denotes the †Ito integral.

The conditional distribution $\pi_t(dx)$ is computed by the Bayes formula:

$$\pi_t(f)=\frac{\rho_t(f)}{\rho_t(1)},\qquad \rho_t(f)=\int f(x_t)\alpha_t(x,Y)\Phi_X(dx),$$

where $Y=(Y_t; 0\leqslant t\leqslant T)$. Moreover, $\rho_t(f)$ satisfies the LSDE

$$\rho_t(f)=\rho_0(f)+\int_0^t\rho_s(Lf)ds+\int_0^t\rho_s(h_s)\rho_s(f)dI_s.$$

The density $\rho_t(x)$ (if x exists) satisfies

$$\rho_t(x)=\rho_0(x)+\int_0^t L^*\rho_s(x)ds+\int_0^t\rho_s(x)\left(\int\rho_s(y)h_s(y)dy\right)dI_s.$$

If $h(t,x)$ is a smooth function, then $\alpha_t(x,y)$ is continuous in y, so that $\pi_t(f)$ or $\rho_t(f)$ is a continuous functional of the observed data $(Y_s; s\leqslant t)$. Thus the filter π_t is a †robust statistic.

Remarks. (i) the signal and noise are not independent if the signal is controlled based on the observed data. In these cases, correction terms are sometimes needed for the SDE of the nonlinear filter. (ii) If the †sample paths of the signal process are not continuous, a similar SDE for a nonlinear filter is valid with L being replaced by some integrodifferential operator. If it is a †Markov chain with finite state, L is the generator of the chain. (iii) Several results are known for the case where the noise W_t is not a Brownian motion but a †Poisson process.

References

[1] W. F. Fleming and R. W. Rishel, Deterministic and stochastic optimal control, Springer, 1975.
[2] R. Bellman, Dynamic programming, Princeton Univ. Press, 1957.
[3] N. V. Krylov, Controlled diffusion processes, Springer, 1980. (Original in Russian, 1977.)
[4] M. Nisio, Stochastic control theory, Indian Statist. Inst. Lect. Notes, Macmillan, 1981.
[5] H. J. Kushner, On the stochastic maximum principle, fixed time of control, J. Math. Anal. Appl., 11 (1965), 78–92.
[6] A. N. Shiryaev, Optimal stopping rules, Springer, 1980. (Original in Russian, 1976.)
[7] A. Bensoussan and J. L. Lions, Sur la théorie du contrôle optimal. I, Temps d'arrêt; II, Contrôle impulsionnel, Hermann, 1977.
[8] R. S. Bucy and R. E. Kalman, New results in linear filtering and prediction theory, J. Basic Eng. ASME, (D) 83 (1961), 95–108.
[9] R. S. Bucy and D. D. Joseph, Filtering for stochastic processes with applications to guidance, Interscience, 1968.
[10] M. Fujisaki, G. Kallianpur, and H. Kunita, Stochastic differential equations for the nonlinear filtering problem, Osaka J. Math., 9 (1972), 19–40.

[11] R. S. Liptzer and A. N. Shiryaev, Statistics of stochastic processes. I, General theory; II, Applications, Springer, 1977, 1978. (Original in Russian, 1974.)
[12] G. Kallianpur, Stochastic filtering theory, Springer, 1980.

406 (XVII.14) Stochastic Differential Equations

A. Introduction

Stochastic differential equations were rigorously formulated by K. Itô [7] in 1942 to construct diffusion processes corresponding to Kolmogorov's differential equations. For this purpose he introduced the notion of stochastic integrals, and thus a differential-integral calculus for sample paths of stochastic processes was established. This theory, often called Itô's stochastic analysis or stochastic calculus, has brought an epoch-making method to the theory of stochastic processes. It provides us with a fundamental tool for describing and analyzing diffusion processes that we can apply effectively to limit theorems and to the probabilistic study of problems in analysis. It also plays an important role in the statistical theory of stochastic processes, such as †stochastic control or †stochastic filtering. Stochastic differential equations on manifolds provide a probabilistic method for differential geometry, sometimes called stochastic differential geometry. Recently, many interesting examples of infinite-dimensional stochastic differential equations have been introduced to describe probabilistic models in physics, biology, etc.

A unified theory of stochastic calculus has been developed in the framework of Doob's martingale theory and this, combined with Stroock and Varadhan's idea of martingale problems, provides an important method in the theory of stochastic processes (→ 262 Martingales).

B. Stochastic Integrals

As is well known, almost all sample paths of a Wiener process are continuous but nowhere differentiable (→ 45 Brownian Motion), and hence integrals with respect to these functions cannot be defined as the usual Stieltjes integrals. But these integrals can be defined by making use of the stochastic nature of Brownian motion. Wiener defined them (the Wiener integrals) for nonrandom integrands, but Itô defined them for a large class of random integrands. Itô's integrals have been extended in the martingale framework by H. Kunita and S. Watanabe, and by others [14, 19], as shown below.

Let $(\Omega, \mathscr{F}, P)$ be a probability space, and let $\mathbf{F}=\{\mathscr{F}_t\}_{t\geqslant 0}$ be an increasing family of σ-subfields of $\mathscr{F}$. Usually we assume that $\{\mathscr{F}_t\}$ is right continuous, i.e., $\mathscr{F}_{t+0}:=\bigcap_{\varepsilon>0}\mathscr{F}_{t+\varepsilon}=\mathscr{F}_t$ for every $t\geqslant 0$. Denote by $\mathscr{M}=\mathscr{M}(\mathbf{F})$ the totality of all continuous square-integrable martingales $X=(X_t)$ relative to $\{\mathscr{F}_t\}$; to be precise, X is an $\{\mathscr{F}_t\}$-martingale such that, with probability 1, $X_0=0$, $t\to X_t$ is continuous and $E(X_t^2)<\infty$ for every $t\geqslant 0$. We introduce the metric $\|X-Y\|=\sum_{k=1}^{\infty}2^{-n}\min(1,\|X_n-Y_n\|_2)$ on $\mathscr{M}$ where $\|\ \|_2$ stands for the $L_2(\Omega,P)$-norm. We always identify two stochastic processes $X=(X_t)$ and $Y=(Y_t)$ if sample functions $t\to X_t$ and $t\to Y_t$ coincide with probability 1. Then, by virtue of Doob's inequality $\|\max_{0\leqslant s\leqslant t}|X_s-Y_s|\,\|_2\leqslant 2\|X_t-Y_t\|_2$ (→ 262 Martingales), $\mathscr{M}$ becomes a complete metric vector space.

Next, by an **integrable increasing process** we mean a process $A=(A_t)$ with the following properties: (i) A is adapted to $\{\mathscr{F}_t\}$, i.e., A_t is $\mathscr{F}_t$-measurable for every $t\geqslant 0$; (ii) with probability 1, $A_0=0$, $t\to A_t$ is continuous and nondecreasing; (iii) A_t ($\geqslant 0$) is integrable for every $t\geqslant 0$, i.e., $E(A_t)<\infty$. We denote by $\mathscr{A}=\mathscr{A}(\mathbf{F})$ the totality of integrable increasing processes. We call a process $V=(V_t)$ an **integrable process of bounded variation** if V is expressed as $V_t=A_t^1-A_t^2$ with $A^1, A^2\in\mathscr{A}$. The totality of integrable processes of bounded variation is denoted by $\mathscr{V}=\mathscr{V}(\mathbf{F})$. It follows from the †Doob-Meyer decomposition theorem that, for every $M, N\in\mathscr{M}$, there exists a unique $V\in\mathscr{V}$ such that $M_tN_t-V_t$ is an $\{\mathscr{F}_t\}$-martingale. We denote this V as $\langle M,N\rangle$. In particular, $\langle M,M\rangle\in\mathscr{A}$, and it is denoted simply by $\langle M\rangle$. $\langle M,N\rangle$ is called the **quadratic variation process** because $\sum_{i=1}^{n}(M_{t_i}-M_{t_{i-1}})(N_{t_i}-N_{t_{i-1}})\to\langle M,N\rangle_t$ in probability as $|\Delta|\to 0$, where $\Delta: t_0=0<t_1<\ldots<t_n=t$ is a partition and $|\Delta|=\max_{1\leqslant i\leqslant n}|t_i-t_{i-1}|$. Brownian motion is the most important example of continuous square-integrable martingales, and this is characterized in our framework as follows. Suppose that a d-dimensional continuous $\{\mathscr{F}_t\}$-adapted process $X=(X_t^i)$ satisfies $M_t^i=X_t^i-X_0^i\in\mathscr{M}$ and $\langle M^i,M^j\rangle_t=\delta^{ij}t$, $i,j=1,2,\ldots,d$. Then X is a d-dimensional Brownian motion such that X_u-X_v, and the $\mathscr{F}_t$ are independent for every $u\geqslant v\geqslant t$. Such a Brownian motion is called an **$\{\mathscr{F}_t\}$-Brownian motion**, and a system of martingales $M^i\in\mathscr{M}$ having this property is often called a system of **$\{\mathscr{F}_t\}$-Wiener martingales**.

Now, we fix $M\in\mathscr{M}$. We denote by $\mathscr{L}_2(M)$ the totality of real, $\{\mathscr{F}_t\}$-adapted, and measurable processes $\Phi=(\Phi(t))$ such that $\|\Phi\|_{t,M}^2=$

$E[\int_0^t \Phi(s)^2 d\langle M\rangle_s] < \infty$ for every $t \geqslant 0$. Two $\Phi_1, \Phi_2 \in \mathscr{L}_2(M)$ are identified if $\|\Phi_1 - \Phi_2\|_{t,M} = 0$ for all $t \geqslant 0$. Since $\| \ \|_{t,M}$ is an L_2-norm on $[0,t] \times \Omega$ with respect to the measure $\mu_M(ds, d\omega) = d\langle M\rangle_s(\omega)P(d\omega)$, it is easy to see that $\mathscr{L}_2(M)$ is a complete metric vector space with the metric $\|\Phi - \Phi'\|_M = \sum_{n=1}^{\infty} 2^{-n}\min(1, \|\Phi - \Phi'\|_{n,M})$, $\Phi, \Phi' \in \mathscr{L}_2(M)$. If $\Phi = (\Phi(t))$ is given, for a partition $0 = t_0 < t_1 < \ldots < t_n \ldots \to \infty$ and $\mathscr{F}_{t_i}$-measurable bounded functions f_i, $i = 0, 1, \ldots$, by

$$\Phi(t,\omega) = \begin{cases} f_0(\omega), & t = 0, \\ f_{i-1}(\omega), & t_{i-1} < t \leqslant t_i, \quad i = 1, 2, \ldots, \end{cases}$$

then $\Phi \in \mathscr{L}_2(M)$ and the totality $\mathscr{L}_0$ of such processes are dense in $\mathscr{L}_2(M)$. If $\Phi \in \mathscr{L}_0$, we define $I^M(\Phi) = (I^M(\Phi)(t))_{t \geqslant 0}$ by

$$I^M(\Phi)(t) = \sum_{i=0}^{n-1} f_i(M_{t_{i+1}} - M_{t_i}) + f_n(M_t - M_{t_n}),$$
$$t_n \leqslant t \leqslant t_{n+1}.$$

Then $I^M(\Phi) \in \mathscr{M}$, and it holds also that $\langle I^M(\Phi), I^N(\Psi)\rangle = \int_0^t \Phi(s)\Psi(s) d\langle M, N\rangle_s$ for $M, N \in \mathscr{M}$ and $\Phi, \Psi \in \mathscr{L}_0$. In particular, $\|I^M(\Phi)(t)\|_2^2 = E[\langle I^M(\Phi)\rangle_t] = \|\Phi\|_{t,M}^2$, and hence $\|I^M(\Phi)\| = \|\Phi\|_M$. This implies that $\Phi \in \mathscr{L}_0 \subset \mathscr{L}_2(M) \to I^M(\Phi) \in \mathscr{M}$ is an isometric linear mapping, and hence it can be extended to $\mathscr{L}_2(M)$ uniquely, preserving the isometric property. $I^M(\Phi) \in \mathscr{M}$ is called the **stochastic integral** of $\Phi \in \mathscr{L}_2(M)$ by $M \in \mathscr{M}$. $I^M(\Phi)(t)$ is often denoted by $\int_0^t \Phi(s) dM_s$, and the random variable $I^M(\Phi)(t)$ obtained by fixing t is also called a stochastic integral.

The definition of stochastic integrals can be extended further by the following localization method. For an $\{\mathscr{F}_t\}$-†progressively measurable process $X = (X_t)$ and $\{\mathscr{F}_t\}$-stopping time σ, the stopped process $X^\sigma = (X_t^\sigma)$ is defined by $X_t^\sigma = X_{t \wedge \sigma}$ ($t \wedge \sigma = \min(t, \sigma)$). It follows from the †optional sampling theorem of Doob that $X^\sigma \in \mathscr{M}$ if $X \in \mathscr{M}$. For $\Phi = (\Phi(t)) \in \mathscr{L}_2(M)$ and an $\{\mathscr{F}_t\}$-stopping time σ, $\Phi_\sigma = (\Phi_\sigma(t))$ defined by $\Phi_\sigma(t) = I_{\{t \leqslant \sigma\}}\Phi(t)$ also belongs to $\mathscr{L}_2(M)$ and it holds that $I^M(\Phi_\sigma) = [I^M(\Phi)]^\sigma$. Keeping these facts in mind, we give the following definition. Let $\mathscr{M}^{\mathrm{loc}} = \{M = (M_t) \mid$ there exists a sequence of $\{\mathscr{F}_t\}$-stopping times σ_n such that $\sigma_n < \infty$, $\sigma_n \uparrow \infty$ as $n \to \infty$ a.s. and $M^{\sigma_n} \in \mathscr{M}$ for every $n = 1, 2, \ldots\}$. $\mathscr{A}^{\mathrm{loc}}$ and $\mathscr{V}^{\mathrm{loc}}$ are defined in a similar way. For $M, N \in \mathscr{M}^{\mathrm{loc}}$, $\langle M, N\rangle \in \mathscr{V}^{\mathrm{loc}}$ is defined to be the unique process in $\mathscr{V}^{\mathrm{loc}}$ such that $\langle M^{\sigma_n}, N^{\sigma_n}\rangle = \langle M, N\rangle^{\sigma_n}$ for a sequence of stopping times $\{\sigma_n\}$ as above, which can be chosen common to M and N. $\langle M, M\rangle$ is denoted by $\langle M\rangle$ as before. We fix $M \in \mathscr{M}^{\mathrm{loc}}$ and set $\mathscr{L}_2^{\mathrm{loc}}(M) = \{\Phi = (\Phi(t)) \mid$ a real, $\{\mathscr{F}_t\}$-adapted and measurable process such that, with probability one, $\int_0^t \Phi(s)^2 d\langle M\rangle_s < \infty$ for every $t \geqslant 0\}$.

For $\Phi \in \mathscr{L}_2^{\mathrm{loc}}(M)$, we can choose a sequence of $\{\mathscr{F}_t\}$-stopping times $\{\sigma_n\}$ such that $\sigma_n < \infty$, $\sigma_n \uparrow \infty$ a.s. and $M^{\sigma_n} \in \mathscr{M}$, $\Phi_{\sigma_n} \in \mathscr{L}_2(M^{\sigma_n})$, $n = 1, 2, \ldots$. For example, set $\sigma_n = \min[n, \inf\{t \mid \langle M\rangle_t + \int_0^t \Phi(s)^2 d\langle M\rangle_s \geqslant n\}]$. Then there exists an $I^M(\Phi) \in \mathscr{M}^{\mathrm{loc}}$ such that $I^M(\Phi)^{\sigma_n} = I^{M^{\sigma_n}}(\Phi_{\sigma_n})$ for $n = 1, 2, \ldots$, which is unique and independent of a particular choice of $\{\sigma_n\}$. $I^M(\Phi)$ is called the **stochastic integral** of $\Phi \in \mathscr{L}_2^{\mathrm{loc}}(M)$ by $M \in \mathscr{M}^{\mathrm{loc}}$. $I^M(\Phi)(t)$ is often denoted by $\int_0^t \Phi(s) dM_s$, and the random variable $I^M(\Phi)(t)$ obtained by fixing t is also called a stochastic integral.

Some of the basic properties of stochastic integrals are: (i) If $M \in \mathscr{M}^{\mathrm{loc}}$, $\Phi \in \mathscr{L}_2^{\mathrm{loc}}(M)$, and $\Psi \in \mathscr{L}_2^{\mathrm{loc}}(I^M(\Phi))$, then $\Phi\Psi \in \mathscr{L}_2^{\mathrm{loc}}(M)$ and $I^M(\Phi\Psi) = I^{I^M(\Phi)}(\Psi)$. (ii) If $M \in \mathscr{M}^{\mathrm{loc}}$ and $\Phi, \Psi \in \mathscr{L}_2^{\mathrm{loc}}(M)$, then for every $\alpha, \beta \in \mathbf{R}$ we have $\alpha\Phi + \beta\Psi = (\alpha\Phi(t) + \beta\Psi(t)) \in \mathscr{L}_2^{\mathrm{loc}}(M)$ and $I^M(\alpha\Phi + \beta\Psi) = \alpha I^M(\Phi) + \beta I^M(\Psi)$. Also if $M, N \in \mathscr{M}^{\mathrm{loc}}$ and $\Phi \in \mathscr{L}_2^{\mathrm{loc}}(M) \cap \mathscr{L}_2^{\mathrm{loc}}(N)$, then for $\alpha, \beta \in \mathbf{R}$ we have $\Phi \in \mathscr{L}_2^{\mathrm{loc}}(\alpha M + \beta N)$ and $I^{\alpha M + \beta N}(\Phi) = \alpha I^M(\Phi) + \beta I^N(\Phi)$. (iii) If $M, N \in \mathscr{M}^{\mathrm{loc}}$, $\Phi \in \mathscr{L}_2^{\mathrm{loc}}(M)$, and $\Psi \in \mathscr{L}_2^{\mathrm{loc}}(N)$, then $\Phi\Psi \in \mathscr{L}_1^{\mathrm{loc}}(\langle M, N\rangle)$ and $\langle I^M(\Phi), I^N(\Psi)\rangle_t = \int_0^t \Phi_s\Psi_s d\langle M, N\rangle_s$. Here, $\mathscr{L}_p^{\mathrm{loc}}(V)$ for $V \in \mathscr{V}^{\mathrm{loc}}$ is defined as follows: With probability 1, $s \to V_s$ is of bounded variation on every finite interval $[0,t]$, the total variation of which is denoted by $|\mathbf{V}|_t$. Then $|\mathbf{V}| \in \mathscr{A}^{\mathrm{loc}}$, and we define $\mathscr{L}_p^{\mathrm{loc}}(V)$ ($p \geqslant 1$) to be the totality of real, $\{\mathscr{F}_t\}$-adapted, and measurable processes $\Phi = (\Phi(t))$ such that, with probability1, $\int_0^t |\Phi(s)|^p d|V|_s < \infty$ for every $t \geqslant 0$. In particular, $\mathscr{L}_2^{\mathrm{loc}}(M) = \mathscr{L}_2^{\mathrm{loc}}(\langle M\rangle)$. (iv) If $M \in \mathscr{M}^{\mathrm{loc}}$, $\Phi \in \mathscr{L}_2^{\mathrm{loc}}(M)$, and σ is an $\{\mathscr{F}_t\}$-stopping time, then $I^M(\Phi_\sigma) = I^{M^\sigma}(\Phi) = I^{M^\sigma}(\Phi_\sigma) = [I^M(\Phi)]^\sigma$. (v) If $\Phi(t) = I_{\{\sigma < t\}} \cdot f$, where σ is an $\{\mathscr{F}_t\}$-stopping time and f is a bounded $\mathscr{F}_\sigma$-measurable random variable, then $\Phi \in \mathscr{L}_2^{\mathrm{loc}}(M)$ for every $M \in \mathscr{M}^{\mathrm{loc}}$ and $I^M(\Phi)(t) = f(M_t - M_{t \wedge \sigma})$. (vi) The definition of stochastic integrals is independent of the increasing family of σ-subfields in the following sense: If $\{\mathscr{G}_t\}$ is another family such that M belongs to the class $\mathscr{M}^{\mathrm{loc}}$ for both $\{\mathscr{F}_t\}$ and $\{\mathscr{G}_t\}$ and Φ belongs to the class $\mathscr{L}_2(M)$ for both $\{\mathscr{F}_t\}$ and $\{\mathscr{G}_t\}$, then $I^M(\Phi)$ is the same whether it is defined with respect to $\{\mathscr{F}_t\}$ or $\{\mathscr{G}_t\}$.

In particular, $N = I^M(\Phi)$, $M \in \mathscr{M}^{\mathrm{loc}}$, $\Phi \in \mathscr{L}_2^{\mathrm{loc}}(M)$, satisfies $\langle N, L\rangle_t = \int_0^t \Phi(s) d\langle M, L\rangle_s$ for all $L \in \mathscr{M}^{\mathrm{loc}}$. Conversely, $N \in \mathscr{M}^{\mathrm{loc}}$ having this property is unique, and hence it coincides with $I^M(\Phi)$. $I^M(\Phi) \in \mathscr{M}$ if and only if $\int_0^t \Phi(s)^2 d\langle M\rangle_s \in \mathscr{A}$.

The above definition of stochastic integrals can be extended with a slight technical modification to the case when M_t is not necessarily continuous [19]. Among such general stochastic integrals, a particularly important role is played by stochastic integrals describing point

processes, including Poisson point processes as an important special case. These stochastic integrals are important in the study of discontinuous processes including †Lévy processes; even in the study of continuous processes, such as diffusion, they provide an important tool for the treatment of excursions [4, 6, 8].

Let $(\Omega, \mathscr{F}, P)$ and $\{\mathscr{F}_t\}$ be as above. By a **continuous semimartingale** with respect to $\{\mathscr{F}_t\}$, or simply a **semimartingale** when there is no danger of confusion, we mean a process $X=(X(t))$ of the following form: $X(t)=X(0)+M(t)+V(t)$, where $X(0)$ is an $\mathscr{F}_0$-measurable random variable, $M=(M(t))\in\mathscr{M}^{\mathrm{loc}}$ and $V=(V(t))\in\mathscr{V}^{\mathrm{loc}}$. M and V are uniquely determined from X, and this decomposition is called the **semimartingale decomposition** of X. M is called the **martingale part** and V the **drift part** of the semimartingale X. A semimartingale X is often called an **Itô process** if $M(t)=\int_0^t\Phi(s)dB(s)$ and $V(t)=\int_0^t\Psi(s)ds$, where $B(t)\in\mathscr{M}$ is an $\{\mathscr{F}_t\}$-Brownian motion, $\Phi\in\mathscr{L}_2^{\mathrm{loc}}$, and $\Psi\in\mathscr{L}_1^{\mathrm{loc}}$. (When $V_t=t$, $\mathscr{L}_p^{\mathrm{loc}}(V)$ is denoted simply by $\mathscr{L}_p^{\mathrm{loc}}$.) A d-dimensional process whose components are semimartingales is called a d-dimensional semimartingale. The following formula, originally due to Itô and extended by Kunita and Watanabe, is of fundamental importance in stochastic calculus.

Itô's formula. Let $X(t)=(X^1(t),\dots,X^d(t))$ be a d-dimensional semimartingale and $X^i(t)=X^i(0)+M^i(t)+V^i(t)$ be the semimartingale decomposition of components. Let $F(x)=F(x^1,\dots,x^d)$ be a C^2-function defined on $\mathbf{R}^d$. Then $F(X(t))$ is also a semimartingale, and we have

$$F(X(t))=F(X(0))+\sum_{i=1}^d\int_0^t D_iF(X(s))dM^i(s)$$
$$+\sum_{i=1}^d\int_0^t D_iF(X(s))dV^i(s)$$
$$+\frac{1}{2}\sum_{i,j=1}^d\int_0^t D_iD_jF(X(s))d\langle M^i,M^j\rangle(s)$$

(where $D_i=\partial/\partial x^i$).

In others words, if $Y(t)=Y(0)+M(t)+V(t)$ is the semimartingale decomposition of $Y(t)=F(X(t))$, then $M(t)=\sum_{i=1}^d\int_0^t D_iF(X(s))dM^i(s)$ and $V(t)=\sum_{i=1}^d\int_0^t D_iF(X(s))dV^i(s)+1/2\sum_{i,j=1}^d\int_0^t D_iD_jF(X(s))d\langle M^i,M^j\rangle(s)$.

We now discuss other important transformations on semimartingales.

Time change. Let $A\in\mathscr{A}^{\mathrm{loc}}$, and assume further that with probability 1, $t\to A_t$ is strictly increasing and $\lim_{t\uparrow\infty}A_t=\infty$. Let $u\to C_u$ be the inverse function of $t\to A_t$, i.e., $C_u=\min\{t\mid A_t\geqslant u\}$. Then for every $u\geqslant 0$, C_u is an $\{\mathscr{F}_t\}$-stopping time. Set $\tilde{\mathscr{F}}_t=\mathscr{F}_{C_t}$, $t\geqslant 0$. For an $\{\mathscr{F}_t\}$-†progressively measurable process $X=(X(t))$, we define $X^A=(X^A(t))$ by $X^A(t)=X(C_t)$ and call it the **time change of** X **determined by** A. Then X^A is progressively measurable with respect to $\{\tilde{\mathscr{F}}_t\}$. If $X:X(t)=X(0)+M(t)+V(t)$ is a semimartingale with respect to $\{\mathscr{F}_t\}$, then X^A is a semimartingale with respect to $\{\tilde{\mathscr{F}}_t\}$, and its semimartingale decomposition is given by $X^A(t)=X(0)+M^A(t)+V^A(t)$. The mappings $M\to M^A$ and $V\to V^A$ are bijections between $\mathscr{M}^{\mathrm{loc}}$ and $\tilde{\mathscr{M}}^{\mathrm{loc}}$ and between $\mathscr{V}^{\mathrm{loc}}$ and $\tilde{\mathscr{V}}^{\mathrm{loc}}$, respectively, where $\tilde{\mathscr{M}}^{\mathrm{loc}}$ and $\tilde{\mathscr{V}}^{\mathrm{loc}}$ are defined relative to $\{\tilde{\mathscr{F}}_t\}$. Furthermore, $\langle M^A,N^A\rangle=\langle M,N\rangle^A$ for every $M,N\in\mathscr{M}^{\mathrm{loc}}$. Noting that $\Phi\in\mathscr{L}_2^{\mathrm{loc}}(M)$ can always be chosen $\{\mathscr{F}_t\}$-progressively measurable (in fact $\{\mathscr{F}_t\}$-†predictable), the mapping $\Phi\to\Phi^A$ defines a bijection between $\mathscr{L}_2^{\mathrm{loc}}(M)$ and $\mathscr{L}_2^{\mathrm{loc}}(M^A)$, and we have $I^{M^A}(\Phi^A)=[I^M(\Phi)]^A$.

Transformation of drift (Girsanov transformation). For $m\in\mathscr{M}^{\mathrm{loc}}$, set $D_m(t)=\exp[m_t-\frac{1}{2}\langle m\rangle_t]$. Then $D_m-1\in\mathscr{M}^{\mathrm{loc}}$, and if m satisfies a certain integrability condition (for example, $E(\exp[\frac{1}{2}\langle m\rangle_t])<\infty$ for every $t\geqslant 0$, in particular, $\langle m\rangle_t\leqslant ct$ for all t for some constant $c>0$), then D_m is a martingale, i.e., $E(D_m(t))=1$ for all $t\geqslant 0$. If $E(D_m(t))=1$ for all t, then there exists a probability $\tilde{P}$ on $(\Omega,\mathscr{F})$ (if $(\Omega,\mathscr{F})$ is a nice measurable space and $\mathscr{F}=\bigvee_{t\geqslant 0}\mathscr{F}_t$, which we can assume without loss of generality) such that $P(A)=E(D_m(t):A)$ for al' $A\in\mathscr{F}_t$, $t\geqslant 0$. Let X be a semimartingale with the decomposition $X(t)=X(0)+M(t)+V(t)$. On the probability space $(\Omega,\mathscr{F},\tilde{P})$ with the same family $\{\mathscr{F}_t\}$, X is still a semimartingale but its semimartingale decomposition is given by $X(t)=X(0)+\tilde{M}(t)+\tilde{V}(t)$, where $\tilde{M}(t)=M(t)-\langle M,m\rangle(t)$ and $\tilde{V}(t)=V(t)+\langle M,m\rangle(t)$. Furthermore, it holds that $\langle\tilde{M},\tilde{N}\rangle=\langle M,N\rangle$, M, $N\in\mathscr{M}^{\mathrm{loc}}$. This result is known as **Girsanov's theorem**. The transformation of probability spaces given above is called a **transformation of drift** or a **Girsanov transformation** since it produces a change as shown above of the drift part in the semimartingale decomposition.

In the discussion above, the increasing family $\{\mathscr{F}_t\}$ was fixed. It is also important to study how the semimartingale character changes under a changing increasing family [12].

C. Stochastic Differentials

In this section, we introduce stochastic differentials of semimartingales and rewrite the results in the previous section in more convenient form. Let $(\Omega,\mathscr{F},P)$ and $\{\mathscr{F}_t\}$ be as above and $\mathscr{M}$, $\mathscr{A}$, $\mathscr{V}$, $\mathscr{M}^{\mathrm{loc}}$, $\mathscr{A}^{\mathrm{loc}}$, $\mathscr{V}^{\mathrm{loc}}$ be defined as in Section B. By $\mathscr{Q}$ we denote the totality of continuous semimartingales relative to $\{\mathscr{F}_t\}$.

For $X \in \mathscr{Q}$, let $X(t)=X(0)+M_X(t)+V_X(t)$ be the semimartingale decomposition. We write formally $X(t)-X(0)=\int_0^t dX(s)$ and call dX (denoted also by dX_t or $dX(t)$) the **stochastic differential** of X. To be precise, dX can be considered as a random interval function $dX(I)=X(t)-X(s)$, $I=(s,t]$ or the equivalence class containing X under the equivalence relation $X \sim Y$ on $\mathscr{Q}$ defined by $X \sim Y$ if and only if $X(t)-X(0)=Y(t)-Y(0)$, $t \geqslant 0$. For X, $Y \in \mathscr{Q}$ and $\alpha, \beta \in \mathbf{R}$, $\alpha dX+\beta dY$ is defined by $d(\alpha X+\beta Y)$ and $dX \cdot dY$ by $d\langle M_X, M_Y\rangle$. Let $d\mathscr{Q}$ be the totality of stochastic differentials of elements in $\mathscr{Q}$ and $d\mathscr{M}$ and $d\mathscr{V}$ be that of elements in $\mathscr{M}^{\mathrm{loc}}$ and $\mathscr{V}^{\mathrm{loc}}$, respectively. $d\mathscr{Q}$ is a commutative algebra under the operations just introduced. Note that $dX \cdot dY \in d\mathscr{V}$ and that $dX \cdot dY=0$ if either of dX and dY is in $d\mathscr{V}$. In particular, $dX \cdot dY \cdot dZ=0$ for every dX, dY, and dZ. Let $\mathscr{B}$ be the totality of $\{\mathscr{F}_t\}$-progressively measurable processes $\Phi=(\Phi(t))$ such that, with probability 1, $\sup_{0 \leqslant s \leqslant t}|\Phi(s)| < \infty$ for every $t \geqslant 0$. Noting that $\mathscr{B} \subset \mathscr{L}_2^{\mathrm{loc}}(M)$ for any $M \in \mathscr{M}^{\mathrm{loc}}$, we define $\Phi \cdot dX \in d\mathscr{Q}$ for $\Phi \in \mathscr{B}$ and $X \in \mathscr{Q}$ to be the stochastic differential of the semimartingale $\int_0^t \Phi(s) dM_X(s)+\int_0^t \Phi(s) dV_X(s)$. $\Phi \cdot dX$ is uniquely determined by Φ and dX. Itô's formula is stated, in this context, as follows: For $X=(X^1, \ldots, X^d)$, $X^i \in \mathscr{Q}$, and $F: \mathbf{R}^d \rightarrow \mathbf{R}$, which is of class C^2, $F(X) \in \mathscr{Q}$, and

$$dF(X)=\sum_{i=1}^d D_i F(X) \cdot dX^i + \frac{1}{2} \sum_{i,j=1}^d D_i D_j F(X) \cdot dX^i \cdot dX^j.$$

We now define another important operation on the space $d\mathscr{Q}$. Noting that $\mathscr{Q} \subset \mathscr{B}$, we define $X \circ dY$ for X, $Y \in \mathscr{Q}$ by

$$X \circ dY = X \cdot dY + \frac{1}{2} dX \cdot dY.$$

This is uniquely determined from X and dY, and is called the **symmetric multiplication** of X and dY. It is also called a **stochastic differential of the Stratonovich type** or **Itô's circle operation** since the notation was introduced by Itô [9]. $\int_0^t X \circ dY$ is called the **stochastic integral of the Stratonovich type**, whereas $\int_0^t X \cdot dY$ is that of the **Itô type**. Under this operation, Itô's formula is rewritten as follows: For $X=(X^1, \ldots, X^d)$, $X^i \in \mathscr{Q}$, and $F: \mathbf{R}^d \rightarrow \mathbf{R}$, which is of class C^3, $F(X)$, $D_i F(X) \in \mathscr{Q}$, and

$$dF(X)=\sum_{i=1}^d D_i F(X) \circ dX^i.$$

This chain rule for stochastic differentials takes the same form as in the ordinary calculus. For this reason symmetric multiplication plays an important role in transferring notions used in ordinary calculus into stochastic calculus and in defining intrinsic (i.e., coordinate-free) notions probabilistically. In particular, it is fundamental to the study of stochastic differential equations on manifolds (→ Section G).

D. Stochastic Differential Equations

Here, we give a general formulation of stochastic differential equations in which the infinitesimal change of the system may depend on the past history of the system; however, equations of **Markovian type**, in which the infinitesimal change of the system depends only on the present state of the system, are considered in most cases. Let W^d be the space of d-dimensional continuous paths: $W^d=C([0,\infty) \rightarrow \mathbf{R}^d)$:= the totality of all continuous functions $w:[0,\infty) \rightarrow \mathbf{R}^d$, endowed with the topology of the uniform convergence on finite intervals and $\mathscr{B}(W^d)$ be the topological σ-field. For each $t \geqslant 0$, define $\rho_t: W^d \rightarrow W^d$ by $(\rho_t w)(s)=w(t \wedge s)$, and let $\mathscr{B}_t(W^d)=\rho_t^{-1}(\mathscr{B}(W^d))$, $t \geqslant 0$. Let $\mathscr{A}^{d,r}$ be the totality of functions $\alpha(t,w)=(\alpha_j^i(t,w))$: $[0,\infty) \times W^d \rightarrow \mathbf{R}^d \otimes \mathbf{R}^r$ (:= the totality of $d \times r$ real matrices) such that each component $\alpha_j^i(x,w)$ $(i=1,2,\ldots,d; j=1,2,\ldots,r)$ is $\mathscr{B}([0,\infty)) \times \mathscr{B}(W^d)$-measurable and $\mathscr{B}_t(W^d)$-measurable for each fixed $t \geqslant 0$. In general, $\alpha_j^i(t,w)$ is called **nonanticipative** if it satisfies the second property above. An important case of $\alpha \in \mathscr{A}^{d,r}$ is when it is given as $\alpha(t,w)=\sigma(t,w(t))$ by a Borel function $a:[0,\infty) \times \mathbf{R}^d \rightarrow \mathbf{R}^d \otimes \mathbf{R}^r$. In this case, α is called **independent of the past history** or **of Markovian type**. For a given $\alpha \in \mathscr{A}^{d,r}$ and $\beta \in \mathscr{A}^{d,1}$, we consider the following stochastic differential equation:

$$(1) \quad dX^i(t)=\sum_{j=1}^r \alpha_j^i(t,X) dB^j(t)+\beta^i(t,X)dt, \qquad i=1,2,\ldots,d,$$

also denoted simply as

$$dX(t)=\alpha(t,X)dB(t)+\beta(t,X)dt.$$

Here $X(t)=(X^1(t), \ldots, X^d(t))$ is a d-dimensional continuous process. $B(t)=(B^1(t), \ldots, B^r(t))$ is a r-dimensional Brownian motion with $B(0)=0$. A precise formulation of equation (1) is as follows. $X=(X(t))$ is called a **solution** of equation (1) if it satisfies the following conditions: (i) X is a d-dimensional, continuous, and $\{\mathscr{F}_t\}$-adapted process defined on a probability space $(\Omega, \mathscr{F}, P)$ with an increasing family $\{\mathscr{F}_t\}$, i.e., $X:\Omega \rightarrow W^d$ which is $\mathscr{F}_t/\mathscr{B}_t(W^d)$-measurable for every $t \geqslant 0$; (ii) $\alpha_j^i(t,X) \in \mathscr{L}_2^{\mathrm{loc}}$, $\beta^i(t,X) \in \mathscr{L}_1^{\mathrm{loc}}$, $i=1,\ldots,d$, $j=1,\ldots,r$ (→ Section B for the definition of $\mathscr{L}_p^{\mathrm{loc}}$); (iii) there exists an r-

dimensional $\{\mathscr{F}_t\}$-Brownian motion $B(t)$ with $B(0)=0$ such that the equality

$$X^i(t)-X^i(0)=\sum_{j=1}^{r}\int_0^t \alpha_j^i(s,X)\,dB^j(s)$$
$$+\int_0^t \beta^i(s,X)\,ds, \qquad i=1,2,\ldots,d,$$

holds with probability 1.

Thus a solution X is always accompanied by a Brownian motion B. To emphasize this, we often call X a solution with the Brownian motion B or call the pair (X,B) itself a solution of (1). In the above definition, a solution is given with reference to an increasing family $\{\mathscr{F}_t\}$. The essential point is that σ-fields $\sigma(B(u)-B(v); u\geqslant v\geqslant t)$ and $\sigma(X(s),B(s); 0\leqslant s\leqslant t)$ are independent for every t: If X satisfies the conditions of solutions stated above, then the specified independence is obvious, and conversely, if this independence is satisfied, then by setting $\mathscr{F}_t=\bigcap_{\varepsilon>0}\sigma(X(s),B(s);0\leqslant s\leqslant t+\varepsilon)$, the conditions of solutions stated above are satisfied. But it is usually convenient to introduce some increasing family $\{\mathscr{F}_t\}$ into the definition of solutions as above. When α and β are of the Markovian type, $\alpha(t,w)=\sigma(t,w(t))$, $\beta(t,w)=b(t,w(t))$, the corresponding equation

(2) $dX(t)=\sigma(t,X(t))\,dB(t)+b(t,X(t))\,dt$

is called a **stochastic differential equation of Markovian type**. Furthermore, if $\sigma(t,x)$ and $b(t,x)$ are independent of t, i.e., $\sigma(t,x)=\sigma(x)$ and $b(t,x)=b(x)$, the equation

(3) $dX(t)=\sigma(X(t))\,dB(t)+b(X(t))\,dt$

is called a stochastic differential equation of **time homogeneous** (or **time-independent**) **Markovian type**.

Next, we define the notions of the uniqueness of solutions. There are two kinds of uniqueness: uniqueness in the sense of law (in distribution) and pathwise uniqueness. When we consider the stochastic differential equations as a means to determine the laws of continuous stochastic processes, uniqueness in the sense of law is sufficient. If, on the other hand, we regard the stochastic differential equation as a means to define the sample paths of solutions as a functional of the accompanying Brownian motion, i.e., if we regard the equation as a machine that produces a solution as an output when we input a Brownian motion, the notion of pathwise uniqueness is more natural and more important. As we shall see, this notion is closely related to the notion of **strong solutions**.

These notions are defined as follows. For a solution $X=(X(t))$ of (1), $X(0)$ is called the **initial value**, its law on $\mathbf{R}^d$ is called the **initial law (distribution)**, and the law of X on W^d is called the **law (distribution) of** X. We say that the **uniqueness in the sense of law of solutions** for (1) holds if the law of any solution X is uniquely determined by its initial law, i.e., if whenever X and X' are two solutions whose initial laws coincide, then the laws of X and X' coincide. In this definition, we restrict ourselves to the solutions whose initial values are nonrandom, i.e., the initial laws are δ-distributions at some points in $\mathbf{R}^d$. Next, we say that the **pathwise uniqueness of solutions** for (1) holds if whenever X and X' are any two solutions defined on the same probability space $(\Omega,\mathscr{F},P)$ with the same increasing family $\{\mathscr{F}_t\}$ and the same r-dimensional $\{\mathscr{F}_t\}$-Brownian motion such that $X(0)=X'(0)$ a.s., then $X(t)=X'(t)$ for all $t\geqslant 0$ a.s. In this definition also, the solutions can be restricted to those having nonrandom initial values.

We say that equation (1) has a **unique strong solution** if there exists a function $F(x,w):\mathbf{R}^d\times W_0^r\to W^d$ ($W_0^r=\{w\in W^r\mid w(0)=0\}$) such that th- following are true: (i) For any solution $(X,$ ' , of (1), $X=F(X(0),B)$ holds a.s.; (ii) for any $\mathbf{R}^d$-valued random variable $X(0)$ and an r-dimensional Brownian motion $B=(B(t))$ with $B(0)=0$ which are mutually independent, $X=F(X(0),B)$ is a solution of (1) with the Brownian motion B and the initial value $X(0)$. If this is the case, $F(x,w)$ itself is a solution of (1) with the initial value x, and with respect to the canonical Brownian motion $B(t,w)=w(t)$ on the r-dimensional Wiener space $(W_0^r,\mathscr{F},P)$, $\mathscr{F}$ is the completion of $\mathscr{B}(W_0^r)$ with respect to the r-dimensional Wiener measure P. If equation (1) has a unique strong solution, then it is clear that pathwise uniqueness holds. Conversely, if pathwise uniqueness holds for (1) and if a solution exists for any given initial law, then equation (1) has a unique strong solution, [6, 25].

The existence of solutions was discussed by A. V. Skorokhod [20]. If the coefficients α and β are bounded and continuous on $[0,\infty)\times W^d$, a solution of (1) exists for any given initial law. This is shown as follows [6]. We first construct approximate solutions by Cauchy's polygonal method and then show that their probability laws are †tight. A limit process in the sense of probability law can be shown to be a solution. The assumption of boundedness above can be weakened, e.g., to the following condition: For every $T>0$, a constant $K_T>0$ exists such that

(4) $$\|\alpha(t,w)\|+\|\beta(t,w)\|\leqslant K_T(1+\|w\|_t),$$
$$t\in[0,T],\quad w\in W^d.$$

Here $\|w\|_t=\max_{0\leqslant s\leqslant t}|w(s)|$. In the case of the

Markovian equation (2), it is sufficient to assume that $\sigma(t,x)$ and $b(t,x)$ are continuous:

(5) $\|\sigma(t,x)\|+\|b(t,x)\|\leqslant K_T(1+|x|)$,

$$t\in[0,T],\quad x\in\mathbf{R}^d.$$

If these conditions are violated, a solution $X(t)$ does not exist globally in general but exists up to a certain time e, called the **explosion time**, such that $\lim_{t\uparrow e}|x(t)|=\infty$ if $e<\infty$. To extend the notion of solutions in such cases, we have to replace the path space W^d by the space $\hat{W}^d$ that consists of all continuous functions $w:[0,\infty)\rightarrow\hat{\mathbf{R}}^d$ $(=\mathbf{R}^d\cup\{\Delta\}=$ the one-point compactification) satisfying $w(t)=\Delta$ for every $t\geqslant e(w)$ $(=\inf\{t\mid w(t)=\Delta\})$.

Now, we list some results on the uniqueness of solutions. First consider the equations of the Markovian type (2), and assume that the coefficients are continuous and satisfy the condition (5). (i) If σ, b are Lipschitz continuous, i.e., for every $N>0$ there exists a constant K_N such that $\|\sigma(t,x)-\sigma(t,y)\|+\|b(t,x)-b(t,y)\|\leqslant K_N|x-y|$, $t\in[0,T]$, $x,y\in B_N:=\{z\in R^d\mid|z|\leqslant N\}$, then the pathwise uniqueness of solutions holds for equation (2). Thus the unique strong solution of (2) exists, and this is constructed directly by Picard's successive approximation (Itô [7,8]). (ii) If $d=1$, σ is Hölder continuous with exponent 1/2 and b is Lipschitz continuous, i.e., for every $N>0$, K_N exists such that

$$|\sigma(t,x)-\sigma(t,y)|^2+|b(t,x)-b(t,y)|\leqslant K_N|x-y|,$$

$$t\in[0,N],\quad x,y\in B_N,$$

then the pathwise uniqueness of solutions holds for equation (2) (T. Yamada and Watanabe [25]). (iii) If the matrix $a(t,x)=\sigma(t,x)\sigma(t,x)^*$ (i.e., $a^{ij}(t,x)=\sum_{k=1}^r\sigma_k^i(t,x)\sigma_k^j(t,x)$) is strictly positive definite, then the uniqueness in the sense of law of the solution for (2) holds (D. W. Stroock and S. R. S. Varadhan [21]). (iv) An example of stochastic differential equations for which the uniqueness in the sense of law holds but the pathwise uniqueness does not hold was given by H. Tanaka as follows: $d=r=1$, $b(t,x)\equiv0$ and $\sigma(t,x)=I_{\{x\geqslant0\}}-I_{\{x<0\}}$. Another example in the non-Markovian cases was given by B. S. Tsirel'son (see below).

Next, consider non-Markovian equations of the following form:

(6) $dX(t)=dB(t)+\beta(t,X)dt$;

i.e., the case $d=r$ and $\alpha(t,w)=I$ (identity matrix). Assume further that $\beta\in\mathscr{A}^{d,1}$ is bounded. Then a solution of (6) exists for any given initial distribution, unique in the sense of law, and it can be constructed by the Girsanov transformation of Section B as follows. On a suitable probability space $(\Omega,\mathscr{F},P)$ with an increasing family $\{\mathscr{F}_t\}$ such that $\mathscr{F}=\bigvee_{t>0}\mathscr{F}_t$, we set up an $\mathscr{F}_0$-measurable, d-dimensional random variable $X(0)$ with a given law and a d-dimensional $\{\mathscr{F}_t\}$-Brownian motion $\tilde{B}=(\tilde{B}(t))$ such that $\tilde{B}(0)=0$. Set $X(t)=X(0)+\tilde{B}(t)$ and $M(t)=\exp[\int_0^t\beta(s,X)d\tilde{B}(s)-\frac{1}{2}\int_0^t|\beta(s,X)|^2ds]$. Then $M(t)$ is an $\{\mathscr{F}_t\}$-martingale, and the probability $\tilde{P}$ on $(\Omega,\mathscr{F})$ is determined by $\tilde{P}(A)=E(M_t;A)$, $A\in\mathscr{F}_t$. By Girsanov's theorem, $B(t)=X(t)-X(0)-\int_0^t\beta(s,X)ds$ is a d-dimensional $\{\mathscr{F}_t\}$-Brownian motion on $(\Omega,\mathscr{F},\tilde{P})$, and hence (X,B) is a solution of (6). Any solution is given in this way and hence the uniqueness in the sense of law holds. But the pathwise uniqueness does not hold in general; an example was given by Tsirel'son [1,6] as follows. Let $\{t_n\}$ be a sequence such that $0<\ldots<t_n<t_{n-1}<t_0=1$ and $\lim_{n\uparrow\infty}t_n=0$. Set

$$\beta(t,w)=\begin{cases}0, & t\geqslant t_0 \text{ and } t=0,\\ \theta\left(\dfrac{w(t_{i+1})-w(t_{i+2})}{t_{i+1}-t_{i+2}}\right), & \\ & t\in[t_{i+1},t_i),\quad i=0,1,2,\ldots,\end{cases}$$

where $\theta(x)=x-[x]$, $x\in R$, is the decimal part of x.

Time changes (→ Section B) are also used to solve some stochastic differential equations [6].

E. Stochastic Differential Equations and Diffusion Processes

In this section we consider equations of time-independent Markovian type (3) only. The time-dependent case can be reduced to the time-independent case by adding one more component $X^{d+1}(t)$ such that $dX^{d+1}(t)=dt$. Further, we assume that coefficients $\sigma(x)\in\mathbf{R}^d\otimes\mathbf{R}^r$ and $b(x)\in\mathbf{R}^d$ are continuous on $\mathbf{R}^d$ and the uniqueness in the sense of law of solutions holds. Let P_x, $x\in\mathbf{R}^d$, be the law on W^d, or on $\hat{W}^d$ if there is an explosion, of a solution with the initial law δ_x (= the unit measure at x). Then $\{P_x\}$ possesses the †strong Markov property with respect to $\{\mathscr{F}_t\}$, where $\mathscr{F}_t$ is a suitable completion of $\mathscr{B}_t(W^d)$ or $\mathscr{B}_t(\hat{W}^d)$, and hence $(W^d,\{\mathscr{F}_t\},P_x)$ or $(\hat{W}^d,\{\mathscr{F}_t\},P_x)$ is a diffusion process †on $\mathbf{R}^d$ (→ 115 Diffusion Processes, 261 Markov Processes).

Let A be the differential operator

$$A=\frac{1}{2}\sum_{i,j=1}^d a^{ij}(x)D_iD_j+\sum_{i=1}^d b^i(x)D_i\qquad(D_i=\partial/\partial x^i)$$

with the domain $C_0^2(\mathbf{R}^d)$ (= the totality of C^2-functions on $\mathbf{R}^d$ with compact supports), where $a^{ij}(x)=\sum_{k=1}^r\sigma_k^i(x)\sigma_k^j(x)$. By Itô's formula,

(7) $f(w(t))-f(w(0))-\int_0^t(Af)(w(s))ds$

is an $\{\mathcal{F}_t\}$-martingale for every $f \in C_0^2(\mathbf{R}^d)$ (we set $f(\Delta)=0$), and this property characterizes the diffusion process. The diffusion is generated by the operator A in this sense. Furthermore, if for some $\lambda > 0$, $(\lambda - A)(C_0^2(\mathbf{R}^d))$ is a dense subset of $C_\infty(\mathbf{R}^d)$ (= the totality of continuous functions f on $\mathbf{R}^d$ such that $\lim_{|x|\to\infty} f(x)=0$) then the †transition semigroup of the diffusion is a †Feller semigroup on $C_\infty(\mathbf{R}^d)$, and its infinitesimal generator A is the closure of $(A, C_0^2(\mathbf{R}^d))$. Hence $u(t,x)= E_x[f(w(t))]$, $f \in C_0^2(R^d)$, is the unique solution of the evolution equation $du/dt = Au$, $u|_{t=0}=f$. Generally, if the coefficients σ and b are sufficiently smooth, we can show, by using the stochastic differential equation (3), that $u(t,x)$ is also smooth for a smooth f and satisfies the heat equation $\partial u/\partial t = Au$. Taking the expectations in (7), we have the relation $E_x[f(w(t))] = f(x) + \int_0^t E_x[Af(w(s))]\,ds$, which implies that the transition probability $P(t,x,dy)$ of the diffusion satisfies the equation $\partial p/\partial t = A^* p$ in (t,y) in a weak sense, where A^* is the adjoint operator of A. If $\partial/\partial t - A^*$ is †hypoelliptic, we can conclude that $P(t,x,dy)$ possesses a smooth density $p(t,x,y)$ by appealing to the theory of partial differential equations. Recently, P. Malliavin showed that a probabilistic method based on the stochastic differential equations can also be applied to this problem effectively, [6, 16, 17].

If $c(t,x)$ is continuous and $v(t,x)$ is sufficiently smooth in (t,x) on $[0,\infty)\times\mathbf{R}^d$, then the following fact, more general than (7), holds:

(8) $$v(t,w(t))\exp\left[\int_0^t c(s,w(s))\,ds\right] - v(0,x) - \int_0^t \exp\left[\int_0^s c(u,w(u))\,du\right] \times (\partial v/\partial t + (A+c)v)(s,w(s))\,ds$$

is a local martingale (i.e., $\in \mathcal{M}^{\mathrm{loc}}$) with respect to $\{\mathcal{F}_t, P_x\}$. By applying the †optional sampling theorem to (8) for a class of $\{\mathcal{F}_t\}$-stopping times, we can obtain the probabilistic representation in terms of the diffusion of solutions for initial or boundary value problems related to the operator A [3, 4].

F. Stochastic Differential Equations with Boundary Conditions

As we saw in the previous section, diffusion processes generated by differential operators can be constructed by stochastic differential equations. A diffusion process on a domain with boundary is generated by a differential operator that describes the behavior of the process inside the domain, and a boundary condition that describes the behavior of the process on the boundary of the domain. For example, consider a reflecting †Brownian motion on the half-line $[0,\infty)$. This is a diffusion process $X=(X_t)$ on $[0,\infty)$ obtained by setting $X_t = |x_t|$ from a 1-dimensional Brownian motion x_t. The corresponding differential operator is $A = \frac{1}{2}d^2/dx^2$, and the boundary condition is $Lu \equiv du/dx|_{x=0} = 0$, that is, the transition expectation $u(t,x) = E_x[f(X_t)]$ is determined by $\partial u/\partial t = Au$, $Lu=0$, and $u|_{t=0} = f$. In constructing such diffusion processes with boundary conditions, stochastic differential equations can be used effectively. In the case of reflecting Brownian motion, it was formulated by Skorokhod in the form

(9) $$dX(t) = dB(t) + d\varphi(t).$$

Here $B(t)$ is a 1-dimensional Brownian motion ($B(0)=0$), $X(t)$ is a continuous process such that $X(t) \geqslant 0$, and $\varphi(t)$ has the following property with probability 1: $\varphi(0)=0$, $t \to \varphi(t)$ is continuous and nondecreasing and increases only on such t that $X(t)=0$, i.e., $\int_0^t I_{\{0\}}(X(s))\,d\varphi(s) = \varphi(t)$. Given a Brownian motion $B(t)$ and a nonnegative random variable $X(0)$ which are mutually independent, $X(t)$ satisfying (9) and with the initial value $X(0)$ is unique and given by $X(t) = X(0) + B(t)$, $t < \sigma_0 = \min\{t \mid X(0) + B(t) = 0\}$ and $X(t) = B(t) - \min_{\sigma_0 \leqslant s \leqslant t} B(s)$, $t \geqslant \sigma_0$ (P. Lévy, Skorokhod; → [6, 18]).

In the case of multidimensional processes, possible boundary conditions were determinea by A. D. Venttsel' [24]. Stochastic differential equations describing these diffusions were formulated by N. Ikeda [5] in the 2-dimensional case and by Watanabe [23] in the general case as follows. Let D be the upper half-space $\mathbf{R}_+^d = \{x = (x^1, \ldots, x^d) \mid x^d \geqslant 0\}$, $\partial D = \{x \mid x^d = 0\}$, and $\mathring{D} = \{x \mid x^d > 0\}$. The general case can be reduced, at least locally, to this case. Suppose that the following system of functions is given: $\sigma(x): D \to \mathbf{R}^d \times \mathbf{R}^r$, $b(x): D \to \mathbf{R}^d$, $\tau(x): \partial D \to \mathbf{R}^{d-1} \times \mathbf{R}^s$, $\beta(x): \partial D \to \mathbf{R}^{d-1}$, and $\rho(x): \partial D \to [0,\infty)$, which are all bounded and continuous. Consider the following stochastic differential equation:

(10) $$\begin{cases} dX^i(t) = \sum_{j=1}^r \sigma_j^i(X(t)) I_{\mathring{D}}(X(t))\,dB^j(t) \\ \qquad + b^i(X(t)) I_{\mathring{D}}(X(t))\,dt \\ \qquad + \sum_{k=1}^s \tau_k^i(X(t)) I_{\partial D}(X(t))\,dM^k(t) \\ \qquad + \beta^i(X(t))\,d\psi(t), \\ \qquad\qquad i = 1, 2, \ldots, d-1, \\ dX^d(t) = \sum_{j=1}^r \sigma_j^d(X(t)) I_{\mathring{D}}(X(t))\,dB^j(t) \\ \qquad + b^d(X(t)) I_{\mathring{D}}(X(t))\,dt + d\varphi(t), \\ I_{\partial D}(X(t))\,dt = \rho(X(t))\,d\varphi(t). \end{cases}$$

By a solution of this equation, we mean a system of continuous semimartingales $\mathfrak{X}=(X(t), B(t), M(t), \varphi(t))$ over a probability space $(\Omega, \mathcal{F}, P)$ with an increasing family $\{\mathcal{F}_t\}$ satisfying the following conditions: (i) $X(t)=(X^1(t), \ldots, X^d(t))$ is D-valued, i.e., $X^d(t)\geqslant 0$; (ii) with probability 1, $\varphi(0)=0$, $t\to\varphi(t)$ is nondecreasing, and $\int_0^t I_{\partial D}(X(s))\,d\varphi(s)=\varphi(t)$; (iii) $B(t)$ and $M(t)$ are r-dimensional and s-dimensional systems of elements in $\mathcal{M}^{\mathrm{loc}}$, respectively, such that $\langle B^i, B^j\rangle_t=\delta^{ij}t$, $\langle B^i, M^m\rangle_t=0$, and $\langle M^m, M^n\rangle_t=\delta^{mn}\varphi(t)$, $i, j=1, \ldots, r$, $m, n=1, \ldots, s$; and finally (iv) the stochastic differentials of these semimartingales satisfy (10).

The processes $B(t)$, $M(t)$, and $\varphi(t)$ are subsidiary, and the process $X(t)$ itself is often called a solution. We say that the uniqueness of solution holds if the law of $X=(X(t))$ is uniquely determined from the law of $X(0)$. As before, the existence and the uniqueness of solutions imply that solutions define a diffusion process on D, and these are guaranteed if, for example, $\min_{x\in\partial D} a^{dd}(x)>0$ and σ, b, τ, β are Lipschitz continuous, [6, 23]. Here, we set $a^{ij}(x)=\sum_{k=1}^r \sigma_k^i(x)\sigma_k^j(x)$ and $\alpha^{ij}(x)=\sum_{k=1}^s \tau_k^i(x)\tau_k^j(x)$. It is a diffusion process generated by the differential operator

$$A=\frac{1}{2}\sum_{i,j=1}^{d} a^{ij}(x)D_iD_j+\sum_{i=1}^{d} b^i(x)D_i,$$

and by the Venttsel' boundary condition,

$$Lu(x)\equiv\frac{1}{2}\sum_{i,j=1}^{d-1} \alpha^{ij}(x)D_iD_ju(x)+\sum_{i=1}^{d-1} \beta^i(x)D_iu(x)$$
$$+D_du(x)-\rho(x)(Au)(x)=0 \text{ on } \partial D.$$

G. Stochastic Differential Equations on Manifolds

Let M be a connected σ-compact C^∞-manifold of dimension d, and let $W_M=C([0,\infty)\to M)$ be the space of all continuous paths in M. If M is not compact, let $\hat{M}=M\cup\{\Delta\}$ be the one-point compactification of M and $\hat{W}_M$ be the space of all continuous paths in $\hat{M}$ with Δ as a †trap. These path spaces are endowed with the σ-fields $\mathcal{B}(W_M)$ and $\mathcal{B}(\hat{W}_M)$, respectively, which are generated by Borel cylinder sets. By a continuous process on M we mean a $(W_M, \mathcal{B}(W_M))$-valued random variable, and by a continuous process on M admitting explosions we mean a $(\hat{W}_M, \mathcal{B}(\hat{W}_M))$-valued random variable. In this section the probability space is taken to be the r-dimensional Wiener space $(W_0^r, \mathcal{F}, P)$ with the increasing family $\{\mathcal{F}_t\}$, where $\mathcal{F}_t$ is generated by $\mathcal{B}_t(W_0^r)$ and P-null sets. Then $w=(w(t))$, $w\in W_0^r$, is an r-dimensional $\{\mathcal{F}_t\}$-Brownian motion.

Suppose that we are given a system of C^∞-vector fields $A_0, A_1, \ldots, A_r$ on M. We consider the following stochastic differential equation on M:

(11) $dX_t=A_k(X_t)\circ dw^k(t)+A_0(X_t)dt.$

(Here, the usual convention for the omission of the summation sign is used.) A precise meaning of equation (11) is as follows: We say that $X=(X_t)$ satisfies equation (11) if X is an $\{\mathcal{F}_t\}$-adapted continuous process on M admitting explosions such that, for any C^∞-function f on M with compact support (we set $f(\Delta)=0$), $f(X_t)$ is a continuous semimartingale satisfying

(12) $df(X_t)=(A_kf)(X_t)\circ dw^k(t)+(A_0f)(X_t)dt,$

where $\circ$ is Itô's circle operation defined in Section C. This is equivalent to saying that $X_t=(X_t^1, \ldots, X_t^d)$, in each local coordinate, is a d-dimensional semimartingale such that

(13) $$dX_t^i=\sigma_k^i(X_t)\circ dw^k(t)+b^i(X_t)dt$$
$$=\sigma_k^i(X_t)dw^k(t)$$
$$+\left[\frac{1}{2}\sum_{k=1}^{r} D_j\sigma_k^i\sigma_k^j+b^i\right](X_t)dt,$$

where $A_k(x)=\sigma_k^i(x)D_i$, $k=1, 2, \ldots, r$, and $A_0(x)=b^i(x)D_i$. By solving the equation in each local coordinate and then putting these solutions together, we can obtain for each $x\in M$ a unique solution X_t of (11) such that $X_0=x$. We can also embed the manifold M in a higher-dimensional Euclidean space and solve the stochastic differential equation there. We denote the solution by $X(t, x, w)$. The law P_x on $\hat{W}_M$ of $[t\to X(t, x, w)]$ defines a diffusion process on M which is generated by the differential operator $A=\frac{1}{2}\sum_{k=1}^r A_k^2+A_0$.

Next, if we consider the mapping $x\to X(t, x, w)$; then, except for w belonging to a set of P-measure 0, the following is valid: For all (t, w) such that $X(t, x_0, w)\in M$, the mapping $x\to X(t, x, w)$ is a diffeomorphism between a neighborhood of x_0 and a neighborhood of $X(t, x_0, w)$. This is based on the following fact for stochastic differential equations on $\mathbf{R}^d$. If in equation (3) the coefficients σ_k^i and b^i are C^∞-functions with bounded derivatives of all orders α, $|\alpha|\geqslant 1$, then, denoting by $X(t, x, w)$ the solution such that $X(0)=x$, we have that $x\to X(t, x, w)$ is, with probability 1, a diffeomorphism of $\mathbf{R}^d$ for all t [13].

Example 1: **Stochastic moving frame** [6, 15]. Let M be a Riemannian manifold of dimension d, $O(M)$ be the orthonormal frame bundle over M, and $L_1, L_2, \ldots, L_d$ be the basic vector fields on $O(M)$, that is,

$$(L_if)(x, \mathbf{e})=\lim_{t\to 0}\frac{1}{t}[f(x_t, \mathbf{e}_t)-f(x, \mathbf{e})],$$
$$i=1, \ldots d,$$

where $\mathbf{e}=(e_1, \ldots, e_d)$ is an orthonormal basis in $T_x(M)$, $x_t=\operatorname{Exp}(te_i)$, i.e., the geodesic such that $x_0=x$ and $\dot{x}=e_i$, and $\mathbf{e}_t$ is the parallel translate of $\mathbf{e}$ along x_t. Let b be a vector field on M and L_b be its horizontal lift on $O(M)$, i.e., L_b is a vector field on $O(M)$ determined by the following two properties: (i) L_b is horizontal and (ii) $d\pi(L_b)=b$, where $\pi: O(M)\rightarrow M$ is the projection. Consider the following stochastic differential equation on $O(M)$:

$$dr(t)=L_i(r(t))\circ dw^i(t)+L_b(r(t))\,dt.$$

Solutions determine a family of (local) diffeomorphisms $r\rightarrow r(t,r,w)=(X(t,r,w),\mathbf{e}(t,r,w))$ on $O(M)$. The law of $[t\rightarrow X(t,r,w)]$ depends only on $x=\pi(r)$, and it defines a diffusion process on M that is generated by the differential operator $\frac{1}{2}\Delta_M+b$ (Δ_M is the Laplace-Beltrami operator). Using this stochastic moving frame $r(t,r,w)$, we can realize a stochastic parallel translation of tensor fields along the paths of Brownian motion on M (a diffusion generated by $\frac{1}{2}\Delta_M$) that was first introduced by Itô [10], and by using it we can treat heat equations for tensor fields by means of a probabilistic method.

Example 2: **Brownian motion on Lie groups.** Let G be a Lie group. A stochastic process $\{g(t)\}$ on G is called a **right-invariant Brownian motion** if it satisfies the following conditions: (i) With probability 1, $g(0)=e$ (the identity), and $t\rightarrow g(t)$ is continuous; (ii) for every $t\geqslant s$, $g(t)g(s)^{-1}$ and $\sigma(g(u); u\leqslant s)$ are independent; and (iii) for every $t\geqslant s$, $g(t)g(s)^{-1}$ and $g(t-s)$ are equally distributed.

Let $A_0, A_1, \ldots, A_r$ be a system of right-invariant vector fields on G, and consider the stochastic differential equation

(14) $\quad dg_t=A_i(g_t)\circ dw^i(t)+A_0(g_t)\,dt.$

Then a solution of (14) with $g_0=e$ exists uniquely and globally; we denote this solution by $g^0(t,w)$. It is a right-invariant Brownian motion G, and conversely, every right-invariant Brownian motion can be obtained in this way. The system of diffeomorphisms $g\rightarrow g(t,g,w)$ defined by the solutions of (14) is given by $g\rightarrow g(t,g,w)=g^0(t,w)g$.

Generally, if M is a compact manifold, the system of diffeomorphisms $g_t: x\rightarrow X(t,x,w)$ defined by equation (11) can be considered as a right-invariant Brownian motion on the infinite-dimensional Lie group consisting of all diffeomorphisms of M [2].

References

[1] B. S. Tsirel'son (Cirel'son), An example of stochastic differential equation having no strong solution, Theory Prob. Appl., 20 (1975), 416–418.

[2] K. D. Elworthy, Stochastic dynamical systems and their flows, Stochastic Analysis, A. Friedman and M. Pinsky (eds.), Academic Press, 1978, 79–95.

[3] A. Friedman, Stochastic differential equations and applications I, II, Academic Press, 1975.

[4] I. I. Gikhman and A. V. Skorokhod, Stochastic differential equations, Springer, 1972.

[5] N. Ikeda: On the construction of two-dimensional diffusion processes satisfying Wentell's boundary conditions and its application to boundary value problems, Mem. Coll. Sci. Univ. Kyoto, (A, Math.) 33 (1961), 367–427.

[6] N. Ikeda and S. Watanabe, Stochastic differential equations and diffusion processes, Kodansha and North-Holland, 1981.

[7] K. Itô, Differential equations determining Markov processes (in Japanese), Zenkoku Shijô Sûgaku Danwakai, 1077 (1942), 1352–1400.

[8] K. Itô, On stochastic differential equations, Mem. Amer. Math. Soc., 4 (1951).

[9] K. Itô, Stochastic differentials, Appl. Math. and Optimization, 1 (1975), 347–381.

[10] K. Itô, The Brownian motion and tensor fields on Riemannian manifold, Proc. Intern. Congr. Math., Stockholm 1962, 536–539.

[11] K. Itô and S. Watanabe, Introduction to stochastic differential equations, Proc. Intern. Symp. SDE, Kyoto, 1976, K. Itô (ed.), Kinokuniya, 1978, i–xxx.

[12] T. Jeulin, Semi-martingales et grossissement d'une filtration, Lecture notes in math. 833, Springer, 1980.

[13] H. Kunita, On the decomposition of solutions of stochastic differential equations, Proc. LMS Symp. Stochastic Integrals, Durham, 1980.

[14] H. Kunita and S. Watanabe, On square integrable martingales, Nagoya Math. J., 30 (1967), 209–245.

[15] P. Malliavin, Géometrie différentielle stochastique, Les Presse de l'Université de Montréal, 1978.

[16] P. Malliavin, Stochastic calculus of variation and hypoelliptic operators, Proc. Intern. Symp. SDE, Kyoto, 1976, K. Itô (ed.), Kinokuniya, 1978, 195–263.

[17] P. Malliavin, C^k-hypoellipticity with degeneracy, Stochastic Analysis, A. Friedman and M. Pinsky (eds.), Academic Press, 1978, 199–214, 327–340.

[18] H. P. McKean, Stochastic integrals, Academic Press, 1969.

[19] P. A. Meyer, Un cours sur les intégrales stochastiques, Lecture notes in math. 511, Springer, 1976, 245–400.

[20] A. V. Skorokhod, Studies in the theory of random processes, Addison-Wesley, 1965.
[21] D. W. Stroock and S. R. S. Varadhan, Diffusion processes with continuous coefficients I, II, Comm. Pure Appl. Math., 22 (1969), 345–400, 479–530.
[22] D. W. Stroock and S. R. S. Varadhan, Multidimensional diffusion processes, Springer, 1979.
[23] S. Watanabe, On stochastic differential equations for multi-dimensional diffusion processes with boundary conditions, I, II, J. Math. Kyoto Univ., 11 (1971), 169–180, 545–551.
[24] A. D. Venttsel' (Wentzell), On boundary conditions for multidimensional diffusion processes, Theory Prob. Appl., 4 (1959), 164–177.
[25] T. Yamada and S. Watanabe, On the uniqueness of solutions of stochastic differential equations, J. Math. Kyoto Univ., 13 (1973), 497–512.

407 (XVII.4) Stochastic Processes

A. Definitions

The theory of stochastic processes was originally involved with forming mathematical models of phenomena whose development in time obeys probabilistic laws. Given a basic †probability space $(\Omega, \mathfrak{B}, P)$ and a set T of real numbers, a family $\{X_t\}_{t\in T}$ of real-valued †random variables defined on $(\Omega, \mathfrak{B}, P)$ is called a **stochastic process** (or simply **process**) over $(\Omega, \mathfrak{B}, P)$, where t is usually called the **time parameter** of the process. For each finite t-set $\{t_1, \ldots, t_n\}$, the †joint distribution of $(X_{t_1}, \ldots, X_{t_n})$ is called a **finite-dimensional distribution** of the process $\{X_t\}_{t\in T}$. Stochastic processes are classified into large groups such as †additive processes (or processes with independent increments), †Markov processes, †Markov chains, †diffusion processes, †Gaussian processes, †stationary processes, †martingales, and †branching processes, according to the properties of their finite-dimensional distributions. This classification is possible because of the following fact, a consequence of Kolmogorov's †extension theorem (→ 341 Probability Measures I): Given a system $\mathscr{P}$ of finite-dimensional distributions satisfying certain †consistency conditions, we can construct a suitable probability measure on the space $W = \mathbf{R}^T$ of real-valued functions on T so that the stochastic process $\{X_t\}_{t\in T}$, obtained by setting $X_t(w)$ = the value of $w \in W$ at t, has $\mathscr{P}$ as its system of finite-dimensional distributions. Now, consider two stochastic processes $\mathscr{X} = \{X_t\}_{t\in T}$ and $\mathscr{Y} = \{Y_t\}_{t\in T}$. $\mathscr{Y}$ is called a **modification** of $\mathscr{X}$ if they are defined over a common probability space $(\Omega, \mathfrak{B}, P)$ and $P(X_t = Y_t) = 1$ $(t \in T)$. Regardless of whether $\mathscr{X}$ and $\mathscr{Y}$ are defined over a common probability space or over different probability spaces, X and Y are said to be **equivalent** or each is said to be a **version** of the other if their finite-dimensional distributions are the same. According to Kolmogorov's extension theorem, every stochastic process has a version over the space $W = \mathbf{R}^T$.

The function $X(\omega)$ of t obtained by fixing ω in a stochastic process $\{X_t\}_{t\in T}$ is called the **sample function (sample process** or **path)** corresponding to ω. In applying various operations to stochastic processes and studying detailed properties of stochastic processes, such as continuity of sample functions, the notions of measurability and separability play important roles. We assume that T is an interval in the real line, and (if needed) that the probability measure P is †complete. Denote by $\mathfrak{F}$ the class of all †Borel subsets of T. A stochastic process $\{X_t\}_{t\in T}$ is said to be **measurable** if the function $X_t(\omega)$ of (t, ω) is $\mathfrak{F} \times \mathfrak{B}$-measurable. Continuity in probability defined in the next paragraph gives a sufficient condition for a stochastic process to have a measurable modification.
A stochastic process $\{X_t\}_{t\in T}$ is said to be **separable** if there exists a countable subset S of T such that

$$P\left\{\omega \,\middle|\, \liminf_{s\to t, s\in S} X_s(\omega) \leqslant X_t(\omega) \leqslant \limsup_{s\to t, s\in S} X_s(\omega) \quad \text{for any } t\in T\right\} = 1.$$

It was proved by J. L. Doob that every stochastic process has a separable modification [6].

Various types of continuity are considered for stochastic processes. $\{X_t\}_{t\in T}$ is said to be **continuous in probability** at $s \in T$ if $P(|X_t - X_s| > \varepsilon) \to 0$ $(t \to s, t \in T)$ for each $\varepsilon > 0$; it is said to be **continuous in the mean (of order** 1) at $s \in T$ if $E(|X_t \to X_s|) \to 0$ $(t \to s, t \in T)$. **Continuity in the mean of order** p (>1) is defined similarly. Continuity in the mean of any order implies continuity in probability. Suppose that $\{X_t\}_{t\in T}$ is separable. Then

$$D_s = \left\{\omega \,\middle|\, X_s(\omega) = \lim_{t\to s, t\in T} X_t(\omega)\right\}^c \quad \text{and} \quad \bigcup_{s\in T} D_s$$

are measurable events. If $P(D_s) > 0$, then $s \in T$ is called a **fixed point of discontinuity**. The condition $P(\bigcup_{s\in T} D_s) = 0$ means that almost all sample functions are continuous. Regularity properties of sample functions of processes, such as continuity or right continuity, have

been studied by many people. The following theorem is due to A. N. Kolmogorov: Let $T=[0,1]$. If

$$E(|X_t-X_s|^{\gamma})\leqslant c|t-s|^{1+\varepsilon}$$

for constants $\gamma>0$, $\varepsilon>0$, and $c>0$, then $\{X_t\}_{t\in T}$ has a modification $\{\tilde{X}_t\}_{t\in T}$ for which almost all sample functions are continuous, and

$$P\left[\lim_{h\downarrow 0} h^{-\delta}\sup_{\substack{|t-s|\leqslant h\\ t,s\in T}}|\tilde{X}_t-\tilde{X}_s|=0\right]=1$$

for any $\delta(0<\delta<\varepsilon/\gamma)$. Each of the following is a sufficient condition for $\mathscr{X}=\{X_t\}_{t\in T}$ to have a modification for which almost all sample paths are right continuous functions with left limits. (i) $\mathscr{X}$ is an additive process which is continuous in probability (P. Lévy [3,4], K. Itô [8]; → 5 Additive Processes B). (ii) $\mathscr{X}$ is a supermartingale that is continuous in probability (Doob [6]; → 262 Martingales C).

B. Increasing Families of σ-Algebras

In the investigation of stochastic processes (especially Markov processes, martingales, and stochastic differential equations), the notion of increasing families of σ-algebras often plays an important role. Let $(\Omega,\mathfrak{B},P)$ be a probability space, and let $T=[0,\infty)$. A family $\{\mathfrak{B}_t\}_{t\in T}$ of σ-subalgebras of $\mathfrak{B}$ is called an **increasing family of σ-algebras** on $(\Omega,\mathfrak{B},P)$ if $\mathfrak{B}_s\subset\mathfrak{B}_t$ for $s<t$. A process $\{X_t\}_{t\in T}$ is said to be **adapted** to $\{\mathfrak{B}_t\}$ if X_t is $\mathfrak{B}_t$-measurable for each $t\in T$. $\{X_t\}$ is said to be **progressively measurable** (or a **progressive process**) with respect to $\{\mathfrak{B}_t\}$ if for every $t\in T$ the mapping $(s,\omega)\mapsto X_s(\omega)$ of $[0,t]\times\Omega$ into $\mathbf{R}$ is measurable with respect to the σ-field $\mathfrak{B}([0,t])\times\mathfrak{B}_t$. A process $\{X_t\}$ with right continuous paths, adapted to $\{\mathfrak{B}_t\}$, is progressively measurable with respect to $\{\mathfrak{B}_t\}$. The same conclusion holds for a process with left continuous paths. A subset A of $[0,\infty)\times\Omega$ is said to be **progressive** if the indicator process $a_t(\omega)=\mathbf{1}_A(t,\omega)$ of A is a progressive process. A random time τ on Ω with values in $[0,\infty]$ is called a **stopping time** (or **Markov time**) if $\{\tau\leqslant t\}\in\mathfrak{B}_t$ for all $t\geqslant 0$. Constants ($\geqslant 0$) are stopping times. If σ and τ are stopping times, then $\min(\sigma,\tau)$ and $\max(\sigma,\tau)$ are also stopping times. The limit of an increasing sequence of stopping times is a stopping time, while the limit of a decreasing sequence of stopping times is a stopping time with respect to $\{\mathfrak{B}_{t+}\}$, where $\mathfrak{B}_{t+}=\bigcap_{s>t}\mathfrak{B}_s$. Let $\mathfrak{B}_\tau$ be the class of $A\in\mathfrak{B}$ such that $A\cap\{\tau\leqslant t\}\in\mathfrak{B}_t$ ($^\forall t\in T$); then it is a σ-algebra if τ is a stopping time. If $\{X_t\}$ is a progressive process and if τ is a stopping time, then $X_\tau\mathbf{1}_{\{\tau<\infty\}}$ is $\mathfrak{B}_\tau$-measurable. An increasing family $\{\mathfrak{B}_t\}$ of σ-algebras on $(\Omega,\mathfrak{B},P)$ is said to be complete if the probability space $(\Omega,\mathfrak{B},P)$ is complete and if all the P-negligible sets belong to $\mathfrak{B}_0$. Fron now on we assume that $\{\mathfrak{B}_t\}$ is complete and right continuous (i.e., $\mathfrak{B}_t=\mathfrak{B}_{t+}$ for all $t\geqslant 0$). Let B be a subset of $\mathbf{R}$ and $\{X_t\}$ be a process. We call $\tau_B=\inf\{t\geqslant 0\,|\,X_t(\omega)\in B\}$ a **hitting time** for B. Measurability of τ_B is not always guaranteed, that is, τ_B is not always a stopping time. G. A. Hunt showed that for a wide class of Markov processes hitting times for analytic sets are stopping times. This result is based on a theorem of G. Choquet on capacitability and was generalized by P. A. Meyer as follows: (i) For every progressively measurable process, hitting times for analytic sets are stopping times; and (ii) for every progressive set A, $D_A=\inf\{t\geqslant 0\,|\,(t,\omega)\in A\}$ is a stopping time. The following notions on measurability are also important. The **predictable σ-algebra** on $[0,\infty)\times\Omega$, denoted by $\mathscr{P}$, is defined to be the least σ-algebra on $[0,\infty)\times\Omega$ with respect to which every process $X_t(\omega)$ that is adapted to $\{\mathfrak{B}_t\}$ and has left continuous paths is measurable in (t,ω). The **well-measurable** or **optional σ-algebra** on $[0,\infty)\times\Omega$, denoted by $\mathscr{O}$, is defined to be the least σ-algebra on $[0,\infty)\times\Omega$ with respect to which every process $X_t(\omega)$ that is adapted to $\{\mathfrak{B}_t\}$ and has right continuous paths with left limits is measurable in (t,ω). A process $\{X_t\}$ defined on Ω is said to be **predictable** (resp. **well-measurable** or **optional**) if the function $(t,\omega)\mapsto X_t(\omega)$ on $[0,\infty)\times\Omega$ is measurable with respect to the predictable σ-algebra $\mathscr{P}$ (resp. the optional σ-algebra $\mathscr{O}$). For further information regarding the notions given in this section → [10].

Up to this point it was assumed that the space in which a process $\{X_t\}_{t\in T}$ takes values, namely, the **state space** of $\{X_t\}_{t\in T}$, is a set of real numbers; but in general, topological spaces or merely measurable spaces can be taken for the state spaces of stochastic processes. The general definitions and results already given can be extended to stochastic processes whose state spaces are †locally compact Hausdorff spaces satisfying the second †countability axiom.

Moreover, the time parameter set T of a process $\{X_t\}_{t\in T}$ need not be a set of real numbers. For example, P. Lévy [12] and H. P. McKean [13] investigated stochastic processes with several-dimensional time; such processes are sometimes called **random fields**. The case in which T is $\mathscr{D}$, $\mathscr{S}$, or in general a space of functions (which is nuclear) has also been investigated (→ Section C). A probabilistic formulation of equilibrium states given by R. L. Dobrushin [14] initiated recent probabilistic study of statistical mechanics. For further information concerning processes with general

time parameter spaces → 136 Ergodic Theory, 176 Gaussian Processes, 340 Probabilistic Methods in Statistical Mechanics.

C. Random Distributions and Generalized Stochastic Processes

The investigation of random distributions was initiated by I. M. Gel'fand [15] and Itô [17]. Denote by $\mathscr{D}$ the space of functions of t $(-\infty < t < \infty)$ of class C^∞ with compact †support, and by $\mathscr{D}'$ the space of †distributions. At function $X(\varphi, \omega)$ of $\omega \in \Omega$ and $\varphi \in \mathscr{D}$ is called a **random distribution** (or **generalized stochastic process**) if $X(\varphi, \omega)$ is a distribution as a function of φ for almost all ω and is measurable as a function of ω for each fixed φ. Denote by $\mathfrak{B}(\mathscr{D}')$ the smallest σ-algebra containing sets of the form $\{y \in \mathscr{D}' \mid y(\varphi) \in E\}$ $(\varphi \in \mathscr{D}$, E is a Borel set). A random distribution is nothing but a $\mathscr{D}'$-valued random variable. For a random distribution $X(\varphi, \omega)$, a probability measure Φ_X on $\mathfrak{B}(\mathscr{D}')$ is induced by

$$\Phi_X(B) = P\{\omega \mid X(\cdot, \omega) \in B\}, \qquad B \in \mathfrak{B}(\mathscr{D}').$$

The functional

$$c(\varphi) = E(e^{iX(\varphi, \omega)}) = \int e^{iy(\varphi)} \Phi_X(dy), \qquad \varphi \in \mathscr{D},$$

is called the **characteristic functional** of $X(\varphi, \omega)$ or Φ_X. The functional $c(\varphi)$ is continuous positive definite, and $c(0) = 1$. Conversely, given a functional $c(\varphi)$ with these properties, a theorem of R. A. Minlos (→ 341 Probability Measures J) states that there exists a unique probability measure Φ on $\mathfrak{B}(\mathscr{D}')$ whose characteristic functional equals $c(\varphi)$. In other words, a random distribution with the characteristic functional $c(\varphi)$ can be constructed over $(\mathscr{D}', \mathfrak{B}(\mathscr{D}'), \Phi)$.

Typical classes of random distributions that have been investigated so far are stationary random distributions and random distributions with independent values at every point. (For stationary ones → 395 Stationary Processes.) A random distribution $X(\varphi, \omega)$ is called a **random distribution with independent values at every point** if $\varphi_1(t)\varphi_2(t) \equiv 0$ implies the independence of $X(\varphi_1, \omega)$ and $X(\varphi_2, \omega)$, that is, $c(\varphi_1 + \varphi_2) = c(\varphi_1)c(\varphi_2)$. A sufficient condition for the functional of the form

$$c(\varphi) = \exp\left(\int_{-\infty}^{\infty} f(\varphi(t), \varphi'(t), \ldots, \varphi^{(k)}(t))\, dt\right)$$

(f is continuous and $f(0, \ldots, 0) = 0$) to be the characteristic functional of a stationary random distribution with independent values at every point is that the function $\exp(sf(x_0, x_1, \ldots, x_k))$ of $(x_0, x_1, \ldots, x_k) \in \mathbf{R}^k$ be positive definite for each $s > 0$ [16]. Under this condition, a concrete representation of f is known [16]. When $k = 0$, a necessary and sufficient condition for $c(\varphi)$ to be the characteristic functional of a stationary random distribution with independent values at every point is that $\exp f(x)$ be the characteristic function of an †infinitely divisible distribution. One such random distribution is the so-called **Guassian white noise**, namely, the distribution derivative of Brownian motion whose characteristic functional is

$$\exp\left(-\frac{1}{2}\int |\varphi(t)|^2\, dt\right).$$

A family $\{X_\varphi\}_{\varphi \in \mathscr{D}}$ of real-valued random variables indexed by $\mathscr{D}$ is called a **random distribution in the wide sense** if X_φ is linear in φ, namely, $X_{a\varphi + b\psi} = aX_\varphi + bX_\psi$ with probability 1 for fixed $\varphi, \psi \in \mathscr{D}$, and real constants a, b, and if $X_\varphi \mapsto 0$ in probability whenever $\varphi \to 0$ in the topology of $\mathscr{D}$. (For a typical class of random distributions in the wide sense → 395 Stationary Processes C.) A random distribution in the wide sense has a modification that is a random distribution.

In the definition of random distributions (in the wide sense) one can replace the space $\mathscr{D}$ by the space $\mathscr{S}$ of rapidly decreasing C^∞-functions or in general by some space Φ of functions that is nuclear. For example, one can define random distributions as $\mathscr{S}'$-valued random variables.

Up to this point random distributions of one variable have been considered. Random distributions of several variables are called generalized random fields and have been investigated by Itô [18], A. M. Yaglom [19], Gel'fand and N. Ya. Vilenkin [16], and others. Moreover, K. Urbanik [20, 21] developed a theory of generalized stochastic processes based on G. Mikusiński's theory instead of L. Schwartz's theory of distributions.

D. Random Measure

Let $(S, \mathfrak{F}, m)$ be any σ-finite measure space, and put $\mathfrak{F}_0 = \{A \in \mathfrak{F} \mid m(A) < \infty\}$. By virtue of †Kolmogorov's extension theorem, there exists a family $\{W(A)\}_{A \in \mathfrak{F}_0}$ of real random variables indexed by $\mathfrak{F}_0$ such that (i) for any mutually disjoint $A_1, \ldots, A_n \in \mathfrak{F}_0$, $\{W(A_1), \ldots, W(A_n)\}$ is independent; (ii) for any $A \in \mathfrak{F}_0$, $W(A)$ is Gaussian distributed with mean 0 and variance $m(A)$; and (iii) for any A, $B \in \mathfrak{F}_0$, $E(W(A)W(B)) = m(A \cap B)$. Similarly, there exists a family $\{N(A)\}_{A \in \mathfrak{F}_0}$ of real-valued random variables indexed by $\mathfrak{F}_0$ such that (i) for any mutually disjoint $A_1, \ldots, A_n \in \mathfrak{F}_0$, $\{N(A_1), \ldots, N(A_n)\}$ is independent; (ii) for any $A \in \mathfrak{F}_0$, $N(A)$ is †Poisson distributed with mean $m(A)$; and (iii) for

any $A, B \in \mathfrak{F}_0$, $E(N(A)N(B)) = m(A \cap B)$. $\{W(A)\}_{A \in \mathfrak{F}_0}$ and $\{N(A)\}_{A \in \mathfrak{F}_0}$ are called a **Gaussian random measure** and a **Poisson random measure** associated with the measure space $(S, \mathfrak{F}, m)$, respectively. By using these random measures, the theory of multiple integrals can be developed.

By a **point function** p on S we mean a mapping $p: D_p \to S$, where the domain D_p is a countable subset of $(0, \infty)$. p defines a counting measure N_p on $(0, \infty) \times S$ such that $N_p((0, t] \times U) = \#\{s \in D_p; s \leqslant t, p(s) \in U\}$ $(t > 0, U \in \mathfrak{F}_0)$. For a point function p and $t > 0$, the shift point function $\Theta_t p$ is defined by $D_{\Theta_t p} = \{s \in (0, \infty); s + t \in D_p\}$ and $(\Theta_t p)(s) = p(s+t)$. Let Π_S be the totality of point functions on S and $\mathscr{B}(\Pi_S)$ be the smallest σ-field on Π_S with respect to which all $p \to N_p((o, t] \times U, t > 0, U \in \mathfrak{F}_0)$ are measurable. A **point process** on S is a $(\Pi_S, \mathscr{B}(\Pi_S))$-valued random variable. Then there exists a point process p on S such that (i) for any $t > 0$, p and $\Theta_t p$ have the same probability law, and (ii) N_p is a Poisson random measure associated with $((0, \infty) \times S, \mathscr{B}(0, \infty) \times \mathscr{B}(\Pi_S), dt \times m(ds))$. The point process p is called the **stationary Poisson point process** with the **characteristic measure** m.

References

[1] A. N. Kolmogorov, Grundbegriffe der Wahrscheinlichkeitsrechnung, Erg. Math., 1933; English translation, Foundations of the theory of probability, Chelsea, 1950.
[2] N. Wiener, Differential-space, J. Math. Phys., 2 (1923), 131–174.
[3] P. Lévy, Processus stochastiques et mouvement brownien, Gauthier-Villars, 1948.
[4] P. Lévy, Théorie de l'addition des variables aléatoires, Gauthier-Villars, 1937.
[5] M. S. Bartlet, An introduction to stochastic processes, Cambridge Univ. Press, 1955.
[6] J. L. Doob, Stochastic processes, Wiley, 1953.
[7] J. L. Doob, Stochastic processes depending on a continuous parameter, Trans. Amer. Math. Soc., 42 (1937), 107–140.
[8] K. Itô, Stochastic processes, Aarhus Univ. lecture notes 16, 1969.
[9] P. A. Meyer, Une présentation de la théorie des ensembles sousliniens; application aux processus stochastiques, Sém. Théorie du Potentiel, 1962–1963, no. 2, Inst. H. Poincaré, Univ. Paris, 1964.
[10] C. Dellacherie and P. A. Meyer, Probabilités et potentiel, Hermann, 1975, chs. I–IV.
[11] C. Dellacherie, Capacités et processus stochastiques, Springer, 1972.
[12] P. Lévy, Le mouvement brownien fonction d'un ou de plusieurs paramètres, Rend. Sem. Math. Univ. Padova, (5) 22 (1963), 24–101.
[13] H. P. McKean, Jr., Brownian motion with a several-dimensional time, Theory Prob. Appl., 8 (1963), 335–354.
[14] R. L. Dobrushin, The description of the random field by its conditional distributions and its regularity conditions, Theory Prob. Appl., 13–14 (1968), 197–224. (Original in Russian, 1968).
[15] I. M. Gel'fand, Generalized random processes (in Russian), Dokl. Akad. Nauk SSSR (N.S.), 100 (1955), 853–856.
[16] I. M. Gel'fand and I. Ya. Vilenkin, Generalized functions IV, Academic Press, 1964. (Original in Russian, 1961.)
[17] K. Itô, Stationary random distributions, Mem. Coll. Sci. Univ. Kyoto, (A) 28 (1954), 209–223.
[18] K. Itô, Isotropic random current, Proc. 3rd Berkeley Symp. Math. Stat. Prob. II, Univ. of California Press, 1956, 125–132.
[19] A. M. Yaglom (Jaglom), Some classes of random fields in n-dimensional space related to stationary random processes, Theory Prob. Appl., 2 (1957), 273–320. (Original in Russian, 1957.)
[20] K. Urbanik, Stochastic processes whose sample functions are distributions, Theory Prob. Appl., 1 (1956), 132–134. (Original in Russian, 1956.)
[21] K. Urbanik, Generalized stochastic processes, Studia Math., 16 (1958), 268–334.

408 (XIX.7) Stochastic Programming

A. General Remarks

Stochastic programming is a method of finding optimal solutions in mathematical programming in its narrow sense (→ 264 Mathematical Programming), when some or all coefficients are stochastic variables with known probability distributions. There are essentially two different types of models in stochastic programming situations: One is **chance-constrained programming** (CCP), and the other is a **two-stage stochastic programming** (TSSP). The difference between them depends mainly on the informational structure of the sequence of observations and decisions. For simplicity, let us here consider stochastic linear programming, which is the best-known model at present.

Let A_0, A be $m \times n$-dimensional matrices and $x, c \in \mathbf{R}^n$ and $b, b_0 \in \mathbf{R}^m$. Suppose further that components of A, b, c are random variables, while those of A_0, b_0 are constants. Consider

the following formally defined linear programming problem: $\min_x\{c'x \mid Ax \leqslant b, x \in X_0\}$, $X_0 = \{x \mid A_0 x \leqslant b_0, x \geqslant 0\}$. Let $(\Omega, \mathfrak{B}, P)$ be a probability space (→ 342 Probability Theory) such that $\{A(\omega), b(\omega), c(\omega)\}$ is a measurable transformation on Ω into $\mathbf{R}^{m\times n+m+n}$.

B. Chance-Constrained Programming (CCP)

This method is based on the assumption that a decision x has to be made in advance of the realization of the random variables. Suppose that $A_i(\omega)$ is the ith row of $A(\omega)$, and $b_i(\omega)$ is the ith component of $b(\omega)$. We call $P(\{\omega \mid A_i(\omega)x \leqslant b_i(\omega)\}) \geqslant \alpha_i$ a **chance constraint**, where α_i is a prescribed fractional value determined by the decision maker according to his attitude toward the constraint $A_i(\omega)x \leqslant b_i(\omega)$: if he attaches importance to it, he will take α_i as great as possible. Defining feasible sets $X_i(\alpha_i)$ and X by $X_i(\alpha_i) = \{x \mid P(\{\omega \mid A_i(\omega)x \leqslant b_i(\omega)\}) \geqslant \alpha_i\}$, $X = X_0 \cap \{\bigcap_{i=1}^m X_i(\alpha_i)\}$, we can formulate CCP as follows: $\min_x\{F(x) \mid x \in X\}$, where $F: X \to R$, is the **certainty equivalent** of the stochastic objective function $c'x$. We have four models of CCP corresponding to different types of $F(x)$: (i) E-model: $F(x) = \bar{c}'x$, $\bar{c} = E_\omega c(\omega)'x$. (ii) V-model: $F(x) = \mathrm{Var}(c(\omega)'x) = x'V_c x$, where V_c is a variance-covariance matrix of $c(\omega)$. (iii) P_1-model: $F(x) = f$, $P(\{\omega \mid c(\omega)'x \leqslant f\}) \geqslant \alpha_0$, $1/2 \leqslant \alpha_0 \leqslant 1$. (iv) P_2-model: $F(x) = P(\{\omega \mid c(\omega)'x \geqslant \gamma\})$ for a given constant γ. In particular, if the components of $A(\omega)$, $b(\omega)$, $c(\omega)$ have a multidimensional normal distribution, the certainty equivalent of the ith chance constraint is derived in the following form: $\bar{A}_i x + \Phi^{-1}(\alpha_i)(x'V_i x + 2w_i'x + v_i^2)^{1/2} \leqslant \bar{b}_i$, where $\bar{A}_i$, V_i, w_i, v_i, $\bar{b}_i$ are expectation vectors or a variance-covariance matrix of $A_i(\omega)$ and $b_i(\omega)$ and $\Phi(t) = \int_{-\infty}^t \exp(-z^2/2)\,dz/\sqrt{2\pi}$. The set $X_i(\alpha_i)$ for this constraint can be shown to be convex for $1/2 \leqslant \alpha_i \leqslant 1$, by using the convexity of the function $\sqrt{x'Vx}$ for a positive semidefinite matrix V. Under the same assumption we can obtain the objective functions $F(x) = \bar{c}'x + \Phi^{-1}(\alpha_0)\sqrt{x'V_c x}$ for the P_1-model and $F(x) = (\bar{c}'x - \gamma)/\sqrt{x'V_c x}$ for the P_2-model. These four models have been shown to be computable by applying convex programming techniques, including the conjugate gradient method. Further studies on the convexity of a more general chance constraint $P(\{\omega \mid A(\omega)x \leqslant b(\omega)\}) \geqslant \alpha$, $0 \leqslant \alpha \leqslant 1$, appear in several articles.

C. Two-Stage Stochastic Programming (TSSP)

This method divides the decision process into two stages. First stage: Before the realization of random variables, one makes a decision x, being allowed to compensate for it after the specification of those values. Second stage: One obtains an optimal compensation $y \in \mathbf{R}^r$ for the given x and the realized values of the random variables. Assuming that $q \in \mathbf{R}^r$ is a random vector in addition to A, b, c, we can formulate TSSP as follows. First stage: $\min_x E_\omega\{(c(\omega)'x + Q(x,\omega)) \mid x \in X)\}$; second stage: $Q(x,\omega) = \min_y\{q(\omega)'y \mid Wy = A(\omega)x - b(\omega), y \geqslant 0\}$, where $X = X_0 \cap K$, $K = \{x \mid Q(x,\omega) < +\infty$ with probability $1\}$ and $q(\omega)'y$ is a loss function for the deviation $A(\omega)x - b(\omega)$. The $m \times n$ matrix W is called a compensation matrix. Several theorems have been proved: (i) K is a closed convex set; (ii) $Q(x) = E_\omega Q(x,\omega)$ is a convex function on K if the random variables in $A(\omega)$, $b(\omega)$, $q(\omega)$ are square integrable; (iii) if P has a density function, then $Q(x)$ has a continuous gradient on K; (iv) when P has a finite discrete probability distribution, a TSSP problem is reduced to a linear programming problem having a dual decomposition structure.

References

[1] A. Charnes and W. W. Cooper, Chance-constrained programming, Management Sci., 6 (1959), 73–79.
[2] G. B. Dantzig and A. Madansky, On the solution of two-stage linear programs under uncertainty, Proc. 4th Berkeley Symp. Math. Statist. Probab. I, Univ. of California Press, 1961.
[3] S. Kataoka, A stochastic programming model, Econometrica, 31 (1963), 181–196.
[4] I. M. Stancu-Minasian and M. J. Wets, A research bibliography in stochastic programming 1955–1975, Operations Res., 24 (1976), 1078–1119.
[5] S. Vajda, Probabilistic programming, Academic Press, 1972.
[6] P. Kall, Stochastic linear programming, Springer, 1976.
[7] V. V. Kolbin, Stochastic programming, Reidel 1977. (Original in Russian, 1977.)

409 (II.7) Structures

A. Examples of Structure

Structure is a unified description of mathematical objects such as †ordered sets, †rings, †linear spaces, †topological spaces, †probability spaces, †manifolds, etc., using only the concepts †set and †relation. The following are examples.

(1) Order. An †ordering on a set A is a binary relation in A (→ 358 Relations) with a †graph α such that (i) if $a \in A$, then $(a, a) \in \alpha$; (ii) if $(a, b) \in \alpha$ and $(b, a) \in \alpha$, then $a = b$; and (iii) if $(a, b) \in \alpha$ and $(b, c) \in \alpha$, then $(a, c) \in \alpha$, where α is an element of the †power set $\mathfrak{P}(A \times A)$. We say that α determines a structure of ordering on the set A.

(2) Law of composition. A **law of composition** on a set A is a mapping from $A \times A$ (or a subset) to A. This mapping is considered as a ternary relation with a graph $\alpha \in \mathfrak{P}(A \times A \times A)$, satisfying the following two conditions (or possibly only (ii)): (i) if $(a, b) \in A \times A$, then $(a, b, c) \in \alpha$ for some $c \in A$; (ii) if $(a, b, c) \in \alpha$ and $(a, b, c') \in \alpha$, then $c = c'$. We say that α determines the structure of a law of composition in A. The †associative law in the law of composition determined by α is given by: (iii) if $(a, b, x) \in \alpha$, $(b, c, y) \in \alpha$, $(x, c, z) \in \alpha$, and $(a, y, z') \in \alpha$, then $z = z'$. A set with conditions (i), (ii), and (iii) is called a **semigroup**.

(3) Operation. An **operation** of a set A on a set B is a mapping from $A \times B$ (or a subset) to B. It is considered as a ternary relation with the graph $\gamma \in \mathfrak{P}(A \times B \times B)$, satisfying the following two conditions (or possibly only (ii)): (i) if $(a, b) \in A \times B$, then $(a, b, c) \in \gamma$ for some $c \in B$; (ii) if $(a, b, c) \in \gamma$ and $(a, b, c') \in \gamma$, then $c = c'$. We say that γ determines the structure of an operation of A on B. Each element α of A is called an **operator** on B; A is called a **domain of operators** on B, and B is called an **A-set**. When B is the main object of consideration, an operation of A on B is sometimes called an **external law of composition** of A on B. The law of composition of A as described in (2) is then called an **internal law of composition** of A. When a domain of operators A on B determined by $\gamma \in \mathfrak{P}(A \times B \times B)$ has an internal law of composition determined by $\alpha \in \mathfrak{P}(A \times A \times A)$, it is usually assumed that the following conditions on α, γ are satisfied: (iii) if $(a, b, x) \in \alpha$, $(b, c, y) \in \gamma$, $(x, c, z) \in \gamma$, and $(a, y, z') \in \gamma$, then $z = z'$. If we denote the law of composition by $(a, b) \rightarrow ab$ and the operation by $(a, b) \rightarrow a \cdot b$, then condition (iii) may be written: (iii') $(ab) \cdot c = a \cdot (b \cdot c)$ for $a, b \in A$ and $c \in B$. When an A-set B with an external law of composition determined by $\gamma \in \mathfrak{P}(A \times B \times B)$ has an internal law of composition determined by $\beta \in \mathfrak{P}(B \times B \times B)$, it is usually assumed that the following condition on β, γ is satisfied: (iv) if $(a, b, x) \in \gamma$, $(a, c, y) \in \gamma$, $(b, c, z) \in \beta$, $(x, y, w) \in \beta$, and $(a, z, w') \in \gamma$, then $w = w'$. According to the notation ab and $a \cdot b$, it is described as: (iv') $(a \cdot b)(a \cdot c) = a \cdot (bc)$ for $a \in A$ and $b, c \in B$.

The mapping $A \times B \rightarrow B$ $(B \times A \rightarrow B)$ is called a **left (right) operation** of A on B. To emphasize leftness or rightness, "left-" or "right-" is attached to corresponding concepts.

(4) Topology. A †topology on a set A is determined by a set $\alpha \in \mathfrak{P}\mathfrak{P}(A)$, called the †system of open sets, satisfying the following conditions: (i) $\varnothing \in \alpha$ and $A \in \alpha$; (ii) if $\beta \subset \alpha$, then $\bigcup \beta \in \alpha$; and (iii) if $\beta \subset \alpha$ is finite, then $\bigcap \beta \in \alpha$. We say that α defines the structure of a topology on the set A (→ 425 Topological Spaces).

B. Mathematical Structures

We now explain the concept of mathematical structure for the case of a †linear space (→ 256 Linear Spaces). A linear space has two basic sets, one of which is a set K of elements called scalars and the other, a set V of elements called †vectors, two laws of compositions in K called addition and multiplication, a law of composition in V called addition, and an operation of K on V called scalar multiplication. The laws of composition and the operation are given by elements of power sets: $\alpha_1, \alpha_2 \in \mathfrak{P}(K \times K \times K)$, $\alpha_3 \in \mathfrak{P}(V \times V \times V)$, and $\alpha_4 \in \mathfrak{P}(K \times V \times V)$; and the basic properties of the linear space, such as $\lambda(a+b) = \lambda a + \lambda b$ $(\lambda \in K, a, b \in V)$, are described as propositions on $K, V, \alpha_1, \ldots, \alpha_4$ and denoted by $P(K, V, \alpha_1, \ldots, \alpha_4), \ldots$.

Up to now, we have been considering a given linear space. To give a description of a linear space in general, we use the symbols $X_1, X_2, \xi_1, \ldots, \xi_4$ instead of the symbols K, $V, \alpha_1, \ldots, \alpha_4$, replace conditions such as $\alpha_1 \in \mathfrak{P}(K \times K \times K), \ldots$ by $\xi_1 \in \mathfrak{P}(X_1 \times X_1 \times X_1), \ldots$, and consider the set Σ of these symbols and formulas:

$$\Sigma: X_1, X_2, \xi_1, \ldots, \xi_4;$$

$$\xi_1 \in \mathfrak{P}(X_1 \times X_1 \times X_1), \ldots,$$

$$\xi_4 \in \mathfrak{P}(X_1 \times X_2 \times X_2).$$

The set Σ is called the **type** of linear space. Similarly, we consider the set Γ of all $P(X_1, X_2, \xi_1, \ldots, \xi_4), \ldots$ corresponding to the basic properties $P(K, V, \alpha_1, \ldots, \alpha_4), \ldots$ of the linear space. The set Γ is called the **axiom system** of the linear space.

In general, let $A_1, \ldots, A_m$ be the **basic sets** (K and V in the preceding example). The **basic concepts** $\alpha_1, \ldots, \alpha_n$ ($\alpha_1, \ldots, \alpha_4$ in the preceding example) are given as elements of finitely generated sets from $A_1, \ldots, A_m$, i.e., elements of sets obtained by a finite number of applications of the operations of forming the †Cartesian product and the †power set from $A_1, \ldots, A_m$. **Basic properties** are given as propositions on $A_1, \ldots, A_m, \alpha_1, \ldots, \alpha_n$. These basic properties and $A_1, \ldots, A_m, \alpha_1, \ldots, \alpha_n$ determine a **mathematical system**. We consider also a type Σ of symbols $X_1, \ldots, X_m$ of basic sets and symbols $\xi_1, \ldots, \xi_n$ of basic concepts, and an

axiom system Γ of basic properties. The pair (Σ, Γ) determines a **mathematical structure**. When we substitute sets $A_1, \ldots, A_m$ for $X_1, \ldots, X_m$ and $\alpha_1, \ldots, \alpha_n$ for $\xi_1, \ldots, \xi_n$, where $X_1, \ldots, X_m$ and $\xi_1, \ldots, \xi_n$ satisfy the axiom system Γ, then $(A_1, \ldots, A_m, \alpha_1, \ldots, \alpha_n)$ is called a mathematical system with the mathematical structure (Σ, Γ), or a **model** of the structure (Σ, Γ). Two mathematical systems are called **similar** or **of the same kind** if they have the same mathematical structure. Groups, rings, topological spaces, etc., are mathematical structures. Mathematical systems are sometimes called **algebraic systems in the wider sense**, and when we consider mainly their laws of composition (and operations), we call them **algebraic systems**. We explain this in further detail in Section C.

C. Algebraic Systems

Algebraic systems are sets with laws of composition and operations satisfying certain axiom systems; the laws of composition and operations and the axiom systems they satisfy determine their type (→ 2 Abelian Groups, 29 Associative Algebras, 42 Boolean Algebra, 67 Commutative Rings, 149 Fields, 151 Finite Groups, 190 Groups, 231 Jorden Algebras, 248 Lie Algebras, 368 Rings). Each algebraic system has its own theory, but general properties and related concepts are dealt with from a common standpoint. From this common ground we often get an insight into concepts from which arose a general theory of mathematical systems. We describe here only algebraic systems, but it is possible to describe similar concepts for mathematical systems. The following is a description based mainly on one law of composition $(a, b) \rightarrow ab$; a similar description is possible for the case of two or more laws of composition.

The law of composition ab is sometimes written $a+b$, $a \cdot b$, $[a, b]$, etc. A mapping $f: A \rightarrow A'$ of similar algebraic systems A and A' is called a **homomorphism** provided that $f(ab) = f(a)f(b)$ $(a, b \in A)$. A' is said to be **homomorphic** to A if there is homomorphism from A onto A'. If f is one-to-one, onto, and its inverse mapping $f^{-1}: A' \rightarrow A$ is also a homomorphism, then f is called an **isomorphism**, A' is said to be **isomorphic** to A, and the relation is written $A \cong A'$. The composition of homomorphisms is a homomorphism, and the identity mapping is an isomorphism. A homomorphism of A to A itself is called an **endomorphism**, and if it is also an isomorphism, then it is an **automorphism**. The set of all endomorphisms of A forms a †semigroup under composition, and the set of automorphisms forms a group under composition. The concept of homomorphism is a fundamental concept appearing in all algebraic systems. A homomorphism is sometimes called a **representation**.

An element e of an algebraic system A with a law of composition ab is called an **identity element** if $ae = ea = a$ (for all $a \in A$). If such an element exists, then it is unique. In the case of a ring, two laws of composition, addition and multiplication, are given. In this case the identity element for multiplication (if it exists) is called the identity element (or †unity element) of the ring. In the case of homomorphism between groups, the identity element is mapped to the identity element, but this does not always hold in general algebraic systems. Since the identity element plays an important role, it is frequently added to the basic concepts. Homomorphism between mathematical systems is generally defined to induce a mapping between basic concepts. A semigroup and a ring with a unity element are called a **unitary semigroup (monoid)** and a **unitary ring**, respectively, and homomorphisms between these systems are restricted to mappings that map the identity element to the identity element.

Let A and A' be similar algebraic systems, and let A be a subset of A'. A is called a **subsystem** of A' if the mapping $f: A \rightarrow A'$ defined by $f(a) = a (a \in A)$ is a homomorphism. A subsystem of a group (ring) is called a †subgroup (†subring), and similarly for other algebraic systems.

An †equivalence relation R in an algebraic system A is called **compatible** with the law of composition if $R(a, a')$ and $R(b, b')$ imply $R(ab, a'b')$. Consider the †quotient set A/R. Then the law of composition in A/R is uniquely determined so that the mapping $f: A \rightarrow A/R$ defined by $a \in f(a)$ is a homomorphism. The algebraic system thus obtained is called a **quotient system**. A quotient system of a group (ring) is a group (ring), called a †quotient group (†quotient ring). Other cases, including those where operations are given, are treated similarly (→ 52 Categories and Functors).

References

[1] N. Bourbaki, Eléments de mathématique, I. Théorie des ensembles, ch. 4, Structures, Actualités Sci. Ind., 1258a, Hermann, second edition, 1966; English translation, Theory of sets, Addison-Wesley, 1968.
[2] C. Chevalley, Fundamental concepts of algebra, Academic Press, 1956.

410 (VI.21) Surfaces

A. The Notion of a Surface

The notion of a surface may be roughly expressed by saying that by moving a curve we get a surface or that the boundary of a solid body is a surface. But these propositions cannot be considered mathematical definitions of a surface. We also make a distinction between surfaces and planes in ordinary language, where we mean by surfaces only those that are not planes. In mathematical language, however, planes are usually included among the surfaces.

A surface can be defined as a 2-dimensional †continuum, in accordance with the definition of a curve as a 1-dimensional continuum. However, while we have a theory of curves based on this definition, we do not have a similar theory of surfaces thus defined (→ 93 Curves).

What is called a surface or a curved surface is usually a 2-dimensional †topological manifold, that is, a topological space that satisfies the †second countability axiom and of which every point has a neighborhood †homeomorphic to the interior of a circular disk in a 2-dimensional Euclidean space. In the following sections, we mean by a surface such a 2-dimensional topological manifold.

B. Examples and Classification

The simplest examples of surfaces are the 2-dimensional †simplex and the 2-dimensional †sphere. Surfaces are generally †simplicially decomposable (or triangulable) and hence homeomorphic to 2-dimensional polyhedra (T. Radó, *Acta Sci. Math. Szeged.* (1925)). A †compact surface is called a **closed surface**, and a noncompact surface is called an **open surface**. A closed surface is decomposable into a finite number of 2-simplexes and so can be interpreted as a †combinatorial manifold. A 2-dimensional topological manifold having a boundary is called a **surface with boundary**. A 2-simplex is an example of a surface with boundary, and a sphere is an example of a closed surface without boundary.

Surfaces are classified as †orientable and †nonorientable. In the special case when a surface is †embedded in a 3-dimensional Euclidean space E^3, whether the surface is orientable or not depends on its having two sides (the "surface" and "back") or only one side. Therefore, in this special case, an orientable surface is called **two-sided**, and a nonorientable surface, **one-sided**. A nonorientable closed surface without boundary cannot be embedded in the Euclidean space E^3 (→ 56 Characteristic Classes, 114 Differential Topology).

The first example of a nonorientable surface (with boundary) is the so-called **Möbius strip** or **Möbius band**, constructed as an †identification space from a rectangle by twisting through 180° and identifying the opposite edges with one another (Fig. 1).

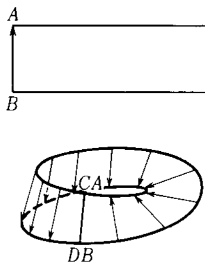


Fig. 1

As illustrated in Fig. 2, from a rectangle $ABCD$ we can obtain a closed surface homeomorphic to the product space $S^1 \times S^1$ by identifying the opposite edges AB with DC and BC with AD. This surface is the so-called 2-dimensional **torus** (or **anchor ring**). In this case, the four vertices A, B, C, D of the rectangle correspond to one point p on the surface, and the pairs of edges AB, DC and BC, AD correspond to closed curves a' and b' on the surface. We use the notation $aba^{-1}b^{-1}$ to represent a torus. This refers to the fact that the torus is obtained from an oriented four-sided polygon by identifying the first side and the third (with reversed orientation), the second side and the fourth (with reversed orientation). Similarly, aa^{-1} represents a sphere (Fig. 3), and $a_1b_1a_1^{-1}b_1^{-1}a_2b_2a_2^{-1}b_2^{-1}$ represents the closed surface shown in Fig. 4.

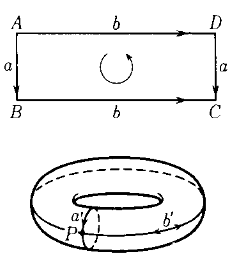


Fig. 2

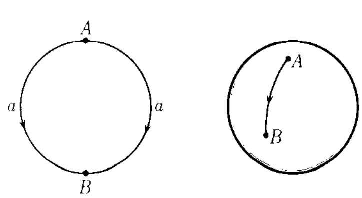


Fig. 3

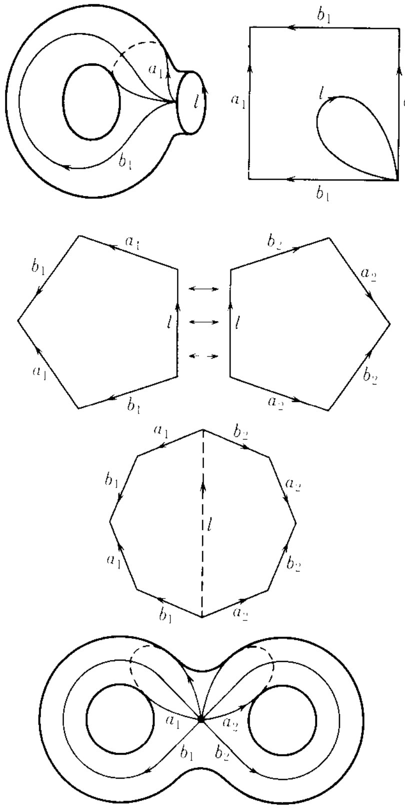


Fig. 4

All closed surfaces without boundary are constructed by identifying suitable pairs of sides of a $2n$-sided polygon in a Euclidean plane E^2. Furthermore, a closed orientable surface without boundary is homeomorphic to the surface represented by aa^{-1} or

$$a_1 b_1 a_1^{-1} b_1^{-1} \dots a_p b_p a_p^{-1} b_p^{-1}. \tag{1}$$

The 1-dimensional †Betti number of this surface is $2p$, the 0-dimensional and 2-dimensional †Betti numbers are 1, the †torsion coefficients are all 0, and p is called the **genus** of the surface. Also, a closed orientable surface of genus p with boundaries $c_1, \dots, c_k$ is represented by

$$w_1 c_1 w_1^{-1} \dots w_k c_k w_k^{-1} a_1 b_1 a_1^{-1} b_1^{-1} \dots a_p b_p a_p^{-1} b_p^{-1} \tag{2}$$

(Fig. 5). A closed nonorientable surface without boundary is represented by

$$a_1 a_1 a_2 a_2 \dots a_q a_q. \tag{3}$$

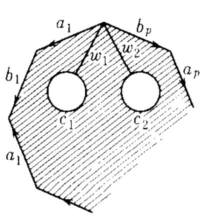


Fig. 5

The 1-dimensional Betti number of this surface is $q-1$, the 0-dimensional and 2-dimensional Betti numbers are 1 and 0, respectively, the 1-dimensional torsion coefficient is 2, the 0-dimensional and 2-dimensional torsion coefficients are 0, and q is called the **genus** of the surface. A closed nonorientable surface of genus q with boundaries $c_1, \dots, c_k$ is represented by

$$w_1 c_1 w_1^{-1} \dots w_k c_k w_k^{-1} a_1 a_1 \dots a_q a_q. \tag{4}$$

Each of forms (1)–(4) is called the **normal form** of the respective surface, and the curves a_i, b_j, w_k are called the **normal sections** of the surface. To explain the notation in (3), we first take the simplest case, aa. In this case, the surface is obtained from a disk by identifying each pair of points on the circumference that are endpoints of a diameter (Fig. 6). The surface aa is then homeomorphic to a †projective plane of which a decomposition into a complex of triangles is illustrated in Fig. 7. On the other hand, $aabb$ represents a surface like that shown in Fig. 8, called the **Klein bottle**. Fig. 9 shows a **handle**, and Fig. 10 shows a **cross cap**.

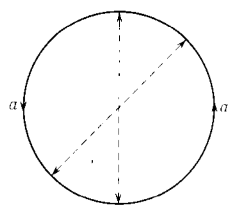


Fig. 6

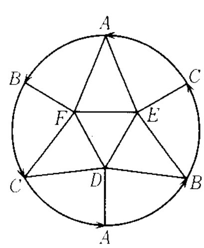


Fig. 7

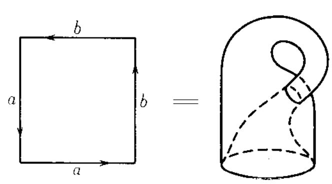


Fig. 8

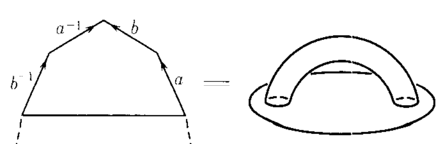


Fig. 9

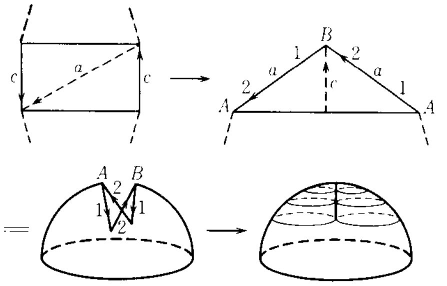


Fig. 10

The last two surfaces have boundaries; a handle is orientable, while a cross cap is nonorientable and homeomorphic to the Möbius strip. If we delete p disks from a sphere and replace them with an equal number of handles, then we obtain a surface homeomorphic to the surface represented in (1), while if we replace the disks by cross caps instead of by handles, then the surface thus obtained is homeomorphic to that represented in (3). Now we decompose the surfaces (1) and (3) into triangles and denote the number of i-dimensional simplexes by α_i $(i=0,1,2)$. Then in view of the †Euler-Poincaré formula, the surfaces (1) and (3) satisfy the respective formulas

$$\alpha_0-\alpha_1+\alpha_2=2(1-p),$$

$$\alpha_0-\alpha_1+\alpha_2=2-q.$$

The †Riemann surfaces of †algebraic functions of one complex variable are always surfaces of type (1), and their genera p coincide with those of algebraic functions.

All closed surfaces are homeomorphic to surfaces of types (1), (2), (3), or (4). A necessary and sufficient condition for two surfaces to be homeomorphic to each other is coincidence of the numbers of their boundaries, their orientability or nonorientability, and their genera (or †Euler characteristic $\alpha^0-\alpha^1+\alpha^2$). This proposition is called the **fundamental theorem of the topology of surfaces**. The †homeomorphism problem of closed surfaces is completely solved by this theorem. The same problem for n $(n\geqslant 3)$ manifolds, even if they are compact, remains open. (For surface area → 246 Length and Area. For the differential geometry of surfaces → 111 Differential Geometry of Curves and Surfaces.)

References

[1] B. Kerékjártó, Vorlesungen über Topologie, Springer, 1923.
[2] H. Seifert and W. Threlfall, Lehrbuch der Topologie, Teubner, 1934 (Chelsea, 1945).
[3] S. Lefschetz, Introduction to topology, Princeton Univ. Press, 1949.
[4] D. Hilbert and S. Cohn-Vossen, Anschauliche Geometrie, Springer, 1932; English translation, Geometry and the imagination, Chelsea, 1952.
[5] W. S. Massey, Algebraic topology: An introduction, Springer, 1967.
[6] E. E. Moise, Geometric topology in dimensions 2 and 3, Springer, 1977.

411 (I.4) Symbolic Logic

A. General Remarks

Symbolic logic (or **mathematical logic**) is a field of logic in which logical inferences commonly used in mathematics are investigated by use of mathematical symbols.

The **algebra of logic** originally set forth by G. Boole [1] and A. de Morgan [2] is actually an algebra of sets or relations; it did not reach the same level as the symbolic logic of today. G. Frege, who dealt not only with the logic of propositions but also with the first-order predicate logic using quantifiers (→ Sections C and K), should be regarded as the real originator of symbolic logic. Frege's work, however, was not recognized for some time. Logical studies by C. S. Peirce, E. Schröder, and G. Peano appeared soon after Frege, but they were limited mostly to propositions and did not develop Frege's work. An essential development of Frege's method was brought about by B. Russell, who, with the collaboration of A. N. Whitehead, summarized his results in *Principia mathematica* [4], which seemed to have completed the theory of symbolic logic at the time of its appearance.

B. Logical Symbols

If A and B are propositions, the propositions (A and B), (A or B), (A implies B), and (not A) are denoted by

$$A\wedge B,\qquad A\vee B,\qquad A\rightarrow B,\qquad \neg A,$$

respectively. We call $\neg A$ the **negation** of A, $A\wedge B$ the **conjunction** (or **logical product**), $A\vee B$ the **disjunction** (or **logical sum**), and $A\rightarrow B$ the **implication** (or B by A). The proposition $(A\rightarrow B)\wedge(B\rightarrow A)$ is denoted by $A\leftrightarrow B$ and is read "A and B are **equivalent**." $A\vee B$ means that at least one of A and B holds. The propositions (For all x, the proposition $F(x)$ holds) and (There exists an x such that $F(x)$ holds) are denoted by $\forall xF(x)$ and $\exists xF(x)$, respectively. A proposition of the form $\forall xF(x)$

is called a **universal proposition**, and one of the form $\exists xF(x)$, an **existential proposition**. The symbols $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$, $\neg$, $\forall$, $\exists$ are called **logical symbols**.

There are various other ways to denote logical symbols, including:

$A \wedge B$: $A \& B$, $A \cdot B$,

$A \vee B$: $A + B$,

$A \rightarrow B$: $A \supset B$, $A \Rightarrow B$,

$A \leftrightarrow B$: $A \rightleftarrows B$, $A \equiv B$, $A \sim B$, $A \supset \subset B$, $A \Leftrightarrow B$,

$\neg A$: $\sim A$, $\bar{A}$,

$\forall xF(x)$: $(x)F(x)$, $\prod xF(x)$, $\bigwedge xF(x)$,

$\exists xF(x)$: $(Ex)F(x)$, $\sum xF(x)$, $\bigvee xF(x)$.

C. Free and Bound Variables

Any function whose values are propositions is called a **propositional function**. $\forall x$ and $\exists x$ can be regarded as operators that transform any propositional function $F(x)$ into the propositions $\forall xF(x)$ and $\exists xF(x)$, respectively. $\forall x$ and $\exists x$ are called **quantifiers**; the former is called the **universal quantifier** and the latter the **existential quantifier**. $F(x)$ is transformed into $\forall xF(x)$ or $\exists xF(x)$ just as a function $f(x)$ is transformed into the definite integral $\int_0^1 f(x)dx$; the resultant propositions $\forall xF(x)$ and $\exists xF(x)$ are no longer functions of x. The variable x in $\forall xF(x)$ and in $\exists xF(x)$ is called a **bound variable**, and the variable x in $F(x)$, when it is not bound by $\forall x$ or $\exists x$, is called a **free variable**. Some people employ different kinds of symbols for free variables and bound variables to avoid confusion.

D. Formal Expressions of Propositions

A formal expression of a proposition in terms of logical symbols is called a **formula**. More precisely, formulas are constructed by the following **formation rules**: (1) If $\mathfrak{A}$ is a formula, $\neg\mathfrak{A}$ is also a formula. If $\mathfrak{A}$ and $\mathfrak{B}$ are formulas, $\mathfrak{A} \wedge \mathfrak{B}$, $\mathfrak{A} \vee \mathfrak{B}$, $\mathfrak{A} \rightarrow \mathfrak{B}$ are all formulas. (2) If $\mathfrak{F}(a)$ is a formula and a is a free variable, then $\forall x\mathfrak{F}(x)$ and $\exists x\mathfrak{F}(x)$ are formulas, where x is an arbitrary bound variable not contained in $\mathfrak{F}(a)$ and $\mathfrak{F}(x)$ is the result of substituting x for a throughout $\mathfrak{F}(a)$.

We use formulas of various scope according to different purposes. To indicate the scope of formulas, we fix a set of formulas, each element of which is called a **prime formula** (or **atomic formula**). The scope of formulas is the set of formulas obtained from the prime formulas by formation rules (1) and (2).

E. Propositional Logic

Propositional logic is the field in symbolic logic in which we study relations between propositions exclusively in connection with the four logical symbols $\wedge$, $\vee$, $\rightarrow$, and $\neg$, called **propositional connectives**.

In propositional logic, we deal only with operations of **logical operators** denoted by propositional connectives, regarding the variables for denoting propositions, called **proposition variables**, only as prime formulas. We examine problems such as: What kinds of formulas are identically true when their proposition variables are replaced by any propositions, and what kinds of formulas can sometimes be true?

Consider the two symbols $\curlyvee$ and $\curlywedge$, read **true** and **false**, respectively, and let $\mathbf{A} = \{\curlyvee, \curlywedge\}$. A univalent function from $\mathbf{A}$, or more generally from a Cartesian product $\mathbf{A} \times \ldots \times \mathbf{A}$, into $\mathbf{A}$ is called a **truth function**. We can regard $\wedge$, $\vee$, $\rightarrow$, $\neg$ as the following truth functions: (1) $A \wedge B = \curlyvee$ for $A = B = \curlyvee$, and $A \wedge B = \curlywedge$ otherwise; (2) $A \vee B = \curlywedge$ for $A = B = \curlywedge$, and $A \vee B = \curlyvee$ otherwise; (3) $A \rightarrow B = \curlywedge$ for $A = \curlyvee$ and $B = \curlywedge$, and $A \rightarrow B = \curlyvee$ otherwise; (4) $\neg A = \curlywedge$ for $A = \curlyvee$, and $\neg A = \curlyvee$ for $A = \curlywedge$.

If we regard proposition variables as variables whose domain is $\mathbf{A}$, then each formula represents a truth function. Conversely, any truth function (of a finite number of independent variables) can be expressed by an appropriate formula, although such a formula is not uniquely determined. If a formula is regarded as a truth function, the value of the function determined by a combination of values of the independent variables involved in the formula is called the **truth value** of the formula.

A formula corresponding to a truth function that takes only $\curlyvee$ as its value is called a **tautology**. For example, $\mathfrak{A} \vee \neg\mathfrak{A}$ and $((\mathfrak{A} \rightarrow \mathfrak{B}) \rightarrow \mathfrak{A}) \rightarrow \mathfrak{A}$ are tautologies. Since a truth function with n independent variables takes values corresponding to 2^n combinations of truth values of its variables, we can determine in a finite number of steps whether a given formula is a tautology. If $\mathfrak{A} \leftrightarrow \mathfrak{B}$ is a tautology (that is, $\mathfrak{A}$ and $\mathfrak{B}$ correspond to the same truth function), then the formulas $\mathfrak{A}$ and $\mathfrak{B}$ are said to be **equivalent**.

F. Propositional Calculus

It is possible to choose some specific tautologies, designate them as axioms, and derive all tautologies from them by appropriately given rules of inference. Such a system is called a **propositional calculus**. There are many ways

to stipulate axioms and rules of inference for a propositional calculus.

The abovementioned propositional calculus corresponds to the so-called classical propositional logic (→ Section L). By choosing appropriate axioms and rules of inference we can also formally construct intuitionistic or other propositional logics. In intuitionistic logic the law of the †excluded middle is not accepted, and hence it is impossible to formalize intuitionistic propositional logic by the notion of tautology. We therefore usually adopt the method of propositional calculus, instead of using the notion of tautology, to formalize intuitionistic propositional logic. For example, V. I. Glivenko's theorem [5], that if a formula $\mathfrak{A}$ can be proved in classical logic, then $\neg\neg\mathfrak{A}$ can be proved in intuitionistic logic, was obtained by such formalistic considerations. A method of extending the classical concepts of truth value and tautology to intuitionistic and other logics has been obtained by S. A. Kripke. There are also studies of logics intermediate between intuitionistic and classical logic (T. Umezawa).

G. Predicate Logic

Predicate logic is the area of symbolic logic in which we take quantifiers in account. Mainly propositional functions are discussed in predicate logic. In the strict sense only single-variable propositional functions are called **predicates**, but the phrase **predicate of** n **arguments** (or n**-ary predicate**) denoting an n-variable propositional function is also employed. Single-variable (or unary) predicates are also called **properties**. We say that a has the property F if the proposition $F(a)$ formed by the property F is true. Predicates of two arguments are called **binary relations**. The proposition $R(a,b)$ formed by the binary relation R is occasionally expressed in the form aRb. Generally, predicates of n arguments are called n**-ary relations**. The domain of definition of a unary predicate is called the **object domain**, elements of the object domain are called **objects**, and any variable running over the object domain is called an **object variable**. We assume here that the object domain is not empty. When we deal with a number of predicates simultaneously (with different numbers of variables), it is usual to arrange things so that all the independent variables have the same object domain by suitably extending their object domains.

Predicate logic in its purest sense deals exclusively with the general properties of quantifiers in connection with propositional connectives. The only objects dealt with in this field are **predicate variables** defined over a certain common domain and object variables running over the domain. Propositional variables are regarded as predicates of no variables. Each expression $F(a_1, \ldots, a_n)$ for any predicate variable F of n variables $a_1, \ldots, a_n$ (object variables designated as free) is regarded as a prime formula ($n = 0, 1, 2, \ldots$), and we deal exclusively with formulas generated by these prime formulas, where bound variables are also restricted to object variables that have a common domain. We give no specification for the range of objects except that it be the common domain of the object variables.

By designating an object domain and substituting a predicate defined over the domain for each predicate variable in a formula, we obtain a proposition. By substituting further an object (object constant) belonging to the object domain for each object variable in a proposition, we obtain a proposition having a definite truth value. When we designate an object domain and further associate with each predicate variable as well as with each object variable a predicate or an object to be substituted for it, we call the pair consisting of the object domain and the association a **model**. Any formula that is true for every model is called an **identically true formula** or **valid formula**. The study of identically true formulas is one of the most important problems in predicate logic.

H. Formal Representations of Mathematical Propositions

To obtain a formal representation of a mathematical theory by predicate logic, we must first specify its object domain, which is a nonempty set whose elements are called **individuals**; accordingly the object domain is called the **individual domain**, and object variables are called **individual variables**. Secondly we must specify **individual symbols**, **function symbols**, and **predicate symbols**, signifying specific individuals, functions, and †predicates, respectively. Here a function of n arguments is a univalent mapping from the Cartesian product $D \times \ldots \times D$ of n copies of the given set to D. Then we define the notion of **term** as in the next paragraph to represent each individual formally. Finally we express propositions formally by formulas.

Definition of terms (formation rule for terms): (1) Each individual symbol is a term. (2) Each free variable is a term. (3) $f(t_1, \ldots, t_n)$ is a term if $t_1, \ldots, t_n$ are terms and f is a function symbol of n arguments. (4) The only terms are those given by (1)–(3).

As a prime formula in this case we use any

formula of the form $F(t_1, \ldots, t_n)$, where F is a predicate symbol of n arguments and $t_1, \ldots, t_n$ are arbitrary terms. To define the notions of term and formula, we need logical symbols, free and bound individual variables, and also a list of individual symbols, function symbols, and predicate symbols.

In pure predicate logic, the individual domain is not concrete, and we study only general forms of propositions. Hence, in this case, predicate or function symbols are not representations of concrete predicates or functions but are **predicate variables** and **function variables**. We also use free individual variables instead of individual symbols. In fact, it is now most common that function variables are dispensed with, and only free individual variables are used as terms.

I. Formulation of Mathematical Theories

To formalize a theory we need **axioms** and **rules of inference**. Axioms constitute a certain specific set of formulas, and a rule of inference is a rule for deducing a formula from other formulas. A formula is said to be **provable** if it can be deduced from the axioms by repeated application of rules of inference. Axioms are divided into two types: **logical axioms**, which are common to all theories, and **mathematical axioms**, which are peculiar to each individual theory. The set of mathematical axioms is called the **axiom system** of the theory.

(I) Logical axioms: (1) A formula that is the result of substituting arbitrary formulas for the proposition variables in a tautology is an axiom. (2) Any formula of the form

$$\forall x \mathfrak{F}(x) \rightarrow \mathfrak{F}(t) \quad \text{or} \quad \mathfrak{F}(t) \rightarrow \exists x \mathfrak{F}(x)$$

is an axiom, where $\mathfrak{F}(t)$ is the result of substituting an arbitrary term t for x in $\mathfrak{F}(x)$.

(II) Rules of inference: (1) We can deduce a formula $\mathfrak{B}$ from two formulas $\mathfrak{A}$ and $\mathfrak{A} \rightarrow \mathfrak{B}$ (**modus ponens**). (2) We can deduce $\mathfrak{A} \rightarrow \forall x \mathfrak{F}(x)$ from a formula $\mathfrak{A} \rightarrow \mathfrak{F}(a)$ and $\exists x \mathfrak{F}(x) \rightarrow \mathfrak{A}$ from $\mathfrak{F}(a) \rightarrow \mathfrak{A}$, where a is a free individual variable contained in neither $\mathfrak{A}$ nor $\mathfrak{F}(x)$ and $\mathfrak{F}(a)$ is the result of substituting a for x in $\mathfrak{F}(x)$.

If an axiom system is added to these logical axioms and rules of inference, we say that a **formal system** is given.

A formal system S or its axiom system is said to be **contradictory** or to contain a **contradiction** if a formula $\mathfrak{A}$ and its negation $\neg \mathfrak{A}$ are provable; otherwise it is said to be **consistent**. Since

$$(\mathfrak{A} \wedge \neg \mathfrak{A}) \rightarrow \mathfrak{B}$$

is a tautology, we can show that any formula is provable in a formal system containing a contradiction. The validity of a proof by **reductio ad absurdum** lies in the fact that

$$(\mathfrak{A} \rightarrow (\mathfrak{B} \wedge \neg \mathfrak{B})) \rightarrow \neg \mathfrak{A}$$

is a tautology. An affirmative proposition (formula) may be obtained by reductio ad absurdum since the formula (of propositional logic) representing the **discharge of double negation**

$$\neg \neg \mathfrak{A} \rightarrow \mathfrak{A}$$

is a tautology.

J. Predicate Calculus

If a formula has no free individual variable, we call it a **closed formula**. Now we consider a formal system S whose mathematical axioms are closed. A formula $\mathfrak{A}$ is provable in S if and only if there exist suitable mathematical axioms $\mathfrak{C}_1, \ldots, \mathfrak{C}_n$ such that the formula

$$(\mathfrak{C}_1 \wedge \ldots \wedge \mathfrak{C}_n) \rightarrow \mathfrak{A}$$

is provable without the use of mathematical axioms. Since any axiom system can be replaced by an equivalent axiom system containing only closed formulas, the study of a formal system can be reduced to the study of pure logic.

In the following we take no individual symbols or function symbols into consideration and we use predicate variables as predicate symbols in accordance with the commonly accepted method of stating properties of the pure predicate logic; but only in the case of **predicate logic with equality** will we use predicate variables and the equality predicate $=$ as a predicate symbol. However, we can safely state that we use function variables as function symbols.

The formal system with no mathematical axioms is called the **predicate calculus**. The formal system whose mathematical axioms are the equality axioms

$$a = a, \quad a = b \ \rightarrow \ (\mathfrak{F}(a) \rightarrow \mathfrak{F}(b))$$

is called the **predicate calculus with equality**.

In the following, by being provable we mean being provable in the predicate calculus.

(1) Every provable formula is valid.

(2) Conversely, any valid formula is provable (K. Gödel [6]). This fact is called the **completeness** of the predicate calculus. In fact, by Gödel's proof, a formula $\mathfrak{A}$ is provable if $\mathfrak{A}$ is always true in every interpretation whose individual domain is of †countable cardinality. In another formulation, if $\neg \mathfrak{A}$ is not provable, the formula $\mathfrak{A}$ is a true proposition in some interpretation (and the individual domain in this case is of countable cardinality). We can

extend this result as follows: If an axiom system generated by countably many closed formulas is consistent, then its mathematical axioms can be considered true propositions by a common interpretation. In this sense, **Gödel's completeness theorem** gives another proof of the †Skolem-Löwenheim theorem.

(3) The predicate calculus is consistent. Although this result is obtained from (1) in this section, it is not difficult to show it directly (D. Hilbert and W. Ackermann [7]).

(4) There are many different ways of giving logical axioms and rules of inference for the predicate calculus. G. Gentzen gave two types of systems in [8]; one is a natural deduction system in which it is easy to reproduce formal proofs directly from practical ones in mathematics, and the other has a logically simpler structure. Concerning the latter, Gentzen proved **Gentzen's fundamental theorem**, which shows that a formal proof of a formula may be translated into a "direct" proof. The theorem itself and its idea were powerful tools for obtaining consistency proofs.

(5) If the proposition $\exists x\mathfrak{A}(x)$ is true, we choose one of the individuals x satisfying the condition $\mathfrak{A}(x)$, and denote it by $\varepsilon x\mathfrak{A}(x)$. When $\exists x\mathfrak{A}(x)$ is false, we let $\varepsilon x\mathfrak{A}(x)$ represent an arbitrary individual. Then

$$\exists x\mathfrak{A}(x)\rightarrow\mathfrak{A}(\varepsilon x\mathfrak{A}(x)) \qquad (1)$$

is true. We consider εx to be an operator associating an individual $\varepsilon x\mathfrak{A}(x)$ with a proposition $\mathfrak{A}(x)$ containing the variable x. Hilbert called it the **transfinite logical choice function**; today we call it **Hilbert's ε-operator** (or **ε-quantifier**), and the logical symbol ε used in this sense **Hilbert's ε-symbol.** Using the ε-symbol, $\exists x\mathfrak{A}(x)$ and $\forall x\mathfrak{A}(x)$ are represented by

$$\mathfrak{A}(\varepsilon x\mathfrak{A}(x)), \quad \mathfrak{A}(\varepsilon x \neg \mathfrak{A}(x)),$$

respectively, for any $\mathfrak{A}(x)$. The system of predicate calculus adding formulas of the form (1) as axioms is essentially equivalent to the usual predicate calculus. This result, called the **ε-theorem**, reads as follows: When a formula $\mathfrak{C}$ is provable under the assumption that every formula of the form (1) is an axiom, we can prove $\mathfrak{C}$ using no axioms of the form (1) if $\mathfrak{C}$ contains no logical symbol ε (D. Hilbert and P. Bernays [9]). Moreover, a similar theorem holds when axioms of the form

$$\forall x(\mathfrak{A}(x)\leftrightarrow\mathfrak{B}(x))\rightarrow\varepsilon x\mathfrak{A}(x)=\varepsilon x\mathfrak{B}(x) \qquad (2)$$

are added (S. Maehara [10]).

(6) For a given formula $\mathfrak{A}$, call $\mathfrak{A}'$ a normal form of $\mathfrak{A}$ when the formula

$$\mathfrak{A}\leftrightarrow\mathfrak{A}'$$

is provable and $\mathfrak{A}'$ satisfies a particular condition. For example, for any formula $\mathfrak{A}$ there is a normal form $\mathfrak{A}'$ satisfying the condition: $\mathfrak{A}'$ has the form

$$Q_1x_1\dots Q_nx_n\mathfrak{B}(x_1,\dots,x_n),$$

where Qx means a quantifier $\forall x$ or $\exists x$, and $\mathfrak{B}(x_1,\dots,x_n)$ contains no quantifier and has no predicate variables or free individual variables not contained in $\mathfrak{A}$. A normal form of this kind is called a **prenex normal form.**

(7) We have dealt with the classical first-order predicate logic until now. For other predicate logics (→ Sections K and L) also, we can consider a predicate calculus or a formal system by first defining suitable axioms or rules of inference. Gentzen's fundamental theorem applies to the intuitionistic predicate calculus formulated by V. I. Glivenko, A. Heyting, and others. Since Gentzen's fundamental theorem holds not only in classical logic and intuitionistic logic but also in several systems of first-order predicate logic or propositional logic, it is useful for getting results in modal and other logics (M. Ohnishi, K. Matsumoto). Moreover, Glivenko's theorem in propositional logic [5] is also extended to predicate calculus by using a rather weak representation (S. Kuroda [12]). G. Takeuti expected that a theorem similar to Gentzen's fundamental theorem would hold in higher-order predicate logic also, and showed that the consistency of analysis would follow if that conjecture could be verified [13]. Moreover, in many important cases, he showed constructively that the conjecture holds partially. The conjecture was finally proved by M. Takahashi [14] by a nonconstructive method. Concerning this, there are also contributions by S. Maehara, T. Simauti, M. Yasuhara, and W. Tait.

K. Predicate Logics of Higher Order

In ordinary predicate logic, the bound variables are restricted to individual variables. In this sense, ordinary predicate logic is called **first-order predicate logic**, while predicate logic dealing with quantifiers $\forall P$ or $\exists P$ for a predicate variable P is called **second-order predicate logic.**

Generalizing further, we can introduce the so-called **third-order predicate logic**. First we fix the individual domain D_0. Then, by introducing the whole class D_1^n of predicates of n variables, each running over the object domain D_0, we can introduce predicates that have D_1^n as their object domain. This kind of predicate is called a **second-order predicate** with respect to the individual domain D_0. Even when we restrict second-order predicates to one-variable predicates, they are divided into vari-

ous types, and the domains of independent variables do not coincide in the case of more than two variables. In contrast, predicates having D_0 as their object domain are called **first-order predicates**. The logic having quantifiers that admit first-order predicate variables is second-order predicate logic, and the logic having quantifiers that admit up to second-order predicate variables is third-order predicate logic. Similarly, we can define further **higher-order predicate logics**.

Higher-order predicate logic is occasionally called **type theory**, because variables arise that are classified into various types. Type theory is divided into **simple type theory** and **ramified type theory**.

We confine ourselves to variables for single-variable predicates, and denote by P such a bound predicate variable. Then for any formula $\mathfrak{F}(a)$ (with a a free individual variable), the formula

$$\exists P \forall x (P(x) \leftrightarrow \mathfrak{F}(x))$$

is considered identically true. This is the point of view in simple type theory.

Russell asserted first that this formula cannot be used reasonably if quantifiers with respect to predicate variables occur in $\mathfrak{F}(x)$. This assertion is based on the point of view that the formula in the previous paragraph asserts that $\mathfrak{F}(x)$ is a first-order predicate, whereas any quantifier with respect to first-order predicate variables, whose definition assumes the totality of the first-order predicates, should not be used to introduce the first-order predicate $\mathfrak{F}(x)$. For this purpose, Russell further classified the class of first-order predicates by their **rank** and adopted the axiom

$$\exists P^k \forall x (P^k(x) \leftrightarrow \mathfrak{F}(x))$$

for the predicate variable P^k of rank k, where the rank i of any free predicate variable occurring in $\mathfrak{F}(x)$ is $\leqslant k$, and the rank j of any bound predicate variable occurring in $\mathfrak{F}(x)$ is $< k$. This is the point of view in ramified type theory, and we still must subdivide the types if we deal with higher-order propositions or propositions of many variables. Even Russell, having started from his ramified type theory, had to introduce the **axiom of reducibility** afterwards and reduce his theory to simple type theory.

L. Systems of Logic

Logic in the ordinary sense, which is based on the **law of the excluded middle** asserting that every proposition is in principle either true or false, is called **classical logic**. Usually, propositional logic, predicate logic, and type theory are developed from the standpoint of classical logic. Occasionally the reasoning of intuitionistic mathematics is investigated using symbolic logic, in which the law of the excluded middle is not admitted (→ 156 Foundations of Mathematics). Such logic is called **intuitionistic logic**. Logic is also subdivided into propositional logic, predicate logic, etc., according to the extent of the propositions (formulas) dealt with.

To express **modal propositions** stating **possibility, necessity**, etc., in symbolic logic, J. Łukaszewicz proposed a propositional logic called **three-valued logic**, having a third truth value, neither true nor false. More generally, **many-valued logics** with any number of truth values have been introduced; classical logic is one of its special cases, **two-valued logic** with two truth values, true and false. Actually, however, many-valued logics with more than three truth values have not been studied much, while various studies in **modal logic** based on classical logic have been successfully carried out. For example, studies of **strict implication** belong to this field.

References

[1] G. Boole, An investigation of the laws of thought, Walton and Maberly, 1854.
[2] A. de Morgan, Formal logic, or the calculus of inference, Taylor and Walton, 1847.
[3] G. Frege, Begriffsschrift, eine der arithmetischen nachgebildete Formalsprache des reinen Denkens, Halle, 1879.
[4] A. N. Whitehead and B. Russell, Principia mathematica I, II, III, Cambridge Univ. Press, 1910–1913; second edition, 1925–1927.
[5] V. Glivenko, Sur quelques points de la logique de M. Brouwer, Acad. Roy. de Belgique, Bulletin de la classe des sciences, (5) 15 (1929), 183–188.
[6] K. Gödel, Die Vollständigkeit der Axiome des logischen Funktionenkalküls, Monatsh. Math. Phys., 37 (1930), 349–360.
[7] D. Hilbert and W. Ackermann, Grundzüge der theoretischen Logik, Springer, 1928, sixth edition, 1972; English translation, Principles of mathematical logic, Chelsea, 1950.
[8] G. Gentzen, Untersuchungen über das logische Schliessen, Math. Z., 39 (1935), 176–210, 405–431.
[9] D. Hilbert and P. Bernays, Grundlagen der Mathematik II, Springer, 1939; second edition, 1970.
[10] S. Maehara, Equality axiom on Hilbert's ε-symbol, J. Fac. Sci. Univ. Tokyo, (I), 7 (1957), 419–435.

[11] A. Heyting, Die formalen Regeln der intuitionistischen Logik I, S.-B. Preuss. Akad. Wiss., 1930, 42–56.
[12] S. Kuroda, Intuitionistische Untersuchungen der formalistischen Logik, Nagoya Math. J., 2 (1951), 35–47.
[13] G. Takeuti, On a generalized logic calculus, Japan. J. Math., 23 (1953), 39–96.
[14] M. Takahashi, A proof of the cut-elimination theorem in simple type-theory, J. Math. Soc. Japan, 19 (1967), 399–410.
[15] S. C. Kleene, Mathematical logic, Wiley, 1967.
[16] J. R. Shoenfield, Mathematical logic, Addison-Wesley, 1967.
[17] R. M. Smullyan, First-order logic, Springer, 1968.

412 (IV.13) Symmetric Riemannian Spaces and Real Forms

A. Symmetric Riemannian Spaces

Let M be a †Riemannian space. For each point p of M we can define a mapping σ_p of a suitable neighborhood U_p of p onto U_p itself so that $\sigma_p(x_t)=x_{-t}$, where x_t ($|t|<\varepsilon, x_0=p$) is any †geodesic passing through the point p. We call M a **locally symmetric Riemannian space** if for any point p of M we can choose a neighborhood U_p so that σ_p is an †isometry of U_p. In order that a Riemannian space M be locally symmetric it is necessary and sufficient that the †covariant differential (with respect to the †Riemannian connection) of the †curvature tensor of M be 0. A locally symmetric Riemannian space is a †real analytic manifold. We say that a Riemannian space M is a **globally symmetric Riemannian space** (or simply **symmetric Riemannian space**) if M is connected and if for each point p of M there exists an isometry σ_p of M onto M itself that has p as an isolated fixed point (i.e., has no fixed point except p in a certain neighborhood of p) and such that σ_p^2 is the identity transformation on M. In this case σ_p is called the **symmetry** at p. A (globally) symmetric Riemannian space is locally symmetric and is a †complete Riemannian space. Conversely, a †simply connected complete locally symmetric Riemannian space is a (globally) symmetric Riemannian space.

B. Symmetric Riemannian Homogeneous Spaces

A †homogeneous space G/K of a connected †Lie group G is a **symmetric homogeneous space** (with respect to θ) if there exists an **involutive automorphism** (i.e., automorphism of order 2) θ of G satisfying the condition $K_\theta^0\subset K\subset K_\theta$, where K_θ is the closed subgroup consisting of all elements of G left fixed by θ and K_θ^0 is the connected component of the identity element of K_θ. In this case, the mapping $aK\rightarrow\theta(a)K$ ($a\in G$) is a transformation of G/K having the point K as an isolated fixed point; more generally, the mapping $\theta_{a_0}: aK\rightarrow a_0\theta(a_0)^{-1}\theta(a)K$ is a transformation of G/K that has an arbitrary given point a_0K of G/K as an isolated fixed point. If there exists a G-invariant Riemannian metric on G/K, then G/K is a symmetric Riemannian space with symmetries $\{\theta_{a_0}\,|\,a_0\in G\}$ and is called a **symmetric Riemannian homogeneous space**. A sufficient condition for a symmetric homogeneous space G/K to be a symmetric Riemannian homogeneous space is that K be a compact subgroup. Conversely, given a symmetric Riemannian space M, let G be the connected component of the identity element of the Lie group formed by all the isometries of M; then M is represented as the symmetric Riemannian homogeneous space $M=G/K$ and K is a compact group. In particular, a symmetric Riemannian space can be regarded as a Riemannian space that is realizable as a symmetric Riemannian homogeneous space.

The Riemannian connection of a symmetric Riemannian homogeneous space G/K is uniquely determined (independent of the choice of G-invariant Riemannian metric), and a geodesic $x_t(|t|<\infty, x_0=a_0K)$ passing through a point a_0K of G/K is of the form $x_t=(\exp tX)a_0K$. Here X is any element of the Lie algebra $\mathfrak{g}$ of G such that $\theta(X)=-X$, where θ also denotes the automorphism of $\mathfrak{g}$ induced by the automorphism θ of G and $\exp tX$ is the †one-parameter subgroup of G defined by the element X. The covariant differential of any G-invariant tensor field on G/K is 0, and any G-invariant †differential form on G/K is a closed differential form.

C. Classification of Symmetric Riemannian Spaces

The †simply connected †covering Riemannian space of a symmetric Riemannian space is also a symmetric Riemannian space. Therefore the problem of classifying symmetric Riemannian spaces is reduced to classifying simply connected symmetric Riemannian spaces M and determining †discontinuous groups of isometries of M. When we take the †de Rham decomposition of such a space M and represent M as the product of a real Euclidean space and a number of simply connected irre-

ducible Riemannian spaces, all the factors are symmetric Riemannian spaces. We say that M is an **irreducible symmetric Riemannian space** if it is a symmetric Riemannian space and is irreducible as a Riemannian space.

A simply connected irreducible symmetric Riemannian space is isomorphic to one of the following four types of symmetric Riemannian homogeneous spaces (here Lie groups are always assumed to be connected):

(1) The symmetric Riemannian homogeneous space $(G \times G)/\{(a,a) \mid a \in G\}$ of the direct product $G \times G$, where G is a simply connected compact †simple Lie group and the involutive automorphism of $G \times G$ is given by $(a,b) \rightarrow (b,a)$ $((a,b) \in G \times G)$. This space is isomorphic, as a Riemannian space, to the space G obtained by introducing a two-sided invariant Riemannian metric on the group G; the isomorphism is induced from the mapping $G \times G \ni (a,b) \rightarrow ab^{-1} \in G$.

(2) A symmetric homogeneous space G/K_θ of a simply connected compact simple Lie group G with respect to an involutive automorphism θ of G. In this case, the closed subgroup $K_\theta = \{a \in G \mid \theta(a) = a\}$ of G is connected. We assume here that θ is a member of the given complete system of representatives of the †conjugate classes formed by the elements of order 2 in the automorphism group of the group G.

(3) The homogeneous space $G^{\mathbf{C}}/G$, where $G^{\mathbf{C}}$ is a complex simple Lie group whose †center reduces to the identity element and G is an arbitrary but fixed maximal compact subgroup of $G^{\mathbf{C}}$.

(4) The homogeneous space G_0/K, where G_0 is a noncompact simple Lie group whose center reduces to the identity element and which has no complex Lie group structure, and K is a maximal compact subgroup of G. In Section D we shall see that (3) and (4) are actually symmetric homogeneous spaces. All four types of symmetric Riemannian spaces are actually irreducible symmetric Riemannian spaces, and G-invariant Riemannian metrics on each of them are uniquely determined up to multiplication by a positive number. On the other hand, (1) and (2) are compact, while (3) and (4) are homeomorphic to Euclidean spaces and not compact. For spaces of types (1) and (3) the problem of classifying simply connected irreducible symmetric Riemannian spaces is reduced to classifying †compact real simple Lie algebras and †complex simple Lie algebras, respectively, while for types (2) and (4) it is reduced to the classification of noncompact real simple Lie algebras (→ Section D) (for the result of classification of these types → Appendix A, Table 5.II). On the other hand, any (not necessarily simply connected) irreducible symmetric Riemannian space defines one of (1)–(4) as its †universal covering manifold; if the covering manifold is of type (3) or (4), the original symmetric Riemannian space is necessarily simply connected.

D. Symmetric Riemannian Homogeneous Spaces of Semisimple Lie Groups

In Section C we saw that any irreducible symmetric Riemannian space is representable as a symmetric Riemannian homogeneous space G/K on which a connected semisimple Lie group G acts †almost effectively (→ 249 Lie Groups). Among symmetric Riemannian spaces, such a space $M = G/K$ is characterized as one admitting no nonzero vector field that is †parallel with respect to the Riemannian connection. Furthermore, if G acts effectively on M, G coincides with the connected component $I(M)^0$ of the identity element of the Lie group formed by all the isometries of M.

We let $M = G/K$ be a symmetric Riemannian homogeneous space on which a connected semisimple Lie group G acts almost effectively. Then G is a Lie group that is †locally isomorphic to the group $I(M)^0$, and therefore the Lie algebra of G is determined by M. Let $\mathfrak{g}$ be the Lie algebra of G, $\mathfrak{k}$ be the subalgebra of $\mathfrak{g}$ corresponding to K, and θ be the involutive automorphism of G defining the symmetric homogeneous space G/K. The automorphism of $\mathfrak{g}$ defined by θ is also denoted by θ. Then $\mathfrak{k} = \{X \in \mathfrak{g} \mid \theta(X) = X\}$. Putting $\mathfrak{m} = \{X \in \mathfrak{g} \mid \theta(X) = -X\}$, we have $\mathfrak{g} = \mathfrak{m} + \mathfrak{k}$ (direct sum of linear spaces), and $\mathfrak{m}$ can be identified in a natural way with the tangent space at the point K of G/K. The †adjoint representation of G gives rise to a representation of K in $\mathfrak{g}$, which induces a linear representation $\mathrm{Ad}_{\mathfrak{m}}(k)$ of K in $\mathfrak{m}$. Then $\{\mathrm{Ad}_{\mathfrak{m}}(k) \mid k \in K\}$ coincides with the †restricted homogeneous holonomy group at the point K of the Riemannian space G/K.

Now let φ be the †Killing form of $\mathfrak{g}$. Then $\mathfrak{k}$ and $\mathfrak{m}$ are mutually orthogonal with respect to φ, and denoting by $\varphi_{\mathfrak{k}}$ and $\varphi_{\mathfrak{m}}$ the restrictions of φ to $\mathfrak{k}$ and $\mathfrak{m}$, respectively, $\varphi_{\mathfrak{k}}$ is a negative definite quadratic form on $\mathfrak{k}$. If $\varphi_{\mathfrak{m}}$ is also a negative definite quadratic form on $\mathfrak{m}$, $\mathfrak{g}$ is a compact real semisimple Lie algebra and G/K is a compact symmetric Riemannian space; in this case we say that G/K is of **compact type**. In the opposite case, where $\varphi_{\mathfrak{m}}$ is a †positive definite quadratic form, G/K is said to be of **noncompact type**. In this latter case, G/K is homeomorphic to a Euclidean space, and if the center of G is finite, K is a maximal compact subgroup of G. Furthermore, the group of isometries $I(G/K)$ of G/K is canonically

isomorphic to the automorphism group of the Lie algebra $\mathfrak{g}$. When G/K is of compact type (noncompact type), there exists one and only one G-invariant Riemannian metric on G/K, which induces in the tangent space $\mathfrak{m}$ at the point K the positive definite inner product $-\varphi_{\mathfrak{m}}(\varphi_{\mathfrak{m}})$.

A symmetric Riemannian homogeneous space G/K_θ of compact type defined by a simply connected compact semisimple Lie group G with respect to an involutive automorphism θ is simply connected. Let $\mathfrak{g}=\mathfrak{m}+\mathfrak{k}_\theta$ be the decomposition of the Lie algebra $\mathfrak{g}$ of G with respect to the automorphism θ of $\mathfrak{g}$, and let $\mathfrak{g}^{\mathbf{C}}$ be the †complex form of $\mathfrak{g}$. Then the real subspace $\mathfrak{g}_\theta=\sqrt{-1}\,\mathfrak{m}+\mathfrak{k}_\theta$ in $\mathfrak{g}^{\mathbf{C}}$ is a real semisimple Lie algebra and a †real form of $\mathfrak{g}^{\mathbf{C}}$. Let G_θ be the Lie group corresponding to the Lie algebra $\mathfrak{g}_\theta$ with center reduced to the identity element, and let K be the subgroup of G_θ corresponding to $\mathfrak{k}_\theta$. Then we get a (simply connected) symmetric Riemannian homogeneous space of noncompact type G_θ/K.

When we start from a symmetric Riemannian space of noncompact type G/K instead of the symmetric Riemannian space of compact type G/K_θ and apply the same process as in the previous paragraphs, taking a simply connected G_θ as the Lie group corresponding to $\mathfrak{g}_\theta$, we obtain a simply connected symmetric Riemannian homogeneous space of compact type. Indeed, each of these two processes is the reverse of the other, and in this way we get a one-to-one correspondence between simply connected symmetric Riemannian homogeneous spaces of compact type and those of noncompact type. This relationship is called **duality** for symmetric Riemannian spaces; when two symmetric Riemannian spaces are related by duality, each is said to be the **dual** of the other.

If one of the two symmetric Riemannian spaces related by duality is irreducible, the other is also irreducible. The duality holds between spaces of types (1) and (3) and between those of types (2) and (4) described in Section C. This fact is based on the following theorem in the theory of Lie algebras, where we identify isomorphic Lie algebras. (i) Complex simple Lie algebras $\mathfrak{g}^{\mathbf{C}}$ and compact real simple Lie algebras $\mathfrak{g}$ are in one-to-one correspondence by the relation that $\mathfrak{g}^{\mathbf{C}}$ is the complex form of $\mathfrak{g}$. (ii) Form the Lie algebra $\mathfrak{g}_\theta$ in the above way from a compact real simple Lie algebra $\mathfrak{g}$ and an involutive automorphism θ of $\mathfrak{g}$. We assume that θ is a member of the given complete system of representatives of conjugate classes of involutive automorphisms in the automorphism group of $\mathfrak{g}$. Then we get from the pair $(\mathfrak{g},\theta)$ a noncompact real simple Lie algebra $\mathfrak{g}_\theta$, and any noncompact real simple Lie algebra is obtained by this process in one and only one way.

Consider a Riemannian space given as a symmetric Riemannian homogeneous space $M=G/K$ with a semisimple Lie group G, and let K be the †sectional curvature of M. Then if M is of compact type the value of K is $\geqslant 0$, and if M is of noncompact type it is $\leqslant 0$. On the other hand, the **rank** of M is the (unique) dimension of a commutative subalgebra of $\mathfrak{g}$ that is contained in and maximal in $\mathfrak{m}$. (For results concerning the group of isometries of M, distribution of geodesics on M, etc. → [3].)

E. Symmetric Hermitian Spaces

A connected †complex manifold M with a †Hermitian metric is called a **symmetric Hermitian space** if for each point p of M there exists an isometric and †biholomorphic transformation of M onto M that is of order 2 and has p as an isolated fixed point. As a real analytic manifold, such a space M is a symmetric Riemannian space of even dimension, and the Hermitian metric of M is a †Kähler metric. Let $I(M)$ be the (not necessarily connected) Lie group formed by all isometries of M, and let $A(M)$ be the subgroup consisting of all holomorphic transformations in $I(M)$. Then $A(M)$ is a closed Lie subgroup of $I(M)$. Let G be the connected component $A(M)^0$ of the identity element of $A(M)$. Then G acts transitively on M, and M is expressed as a symmetric Riemannian homogeneous space G/K.

Under the de Rham decomposition of a simply connected symmetric Hermitian space (regarded as a Riemannian space), all the factors are symmetric Hermitian spaces. The factor that is isomorphic to a real Euclidean spaces as a Riemannian space is a symmetric Hermitian space that is isomorphic to the complex Euclidean space $\mathbf{C}^n$. A symmetric Hermitian space defining an irreducible symmetric Riemannian space is called an **irreducible symmetric Hermitian space**. The problem of classifying symmetric Hermitian spaces is thus reduced to classifying irreducible symmetric Hermitian spaces.

In general, if the symmetric Riemannian space defined by a symmetric Hermitian space M is represented as a symmetric Riemannian homogeneous space G/K by a connected semisimple Lie group G acting effectively on M, then M is simply connected, G coincides with the group $A(M)^0$ introduced in the previous paragraph, and the center of K is not a †discrete set. In particular, an irreducible symmetric Hermitian space is simply connected. Moreover, in order for an irreducible symmetric Riemannian homogeneous space G/K to be defined by an irreducible symmetric Hermitian

space M, it is necessary and sufficient that the center of K not be a discrete set. If G acts effectively on M, then G is a simple Lie group whose center is reduced to the identity element, and the center of K is of dimension 1. For a space G/K satisfying these conditions, there are two kinds of structures of symmetric Hermitian spaces defining the Riemannian structure of G/K.

As follows from the classification of irreducible symmetric Riemannian spaces, an irreducible Hermitian space defines one of the following symmetric Riemannian homogeneous spaces, and conversely, each of these homogeneous spaces is defined by one of the two kinds of symmetric Hermitian spaces.

(I) The symmetric homogeneous space G/K of a compact simple Lie group G with respect to an involutive automorphism θ such that the center of G reduces to the identity element and the center of K is not a discrete set. Here θ may be assumed to be a representative of a conjugate class of involutive automorphisms in the automorphism group of G.

(II) The homogeneous space G_0/K of a noncompact simple Lie group G_0 by a maximal compact subgroup K such that the center of G_0 reduces to the identity element and the center of K is not a discrete set.

An irreducible symmetric Hermitian space of type (I) is compact and is isomorphic to a †rational algebraic variety. An irreducible symmetric Hermitian space of type (II) is homeomorphic to a Euclidean space and is isomorphic (as a complex manifold) to a bounded domain in $\mathbf{C}^n$ (Section F).

By the same principle as for irreducible symmetric Riemannian spaces, a duality holds for irreducible symmetric Hermitian spaces which establishes a one-to-one correspondence between the spaces of types (I) and (II). Furthermore, an irreducible symmetric Hermitian space M_b of type (II) that is dual to a given irreducible symmetric Hermitian space $M_a = G/K$ of type (I) can be realized as an open complex submanifold of M_a in the following way. Let $G^{\mathbf{C}}$ be the connected component of the identity element in the Lie group formed by all the holomorphic transformations of M_a. Then $G^{\mathbf{C}}$ is a complex simple Lie group containing G as a maximal compact subgroup, and the complex Lie algebra $\mathfrak{g}^{\mathbf{C}}$ of $G^{\mathbf{C}}$ contains the Lie algebra $\mathfrak{g}$ of G as a real form. Let θ be the involutive automorphism of G defining the symmetric homogeneous space G/K, and let $\mathfrak{g} = \mathfrak{m} + \mathfrak{k}$ be the decomposition of $\mathfrak{g}$ determined by θ. We denote by G_0 the real subgroup of $G^{\mathbf{C}}$ corresponding to the real form $\mathfrak{g}_0 = \sqrt{-1}\,\mathfrak{m} + \mathfrak{k}$ of $\mathfrak{g}^{\mathbf{C}}$. Then G_0 (i) is a closed subgroup of $G^{\mathbf{C}}$ whose center reduces to the identity element and (ii) contains K as a maximal compact subgroup. By definition the space M_b is then given by G_0/K. Now the group G_0 acts on M_a as a subgroup of $G^{\mathbf{C}}$, and the orbit of G_0 containing the point K of M_a is an open complex submanifold that is isomorphic to M_b (as a complex manifold). M_a regarded as a complex manifold can be represented as the homogeneous space $G^{\mathbf{C}}/U$ of the complex simple Lie group $G^{\mathbf{C}}$.

F. Symmetric Bounded Domains

We denote by D a bounded domain in the complex Euclidean space $\mathbf{C}^n$ of dimension n. We call D a **symmetric bounded domain** if for each point of D there exists a holomorphic transformation of order 2 of D onto D having the point as an isolated fixed point. On the other hand, the group of all holomorphic transformations of D is a Lie group, and D is called a **homogeneous bounded domain** if this group acts transitively on D. A symmetric bounded domain is a homogeneous bounded domain. The following theorem gives more precise results: On a bounded domain D, †Bergman's kernel function defines a Kähler metric that is invariant under all holomorphic transformations of D. If D is a symmetric bounded domain, D is a symmetric Hermitian space with respect to this metric, and its defining Riemannian space is a symmetric Riemannian homogeneous space of noncompact type G/K with semisimple Lie group G. Conversely, any symmetric Hermitian space of noncompact type is isomorphic (as a complex manifold) to a symmetric bounded domain. When D is isomorphic to an irreducible symmetric Hermitian space, we call D an **irreducible symmetric bounded domain**. A symmetric bounded domain is simply connected and can be decomposed into the direct product of irreducible symmetric bounded domains.

The connected component of the identity element of the group of all holomorphic transformations of a symmetric bounded domain D is a semisimple Lie group that acts transitively on D. Conversely, D is a symmetric bounded domain if a connected semisimple Lie group, or more generally, a connected Lie group admitting a two-sided invariant †Haar measure, acts transitively on D. Homogeneous bounded domains in $\mathbf{C}^n$ are symmetric bounded domains if $n \leqslant 3$ but not necessarily when $n \geqslant 4$.

G. Examples of Irreducible Symmetric Riemannian Spaces

Here we list irreducible symmetric Riemannian spaces of types (2) and (4) (→ Section C) that

can be represented as homogeneous spaces of classical groups, using the notation introduced by E. Cartan. We denote by $M_u = G/K$ a simply connected irreducible symmetric Riemannian space of type (2), where G is a group that acts almost effectively on M_u and K is the subgroup given by $K = K_\theta^0$ for an involutive automorphism θ of G. For such an M_u, the space of type (4) that is dual to M_u is denoted by $M_\theta = G_\theta/K$. Clearly $\dim M_u = \dim M_\theta$. (For the dimension and rank of M_u and for those M_u that are represented as homogeneous spaces of simply connected †exceptional compact simple Lie groups → Appendix A, Table 5.III.) In this section (and also in Appendix A, Table 5.III), $O(n)$, $U(n)$, $Sp(n)$, $SL(n, \mathbf{R})$, and $SL(n, \mathbf{C})$ are the †orthogonal group of degree n, the †unitary group of degree n, the †symplectic group of degree $2n$, and the real and complex †special linear groups of degree n, respectively. Let $SO(n) = SL(n, \mathbf{R}) \cap O(n)$ and $SU(n) = SL(n, \mathbf{C}) \cap U(n)$. We put

$$I_{p,q} = \begin{pmatrix} -I_p & 0 \\ 0 & I_q \end{pmatrix}, \quad J_n = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix},$$

$$K_{p,q} = \begin{pmatrix} -I_p & 0 & 0 & 0 \\ 0 & I_q & 0 & 0 \\ 0 & 0 & -I_p & 0 \\ 0 & 0 & 0 & I_q \end{pmatrix},$$

where I_p is the $p \times p$ unit matrix.

Type AI. $M_u = SU(n)/SO(n)$ $(n > 1)$, where $\theta(s) = \bar{s}$ (with $\bar{s}$ the complex conjugate matrix of s). $M_\theta = SL(n, \mathbf{R})/SO(n)$.

Type AII. $M_u = SU(2n)/Sp(n)$ $(n > 1)$, where $\theta(s) = J_n s J_n^{-1}$. $M_\theta = SU^*(2n)/Sp(n)$. Here $SU^*(2n)$ is the subgroup of $SL(2n, \mathbf{C})$ formed by the matrices that commute with the transformation $(z_1, \ldots, z_n, z_{n+1}, \ldots, z_{2n}) \to (\bar{z}_{n+1}, \ldots, \bar{z}_{2n}, -\bar{z}_1, \ldots, -\bar{z}_n)$ in $\mathbf{C}^n$; $SU^*(2n)$ is called the **quaternion unimodular group** and is isomorphic to the commutator group of the group of all regular transformations in an n-dimensional vector space over the quaternion field $\mathbf{H}$.

Type AIII. $M_u = SU(p+q)/S(U_p \times U_q)$ $(p \geqslant q \geqslant 1)$, where $S(U_p \times U_q) = SU(p+q) \cap (U(p) \times U(q))$, with $U(p) \times U(q)$ being canonically identified with a subgroup of $U(p+q)$, and $\theta(s) = I_{p,q} s I_{p,q}$. This space M_u is a †complex Grassmann manifold. $M_\theta = SU(p,q)/S(U_p \times U_q)$, where $SU(p,q)$ is the subgroup of $SL(p+q, \mathbf{C})$ consisting of matrices that leave invariant the Hermitian form $z_1\bar{z}_1 + \ldots + z_p\bar{z}_p - z_{p+1}\bar{z}_{p+1} - \ldots - z_{p+q}\bar{z}_{p+q}$.

Type AIV. This is the case $q = 1$ of type AIII. M_u is the $(n-1)$-dimensional complex projective space, and M_θ is called a **Hermitian hyperbolic space**.

Type BDI. $M_u = SO(p+q)/SO(p) \times SO(q)$ $(p \geqslant q \geqslant 1, p > 1, p+q \neq 4)$, where $\theta(s) = I_{p,q} s I_{p,q}$. M_u is the †real Grassmann manifold formed by the oriented p-dimensional subspaces in $\mathbf{R}^{p+q}$. $M_\theta = SO_0(p,q)/SO(p) \times SO(q)$, where $SO(p,q)$ is the subgroup of $SL(n, \mathbf{R})$ consisting of matrices that leave invariant the quadratic form $x_1^2 + \ldots + x_p^2 - x_{p+1}^2 - \ldots - x_{p+q}^2$, and $SO_0(p,q)$ is the connected component of the identity element.

Type BDII. This is the case $q = 1$ of type BDI. M_u is the $(n-1)$-dimensional sphere, and M_θ is called a **real hyperbolic space**.

Type DIII. $M_u = SO(2n)/U(n)$ $(n > 2)$, where $U(n)$ is regarded as a subgroup of $SO(2n)$ by identifying $s \in U(n)$ with

$$\begin{pmatrix} \operatorname{Re} s & \operatorname{Im} s \\ -\operatorname{Im} s & \operatorname{Re} s \end{pmatrix} \in SO(2n),$$

and $\theta(s) = J_n s J_n^{-1}$. $M_\theta = SO^*(2n)/U(n)$. Here $SO^*(2n)$ denotes the group of all complex orthogonal matrices of determinant 1 leaving invariant the skew-Hermitian form $z_1\bar{z}_{n+1} - z_{n+1}\bar{z}_1 + z_2\bar{z}_{n+2} - z_{n+2}\bar{z}_2 + \ldots + z_n\bar{z}_{2n} - z_{2n}\bar{z}_n$; this group is isomorphic to the group of all linear transformations leaving invariant a nondegenerate skew-Hermitian form in an n-dimensional vector space over the quaternion field $\mathbf{H}$.

Type CI. $M_u = Sp(n)/U(n)$ $(n \geqslant 1)$, where $U(n)$ is considered as a subgroup of $Sp(n)$ by the identification $U(n) \subset SO(2n)$ explained in type DIII and $\theta(s) = \bar{s} (= J_n s J_n^{-1})$. $M_\theta = Sp(n, \mathbf{R})/U(n)$, where $Sp(n, \mathbf{R})$ is the real symplectic group of degree $2n$.

Type CII. $M_u = Sp(p+q)/Sp(p) \times Sp(q)$ $(p \geqslant q \geqslant 1)$, where $Sp(p) \times Sp(q)$ is identified with a subgroup of $Sp(p+q)$ by the mapping

$$\left(\begin{pmatrix} A_1 & B_1 \\ C_1 & D_1 \end{pmatrix}, \begin{pmatrix} A_2 & B_2 \\ C_2 & D_2 \end{pmatrix}\right) \to \begin{pmatrix} A_1 & 0 & B_1 & 0 \\ 0 & A_2 & 0 & B_2 \\ C_1 & 0 & D_1 & 0 \\ 0 & C_2 & 0 & D_2 \end{pmatrix}$$

and $\theta(s) = K_{p,q} s K_{p,q}$. $M_\theta = Sp(p,q)/Sp(p) \times Sp(q)$. Here $Sp(p,q)$ is the group of complex symplectic matrices of degree $2(p+q)$ leaving invariant the Hermitian form $(z_1, \ldots, z_{p+q}) K_{p,q} {}^t(\bar{z}_1, \ldots, \bar{z}_{p+q})$; this group is interpreted as the group of all linear transformations leaving invariant a nondegenerate Hermitian form of index p in a $(p+q)$-dimensional vector space over the quaternion field $\mathbf{H}$. For $q = 1$, M_u is the quaternion projective space, and M_θ is called the **quaternion hyperbolic space**.

Among the spaces introduced here, there are some with lower p, q, n that coincide (as Riemannian spaces) (→ Appendix A, Table 5.III).

H. Space Forms

A Riemannian manifold of †constant curvature is called a **space form**; it is said to be **spherical**,

Euclidean, or **hyperbolic** according as the constant curvature K is positive, zero, or negative. A space form is a locally symmetric Riemannian space; a simply connected complete space form is a sphere if $K>0$, a real Euclidean space if $K=0$, and a real hyperbolic space if $K<0$. More generally, a complete spherical space form of even dimension is a sphere or a projective space, and one of odd dimension is an orientable manifold. A complete 2-dimensional Euclidean space form is one of the following spaces: Euclidean plane, cylinder, torus, †Möbius strip, †Klein bottle. Except for these five spaces and the 2-dimensional sphere, any †closed surface is a 2-dimensional hyperbolic space form (for details about space forms → [6]).

I. Examples of Irreducible Symmetric Bounded Domains

Among the irreducible symmetric Riemannian spaces described in Section H, those defined by irreducible symmetric Hermitian spaces are of types AIII, DIII, BDI ($q=2$), and CI. We list the irreducible symmetric bounded domains that are isomorphic to the irreducible Hermitian spaces defining these spaces. Positive definiteness of a matrix will be written $\gg 0$.

Type $\mathbf{I}_{m,m'}$ ($m'\geqslant m\geqslant 1$). The set of all $m\times m'$ complex matrices Z satisfying the condition $I_{m'}-{}^t\bar{Z}Z\gg 0$ is a symmetric bounded domain in $\mathbf{C}^{mm'}$, which is isomorphic (as a complex manifold) to the irreducible symmetric Hermitian space defined by M_θ of type AIII ($p=m$, $q=m'$).

Type $\mathbf{II}_m$ ($m\geqslant 2$). The set of all $m\times m$ complex †skew-symmetric matrices Z satisfying the condition $I_m-{}^t\bar{Z}Z\gg 0$ is a symmetric bounded domain in $\mathbf{C}^{m(m-1)/2}$ corresponding to the type DIII ($n=m$).

Type $\mathbf{III}_m$ ($m\geqslant 1$). The set of all $m\times m$ complex symmetric matrices satisfying the condition $I_m-{}^t\bar{Z}Z\gg 0$ is a symmetric bounded domain in $\mathbf{C}^{m(m+1)/2}$ corresponding to the type CI ($n=m$). This bounded domain is holomorphically isomorphic to the †Siegel upper half-space of degree m.

Type $\mathbf{IV}_m$ ($m\geqslant 1, m\neq 2$). This bounded domain in $\mathbf{C}^m$ is formed by the elements $(z_1,\dots,z_m)$ satisfying the condition $|z_1|^2+\dots+|z_m|^2<(1+|z_1^2+\dots+z_m^2|)/2<1$, and corresponds to the type BDI ($p=m, q=2$).

Among these four types of bounded domains, the following complex analytic isomorphisms hold: $\mathrm{I}_{1,1}\cong\mathrm{II}_2\cong\mathrm{III}_1\cong\mathrm{IV}_1$, $\mathrm{II}_3\cong\mathrm{I}_{1,3}$, $\mathrm{IV}_3\cong\mathrm{III}_2$, $\mathrm{IV}_4\cong\mathrm{I}_{2,2}$, $\mathrm{IV}_6\cong\mathrm{II}_4$. (For details about these symmetric bounded domains → [2].) There are two more kinds of irreducible symmetric bounded domains, which are represented as homogeneous spaces of exceptional Lie groups.

J. Weakly Symmetric Riemannian Spaces

A generalization of symmetric Riemannian space is the notion of weakly symmetric Riemannian space introduced by Selberg. Let M be a Riemannian space. M is called a **weakly symmetric Riemannian space** if a Lie subgroup G of the group of isometries $I(M)$ acts transitively on M and there exists an element $\mu\in I(M)$ satisfying the relations (i) $\mu G\mu^{-1}=G$; (ii) $\mu^2\in G$; and (iii) for any two points x, y of M, there exists an element m of G such that $\mu x=my$, $\mu y=mx$. A symmetric Riemannian space M becomes a weakly symmetric Riemannian space if we put $G=I(M)$ and $\mu=$ the identity transformation; as the element m in condition (iii) we can take the symmetry σ_p at the midpoint p on the geodesic arc joining x and y. There are, however, weakly symmetric Riemannian spaces that do not have the structure of a symmetric Riemannian space. An example of such a space is given by $M=G=SL(2,\mathbf{R})$ with a suitable Riemannian metric, where μ is the inner automorphism defined by

$$\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}$$

(Selberg [4]). On a weakly symmetric Riemannian space, the ring of all G-invariant differential-integral operators is commutative; this fact is useful in the theory of spherical functions (→ 437 Unitary Representations).

References

[1] S. Helgason, Differential geometry, Lie groups, and symmetric spaces, Academic Press, 1978.
[2] C. L. Siegel, Analytic functions of several complex variables, Princeton Univ. Press, 1950.
[3] J.-L. Koszul, Exposés sur les espaces homogènes symétriques, São Paulo, 1959.
[4] A. Selberg, Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with application to Dirichlet series, J. Indian Math. Soc., 20 (1956), 47–87.
[5] E. Cartan, Sur certaines formes riemanniennes remarquables des géométries à groupe fondamental simple, Ann. Sci. Ecole Norm. Sup., 44 (1927), 345–467 (Oeuvres complètes, Gauthier-Villars, 1952, pt. I, vol. 2, 867–989).
[6] E. Cartan, Sur les domaines bornés homogènes de l'espace de n-variables complexes, Abh. Math. Sem. Univ. Hamburg, 11 (1936), 116–162 (Oeuvres complètes, Gauthier-Villars, 1952, pt. I, vol. 2, 1259–1305).

[7] J. A. Wolf, Spaces of constant curvature, McGraw-Hill, 1967.
[8] S. Kobayashi and K. Nomizu, Foundations of differential geometry II, Interscience, 1969.
[9] O. Loos, Symmetric spaces. I, General theory; II, Compact spaces and classification, Benjamin, 1969.

413 (VII.7) Symmetric Spaces

A †Riemannian manifold M is called a **symmetric Riemannian space** if M is connected and if for each $p \in M$ there exists an involutive †isometry σ_p of M that has p as an isolated fixed point. For the classification and the group-theoretic properties of symmetric Riemannian spaces → 412 Symmetric Riemannian Spaces and Real Forms. We state here the geometrical properties of a symmetric Riemannian space M. Let M be represented by G/K, a †symmetric Riemannian homogeneous space. The †Lie algebras of G and K are denoted by $\mathfrak{g}$ and $\mathfrak{k}$ respectively. Let us denote by τ_a the †left translation of M defined by $a \in G$, and by X^* the vector field on M generated by $X \in \mathfrak{g}$. We denote by θ the differential of the involutive automorphism of G defining G/K and identify the subspace $\mathfrak{m} = \{X \in \mathfrak{g} \mid \theta(X) = -X\}$ of $\mathfrak{g}$ with the tangent space $T_o(M)$ of M at the origin $o = K$ of M. The †representation of $\mathfrak{k}$ on $\mathfrak{m}$ induced from the †adjoint representation of $\mathfrak{g}$ is denoted by $\mathrm{ad}_{\mathfrak{m}}$.

A. Riemannian Connections

M is a complete real analytic †homogeneous Riemannian manifold. If M is a †symmetric Hermitian space, it is a †homogeneous Kählerian manifold. The †Riemannian connection ∇ of M is the †canonical connection of the homogeneous space G/K and satisfies $\nabla_Y X^* = [X, Y]$ $(Y \in \mathfrak{m})$ for each $X \in \mathfrak{k}$ and $\nabla_Y X^* = 0$ $(Y \in \mathfrak{m})$ for each $X \in \mathfrak{m}$. For each $X \in \mathfrak{m}$, the curve γ_X of M defined by $\gamma_X(t) = (\exp tX)o$ $(t \in \mathbf{R})$ is a †geodesic of M such that $\gamma_X(0) = o$ and $\dot{\gamma}_X(0) = X$. In particular, the †exponential mapping Exp_o at o is given by $\mathrm{Exp}_o X = (\exp X)o$ $(X \in \mathfrak{m})$. For each $X \in \mathfrak{m}$, the †parallel translation along the geodesic arc $\gamma_X(t)$ $(0 \leqslant t \leqslant t_0)$ coincides with the differential of $\tau_{\exp t_0 X}$. If M is compact, for each $p \in M$ there exists a smooth simply closed geodesic passing through p. Any G-invariant tensor field on M is †parallel with respect to ∇. Any G-invariant †differential form on M is closed. The Lie algebra $\mathfrak{h}$ of the †restricted homogeneous holonomy group of M at o coincides with $\mathrm{ad}_{\mathfrak{m}}[\mathfrak{m}, \mathfrak{m}]$. If the group $I(M)$ of all isometries of M is †semisimple, one has $\mathfrak{h} = \{A \in \mathfrak{gl}(\mathfrak{m}) \mid A \cdot g_o = 0, A \cdot R_o = 0\} = \mathrm{ad}_{\mathfrak{m}}\mathfrak{k}$. Here, g_o and R_o denote the values at o of the Riemannian metric g and the †Riemannian curvature R of M, respectively, and $A \cdot$ is the natural action of A on the tensors over $\mathfrak{m}$. If, moreover, M is a symmetric Hermitian space, the value J_0 at o of the †almost complex structure J of M belongs to the center of $\mathfrak{h}$. In general, $\mathfrak{h} = \{0\}$ if and only if M is †flat, and $\mathfrak{h}$ has no nonzero invariant on $\mathfrak{m}$ if and only if $I(M)$ is semisimple.

B. Riemannian Curvature Tensors

The Riemannian curvature tensor R of M is parallel and satisfies $R_0(X, Y) = -\mathrm{ad}_{\mathfrak{m}}[X, Y]$ $(X, Y \in \mathfrak{m})$. Assume that $\dim M \geqslant 2$ in the following. Let P be a 2-dimensional subspace of $\mathfrak{m}$, and $\{X, Y\}$ an orthonormal basis of P with respect to g_o. Then the †sectional curvature $K(P)$ of P is given by $K(P) = g_o([[X, Y], X], Y)$. $K = 0$ everywhere if and only if M is flat. If M is of †compact type (resp. of †noncompact type), then $K \geqslant 0$ (resp. $K \leqslant 0$) everywhere. $K > 0$ (resp. $K < 0$) everywhere if and only if the †rank of M is 1 and M is of compact type (resp. of noncompact type). For any four points p, q, p', q' of a manifold M of any of these types satisfying $d(p,q) = d(p', q')$, d being the †Riemannian distance of M, there exists a $\phi \in I(M)$ such that $\phi(p) = p'$ and $\phi(q) = q'$. Other than the aforementioned M's, the only Riemannian manifolds having this property are circles and Euclidean spaces. If $K > 0$ everywhere, any geodesic of M is a smooth simply closed curve and all geodesics are of the same length. For a symmetric Hermitian space M, the †holomorphic sectional curvature H satisfies $H = 0$ (resp. $H > 0$, $H < 0$) everywhere if and only if M is flat (resp. of compact type, of noncompact type).

C. Ricci Tensors

The †Ricci tensor S of M is parallel. If $\varphi_{\mathfrak{m}}$ denotes the restriction to $\mathfrak{m} \times \mathfrak{m}$ of the †Killing form φ of $\mathfrak{g}$, the value S_o of S at o satisfies $S_o = -\frac{1}{2}\varphi_{\mathfrak{m}}$. If M is †irreducible, it is an †Einstein space. $S = 0$ (resp. positive definite, negative definite, nondegenerate) everywhere if and only if M is flat (resp. M is of compact type, M is of noncompact type, $I(M)$ is semisimple). If M is a †symmetric bounded domain and g is the †Bergman metric of M, one has $S = -g$.

D. Symmetric Riemannian Spaces of Noncompact Type

Let M be of noncompact type. For each $p \in M$, p is the only fixed point of the †symmetry σ_p, and the exponential mapping at p is a diffeomorphism from $T_p(M)$ to M. In particular, M is diffeomorphic to a Euclidean space. For each pair $p, q \in M$, a geodesic arc joining p and q is unique up to parametrization. For each $p \in M$ there exists neither a †conjugate point nor a †cut point of p. If M is a symmetric Hermitian space, that is, if it is a symmetric bounded domain, then it is a †Stein manifold and holomorphically homeomorphic to a †Siegel domain.

E. Groups of Isometries

The isotropy subgroup at o in $I(M)$ is denoted by $I_o(M)$. Then the smooth mapping $I_o(M) \times \mathfrak{m} \rightarrow I(M)$ defined by the correspondence $\phi \times X \mapsto \phi\tau_{\exp X}$ is surjective, and it is a diffeomorphism if M is of noncompact type. If M is of noncompact type, $I(M)$ is isomorphic to the group $A(\mathfrak{g})$ of all automorphisms of $\mathfrak{g}$ in a natural way, and $I_o(M)$ is isomorphic to the subgroup $A(\mathfrak{g}, \mathfrak{k}) = \{\phi \in A(\mathfrak{g}) \mid \phi(\mathfrak{k}) = \mathfrak{k}\}$ of $A(\mathfrak{g})$, provided that G acts almost effectively on M. Moreover, in this case the center of the identity component $I(M)^0$ of $I(M)$ reduces to the identity, and the isotropy subgroup at a point in $I(M)^0$ is a maximal compact subgroup of $I(M)^0$. If $I(M)$ is semisimple, any element of $I(M)^0$ may be represented as a product of an even number of symmetries of M. In the following, let M be a symmetric Hermitian space, and denote by $A(M)$ (resp. $H(M)$) the group of all holomorphic isometries (resp. all holomorphic homeomorphisms) of M, and by $A(M)^0$ and $H(M)^0$ their identity components. All these groups act transitively on M. If M is compact or if $I(M)$ is semisimple, one has $A(M)^0 = I(M)^0$. If $I(M)$ is semisimple, M is simply connected and the center of $I(M)^0$ reduces to the identity. If M is of compact type, M is a †rational †projective algebraic manifold, and $H(M)^0$ is a complex semisimple Lie group whose center reduces to the identity, and it is the †complexification of $I(M)^0$. In this case, the isotropy subgroup at a point in $H(M)^0$ is a †parabolic subgroup of $H(M)^0$. If M is of noncompact type, one has $H(M)^0 = I(M)^0$. In the following we assume that G is compact.

F. Cartan Subalgebras

A maximal Abelian †Lie subalgebra in $\mathfrak{m}$ is called a **Cartan subalgebra** for M. Cartan subalgebras are conjugate to each other under the †adjoint action of K. Fix a Cartan subalgebra $\mathfrak{a}$ and introduce an inner product $(\ ,\ )$ on $\mathfrak{a}$ by the restriction to $\mathfrak{a} \times \mathfrak{a}$ of g_o. For an element α of the dual space $\mathfrak{a}^*$ of $\mathfrak{a}$, we put $\mathfrak{m}_\alpha = \{X \in \mathfrak{m} \mid [H, [H, X]] = -\alpha(H)^2 X$ for any $H \in \mathfrak{a}\}$. The subset $\Sigma = \{\alpha \in \mathfrak{a}^* - \{0\} \mid \mathfrak{m}_\alpha \neq \{0\}\}$ of $\mathfrak{a}^*$ is called the **root system** of M (relative to $\mathfrak{a}$). We write $m_\alpha = \dim \mathfrak{m}_\alpha$ for $\alpha \in \Sigma$. The subset $D = \{H \in \mathfrak{a} \mid \alpha(H) \in \pi\mathbf{Z}$ for some $\alpha \in \Sigma\}$ of $\mathfrak{a}$ is called the **diagram** of M. A connected component of $\mathfrak{a} - D$ is called a **fundamental cell** of M. The quotient group W of the normalizer of $\mathfrak{a}$ in K modulo the centralizer of $\mathfrak{a}$ in K is called the **Weyl group** of M. W is identified with a finite group of orthogonal transformations of $\mathfrak{a}$.

G. Conjugate Points

For a geodesic arc γ with the initial point o, any †Jacobi field along γ that vanishes at o and the end point of γ is obtained as the restriction to γ of the vector field X^* generated by an element $X \in \mathfrak{k}$. For $H \in \mathfrak{a} - \{0\}$, $\mathrm{Exp}_o H$ is a conjugate point to o along the geodesic γ_H if and only if $\alpha(H) \in \pi\mathbf{Z} - \{0\}$ for some $\alpha \in \Sigma$. In this case, the multiplicity of the conjugate point $\mathrm{Exp}_o H$ is equal to $\frac{1}{2}\sum_{\alpha \in \Sigma, \alpha(H) \in \pi\mathbf{Z} - \{0\}} M_\alpha$. From this fact and Morse theory (→ 279 Morse Theory), we get a †cellular decomposition of the †loop space of M. The set of all points conjugate to o coincides with $K\,\mathrm{Exp}_o D$ and is stratified to a disjoint union of a finite number of connected regular submanifolds with dimension $\leqslant \dim M - 2$.

H. Cut Points

We define a †lattice group Γ of $\mathfrak{a}$ by $\Gamma = \{A \in \mathfrak{a} \mid \mathrm{Exp}_o A = o\}$, and put $C_\mathfrak{a} = \{H \in \mathfrak{a} \mid \mathrm{Max}_{A \in \Gamma - \{0\}} 2(H, A)/(A, A) = 1\}$. Then, for $H \in \mathfrak{a} - \{0\}$, $\mathrm{Exp}_o H$ is a cut point of o along the geodesic γ_H if and only if $H \in C_\mathfrak{a}$. The set C_o of all cut points of o coincides with $K\,\mathrm{Exp}_o C_\mathfrak{a}$ and is stratified to a disjoint union of a finite number of connected regular submanifolds with dimension $\leqslant \dim M - 1$. The set of all points †first conjugate to o coincides with C_o if and only if M is simply connected.

I. Fundamental Groups

Let Γ_0 denote the subgroup of $\mathfrak{a}$ generated by $\{(2\pi/(\alpha, \alpha))\alpha \mid \alpha \in \Sigma\}$, identifying $\mathfrak{a}^*$ with $\mathfrak{a}$ by means of the inner product $(\ ,\ )$ of $\mathfrak{a}$. This is a subgroup of Γ. We regard Γ as a subgroup of the group $I(\mathfrak{a})$ of all motions of $\mathfrak{a}$ by parallel

translations. The subgroup $\tilde{W}=W\Gamma$ of $I(\mathfrak{a})$ generated by Γ and the Weyl group W is called the **affine Weyl group** of M. $\tilde{W}$ leaves the diagram D invariant and acts transitively on the set of all fundamental cells of M. Take a fundamental cell σ such that its closure $\bar{\sigma}$ contains 0, and put $\tilde{W}_\sigma=\{w\in\tilde{W}\mid w(\sigma)=\sigma\}$. Then the fundamental group $\pi_1(M)$ of M is an †Abelian group isomorphic to the groups $\tilde{W}_\sigma$ and Γ/Γ_0. $\pi_1(M)$ is a finite group if and only if M is of compact type. In this case, the order of $\pi_1(M)$ is equal to the cardinality of the set $\Gamma\cap\bar{\sigma}$ as well as to the index $[\Gamma:\Gamma_0]$. Moreover, if we denote by $\tilde{W}_\sigma^*$ the group $\tilde{W}_\sigma$ for the symmetric Riemannian space $M^*=G^*/K^*$ defined by the †adjoint group G^* of G and $K^*=\{a\in G^*\mid a\theta=\theta a\}$, then $\tilde{W}_\sigma$ is isomorphic to a subgroup of $\tilde{W}_\sigma^*$. If M is irreducible, $\tilde{W}_\sigma^*$ is isomorphic to a subgroup of the group of all automorphisms of the †extended Dynkin diagram of the root system Σ.

J. Cohomology Rings

Let $P(\mathfrak{g})$ (resp. $P(\mathfrak{k})$) be the †graded linear space of all †primitive elements in the †cohomology algebra $H(\mathfrak{g})$ of $\mathfrak{g}$ (resp. $H(\mathfrak{k})$ of $\mathfrak{k}$), and $P(\mathfrak{g},\mathfrak{k})$ the intersection of $P(\mathfrak{g})$ with the image of the natural homomorphism $H(\mathfrak{g},\mathfrak{k})\to H(\mathfrak{g})$, where $H(\mathfrak{g},\mathfrak{k})$ denotes the relative cohomology algebra for the pair $(\mathfrak{g},\mathfrak{k})$. Then one has $\dim P(\mathfrak{g},\mathfrak{k})+\dim P(\mathfrak{k})=\dim P(\mathfrak{g})$. Denote by $\Lambda P(\mathfrak{g},\mathfrak{k})$ the exterior algebra over $P(\mathfrak{g},\mathfrak{k})$. The †graded algebra of all G-invariant polynomials on $\mathfrak{g}$ (resp. all K-invariant polynomials on $\mathfrak{k}$) is denoted by $I(G)$ (resp. $I(K)$), where the degree of a homogeneous polynomial with degree p is defined to be $2p$. We denote by $I^+(G)$ the ideal of $I(G)$ consisting of all $f\in I(G)$ such that $f(0)=0$, and regard $I(K)$ as an $I^+(G)$-module through the restriction homomorphism. Then the †real cohomology ring $H(M)$ of M is isomorphic to the tensor product $\Lambda P(\mathfrak{g},\mathfrak{k})\otimes(I(K)/I^+(G)I(K))$. If K is connected and the †Poincaré polynomials of $P(\mathfrak{g})$, $P(\mathfrak{k})$, and $P(\mathfrak{g},\mathfrak{k})$ are $\sum_{i=1}^r t^{2m_i-1}$, $\sum_{i=1}^s t^{2n_i-1}$, and $\sum_{i=s+1}^r t^{2m_i-1}$, respectively, then the Poincaré polynomial of $H(M)$ is given by $\prod_{i=s+1}^r(1+t^{2m_i-1})\prod_{i=1}^s(1-t^{2m_i})\prod_{i=1}^s(1-t^{2n_i})^{-1}$.

References

[1] E. Cartan, Sur certaines formes riemannienes remarquables des géométries à groupe fondamental simple, Ann. Sci. Ecole Norm. Sup., 44 (1927), 345–467.
[2] S. Helgason, Differential geometry, Lie groups, and symmetric spaces, Academic Press, 1978.
[3] S. Kobayashi and K. Nomizu, Foundations of differential geometry II, Interscience, 1969.
[4] H. C. Wang, Two point homogeneous spaces, Ann. Math., (2) 55 (1952), 177–191.
[5] A. Korányi and J. A. Wolf, Realization of Hermitian symmetric spaces as generalized half-planes, Ann. Math., (2) 81 (1965), 265–288.
[6] R. Bott and H. Samelson, Applications of the theory of Morse to symmetric spaces, Amer. J. Math., 80 (1958), 964–1029.
[7] T. Sakai, On cut loci of compact symmetric spaces, Hokkaido Math. J., 6 (1977), 136–161.
[8] M. Takeuchi, On conjugate loci and cut loci of compact symmetric spaces I, Tsukuba J. Math., 2 (1978), 35–68.
[9] R. Crittenden, Minimum and conjugate points in symmetric spaces, Canad. J. Math., 14 (1962), 320–328.
[10] J. L. Koszul, Sur un type d'algèbre différentielles avec la transgression, Colloque de Topologie (Espaces fibrès), Brussels, 1950, 73–81.

414 (XX.1) Systems of Units

A. International System of Units

Units representing various physical quantities can be derived from a certain number of **fundamental (base) units**. By a **system of units** we mean a system of fundamental units. Various systems of units have been used in the course of the development of physics. Today, the standard is set by the **international system of units** (système international d'unités; abbreviated SI) [1], which has been developed in the spirit of the meter-kilogram system. This system consists of the seven fundamental units listed in Table 1, units induced from them, and unit designations with prefixes representing the powers of 10 where necessary. It also contains two **auxiliary units** for plane and solid angles, and a large number of derived units [1].

B. Systems of Units in Mechanics

Units in mechanics are usually derived from length, mass, and time, and SI uses the meter, kilogram, and second as base units. Neither the CGS system, derived from centimeter, gram, and second, nor the **system of gravitational units**, derived from length, force, and time, are recommended for general use by

Table 1

Quantity	SI unit	Symbol	Description
Length	meter	m	The meter is the length equal to 1,650,763.73 wavelengths in vacuum of the radiation corresponding to the transmission between the levels $2p^{10}$ and $5d^5$ of the krypton-86 atom.
Mass	kilogram	kg	The kilogram is equal to the mass of the international prototype of the kilogram.
Time	second	s	The second is the duration of 9,192,631,770 periods of the radiation corresponding to the transmission between the two hyperfine levels of the ground state of the cesium-133 atom.
Intensity of electric current	ampere	A	The ampere is the intensity of the constant current maintained in two parallel, rectilinear conductors of infinite length and of negligible circular section, placed 1 m apart in vacuum, and producing a force between them equal to 2×10^{-7} newton ($m \cdot kg \cdot s^{-2}$) per meter of length.
Temperature	kelvin	K	The kelvin, the unit of thermodynamical temperature, is 1/273.16 of the thermodynamical temperature of the triple point of water.
Amount of substance	mole	mol	The mole is the amount of substance of a system containing as many elementary entities as there are atoms in 0.012 kg of carbon-12.
Luminous intensity	candela	cd	The candela is the luminous intensity in a given direction of a source emitting monochromatic radiation of frequency 540×10^{12} hertz ($= s^{-1}$), the radiant intensity of which in that direction is 1/683 watt per steradian. (This revised definition of candela was adopted in 1980.)

Table 2

Quantity	SI unit	Symbol	Unit in terms of SI base or derived units
Frequency	hertz	Hz	$1\ Hz = 1\ s^{-1}$
Force	newton	N	$1\ N = 1\ kg \cdot m/s^2$
Pressure and stress	pascal	Pa	$1\ Pa = 1\ N/m^2$
Work, energy, quantity of heat	joule	J	$1\ J = 1\ N \cdot m$
Power	watt	W	$1\ W = 1\ J/s$
Quantity of electricity	coulomb	C	$1\ C = 1\ A \cdot s$
Electromotive force, potential difference	volt	V	$1\ V = 1\ W/A$
Electric capacitance	farad	F	$1\ F = 1\ C/V$
Electric resistance	ohm	Ω	$1\ \Omega = 1\ V/A$
Electric conductance	siemens	S	$1\ S = 1\ \Omega^{-1}$
Flux of magnetic induction magnetic flux	weber	Wb	$1\ Wb = 1\ V \cdot s$
Magnetic induction, magnetic flux density	tesla	T	$1\ T = 1\ Wb/m^2$
Inductance	henry	H	$1\ H = 1\ Wb/A$
Luminous flux	lumen	lm	$1\ lm = 1\ cd \cdot sr$
Illuminance	lux	lx	$1\ lx = 1\ lm/m^2$
Activity	becquerel	Bq	$1\ Bq = 1\ s^{-1}$
Adsorbed dose	gray	Gy	$1\ Gy = 1\ J/kg$
Radiation dose	sievert	Sv	$1\ Sv = 1\ J/kg$

the SI Committee. Besides the base units, minute, hour, and day, degree, minute, and second (angle), liter, and ton have been approved by the SI Committee. Units such as the electron volt, atomic mass unit, astronomical unit, and parsec (not SI) are empirically defined and have been approved. Several other units, such as nautical mile, knot, are (area), and bar, have been provisionally approved.

C. System of Units in Thermodynamics

The base unit for temperature is the degree Kelvin (°K; formerly called the absolute temperature). Degree Celsius (°C), defined by $t = T - 273.15$, where T is in °K, is also used. The unit of heat is the joule J, the same as the unit for other forms of energy. Formerly, one calorie was defined as the quantity of heat that must be supplied to one gram of water to raise its temperature from 14.5°C to 15.5°C; now one calorie is defined by 1 cal = 4.1855 J.

D. Systems of Units in Electricity and Magnetism

Three distinct systems of units have been developed in the field of electricity and magnetism: the electrostatic system, which originates from Coulomb's law for the force between two electric charges and defines magnetic quantities by means of the Biot-Savart law; the electromagnetic system, which originates from Coulomb's law for magnetism; and the Gaussian system, in which the dielectric constant and permeability are taken to be nondimensional. At present, however, the rationalized MKSA system of units is adopted as the international standard. It uses the **derived units** listed in Table 2 (taken from [2]), where the derived units with proper names in other fields are also listed.

E. Other Units

In the field of photometry, the following definition was adopted in 1948: One candela (cd) ($\fallingdotseq 0.98$ old candle) is defined as $1/(6 \times 10^5)$ of the luminous intensity in the direction normal to a plane surface of 1 m^2 area of a black body at the temperature of the solidifying point of platinum. The total luminous flux emanating uniformly in all directions from a source of luminous intensity 1 cd is defined as 4π lumen (lm). One lux (lx) is defined as the illuminance on a surface area of 1 m^2 produced by a luminous flux of 1 cd uniformly incident on the surface. In 1980, the definition was revised as shown in Table 1.

For theoretical purposes, a system of units called the absolute system of units is often used, in which units of mass, length, and time are chosen so that the values of universal constants, such as the universal gravitational constant, speed of light, Planck's constant, and Boltzmann's constant, are equal to 1.

References

[1] Bureau International des Points et Mesures, Le système international d'unités, 1970, fourth revised edition, 1981.
[2] R. G. Lerner and G. L. Trigg (eds.), Encyclopedia of physics, Addison-Wesley, 1981.

T

415 (XXI.41) Takagi, Teiji

Teiji Takagi (April 21, 1875–February 28, 1960) was born in Gifu Prefecture, Japan. After graduation from the Imperial University of Tokyo in 1897, he continued his studies in Germany, first with Frobenius in Berlin and then with Hilbert in Göttingen. He returned to Japan in 1901 and taught at the Imperial University of Tokyo until 1936, when he retured. He died in Tokyo of cerebral apoplexy.

Since his student years he had been interested in Kronecker's conjecture on †Abelian extensions of imaginary quadratic number fields. He solved it affirmatively for the case of $\mathbf{Q}(\sqrt{-1})$ while still in Göttingen and presented this result as his doctoral thesis. During World War I, he pursued his research in the theory of numbers in isolation from Western countries. It developed into †class field theory, a beautiful general theory of Abelian extensions of algebraic number fields. This was published in 1920, and was complemented by his 1922 paper on the †reciprocity law of power residues and then by †Artin's general law of reciprocity published in 1927. Besides these arithmetical works, he also published papers on algebraic and analytic subjects and on the foundations of the theories of natural numbers and of real numbers. His book (in Japanese) on the history of mathematics in the 19th century and his *General course of analysis* (also in Japanese) as well as his teaching and research activities at the University exercised great influence on the development of mathematics in Japan.

Reference

[1] S. Kuroda (ed.), The collected papers of Teiji Takagi, Iwanami, 1973.

416 (XI.16) Teichmüller Spaces

Consider the set $\mathbf{M}_g$ consisting of the conformal equivalence classes of closed Riemann surfaces of genus g. In 1859 Riemann stated, without rigorous proof, that $\mathbf{M}_g$ is parametrized by $m(g)$ ($=0$ if $g=0$, $=1$ if $g=1$, $=3g-3$ if $g\geqslant 2$) complex parameters (→ 11 Algebraic Functions). Later, the introduction of a topology and $m(g)$-dimensional complex structure on $\mathbf{M}_g$ were discussed rigorously in various ways. The following explanation of these methods is due to O. Teichmüller [1, 2], L. V. Ahlfors [3, 4], and L. Bers [5–7]. For the algebraic-geometric approach → 9 Algebraic Curves.

The trivial case $g=0$ is excluded, since $\mathbf{M}_0$ consists of a single point. Take a closed Riemann surface $\mathfrak{R}_0$ of genus $g\geqslant 1$, and consider the pairs $(\mathfrak{R}, H)$ consisting of closed Riemann surfaces $\mathfrak{R}$ of the same genus g and the †homotopy classes H of orientation-preserving homeomorphisms of $\mathfrak{R}_0$ into $\mathfrak{R}$. Two pairs $(\mathfrak{R}, H)$ and $(\mathfrak{R}', H')$ are defined to be conformally equivalent if the homotopy class $H'H^{-1}$ contains a conformal mapping. The set $\mathbf{T}_g$ consisting of the conformal equivalence classes $\langle\mathfrak{R}, H\rangle$ is called the **Teichmüller space** (with center at $\mathfrak{R}_0$). Let $\mathfrak{H}_g$ be the group of homotopy classes of orientation-preserving homeomorphisms of $\mathfrak{R}_0$ onto itself. $\mathfrak{H}_g$ is a transformation group acting on $\mathbf{T}_g$ in the sense that each $\eta\in\mathfrak{H}_g$ induces the transformation $\langle\mathfrak{R}, H\rangle\rightarrow\langle\mathfrak{R}, H\eta\rangle$. It satisfies $\mathbf{T}_g/\mathfrak{H}_g=\mathbf{M}_g$. The set $\mathfrak{J}_g$ of elements of $\mathfrak{H}_g$ fixing every point of $\mathbf{T}_g$ consists only of the unity element if $g\geqslant 3$ and is a normal subgroup of order 2 if $g=1, 2$. For the remainder of this article we assume that $g\geqslant 2$. The case $g=1$ can be discussed similarly, and the result coincides with the classical one: $\mathbf{T}_1$ can be identified with the upper half-plane and $\mathfrak{H}_1/\mathfrak{J}_1$ is the †modular group.

Denote by $B(\mathfrak{R}_0)$ the set of measurable invariant forms $\mu\,d\bar{z}\,dz^{-1}$ with $\|\mu\|_\infty<1$. For every $\mu\in B(\mathfrak{R}_0)$ there exists a pair $(\mathfrak{R}, H)$ for which some $h\in H$ satisfies $h_{\bar{z}}=\mu h_z$ (→ 352 Quasiconformal Mappings). This correspondence determines a surjection $\mu\in B(\mathfrak{R}_0)\mapsto\langle\mathfrak{R}, H\rangle\in\mathbf{T}_g$. Next, if $Q(\mathfrak{R}_0)$ denotes the space of holomorphic quadratic differentials $\varphi\,dz^2$ on $\mathfrak{R}_0$, a mapping $\mu\in B(\mathfrak{R}_0)\mapsto\varphi\in Q(\mathfrak{R}_0)$ is obtained as follows: Consider μ on the universal covering space U ($=$upper half-plane) of $\mathfrak{R}_0$. Extend it to U^* ($=$lower half-plane) by setting $\mu=0$, and let f be a quasiconformal mapping f of the plane onto itself satisfying $f_{\bar{z}}=\mu f_z$. Take the †Schwarzian derivative $\psi=\{f, z\}$ of the holomorphic function f in U^*. The desired φ is given by $\varphi(z)=\overline{\psi(\bar{z})}$ on U. It has been verified that two μ induce the same φ if and only if the same $\langle\mathfrak{R}, H\rangle$ corresponds to μ. Consequently, an injection $\langle\mathfrak{R}, H\rangle\in\mathbf{T}_g\mapsto\varphi\in Q(\mathfrak{R}_0)$ is obtained. Since $Q(\mathfrak{R}_0)=\mathbf{C}^{m(g)}$ by the Riemann-Roch theorem, this injection yields an embedding $\mathbf{T}_g\subset\mathbf{C}^{m(g)}$, where $\mathbf{T}_g$ is shown to be a domain.

As a subdomain of $\mathbf{C}^{m(g)}$, the Teichmüller space is an $m(g)$-dimensional complex analytic manifold. It is topologically equivalent to the unit ball in real $2m(g)$-dimensional space and is a bounded †domain of holomorphy in $\mathbf{C}^{m(g)}$.

Let $\{\alpha_1, \ldots, \alpha_{2g}\}$ be a 1-dimensional homology basis with integral coefficients in $\mathfrak{R}_0$ such that the intersection numbers are $(\alpha_i, \alpha_j)=(\alpha_{g+i}, \alpha_{g+j})=0$, $(\alpha_i, \alpha_{g+j})=\delta_{ij}$, $i, j=1, \ldots, g$.

Given an arbitrary $\langle\mathfrak{R}, H\rangle \in \mathbf{T}_g$, consider the †period matrix Ω of $\mathfrak{R}$ with respect to the homology basis $H\alpha_1, \ldots, H\alpha_{2g}$ and the basis $\omega_1, \ldots, \omega_g$ of †Abelian differentials of the first kind with the property that $\int_{H\alpha_i} \omega_j = \delta_{ij}$. Then Ω is a holomorphic function on $\mathbf{T}_g$. Furthermore, the analytic structure of the Teichmüller space introduced previously is the unique one (with respect to the topology defined above) for which the period matrix is holomorphic.

$\mathfrak{H}_g$ is a properly discontinuous group of analytic transformations, and therefore $\mathbf{M}_g$ is an $m(g)$-dimensional normal †analytic space. $\mathfrak{H}_g$ is known to be the whole group of the holomorphic automorphisms of $\mathbf{T}_g$ (Royden [8]); thus $\mathbf{T}_g$ is not a †symmetric space.

To every point τ of the Teichmüller space, there corresponds a Jordan domain $D(\tau)$ in the complex plane in such a way that the fiber space $\mathbf{F}_g = \{(\tau, z) \mid z \in D(\tau), \tau \in \mathbf{T}_g \subset \mathbf{C}^{m(g)}\}$ has the following properties: $\mathbf{F}_g$ is a bounded domain of holomorphy of $\mathbf{C}^{m(g)+1}$. It carries a properly discontinuous group $\mathfrak{G}_g$ of holomorphic automorphisms, which preserves every fiber $D(\tau)$ and is such that $D(\tau)/\mathfrak{G}_g$ is conformally equivalent to the Riemann surface corresponding to τ. $\mathbf{F}_g$ carries holomorphic functions $F_j(\tau, z)$, $j = 1, \ldots, 5g-5$ such that for every τ the functions $F_j/F_1, j = 2, \ldots, 5g-5$ restricted to $D(\tau)$ generate the meromorphic function field of the Riemann surface $D(\tau)/\mathfrak{G}_g$.

By means of the †extremal quasiconformal mappings, it can be verified that $\mathbf{T}_g$ is a complete metric space. The metric is called the **Teichmüller metric**, and is known to be a Kobayashi metric.

The Teichmüller space also carries a naturally defined Kähler metric, which for $g=1$ coincides with the †Poincaré metric if $\mathbf{T}_1$ is identified with the upper half-plane. The †Ricci curvature, †holomorphic sectional cruvature, and †scalar curvature are all negative (Ahlfors [9]).

By means of the quasiconformal mapping f, which we considered previously in order to construct the correspondence $\mu \mapsto \varphi$, it is possible to regard the Teichmüller space as a space of quasi-Fuchsian groups (→ 234 Kleinian Groups). To the boundary of $\mathbf{T}_g$, it being a bounded domain in $\mathbf{C}^{m(g)}$, there correspond various interesting Kleinian groups, which are called †boundary groups (Bers [10], Maskit [11]).

The definition of Teichmüller spaces can be extended to open Riemann surfaces $\mathfrak{R}_0$ and, further, to those with signatures. A number of propositions stated above are valid to these cases as well. In particular, the Teichmüller space for the case where $\mathfrak{R}_0$ is the unit disk is called the **universal Teichmüller space**. It is a bounded domain of holomorphy in an infinite-dimensional Banach space and is a symmetric space. Every Teichmüller space is a subspace of the universal Teichmüller space.

References

[1] O. Teichmüller, Extremale quasikonforme Abbildungen und quadratische Differentiale, Abh. Preuss. Akad. Wiss., 1939.
[2] O. Teichmüller, Bestimmung der extremalen quasikonformen Abbildung bei geschlossenen orientierten Riemannschen Flächen, Abh. Preuss. Akad. Wiss., 1943.
[3] L. V. Ahlfors, The complex analytic structure of the space of closed Riemann surfaces, Analytic functions, Princeton Univ. Press, 1960, 45–66.
[4] L. V. Ahlfors, Lectures on quasiconformal mappings, Van Nostrand, 1966.
[5] L. Bers, Spaces of Riemann surfaces. Proc. Intern. Congr. Math., Edinburgh, 1958, 349–361.
[6] L. Bers, On moduli of Riemann surfaces, Lectures at Forschungsinstitut für Mathematik, Eidgenössische Technische Hochschule, Zürich, 1964.
[7] L. Bers, Uniformization, moduli, and Kleinian groups, Bull. London Math. Soc., 4 (1972), 257–300.
[8] H. L. Royden, Automorphisms and isometries of Teichmüller spaces, Advances in the Theory of Riemann Surfaces, Princeton Univ. Press, 1971, 369–383.
[9] L. V. Ahlfors, Curvature properties of Teichmüller's space, J. Analyse Math., 9 (1961), 161–176.
[10] L. Bers, On boundaries of Teichmüller spaces and on Kleinian groups I, Ann. Math., (2) 91 (1970), 570–600.
[11] B. Maskit, On boundaries of Teichmüller spaces and on Kleinian groups II, Ann. Math., (2) 91 (1970), 608–638.

417 (VII.5) Tensor Calculus

A. General Remarks

In a †differentiable manifold with an †affine connection (in particular, in a †Riemannian manifold), we can define an important operator on tensor fields, the operator of covariant differentiation. The **tensor calculus** is a differential calculus on a differentiable manifold that deals with various geometric objects and differential operators in terms of covariant differentiation, and it provides an important tool for studying geometry and analysis on a differentiable manifold.

B. Covariant Differential

Let M be an n-dimensional smooth manifold. We denote by $\mathfrak{F}(M)$ the set of all smooth functions on M and by $\mathfrak{X}^r_s(M)$ the set of all smooth tensor fields of type (r,s) on M. $\mathfrak{X}^1_0(M)$ is the set of all smooth vector fields on M, and we denote it simply by $\mathfrak{X}(M)$.

In the following we assume that an affine connection ∇ is given on M. Then we can define the **covariant differential** of tensor fields on M with respect to the connection (→ 80 Connections). We denote the **covariant derivative** of a tensor field K in the direction of a vector field X by $\nabla_X K$ and the covariant differential of K by ∇K. The operator ∇_X maps $\mathfrak{X}^r_s(M)$ into itself and has the following properties:

(1) $\nabla_{X+Y}=\nabla_X+\nabla_Y$, $\nabla_{fX}=f\nabla_X$,
(2) $\nabla_X(K+K')=\nabla_X K+\nabla_X K'$,
(3) $\nabla_X(K\otimes K')=(\nabla_X K)\otimes K'+K\otimes(\nabla_X K')$,
(4) $\nabla_X f=Xf$,
(5) ∇_X commutes with contraction of tensor fields, where K and K' are tensor fields on M, X, $Y\in\mathfrak{X}(M)$ and $f\in\mathfrak{F}(M)$.

The **torsion tensor** T and the **curvature tensor** R of the affine connection ∇ are defined by

$$T(X,Y)=\nabla_X Y-\nabla_Y X-[X,Y],$$

$$R(X,Y)Z=\nabla_X(\nabla_Y Z)-\nabla_Y(\nabla_X Z)-\nabla_{[X,Y]}Z$$

for vector fields X, Y, and Z. The torsion tensor is of type $(1,2)$, and the curvature tensor is of type $(1,3)$. Some authors define $-R$ as the curvature tensor. We here follow the convention used in [1–6], while in [7, 8] the sign of the curvature tensor is opposite. The torsion tensor and the curvature tensor satisfy the identities

$$T(X,Y)=-T(Y,X),\qquad R(X,Y)=-R(Y,X),$$

$$\begin{aligned}&R(X,Y)Z+R(Y,Z)X+R(Z,X)Y\\&=(\nabla_X T)(Y,Z)+(\nabla_Y T)(Z,X)+(\nabla_Z T)(X,Y)\\&\quad+T(T(X,Y),Z)+T(T(Y,Z),X)\\&\quad+T(T(Z,X),Y),\end{aligned}$$

$$\begin{aligned}&(\nabla_X R)(Y,Z)+(\nabla_Y R)(Z,X)+(\nabla_Z R)(X,Y)\\&=R(X,T(Y,Z))+R(Y,T(Z,X))\\&\quad+R(Z,T(X,Y)).\end{aligned}$$

The last two identities are called the **Bianchi identities**.

The operators ∇_X and ∇_Y for two vector fields X and Y are not commutative in general, and they satisfy the following formula, the **Ricci formula**, for a tensor field K:

$$\nabla_X(\nabla_Y K)-\nabla_Y(\nabla_X K)-\nabla_{[X,Y]}K=R(X,Y)\cdot K,$$

where in the right-hand side $R(X,Y)$ is regarded as a derivation of the tensor algebra $\sum_{r,s}\mathfrak{X}^r_s(M)$.

A **moving frame** of M on a neighborhood U is, by definition, an ordered set $(e_1,\ldots,e_n)$ of n vector fields on U such that $e_1(p),\ldots,e_n(p)$ are linearly independent at each point $p\in U$. For a moving frame $(e_1,\ldots,e_n)$ of M on a neighborhood U we define n differential 1-forms $\theta^1,\ldots,\theta^n$ by $\theta^i(e_j)=\delta^i_j$, and we call them the **dual frame** of $(e_1,\ldots,e_n)$. For a tensor field K of type (r,s) on M, we define n^{r+s} functions $K^{i_1\ldots i_r}_{j_1\ldots j_s}$ on U by

$$K^{i_1\ldots i_r}_{j_1\ldots j_s}=K(e_{j_1},\ldots,e_{j_s},\theta^{i_1},\ldots,\theta^{i_r})$$

and call these functions the components of K with respect to the moving frame $(e_1,\ldots,e_n)$.

Since the covariant differentials ∇e_j are tensor fields of type $(1,1)$, n^2 differential 1-forms ω^i_j are defined by

$$\nabla e_j=\omega^i_j\otimes e_i,$$

where in the right-hand side (and throughout the following) we adopt **Einstein's summation convention**: If an index appears twice in a term, once as a superscript and once as a subscript, summation has to be taken on the range of the index. (Some authors write the above equation as $de_j=\omega^i_j e_i$ or $De_j=\omega^i_j e_i$.) We call these 1-forms ω^i_j the **connection forms** of the affine connection with respect to the moving frame $(e_1,\ldots,e_n)$. The torsion forms Θ^i and the curvature forms Ω^i_j are defined by

$$\Theta^i=d\theta^i+\omega^i_j\wedge\theta^j,\qquad\Omega^i_j=d\omega^i_j+\omega^i_k\wedge\omega^k_j.$$

These equations are called the **structure equation** of the affine connection ∇. If we denote the components of the torsion tensor and the curvature tensor with respect to $(e_1,\ldots,e_n)$ by T^i_{jk} and R^i_{jkl} $(=\theta^i(R(e_k,e_l)e_j))$, respectively, then they satisfy the relations

$$\Theta^i=\frac{1}{2}T^i_{jk}\theta^j\wedge\theta^k,\qquad\Omega^i_j=\frac{1}{2}R^i_{jkl}\theta^k\wedge\theta^l.$$

Using these forms, the Bianchi identities are written as

$$d\Theta^i+\omega^i_j\wedge\Theta^j=\Omega^i_j\wedge\theta^j,$$

$$d\Omega^i_j+\omega^i_k\wedge\Omega^k_j-\omega^k_j\wedge\Omega^i_k=0.$$

Let K be a tensor field of type (r,s) on M and $K^{i_1\ldots i_r}_{j_1\ldots j_s}$ be the components of K with respect to $(e_1,\ldots,e_n)$. We define the covariant differential $DK^{i_1\ldots i_r}_{j_1\ldots j_s}$ and the covariant derivative $K^{i_1\ldots i_r}_{j_1\ldots j_s,k}$ by

$$DK^{i_1\ldots i_r}_{j_1\ldots j_s}=K^{i_1\ldots i_r}_{j_1\ldots j_s,k}\theta^k=dK^{i_1\ldots i_r}_{j_1\ldots j_s}+\sum_{v=1}^{r}K^{i_1\ldots a\ldots i_r}_{j_1\ldots j_s}\omega^{i_v}_a$$

$$-\sum_{v=1}^{s}K^{i_1\ldots i_r}_{j_1\ldots a\ldots j_s}\omega^a_{j_v},$$

Then $K^{i_1\ldots i_r}_{j_1\ldots j_s,k}$ are the components of ∇K with respect to the moving frame $(e_1,\ldots,e_n)$. Some authors write $\nabla_k K^{i_1\ldots i_r}_{j_1\ldots j_s}$ instead of $K^{i_1\ldots i_r}_{j_1\ldots j_s}$ [5, 6].

Using components, the Bianchi identities are written as

$$R^h_{ijk}+R^h_{jki}+R^h_{kij}=T^h_{ij,k}+T^h_{jk,i}+T^h_{ki,j}$$
$$+T^h_{ai}T^a_{jk}+T^h_{aj}T^a_{ki}+T^h_{ak}T^a_{ij},$$

$$R^h_{ijk,l}+R^h_{ikl,j}+R^h_{ilj,k}=R^h_{iak}T^a_{jl}+R^h_{iaj}T^a_{lk}+R^h_{ial}T^a_{kj}.$$

The Ricci formula is written as

$$K^{i_1\ldots i_r}_{j_1\ldots j_s,kl}-K^{i_1\ldots i_r}_{j_1\ldots j_s,lk}=\sum_{v=1}^{r}R^{i_v}_{alk}K^{i_1\ldots a\ldots i_r}_{j_1\ldots\ldots\ldots j_s}$$
$$-\sum_{v=1}^{s}R^a_{j_v lk}K^{i_1\ldots\ldots\ldots i_r}_{j_1\ldots a\ldots j_s}$$
$$+T^a_{kl}K^{i_1\ldots i_r}_{j_1\ldots j_s,a}.$$

Let $(x^1,\ldots,x^n)$ be a local coordinate system defined on a neighborhood U of M. Then $(\partial/\partial x^1,\ldots,\partial/\partial x^n)$ is a moving frame of M on U, and we call it the **natural moving frame** associated with the coordinate system $(x^1,\ldots,x^n)$. Components of a tensor field with respect to the natural moving frame $(\partial/\partial x^1,\ldots,\partial/\partial x^n)$ are often called components with respect to the coordinate system $(x^1,\ldots,x^n)$. We define an n^3 function Γ^i_{kj} on U by $\omega^i_j=\Gamma^i_{kj}dx^k$, where ω^i_j are the connection forms for the natural moving frame. Γ^i_{kj} are called the coefficients of the affine connection ∇. The components of the torsion tensor and the curvature tensor with respect to $(x^1,\ldots,x^n)$ are given by

$$T^i_{jk}=\Gamma^i_{jk}-\Gamma^i_{kj},$$

$$R^h_{ijk}=\partial_j\Gamma^h_{ki}-\partial_k\Gamma^h_{ji}+\Gamma^a_{ki}\Gamma^h_{ja}-\Gamma^a_{ji}\Gamma^h_{ka},$$

where $\partial_i=\partial/\partial x^i$.

With respect to the foregoing coordinate system, the components $K^{i_1\ldots i_r}_{j_1\ldots j_s,k}$ of the covariant differential ∇K of a tensor field K of type (r,s) are given by

$$K^{i_1\ldots i_r}_{j_1\ldots j_s,j}=\partial_j K^{i_1\ldots i_r}_{j_1\ldots j_s}+\sum_{v=1}^{r}\Gamma^{i_v}_{ja}K^{i_1\ldots a\ldots i_r}_{j_1\ldots j_s}$$
$$-\sum_{v=1}^{s}\Gamma^a_{jj_v}K^{i_1\ldots i_r}_{j_1\ldots a\ldots j_s}.$$

C. Covariant Differential of Tensorial Forms

A **tensorial p-form** of type (r,s) on a manifold M is an alternating $\mathfrak{F}(M)$-multilinear mapping of $\overbrace{\mathfrak{X}(M)\times\ldots\times\mathfrak{X}(M)}^{p}$ to $\mathfrak{X}^r_s(M)$. A tensorial p-form of type $(0,0)$ is a differential p-form in the usual sense. A tensorial p-form of type $(1,0)$ is often called a **vectorial p-form**.

If an affine connection ∇ is provided on M, we define the covariant differential of tensorial forms. Let α be a tensorial p-form of type (r,s). The covariant differential $D\alpha$ of α is a tensorial $(p+1)$-form of type (r,s) and is defined by

$$(p+1)D\alpha(X_1,\ldots,X_{p+1})$$
$$=\sum_{i=1}^{p+1}(-1)^{i-1}\nabla_{X_i}(\alpha(X_1,\ldots,\hat{X}_i,\ldots,X_{p+1}))$$
$$+\sum_{i<j}(-1)^{i+j}\alpha([X_i,X_j],$$
$$X_1,\ldots,\hat{X}_i,\ldots,\hat{X}_j,\ldots,X_{p+1}),$$

where $\hat{X}_i$ means that X_i is deleted. If α is of type $(0,0)$, $D\alpha$ coincides with the usual exterior differential $d\alpha$.

The simplest example of a tensorial form is the identity mapping of $\mathfrak{X}(M)$, which will be denoted by θ. Some authors write this vectorial form as dp or dx, where p or x expresses an arbitrary point of a manifold. We call θ the **canonical vectorial form** of M. The torsion tensor T can be regarded as a vectorial 2-form, and we have $2D\theta=T$. The curvature tensor R can be regarded as a tensorial 2-form of type $(1,1)$, i.e., $(X,Y)\rightarrow R(X,Y)\in\mathfrak{X}^1_1(M)$, and the Bianchi identities are written as $DT=R\wedge\theta$, $DR=0$, where the exterior product $R\wedge\alpha$ of R and a tensorial p-form α is defined by

$$(p+1)(p+2)(R\wedge\alpha)(X_1,\ldots,X_{p+2})$$
$$=2\sum_{i<j}(-1)^{i+j-1}R(X_i,X_j)\alpha(X_1,\ldots,\hat{X}_i,\ldots,\hat{X}_j,$$
$$\ldots,X_{p+2}).$$

In general, $2D^2\alpha=R\wedge\alpha$ holds for an arbitrary tensorial form α.

Let $(e_1,\ldots,e_n)$ be a moving frame of M on a neighborhood U and $\theta^1,\ldots,\theta^n$ be its dual frames. A tensorial p-form α of type (r,s) is written as

$$\alpha=\alpha^{i_1\ldots i_r}_{j_1\ldots j_s}\otimes e_{i_1}\otimes\ldots\otimes e_{i_r}\otimes\theta^{j_1}\otimes\ldots\otimes\theta^{j_s},$$

on U, where the $\alpha^{i_1\ldots i_r}_{j_1\ldots j_s}$ are the usual differential p-forms on U. We call them the components of α with respect to $(e_1,\ldots,e_n)$. Then the components of $D\alpha$, which we denote by $D\alpha^{i_1\ldots i_r}_{j_1\ldots j_s}$, are given by

$$D\alpha^{i_1\ldots i_r}_{j_1\ldots j_s}=d\alpha^{i_1\ldots i_r}_{j_1\ldots j_s}+\sum_{v=1}^{r}\omega^{i_v}_a\wedge\alpha^{i_1\ldots a\ldots i_r}_{j_1\ldots j_s}$$
$$-\sum_{v=1}^{s}\omega^a_{j_v}\wedge\alpha^{i_1\ldots i_r}_{j_1\ldots a\ldots j_s}.$$

Then we have

$$D^2\alpha^{i_1\ldots i_r}_{j_1\ldots j_s}=\sum_{v=1}^{r}\Omega^{i_v}_a\wedge\alpha^{i_1\ldots a\ldots i_r}_{j_1\ldots j_s}-\sum_{v=1}^{s}\Omega^a_{j_v}\wedge\alpha^{i_1\ldots i_r}_{j_1\ldots a\ldots j_s}.$$

This is an expression of $2D^2\alpha=R\wedge\alpha$ in terms of components. The components of the canonical vectorial form θ are the dual forms $\theta^1,\ldots,\theta^n$ of $(e_1,\ldots,e_n)$, and we have $D\theta^i=\Theta^i$, which means that the components of $D\theta$ are the torsion forms Θ^i.

D. Tensor Fields on a Riemannian Manifold

Let (M, g) be an n-dimensional Riemannian manifold (→ 364 Riemannian Manifolds). The fundamental tensor g defines a one-to-one correspondence between vector fields and differential 1-forms. A differential 1-form α which corresponds to a vector field X is defined by $\alpha(Y)=g(X, Y)$ for any vector field Y. This correspondence is naturally extended to a one-to-one correspondence between $\mathfrak{X}^r_s(M)$ and $\mathfrak{X}^{r'}_{s'}(M)$, where $r+s=r'+s'$. Let $(e_1, \ldots, e_n)$ be a moving frame of M on a neighborhood U and g_{ij} be the components of g with respect to the moving frame. Let (g^{ij}) be the inverse matrix of the matrix (g_{ij}). The g^{ij} are the components of a symmetric contravariant tensor field of order 2. Let X^i be the components of a vector field X and α_i be the components of the differential 1-form α corresponding to X. Then X^i and α_i satisfy the relations $\alpha_i=g_{ij}X^j$ and $X^i=g^{ij}\alpha_j$. If K^h_{ij} are the components of a tensor field K of type (1, 2) (here taken for simplicity), then

$$K_{hij}=K^a_{ij}g_{ah}, \quad K^{hi}_j=K^h_{aj}g^{ai},$$

$$K^{hij}=K^h_{ab}g^{ai}g^{bj}, \ldots,$$

are the components of a tensor field of type (0, 3), (2, 1), (3, 0), ..., respectively, all of which correspond to K. We call this process of obtaining the components of the corresponding tensor fields from the components of a given tensor field **raising the subscripts** and **lowering the superscripts** by means of the fundamental tensor g.

On a Riemannian manifold, we use the †Riemannian connection, unless otherwise stated. The covariant derivative with respect to the Riemannian connection is given by

$$\begin{aligned}2g(\nabla_X Y, Z) &= Xg(Y, X)+Yg(X, Z)-Zg(X, Y)\\ &\quad +g([X, Y], Z)-g([X, Z], Y)\\ &\quad -g(X, [Y, Z])\end{aligned}$$

for vector fields X, Y, and Z. The coefficients of the Riemannian connection with respect to a local coordinate system $(x^1, \ldots, x^n)$ are usually written as $\left\{ {i \atop kj} \right\}$, called the **Christoffel symbols**, which are given by $\left\{ {i \atop kj} \right\}=g^{ia}(\partial_k g_{ja}+\partial_j g_{ka}-\partial_a g_{kj})/2$. The curvature tensor R of the Riemannian connection satisfies the identities

$$R(X, Y)Z+R(Y, Z)X+R(Z, X)Y=0,$$

$$(\nabla_X R)(Y, Z)+(\nabla_Y R)(Z, X)+(\nabla_Z R)(X, Y)=0,$$

$$R(X, Y)=-R(Y, X),$$

$$\begin{aligned}g(R(X, Y)Z, W) &= g(R(Z, W)X, Y)\\ &= -g(Z, R(X, Y)W),\end{aligned}$$

$$\begin{aligned}g(R(X, Y)Z, W)+g(R(X, Z)W, Y)&\\ +g(R(X, W)Y, Z)&=0.\end{aligned}$$

In terms of the components, these identities are

$$R^h_{ijk}+R^h_{jki}+R^h_{kij}=0,$$

$$R^h_{ijk,l}+R^h_{ikl,j}+R^h_{ilj,k}=0,$$

$$R^h_{ijk}=-R^h_{ikj}, \; R_{hijk}=R_{jkhi}=-R_{ihjk},$$

$$R_{hijk}+R_{hjki}+R_{hkij}=0,$$

where $R_{hijk}=R^a_{ijk}g_{ah}$.

The †Ricci tensor S of the Riemannian manifold is a tensor field of type (0, 2) defined by

$$S(X, Y)=\text{trace of the mapping } Z \rightarrow R(Z, X)Y$$

for vector fields X and Y. The components S_{ji} of the Ricci tensor are given by $S_{ji}=R^a_{jai}$. The †scalar curvature k of the Riemannian manifold M is a scalar on M defined by $k=g^{ji}S_{ji}$. The Ricci tensor and the scalar curvature satisfy the identities

$$S(X, Y)=S(Y, X) \quad \text{or} \quad S_{ji}=S_{ij},$$

$$S_{ij,k}-S_{ik,j}=R^a_{ikj,a}, \qquad 2g^{jk}S_{ij,k}=\partial_i k.$$

For a moving frame of a Riemannian manifold, it is convenient to use an **orthonormal moving frame**. A moving frame $(e_1, \ldots, e_n)$ is orthonormal if $e_1, \ldots, e_n$ satisfy $g(e_i, e_j)=\delta_{ij}$. Since the components of the fundamental tensor with respect to an orthonormal moving frame are δ_{ij}, raising or lowering the indices does not change the values of the components. Some authors write all the indices as subscripts. Also they write the dual 1-forms, the connection forms, and the curvature forms as θ_i, ω_{ji}, and Ω_{ji}, respectively, instead of θ^i, ω^i_j, and Ω^i_j. With respect to an orthonormal moving frame, the connection forms ω^i_j and the curvature forms Ω^i_j satisfy

$$\omega^i_j+\omega^j_i=0 \quad \text{and} \quad \Omega^i_j+\Omega^j_i=0.$$

On a Riemannian manifold, the divergence of a vector field and the operators d, δ, and Δ on differential forms (→ 194 Harmonic Integrals) can be expressed by using the covariant derivatives with respect to the Riemannian connection.

If X^i are the components of a vector field X with respect to a local coordinate system $(x^1, \ldots, x^n)$, the divergence $\operatorname{div} X$ of X is given by $\operatorname{div} X=X^i{}_{,i}$.

Let α be a differential p-form on M. α is written locally in the form $\alpha=(1/p!)\alpha_{i_1 \ldots i_p}dx^{i_1} \wedge \ldots \wedge dx^{i_p}$, where the coefficients $\alpha_{i_1 \ldots i_p}$ are skew-symmetric in all the indices. We call $\alpha_{i_1 \ldots i_p}$ the components of α with respect to the coordinate system. Since α is regarded as an alternating tensor field of type $(0, p)$, we can define the covariant differential $\nabla\alpha$ of α. Then the components of $d\alpha$, $\delta\alpha$, and $\Delta\alpha$ are

given by

$$(d\alpha)_{i_1\dots i_{p+1}}=\sum_{v=1}^{p+1}(-1)^{v-1}\alpha_{i_1\dots\hat{i}_v\dots i_{p+1},i_v},$$

$$(\delta\alpha)_{i_1\dots i_{p-1}}=-g^{ab}\alpha_{ai_1\dots i_{p-1},b},$$

$$(\Delta\alpha)_{i_1\dots i_p}=-g^{ab}\left[\alpha_{i_1\dots i_p,ab}-\sum_{v=1}^{p}S_{i_va}\alpha_{i_1\dots b\dots i_p}-\sum_{v<w}R^c_{ai_vi_w}\alpha_{i_1\dots b\dots c\dots i_p}\right].$$

For a smooth function f and a differential 1-form β we have

$$\Delta f=-\frac{1}{\sqrt{g}}\partial_i(g^{ij}\sqrt{g}\,\partial_j f),$$

$$(\Delta\beta)_i=-g^{ab}[\beta_{i,ab}-S_{ia}\beta_b],$$

where $g=\det(g_{ij})$.

E. Van der Waerden–Bortolotti Covariant Differential

Let E be a finite dimensional smooth †vector bundle over a smooth manifold M and $\Gamma(E)$ be an $\mathfrak{F}(M)$-module of all smooth sections of E. A connection ∇' in E is a mapping of $\mathfrak{X}(M)\times\Gamma(E)$ to $\Gamma(E)$ such that

(1) $\nabla'_X(\xi+\eta)=\nabla'_X\xi+\nabla'_X\eta$,

(2) $\nabla'_X(f\xi)=Xf\cdot\xi+f\nabla'_X\xi$,

(3) $\nabla'_{X+Y}\xi=\nabla'_X\xi+\nabla'_Y\xi$,

(4) $\nabla'_{fX}\xi=f\nabla'_X\xi$,

for X, $Y\in\mathfrak{X}(M)$, $\xi,\eta\in\Gamma(E)$, and $f\in\mathfrak{F}(M)$. $\nabla'_X\xi$ is called the covariant derivative of ξ in the direction X.

An element K of $\mathfrak{X}^r_s(M)\otimes\Gamma(E)$ is called a **tensor field of type** (r,s) **with values in** E (or simply an E-valued tensor field of type (r,s)). K can be regarded as an $\mathfrak{F}(M)$-linear mapping of $\mathfrak{X}^s_r(M)$ to $\Gamma(E)$ or an $\mathfrak{F}(M)$-multilinear mapping of $\overbrace{\mathfrak{X}(M)\times\dots\times\mathfrak{X}(M)}^{s}$ to $\mathfrak{X}^r_0(M)\otimes\Gamma(E)$. For a given $\xi\in\Gamma(E)$, a mapping $X\to\nabla'_X\xi$ defines a tensor field of type $(0,1)$ with values in E which we call the covariant differential of ξ, denoted by $\nabla'\xi$.

The curvature tensor R' of ∇' is a tensor field of type $(0,2)$ with values in $E^*\otimes E$ (E^* is the dual vector bundle of E), and is defined by

$$R'(X,Y)\xi=\nabla'_X(\nabla'_Y\xi)-\nabla'_Y(\nabla'_X\xi)-\nabla'_{[X,Y]}\xi$$

for any vector fields X and Y and any $\xi\in\Gamma(E)$.

If an affine connection ∇ is given on M, we can define the **van der Waerden–Bortolotti covariant derivative** $\bar{\nabla}_XK$ for ∇ and ∇' of a tensor field K of type (r,s) with values in E. It is defined by

$$(\bar{\nabla}_XK)(S)=\nabla'_X(K(S))-K(\nabla_XS)$$

for any $S\in\mathfrak{X}^s_r(M)$. If we regard $\xi\in\Gamma(E)$ as an E-valued tensor field of type $(0,0)$, we have $\bar{\nabla}_X\xi=\nabla'_X\xi$. The covariant derivative $\bar{\nabla}_XR'$ of the curvature tensor R' of ∇' is a tensor field of type $(0,2)$ with values in $E^*\otimes E$ is defined by

$$(\bar{\nabla}_XR')(Y,Z)\xi=\nabla'_X(R'(Y,Z)\xi)-R'(\nabla_XY,Z)\xi-R'(Y,\nabla_XZ)\xi-R'(Y,Z)\nabla'_X\xi.$$

The Bianchi identity is written as

$$(\bar{\nabla}_XR')(Y,Z)+(\bar{\nabla}_YR')(Z,X)+(\bar{\nabla}_ZR')(X,Y)=R'(X,T(Y,Z))+R'(Y,T(Z,X))+R'(Z,T(X,Y)),$$

where T is the torsion tensor of ∇. The Ricci formula is given by

$$(\bar{\nabla}_X(\bar{\nabla}_YK))(S)-(\bar{\nabla}_Y(\bar{\nabla}_XK))(S)-(\bar{\nabla}_{[X,Y]}K)(S)=R'(X,Y)\cdot K(S)-K(R(X,Y)\cdot S),$$

where R is the curvature tensor of ∇, $K\in\mathfrak{X}^r_s(M)\otimes\Gamma(E)$ and $S\in\mathfrak{X}^s_r(M)$.

In the following we assume that the fiber of E is of finite dimension m. A moving frame of E on a neighborhood U of M is an ordered set $(\xi_1,\dots,\xi_m)$ of local sections $\xi_1,\dots,\xi_m$ on U such that $\xi_1(p),\dots,\xi_m(p)$ are linearly independent at each point p of U. Let $(e_1,\dots,e_n)$ be a moving frame of M on U. Then an E-valued tensor field K of type (r,s) is locally written as

$$K^{i_1\dots i_r\alpha}_{j_1\dots j_s}e_{i_1}\otimes\dots\otimes e_{i_r}\otimes\theta^{j_1}\otimes\dots\otimes\theta^{j_s}\otimes\xi_\alpha,$$

where $\theta^1,\dots,\theta^n$ are the dual 1-forms of $(e_1,\dots,e_n)$. The $n^{r+s}m$ functions $K^{i_1\dots i_r\alpha}_{j_1\dots j_s}$ on U are called the components of K with respect to $(e_1,\dots,e_n)$ and $(\xi_1,\dots,\xi_m)$. We define the connection forms ω'^α_β of the connection ∇' by $\nabla'\xi_\beta=\omega'^\alpha_\beta\otimes\xi_\alpha$. Then the curvature forms Ω'^α_β are defined by

$$\Omega'^\alpha_\beta=d\omega'^\alpha_\beta+\omega'^\alpha_\lambda\wedge\omega'^\lambda_\beta=\tfrac{1}{2}R^\alpha_{\beta ji}\theta^j\wedge\theta^i,$$

where $R^\alpha_{\beta ji}$ are the components of the curvature tensor R', i.e., $R'(e_j,e_i)\xi_\beta=R^\alpha_{\beta ji}\xi_\alpha$.

For a given tensor field K of type (r,s) with values in E, the mapping $X\to\bar{\nabla}_XK$ defines a tensor field $\bar{\nabla}K$ of $(r,s+1)$ with values in E which we call the van der Waerden–Bortollotti covariant differential of K. Then if $K^{i_1\dots i_r\alpha}_{j_1\dots j_s}$ are the components of K with respect to $(e_1,\dots,e_n)$ and $(\xi_1,\dots,\xi_m)$, the components $K^{i_1\dots i_r\alpha}_{j_1\dots j_s,k}$ of $\bar{\nabla}K$ are given by

$$K^{i_1\dots i_r\alpha}_{j_1\dots j_s,k}\theta^k=dK^{i_1\dots i_r\alpha}_{j_1\dots j_s}+\sum_{v=1}^{r}K^{i_1\dots a\dots i_r\alpha}_{j_1\dots j_s}\omega^{i_v}_a-\sum_{v=1}^{s}K^{i_1\dots i_r\alpha}_{j_1\dots a\dots j_s}\omega^a_{j_v}+K^{i_1\dots i_r\beta}_{j_1\dots j_s}\omega'^\alpha_\beta.$$

Let f be a smooth mapping of M into a smooth manifold M'. The differential f_*(or df) can be regarded as a tensor field of type $(0,1)$ with values in $f^*T(M')$. Assume that M (resp. M') has a Riemannian metric g (resp. g'). We denote the Riemannian connection of M by ∇.

From the Riemannian connection of M' a connection ∇' in $f^*T(M')$ can be defined. Let $(y^1, \dots, y^m)$ be a local coordinate system of M' on a neighborhood V and $(x^1, \dots, x^n)$ be a local coordinate system on a neighborhood U of M such that $f(U) \subset V$. Put $\xi_\alpha(p) = (\partial/\partial y^\alpha)(f(p))$ for a point $p \in U$. Then $(\xi_1, \dots, \xi_m)$ is a moving frame of $f^*T(M')$. The components of f_* with respect to $(\partial/\partial x^1, \dots, \partial/\partial x^n)$ and $(\xi_1, \dots, \xi_m)$ are given by $f^\alpha(p) = (\partial y^\alpha/\partial x^i)(p)$. The Laplacian Δf of the mapping f is a tensor field of type $(0,0)$ with values in $f^*T(M')$ and is defined by $(\Delta f)^\alpha = g^{ij} f^\alpha_{i,j}$. If $\Delta f = 0$, the mapping f is called a harmonic mapping (→ 195 Harmonic Mappings).

F. Tensor Fields on a Submanifold

Consider an n-dimensional smooth manifold M immersed in an $(n+m)$-dimensional Riemannian manifold $(\overline{M}, \overline{g})$. If we denote the immersion $M \to \overline{M}$ by f, then $g = f^*\overline{g}$ is a Riemannian metric on M, and we denote its Riemannian connection by ∇. The induced bundle $f^*T(\overline{M})$ splits into the sum of the tangent bundle $T(M)$ of M and the normal bundle $T^\perp(M)$. The Riemannian connection on $\overline{M}$ induces connections in $f^*T(\overline{M})$ and in $T^\perp(M)$ which are denoted by $\overline{\nabla}$ and $\nabla^\perp$, respectively. The van der Waerden–Bortolotti covariant derivative for ∇ and $\nabla^\perp$ is denoted by $\tilde{\nabla}$.

For vector fields X and Y on M, the tangential part of $\overline{\nabla}_X Y$ (here we regard Y as a section of $f^*T(\overline{M})$) is $\nabla_X Y$, and we denote the normal part of $\overline{\nabla}_X Y$ by $h(X, Y)$. Then h is a symmetric tensor field of type $(0,2)$ with values in $T^\perp(M)$, and we call h the **second fundamental tensor** of the immersion f. For $\xi \in \Gamma(T^\perp(M))$, the tangential part of $\overline{\nabla}_X \xi$ (here ξ is also regarded as a section of $f^*T(\overline{M})$) is denoted by $-A_\xi X$ and the normal part of $\overline{\nabla}_X \xi$ is $\nabla^\perp_X \xi$. Thus we have

$$\overline{\nabla}_X Y = \nabla_X Y + h(X, Y), \quad \overline{\nabla}_X \xi = -A_\xi X + \nabla_X \xi.$$

h and A are related by

$$\overline{g}(h(X, Y), \xi) = g(A_\xi X, Y).$$

We have the following formulas, called the equations of Gauss, Codazzi, and Ricci:

$$\overline{g}(\overline{R}(X, Y)Z, W) = g(R(X, Y)Z, W) + \overline{g}(h(X, Z), h(Y, W)) - \overline{g}(h(X, W), h(Y, X)),$$

$$\overline{g}(\overline{R}(X, Y)Z, \xi) = \overline{g}((\tilde{\nabla}_X h)(Y, Z), \xi) - \overline{g}((\tilde{\nabla}_Y h)(X, Z), \xi),$$

$$\overline{g}(\overline{R}(X, Y)\xi, \eta) = \overline{g}(R^\perp(X, Y)\xi, \eta) + g([A_\xi, A_\eta]X, Y),$$

for $X, Y, Z, W \in X(M)$ and $\xi, \eta \in \Gamma(T^\perp(M))$, where R, $\overline{R}$, and $R^\perp$ are the curvature tensors of ∇, $\overline{\nabla}$, and $\nabla^\perp$, respectively.

For the manifold M immersed in $\overline{M}$, we use a moving frame $(e_1, \dots, e_n, \xi_1, \dots, \xi_m)$ such that $(e_1, \dots, e_n)$ is an orthonormal moving frame of M on a neighborhood U and $(\xi_1, \dots, \xi_m)$ is a moving frame of $T^\perp(M)$ on U with $\overline{g}(\xi_\alpha, \xi_\beta) = \delta_{\alpha\beta}$. Then we can define the connection forms ω^i_j for ∇ and ω^α_β for $\nabla^\perp$. If we extend $(e_1, \dots, e_n, \xi_1, \dots, \xi_m)$ to an orthonormal moving frame $(\overline{e}_1, \dots, \overline{e}_{n+m})$ of $\overline{M}$ such that $\overline{e}_i(p) = e_i(p)$ $(i = 1, \dots, n)$ and $\overline{e}_{n+\alpha}(p) = \xi_\alpha(p)$ $(\alpha = 1, \dots, m)$ for $p \in U$, then the restriction $f^*\overline{\theta}^A$ and $f^*\overline{\omega}^A_B$ of the dual 1-forms and the connection forms of $\overline{M}$ with respect to $(\overline{e}_1, \dots, \overline{e}_{n+m})$ satisfy the relations

$$f^*\overline{\theta}^i = \theta^i, \quad f^*\overline{\theta}^{n+\alpha} = 0, \quad f^*\overline{\omega}^i_j = \omega^i_j,$$

$$f^*\overline{\omega}^{n+\alpha}_{n+\beta} = \omega^\alpha_\beta, \quad f^*\overline{\omega}^{n+\alpha}_i = \sum_j h^\alpha_{ij}\theta^j,$$

where h^α_{ij} are the components of the second fundamental tensor h with respect to $(e_1, \dots, e_n, \xi_1, \dots, \xi_m)$.

The components $h^\alpha_{ij,k}$ of the covariant differential $\tilde{\nabla}h$ of h are defined by

$$h^\alpha_{ij,k}\theta^k = dh^\alpha_{ij} - h^\alpha_{aj}\omega^a_i - h^\alpha_{ia}\omega^a_j + h^\beta_{ij}\omega^\alpha_\beta.$$

In terms of the components, the equations of Gauss, Codazzi, and Ricci are given by

$$\overline{R}_{hijk} = R_{hijk} + \sum_\alpha (h^\alpha_{ij}h^\alpha_{hk} - h^\alpha_{ik}h^\alpha_{hj}),$$

$$\overline{R}^\alpha_{ijk} = h^\alpha_{ik,j} - h^\alpha_{ij,k},$$

$$\overline{R}^\alpha_{\beta jk} = R^\perp{}^\alpha{}_{\beta jk} - \sum_a (h^\alpha_{ja}h^\beta_{ak} - h^\beta_{ja}h^\alpha_{ak}).$$

Let $(x^1, \dots, x^n)$ be a local coordinate system on a neighborhood U of M and $(y^1, \dots, y^{n+m})$ be a local coordinate system on a neighborhood V of $\overline{M}$ such that $f(U) \subset V$. Regarding the differential f_* of the immersion f as a tensor field of type $(0,1)$ with values in $f^*T(\overline{M})$, we denote the components of f_* with respect to $(x^1, \dots, x^n)$ and $(y^1, \dots, y^{n+m})$ by B^A_i $(i = 1, \dots, n;\ A = 1, \dots, n+m)$. Then we have $B^A_i = \partial y^A/\partial x^i$. We denote by ∇' the van der Waerden–Bortolotti covariant derivative for ∇ and $\overline{\nabla}$. Then the components $B^A_{i,j}$ of $\nabla' f_*$ are given by

$$B^A_{i,j} = \partial_j B^A_i - \left\{{a \atop ji}\right\} B^A_a + B^C_j B^B_i \left\{{A \atop CB}\right\},$$

where $\partial_j = \partial/\partial x^j$, $\left\{{h \atop ji}\right\}$, and $\left\{{A \atop CB}\right\}$ are the Christoffel symbols of the Riemannian metrics g and $\overline{g}$, respectively.

Let $(\xi_1, \dots, \xi_m)$ be an orthonormal moving frame of $T^\perp(M)$ on U and ξ^A_α be the components of ξ_α with respect to $(y^1, \dots, y^{n+m})$. Then we have

$$B^A_{i,j} = h^\alpha_{ij}\xi^A_\alpha,$$

where h_{ij} are the components of the second

fundamental tensor with respect to $(\partial/\partial x^1, \ldots, \partial/\partial x^n)$ and $(\xi_1, \ldots, \xi_m)$.

A tensor field K with values in $T^{\perp}(M)$ can be regarded as a tensor field with values in $f^*T(M)$, and $\tilde{\nabla}K$ is the normal component of $\nabla' K$. For example, if we regard the second fundamental tensor h as a tensor field with values in $f^*T(\overline{M})$, the components of h with respect to the coordinates $(x^1, \ldots, x^n)$ and $(y^1, \ldots, y^{n+m})$ are equal to $B^A_{i,j}$, and we have

$h^{\alpha}_{ij,k} = B^A_{i,jk} \xi^B_{\alpha} \bar{g}_{AB}$.

References

[1] E. Cartan, Leçon sur la géometrie des espaces de Riemann, Gauthier-Villars, second edition, 1946; English translation, Geometry of Riemannian Spaces, Math-Sci Press, 1983.
[2] B.-Y. Chen, Geometry of submanifolds, Dekker, 1973.
[3] S. Helgason, Differential geometry, Lie groups, and symmetric spaces, Academic Press, 1978.
[4] S. Kobayashi and K. Nomizu, Foundations of differential geometry, Interscience, I, 1963; II, 1969.
[5] J. A. Schouten, Ricci-calculus, Springer, second edition, 1954.
[6] K. Yano, Integral formulas in Riemannian geometry, Dekker, 1970.
[7] G. de Rham, Variétés différentiables, Actualités Sci. Ind., Hermann, second edition, 1960.
[8] L. P. Eisenhart, Riemannian geometry, Princeton Univ. Press, second edition, 1949.
[9] R. L. Bishop and S. I. Goldberg, Tensor analysis on manifolds, Macmillan, 1968.

418 (IX.20) Theory of Singularities

A. Introduction

Let $f_1, f_2, \ldots, f_r$ be †holomorphic functions defined in an open set U of the complex space $\mathbf{C}^n$. Let X be the analytic set $f_1^{-1}(0) \cap \ldots \cap f_r^{-1}(0)$. Let $z_0 \in X$, and let $g_1, \ldots, g_s$ be a system of generators of the ideal $\mathscr{I}(X)_{z_0}$ of the germs of the holomorphic functions which vanish identically on a neighborhood of z_0 in X. z_0 is called a **simple point** of X if the matrix $(\partial g_i/\partial z_j)$ attains its maximal rank, say k, at $z = z_0$. In this case, X is a †complex manifold of dimension $n-k$ near z_0. Otherwise, z_0 is called a **singular point** of X.

B. Resolution of Singularities

Let X be a complex analytic space, and let Y be its singular locus. A **resolution of the singularity** of X is a pair of a complex manifold $\tilde{X}$ and a proper surjective holomorphic mapping $\pi: \tilde{X} \to X$ such that the restriction $\pi|_{\tilde{X} - \pi^{-1}(Y)}$ is biholomorphic and $\tilde{X} - \pi^{-1}(Y)$ is dense in $\tilde{X}$. H. Hironaka proved that there exists a resolution for any X such that $\pi^{-1}(Y)$ is a divisor in $\tilde{X}$ with only †normal crossings [16, 17].

Suppose that a compact connected analytic subset $\tilde{Y}$ of a complex manifold $\tilde{X}$ has a †strongly pseudoconvex neighborhood in $\tilde{X}$. Then the contraction $\tilde{X}/\tilde{Y}$ naturally has a structure of a †normal complex analytic variety such that the projection $\tilde{X} \to \tilde{X}/\tilde{Y}$ is a resolution of $\tilde{X}/\tilde{Y}$ (H. Grauert [14]).

C. Two-Dimensional Singularities

Let X be a normal 2-dimensional analytic space. Then the singular points of X are discrete.

Among the resolutions of X, there exists a unique resolution $\pi: \tilde{X} \to X$ with the following universal property: For any resolution $\pi': \tilde{X}' \to X$, there exists a unique mapping $\rho: \tilde{X}' \to \tilde{X}$ with $\pi' = \pi \circ \rho$. This resolution is called the **minimal resolution**.

Let $\pi: \tilde{X} \to X$ be a resolution of a singular point x of X, and let A_i $(i = 1, \ldots, m)$ be the irreducible components of $\pi^{-1}(x)$. The matrix $(A_i \cdot A_j)$ of the †intersection numbers is known to be negative definite (P. Du Val [12]).

The resolution $\pi: \tilde{X} \to X$ is called **good** if (i) each A_i is nonsingular, (ii) $A_i \cap A_j$ $(i \neq j)$ is at most one point and the intersection is transverse and (iii) no three A_i's meet at a point. For a given good resolution $\pi: \tilde{X} \to X$, we associate a diagram in which the vertices v_i $(i = 1, \ldots, m)$ correspond to A_i $(i = 1, \ldots, m)$ and v_i and v_j are joined by a segment if and only if $A_i \cap A_j \neq \varnothing$.

The **geometric genus** $p_g(X, x)$ of a singular point $x \in X$ is the dimension of the †stalk at x of the first direct image sheaf $R^1\pi_* \mathcal{O}_{\tilde{X}}$, where $\pi: \tilde{X} \to X$ is a resolution of $x \in X$ and $\mathcal{O}_{\tilde{X}}$ is the †structure sheaf of $\tilde{X}$. The definition is independent of the choice of the resolution, and $p_g(X, x)$ is a finite integer.

Among the positive cycles of the form $Z = \sum_{i=1}^n n_i A_i$ (i.e., $n_i \geqslant 0$) such that $Z \cdot A_i < 0$ for each $i = 1, \ldots, m$, there exists a smallest one Z_0, which is called the **fundamental cycle** [3].

(1) **Rational singularities.** A singular point x of X is called **rational** if $p_g(X, x) = 0$. (The singularity (X, x) is also called rational even when $\dim X \geqslant 3$ if the direct image sheaf $R^i\pi_* \mathcal{O}_{\tilde{X}} = 0$ for $i > 0$.)

For a rational singularity $x \in X$, the †multiplicity of X at x equals $-Z_0^2$ and the local embedding dimension of X at x is $-Z_0^2+1$. Hence a rational singularity with multiplicity 2, which is called a **rational double point**, is a hypersurface singularity. The following weighted homogeneous polynomials (→ Section D) give the complete list of the defining equations up to analytic isomorphism:

$A_n: x^{n+1}+y^2+z^2$,

weights $(1/(n+1),\ 1/2,\ 1/2)$, $n \geqslant 1$;

$D_n: x^{n-1}+xy^2+z^2$,

weights $(1/(n-1),\ (n-2)/2(n-1),\ 1/2)$, $n \geqslant 4$;

$E_6: x^4+y^3+z^2$,

weights $(1/4,\ 1/3,\ 1/2)$;

$E_7: x^3y+y^3+z^2$,

weights $(2/9,\ 1/3,\ 1/2)$;

$E_8: x^5+y^3+z^2$,

weights $(1/5,\ 1/3,\ 1/2)$,

where the labels appearing at the left are given according to the coincidence of the diagram of the respective minimal resolutions and the †Dynkin diagrams. Rational double points have many different characterizations [11].

The generic part of the singular locus of the unipotent variety of a †complex simple Lie group G (= the orbit of the subregular †unipotent elements in G) is locally expressed as the product of a rational double point and a polydisk. The †universal deformation of a rational double point and its †simultaneous resolution are constructed by restricting the following diagram on a transverse slice to the subregular unipotent orbit (Brieskorn [7]; [34]):

$$\begin{array}{ccc} Y & \longrightarrow & G \\ \downarrow & & \downarrow \\ T & \longrightarrow & T/W \end{array}$$

where T is a †Cartan subgroup of G with the action of the Weyl group W, $G \to T/W$ is the quotient mapping by the †adjoint action of G and $Y = \{(x,B) \mid x \in G$ and B is a †Borel subgroup of G with $x \in B\}$, and other morphisms are defined naturally so that the diagram commutes. Here, $Y \to T$ is the simultaneous resolution of the morphism $G \to T/W$.

(2) **Quotient singularities.** A singular point $x \in X$ is called a **quotient singularity** if there exists a neighborhood of x which is analytically isomorphic to an orbit space U/G, where U is a neighborhood of 0 in $\mathbf{C}^2$ and G is a finite group of analytic automorphisms of U with the unique fixed point 0. The quotient singularities are rational, and their resolutions have been well studied [6]. U/G has a rational double point at 0 if and only if G is conjugate to a nontrivial finite subgroup of $SU(2)$.

(3) **Elliptic singularities.** The singularity (X,x) is called **minimally elliptic** if $p_g(X,x)=1$ and (X,x) is Gorenstein [23]. The following are examples of minimally elliptic singularities.

A singular point $x \in X$ is called **simply elliptic** if the exceptional set A of the minimal resolution is a smooth †elliptic curve [33]. When $A^2 = -1, -2, -3$, (X,x) is a hypersurface singularity given by the following weighted homogeneous polynomials:

$\tilde{E}_6: x^3+y^3+z^3+axyz$,

weights $(1/3,\ 1/3,\ 1/3)$, $A^2=-3$;

$\tilde{E}_7: x^4+y^4+z^2+axyz$,

weights $(1/4,\ 1/4,\ 1/2)$, $A^2=-2$;

$\tilde{E}_8: x^6+y^3+z^2+axyz$,

weights $(1/6,\ 1/3,\ 1/2)$, $A^2=-1$,

(4) **Cusp singularities.** A singular point $x \in X$ is called a **cusp singularity** if the exceptional set of the minimal resolution is either a single rational curve with a †node or a cycle of smooth rational curves. Cusp singularities appear as the boundary of †Hilbert modular surfaces [18]. The hypersurface cusp singularities are given by the polynomials

$T_{p,q,r}: x^p+y^q+z^r+axyz$,

where $1/p+1/q+1/r<1$ and $a \neq 0$.

D. The Milnor Fibration for Hypersurface Singularities

Let V be an analytic set in $\mathbf{C}^N$, and take a point $z_0 \in V$. Let $S_\varepsilon = S(z_0, \varepsilon)$ be a $(2N-1)$-dimensional sphere in $\mathbf{C}^N$ with center z_0 and radius $\varepsilon > 0$, and let $K_\varepsilon = V \cap S_\varepsilon$. If ε is sufficiently small, the topological type of the pair $(S_\varepsilon, K_\varepsilon)$ is independent of ε [27]. By virtue of this fact, the study of singular points constitutes an important aspect of the application of topology to the theory of functions of several complex variables.

A singular point z_0 of V is said to be **isolated** if, for some open neighborhood W of z_0 in $\mathbf{C}^N$, $W \cap V - \{z_0\}$ is a smooth submanifold of $W - \{z_0\}$. In that case, K_ε is a closed smooth submanifold of S_ε, and the diffeomorphism type of $(S_\varepsilon, K_\varepsilon)$ is independent of (sufficiently small) $\varepsilon > 0$. So far, the topological study of such singular points has been primarily focused on isolated singularities. When V is a plane curve, that is, $N=2$ and $r=1$, all the singular points of V are isolated, and the submanifold K_ε of the 3-sphere S_ε can be described as an iterated torus link, where type numbers are

completely determined by the †Puiseaux expansion of the defining equation f of V at the point z_0 [5]. In 1961, D. Mumford, using a resolution argument, showed that if an algebraic surface V is †normal at z_0 and if the closed 3-manifold K_ε is simply connected, then K_ε is diffeomorphic to the 3-sphere and z_0 is nonsingular [29]. The following theorem in the higher-dimensional case is due to E. Brieskorn [8] (1966):

Every †homotopy $(2n-1)$-sphere $(n\neq 2)$ that is a boundary of a †π-manifold is diffeomorphic to the K_ε of some complex hypersurface defined by an equation of the form $f(z)=z_1^{a_1}+\ldots+z_{n+1}^{a_{n+1}}=0$ at the origin in $\mathbf{C}^{n+1}$, provided that $n\neq 2$. The hypersurface of this type is called the **Brieskorn variety**. Inspired by Brieskorn's method, J. W. Milnor developed topological techniques for the study of hypersurface singularities and obtained results such as the **Milnor fibering theorem**, which can be briefly stated as follows:

Suppose that V is defined by a single equation $f(z)=0$ in the neighborhood of $z_0\in\mathbf{C}^{n+1}$. Then there is an associated smooth †fiber bundle $\varphi: S_\varepsilon - K_\varepsilon \to S^1$, where $\varphi(z)=f(z)/|f(z)|$ for $z\in S_\varepsilon-K_\varepsilon$. The fiber $F=\varphi^{-1}(p)$ $(p\in S^1)$ has the homotopy type of a finite CW-complex of dimension n, and K_ε is $(n-2)$-connected.

Suppose that z_0 is an isolated critical point of f. Then F has the homotopy type of a †bouquet of spheres of dimension n [27]. **The Milnor number** $\mu(f)$ of f is defined by the nth Betti number of F, and it is equal to $\dim_{\mathbf{C}}\mathcal{O}_{\mathbf{C}^{n+1},z_0}/(\partial f/\partial z_1,\ldots,\partial f/\partial z_{n+1})$, where $\mathcal{O}_{\mathbf{C}^{n+1},z_0}$ is the ring of the germs of analytic functions of $n+1$ variables at $z=z_0$. **The Milnor monodromy** h_* is the automorphism of $H_n(F)$ that is induced by the action of the canonical generator of the fundamental group of the base space S^1. The †Lefschetz number of h_* is zero if z^0 is a singular point of V. Let $\Delta(t)$ be the characteristic polynomial of h_*. Then K_ε is a homology sphere if and only if $\Delta(1)=\pm 1$ [27]. It is known that $\Delta(t)$ is a product of †cyclotomic polynomials.

The diffeomorphism class of $(S_\varepsilon, K_\varepsilon)$ is completely determined by the congruence class of the linking matrix $L(e_i,e_j)$ $(1\leqslant i,j\leqslant \mu(f))$, where $e_1,\ldots,e_{\mu(f)}$ is an integral basis of $H_n(F)$ and $L(e_i,e_j)$ is the †linking number [21, 10].

The Milnor fibration is also described in the following way. Let $E(\varepsilon,\delta)$ be the intersection of $f^{-1}(D_\delta^*)$ and $B(\varepsilon)$, the open disk of radius ε and center z_0, where D_δ^* is $\{\eta\in\mathbf{C}\,|\,0<|\eta|<\delta\}$. The restriction of f to $E(\varepsilon,\delta)$ is a †locally trivial fibration over D_δ^* if δ is sufficiently smaller than ε [27].

Let $f(z)$ be an analytic function; suppose that $f(0)=0$ and let $\sum_{p\in\mathbf{N}^{n+1}}a_p z^p$ be the Taylor expansion of f at $z=0$. Let $\Gamma_+(f)$ be the convex hull of the union of $\{p+(\mathbf{R}^+)^{n+1}\}$ for $p\in\mathbf{N}^{n+1}\subset\mathbf{R}^{n+1}$ with $a_p\neq 0$, where $\mathbf{R}^+=\{x\in\mathbf{R}\,|\,x\geqslant 0\}$, and let $\Gamma(f)$ be the union of compact faces of $\Gamma_+(f)$. We call $\Gamma(f)$ the **Newton boundary** of f in the **coordinates** $z_1,\ldots,z_{n+1}$. For a closed face Δ of $\Gamma(f)$ of any dimension, let $f_\Delta(z)=\sum_{p\in\Delta}a_p z^p$. We say that f has a **non-degenerate Newton boundary** if $(\partial f_\Delta/\partial z_1,\ldots,\partial f_\Delta/\partial z_{n+1})$ is a nonzero vector for any $z\in(\mathbf{C}^*)^{n+1}$ and any $\Delta\in\Gamma(f)$. Suppose that f has a non-degenerate Newton boundary and 0 is an isolated critical point of f. Then the Milnor fibration of f is determined by $\Gamma(f)$ and $\mu(f)$, and the characteristic polynomial can be explicitly computed by $\Gamma(f)$ [22, 38].

$f(z)$ is called **weighted homogeneous** if there exist positive rational numbers $r_1,\ldots,r_{n+1}$, which are called **weights**, such that $a_p=0$ if $\sum_{i=1}^{n+1}p_i r_i\neq 1$. An analytic function $f(z)$ with an isolated critical point at 0 is weighted homogeneous in suitable coordinates if and only if f belongs to the ideal $(\partial f/\partial z_1,\ldots,\partial f/\partial z_{n+1})$ (K. Saito [32]). Suppose that $f(z)$ is a weighted homogeneous polynomial with an isolated critical point at 0. Then the Milnor fibration of f is uniquely determined by the weights, and $\mu(f)=\prod_{i=1}^{n+1}\left(\frac{1}{r_i}-1\right)$. The surface $f^{-1}(0)$ for $n=2$ is a rational double point if and only if $\sum_{i=1}^{3}r_i>1$.

E. Unfolding Theory

An **unfolding** of a germ of an analytic function $f(z)$ at 0 is a germ of an analytic function $F(z,t)$, where $t\in\mathbf{C}^m$ (m is finite) such that $F(z,0)=f(z)$. We assume that f has an isolated critical point at 0. Among all the unfoldings of f, there exists a universal one, in a suitable sense, that is unique up to a local analytic isomorphism. It is called the **universal unfolding** of f [36, 37, 26] (→ 51 Catastrophe Theory). Explicitly it can be given by $F(z,t)=f(z)+t_1\varphi_1(z)+\ldots+t_\mu\varphi_\mu(z)$, where $\varphi_i(z)$ $(i=1,\ldots,\mu)$ are holomorphic functions which form a $\mathbf{C}$-basis of the Jacobi ring $\mathcal{O}_{\mathbf{C}^{n+1},0}/(\partial f/\partial z_1,\ldots,\partial f/\partial z_{n+1})$ $(\mu=\mu(f))$.

In the universal unfolding $F(z,t)$ of f, the set of points (z_0,t_0) such that $F(z,t_0)$ has an isolated critical point at z_0 with the Milnor number $\mu(f)$ and $F(z_0,t_0)=0$ forms an analytic set at $(z,t)=0$. **The modulus number** of f is the dimension of this set at 0. This set is sometimes called the **μ-constant stratum**. Let g be a germ of an analytic function. g is said to be **adjacent** to f (denoted by $f\to g$), if there exists a sequence of points $(z(m),t(m))$ in $\mathbf{C}^{n+1}\times\mathbf{C}^\mu$ that converges to the origin such that the term of $F(z,t(m))$ at $z(m)$ is equivalent to g. Adjacency relations are important for the

understanding of the degeneration phenomena of functions. The unfolding theory can be considered in exactly the same way as that for the germ of a real-valued smooth function that is finitely determined [36, 26].

The germs of analytic functions with modulus number 0, 1, and 2 are called **simple**, **unimodular**, and **bimodular**, respectively. They were classified by V. I. Arnold [1] (→ Appendix A, Table 5.V). Simple germs correspond to the equations for the rational double points, and unimodular germs define simply elliptic singularities or cusp singularities. Every unimodular or bimodular germ defines a singularity with $p_g = 1$.

F. Picard-Lefschetz Theory

Let $f(z)$ be a holomorphic function such that $f(0)=0$ and 0 is an isolated critical point with the Milnor number μ. Let $F(z,t)$ be a universal unfolding of f at 0. Let $f: E(\varepsilon,\delta)\to D_\delta^*$ be the Milnor fibration of f by the second description in Section D. There exists a positive number r and a codimension 1 analytic subset Δ (called the **bifurcation set**) of $B'(r)$, the open disk of radius r with the center 0 in the parameter space $\mathbf{C}^\mu$, such that for any $t_0 \in B'(r)-\Delta$, $f_{t_0} = F|_{B(\varepsilon)\times t_0}$ has μ different nondegenerate critical points in $B(\varepsilon)$. Let $p_1, \dots, p_\mu$ be the critical points of f_{t_0}. For each p_i, one can choose local coordinates $(y_1, \dots, y_{n+1})$ so that $f_{t_0}(y) = f_{t_0}(p_i) + y_1^2 + \dots + y_{n+1}^2$. Such an f_{t_0} is called a **Morsification** of f.

Let B_i be a small disk with center p_i in $\mathbf{C}^{n+1}$. Then for any q_i which is near enough to $f_{t_0}(p_i)$, the intersection $f_{t_0}^{-1}(q_i)\cap B_i$ is diffeomorphic to the tangent disk bundle of the sphere S^n. The **vanishing cycle** e_i is the corresponding n-dimensional homology class of $f_{t_0}^{-1}(q_i)\cap B_i$. (We fix q_i.) The self-intersection number of e_i is given by

$$\langle e_i, e_i\rangle = \begin{cases} 2(-1)^{n(n-1)/2}, & n \text{ even}, \\ 0, & n \text{ odd}. \end{cases}$$

For a sufficiently small $t_0 \in B'(r)-\Delta$, one has the following: (i) $|f_{t_0}(p_i)|<\delta$; (ii) the restriction of f_{t_0} to E is a fiber bundle over D', where $D' = \{w\in\mathbf{C} \mid |w|\leqslant\delta$, and $w\neq f_{t_0}(p_i)$ for $i=1,\dots,\mu\}$ and $E = f_{t_0}^{-1}(D')\cap B(\varepsilon)$; (iii) the restriction of the above fibration to $\{w \mid |w|=\delta\}$ is equivalent to the restriction of the Milnor fibration of f to $\{w \mid |w|=\delta\}$. Let w_0 be a fixed point of D', and let $F = f_{t_0}^{-1}(w_0)\cap E$. Then F is diffeomorphic to the Milnor fiber of f. Let l_i be a simple path from w_0 to q_i, and let γ_i be the loop $|w - f_{t_0}(p_i)| = |q_i - f_{t_0}(p_i)|$. We suppose that the union of the l_i is contractible to w_0. By parallel translation of the vanishing cycle e_i along l_i, we consider $e_i \in H_n(F)$. The collection $\{e_i \mid i = 1,\dots,\mu\}$ is an integral basis of $H_n(F)$, which is called a **strongly distinguished basis** (Fig. 1).

Now let h_i be the linear transformation of $H_n(F)$ that is induced by the parallel translation along $l_i\gamma_i l_i^{-1}$. The **Picard-Lefschetz formula** says that

$$h_i(e) = e - (-1)^{n(n-1)/2}\langle e, e_i\rangle\cdot e_i \quad \text{for} \quad e\in H_n(F).$$

Here $\langle\ ,\ \rangle$ is the intersection number in $H_n(F)$. For n even, h_i is a †reflection.

The Milnor monodromy h_* of f is equal to the composition $h_1 \dots h_\mu$ under a suitable ordering of the h_i. The subgroup of the group of linear isomorphisms of $H_n(F)$ generated by $h_1, \dots, h_\mu$ is called the **total monodromy group**.

When f is a simple germ and $n\equiv 2 \bmod 4$, the total monodromy group is isomorphic to the †Weyl group of the corresponding Dynkin diagram. Even-dimensional simple singularities are the only ones for which the monodromy group is finite. These are also characterized as the singularities with definite intersection forms.

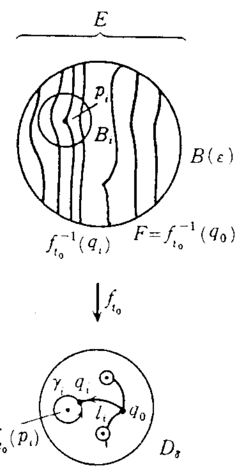


Fig. 1

G. Stratification Theory

The notion of **Whitney stratification** was first introduced by H. Whitney to study the singularities of analytic varieties [39] and was developed by R. Thom for the general case [37].

Let X and Y be submanifolds of the space $\mathbf{R}^n$. We say that the pair (X, Y) satisfies the **Whitney condition (b) at a point** $y\in Y$ if the following holds: Let x_i $(i=1,2,\dots)$ and y_i $(i=1,2,\dots)$ be sequences in X and Y, respectively, that converge to y. Suppose that the tangent space $T_{x_i}X$ converges to a plane T in the corresponding Grassmannian space and the secant $\overline{x_i y_i}$ converges to a line L. Then $L\subset T$. We say that (X, Y) satisfies the **Whitney condition (b)** if it satisfies the Whitney con-

dition (b) at any point $y \in Y$. Let h be a local diffeomorphism of a neighborhood of y. One can see that $(h(X), h(Y))$ satisfies the Whitney condition (b) at $h(y)$ if (X, Y) satisfies it at y. Thus the Whitney condition can be considered for a pair of submanifolds X and Y of a manifold M using a local coordinate system. Let S be a subset of a manifold M, and let $\mathcal{S}$ be a family of submanifolds of M. $\mathcal{S}$ is called a **Whitney prestratification** of S if $\mathcal{S}$ is a locally finite disjoint cover of S satisfying the following: (i) For any $X \in \mathcal{S}$, the frontier $\bar{X} - X$ is a union of $Y \in \mathcal{S}$; (ii) for any pair (X, Y) $(X, Y \in \mathcal{S})$, the Whitney condition (b) is satisfied. A submanifold X in $\mathcal{S}$ is called a **stratum**. There exists a canonical partial order in $\mathcal{S}$ that is defined by $X < Y$ if and only if $X \subset \bar{Y} - Y$.

Let V be an analytic variety, and let $\mathcal{S}$ be an analytic stratification of V that satisfies the frontier condition (i). Then there exists a Whitney prestratification $\mathcal{S}'$ that is finer than $\mathcal{S}$ (Whitney [39]).

For a given Whitney prestratification $\mathcal{S}$, one can construct the following **controlled tubular neighborhood system**: For each $X \in \mathcal{S}$, a †tubular neighborhood $|T_X|$ of X in M and the projection $\pi_X : |T_X| \to X$ and a tubular function $\rho_X : |T_X| \to \mathbf{R}^+$ (= the square of a norm under the identification of $|T_X|$ with the †normal disk bundle of X) are given such that the commutation relations

$$\pi_X \cdot \pi_Y(m) = \pi_X(m), \ \rho_X \pi_Y(m) = \rho_X(m) \qquad \text{for } m \in M, X < Y.$$

are satisfied whenever both sides are defined.

By virtue of this, the notions of vector fields and their integral curves can be defined on a Whitney prestratified set so that several important results on a differentiable manifold can be generalized to the case of stratified sets. For example, the following is **Thom's first isotopy lemma**: Let M and P be differentiable manifolds, and let $(S, \mathcal{S})$ be a Whitney prestratified subset of M. Let $f: S \to P$ be a continuous mapping that is the restriction of a differentiable mapping from M to P. Suppose that the restriction of f to each stratum X of $\mathcal{S}$ is a proper submersion onto P. Then $f: S \to P$ is a fiber bundle [37].

H. b-Functions

Let $f(z)$ be a germ of an analytic function in $\mathbf{C}^{n+1}$ with $f(0) = 0$. The **b-function** of f at 0 is the monic polynomial $b_f(s)$ of lowest degree among all polynomials $b(s)$ with the following property [4, 20]: There exists a differential operator $P(z, \partial/\partial z, s)$, which is a polynomial in s, such that $b(s) f^s(z) = P(z, \partial/\partial z, s) f^{s+1}(z)$. Since $b_f(s)$ is always divisible by $s+1$, we define $\tilde{b}_f(s) = b_f(s)/(s+1)$. All the roots of $\tilde{b}_f(s) = 0$ are negative rational numbers (M. Kashiwara [20]. When f has an isolated critical point at 0, the set $\{\exp(2\pi i\alpha) \mid \alpha$ is a root of $b_f(s) = 0\}$ coincides with the set of eigenvalues of the Milnor monodromy [25].

The name "b-function" is due to M. Sato. He first introduced it in the study of †prehomogeneous vector spaces. Some authors call it the **Bernstein (Bernshteĭn) polynomial**.

I. Hyperplane Sections

Let V be an algebraic variety of complex dimension k in the complex projective space $\mathbf{P}^n$. Let L be a hyperplane that contains the singular points of V. Then the †relative homotopy group $\pi_i(V, V \cap L)$ is zero for $i < k$. Thus the same assertion is true for the †relative homology groups (S. Lefschetz [24]; [28]).

Let f be a holomorphic function defined in the neighborhood of $0 \in \mathbf{C}^{n+1}$ and $f(0) = 0$. Let H be the hypersurface $f^{-1}(0)$. There exists a †Zariski open subset U of the space $(= \mathbf{P}^n)$ of hyperplanes such that for each $L \in U$, there exists a positive number ε such that $\pi_i(B(r) - H, (B(r) - H) \cap L) = 0$ for $i < n$ and $0 < r \leqslant \varepsilon$, where $B(r)$ is a disk of radius r (D. T. Lê and H. Hamm [15]). This implies the following theorem of Zariski: Let V be a hypersurface of $\mathbf{P}^n$, and let $\mathbf{P}^2$ be a general plane in $\mathbf{P}^n$. Then the fundamental group of $\mathbf{P}^n - V$ is isomorphic to the fundamental group of $\mathbf{P}^2 - C$, where $C = V \cap \mathbf{P}^2$. The fundamental group of $\mathbf{P}^2 - C$ is an Abelian group if C is a nodal curve [9, 13].

Suppose that f has an isolated critical point at 0. Let $\mu^{(n+1)}$ be the Milnor number $\mu(f)$. Take a generic hyperplane L. The Milnor number of $f|_L$ is well defined, and we let $\mu^{(n)} = \mu(f|_L)$. Similarly one can define $\mu^{(i)}$ of f and let $\mu^* = (\mu^{(n+1)}, \mu^{(n)}, \ldots, \mu^{(1)})$. Let $f_t(z)$ be a deformation of f. Each f_t has an isolated critical point at 0, and t is a point of a disk D of the complex plane. Let $W = \{(z, t) \mid f_t(z) = 0\}$ and $D' = \{0\} \times D$. $W - D'$ and D' satisfy the Whitney condition (b) if and only if $\mu^*(f_t)$ is invariant under the deformation [35]. The Whitney condition (b) implies topological triviality of the deformation.

References

[1] V. I. Arnol'd, Local normal forms of functions, Inventiones Math., 35 (1976), 87–109.
[2] V. I. Arnol'd, Normal forms for functions near degenerate critical points, the Weyl group of A_k, D_k, E_k and Lagrangian singularities, Functional Anal. Appl., 6 (1972), 254–272. (Original in Russian, 1972.)

[3] M. Artin, On isolated rational singularities of surfaces, Amer. J. Math., 88 (1966), 129–136.
[4] I. N. Bernshteĭn, The analytic continuation of generalized functions with respect to parameter, Functional Anal. Appl., 6 (1972), 273–285. (Original in Russian, 1972.)
[5] K. Brauer, Zur Geometrie der Funktionen zweier komplexen Veränderlichen III, Abh. Math. Sem. Hamburg, 6 (1928), 8–54.
[6] E. Brieskorn, Rationale Singularitäten Komplexer Flächen, Inventiones Math., 4 (1968), 336–358.
[7] E. Brieskorn, Singular elements of semi-simple algebraic groups, Actes Congrès Int. Math., 1970, vol. 2, 279–284.
[8] E. Brieskorn, Beispiele zur Differentialtopologie, Inventiones Math., 2 (1966), 1–14.
[9] P. Deligne, Le groupe fondamental du complément d'une courbe plane n'ayant que des points doubles ordinaires est abelian, Sém. Bourbaki, no. 543, 1979–1980.
[10] H. Durfee, Fibered knots and algebraic singularities, Topology, 13 (1974), 47–59.
[11] A. H. Durfee, Fifteen characterizations of rational double points and simple critical points, Enseignement Math., 25 (1979), 131–163.
[12] P. Du Val, On isolated singularities of surfaces which do not affect the condition of adjunction I–III, Proc. Cambridge Philos. Soc., 30 (1933/34), 453–465, 483–491.
[13] W. Fulton, On the fundamental group of the complement of a node curve, Ann. Math., (2) 111, (1980), 407–409.
[14] H. Grauert, Über Modifikationen und exzeptionelle anaytische Mengen, Math. Ann., 146 (1962), 331–368.
[15] H. Hamm and D. T. Lê, Un théorème de Zariski du type de Lefschetz, Ann. Sci. Ecole Norm. Sup., 3, 1973.
[16] H. Hironaka, Resolution of singularities of an algebraic variety over a field of characteristic zero, Ann. Math., (2) 79 (1964), 109–326.
[17] H. Hironaka, Bimeromorphic smoothing of a complex analytic space, Acta Math. Vietnamica, 2, no. 2 (1977), 106–168.
[18] F. Hirzebruch, Hilbert modular surfaces, Enseignement Math., 19 (1973), 183–281.
[19] P. Holm, Real and complex singularities, Sijthoff and Noordhoff, 1976.
[20] M. Kashiwara, B-functions and holonomic systems, Inventiones Math., 38 (1976), 33–54.
[21] M. Kato, A classification of simple spinnable structure on a 1-connected Alexander manifold, J. Math. Soc. Japan, 26 (1974), 454–463.
[22] A. G. Kouchnirenko, Polyèdre de Newton et nombres de Milnor, Inventiones Math., 32 (1976), 1–31.
[23] H. Laufer, On minimally elliptic singularities, Amer. J. Math., 99 (1977), 1257–1295.
[24] S. Lefschetz, L'analysis situs et la géométrie algébrique, Gauthiers-Villars, 1924.
[25] B. Malgrange, Le polynome de Bernstein d'une singularité isolée, Lecture notes in math. 459, Springer, 1976, 98–119.
[26] J. N. Mather, Stability of C^∞-mappings. I, Ann. Math., (2) 87 (1968), 89–104; II, (2) 89 (1969), 254–291; III, Publ. Math. Inst. HES, 35 (1970), 301–336; IV, Lecture notes in math. 192, Springer, 1971, 207–253.
[27] J. W. Milnor, Singular points of complex hypersurfaces, Ann. math. studies 61, Princeton Univ. Press, 1968.
[28] J. W. Milnor, Morse theory, Ann. math. studies 51, Princeton Univ. Press, 1963.
[29] D. Mumford, The topology of normal singularities of an algebraic surface and a criterion for simplicity, Publ. Math. Inst. HES, 9 (1961), 5–22.
[30] F. Pham, Formules de Picard-Lefschetz généralisées et ramification des intégrales, Bull. Soc. Math. France, 93 (1965), 333–367.
[31] E. Picard and S. Simart, Traité des fonctions algébriques de deux variables I, Gauthier-Villars, 1897.
[32] K. Saito, Quasihomogene isolierte Singularitäten von Hyperflächen, Inventiones Math., 14 (1971), 123–142.
[33] K. Saito, Einfach elliptische Singularitäten, Inventiones Math., 23 (1974), 289–325.
[34] P. Slodowy, Simple singularities and simple algebraic groups, Lecture notes in math. 815, Springer, 1980.
[35] B. Teissier, Cycles evanescents, sections planes et conditions de Whitney, Astérisque, 78 (1973), 285–362.
[36] R. Thom, La stabilité topologique des applications polynomiales, Enseignement Math., 8 (1962), 24–33.
[37] R. Thom, Ensembles et morphismes stratifiés, Bull. Amer. Math. Soc., 75 (1969), 240–284.
[38] A. N. Varchenko, Zeta function of the monodromy and Newton's diagram, Inventiones Math., 37 (1976), 253–262.
[39] H. Whitney, Tangents to an analytic variety, Ann. Math., (2) 81 (1964), 496–549.
[40] O. Zariski, Algebraic surfaces, Erg. math. 61, Springer, 1971.

419 (XX.18) Thermodynamics

A. Basic Concepts and Postulates

Thermodynamics traditionally focuses its attention on a particular class of states of a

given system called (thermal) equilibrium states, although a more recent extension, called the thermodynamics of irreversible processes, deals with certain nonequilibrium states. In a simple system, an **equilibrium state** is completely specified (up to the shape of the volume it occupies) by the volume V (a positive real number), the **mole numbers** $N_1, \ldots, N_r$ (nonnegative reals) of its chemical components, and the **internal energy** U (real). (More variables might be needed if the system were, e.g., inhomogeneous, anisotropic, electrically charged, magnetized, chemically not inert, or acted on by electric, magnetic, or gravitational fields.) This means that any of the quantities associated with equilibrium states (called **thermodynamical quantities**) of a simple system under consideration is a function of V, $N_1, \ldots, N_r$, and U.

When n copies of the same state are put next to each other and the dividing walls are removed, V, $N_1, \ldots, N_r$, and U for the new state will be n times the old values of these variables under the assumptions that each volume is sufficiently large and that the effects of the boundary walls can be neglected. Thermodynamical quantities behaving in this manner are called **extensive**. Those that are invariant under the foregoing procedure are called **intensive**. More precisely, the thermodynamic variables are defined by homogeneity of degree 1 and 0 as functions of V, $N_1, \ldots, N_r$, and U.

By a shift of the position of the boundary (called an **adiabatic wall** if energy and chemical substances do not move through it) or by transport of energy through the boundary (called a **diathermal wall** if this is allowed) or by transport of chemical components through the boundary (called a **permeable membrane**) (in short, by thermodynamical processes), these variables can change their values. If these shifts or transports are not permitted (especially for a composite system consisting of several simple systems, at its boundary with the outside), the system is called **closed**. Otherwise it is called **open**.

Those equilibrium states that do not undergo any change when brought into contact with each other across an immovable and impermeable diathermal wall (called a **thermal contact**) form an equivalence class. This is sometimes called the **0th law of thermodynamics**. The equivalence class, called the **temperature** of states belonging to it, is an intensive quantity.

The force needed to keep a movable wall at rest, divided by the area of the wall, is called the **pressure**. It is another intensive quantity. For a (slow) change of the volume by an amount dV under a constant pressure P, mechanical work of amount $-P\,dV$ is done on the system. Together with a possible change of the internal energy, say of amount dU, the amount

$$\delta Q = dU - P\,dV \tag{1}$$

of energy is somehow gained (if it is positive) or lost (if it is negative) by the system. This amount of energy is actually transported from or to a neighboring system through diathermal walls so that the total energy for a bigger closed composite system is conserved. This is called the **first law of thermodynamics**, and δQ is called the **heat gain** or **loss** by the system.

If two states of different temperatures T_1 and T_2 are brought into thermal contact, energy is transferred from one, say T_1, to the other (called heat transfer). This defines a binary class relation denoted by $T_1 > T_2$. The Clausius formulation of the **second law of thermodynamics** says that it is impossible to make a positive heat transfer from a state of lower temperature to another state of higher temperature without another change elsewhere. By considering a certain composite system, one reaches the conclusion that there exists a labeling of temperatures by positive real numbers T, called the **absolute temperature**, for which the following is an exact differential:

$$\delta Q/T = (dU - P\,dV)/T = dS. \tag{2}$$

The integral S is an extensive quantity, called the **entropy**. Furthermore, the sum of the entropies of component simple systems in an isolated composite system is nondecreasing during any thermodynamic process, and the following **entropy maximum principle** holds: An isolated composite system reaches an equilibrium at those values of extensive parameters that maximize the sum of the entropies of component simple systems (for constant total energy and volume and within the set of allowed states under a given constraint).

A relation expressing the entropy of a given system as a function of the extensive parameters (specifying equilibrium states) is known as the **fundamental relation** of the system. If it is given as a continuous and differentiable homogeneous function of V, $N_1, \ldots, N_r$, and U and is monotone increasing in U for fixed V, $N_1, \ldots, N_r$, then one can develop the thermodynamics of the system based on the above entropy maximum principle. A relation expressing an intensive parameter as a function of some other independent variables is called an **equation of state**.

Another postulate, which is much less frequently used, is the **Nernst postulate** or the **third law of thermodynamics**, which says that the entropy vanishes at the vanishing absolute temperature.

B. Various Coefficients and Relationships

The partial derivative $\partial/\partial x$ of a function $f(x, y, \ldots)$ with respect to the variable x with the variables $y, \ldots$ fixed is denoted by $(\partial f/\partial x)_{y,\ldots}$. We abbreviate $N_1, \ldots, N_r$ as N in the following.

If the fundamental relation is written as $U = U(V, N_1, \ldots, N_r, S)$ (instead of S being represented as a function of the other quantities), then (2) implies

$$(\partial U/\partial S)_{V,N} = T, \quad (\partial U/\partial V)_{N,S} = -P.$$

The other first-order partial derivatives of U are

$$\mu_j = (\partial U/\partial N_j)_{V,N_1,\ldots,\hat{j},\ldots,N_r,S},$$

with μ_j called the **chemical potential** (or electrochemical potential) of the jth component.

If a system is surrounded by an adiabatic wall (i.e., the system is thermally isolated) and goes through a gradual reversible change (**quasistatic adiabatic process**), then the entropy has to stay constant. If a system is in thermal contact through a diathermal wall with a large system (called the **heat bath**) whose temperature is assumed to remain unchanged during the thermal contact, then the temperature of the system itself remains constant (an **isothermal process**). The decrease of the volume per unit increase of pressure under the latter circumstance is called the **isothermal compressibility** and is given by

$$\kappa_T = -V^{-1}(\partial V/\partial P)_{T,N}.$$

Under constant pressure, the increase of the volume per unit increase of the temperature is called the **coefficient of thermal expansion** and is given by

$$\alpha = V^{-1}(\partial V/\partial T)_{P,N}.$$

Under constant pressure, the amount of (quasistatic) heat transfer into the system per mole required to produce a unit increase of temperature is called the **specific heat at constant pressure** and is given by

$$c_P = N^{-1}T(\partial S/\partial T)_{P,N},$$

where $N = N_1 + \ldots + N_r$. The same quantity under constant volume is called the **specific heat at constant volume** and is given by

$$c_v = N^{-1}T(\partial S/\partial T)_{V,N}.$$

The positivity of c_v is equivalent to the convexity of energy as a function of entropy for fixed values of V and N.

Because of the first-order homogeneity of an extensive quantity as a function of other extensive variables, one can derive an **Euler relation**, such as

$$U = TS - PV + \mu_1 N_1 + \ldots + \mu_r N_r,$$

for a simple system. Its differential form implies the following **Gibbs-Duhem relation**:

$$S\,dT - V\,dP + N_1\,d\mu_1 + \ldots + N_r\,d\mu_r = 0.$$

Because of the identity

$$\left(\frac{\partial}{\partial x}\right)\left(\frac{\partial f}{\partial y}\right) = \left(\frac{\partial}{\partial y}\right)\left(\frac{\partial f}{\partial x}\right),$$

there arise relationships among second derivatives, known as the **Maxwell relations**:

$$(\partial T/\partial V)_{S,N} = -(\partial P/\partial S)_{V,N},$$

$$(\partial V/\partial S)_{P,N} = (\partial T/\partial P)_{S,N},$$

$$(\partial S/\partial V)_{T,N} = (\partial P/\partial T)_{V,N},$$

$$(\partial S/\partial P)_{T,N} = -(\partial V/\partial T)_{P,N}.$$

By computing the Jacobian of transformations of variables, further relations can be obtained. For example,

$$c_P = c_v + N^{-1}TV\alpha^2/\kappa_T.$$

C. Legendre Transform and Variational Principles

The **Legendre transform** of a function $f(x_1, \ldots, y_1, \ldots)$ relative to the variables x is given by

$$g(p_1, \ldots, y_1, \ldots) = f - \sum_j x_j p_j$$

as a function of the variables $p_j = \partial f/\partial x_j$ and y. The original variables x can be recovered as $-x_j = \partial g/\partial p_j$.

In terms of Legendre transforms, the entropy maximum principle can be reformulated in various forms:

Energy minimum principle: For given values of the total entropy and volume, the equilibrium is reached at those values of unconstrained parameters that minimize the total energy. This principle is applicable in reversible processes where the total entropy stays constant.

Helmholtz free energy minimum principle: For given values of the temperature (equal to that of a heat bath in thermal contact with the system) and the total volume, the equilibrium is reached at those values of the unconstrained parameters that minimize the total Helmholtz free energy, where the **Helmholtz free energy** for a simple system is defined as a function of $T, V, N_1, \ldots, N_r$ by

$$F = U - TS,$$

$$dF = -S\,dT - P\,dV + \mu_1\,dN_1 + \mu_r\,dN_r.$$

Enthalpy minimum principle: For given values of the pressure and the total entropy,

the equilibrium is reached at those values of unconstrained parameters that minimize the total enthalpy, where the **enthalpy** for a simple system is defined as a function of S, P, $N_1, \ldots, N_r$ by

$$H = U + PV,$$

$$dH = T\,dS + V\,dP + \mu_1\,dN_1 + \ldots + \mu_r\,dN_r.$$

Gibbs free energy minimum principle: For constant temperature and pressure, the equilibrium is reached at those values of unconstrained parameters that minimize the total Gibbs free energy, where the **Gibbs free energy** for a simple system is given as a function of T, P, $N_1, \ldots, N_r$ by

$$G = U - TS + PV,$$

$$dG = -S\,dT + V\,dP + \mu_1\,dN_1 + \ldots + \mu_r\,dN_r.$$

References

[1] M. W. Zemansky, Heat and thermodynamics, McGraw-Hill, 1951.
[2] H. B. Callen, Thermodynamics, Wiley, 1960.
[3] J. W. Gibbs, The collected works of J. Willard Gibbs, vol. 1, Thermodynamics, Yale Univ. Press, 1948.
[4] M. Planck, Treatise on thermodynamics, Dover, 1945. (Original in German, seventh edition, 1922.)
[5] E. A. Guggenheim, Thermodynamics, North-Holland, 1949.
[6] F. C. Andrews, Thermodynamics: Principles and applications, Wiley, 1971.

420 (XX.8) Three-Body Problem

A. n-Body Problem and Classical Integrals

In the **n-body problem**, we study the motions of n particles $P_i(x_i, y_i, z_i)$ $(i = 1, 2, \ldots, n)$ with arbitrary masses $m_i(>0)$ following †Newton's law of motion,

$$m_i \frac{d^2 w_i}{dt^2} = \frac{\partial U}{\partial w_i}, \qquad i = 1, 2, \ldots, n, \tag{1}$$

where w_i is any one of x_i, y_i, or z_i,

$$U = \sum_{i \neq j} k^2 m_i m_j / r_{ij},$$

with k^2 the gravitation constant, and

$$r_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}.$$

Although the one-body and two-body problems have been completely solved, the problem has not been solved for $n > 2$. The **three-body problem** is well known and is important both in celestial mechanics and in mathematics. For $n > 3$ the problem is called the **many-body problem**.

The equations (1) have the so-called ten classical integrals, that is, the **energy integral** $\sum_i (m_i/2)((\dot{x}_i)^2 + (\dot{y}_i)^2 + (\dot{z}_i)^2) - U = \text{constant}$ $(\dot{w} = dw/dt)$, six **integrals of the center of mass** $\sum_i m_i \dot{w}_i = \text{constant}$, $\sum_i m_i w_i = (\sum_i m_i \dot{w}_i)t + \text{constant}$, and three **integrals of angular momentum** $\sum_i m_i(u_i \dot{w}_i - w_i \dot{u}_i) = \text{constant}$ $(u \neq w)$. Using these integrals and eliminating the time t and the ascending node by applying Jacobi's method, the order of the equations (1) can be reduced to $6n - 12$. H. Bruns proved that algebraic integrals cannot be found except for the classical integrals, and H. Poincaré showed that there is no other single-valued integral (Bruns, *Acta Math.*, 11 (1887); Poincaré [2, I, ch. 5]). These results are called **Poincaré-Bruns theorems**. Therefore we cannot hope to obtain general solutions for the equations (1) by †quadrature. General solutions for $n \geqslant 3$ have not been discovered except for certain specific cases.

B. Particular Solutions

Let r_i be the position vector of the particle P_i with respect to the center of mass of the n-body system. A configuration $r \equiv \{r_1, \ldots, r_n\}$ of the system is said to form a **central figure** (or **central configuration**) if the resultant force acting on each particle P_i is proportional to $m_i r_i$, where each proportionality constant is independent of i. The proportionality constant is uniquely determined as $-U/\sum_{i=1}^n m_i r_i^2$ by the configuration of the system. A configuration r is a central figure if and only if r is a †critical point of the mapping $r \mapsto U^2(r)\sum_{i=1}^n m_i r_i^2$ [5, 6]. A rotation of the system, in planar central figure, with appropriate angular velocity is a particular solution of the planar n-body problem.

Particular solutions known for the three-body problem are the **equilateral triangle solution** of Lagrange and the **straight line solution** of Euler. They are the only solutions known for the case of arbitrary masses, and their configuration stays in the central figure throughout the motion.

C. Domain of Existence of Solutions

The solutions for the three-body problem are analytic, except for the collison case, i.e., the case where $\min r_{ij} = 0$, in a strip domain enclosing the real axis of the t-plane (Poincaré, P.

Painlevé). K. F. Sundman proved that when two bodies collide at $t=t_0$, the solution is expressed as a power series in $(t-t_0)^{1/3}$ in a neighborhood of t_0, and the solution which is real on the real axis can be uniquely and analytically continued across $t=t_0$ along the real axis. When all three particles collide, the total angular momentum f with respect to the center of mass must vanish (and the motion is planar) (**Sundman's theorem**); so under the assumption $f\neq 0$, introducing $s=\int^t(U+1)dt$ as a new independent variable and taking it for granted that any binary collision is analytically continued, we see that the solution of the three-body problem is analytic on a strip domain $|\mathrm{Im}\, s|<\delta$ containing the real axis of the s-plane. The conformal mapping

$$\omega=(\exp(\pi s/2\delta)-1)/(\exp(\pi s/2\delta)+1)$$

maps the strip domain onto the unit disk $|\omega|<1$, where the coordinates of the three particles w_κ, their mutual distances r_{kl}, and the time t are all analytic functions of ω and give a complete description of the motion for all real time (Sundman, *Acta Math.*, 36 (1913); Siegel and Moser [7]).

When a triple collision occurs at $t=t_0$, G. Bisconcini, Sundman, H. Block, and C. L. Siegel showed that as $t\to t_0$, (i) the configuration of the three particles approaches asymptotically the Lagrange equilateral triangle configuration or the Euler straight line configuration, (ii) the collision of the three particles takes place in definite directions, and (iii) in general the triple-collision solution cannot be analytically continued beyond $t=t_0$.

D. Final Behavior of Solutions

Suppose that the center of mass of the three-body system is at rest. The motion of the system was classified by J. Chazy into seven types according to the asymptotic behavior when $t\to+\infty$, provided that the angular momentum f of the system is different from zero. In terms of the †order of the three mutual distances r_{ij} (for large t) these types are defined as follows:

(i) H^+: **Hyperbolic motion.** $r_{ij}\sim t$.
(ii) HP^+: **Hyperbolic-parabolic motion.** r_{13}, $r_{23}\sim t$ and $r_{12}\sim t^{2/3}$.
(iii) HE^+: **Hyperbolic-elliptic motion.** r_{13}, $r_{23}\sim t$ and $r_{12}<a$ (a = finite).
(iv) P^+: **Parabolic motion.** $r_{ij}\sim t^{2/3}$.
(v) PE^+: **Parabolic-elliptic motion.** r_{13}, $r_{23}\sim t^{2/3}$ and $r_{12}<a$.
(vi) L^+: **Lagrange-stable motion** or **bounded motion.** $r_{ij}<a$.
(vii) OS^+: **Oscillating motion.** $\overline{\lim}_{t\to\infty}\sup r_{ij}=\infty$, $\underline{\lim}_{t\to\infty}\sup r_{ij}<\infty$.

Define H^-, HE^-, etc. analogously but with $t\to-\infty$. There are three classes for each of the motions HP, HE, and PE, depending on which of the three bodies separates from the other two bodies and recedes to infinity, denoted by HP_i, HE_i, PE_i ($i=1,2,3$), respectively. The energy constant h is positive for H- and HP-motion, zero for P-motion, and negative for PE-, L-, and OS-motion. For HE-motion, h may be positive, zero, or negative.

We say that a **partial capture** takes place when the motion is H^- for $t\to-\infty$ and HE_i^+ for $t\to+\infty$ (for $h>0$), and a **complete capture** when the motion is HE_i^- for $t\to-\infty$ and L^+ for $t\to+\infty$ (for $h<0$). We say also that an **exchange** takes place when HE_i^- for $t\to-\infty$ and HE_j^+ for $t\to+\infty$ $(t\neq j)$. The probability of complete capture in the domain $h<0$ is zero (J. Chazy, G. A. Merman).

E. Perturbation Theories

The radius of convergence in the s-plane for Sundman's solution is too small and the convergence is too slow in the ω-plane to make it possible to compute orbits of celestial bodies, and for that purpose a perturbation method is usually adopted. When the masses $m_2,\dots,m_n$ are negligibly small compared with m_1 for the n-body problem, the motion of the nth body is derived as the solution of the two-body problem for m_1 and m_n by assuming $m_2=\dots=m_{n-1}=0$ as a first approximation, and then the deviations of the true orbit from the ellipse are derived as †perturbations. In the **general theory of perturbations** the deviations are derived theoretically by developing a disturbing function, whereas in the **special theory of perturbations** they are computed by numerical integration. In general perturbation theory, problems concerning convergence of the solution are important, and it becomes necessary to simplify the disturbing function in dealing with the actual relations among celestial bodies. Specific techniques have to be developed in order to compute perturbations for lunar motion, motions of characteristic asteroids, and motions of satellites (e.g., the system of the Sun, Jupiter, and Jovian satellites).

F. The Restricted Three-Body Problem

Since the three-body problem is very difficult to handle mathematically, mathematical interest has been concentrated on the **restricted three-body problem** (in particular, the planar problem) since Hill studied lunar theory in the 19th century. For the restricted three-body problem, the third body, of zero mass, cannot have any influence on the motion of the other

two bodies, which are of finite masses and which move uniformly on a circle around the center of mass. In the planar case, let us choose units so that the total mass, the angular velocity of the two bodies about their center of mass, and the gravitation constant are all equal to 1, and let (q_1, q_2) be the coordinates of the third body with respect to a rotating coordinate system chosen in such a way that the origin is at the center of mass and the two bodies of finite masses μ and $1-\mu$ are always fixed on the q_1-axis. Then the equations of motion for the third body are given by a Hamiltonian system:

$$\frac{dq_i}{dt}=\frac{\partial H}{\partial p_i}, \quad \frac{dp_i}{dt}=-\frac{\partial H}{\partial q_i}, \quad i=1,2, \qquad (2)$$

with

$$H=\frac{1}{2}(p_1^2+p_2^2)+q_2p_1-q_1p_2-U(q_1,q_2),$$

$$U=\frac{1-\mu}{\sqrt{(q_1+\mu)^2+q_2^2}}+\frac{\mu}{\sqrt{(q_1+\mu-1)^2+q_2^2}}.$$

The equations (2) have the energy integral $H(p,q)=$constant, called **Jacobi's integral.** Siegel showed that there is no other algebraic integral, and it can be proved by applying Poincaré's theorem that there is no other single-valued integral. Regularization of the two singular points for the equations (2) and solutions passing through the singular points were studied by T. Levi-Civita, and solutions tending to infinity were studied by B. O. Koopman.

After reducing the number of variables by means of the Jacobi integral, the equations (2) give rise to a flow in a 3-dimensional manifold of which the topological type was clarified by G. D. Birkhoff (*Rend. Circ. Mat. Palermo*, 39 (1915)). Since this flow has an †invariant measure, the equations have been studied topologically, and important results for the restricted three-body problem, particularly on periodic solutions, have been obtained.

G. Stability of Equilateral Triangular Solutions

Suppose that the origin $q_i=p_i=0$ is an †equilibrium point for an autonomous Hamiltonian system with two degrees of freedom:

$$\frac{dq_i}{dt}=\frac{\partial H}{\partial p_i}, \quad \frac{dp_i}{dt}=-\frac{\partial H}{\partial q_i}, \quad i=1,2,$$

with the Hamiltonian H being analytic at the origin. When the †eigenvalues of the corresponding linearized system are purely imaginary and distinct, denoted by $\pm\lambda_1$, $\pm\lambda_2$, and $\lambda_1k_1+\lambda_2k_2\neq0$ for $0<|k_1|+|k_2|\leqslant4$ (where k_i is an integer), we can find suitable coordinates ξ_i, η_i so that the Hamiltonian H takes the form

$$H=\lambda_1\zeta_1+\lambda_2\zeta_2+\frac{1}{2}(c_{11}\zeta_1^2+2c_{12}\zeta_1\zeta_2+c_{22}\zeta_2^2)$$
$$+H_5+\ldots$$

with $\zeta_i\equiv\xi_i\eta_i$ and real c_{ij}. It is necessary that $\eta_i=\sqrt{-1}\,\bar{\xi}_i$ for the solutions to be real. In addition, if the condition

$$D\equiv c_{11}\lambda_2^2-2c_{12}\lambda_1\lambda_2+c_{22}\lambda_1^2\neq0$$

is satisfied, then the origin is a †stable equilibrium point of the original system (V. I. Arnol'd, J. Moser) [7].

For Lagrange equilateral triangular solutions of the planar restricted three-body problem, the eigenvalues λ of the linearized system derived from (2) are given as roots of the equation $\lambda^4+\lambda^2+(27/4)\mu(1-\mu)=0$ and are purely imaginary if $\mu(1-\mu)<1/27$. Applying the Arnol'd-Moser result, A. M. Leontovich and A. Deprit and Bartholomé showed that the Lagrange equilibrium points are stable for μ such that $0<\mu<\mu_0$, where μ_0 is the smaller root of $27\mu(1-\mu)=1$, excluding three values: μ_1, μ_2 at which $\lambda_1k_1+\lambda_2k_2=0$ $|k_1|+|k_2|\leqslant4$ and μ_3 at which $D=0$.

Arnol'd proved that if the masses $m_2, \ldots, m_n$ are negligibly small in comparison with m_1, the motion of the n-body system is †quasi-periodic for the majority of initial conditions for which the eccentricities and inclinations of the osculating ellipses are small.

References

[1] Y. Hagihara, Celestial mechanics. I, Dynamical principles and transformation theory, MIT Press, 1970; II, Perturbation theory, pts. 1, 2, MIT 1972; III, Differential equations in celestial mechanics, pts. 1, 2, Japan Society for the Promotion of Science, 1974; IV, Periodic and quasi-periodic solutions, pts. 1, 2, Japan Society, 1975; V, Topology of the three-body problem, pts. 1, 2, Japan Society, 1976.
[2] H. Poincaré, Les méthodes nouvelles de la mécanique céleste I–III, Gauthier-Villars, 1892–1899.
[3] E. T. Whittaker, A treatise on the analytical dynamics of particles and rigid bodies, Cambridge Univ. Press, fourth edition, 1937.
[4] H. Happel, Das Dreikörperproblem, Koehler, 1941.
[5] A. Wintner, The analytical foundations of celestial mechanics, Princeton Univ. Press, 1947.
[6] R. Abraham and J. E. Marsden, Foundations of mechanics, Benjamin, second edition, 1978.
[7] C. L. Siegel and J. Moser, Lectures on celestial mechanics, Springer, 1971.

[8] V. M. Alekseev (Alexeyev), Sur l'allure finale de mouvement dans le problème des trois corps, Actes Congrès Intern. Math., 1970, vol. 2, 893–907.
[9] V. Szebehely, Theory of orbits, Academic Press, 1967.

421 (XVIII.11) Time Series Analysis

A. Time Series

A time series is a sequence of observations ordered in time. Here we assume that measurements are quantitative and the times of measurements are equally spaced. We consider this sequence to be a realization of a stochastic process X_t (→ 407 Stochastic Processes). Usually time series analysis means a statistical analysis based on samples drawn from a stationary process (→ 395 Stationary Processes) or a related process. In what follows we denote the sample by $\mathbf{X}=(X_1, X_2, \ldots, X_T)'$.

B. Statistical Inference of the Autocorrelation

Let us assume X_t (t an integer) to be real-valued and weakly stationary (→ 395 Stationary Processes) and for simplicity $EX_t=0$ and consider the estimation of the **autocorrelation** $\rho_h=R_h/R_0$ of time lag h, where $R_l=EX_tX_{t+l}$. We denote the **sample autocovariance** of time lag h as

$$\tilde{R}_h=\frac{1}{T-|h|}\sum_{t=1}^{t-|h|}X_tX_{t+|h|},$$

and define the **serial correlation coefficient** of time lag h by $\tilde{\rho}_h=\tilde{R}_h/\tilde{R}_0$. It can be shown that the joint distribution of $\{\sqrt{T}(\tilde{\rho}_h-\rho_h)\mid 1\leqslant h\leqslant H\}$ tends to an H-dimensional †normal (Gaussian) distribution with mean vector $\mathbf{0}$, if one assumes that X_t is expressed as $X_t=\sum_{j=-\infty}^{\infty}b_j\xi_{t-j}$, where $\sum_{j=-\infty}^{\infty}|b_j|<+\infty$, $\sum_{j=-\infty}^{\infty}|j|^{1/2}b_j^2<+\infty$, and the ξ_t are independently and identically distributed random variables with $E\xi_t=0$ and $E\xi_t^4<+\infty$.

When X_t is an autoregressive process of order K (→ Section D) and also a †Gaussian process, it can be shown that the asymptotic distribution of $\{\sqrt{T}(\tilde{\rho}_h-\rho_h)\mid 1\leqslant h\leqslant K\}$ as $T\to\infty$ is equal to the asymptotic distribution of $\{\sqrt{T}(\hat{\rho}_h-\rho_h)\mid 1\leqslant h\leqslant K\}$, where $\hat{\rho}_h$ is the †maximum likelihood estimator of ρ_h. In general, it is difficult to obtain the maximum likelihood estimator of ρ_h. The statistical properties of other estimators of ρ_h, e.g., an estimator constructed by using $\operatorname{sgn}(X_t)$ ($\operatorname{sgn}(y)$ means 1 ($y>0$), 0 ($y=0$), -1 ($y<0$)) have also been investigated.

Testing hypotheses concerning autocorrelation can be carried out by using the above results. Let us now consider the problem of testing the hypothesis that X_t is a †white noise. Assume that X_t is a Gaussian process and that a white noise with $EX_t^2=\sigma^2$ exists, and define $\tilde{C}_h=\sum_{t=1}^{T}(X_t-\bar{X})(X_{t+h}-\bar{X})$ and $\tilde{\gamma}_h=\tilde{C}_h/\tilde{C}_0$ for $h\geqslant 0$, where $X_{T+j}=X_j$ and $\bar{X}=\sum_{t=1}^{T}X_t/T$. Then the probability density function of $\tilde{\gamma}_1$ can be obtained and it can be shown that

$$P(\tilde{\gamma}_1>\gamma)=\sum_{j=1}^{m}(\lambda_j-\gamma)^{(T-3)/2}\frac{1}{\Lambda_j},\qquad \lambda_{m+1}\leqslant\gamma\leqslant\lambda_m,$$

where $\lambda_j=\cos 2\pi j/T$ and

$$\Lambda_j=\prod_{\substack{k=1\\(k\neq j)}}^{(T-1)/2}(\lambda_j-\lambda_k),\qquad T=3,5,\ldots,$$

$$\Lambda_j=\prod_{\substack{k=1\\(k\neq j)}}^{T/2-1}(\lambda_j-\lambda_k)\sqrt{1+\lambda_j},\qquad T=4,6,\ldots,$$

$$1\leqslant m\leqslant(T-3)/2\qquad\text{if } T \text{ is odd},$$

$$1\leqslant m\leqslant T/2-1\qquad\text{if } T \text{ is even}.$$

This can be used to obtain a test of significance.

C. Statistical Inference of the Spectrum

To find the periodicities of a real-valued †weakly stationary process X_t with mean 0, the statistic, called the **periodogram**,

$$I_T(\lambda)=\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}X_te^{-2\pi it\lambda}\right|^2$$

is used. If X_t is expressed as

$$X_t=\sum_{l=1}^{L}\{m_l\cos 2\pi\lambda_lt+m_l'\sin 2\pi\lambda_lt\}+Y_t,$$

where $\{m_l\}$, $\{m_l'\}$, and $\{Y_t\}$ are mutually independent random variables with $Em_l=Em_l'=0$ and $V(m_l)=V(m_l')=\sigma_l^2$ and $\{Y_t\}$ is independent and identically distributed with means 0 and finite variances σ^2, the distribution of $I_T(\lambda)$ converges to a distribution with finite mean and finite variance at $\lambda\neq\pm\lambda_l$ for $1\leqslant l\leqslant L$ when T tends to infinity. On the other hand, the magnitude of $I_T(\lambda)$ is of the order of T at $\lambda=\pm\lambda_l$, $1\leqslant l\leqslant L$. This means that we can find the periodicities of X_t by using $I_T(\lambda)$. When $X_t=Y_t$, we find that the distribution of $2I_T(\lambda)/\sigma^2$ (when $\lambda\neq 0$, $\pm 1/2$) or $I_T(\lambda)/\sigma^2$ (when $\lambda=0$ or $\pm 1/2$) tends to the †χ^2 distribution with degrees of freedom 2 or 1, respectively, and $I(\mu_1), I(\mu_2), \ldots, I(\mu_M)$ are asymptotically independent random variables for $0\leqslant|\mu_1|<|\mu_2|<\ldots<|\mu_M|\leqslant 1/2$ when $T\to\infty$. Applying this result, we can test for periods in the data.

Let $f(\lambda)$ be the spectral density function of a real-valued weakly stationary process X_t. In general, the variance of $|\sum_{t=1}^{T} X_t e^{-2\pi i t\lambda}/\sqrt{T}|$ does not tend to 0 as T tends to infinity; hence $I_T(\lambda)$ cannot be used as a good estimator for the spectral density. To obtain an estimate of $f(\lambda)$, several estimators defined by using weight functions have been proposed by several authors. Let $W_T(\lambda)$ be a weight function defined on $(-\infty, \infty)$, and construct a statistic $\tilde{f}(\lambda)=\int_{-1/2}^{1/2} I_T(\mu)W_T(\lambda-\mu)d\mu$. Let us use $\tilde{f}(\mu)$ for the estimation of $f(\lambda)$. $W_T(\lambda)$ is called a **window**. An important class of $W_T(\lambda)$ is as follows. Let $W(\lambda)$ be continuous, $W(\lambda)=W(-\lambda)$, $W(0)=1$, $|W(\lambda)|<1$, and $\int_{-\infty}^{\infty} W(\lambda)^2 d\lambda < +\infty$, and let H be a positive integer depending on T such that $H\to\infty$ and $H/T\to 0$ as $T\to\infty$. Putting $w_j=W(j/H)$, we define $W_T(\lambda)$ by $W_T(\lambda)=\sum_{j=-T+1}^{T-1} w_j e^{-2\pi i j\lambda}$. Then $\tilde{f}(\lambda)$ can be expressed as $\tilde{f}(\lambda)=\sum_{h=-T+1}^{T-1} \tilde{\tilde{R}}_h w_h e^{-2\pi i h\lambda}$, where $\tilde{\tilde{R}}_h=\sum_{t=1}^{T-h} X_{t+h}X_t/T$ for $h\geqslant 0$ and $\tilde{\tilde{R}}_h=\sum_{t=|h|+1}^{T} X_{t+h}X_t/T$ for $h<0$. Let X_t be stationary to the fourth order (→ 395 Stationary Processes) and satisfy

$$\sum_{h=-\infty}^{\infty} |R_h| < +\infty,$$

$$\sum_{h,l,p=-\infty}^{\infty} |C_{o,h,l,p}| < +\infty,$$

where $C_{o,h,l,p}$ is the fourth-order joint †cumulant of X_t, X_{t+h}, X_{t+l}, and X_{t+p}. Then we have

$$\lim_{T\to\infty}\frac{T}{H}V(\tilde{f}(0))=2f(0)^2\int_{-\infty}^{\infty} W(\lambda)^2 d\lambda,$$

$$\lim_{T\to\infty}\frac{T}{H}V(\tilde{f}(\pm 1/2))=2f(1/2)^2\int_{-\infty}^{\infty} W(\lambda)^2 d\lambda,$$

$$\lim_{T\to\infty}\frac{T}{H}V(\tilde{f}(\lambda))=f(\lambda)^2\int_{-\infty}^{\infty} W(\lambda)^2 d\lambda, \qquad \lambda\neq 0, \quad \pm 1/2,$$

$$\lim_{T\to\infty}\frac{T}{H}\operatorname{Cov}(\tilde{f}(\lambda),\ \tilde{f}(\mu))=0, \quad \lambda\neq\mu. \qquad (1)$$

$\{w_h\}$ or $W_T(\lambda)$ should have an optimality, e.g., to minimize the mean square error of $\tilde{f}(\lambda)$. But, generally, it is difficult to obtain such a $\{w_h\}$ or $W_T(\lambda)$.

Several authors have proposed specific types of windows. The following are some examples: (i) (Bartlett) $w_h=(1-|h|/H)$ for $|h|\leqslant H$ and $w_h=0$ for $|h|>H$; (ii) (Tukey) $w_h=\sum_{l=-\infty}^{\infty} a_l\cos(\pi lh/H)$ for $|h|\leqslant H$ and $w_h=0$ for $|h|>H$, where the a_l are constants such that $\sum_{l=-\infty}^{\infty}|a_l|<+\infty$, $\sum_{l=-\infty}^{\infty} a_l=1$ and $a_l=a_{-l}$. The Hanning and Hamming windows are $a_0=0.50$, $a_1=a_{-1}=0.25$, and $a_l=0$ for $|l|\geqslant 2$ and $a_0=0.54$, $a_1=a_{-1}=0.23$, and $a_l=0$ for $|l|\geqslant 2$, respectively [2]. Let $X_t=\sum_{j=-\infty}^{\infty} b_j\varepsilon_{t-j}$, where $\sum_{j=-\infty}^{\infty}|b_j|<+\infty$ and the ε_t are independently and identically distributed random variables with $E\varepsilon_t=0$ and $E\varepsilon_t^4<+\infty$. Let $\{\lambda_j \mid 1\leqslant j\leqslant M\}$ be arbitrary real numbers such that $0\leqslant\lambda_1<\lambda_2<\ldots<\lambda_M\leqslant 1/2$, where M is an arbitrary positive integer. Then the joint distribution of $\{\sqrt{T/H}(\tilde{f}(\lambda_\nu)-E\tilde{f}(\lambda_\nu)) \mid 1\leqslant\nu\leqslant M\}$ tends to the normal distribution with means 0 and covariance matrix Σ, which is defined by (1). Let us assume, furthermore, that $\lim_{x\to 0}(1-w(x))/|x|^q=C$ and $\sum_{h=-\infty}^{\infty}|h|^p|R_h|<+\infty$, where C, q, and p are some positive constants satisfying the following conditions: (i) when $p\geqslant q$, $H^q/T\to 0$ $(p\geqslant 1)$ and $H^{q+1-p}/T\to 0$ $(p\geqslant 1)$ as $T\to\infty$ and $\lim_{T\to\infty} T/H^{2q+1}$ is finite; (ii) when $p<q$, $H^p/T\to 0$ $(p\geqslant 1)$ and $H/T\to 0$ $(p\leqslant 1)$ as $T\to\infty$ and $\lim_{T\to\infty} T/H^{2p+1}=0$. Then $\sqrt{T/H}(\tilde{f}(\lambda_\nu)-E\tilde{f}(\lambda_\nu))$ in the results above can be replaced by $\sqrt{T/H}(\tilde{f}(\lambda_\nu)-f(\lambda_\nu))$.

Estimation of higher-order spectra, particularly the bispectrum, has also been discussed. Let X_t be a weakly stationary process with mean 0, and let its spectral decomposition be given by $X_t=\int_{-1/2}^{1/2} e^{2\pi i t\lambda}dZ(\lambda)$ (→ 395 Stationary Processes). We assume that X_t is a weakly stationary process of degree 3 and put $R_{h_1,h_2}=EX_tX_{t+h_1}X_{t+h_2}$ for any integers h_1 and h_2. Then we have

$$R_{h_1,h_2}=\int_{1/2}^{1/2}\int_{-1/2}^{1/2} e^{2\pi i(h_1\lambda_1+h_2\lambda_2)}dF(\lambda_1,\lambda_2).$$

Symbolically, $dF(\lambda_1,\lambda_2)=EdZ(\lambda_1)dZ(\lambda_2)dZ(-\lambda_1-\lambda_2)$. If $F(\lambda_1,\lambda_2)$ is absolutely continuous with respect to the Lebesgue measure of $\mathbf{R}^2$ and $\partial F(\lambda_1,\lambda_2)/\partial\lambda_1\partial\lambda_2=f(\lambda_1,\lambda_2)$, we call $f(\lambda_1,\lambda_2)$ the **bispectral density function**. When X_t is Gaussian, $R_{h_1,h_2}=0$ and $f(\lambda_1,\lambda_2)=0$ for any h_1, h_2 and any λ_1, λ_2. $f(\lambda_1,\lambda_2)$ can be considered to give a kind of measure of the departure from a Gaussian process or a kind of nonlinear relationship among waves of different frequencies. We can construct an estimator for $f(\lambda_1,\lambda_2)$ by using windows as in the estimation of a spectral density [3].

D. Statistical Analysis of Parametric Models

When we assume merely that X_t is a stationary process and nothing further, then X_t contains infinite-dimensional unknown parameters. In this case, it may be difficult to develop a satisfactory general theory for statistical inference about X_t. But in most practical applications of time series analysis, we can safely assume at least some of the time dependences to be known. For this reason, we can often use a model with finite-dimensional parameters. This means, mainly, that the moments (usually, second-order moments) or the spectral density are assumed to be expressible in terms of finite-dimensional parameters. As examples of such

models, autoregressive models, moving average models, and autoregressive moving average models are widely used.

A process X_t is called an **autoregressive process** of order K if X_t satisfies a difference equation $\sum_{k=0}^{K} a_k X_{t-k} = \xi_t$, where the a_k are constants, $a_0 = 1$, $a_K \neq 0$, and the ξ_t are mutually uncorrelated with $E\xi_t = 0$ and $V(\xi_t) = \sigma_\xi^2 > 0$. We usually assume that X_t is a weakly stationary process with $EX_t = 0$. We sometimes use the notation AR(K) to express a weakly stationary and autoregressive process of order K. Let $\{\xi_t\}$ be as above. If X_t is expressed as $X_t = \sum_{l=0}^{L} b_l \xi_{t-l}$, where the b_l are constants, $b_0 = 1$ and $b_L \neq 0$, X_t is called a **moving average process** of order L (MA(L) process). Furthermore, if X_t is weakly stationary with $EX_t = 0$ and expressed as $\sum_{k=0}^{K} a_k X_{t-k} = \sum_{l=0}^{L} b_l \xi_{t-l}$ with $a_0 = 1$, $b_0 = 1$, and $a_K b_L \neq 0$, then X_t is called an **autoregressive moving average process** of order (K, L) (ARMA(K, L) process). Let $A(Z)$ and $B(Z)$ be two polynomials of Z such that $A(Z) = \sum_{k=0}^{K} a_k Z^{K-k}$ and $B(Z) = \sum_{l=0}^{L} b_l Z^{L-l}$, and let $\{\alpha_k | 1 \leqslant k \leqslant K\}$ and $\{\beta_l | 1 \leqslant l \leqslant L\}$ be the solutions of the associated polynomial equations $A(Z) = 0$ and $B(Z) = 0$, respectively, we assume that $|\alpha_k| < 1$ for $1 \leqslant k \leqslant K$ and $|\beta_l| < 1$ for $1 \leqslant l \leqslant L$. This condition implies that X_t is purely nondeterministic. Let the observed sample be $\{X_t | 1 \leqslant t \leqslant T\}$. If we assume that X_t is Gaussian and an ARMA(K, L) process, we can show that the †maximum likelihood estimators $\{\hat{a}_k\}$ and $\{\hat{b}_l\}$ of $\{a_k\}$ and $\{b_l\}$ are †consistent and asymptotically efficient when $T \to \infty$ ("asymptotically efficient" means that the covariance matrix of the distribution of the estimators is asymptotically equal to the inverse of the information matrix) [5] (→ 399 Statistical Estimation D). Furthermore, if X_t is an AR(K) process, the joint distribution of $\{\sqrt{T}(\hat{a}_k - a_k) | 1 \leqslant k \leqslant K\}$ tends to a K-dimensional normal distribution with means 0, and this distribution is the same as the one to which the distribution of the †least-square estimators $\{\hat{\hat{a}}_k\}$ minimizing $Q = \sum_{t=K+1}^{T} (X_t + \sum_{k=1}^{K} a_k X_{t-k})^2$ tends when $T \to \infty$. If X_t is a MA(L) or ARMA(K, L) process ($L \geqslant 1$), the likelihood equations are complicated and cannot be solved directly. Many approximation methods have been proposed to obtain the estimates.

When X_t is an AR(K) process with $|\alpha_k| < 1$ for $1 \leqslant k \leqslant K$, R_t satisfies $\sum_{k=0}^{K} a_k R_{h-k} = 0$ for $h \geqslant 1$. These are often called the **Yule-Walker equations**. R_h can be expressed as $R_h = \sum_{j=1}^{K} C_j \alpha_j^h$ if the α_k are distinct and $a_K \neq 0$, where $\{C_j\}$ are constants and determined by R_h for $0 \leqslant h \leqslant K-1$. When X_t is an ARMA(K, L) process, $\sum_{k=0}^{K} a_k R_{h-k} = 0$ for $h \geqslant L+1$, and the C_j of $R_h = \sum_{j=1}^{k} C_j \alpha_j^h$ are determined by $\{R_h | 0 \leqslant h \leqslant \max(K, L)\}$.

The spectral density is expressed as $f(\lambda) = \sigma_\xi^2 |B(e^{2\pi i\lambda})|^2 / |A(e^{2\pi i\lambda})|^2$. If X_t is Gaussian, the maximum likelihood estimator of $f(\lambda)$ is asymptotically equal to the statistic obtained by replacing σ_ξ^2, $\{b_l\}$, and $\{a_k\}$ in $f(\lambda)$ with $\hat{\sigma}_\xi^2$, $\{\hat{b}_l\}$, and $\{\hat{a}_k\}$, respectively, where $\hat{\sigma}_\xi^2$ is the maximum likelihood estimator of σ_ξ^2, when $T \to \infty$.

When we analyze a time series and intend to fit an ARMA(K, L) model, we have to determine the values of K and L. For AR(K) models, many methods have been proposed to determine the value of K. Some examples are: (i) (Quenouille) Let $(Z^K A(1/Z))^2 = \sum_{j=0}^{2K} A_j Z^j$, and $G_K = \sum_{K-j=0}^{2K} \hat{A}_j(\tilde{R}_j / \tilde{R}_0)$, where $\hat{A}_j$ is obtained by replacing $\{a_k\}$ in A_j by $\{\hat{a}_k\}$, and we construct the statistic $\chi_f^2 = \sum_{l=1}^{f} G_{K+l}$. Then χ_f^2 has a †χ^2 distribution asymptotically with f degrees of freedom under the assumption that $K \geqslant K_0$, where K_0 is the true order, as $T \to \infty$. Using this fact, we can determine the order of an AR model. (ii) (Akaike) We consider choosing an order K satisfying $K_L \leqslant K \leqslant K_M$, where K_L and K_M are minimum order and maximum order, respectively, specified a priori. Then we construct the statistic $\mathrm{AIC}(K) = (T-K)\log \hat{\hat{\sigma}}_\xi^2(K) + 2K$, where

$$\hat{\hat{\sigma}}_\xi^2(K) = \sum_{t=K+1}^{T} (X_t + \hat{\hat{a}}_1 X_{t-1} + \ldots + \hat{\hat{a}}_K X_{t-K})^2 / T$$

and $\{\hat{\hat{a}}_k | 1 \leqslant k \leqslant K\}$ are the least square estimators of the autoregressive coefficients of an AR(K) model fitting X_t. Calculate AIC(K) for $K = K_L, K_L + 1, \ldots, K_M$. If AIC($K$) has the minimum value at $K = \hat{K}$, we determine the order to be $\hat{K}$ [6] (→ 403 Statistical Models F). Parzen proposed another method by using the criterion autoregressive transfer function (CAT). Here $\mathrm{CAT}(K) = 1 - \tilde{\sigma}^2(\infty)/\tilde{\sigma}_\xi^2(K) + K/T$, where $\tilde{\sigma}_\xi^2(K) = (T/(T-K))\hat{\hat{\sigma}}_\xi^2(K)$ and $\tilde{\sigma}^2(\infty)$ is an estimator of $\sigma^2(\infty) = \exp(\int_{-1/2}^{1/2} \log f(\lambda) d\lambda)$ [7]. (iii) We can construct a test statistic for the null hypothesis AR(K) against the alternative hypothesis AR($K+1$) (Jenkins) or use a multiple decision procedure (T. W. Anderson [8]).

Not much is known about the statistical properties of the above methods, and few comparisons have been made among them.

Another parametric model is an exponential model for the spectrum. The spectral density is expressed by $f(\lambda) = C^2 \exp\{2\sum_{k=1}^{K} \theta_k \cos(2\pi k\lambda)\}$, where the θ_k and C are constants.

We now discuss some general theories of estimation for finite-dimensional-parameter models. Let X_t be a real-valued Gaussian process of mean 0 and of spectral density $f(\lambda)$ which is continous and positive in $[-1/2, 1/2]$, and let the moving average representation of X_t be $X_t = \sum_{l=0}^{\infty} b_l \xi_{t-l}$, where ξ_t is a white noise and $\sigma_\xi^2 = E\xi_t^2$. We assume that $f(\lambda)/\sigma_\xi^2 = g(\lambda)$

depends only on M parameters $\boldsymbol{\theta}=(\theta_1,\theta_2,\ldots,\theta_M)'$ which are independent of σ_ξ^2. Then the logarithm of the †likelihood function can be approximated by $-(1/2)\{T\log 2\pi\sigma_\xi^2+\mathbf{X}'\Sigma_T^{-1}(\boldsymbol{\theta})\mathbf{X}/\sigma_\xi^2\}$ by ignoring the lower-order terms in T, where $\sigma_\xi^2\Sigma_T(\boldsymbol{\theta})$ is the covariance matrix of $\mathbf{X}$. Usually, it is difficult to find an explicit expression for each element of $\Sigma_T^{-1}(\boldsymbol{\theta})$. Another approximation for the logarithm of the likelihood function is given by

$$-\frac{T}{2}\int_{-1/2}^{1/2}\left[\log f(\lambda)+\frac{I_T(\lambda)}{f(\lambda)}\right]d\lambda.$$

Under mild conditions on the regularity of $g(\lambda)$, the estimators $\hat{\boldsymbol{\theta}}=(\hat\theta_1,\hat\theta_2,\ldots,\hat\theta_M)$ and $\hat\sigma_\xi^2$, obtained as the solutions of the likelihood equations, are †consistent and asymptotically normal as T tends to infinity. This means that the distribution of $\sqrt{T}(\hat\sigma_\xi^2-\sigma_\xi^2)$ is asymptotically normal and $\sqrt{T}(\hat\sigma_\xi^2-\sigma_\xi^2)$ and $\sqrt{T}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})$ are asymptotically independent. The asymptotic distribution of $\sqrt{T}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})$ is the normal distribution $N(\mathbf{0},\Gamma^{-1})$, where the (k,l)-component Γ_{kl} of Γ is given by

$$\Gamma_{kl}=\frac{1}{2}\int_{-1/2}^{1/2}\left(\frac{\partial\log g(\lambda)}{\partial\theta_k}\cdot\frac{\partial\log g(\lambda)}{\partial\theta_l}\right)_{\boldsymbol{\theta}}d\lambda.$$

E. Statistical Analysis of Multiple Time Series

Let $\mathbf{X}_t=(X_t^{(1)},X_t^{(2)},\ldots,X_t^{(p)})'$ be a complex-valued weakly stationary process with $E\mathbf{X}_t=\mathbf{0}$ and $E\mathbf{X}_t\bar{\mathbf{X}}_s'=R_{t-s}$. R_{t-s} is the $p\times p$ matrix whose (k,l)-component is $R_{t-s}^{(k,l)}=EX_t^{(k)}\bar X_s^{(l)}$. We discuss the case when t is an integer. R_h has the spectral representation

$$R_h=\int_{-1/2}^{1/2}e^{2\pi ih\lambda}dF(\lambda),$$

where $F(\lambda)$ is a $p\times p$ matrix and $F(\lambda_1)-F(\lambda_2)$, $\lambda_1\geqslant\lambda_2$, is Hermitian nonnegative. Let $f^{k,l}(\lambda)$ be the (k,l)-component of the spectral density matrix $f(\lambda)$, i.e., $F_a(\lambda)=\int_{-1/2}^{\lambda}f(\mu)d\mu$, of the absolutely continuous part in the Lebesgue decomposition of $F(\lambda)$. The function $f^{k,l}(\lambda)$ for $k\neq l$ is called the **cross spectral density function.** $f^{k,l}(\lambda)$ represents a kind of correlation between the wave of frequency λ included in $X_t^{(k)}$ and the one included in $X_t^{(l)}$.

Let $\mathbf{X}_t=(X_t^{(1)},X_t^{(2)},\ldots,X_t^{(p)})'$ and $\mathbf{Y}_t=(Y_t^{(1)},Y_t^{(2)},\ldots,Y_t^{(q)})'$ be two complex-valued weakly stationary processes with $E\mathbf{X}_t=0$, $E\mathbf{Y}_t=0$, $E\mathbf{X}_t\bar{\mathbf{X}}_s'=R_{t-s}^{\mathbf{X}}$, $E\mathbf{Y}_t\bar{\mathbf{Y}}_s'=R_{t-s}^{\mathbf{Y}}$ and $E\mathbf{X}_t\bar{\mathbf{Y}}_s'=R_{t-s}^{\mathbf{XY}}$. We assume $\mathbf{Y}_t=\sum_{s=-\infty}^{\infty}A_s\mathbf{X}_{t-s}$, where A_s is a $q\times p$ matrix whose components are constants depending on s. Put $A(\lambda)=\sum_{s=-\infty}^{\infty}A_se^{-2\pi is\lambda}$. $A(\lambda)$ should exist in the sense of mean square convergence with respect to the spectral distribution function F for $\mathbf{X}_t$.

The function $A(\lambda)$ is called the matrix **frequency response function.**

As a measure of the strength of association between $X_t^{(k)}$ and $X_t^{(l)}$ at frequency λ, we introduce the quantity $\gamma^{k,l}(\lambda)=|f^{k,l}(\lambda)|^2/f^{k,k}(\lambda)f^{l,l}(\lambda)$. $\gamma^{k,l}(\lambda)$ is called the **coherence**. Let $X_t^{(k)}=\sum_{s=-\infty}^{\infty}a_s^{k,l}X_{t-s}^{(l)}+\eta_t$, where η_t is a weakly stationary process with mean 0 and uncorrelated with $X_s^{(l)}$, $-\infty<s<\infty$. If $E|\eta_t|^2=0$, $\gamma^{k,l}(\lambda)=1$. If $E|\sum_{-\infty}^{\infty}a_s^{k,l}X_{t-s}^{(l)}|^2=0$, $\gamma^{k,l}(\lambda)=0$. Generally, we have $0\leqslant\gamma^{k,l}(\lambda)\leqslant 1$.

For the estimation of $F(\lambda)$, $A(\lambda)$, and $\gamma^{k,l}(\lambda)$, the theories have been similar to those for the estimation of the spectral density of a scalar time series. For example, an estimator of $f(\lambda)$ is given [11] in the form

$$\hat f(\lambda)=\sum_{h=-(T-1)}^{T-1}\tilde{\bar{R}}_hw_he^{-2\pi ih\lambda},$$

where

$$\tilde{\bar{R}}_h=\sum_{t=1}^{T-|h|}\mathbf{X}_{t+|h|}\bar{\mathbf{X}}_t'/T$$

and the w_h are the same as in Section C.

We can define an autoregressive, moving average, or autoregressive moving average process in a similar way as for a scalar time series. The a_k and b_l in Section D should be replaced by $p\times p$ matrices and the associated polynomial equations $A(Z)=0$ and $B(Z)=0$ should be understood in the vector sense [11]. There are problems with determining the coefficients uniquely or identifying an ARMA(K,L) model, and these problems have been discussed to some extent.

F. Statistical Inference of the Mean Function

Let X_t be expressed as $X_t=m_t+Y_t$, where m_t is a real-valued deterministic function of t and Y_t is a real-valued weakly stationary process with mean 0 and spectral distribution function $F(\lambda)$. This means that $EX_t=m_t$. We consider the case when $m_t=\sum_{j=1}^{M}C_j\varphi_t^{(j)}$, where $\mathbf{C}=(C_1,C_2,\ldots,C_M)'$ is a vector of unknown coefficients and $\boldsymbol{\varphi}_t=(\varphi_t^{(1)},\varphi_t^{(2)},\ldots,\varphi_t^{(M)})'$ is a set of known (regression) functions.

Let us construct †linear unbiased estimators $\{\tilde C_j=\sum_{t=1}^{T}\gamma_{jt}X_t\,|\,1\leqslant j\leqslant M\}$ for the coefficients C_j, where the γ_{jt} are known constants. Put $\Phi=(\boldsymbol{\varphi}_1,\boldsymbol{\varphi}_2,\ldots,\boldsymbol{\varphi}_T)'$. Then the †least squares estimator of $\mathbf{C}$ is given by $\hat{\mathbf{C}}=(\Phi'\Phi)^{-1}\Phi'\mathbf{X}$ when $\Phi'\Phi$ is nonsingular. Let Σ be the covariance matrix of $\mathbf{X}$. Then the †best linear unbiased estimator is $\hat{\mathbf{C}}^*=(\Phi'\Sigma^{-1}\Phi)^{-1}\Phi'\Sigma^{-1}\mathbf{X}$. We put $\|\varphi^{(j)}\|_T^2=\sum_{t=1}^{T}(\varphi_t^{(j)})^2$ and assume that $\lim_{T\to\infty}\|\varphi^{(j)}\|_T^2=\infty$, $\lim_{T\to\infty}\|\varphi^{(j)}\|_{T+h}^2/\|\varphi^{(j)}\|_T^2=1$ for $1\leqslant j\leqslant M$ and any fixed h and assume the existence of $\psi_h^{(j,k)}=\lim_{T\to\infty}\sum_{t=1}^{T}\varphi_{t+h}^{(j)}\varphi_t^{(k)}/\|\varphi^{(j)}\|_T\|\varphi^{(k)}\|_T$ for $1\leqslant j,k\leqslant M$. We also assume

that $F(\lambda)$ is absolutely continuous and $F'(\lambda)=f(\lambda)$ is positive and piecewise continuous. Let ψ_h be the $M\times M$ matrix whose (j,k)-component is $\psi_h^{(j,k)}$. Then ψ_h can be represented by

$$\psi_h=\int_{-1/2}^{1/2} e^{2\pi ih\lambda}\,dG(\lambda),$$

where $G(\lambda)-G(\mu)$ is a nonnegative definite matrix for $\lambda>\mu$. Assume that $\psi_0=G(1/2)-G(-1/2)$ is nonsingular and put $H(\lambda)=\psi_0^{-1/2}G(\lambda)\psi_0^{-1/2}$, and for any set S, $H(S)=\int_S H(d\lambda)$. Suppose further that $S_1, S_2, \ldots, S_q$ are q sets such that $H(S_j)>0$, $\sum_{j=1}^q H(S_j)=I$, $H(S_j)H(S_k)=0, j\neq k$, and for any j there is no subset $S_j'\subset S_j$ such that $H(S_j')>0$, $H(S_j-S_j')>0$ and $H(S_j')H(S_j-S_j')=0$. We have $q\leqslant M$. It can be shown that the spectrum of the regression can be decomposed into such disjoint sets $S_1,\ldots,S_q$. Then we can show that $\hat{\mathbf{C}}$ is asymptotically efficient in the sense that the asymptotic covariance matrix of $\hat{\mathbf{C}}$ is equivalent to that of $\hat{\mathbf{C}}^*$ if and only if $f(\lambda)$ is constant on each of the elements S_j. Especially, if $\psi_t^{(j)}=t^je^{2\pi it\mu_j}$, $\hat{\mathbf{C}}$ is asymptotically efficient.

G. Nonstationary Models

It is difficult to develop a statistical theory for a general class of nonstationary time series, but some special types of nonstationary processes have been investigated more or less in detail. Let X_t (t an integer) be a real-valued stochastic process and ∇ be the backward difference operator defined by $\nabla X_t=X_t-X_{t-1}$ and $\nabla^d X_t=\nabla(\nabla^{d-1}X_t)$ for $d\geqslant 2$. We assume that X_t is defined for $t\geqslant t_0$ (t_0 a finite integer), and $EX_t^2<+\infty$. For analyzing a nonstationary time series, Box and Jenkins introduced the following model: For a positive integer d, $Y_t=\nabla^d X_t$, $t\geqslant t_0+d$, is stationary and is an autoregressive moving average process of order (K,L) for $t\geqslant t_0+d+\max(K,L)$. They called such an X_t an **autoregressive integrated moving average process** of order (K,d,L) and denoted it by ARIMA(K,d,L). The word "integrated" means a kind of summation; in fact, X_t can be expressed as a sum of the weakly stationary process Y_t, i.e.,

$$X_t=X_0+(\nabla X_0)t+(\nabla^2X_0)\left(\sum_{s_2=1}^{t}\sum_{s_1=1}^{s_2}\right)+\ldots$$
$$+(\nabla^{d-1}X_0)\left(\sum_{s_{d-1}=1}^{t}\cdots\sum_{s_1=1}^{s_2}\right)$$
$$+\sum_{s_d=1}^{t}\sum_{s_{d-1}=1}^{s_d}\cdots\sum_{s_1=1}^{s_2}Y_{s_1}$$

when $t_0=-d+1$. Using this model, methods of forecasting and of model identification and estimation can be discussed [13].

Another nonstationary model is based on the concept of evolutionary spectra [14]. In this approach, spectral distribution functions are taken to be time-dependent. Let X_t be a complex-valued stochastic process (t an integer) with $EX_t=0$ and $R_{t,s}=EX_t\bar{X}_s$. In the following, we write simply $\int$ for $\int_{-1/2}^{1/2}$. We now restrict our attention to the class of X_t for which there exist functions $\{u_t(\lambda)\}$ defined on $[-1/2,1/2]$ such that $R_{t,s}$ can be expressed as $R_{t,s}=\int u_t(\lambda)\bar{u}_s(\lambda)\,d\mu(\lambda)$, where $\mu(\lambda)$ is a measure. $u_t(\lambda)$ should satisfy $\int|u_t(\lambda)|^2\,d\mu(\lambda)<+\infty$. Then X_t admits a representation of the form $X_t=\int u_t(\lambda)\,dZ(\lambda)$, where $Z(\lambda)$ is a process with orthogonal increments and $E|dZ(\lambda)|^2=d\mu(\lambda)$. If $u_t(\lambda)$ is expressed as $u_t(\lambda)=\gamma_t(\lambda)e^{2\pi i\theta(\lambda)t}$ and $\gamma_t(\lambda)$ is of the form $\gamma_t(\lambda)=\int e^{2\pi itw}\,d\Gamma_\lambda(w)$ with $|d\Gamma_\lambda(w)|$ having the absolute maximum at $w=0$, we call $u_t(\lambda)$ an oscillatory function and X_t an oscillatory process. The evolutionary power spectrum $dF_t(\lambda)$ is defined by $dF_t(\lambda)=|\gamma_t(\lambda)|^2\,d\mu(\lambda)$.

Other models, such as an autoregressive model whose coefficients vary with time or whose associated polynomial has roots outside the unit circle, have also been discussed.

References

[1] E. J. Hannan, Time series analysis, Methuen, 1960.
[2] R. B. Blackman and J. W. Tukey, The measurement of power spectra from the point of view of communications engineering, Dover, 1959.
[3] D. R. Brillinger, Time series: Data analysis and theory, Holt, Rinehart and Winston, 1975.
[4] M. S. Bartlett, An introduction to stochastic processes with special reference to methods and applications, Cambridge Univ. Press, second edition, 1966.
[5] P. Whittle, Estimation and information in stationary time series, Ark. Mat.. 2 (1953), 423–434.
[6] H. Akaike, Information theory and an extension of the maximum likelihood principle, 2nd Int. Symp. Information Theory, B. N. Petrov and F. Csáki (eds.), Akadémiai Kiadó, 1973, 267–281.
[7] E. Parzen, Some recent advances in time series modeling, IEEE Trans. Automatic Control, AC-19 (1974), 723–730.
[8] T. W. Anderson, The statistical analysis of time series, Wiley, 1971.
[9] P. Whittle, Hypothesis testing in time series analysis, Almqvist and Wiksell, 1951.
[10] H. B. Mann and A. Wald, On the statistical treatment of linear stochastic difference equations, Econometrica, 11 (1943), 173–220.

[11] E. J. Hannan, Multiple time series, Wiley, 1970.
[12] U. Grenander and M. Rosenblatt, Statistical analysis of stationary time series, Wiley, 1957.
[13] G. E. P. Box and G. M. Jenkins, Time series analysis: Forecasting and control, Holden-Day, revised edition, 1976.
[14] M. B. Priestley, Evolutionary spectra and nonstationary processes, J. Roy. Statist. Soc., ser. B, 27 (1965), 204–237.

422 (IV.7) Topological Abelian Groups

A. Introduction

A commutative topological group is called a **topological Abelian group**. Throughout this article, except in Section L, all topological groups under consideration are locally compact Hausdorff topological Abelian groups and are simply called groups (→ 423 Topological Groups).

B. Characters

A **character** of a group is a continuous function $\chi(x)$ $(x \in G)$ that takes on as values complex numbers of absolute value 1 and satisfies $\chi(xy)=\chi(x)\chi(y)$. Equivalently, χ is a 1-dimensional and therefore an irreducible †unitary representation of G. Conversely any irreducible unitary representation of G is 1-dimensional. Indeed, for a topological Abelian group, the set of its characters coincides with the set of its irreducible unitary representations. If the product of two characters χ, χ' is defined by $\chi\chi'(x)=\chi(x)\chi'(x)$, then the set of all characters forms the **character group** $C(G)$ of G. With †compact-open topology, $C(G)$ itself becomes a locally compact topological Abelian group.

C. The Duality Theorem

For a fixed element x of G, $\chi(x)$ $(\chi \in C(G))$ is a character of $C(G)$, namely, an element of $CC(G)$. Denote this character of $C(G)$ by $x(\chi)$, and consider the correspondence $G \ni x \to x(\chi)$. That this correspondence is one-to-one follows from the fact that any locally compact G has †sufficiently many irreducible unitary representations (→ 437 Unitary Representations) and the fact that if G is an Abelian group, then any irreducible unitary representation of G is a character of G. Furthermore, any character of $C(G)$ is given as one of the $x(\chi)$; indeed, by this correspondence, we have $G \cong CC(G)$ (**Pontryagin's duality theorem**).

By the duality theorem, each of G and $C(G)$ is isomorphic to the character group of the other. In this sense, G and $C(G)$ are said to be **dual** to each other.

D. Correspondence between Subgroups

Let $G, G'=C(G)$ be groups that are dual to each other. Given a closed subgroup g of G, the set of all χ' such that $\chi'(x)=1$ for all x in g forms a closed subgroup of G', usually denoted by (G',g). The definition of (G,g') is similar. Then $g \leftrightarrow (G',g)=g'$ gives a one-to-one correspondence between the closed subgroups of G and those of G'. If $g_1 \supset g_2$, then g_1/g_2 and $(G',g_2)/(G',g_1)$ are dual to each other. If the group operations of G, G' are written in additive form, with 0 for the identity, then $x(\chi')=1$ is written as $x(\chi')=0$. In this sense, (G',g) is called the **annihilator** (or **annulator**) of g.

E. The Structure Theorem

Let $\mathfrak{A}$ be the set of all groups (more precisely, of all locally compact Hausdorff topological Abelian groups). If $G_1, G_2 \in \mathfrak{A}$, then the direct product $G_1 \times G_2 \in \mathfrak{A}$, and if $G \in \mathfrak{A}$ and H is a closed subgroup of G, then $H \in \mathfrak{A}$ and $G/H \in \mathfrak{A}$. In addition, if H is a closed subgroup of a group G such that $H \in \mathfrak{A}$ and $G/H \in \mathfrak{A}$, then $G \in \mathfrak{A}$. In other words, $\mathfrak{A}$ is closed under the operations of forming direct products, closed subgroups, quotient groups, and †extensions by members of $\mathfrak{A}$. Furthermore, the operation C that assigns to each element of $\mathfrak{A}$ its dual element is a reflexive correspondence of $\mathfrak{A}$ onto $\mathfrak{A}$, and if $G \supset H$, the annihilator $(C(G), H)$ of H is a closed subgroup of $C(G)$. Also, $C(G/H) \cong (C(G), H)$, $C(H) \cong C(G)/(C(G), H)$. Furthermore, $C(G_1 \times G_2) \cong C(G_1) \times C(G_2)$. Finally, $H=(G,(C(G),H))$ (**reciprocity of annihilators**).

Typical examples of groups in $\mathfrak{A}$ are the additive group $\mathbf{R}$ of real numbers, the additive group $\mathbf{Z}$ of rational integers, the 1-dimensional †torus group $\mathbf{T}=\mathbf{R}/\mathbf{Z}$, and finite Abelian groups $\mathbf{F}$. The torus group $\mathbf{T}$ is also isomorphic to the multiplicative group $U(1)$ of complex numbers of absolute value 1. The direct product $\mathbf{R}^n$ of n copies of $\mathbf{R}$ is the **vector group** of dimension n, and the direct product $\mathbf{T}^n$ of n copies of $\mathbf{T}$ is the **torus** (or **torus group**) of dimension n (or n-**torus**). Both $\mathbf{T}^n$ and $\mathbf{F}$ are compact, while $\mathbf{R}^n$ and $\mathbf{Z}^n$ are not. We have $C(\mathbf{R})=\mathbf{R}$, $C(\mathbf{T})=\mathbf{Z}$, $C(\mathbf{Z})=\mathbf{T}$. Any finite Abelian group $\mathbf{F}$ is isomorphic to its character group $C(\mathbf{F})$. The direct product of a finite

number of copies of **R**, **T**, **Z**, and a finite Abelian group **F**, namely, a group of the form $\mathbf{R}^l \times \mathbf{T}^m \times \mathbf{Z}^n \times \mathbf{F}$, is called an **elementary topological Abelian group**.

Any group in $\mathfrak{A}$ is isomorphic to the direct product of a vector group of some dimension and the extension of a compact group by a discrete group (the **structure theorem**). Hence, if the effect of the operation C is explicitly known, then the problem of finding the structure of groups in $\mathfrak{A}$ is reduced to the problem concerning discrete groups alone. For the structure of groups in $\mathfrak{A}$, the following theorem is known: If $G \in \mathfrak{A}$ is generated by a compact neighborhood of the identity e, then G is isomorphic to the direct product of a compact subgroup K and a group of the form $\mathbf{R}^n \times \mathbf{Z}^m$ (n, m are nonnegative integers). Then any compact subgroup of G is contained in K, which is the unique maximal compact subgroup of G. A group $G \in \mathfrak{A}$ generated by a compact neighborhood of e is the †projective limit of elementary topological Abelian groups. L. S. Pontryagin first proved a structure theorem of this type and then the duality theorem.

F. Compact Elements

An element a of a group $G \in \mathfrak{A}$ is called a **compact element** if the cyclic group $\{a^n \mid n \in \mathbf{Z}\}$ generated by a is contained in a compact subset of G. The set C_0 of all compact elements of G is a closed subgroup of G, and the quotient group G/C_0 does not contain any compact element other than the identity. In particular, if G is generated by a compact neighborhood of the identity, then C_0 coincides with the maximal compact subgroup K of G. Let C_0 be the set of all compact elements of a group $G \in \mathfrak{A}$. The annihilator $(C(G), C_0)$ is a connected component of the character group $C(G)$ of G. If G is a discrete group, then a compact element of G is an element of G of finite order.

G. Compact Groups and Discrete Groups

Suppose that two groups $G, X \in \mathfrak{A}$ are dual to each other. Then one group is compact if and only if the other group is discrete. By the duality theorem, the properties of a compact Abelian group G can be stated, in principle, through the properties of the discrete Abelian group $C(G)$. The following are a few such examples. Let G be a compact Abelian group. Then its †dimension is equal to the †rank of the discrete Abelian group $C(G)$. A subgroup Y of a discrete Abelian group X is called a **divisible subgroup** if the quotient group X/Y contains no element of finite order other than the identity. A compact Abelian group G is locally connected if and only if any finite subset of the character group $C(G)$ is contained in some divisible subgroup of $C(G)$ generated by a finite number of elements. Hence if a compact locally connected Abelian group G has an †open basis consisting of a countable number of open sets, then G is of the form $\mathbf{T}^a \times \mathbf{F}$, where **F** is a finite Abelian group and $\mathbf{T}^a$ is the direct product of an at most countable number of 1-dimensional torus groups **T**.

H. Dual Decomposition into Direct Products

Let G be a compact or discrete Abelian group, and let $\mathfrak{M} = \{H_\alpha \mid \alpha \in A\}$ be a family of closed subgroups of G. Let $\Delta(\mathfrak{M}) = \bigcap_{\alpha \in A} H_\alpha$, and denote by $\Sigma(\mathfrak{M})$ the smallest closed subgroup of G containing $\bigcup_{\alpha \in A} H_\alpha$. Then, with $\Omega = \{(C(G), H_\alpha) \mid \alpha \in A\}$, the relations $\Delta(\Omega) = (C(G), \Sigma(\mathfrak{M}))$ and $\Sigma(\Omega) = (C(G), \Delta(\mathfrak{M}))$ hold. Furthermore, suppose that G is decomposed into the direct product $G = \prod_{\alpha \in A} H_\alpha$, and for each $\alpha \in A$ put $K_\alpha = \Sigma(\mathfrak{M} - \{H_\alpha\})$, $X_\alpha = (C(G), K_\alpha)$. Then X_α is the character group of H_α, and $C(G)$ can be decomposed into the direct product $C(G) = \prod_{\alpha \in A} X_\alpha$. This decomposition of $C(G)$ into a direct product is called the **dual direct product decomposition** corresponding to the decomposition $G = \prod_{\alpha \in A} H_\alpha$.

I. Orthogonal Group Pairs

Suppose that for two groups G, G' there exists a mapping $(x, x') \rightarrow xx'$ of the Cartesian product $G \times G'$ into the set $U(1)$ of all complex numbers of absolute value 1 such that

$$(x_1 x_2)x' = (x_1 x')(x_2 x'),$$

$$x(x_1' x_2') = (xx_1')(xx_2').$$

Then G, G' are said to form a **group pair**. Suppose that G, G' form a group pair, and consider xx' to be a function $x(x')$ in x'. If two functions $x_1(x')$ and $x_2(x')$ coincide only when $x_1 = x_2$ and the same is true when the roles of G and G' are interchanged, then G, G' are said to form an **orthogonal group pair**. If G is a compact Abelian group, G' is a discrete Abelian group, and G, G' form an orthogonal group pair, then G, G' are dual to each other.

J. Commutative Lie Groups

An elementary topological Abelian group $\mathbf{R}^l \times \mathbf{T}^m \times \mathbf{Z}^n \times \mathbf{F}$ is a commutative †Lie group. Conversely, any commutative Lie group G

generated by a compact neighborhood of the identity is isomorphic to an elementary topological Abelian group. In particular, any connected commutative Lie group G is isomorphic to $\mathbf{R}^l \times \mathbf{T}^m$ for some l and m. A closed subgroup H of the vector group $\mathbf{R}^n$ of dimension n is isomorphic to $\mathbf{R}^p \times \mathbf{Z}^q$ $(0 \leqslant p+q \leqslant n)$. More precisely, there exists a basis $a_1, \ldots, a_n$ of the vector group $\mathbf{R}^n$ such that $H = \{\sum_{i=1}^{p} x_i a_i + \sum_{j=p+1}^{r} n_j a_j \mid x_i \in \mathbf{R}, n_j \in \mathbf{Z}\}$. Hence the quotient groups of $\mathbf{R}^n$ that are †separated topological groups are all isomorphic to groups of the form $\mathbf{R}^l \times \mathbf{T}^m$ $(0 \leqslant l+m \leqslant n)$. Any closed subgroup of the torus group $\mathbf{T}^n$ of dimension n is isomorphic to a group of the form $\mathbf{T}^p \times \mathbf{F}$ $(0 \leqslant p \leqslant n)$, where $\mathbf{F}$ is a finite Abelian group. Hence the quotient groups of $\mathbf{T}^n$ that are separated topological groups are all isomorphic to $\mathbf{T}^m$ $(0 \leqslant m \leqslant n)$. A †regular linear transformation of the linear space $\mathbf{R}^n$ is a continuous automorphism of the vector group $\mathbf{R}^n$, and in fact, any continuous automorphism of $\mathbf{R}^n$ is given by a regular linear transformation. Indeed, the group of all continuous automorphisms of $\mathbf{R}^n$ is isomorphic to the †general linear group $GL(n, \mathbf{R})$ of degree n. Any continuous automorphism of the torus group $\mathbf{T}^n = \mathbf{R}^n/\mathbf{Z}^n$ of dimension n is given by a regular linear transformation φ of $\mathbf{R}^n$ such that $\varphi(\mathbf{Z}^n) = \mathbf{Z}^n$. Hence the group of continuous automorphisms of $\mathbf{T}^n$ is isomorphic to the multiplicative group of all $n \times n$ matrices, with determinant ± 1 and with entries in the set of rational integers.

K. Kronecker's Approximation Theorem

Let H be a subgroup of a group $G \in \mathfrak{A}$ (not necessarily closed). Then $(G,(C(G), H))$ coincides with the closure $\bar{H}$ of H. In particular, H is †dense in G if and only if the annihilator $(C(G), H)$ consists of the identity alone. Now let $G = \mathbf{R}^n$ and let H be the subgroup of $\mathbf{R}^n$ generated by $\theta = (\theta_1, \ldots, \theta_n) \in \mathbf{R}^n$ and the natural †basis $e_1 = (1, 0, \ldots, 0), \ldots, e_n = (0, \ldots, 0, 1)$ of $\mathbf{R}^n$. Then H is dense in $\mathbf{R}^n$ if and only if $(\mathbf{R}^n, H) = \{0\}$; that is, $\theta_1, \ldots, \theta_n, 1$ are linearly independent over the rational number field $\mathbf{Q}$ (**Kronecker's approximation theorem**). This theorem implies that the torus group $\mathbf{T}^n$ of dimension n has a cyclic subgroup and a 1-parameter subgroup that are both dense in $\mathbf{T}^n$.

L. Linear Topology

Consider the discrete topology in a field Ω. Suppose that an Ω-module G has a topology that satisfies †Hausdorff's separation axiom and is such that a base for the neighborhood system of the zero element 0 consists of Ω-submodules, and suppose that G together with this topology constitutes a topological Abelian group. Then this topology is called a **linear topology**. If a linear topology is restricted to a Ω-submodule, then it is also a linear topology. If G is of finite rank, then any linear topology is the discrete topology. The discrete topology on G is a linear topology. Let H be a Ω-submodule. Then the subset $V = H + g$ of G obtained by translating H by an element g of G is called a **linear variety** in G. If V is a linear variety, then $\bar{V}$ is also a linear variety. If Ω-modules G, G' have linear topologies, a homomorphism of G into G' is always assumed to be open and continuous with respect to these topologies. A linear variety V in G is said to be **linearly compact** if, for any system $\{V_\alpha\}$ of linear varieties closed in V with the †finite intersection property, we have $\bigcap_\alpha V_\alpha \neq \varnothing$. In this case V is closed in G. If linearly compact Ω-submodules can be chosen as a base for the neighborhood system of the zero element of G, we say that G is **locally linearly compact**. The set $C_\Omega(G)$ of homomorphisms of an Ω-module G with linear topology into Ω is also an Ω-module. For any linearly compact Ω-submodule H of G, let $U(H) = \{\chi \mid \chi(g) = 0, g \in H\}$. Then, with $\{U(H)\}$ as a base for the neighborhood system, a linear topology can be introduced in $C_\Omega(G)$. According as G is discrete, linearly compact, or locally linearly compact, $C_\Omega(G)$ is linearly compact, discrete, or locally linearly compact. Let G, H be Ω-modules each of which has a linear topology, and let $\varphi: G \ni g \rightarrow \varphi_g \in C_\Omega(H)$, $\psi: H \ni h \rightarrow \psi_h \in C_\Omega(G)$ be homomorphisms such that $\varphi_g(h) = \psi_h(g)$. Then if one of φ, ψ is an isomorphism, so is the other. This is an analog of the Pontryagin duality theorem and is called the **duality theorem for Ω-modules**. In particular, a linearly compact Ω-module is the direct sum of 1-dimensional spaces (S. Lefschetz [3]).

References

[1] L. S. Pontryagin, Topological groups, first edition, Princeton Univ. Press, 1939; second edition, Gordon & Breach, 1966 (translation of the 1954 Russian edition).
[2] A. Weil, L'intégration dans les groupes topologiques et ses applications, Actualités Sci. Ind., Hermann, 1940.
[3] S. Lefschetz, Algebraic topology, Amer. Math. Soc. Colloq. Publ., 1942.
[4] W. Rudin, Fourier analysis on groups, Interscience, 1962.
[5] E. Hewitt and K. A. Ross, Abstract harmonic analysis, Springer, I, 1963; II, 1970.

423 (IV.6) Topological Groups

A. Definitions

If a †group G has the structure of a †topological space such that the mapping $(x, y) \rightarrow xy$ (product) of the Cartesian product $G \times G$ into G and the mapping $x \rightarrow x^{-1}$ (inverse) of G into G are both continuous, then G is called a **topological group**. The group G without a topological structure is called the **underlying group** of the topological group G, and the topological space G is called the **underlying topological space** of the topological group G. Let G, G' be topological groups. A mapping f of G into G' is called an **isomorphism** of the topological group G onto the topological group G' if f is an †isomorphism of the underlying group G onto the underlying group G' and also a †homeomorphism of the underlying topological space G onto the underlying topological space G'. Two topological groups are said to be **isomorphic** if there exists an isomorphism of one onto the other.

B. Neighborhood Systems

Let $\mathfrak{N}$ be the †neighborhood system of the identity e of a topological group G. Namely, $\mathfrak{N}$ consists of all subsets of G each of which contains an open set containing the element e. Then $\mathfrak{N}$ satisfies the following six conditions: (i) If $U \in \mathfrak{N}$ and $U \subset V$, then $V \in \mathfrak{N}$. (ii) If U, $V \in \mathfrak{N}$, then $U \cap V \in \mathfrak{N}$. (iii) If $U \in \mathfrak{N}$, then $e \in U$. (iv) For any $U \in \mathfrak{N}$, there exists a $W \in \mathfrak{N}$ such that $WW = \{xy \mid x, y \in W\} \subset U$. (v) If $U \in \mathfrak{N}$, then $U^{-1} \in \mathfrak{N}$. (vi) If $U \in \mathfrak{N}$ and $a \in G$, then $aUa^{-1} \in \mathfrak{N}$. Conversely, if a nonempty family $\mathfrak{N}$ of subsets of a group G satisfies conditions (i)–(vi), then there exists a †topology $\mathfrak{O}$ of G such that $\mathfrak{N}$ is the neighborhood system of e and G is a topological group with this topology. Moreover, such a topology is uniquely determined by $\mathfrak{N}$. †Left translation $x \rightarrow ax$ and †right translation $x \rightarrow xa$ in a topological group G are homeomorphisms of G onto G; thus if $\mathfrak{N}$ is the neighborhood system of the identity e, then $a\mathfrak{N} = \mathfrak{N}a$ is the neighborhood system of a, where $a\mathfrak{N} = \{aU \mid U \in \mathfrak{N}\}$.

If the underlying topological space of a topological group G is a †Hausdorff space, G is called a **T_2-topological group (Hausdorff topological group** or **separated topological group)**. If the underlying topological space of a topological group G is a †T_0-topological space, then, as is easily seen, it is a †T_1-topological space. If it is a T_1-topological space, then by the fact that the topology may be defined by a †uniformity, it is a †completely regular space, hence, in particular, a Hausdorff space (→ Section G). Thus a topological group whose underlying topological space is a T_0-topological space is a T_2-topological group.

C. Direct Product of Topological Groups

Consider a family $\{G_\alpha\}_{\alpha \in A}$ of topological groups. The Cartesian product $G = \prod_{\alpha \in A} G_\alpha$ of the underlying groups of G_α is a topological group with the †product topology of the underlying topological spaces of G_α. This topological group $G = \prod_{\alpha \in A} G_\alpha$ is called the **direct product** of topological groups G_α $(\alpha \in A)$.

D. Subgroups

Let H be a subgroup of the underlying group of a topological group G. Then H is a topological group with the topology of a †topological subspace of G (†relative topology). This topological group H is called a **subgroup** of G. A subgroup that is a closed (open) set is called a **closed (open) subgroup**. Any open subgroup is also a closed subgroup. For any subgroup H of a topological group G, the closure $\bar{H}$ of H is also a subgroup. If H is a normal subgroup, so is $\bar{H}$. If H is commutative, so is $\bar{H}$. In a T_2-topological group G, the †centralizer $C(M) = \{x \in G \mid xm = mx \ (m \in M)\}$ of a subset M of G is a closed subgroup of G. In particular, the †center $C = C(G)$ of a T_2-topological group is a closed normal subgroup.

E. Quotient Spaces

Given a subgroup H of a topological group G, let $G/H = \{aH \mid a \in G\}$ be the set of †left cosets, and let p be the canonical surjection $p(a) = aH$ of G onto G/H. Consider the †quotient topology on G/H, namely, the strongest topology such that p is a continuous mapping. Since a subset A of G/H is open when $p^{-1}(A)$ is an open set of G, p is also an †open mapping. The set G/H with this topology is called the **left quotient space** (or **left coset space**) of G by H. The **right quotient space** (or **right coset space**) $H \backslash G = \{Ha \mid a \in G\}$ is defined similarly. The quotient space G/H is discrete if and only if H is an open subgroup of G. The quotient space is a Hausdorff space if and only if H is a closed subgroup. If G/H and H are both †connected, then G itself is connected. If G/H and H are both †compact, then G is compact. If H is a closed subgroup of G and G/H, H are both †locally compact, then G is locally compact.

Suppose that H is a normal subgroup of a topological group G. Then the quotient group

G/H is a topological group with the topology of the quotient space G/H. This topological group is called the **quotient group** of the topological group G by the normal subgroup H.

F. Connectivity

The †connected component G_0 containing the identity e of a topological group G is a closed normal subgroup of G. The connected component that contains an element $a \in G$ is the coset $aG_0 = G_0 a$. G_0 is called the **identity component** of G. The quotient group G/G_0 is †totally disconnected. A connected topological group G is generated by any neighborhood U of the identity. Namely, any element of G can be expressed as the product of a finite number of elements in U. Totally disconnected (in particular, discrete) normal subgroups of a connected topological group G are contained in the center of G.

G. Uniformity

Let $\mathfrak{N}_0$ be the neighborhood system of the identity of a topological group G, and let $U_l = \{(x, y) \in G \times G \mid y \in xU\}$ for $U \in \mathfrak{N}_0$. Then a †uniformity having $\{U_l \mid U \in \mathfrak{N}_0\}$ as a base is defined on G. This uniformity is called the **left uniformity** of G. Left translation $x \rightarrow ax$ of G is †uniformly continuous with respect to the left uniformity. The **right uniformity** is defined similarly by $U_r = \{(x, y) \mid y \in Ux\}$. These two uniformities do not necessarily coincide. The mapping $x \rightarrow x^{-1}$ is a †uniform isomorphism of G considered as a uniform space with respect to the left uniformity onto the same group G considered as a uniform space with respect to the right uniformity. A topological group G is thus a †uniform space under a uniformity †compatible with its topology, and hence it is a completely regular space if the underlying topological space is a T_1-space.

H. Completeness

If a topological group G is †complete with respect to the left uniformity, then it is also complete with respect to the right uniformity, and conversely. In this case the topological group G is said to be **complete**. A locally compact T_2-topological group is complete. If a T_2-topological group G is isomorphic to a dense subgroup of a complete T_2-topological group $\hat{G}$, then $\hat{G}$ is called the **completion** of G, and G is said to be **completable**. A T_2-topological group G is not always completable. For a T_2-topological group G to be completable it is necessary and sufficient that any †Cauchy filter of G considered as a uniform space with respect to the left uniformity is mapped to a Cauchy filter of the same uniform space G under the mapping $x \rightarrow x^{-1}$. Then the completion $\hat{G}$ of G is uniquely determined up to isomorphism. A commutative T_2-topological group always has a completion $\hat{G}$, and $\hat{G}$ is also commutative. If each point of a T_2-topological group G has a †totally bounded neighborhood, there exists a completion $\hat{G}$, and $\hat{G}$ is locally compact.

I. Metrization

If a †metric can be introduced in a T_2-topological group G so that the metric gives the topology of G, then G is said to be **metrizable**. For a T_2-topological group G to be metrizable it is necessary and sufficient that G satisfy the †first axiom of countability. Then the metric can be chosen so that it is **left invariant**, i.e., invariant under left translation. Similarly, it can be chosen so that it is right invariant. In particular, the topology of a compact T_2-topological group that satisfies the first axiom of countability can be given by a metric that is both left and right invariant.

J. Isomorphism Theorems

Let G and G' be topological groups. If a homomorphism f of the underlying group of G into the underlying group of G' is a continuous mapping of the underlying topological space of G into that of G', f is called a **continuous homomorphism**. If f is a continuous open mapping, f is called a **strict morphism** (or **open continuous homomorphism**). A continuous homomorphism of a †paracompact locally compact topological group onto a locally compact T_2-topological group is an open continuous homomorphism.

A topological group G' is said to be **homomorphic** to a topological group G if there exists an open continuous homomorphism f of G onto G'. Let N denote the kernel $f^{-1}(e)$ of f. Then the quotient group G/N is isomorphic to G', with G/N and G' both considered as topological groups (**homomorphism theorem**). Let f be an open continuous homomorphism of a topological group G onto a topological group G', and let H' be a subgroup of G'. Then $H = f^{-1}(H')$ is a subgroup of G, and the mapping φ defined by $\varphi(gH) = f(g)H'$ is a homeomorphism of the quotient space G/H onto G'/H'. In particular, if H' is a normal subgroup, then H is also a normal subgroup and φ is an isomorphism of the quotient group G/H onto G'/H' as topological groups (**first isomorphism theorem**). Let H and N be subgroups of a topo-

logical group G such that $HN=NH$. Then the canonical mapping $f: h(H\cap N)\to hN$ of the quotient space $H/H\cap N$ to HN/N is a continuous bijection but not necessarily an open mapping. In particular, if N is a normal subgroup of the group HN, then f is a continuous homomorphism. In addition, if f is an open mapping, the quotient groups $H/H\cap N$ and HN/N are isomorphic as topological groups (**second isomorphism theorem**). For example, f is an open mapping (1) if N is compact or (2) if G is locally compact, HN and N are closed subgroups of G, and H is the union of a countable number of compact subsets. Let H be a subgroup of a topological group G and N be a normal subgroup of G such that $H\supset N$. Then the canonical mapping of the quotient space $(G/N)/(H/N)$ onto G/H is a homeomorphism. In particular, if H is also a normal subgroup, the quotient groups $(G/N)/(H/N)$ and G/H are isomorphic as topological groups (**third isomorphism theorem**).

K. The Projective Limit

Let $\{G_\alpha\}_{\alpha\in A}$ be a family of topological groups indexed by a †directed set A, and suppose that if $\alpha\leqslant\beta$, there exists a continuous homomorphism $f_{\alpha\beta}: G_\beta\to G_\alpha$ such that $f_{\alpha\gamma}=f_{\alpha\beta}\circ f_{\beta\gamma}$ if $\alpha\leqslant\beta\leqslant\gamma$. Then the collection $\{G_\alpha, f_{\alpha\beta}\}$ of the family $\{G_\alpha\}_{\alpha\in A}$ of topological groups together with the family $\{f_{\alpha\beta}\}$ of mappings is called a **projective system** of topological groups. Consider the direct product $\prod_{\alpha\in A} G_\alpha$ of topological groups $\{G_\alpha\}$, and denote by G the set of all elements $x=\{x_\alpha\}_{\alpha\in A}$ of $\prod G_\alpha$ that satisfy $x_\alpha=f_{\alpha\beta}(x_\beta)$ for $\alpha\leqslant\beta$. Then G is a subgroup of $\prod G_\alpha$. The topological group G obtained in this way is called the **projective limit** of the projective system $\{G_\alpha, f_{\alpha\beta}\}$ of topological groups and is denoted by $G=\varprojlim G_\alpha$. If each G_α is a T_2-topological (resp. complete) group, then G is also a T_2-topological (complete) group.

Now consider another projective system $\{G'_\alpha, f'_{\alpha\beta}\}$ of topological groups indexed by the same A, and consider continuous homomorphisms $u_\alpha: G_\alpha\to G'_\alpha$ such that $u_\alpha\circ f_{\alpha\beta}=f'_{\alpha\beta}\circ u_\beta$ for $\alpha\leqslant\beta$. Then there exists a unique continuous homomorphism u of $G=\varprojlim G_\alpha$ into $G'=\varprojlim G'_\alpha$ such that for any $\alpha\in A$, $u_\alpha\circ f_\alpha=f'_\alpha\circ u$ holds, where $f_\alpha(f'_\alpha)$ is the restriction to $G(G')$ of the projection of $\prod G_\alpha$ ($\prod G'_\alpha$) onto $G_\alpha(G'_\alpha)$. The homomorphism u is called the **projective limit** of the family $\{u_\alpha\}$ of continuous homomorphisms and is denoted by $u=\varprojlim u_\alpha$. Let G be a T_2-topological group, and let $\{H_\alpha\}_{\alpha\in A}$ be a decreasing sequence ($H_\alpha\supset H_\beta$ for $\alpha\leqslant\beta$) of closed normal subgroups of G. Consider the quotient group G/H_α, and let $f_{\alpha\beta}$ be the canonical mapping $gH_\beta\to gH_\alpha$ of G_β to G_α for $\alpha\leqslant\beta$. Then $\{G_\alpha, f_{\alpha\beta}\}$ is a projective system of topological groups. Let f_α be the projection of G onto $G_\alpha=G/H_\alpha$, and let $f=\varprojlim f_\alpha$. Now assume that any neighborhood of the identity of G contains some H_α and that some H_γ is complete. Then $f=\varprojlim f_\alpha$ is an isomorphism of G onto $\varprojlim G/H_\alpha$ as topological groups. (For a general discussion of the topological groups already discussed → [1,4].)

L. Locally Compact Groups

For the rest of this article, all topological groups under consideration are assumed to be T_2-topological groups. The identity component G_0 of a locally compact group G is the intersection of all open subgroups of G. In particular, any neighborhood of the identity of a totally disconnected locally compact group contains an open subgroup. A totally disconnected compact group is a projective limit of finite groups with discrete topology.

A T_1-topological space L is called a **local Lie group** if it satisfies the following six conditions: (i) There exist a nonempty subset M of $L\times L$ and a continuous mapping $\mu: M\to L$, called **multiplication** ($\mu(a,b)$ is written as ab). (ii) If (a,b), (ab,c), (b,c), (a,bc) are all in M, then $(ab)c=a(bc)$. (iii) L contains an element e, called the **identity**, such that $L\times\{e\}\subset M$ and $ae=a$ for all $a\in L$. (iv) There exists a nonempty open subset N of L and a continuous mapping $v: N\to L$ such that $av(a)=e$ for all $a\in N$. (v) There exist a neighborhood U of e in L and a homeomorphism f of U into a neighborhood V of the origin in the Euclidean space $\mathbf{R}^n$. (vi) Let D be the open subset of $V\times V$ defined by $D=\{(x,y)\in V\times V \mid (f^{-1}(x), f^{-1}(y))\in M, f^{-1}(x), f^{-1}(y)\in U\}$. Then the function $F: D\to V$ defined by $F(x,y)=f\mu(f^{-1}(x), f^{-1}(y))$ is of †class C^ω.

For any neighborhood U of the identity e of a connected locally compact group G, there exist a compact normal subgroup K and a subset L that is a local Lie group under the †induced topology and the group operations of G such that the product LK is a neighborhood of e contained in U. Furthermore, under $(l,k)\to lk$, LK is homeomorphic to the product space $L\times K$. Any compact subgroup of a connected locally compact group G is contained in a maximal compact subgroup, and maximal compact subgroups of G are †conjugate. For a maximal compact subgroup K of G, there exists a finite number of subgroups $H_1, \ldots, H_r$ of G, each of which is isomorphic to the additive group of real numbers such that $G=KH_1\ldots H_r$, and the mapping $(k,h_1,\ldots,h_r)\to kh_1\ldots h_r$ is a homeomorphism of the direct product $K\times H_1\times\ldots\times H_r$ onto G. Any locally compact group has a left-invariant positive

measure and a right-invariant positive measure, which are uniquely determined up to constant multiples (→ 225 Invariant Measures). Using these measures, the theory of harmonic analysis on the additive group **R** of real numbers can be extended to that on G (→ 69 Compact Groups; 192 Harmonic Analysis; 422 Topological Abelian Groups; 437 Unitary Representations).

M. Locally Euclidean Groups

Suppose that each point of a topological group G has a neighborhood homeomorphic to an open set of a given Euclidean space. Then G is called a **locally Euclidean group**. If the underlying topological space of a topological group has the structure of a †real analytic manifold such that the group operation $(x, y) \to xy^{-1}$ is a real analytic mapping, then G is called a †**Lie group**. A Lie group is a locally Euclidean group.

N. Hilbert's Fifth Problem

Hilbert's fifth problem asks if every locally Euclidean group is a Lie group (→ 196 Hilbert). This problem was solved affirmatively in 1952; it was proved that any †locally connected finite-dimensional locally compact group is a Lie group (D. Montgomery and L. Zippin [3]). In connection with this, the relation between Lie groups and general locally compact groups has been studied, and the following results have been obtained: A necessary and sufficient condition for a locally compact group to be a Lie group is that there exist a neighborhood of the identity e that does not contain any subgroup (or any normal subgroup) other than $\{e\}$. A locally compact group has an open subgroup that is the projective limit of Lie groups. Hilbert's fifth problem is closely related to the following problem: Find the conditions for a †topological transformation group operating †effectively on a manifold to be a Lie group (→ 431 Transformation Groups).

O. Covering Groups

Let $\mathfrak{G}$ be the collection of all †arcwise connected and †locally arcwise connected T_2-topological groups. Suppose that $G^* \in \mathfrak{G}$ is a †covering space of $G \in \mathfrak{G}$ and the †covering mapping $f: G^* \to G$ is an open continuous homomorphism, with G^* and G considered as topological groups. Then G^* (or, more precisely, (G^*, f)) is called a **covering group** of G. Then the kernel $f^{-1}(e) = D$ of f is a discrete subgroup contained in the center of G^*, and G^*/D and G, considered as topological groups, are isomorphic to each other. Let $\pi_1(G)$ be the †fundamental group of G. The natural homomorphism $f^*: \pi_1(G^*) \to \pi_1(G)$ induced by f is an injective homomorphism, and if we identify $\pi_1(G^*)$ with the subgroup $f^*(\pi_1(G^*))$ of $\pi_1(G)$, we have $D \cong \pi_1(G)/\pi_1(G^*)$. Conversely, if D is any discrete subgroup contained in the center of $G^* \in \mathfrak{G}$, then G^* is a covering group of $G = G^*/D$. For any covering space (G^*, f) of $G \in \mathfrak{G}$, a multiplication law can be introduced in G^* so that G^* is a topological group belonging to $\mathfrak{G}$ and (G^*, f) is a covering group of G. In particular, any $G \in \mathfrak{G}$ has a †simply connected covering group $(\tilde{G}, \varphi)$. Then for any covering group (G^*, f) of G, there exists a homomorphism $f^*: \tilde{G} \to G^*$, and $(\tilde{G}, f^*)$ is a covering group of G^*. Furthermore, $\varphi = f \circ f^*$. Hence, in particular, any simply connected covering group of G is isomorphic to $\tilde{G}$, with G and $\tilde{G}$ considered as topological groups. This simply connected covering group $(\tilde{G}, \varphi)$ is called the **universal covering group**.

Let G and G' be topological groups, and let e and e' be their identities. A homeomorphism f of a neighborhood U of e onto a neighborhood U' of e' is called a **local isomorphism** of G to G' if it satisfies the following two conditions: (i) If a, b, ab are all contained in U, then $f(ab) = f(a)f(b)$. (ii) Let $f^{-1} = g$, then if $a', b', a'b' \in U'$, $g(a'b') = g(a')g(b')$ holds. If there exists a local isomorphism of G to G', we say that G and G' are **locally isomorphic**. If G^* is a covering group of G, then G^* and G are locally isomorphic. For two topological groups G and G' to be locally isomorphic it is necessary and sufficient that the universal covering groups of G and G' be isomorphic. For two connected Lie groups to be locally isomorphic it is necessary and sufficient that their †Lie algebras be isomorphic.

Let f be a mapping of a neighborhood U of the identity of a topological group G into a group H such that if a, b, ab are all contained in U, then $f(ab) = f(a)f(b)$. Then f is called a **local homomorphism** of G into H and U is called its **domain**. A local homomorphism of a simply connected group $G \in \mathfrak{G}$ into a group H can be extended to a homomorphism of G into H if the domain is connected [2, 4].

P. Topological Rings and Fields

If a ring R has the structure of a topological group such that $(x, y) \to x + y$ (sum) and $(x, y) \to xy$ (product) are both continuous mappings of $R \times R$ into R, then R is called a **topological ring**. If a topological ring K is a field (not necessarily commutative) such that $x \to x^{-1}$

(inverse element) is a continuous mapping of $K^* = K - \{0\}$ into K^*, then K is called a **topological field**. Let us assume that K is a topological field that is a locally compact Hausdorff space and is not discrete. If K is connected, then K is a †division algebra of finite rank over the field $\mathbf{R}$ of real numbers; hence it is isomorphic to the field $\mathbf{R}$ of real numbers, the field $\mathbf{C}$ of complex numbers, or the †quaternion field $\mathbf{H}$. If K is not connected, then K is totally disconnected and is isomorphic to a division algebra of finite rank over the †p-adic number field $\mathbf{Q}_p$ or a division algebra of finite rank over the †formal power series field with coefficients in a finite field [4].

For various important classes of topological groups → 69 Compact Groups; 249 Lie Groups; 422 Topological Abelian Groups; 424 Topological Linear Spaces.

References

[1] N. Bourbaki, Eléments de mathématique, Topologie générale, ch. 3, Actualités Sci. Ind., 1143c, Hermann, third edition, 1960; English translation, General topology, pt. 1, Addison-Wesley, 1966.
[2] C. Chevalley, Theory of Lie groups I, Princeton Univ. Press, 1946.
[3] D. Montgomery and L. Zippin, Topological transformation groups, Interscience, 1955.
[4] L. S. Pontryagin, Topological groups, first edition, Princeton Univ. Press, 1939, second edition, Gordon & Breach, 1966. (Second English edition translated from the second Russian edition, 1954.)

424 (XII.5) Topological Linear Spaces

A. Definition

A †linear space E over the real or complex number field K is said to be a **topological linear space**, **topological vector space**, or **linear topological space** if E is a †topological space and the basic operations $x+y$ and αx ($x, y \in E$, $\alpha \in K$) in the linear space are continuous as mappings of $E \times E$ and $K \times E$, respectively, into E. The coefficient field K may be a general †topological field, although it is usually assumed to be the real number field $\mathbf{R}$ or the complex number field $\mathbf{C}$, and accordingly E is called a **real topological linear space** or a **complex topological linear space**. Topological linear spaces are generalizations of †normed linear spaces and play an important role in the study of †function spaces, such as the †space of distributions, that are not †Banach spaces.

Each topological linear space E is equipped with a †uniform topology in which translations of the neighborhoods of zero form a †uniform family of neighborhoods, and the addition $x+y$ and the multiplication αx by a scalar α are uniformly continuous relative to this uniform topology. In particular, if for each $x \neq 0$ there is a neighborhood of the origin that does not contain x, then E satisfies the †separation axiom T_1 and hence is a †completely regular space. The †completion $\hat{E}$ of E is also a topological linear space.

We assume in this article that K is the real or complex number field and E is a topological linear space over K satisfying the axiom of T_1-spaces. Then E is finite-dimensional if and only if E has a †totally bounded neighborhood of zero. The topology of E is †metrizable if and only if it satisfies the †first countability axiom.

B. Linear Functional

A K-valued function $f(x)$ on E is said to be a **linear functional** if it satisfies (i) $f(x+y)=f(x)+f(y)$ and (ii) $f(\alpha x)=\alpha f(x)$. A linear functional that is continuous relative to the topologies of E and K is said to be a continuous linear functional. (Sometimes continuous linear functionals are simply called linear functionals, while abstract linear functionals are called **algebraic linear functionals**.) The following three statements are equivalent for linear functionals $f(x)$: (i) $f(x)$ is continuous; (ii) the half-space $\{x \in E \mid \mathrm{Re} f(x) > 0\}$ is open; (iii) the hyperplane $\{x \in E \mid f(x)=0\}$ is closed.

C. The Hahn-Banach Theorem

A linear functional $f(x)$ defined on a linear subspace F of E can be extended to a continuous linear functional on E if and only if there exists an open †convex neighborhood V of the origin in E that is disjoint with $\{x \in F \mid f(x)=1\}$. Furthermore, if $f(x)$ can be extended, at least one extension $f(x)$ never takes the value 1 on V (**Hahn-Banach theorem**).

D. Dual Spaces

The set E' of all continuous linear functionals on E is called the **dual space** of E. It is often denoted by E^* and is also called the **conjugate space** or **adjoint space**. It forms a linear space when $f+g$ and αf ($f, g \in E'$, $\alpha \in K$) are defined by $(f+g)(x)=f(x)+g(x)$ and $(\alpha f)(x)=\alpha(f(x))$ for $x \in E$.

E. Locally Convex Spaces

A topological linear space is said to be **locally convex** if it has a family of convex sets as a †base of the neighborhood system of 0. It follows from the Hahn-Banach theorem that for each $x \neq 0$ in a locally convex space E there is a continuous linear functional f such that $f(x) \neq 0$. A subset M of E is said to be **circled** if M contains $\alpha M = \{\alpha x \mid x \in M\}$ whenever $|\alpha| \leqslant 1$. A set that is both circled and convex is called **absolutely convex**. In a locally convex space, a family of absolutely convex and closed sets can be chosen as a base of the neighborhood system of the origin. Let A and B be subsets of E. A is said to **absorb** B if there is an $\alpha > 0$ such that $\alpha A \supset B$. A set V that absorbs every point $x \in E$ is called **absorbing**. Neighborhoods of 0 are absorbing.

F. Seminorms

A real-valued function $p(x)$ on E is said to be a **seminorm** (or **pseudonorm**) if it satisfies (i) $0 \leqslant p(x) < +\infty$ $(x \in E)$; (ii) $p(x+y) \leqslant p(x) + p(y)$; and (iii) $p(\alpha x) = |\alpha| p(x)$. The relation $V = \{x \mid p(x) \leqslant 1\}$ gives a one-to-one correspondence between seminorms $p(x)$ and absolutely convex absorbing sets V whose intersection with any line through the origin is closed. In terms of seminorms, the Hahn-Banach theorem states: Let E be a linear space on which a seminorm $p(x)$ is given. If a linear functional $f(x)$ defined on a linear subspace F of E satisfies $|f(x)| \leqslant p(x)$ on F, then $f(x)$ can be extended to the whole space E in such a way that the inequality holds on E.

The topology of a locally convex space is determined by the family of continuous seminorms on it. Conversely, if there is a family of seminorms $\{p_\lambda(x)\}$ $(\lambda \in \Lambda)$ on a linear space E over K that satisfies (iv) $p_\lambda(x) = 0$ for all λ implies $x = 0$, then there exists on E the weakest locally convex topology that renders the seminorms continuous. This topology is called the locally convex topology determined by $\{p_\lambda(x)\}$.

We assume that E is a locally convex space whose topology is determined by the family of seminorms $\{p_\lambda(x)\}$ $(\lambda \in \Lambda)$. Then a †net x_ν of E converges to x if and only if $p_\lambda(x_\nu - x) \rightarrow 0$ for all $\lambda \in \Lambda$. If F is a locally convex space whose topology is determined by the family of seminorms $\{q_\mu(y)\}$, then a necessary and sufficient condition for a linear mapping $u: E \rightarrow F$ to be continuous is that for every $q_\mu(y)$ there exist a finite number of $\lambda_1, \ldots, \lambda_n \in \Lambda$ and a constant C such that $q_\mu(u(x)) \leqslant C(p_{\lambda_1}(x) + \ldots + p_{\lambda_n}(x))$ $(x \in E)$.

A set is said to be **bounded** if it is absorbed by every neighborhood of zero. When the topology of E is determined by the family $\{p_\lambda(x)\}$ of seminorms a set B is bounded if and only if every p_λ is bounded on B. Totally bounded sets are bounded. The unit ball in a normed space is bounded. Conversely, a locally convex space is normable if it has a bounded neighborhood of 0. A locally convex space is called **quasicomplete** if every bounded closed set is complete. Since Cauchy sequences $\{x_n\}$ are totally bounded, all Cauchy sequences converge in a quasicomplete space (i.e., the space is sequentially complete).

G. Pairing of Linear Spaces

Let E and F be linear spaces over the same field K. A K-valued function $B(x, y)$ $(x \in E, y \in F)$ on $E \times F$ is called a **bilinear functional** or **bilinear form** if for each fixed $y \in F$ (resp. $x \in E$), it is a linear functional of x (resp. y). When a bilinear functional $\langle x, y \rangle$ on $E \times F$ is given so that $\langle x, y \rangle = 0$ for all $y \in F$ (all $x \in E$) implies $x = 0$ $(y = 0)$, then E and F are said to form a (separated) **pairing** relative to the **inner product** $\langle x, y \rangle$. A locally convex space E and its dual space E' form a pairing relative to the natural inner product $\langle x, x' \rangle = x'(x)$ $(x \in E, x' \in E')$.

H. Weak Topologies

When E and F form a pairing relative to an inner product $\langle x, y \rangle$, the locally convex topology on E determined by the family of seminorms $\{|\langle x, y \rangle| \mid y \in F\}$ is called the **weak topology (relative to the pairing** $\langle E, F \rangle$**)** and is denoted by $\sigma(E, F)$. A net x_ν in E is said to **converge weakly** if it converges in the weak topology. When E and E' are a locally convex space and its dual space, $\sigma(E, E')$ is called the **weak topology** of E, and $\sigma(E', E)$ the **weak* topology** of E'. The weak topology on a locally convex space E is weaker than the original topology on E. Consequently, a weakly closed set is closed. If the set is convex, the converse holds, and hence a convex closed set is weakly closed. Also, boundedness is preserved if we replace the original topology by the weak topology. Thus a weakly bounded set is bounded.

Let E and F form a pairing relative to $\langle x, y \rangle$, and let A be a subset of E. Then the set A° of points $y \in F$ satisfying $\mathrm{Re}\langle x, y \rangle \geqslant -1$ for all $x \in A$ is called the **polar** of A (relative to the pairing). If A is absolutely convex, A° is also absolutely convex and is the set of points y such that $|\langle x, y \rangle| \leqslant 1$ for all $x \in A$. If A is a convex set containing zero, its (weak) closure is equal to the **bipolar** $A^{\circ\circ} = (A^\circ)^\circ$ (**bipolar theorem**). In general, let A be a subset of a

topological linear space E. We call the smallest closed convex set containing A the **closed convex hull** of A. If E is locally convex, the bipolar $A^{\circ\circ}$ relative to E' coincides with the closed convex hull of $A \cup \{0\}$.

A subset B of the dual space E' is †equicontinuous on E if and only if it is contained in the polar V° of a neighborhood V of 0 in E. Also, V° is weak*- compact in E' (**Banach-Alaoglu theorem**).

I. Barreled Spaces and Bornological Spaces

An absorbing absolutely convex closed set in a locally convex space E is called a **barrel**. In a sequentially complete space (hence in a quasi-complete space also), a barrel absorbs every bounded set. A locally convex space is said to be **barreled** if each barrel is a neighborhood of 0. A locally convex space is said to be **quasi-barreled** (or **evaluable**) if each barrel that absorbs every bounded set is a neighborhood of 0. Furthermore, a locally convex space is said to be **bornological** if each absolutely convex set that absorbs every bounded set is a neighborhood of 0. Bornological spaces are quasi-barreled. However, they are not necessarily barreled. Furthermore, barreled spaces are not necessarily bornological. A metrizable locally convex space, i.e., a space whose topology is determined by a countable number of seminorms, is bornological. A complete metrizable locally convex space is called a **locally convex Fréchet space** (**(F)-space** or simply **Fréchet space**). To distinguish it from Fréchet space as in 37 Banach Spaces, it is sometimes called a **Fréchet space in the sense of Bourbaki**. (F)-spaces are bornological and barreled.

A continuous linear mapping $u: E \to F$ of one locally convex space into another maps each bounded set of E to a bounded set in F. Conversely, if E is bornological, then each linear mapping that maps every bounded sequence to a bounded set is continuous.

J. The Banach-Steinhaus Theorem

In the dual space of a barreled space E, each (weak*-)bounded set is equicontinuous. Thus if a sequence of continuous linear mappings u_n of E into a locally convex space F converges at each point of E, then u_n converges uniformly on each totally bounded set of E, and the limit linear mapping is continuous (**Banach-Steinhaus theorem**).

K. The S-Topology

Let E and F be paired linear spaces relative to the inner product $\langle x, y\rangle$. When a family S of (weakly) bounded sets of F generates a dense subspace of F, the family of seminorms $\{\sup_{y\in B}|\langle x, y\rangle| \mid B\in S\}$ determines a locally convex topology on E. This is called the **S-topology** or **topology of uniform convergence on members of** S, because $x_\nu \to x$ in the S-topology is equivalent to the uniform convergence of $\langle x_\nu, y\rangle \to \langle x, y\rangle$ on each $B \in S$. The space E with the S-topology is denoted by E_S. The weak topology is the same as the topology of pointwise convergence. The S-topology in which S is the family of all bounded sets in F is called the **strong topology** and is denoted by $\beta(E, F)$. The dual space E' of a locally convex space E is usually regarded as a locally convex space with the strong topology $\beta(E', E)$. It is called the **strong dual space**. The topology of a locally convex space E is that of uniform convergence on equicontinuous sets of E'. The topology of a barreled space E coincides with the strong topology $\beta(E, E')$.

L. Grothendieck's Criterion of Completeness

Let E and F be paired spaces, and let S be a family of absolutely convex bounded sets of F such that: (i) the sets of S generate F; (ii) if B_1, $B_2 \in S$, then there is a $B_3 \in S$ such that $B_3 \supset B_1$ and $B_3 \supset B_2$. Then E_S is complete if and only if each algebraic linear functional $f(y)$ on F that is weakly continuous on every $B \in S$ is expressed as $f(y) = \langle x, y\rangle$ for some $x \in E$. When E_S is not complete, the space of all linear functionals satisfying this condition gives the completion $\hat{E}_S$ of E_S.

M. Mackey's Theorem

Let E, F, and S satisfy the same conditions as in Section L. Then the dual space of E_S is equal to the union of the weak completions of λB, where $\lambda > 0$ and $B \in S$ (**Mackey's theorem**).

N. The Mackey Topology

When E and F form a pairing, the topology on E of uniform convergence on convex weakly compact sets of F is called the **Mackey topology** and is denoted by $\tau(E, F)$. The dual space of E endowed with a locally convex topology T coincides with F if and only if T is stronger than the weak topology $\sigma(E, F)$ and weaker than the Mackey topology $\tau(E, F)$ (**Mackey-Arens theorem**). A locally convex space is said to be a **Mackey space** if the topology is equal to the Mackey topology $\tau(E, E')$. Every quasi-barreled space is a Mackey space.

O. Reflexivity

Let E be a locally convex space. The dual space E'' of the dual space E' equipped with the strong topology contains the original space E. We call E **semireflexive** if $E''=E$, and **reflexive** if in addition the topology of E coincides with the strong topology $\beta(E,E')$. E is semireflexive if and only if every bounded weakly closed set of E is weakly compact. E is reflexive if and only if E is semireflexive and (quasi)barreled.

A barreled space in which every bounded closed set is compact is called a **Montel space** or **(M)-space**. (M)-spaces are reflexive, and their strong dual spaces are also (M)-spaces.

Many of the function spaces that appear in applications are (F)-spaces or their dual spaces. For these spaces detailed consequences of the countability axiom are known [7, 8]. A convex set C in the dual space E' of an (F)-space E is weak*-closed if and only if for every neighborhood V of 0 in E, $C\cap V^{\circ}$ is weak*-closed (**Kreĭn-Shmul'yan theorem**). The strong dual space E' of an (F)-space E is (quasi)barreled if and only if it is bornological. In particular, the dual space of a reflexive (F)-space is bornological.

P. (DF)-Spaces

A locally convex space is called a **(DF)-space** if it satisfies: (i) There is a countable base of bounded sets (i.e., every bounded set is included in one of them); (ii) if the intersection V of a countable number of absolutely convex closed neighborhoods of 0 absorbs every bounded set, then V is also a neighborhood of 0. The dual space of an (F)-space is a (DF)-space, and the dual space of a (DF)-space is an (F)-space. A linear mapping of a (DF)-space E into a locally convex space F is continuous if and only if its restriction to every bounded set of E is continuous. A quasicomplete (DF)-space is complete.

Q. Bilinear Mappings

A bilinear mapping $b(x,y)$ on locally convex spaces E and F $(x\in E, y\in F)$ to a locally convex space G is said to be **separately continuous** if for each fixed $y\in F$ $(x\in E)$ it is continuous as a function of x (y). The linear mappings obtained from $b(x,y)$ by fixing x (y) are denoted by $b_x(y)$ $(b_y(x))$. We call $b(x,y)$ **hypocontinuous** if for each bounded set B of E and B' of F, $\{b_x(y)\mid x\in B\}$ and $\{b_y(x)\mid y\in B'\}$ are equicontinuous. A continuous bilinear mapping is hypocontinuous. However, the converse is not always true. A separately continuous bilinear mapping is not necessarily hypocontinuous. If both E and F are barreled, however, then every separately continuous mapping is hypocontinuous. If E is an (F)-space and F is metrizable, then every separately continuous bilinear mapping is continuous. Similarly, if both E and F are (DF)-spaces, then every hypocontinuous bilinear mapping is continuous.

R. Tensor Products

It is possible to introduce many topologies in the †tensor product $E\otimes F$ of locally convex spaces E and F. The **projective topology** (or **topology** π) is defined to be the strongest topology such that the natural bilinear mapping $E\times F\to E\otimes F$ is continuous. The dual space of $E\otimes_\pi F$ is identified with the space $B(E,F)$ of all continuous bilinear functionals on $E\times F$. The completion of $E\otimes_\pi F$ is denoted by $E\hat{\otimes} F$. The **topology of biequicontinuous convergence** (or **topology** ε) is defined to be the topology of uniform convergence on sets $V^{\circ}\times U^{\circ}$, where V and U are neighborhoods of 0 in E and F, respectively, considering the elements of $E\otimes F$ as linear functionals on $E'\otimes F'$ by the natural pairing of $E\otimes F$ and $E'\otimes F'$. The completion of $E\otimes_\varepsilon F$ is denoted by $E\hat{\hat{\otimes}} F$. The dual space of $E\otimes_\varepsilon F$ coincides with the subspace $J(E,F)$ of $B(E,F)$ composed of the union of the absolute convex hulls of the products $V^{\circ}\otimes U^{\circ}$ of equicontinuous sets. The elements of $J(E,F)$ are called **integral bilinear functionals**.

Closely related to $E\hat{\hat{\otimes}} F$ is L. Schwartz's ε **tensor product** $E\varepsilon F$ [12]. (They coincide if E and F are complete and if E or F has the †approximation property.) $E\varepsilon F$ can be regarded as (i) a space of vector-valued functions if E is a space of functions and F is an abstract locally convex space, especially a space of functions of two variables if E and F are, respectively, spaces of functions of one variable, and (ii) a space of operators $G\to F$ if E is the dual space G' of a locally convex space G.

S. Nuclear Spaces

Let E be a locally convex space, V be an absolutely convex closed neighborhood of the origin, and $p(x)$ be the seminorm corresponding to V. Then we denote by E_V the normed space with norm $p(x)$ obtained from E by identifying the two elements x and y with $p(x-y)=0$. If $U\subset V$, then a natural linear mapping $\varphi_{U,V}:E_U\to E_V$ is defined.

A locally convex space E is said to be a

nuclear space (resp. **Schwartz space** or simply **(S)-space**) if for each absolutely convex closed neighborhood V of 0 there is another U such that $\varphi_{U,V}$ is a †nuclear operator (resp. †compact operator) as an operator of E_U into the completion of E_V. A nuclear space or (S)-space is an (M)-space if it is quasicomplete and quasibarreled. A locally convex space E is a nuclear space if and only if the topologies π and ε coincide on the tensor product $E \otimes F$ with any locally convex space F. Accordingly, it follows that $B(E, F) = J(E, F)$. This can be regarded as a generalization of Schwartz's **kernel theorem**, which says that every separately continuous bilinear functional on $\mathcal{D}_x \times \mathcal{D}_y$ is represented by an integral with kernel in $\mathcal{D}'_{xy}$. The theory of topological tensor products and nuclear spaces is due to Grothendieck [9].

A locally convex space E is a nuclear (F)-space if and only if E is isomorphic to a closed subspace of $C^\infty(-\infty, \infty)$ (T. Kōmura and Y. Kōmura, 1966). An example of a nuclear (F)-space without basis is known (B. S. Mityagin and N. M. Zobin, 1974).

T. Gel'fand Triplet

Let H and L be Hilbert spaces. If L is a dense subspace of H and the injection $L \to H$ is a †Hilbert-Schmidt operator, then $H = H'$ is regarded as a dense subspace of L' and the injection $H' \to L'$ is a Hilbert-Schmidt operator. In this case, (L, H, L') is called a **Gel'fand triplet** (or a **rigged Hilbert space**).

A subset of H is called a cylindrical set if it is expressed in the form $P_F^{-1}(B)$ by the orthogonal projection P_F onto a finite-dimensional subspace F and a Borel subset B of F. If a finitely additive positive measure μ with $\|\mu\|_1 = 1$ defined on the cylindrical sets of H satisfies (i) μ is countably additive on cylindrical sets for a fixed F and (ii) for any $\varepsilon > 0$ there exists a $\delta > 0$ such that $\|x\| < \delta$ implies $\mu\{y \in H \mid |\langle x, y\rangle| \geqslant 1\} < \varepsilon$, then μ is the restriction of a countably additive measure $\tilde{\mu}$ defined on the Borel subsets of L' (**Minlos's theorem**, 1959).

Let T be a self-adjoint operator in H. Then T has a natural extension $\tilde{T}$ in L' and almost every continuous spectrum λ of T has an associated eigenvector x_λ in L': $\tilde{T}x_\lambda = \lambda x_\lambda$, $x_\lambda \in L'$.

U. The Extreme Point Theorem

Let A be a subset of a linear space E. A point $x \in A$ is said to be an **extreme point** if x is an extreme point of any real segment containing x and contained in A. If A is a compact convex subset of a locally convex space E, A is the convex closed hull of (i.e., smallest convex closed set containing) the set of its extreme points (**Kreĭn-Milman theorem**). In applications it is important to know whether every point of A is represented uniquely as an integral of extreme points. For a metrizable convex compact subset A of a locally convex space E, the following two conditions are equivalent (**Choquet's theorem**): (i) A is a **simplex**, i.e., if we put $\tilde{A} = \{(\lambda x, \lambda) \mid x \in A, \lambda > 0\} \subset E \times \mathbf{R}^1$, the vector space $\tilde{A} - \tilde{A}$ becomes a †lattice with positive cone $\tilde{A}$; (ii) for any $x \in A$ there exists a unique positive measure μ on A with $\|\mu\|_1 = 1$ such that $l(x) = \int_A l(y) d\mu(y)$ $(l \in E')$ and the support of μ is contained in the set of extreme points of A.

V. Weakly Compact Set

A subset of a quasicomplete locally convex space is relatively weakly compact if and only if every sequence in the set has a weak accumulation point (**Eberlein's theorem**). If E is a metrizable locally convex space, every weakly compact set of E is weakly sequentially compact (**Shmul'yan's theorem**). If E is a quasicomplete locally convex space, the convex closed hull of any weakly compact subset is weakly compact (**Kreĭn's theorem**). If E is not quasicomplete, this is not necessarily true.

W. Permanence

Each subspace, quotient space, direct product, direct sum, projective limit, and inductive limit (of a family) of locally convex spaces has a unique natural locally convex topology. These spaces, except for quotient spaces and inductive limits, are separated, and a quotient space E/A is separated if and only if the subspace A is closed. The limit of a sequence $E_1 \subset E_2 \subset \ldots$ is said to be a **strictly inductive limit** if E_n has the induced topology as a subspace of E_{n+1}. If E is a strictly inductive limit of a sequence E_n such that E_n is closed in E_{n+1} or if E is the inductive limit of a sequence $E_1 \subset E_2 \subset \ldots$ such that the mapping $E_n \to E_{n+1}$ maps a neighborhood of 0 to a relatively weakly compact set, then E is separated and each bounded set of E is the image of a bounded set in some E_n. If $E = \bigcup E_n$ is the strictly inductive limit of the sequence $\{E_n\}$, then the topology of E_n coincides with the relative topology of $E_n \subset E$. The strictly inductive limit of a sequence of (F)-spaces is called an **(LF)-space**.

Any complete locally convex space (resp. any locally convex space) is (resp. a dense linear subspace of) the projective limit of Banach spaces. Every (F)-space E is the projective limit of a sequence of Banach spaces $E_1 \leftarrow E_2 \leftarrow \ldots$. In particular, E is said to be a **count-**

ably normed space if the mappings $E \to E_n$ are one-to-one and $\|x\|_n \leqslant \|x\|_{n+1}$ for all $x \in E$ with E considered as a subspace E_n. We call E a **countably Hilbertian space** if, in particular, the E_n are †Hilbert spaces. An (F)-space with at least one continuous norm is a nuclear space if and only if it is a countably Hilbertian space such that the mappings $E_{n+1} \to E_n$ are Hilbert-Schmidt operators or nuclear operators.

A locally convex space is bornological if and only if it is the inductive limit of normed spaces. A locally convex space is said to be **ultrabornological** if it is the inductive limit of Banach spaces, or in particular, if it is quasi-complete and bornological.

Properties of spaces, such as being complete, quasicomplete, semireflexive, or having every bounded closed set compact, are inherited by closed subspaces, direct products, projective limits, direct sums, and strictly inductive limits formed from the original spaces, and properties of spaces, such as being Mackey, quasibarreled, barreled, and bornological, are inherited by quotient spaces, direct sums, inductive limits, and direct products formed from the spaces. (For direct products of high power of bornological spaces, unsolved problems still exist concerning the inheritance of properties.) Quotient spaces of (F)-spaces are (F)-spaces, but quotient spaces of general complete spaces are not necessarily complete. There are examples of a Montel (F)-space whose quotient space is not reflexive and a Montel (DF)-space whose closed subspace is neither a Mackey space nor a (DF)-space. The property of being a Schwartz space or a nuclear space is inherited by the completions, subspaces, quotient spaces of closed subspaces, direct products, projective limits, direct sums of countable families, and inductive limits of countable families formed from such spaces. Tensor products of nuclear spaces are nuclear spaces. Y. Kōmura gave an example of a noncomplete space that is quasicomplete, bornological, and nuclear (and hence a Montel space).

X. The Open Mapping Theorem and the Closed Graph Theorem

Let E and F be topological linear spaces. The statement that every continuous linear mapping of E onto F is open is called the **open mapping theorem** (or **homomorphism theorem**), and the statement that every linear mapping of F into E is continuous if its graph is closed in $F \times E$ is called the **closed graph theorem**. These theorems hold if both E and F are complete and metrizable (S. Banach).

A locally convex space is said to be **B-complete** (or **fully complete**) if a subspace C of E' is weak*-closed whenever $C \cap V^{\circ}$ is weak*-closed for every neighborhood V of 0 in E. (F)-spaces and the dual spaces of reflexive (F)-spaces are B-complete. B-complete spaces are complete, and closed subspaces and quotient spaces by closed subspaces of B-complete spaces are B-complete. If E is B-complete and F is barreled, then the open mapping theorem and the closed graph theorem hold (V. Pták).

Both theorems hold also if F is ultrabornological and E is a locally convex space obtained from a family of (F)-spaces after a finite number of operations of taking closed subspaces, quotient spaces by closed spaces, direct products of countable families, projective limits of countable families, direct sums of countable families, and inductive limits of countable families. This was conjectured by Grothendieck and proved by W. Słowikowski (1961) and D. A. Raikov. Later, L. Schwartz, A. Martineau, M. De Wilde, W. Robertson, and M. Nakamura simplified the proof and enlarged the class of spaces E [15].

(LF)-spaces, the dual spaces of Schwartz (F)-spaces, and the space $\mathscr{D}'$ of distributions are examples of spaces E described in the previous paragraph.

Y. Nonlocally Convex Spaces

The space L_p for $0<p<1$ shows that nonlocally convex spaces are meaningful in functional analysis. Recently, the Banach-Steinhaus theorem, closed graph theorems, etc. have been investigated for nonlocally convex topological linear spaces [13].

Z. Diagram of Topological Linear Spaces

The spaces in Fig. 1 are all locally convex spaces over the real number field or the complex number field and satisfy the separation axiom T_1. The notation $A \to B$ means that spaces with property A have property B. Main properties of dual spaces are listed in Table 1.

References

[1] N. Bourbaki, Eléments de mathématique, Espaces vectoriels topologiques, Actualités Sci. Ind., Hermann, 1189a, 1966; 1229b, 1967; 1230a, 1955.
[2] A. Grothendieck, Espaces vectoriels topologiques, Lecture notes, São Paulo, 1954.
[3] G. Köthe, Topologische lineare Räume I, Springer, second edition, 1966; English translation, Topological vector spaces I, II, Springer, 1969, 1979.

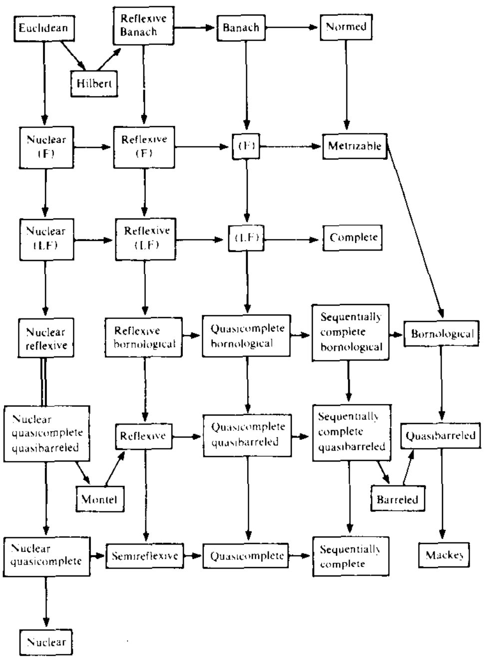


Fig. 1
Topological linear spaces.

Table 1

E	E'
Semireflexive	Barreled
Reflexive	Reflexive
Quasibarreled	Quasicomplete
Bornological	Complete
Reflexive, (F)	Bornological
(F)	(DF)
(DF)	(F)
(M)	(M)
Nuclear, (LF) or (DF)	Nuclear, reflexive
Complete, (S)	Ultrabornological

[4] I. M. Gel'fand et al., Generalized functions. II, Functions and generalized function spaces (with G. E. Shilov), Academic Press, 1968; IV, Applications of harmonic analysis (with N. Ya. Vilenkin), Academic, 1964. (Original in Russian, 1958–1966.)
[5] A. Pietsch, Nukleare lokalkonvexe Räume, Berlin, 1965.
[6] A. P. Robertson and W. Robertson, Topological vector spaces, Cambridge Univ. Press, 1964.
[7] J. Dieudonné and L. Schwartz, La dualité dans les espaces ($\mathfrak{F}$) et ($\mathfrak{L}\mathfrak{F}$), Ann. Inst. Fourier, 1 (1949), 61–101.
[8] A. Grothendieck, Sur les espaces (F) et (DF), Summa Brasil. Math., 3 (1954), 57–123.
[9] A. Grothendieck, Produits tensoriels topologiques et espaces nucléaires, Mem. Amer. Math. Soc., 1955.
[10] G. Choquet, Le théorème de représentation intégrale dans les ensembles convexes compacts, Ann. Inst. Fourier, 10 (1960), 333–344.
[11] J. L. Kelley, I. Namioka, et al., Linear topological spaces, Van Nostrand, 1963.
[12] L. Schwartz, Théorie des distributions à valeurs vectorielles I, Ann. Inst. Fourier, 7 (1957), 1–141.
[13] N. Adasch, B. Ernst, and D. Keim, Topological vector spaces: The theory without convexity conditions, Lecture notes in math. 639, Springer, 1978.
[14] F. Treves, Topological vector spaces, distributions and kernels, Academic Press, 1967.
[15] M. De Wilde, Closed graph theorems and webbed spaces, Pitman, 1978.

425 (II.16) Topological Spaces

A. Introduction

Convergence and continuity, as well as the algebraic operations on real numbers, are fundamental notions in analysis. In an abstract space too, it is possible to provide an additional structure so that convergence and continuity can be defined and a theory analogous to classical analysis can be developed. Such a structure is called a **topological structure** (for a precise definition, → Section B). There are several ways of giving a topology to a space. One method is to axiomatize the notion of convergence (M. Fréchet [1], 1906; → 87 Convergence). However, defining a topology in terms of either a neighborhood system (due to F. Hausdorff [3], 1914), a closure operation (due to C. Kuratowski, *Fund. Math.*, 3 (1922)), or a family of open sets is more common.

B. Definition of a Topology

Let X be a set. A **neighborhood system** for X is a function $\mathfrak{U}$ that assigns to each point x of X, a family $\mathfrak{U}(x)$ of subsets of X subject to the following axioms (**U**):
(1) $x \in U$ for each U in $\mathfrak{U}(x)$.
(2) If $U_1, U_2 \in \mathfrak{U}(x)$, then $U_1 \cap U_2 \in \mathfrak{U}(x)$.
(3) If $U \in \mathfrak{U}(x)$ and $U \subset V$, then $V \in \mathfrak{U}(x)$.
(4) For each U in $\mathfrak{U}(x)$, there is a member W of $\mathfrak{U}(x)$ such that $U \in \mathfrak{U}(y)$ for each y in W.

A **system of open sets** for a set X is a family $\mathfrak{O}$ of subsets of X satisfying the following axioms (**O**):

(1) $X, \varnothing \in \mathfrak{O}$.
(2) If $O_1, O_2 \in \mathfrak{O}$, then $O_1 \cap O_2 \in \mathfrak{O}$.
(3) If $O_\lambda \in \mathfrak{O}$ $(\lambda \in \Lambda)$, then $\bigcup_{\lambda \in \Lambda} O_\lambda \in \mathfrak{O}$.

A **system of closed sets** for a space X is a family $\mathfrak{F}$ of subsets of X satisfying the following axioms **(F)**:
(1) $X, \varnothing \in \mathfrak{F}$.
(2) If $F_1, F_2 \in \mathfrak{F}$, then $F_1 \cup F_2 \in \mathfrak{F}$.
(3) If $F_\lambda \in \mathfrak{F}$ $(\lambda \in \Lambda)$, then $\bigcap_{\lambda \in \Lambda} F_\lambda \in \mathfrak{F}$.

A **closure operator** for a space X is a function that assigns to each subset A of X, a subset A^a of X satisfying the following axioms **(C)**:
(1) $\varnothing^a = \varnothing$.
(2) $(A \cup B)^a = A^a \cup B^a$.
(3) $A \subset A^a$.
(4) $A^a = A^{aa}$.

An **interior operator** for a space X is a function that assigns to each subset A of X a subset A^i of X satisfying the following axioms **(I)**:
(1) $X^i = X$.
(2) $(A \cap B)^i = A^i \cap B^i$.
(3) $A^i \subset A$.
(4) $A^{ii} = A^i$.

Any one of these five structures for a set X, i.e., a structure satisfying any one of **(U)**, **(O)**, **(F)**, **(C)**, or **(I)**, determines the four other structures in a natural way. For instance, assume that a system of open sets $\mathfrak{O}$ satisfying **(O)** is given. In this case, each member of $\mathfrak{O}$ is called an **open set**. A subset U of X is called a **neighborhood** of a point x in X provided that there is an open set O such that $x \in O \subset U$. If $\mathfrak{U}(x)$ is the family of all neighborhoods of x, the function $x \to \mathfrak{U}(x)$ satisfies **(U)**. The complement of an open set in X is called a **closed set**. The family $\mathfrak{F}$ of all closed sets satisfies **(F)**. Given a subset A of X, the intersection A^a of the family of all closed sets containing A is called the **closure** of A, and each point of A^a is called an **adherent point** of A. The closure A^a is the smallest closed set containing A, and the function $A \to A^a$ satisfies **(A)**. The closure A^a is also denoted by $\bar{A}$ or $\operatorname{Cl} A$. Dually, there is a largest open subset A^i of A. The set A^i (also denoted by A° or $\operatorname{Int} A$) is called the **interior** of A, and each point of A° is called an **interior point** of A. The closure and interior are related by $A^\circ = X - \overline{(X - A)}$ and $\bar{A} = X - (X - A)^\circ$. The correspondence $A \to A^\circ$ satisfies **(I)**. Conversely, open sets can be characterized variously as follows:

$$\begin{aligned} A \text{ is open} &\Leftrightarrow A \in \mathfrak{U}(x) \text{ for each } x \text{ in } A \\ &\Leftrightarrow X - A \in \mathfrak{F} \\ &\Leftrightarrow \overline{(X - A)} = X - A \\ &\Leftrightarrow A^\circ = A. \end{aligned}$$

When a structure satisfying **(U)**, **(F)**, **(C)**, or **(I)** is given, one of the four characterizations of open sets can be used to define a system of open sets satisfying **(O)** and hence the other structure.

A **topological structure** or simply a **topology** for a space X is any of these five structures for X. If two topologies τ_1 and τ_2 for X give rise to identical systems of open sets, then τ_1 and τ_2 are considered to be identical. For this reason "topology" frequently means simply "system of open sets" in the literature. A **topological space** is a set X provided with a topology τ and is denoted by (X, τ) or simply X when there is no ambiguity.

C. Examples

(1) Discrete Topology. Let X be a set, and let the system $\mathfrak{O}$ of open sets be the family of all subsets of X. The resulting topology is called the **discrete topology**, and X with the discrete topology is a **discrete topological space**. In this space, $\bar{A} = A^\circ = A$ for each subset A, and A is a neighborhood of each of its points.

(2) Trivial Topology. The **trivial** (or **indiscrete**) **topology** for a set X is defined by the system of open sets which consists of X and $\varnothing$ only. If $A \subsetneqq X$, then $A^\circ = \varnothing$, and if $A \neq \varnothing$, then $\bar{A} = X$. Each point of X has only one neighborhood, X itself.

(3) Metric Topology. Let (X, ρ) be a †metric space, i.e., a set X provided with a †metric ρ. For a positive number ε, the ε-neighborhood of a point x is defined to be the set $U_\varepsilon(x) = \{y \mid y \in X, \rho(x, y) < \varepsilon\}$. Let $\mathfrak{U}(x)$ be the family of all sets V such that $U_\varepsilon(x) \subset V$ for some ε; then the assignment $x \to \mathfrak{U}(x)$ satisfies **(U)** and hence defines a topology. This topology is the **metric topology** for the metric space (X, ρ).

(4) Order Topology. Let X be a set †linearly ordered by $\leqslant$. For each point x in X, let $\mathfrak{U}(x)$ be the family of all subsets U such that $x \in \{y \mid a < y < b\} \subset U$ for some a, b. The function $x \to \mathfrak{U}(x)$ satisfies **(U)** and defines the **order topology** for the linearly ordered set X.

(5) Convergence and Topology. We can define the notion of convergence in a topological space, and conversely we can define a topology using convergence as a primitive notion (→ 87 Convergence). In particular, for a metric space, the metric topology can be defined in terms of convergent sequences (→ 273 Metric Spaces).

D. Generalized Topological Spaces

When a space X is equipped with a closure operator that does not satisfy all of **(C)**, the

space is called a **generalized topological space** by some authors. Topological implications of each axiom in (C) have been investigated for such spaces.

E. Local Bases

Let X be a topological space, and let x be a point of X. A collection $\mathfrak{U}_0(x)$ of neighborhoods of x is called a **base for the neighborhood system (fundamental system of neighborhoods** of a point x or **local base** at x) if each neighborhood of x contains a member of $\mathfrak{U}_0(x)$. Let $\{\mathfrak{U}_0(x)|x\in X\}$ be a system of local bases; then the system has the following properties $(\mathbf{U}_0)$:
(1) For each V in $\mathfrak{U}_0(x)$, $x\in V\subset X$.
(2) If $V_1, V_2\in\mathfrak{U}_0(x)$, then there is a V_3 in $\mathfrak{U}_0(x)$ such that $V_3\subset V_1\cap V_2$.
(3) For each V in $\mathfrak{U}_0(x)$, there exists a $W\subset V$ in $\mathfrak{U}_0(x)$ such that for each y in W, V contains some member of $\mathfrak{U}_0(y)$.

Conversely, suppose that $\{\mathfrak{U}_0(x)|x\in X\}$ is a system satisfying $(\mathbf{U}_0)$. For each x in X, let $\mathfrak{U}(x)$ consist of all subsets V of X such that $V\supset U$ for some U in $\mathfrak{U}_0(x)$. Then the system $\{\mathfrak{U}(x)|x\in X\}$ satisfies (U) and therefore defines a topology for X. This topology is called the topology determined by the system $\{\mathfrak{U}_0(x)|x\in X\}$.

For instance, in a metric space X, the set of ε-neighborhoods of $x(\varepsilon>0)$ is a local base at x with respect to the metric topology. In an arbitrary topological space, the collection of all open sets containing x, i.e., the **open neighborhoods** of x, is a local base at x.

Two systems satisfying $(\mathbf{U}_0)$ are called **equivalent** if they determine the same topology. For systems $\{\mathfrak{U}_0(x)|x\in X\}$ and $\{\mathfrak{B}_0(x)|x\in X\}$ to be equivalent it is necessary and sufficient that for each x in X each member of $\mathfrak{U}_0(x)$ contain a member of $\mathfrak{B}_0(x)$ and each member of $\mathfrak{B}_0(x)$ contain a member of $\mathfrak{U}_0(x)$.

Sometimes the word "neighborhood" stands for a member of a local base or for an open neighborhood. However, this convention is not used here.

F. Bases and Subbases

A family $\mathfrak{O}_0$ of open sets of a topological space X is called a **base for the topology (base for the space**, or **open base**) if each open set is the union of a subfamily of $\mathfrak{O}_0$. A base $\mathfrak{O}_0$ for the topology of a topological space X has the following properties $(\mathbf{O}_0)$:
(1) $\bigcup\mathfrak{O}_0=X$.
(2) If $W_1, W_2\in\mathfrak{O}_0$ and $x\in W_1\cap W_2$, then there is a W_3 in $\mathfrak{O}_0$ such that $x\in W_3\subset W_1\cap W_2$.

Conversely, if a family $\mathfrak{O}_0$ of subsets of a set X satisfies $(\mathbf{O}_0)$, then $\mathfrak{O}_0$ is a base for a unique topology. A member of $\mathfrak{O}_0$ is called a **basic open set**.

A family $\mathfrak{O}_{00}$ of open sets of a topological space X is a **subbase for the topology** (or **subbase for the space**) if the family of all finite intersections of members of $\mathfrak{O}_{00}$ is a base for the topology. If $\mathfrak{O}_{00}$ a subbase for the topology of a topological space X, then $\bigcup\mathfrak{O}_{00}=X$. Conversely, if $\mathfrak{O}_{00}$ is a family of subsets of a set X such that $\bigcup\mathfrak{O}_{00}=X$, then the family of all finite intersections of members of $\mathfrak{O}_{00}$ is a base for a unique topology τ. A subset of X is open for τ if and only if it is the union of a family of finite intersections of members of $\mathfrak{O}_{00}$. The system of open sets relative to τ is said to be **generated** by the family $\mathfrak{O}_{00}$. Thus any family of sets defines a topology for its union.

A set $\mathfrak{F}$ of subsets of a topological space is called a **network** if for each point x and its neighborhood U there is a member $F\in\mathfrak{F}$ such that $x\in F\subset U$ (A. V. Arkhangel'skiĭ, 1959). If all $F\in\mathfrak{F}$ are required to be open, the network $\mathfrak{F}$ is exactly an open base.

G. Continuous Mappings

A mapping f on a topological space X into a topological space Y is called **continuous** at a point a of X if it satisfies one of the following equivalent conditions:
(1) For each neighborhood V of $f(a)$, there is a neighborhood U of a such that $f(U)\subset V$. (1′) For each neighborhood V of $f(a)$, the inverse image $f^{-1}(V)$ is a neighborhood of a.
(2) For an arbitrary subset A of X such that $a\in\bar{A}$, $f(a)\in\overline{f(A)}$.

Continuity can also be defined in terms of convergence (→ 87 Convergence).

If f is continuous at each point of X, f is said to be **continuous**. Continuity of f is equivalent to each of the following conditions:
(1) For each open subset O of Y, the inverse image $f^{-1}(O)$ is open in X.
(1′) The inverse image under f of each member of a subbase for the topology of Y is open in X.
(2) For each closed subset F of Y, the inverse image $f^{-1}(F)$ is closed.
(3) For each subset A of X, $f(\bar{A})\subset\overline{f(A)}$.

The image $f(X)$ of X under a continuous mapping f is called a **continuous image** of X. Let X, Y, and Z be topological spaces, and let $f:X\rightarrow Y$ and $g:Y\rightarrow Z$ be mappings. If f is continuous at a point a of X and g is continuous at $f(a)$, then the composite mapping $g\circ f:X\rightarrow Z$ is continuous at the point a. Hence if f and g are continuous, so is $g\circ f$.

When a continuous mapping $f:X\rightarrow Y$ is

†bijective and f^{-1} is continuous, the mapping f is called a **homeomorphism** (named by H. Poincaré, 1895) or **topological mapping**. Two topological spaces X and Y are **homeomorphic**, $X \approx Y$, if there is a homeomorphism $f: X \rightarrow Y$.

The relation of being homeomorphic is an †equivalence relation. A property which, when held by a topological space, is also held by each space homeomorphic to it is a **topological property** or **topological invariant**. The problem of deciding whether or not given spaces are homeomorphic is called the **homeomorphism problem**.

A mapping $f: X \rightarrow Y$ is called **open** (resp. **closed**) if the image under f of each open (resp. closed) subset of X is open (closed) in Y. A continuous bijection that is either open or closed is a homeomorphism.

A continuous surjection $f: X \rightarrow Y$ is called a **quotient mapping** if $U \subset Y$ is open whenever $f^{-1}(U)$ is open (→ Section L). If moreover $f | f^{-1}(S)$ is quotient for each $S \subset Y$ as a mapping from the subspace (→ Section J) $f^{-1}(S)$ onto the subspace S, then f is called a **hereditarily quotient mapping**. Open or closed continuous mappings are hereditarily quotient mappings.

H. Comparison of Topologies

When a set X is provided with two topologies τ_1 and τ_2 and the identity mapping: $(X, \tau_1) \rightarrow (X, \tau_2)$ is continuous, the topology τ_1 is said to be **stronger** (**larger** or **finer**) than the topology τ_2, τ_2 is said to be **weaker** (**smaller** or **coarser**) than τ_1, and the notation $\tau_1 \geqslant \tau_2$ or $\tau_2 \leqslant \tau_1$ is used. Let $\mathfrak{O}_i$, $\mathfrak{F}_i$, $\mathfrak{U}_i$, and a_i be the system of open sets, system of closed sets, neighborhood system, and closure operation for X relative to the topology τ_i $(i=1,2)$, respectively. Then each of the following is equivalent to the statement $\tau_1 \geqslant \tau_2$:

(1) $\mathfrak{O}_1 \supset \mathfrak{O}_2$.

(2) $\mathfrak{F}_1 \supset \mathfrak{F}_2$.

(3) For each x in X, $\mathfrak{U}_1(x) \supset \mathfrak{U}_2(x)$.

(4) $A^{a_1} \subset A^{a_2}$ for each subset A of X.

Let S be the family of all topologies for X. Then S is ordered by the relation $\geqslant$. The discrete topology is the strongest topology for X. If $\{\tau_\lambda | \lambda \in \Lambda\}$ is a subfamily of S, then among the topologies stronger than each τ_λ, there is a weakest one $\tau_1 = \sup\{\tau_\lambda | \lambda \in \Lambda\}$. Similarly, among the topologies weaker than each τ_λ, there is a strongest one $\tau_2 = \inf\{\tau_\lambda | \lambda \in \Lambda\}$. In fact, let $\mathfrak{O}_\lambda$ be the family of all open sets relative to τ_λ; then the system of open sets for τ_1 is generated by $\bigcup_{\lambda \in \Lambda} \mathfrak{O}_\lambda$, and the system of open sets for τ_2 is precisely $\bigcap_{\lambda \in \Lambda} \mathfrak{O}_\lambda$. The family S is therefore a †complete lattice.

I. Induced Topology

Let f be a mapping from a set X into a topological space Y. Then the family $\{f^{-1}(O) | O$ is open in $Y\}$ satisfies axioms (**O**) and defines a topology for X. This topology is called the **topology induced by** f (or simply **induced topology**), and it is characterized as the weakest one among the topologies for X relative to which the mapping f is continuous.

J. Subspaces

Let (X, τ) be a topological space and M be a subset of X. The topology for M induced by the inclusion mapping $f: M \rightarrow X$, i.e., the mapping f defined by $f(x) = x$ for each x in M, is called the **relativization** of τ to M or the **relative topology**. The set M provided with the relative topology is called a **subspace** of the topological space (X, τ). Topological terms, when applied to a subspace, are frequently preceded by the adjective "relative" to avoid ambiguity. Thus a **relative neighborhood** of a point x in M is a set of the form $M \cap U$, where U is a neighborhood of x in X. A **relatively open** (**relatively closed**) set in M is a set of the form $M \cap A$, where A is open (closed) in X. For a subset A of M, the relative closure of A in M is $M \cap \bar{A}$, where $\bar{A}$ is the closure of A in X. A mapping $f: X \rightarrow Y$ is called an **embedding** if f is a homeomorphism from X to the subspace $f(X)$, and in this case X is said to be **embedded** into Y. A property P is said to hold **locally** on a topological space X if each point x of X has a neighborhood U such that the property P holds on the subspace U. A subset A of X is **locally closed** if for each point x of X, there exists a neighborhood V of x such that $V \cap A$ is relatively closed in V. A subset of X is locally closed if and only if it can be represented as $O \cap F$, where O is open and F is closed in X.

K. Product Spaces

Let X be a set, and for each member λ of an index set Λ, let f_λ be a mapping of X into a topological space X_λ. Then there is a weakest topology for X that makes each f_λ continuous. In fact, this topology is $\sup\{\tau_\lambda\}$, where τ_λ is the topology for X induced by f_λ. In particular, let $\{X_\lambda | \lambda \in \Lambda\}$ be a family of topological spaces, and let X be the Cartesian product $\prod_{\lambda \in \Lambda} X_\lambda$. Then the weakest topology for X such that each projection $\mathrm{pr}_\lambda: X \rightarrow X_\lambda$ is continuous is called the **product topology** or **weak topology**. The Cartesian product $\prod_{\lambda \in \Lambda} X_\lambda$ equipped with the product topology is called the **product topological space** or simply the **product space**

or **direct product** of the family $\{X_\lambda | \lambda \in \Lambda\}$ of topological spaces. If $\mathfrak{O}$ is the family of all open subsets of X_λ, the union $\bigcup_\lambda \mathrm{pr}_\lambda^{-1}(\mathfrak{O}_\lambda)$ is a subbase for the product topology. If $x=\{x_\lambda\}$ is a point of X, then sets of the type $\bigcap_{j=1}^n \mathrm{pr}_j^{-1}(U_j)=\prod_{\lambda\neq\lambda_1,\ldots,\lambda_n} X_\lambda \times U_1 \times \ldots \times U_n$ form a local base at x for the product topology, where $\lambda_1,\ldots,\lambda_n \in \Lambda$ and U_j is a neighborhood of x_{λ_j}. Each projection $\mathrm{pr}_\lambda: X \rightarrow X_\lambda$ is continuous and open, and a mapping f from a topological space Y into the product space $\prod_\lambda X_\lambda$ is continuous if and only if $\mathrm{pr}_\lambda \circ f: Y \rightarrow X_\lambda$ is continuous for each λ. Given a family $\{f_\lambda\}$ of continuous mappings $f_\lambda: X_\lambda \rightarrow Y_\lambda$, the product mapping $\prod_\lambda f_\lambda: \prod_\lambda X_\lambda \rightarrow \prod_\lambda Y_\lambda$ is continuous with respect to the product topologies.

For the Cartesian product $\prod_\lambda X_\lambda$ of a family $\{X_\lambda | \lambda \in \Lambda\}$ of topological spaces, there is another topology called the **box topology** (or **strong topology**). A base for the box topology is the family of all sets $\prod_\lambda O_\lambda$, where O_λ is open in X_λ for each λ. For a point $x=\{x_\lambda\}$, the family of all sets of the form $\prod_\lambda U_\lambda$ is a local base at x relative to the box topology, where U_λ is a neighborhood of x_λ for each λ. With respect to the box topology, each projection $\mathrm{pr}_\lambda: \prod_\lambda X_\lambda \rightarrow X_\lambda$ is continuous and open, and the product mapping $\prod f_\lambda: \prod_\lambda X_\lambda \rightarrow \prod_\lambda Y_\lambda$ of a family $\{f_\lambda\}$ of continuous mappings $f_\lambda: X_\lambda \rightarrow Y_\lambda$ is continuous. For a finite product of topological spaces, the product topology agrees with the box topology, but for an arbitrary product the product topology is weaker than the box topology. For the Cartesian product of topological spaces the usual topology considered is the product topology rather than the box topology.

L. Quotient Spaces

Let f be a mapping of a topological space X onto a set Y. The **quotient topology** for Y (relative to the mapping f) is the strongest topology for Y such that f is continuous. A subset O of Y is open relative to the quotient topology if and only if $f^{-1}(O)$ is open. Given an equivalence relation $\sim$ on a topological space X, the †quotient set $Y=X/\sim$ provided with the quotient topology relative to the projection $\varphi: X \rightarrow Y$ is called the **quotient topological space** (or simply **quotient space**). A mapping f from the quotient space $Y=X/\sim$ into a topological space is continuous if and only if $f \circ \varphi$ is continuous.

A **partition** of a space X is a family $\{A_\lambda | \lambda \in \Lambda\}$ of pairwise disjoint subsets of X such that $\bigcup_\lambda A_\lambda = X$. A partition $\{A_\lambda\}$ of a topological space X determines an equivalence relation $\sim$ on X such that the family $\{A_\lambda\}$ is precisely the family of all equivalence classes under $\sim$, and therefore the partition determines the quotient space $Y=X/\sim$. This space is called the **identification space** of X by the given partition. Each member A_λ of the partition can be regarded as a point of Y, and the projection $\varphi: X \rightarrow Y$ satisfies $\varphi(x)=A_\lambda$ whenever $x \in A_\lambda$. A partition $\{A_\lambda | \lambda \in \Lambda\}$ of a topological space is called **upper semicontinuous** if for each A_λ and each open set U containing A_λ, there is an open set V such that $A_\lambda \subset V \subset U$, and V is the union of members of $\{A_\lambda | \lambda \in \Lambda\}$. A partition $\{A_\lambda | \lambda \in \Lambda\}$ is upper semicontinuous if and only if the projection $\varphi: X \rightarrow Y = \{A_\lambda | \lambda \in \Lambda\}$ is a closed mapping.

M. Topological Sums

Let X be a set, and for each member λ of an index set Λ, let f_λ be a mapping of a topological space X_λ to X. Then the family $\{O \subset X | f_\lambda^{-1}(O)$ is open for any $\lambda\}$ satisfies the axioms of the open sets. This topology τ is characterized as the strongest one for X that makes each f_λ continuous. A mapping g on X with τ to a topological space Y is continuous if and only if $g \circ f_\lambda: X_\lambda \rightarrow Y$ is continuous for each $\lambda \in \Lambda$. The simplest is the case where X is the disjoint union of X_λ and f_λ is the inclusion mapping. Then we call the topological space X the **direct sum** or the **topological sum** of $\{X_\lambda\}$ and denote it by $\oplus X_\lambda$ or $\coprod X_\lambda$. More generally let the set X be the union of topological spaces $\{X_\lambda\}_{\lambda \in \Lambda}$ such that for each λ and $\mu \in \Lambda$ the relative topologies of $X_\lambda \cap X_\mu$ from X_λ and X_μ coincide. Then we call the topology τ the **weak topology** with respect to $\{X_\lambda\}$. If $X_\lambda \cap X_\mu$ is closed (resp. open) in X_μ for any μ, then X_λ is closed (resp. open) in X and the original topology of X_λ coincides with the relative topology. If, moreover, for each subset Γ of Λ, $F=\bigcup_{\lambda \in \Gamma} X_\lambda$ is closed and the weak topology of F with respect to $\{X_\lambda\}_{\lambda \in \Gamma}$ coincides with the relative topology induced by τ, then X with τ is said to have the **hereditarily weak topology** with respect to $\{X_\lambda\}$ (or to be **dominated** by $\{X_\lambda\}$). A topological space has the hereditarily weak topology with respect to any locally finite closed covering, and every CW-complex (→ 70 Complexes) has the hereditarily weak topology with respect to the covering of all finite subcomplexes.

When $\{X_n\}$ is an increasing sequence of topological spaces such that each X_n is a subspace of X_{n+1}, then the union $X=\bigcup X_n$ with the weak topology is called the **inductive limit** of $\{X_n\}$ and is denoted by $\varinjlim X_n$. Each X_n may again be regarded as a subspace of X.

N. Baire Spaces

For a subset A of a topological space X, the set $X-\bar{A}$ is called the **exterior** of A, and the set $\bar{A}\cap\overline{X-A}$ is called the **boundary** of A, denoted by $\mathrm{Bd}\,A$, $\mathrm{Fr}\,A$, or ∂A. A point belonging to the exterior (boundary) of A is an **exterior point** (**boundary point** or **frontier point**) of A. If the closure of A is X, then A is said to be **dense** in X. When $X-A$ is dense in X, i.e., when the interior of A is empty, A is called a **boundary set** (or **border set**), and if the closure $\bar{A}$ is a boundary set, A is said to be **nowhere dense**. The union of a countable family of nowhere dense sets is called a **set of the first category** (or **meager set**). A set that is not of the first category is called a **set of the second category** (or **nonmeager set**). The complement of a set of the first category is called a **residual set**. In the space $\mathbf{R}$ of real numbers, the set $\mathbf{Q}$ of all rational numbers is of the first category, and the set $\mathbf{R}-\mathbf{Q}$ of all irrational numbers is of the second category. Both $\mathbf{Q}$ and $\mathbf{R}-\mathbf{Q}$ are dense in X and hence are boundary sets. The union of a finite family of nowhere dense sets is nowhere dense, and the union of a countable family of sets of the first category is also of the first category. A subset A of X is nowhere dense in X if and only if for each open set O, $O\cap A$ is not dense in O.

A topological space X is called a **Baire space** (Baire, 1899) if each subset of X of the first category has an empty interior. Each of the following conditions is necessary and sufficient for a space X to be a Baire space:
(1) Each nonempty open subset of X is of the second category.
(2) If $F_1, F_2, \ldots$ is a sequence of closed subsets of X such that the union $\bigcup_{n=1}^{\infty} F_n$ has an interior point, then at least one F_n has an interior point.
(3) If $O_1, O_2, \ldots$ is a sequence of dense open subsets of X, then the intersection $\bigcap_{n=1}^{\infty} O_n$ is dense in X.

An open subset of a Baire space is a Baire space for the relative topology. A topological space that is homeomorphic to a complete metric space (→ 436 Uniform Spaces I) is a Baire space (**Baire-Hausdorff theorem**). A locally compact Hausdorff space (→ Section V) is also a Baire space. The class of Čech-complete completely regular spaces (→ Section T) includes both of these spaces, but there are also Baire spaces that are not in the class. A subset A of a topological space is said to satisfy **Baire's condition** or to have the **Baire property** if there exist an open set O and sets P_1, P_2 of the first category such that $A=(O\cup P_1)-P_2$. A †Borel set satisfies Baire's condition.

O. Accumulation Points

A point x is called an **accumulation point**, or a **cluster point** of a subset A of a topological space X if $x\in\overline{A-\{x\}}$. The set of all accumulation points of a set A is called the **derived set** of A and is denoted by A' or A^d. A point x belongs to A' if and only if each neighborhood of x contains a point of A other than x itself. A point belonging to the set $A^s=A-A'$ is called an **isolated point** of A, and a set A consisting of isolated points only, i.e., $A=A^s$, is said to be **discrete**. If each nonempty subset of A contains an isolated point, then A is said to be **scattered**; and if A does not possess an isolated point, i.e., $A\subset A'$, then A is said to be **dense in itself**. The largest subset of A which is dense in itself is called the **kernel** of A. If $A=A'$, then A is called a **perfect set**.

If x is an accumulation point of A, then for each neighborhood U of x, $U\cap(A-\{x\})\neq\varnothing$. Furthermore, it is possible to classify an accumulation point of A according to the †cardinality of $U\cap(A-\{x\})$. A point x is called a **condensation point** of a set A if for each neighborhood U of x, the set $U\cap A$ is uncountable. A point x is a **complete accumulation point** of A if for each neighborhood U of x, the set $U\cap A$ has the same cardinality as A.

P. Countability Axioms

A topological space X satisfies the **first countability axiom** if each point x of X has a countable local base (F. Hausdorff [3]). Metric spaces satisfy the first countability axiom. In fact, the family of $(1/n)$-neighborhoods ($n=1,2,\ldots$) of a point is a local base of the point. The topology of a topological space that satisfies the first countability axiom is completely determined by convergent sequences. For instance, the closure of a subset A of such a space consists of all limits of sequences in A (→ 87 Convergence). A topological space X is said to satisfy the **second countability axiom** or to be **perfectly separable** if there is a countable base for the topology. †Euclidean spaces satisfy the second countability axiom. If X contains a countable dense subset, X is said to be **separable**. A space that satisfies the second countability axiom satisfies the first and is also a separable Lindelöf space (→ Section S). However, the converse is not true. Each of the following properties is independent of the others: separability, the first countability axiom, and the Lindelöf property. If a metric space is separable, then it satisfies the second countability axiom. There are metric spaces that are not separable.

Q. Separation Axioms

Topological spaces that are commonly encountered usually satisfy some of the following separation axioms.

($\mathbf{T}_0$) **Kolmogorov's axiom**. For each pair of distinct points, there is a neighborhood of one point of the pair that does not contain the other.

($\mathbf{T}_1$) **The first separation axiom** or **Fréchet's axiom**. For each pair x, y of distinct points, there are neighborhoods U of x and V of y such that $x \notin V$ and $y \notin U$.

Axiom ($\mathbf{T}_1$) can be restated as follows:

($\mathbf{T}'_1$) For each point x of the space, the singleton $\{x\}$ is closed.

($\mathbf{T}_2$) **The second separation axiom** or **Hausdorff's axiom** [3]. For each pair x, y of distinct points of the space X, there exist disjoint neighborhoods of x and y.

Axiom ($\mathbf{T}_2$) is equivalent to the following:

($\mathbf{T}'_2$) In the product space $X \times X$ the diagonal set Δ is closed.

($\mathbf{T}_3$) **The third separation axiom** or **Vietoris's axiom** (*Monatsh. Math. Phys.*, 31 (1921)). Given a point x and a subset A such that $x \notin \bar{A}$, there exist disjoint open sets O_1 and O_2 such that $x \in O_1$ and $A \subset O_2$. (In this case, the sets $\{x\}$ and A are said to be **separated** by open sets.)

Axiom ($\mathbf{T}_3$) can be restated as ($\mathbf{T}'_3$) or ($\mathbf{T}''_3$):

($\mathbf{T}'_3$) For each point x of the space, there is a local base at x consisting of closed neighborhoods of x.

($\mathbf{T}''_3$) An arbitrary closed set and a point not belonging to it can be separated by open sets.

($\mathbf{T}_4$) **The fourth separation axiom** or **Tietze's first axiom** (*Math. Ann.*, 88 (1923)). Two disjoint closed sets F_1 and F_2 can be separated by open sets, i.e., there exist disjoint open sets O_1 and O_2 such that $F_1 \subset O_1$ and $F_2 \subset O_2$.

($\mathbf{T}_5$) **Tietze's second axiom**. Whenever two subsets A_1 and A_2 satisfy $A_1 \cap \bar{A}_2 = \bar{A}_1 \cap A_2 = \varnothing$, A_1 and A_2 can be separated by open sets.

It is easily seen that ($\mathbf{T}_5$) $\Rightarrow$ ($\mathbf{T}_4$), ($\mathbf{T}_0$) and ($\mathbf{T}_3$) $\Rightarrow$ ($\mathbf{T}_2$), ($\mathbf{T}_4$) and ($\mathbf{T}_1$) $\Rightarrow$ ($\mathbf{T}_3$). Axiom ($\mathbf{T}_4$) is equivalent to each of ($\mathbf{T}'_4$) and ($\mathbf{T}''_4$):

($\mathbf{T}'_4$) Whenever F_1 and F_2 are disjoint closed subsets, there exists a continuous function f on the space into the interval $[0, 1]$ such that f is identically 0 on F_1 and 1 on F_2.

($\mathbf{T}''_4$) Each real-valued continuous function defined on a closed subspace can be extended to a real-valued continuous function on the entire space.

The implications ($\mathbf{T}_4$) $\Rightarrow$ ($\mathbf{T}'_4$) and ($\mathbf{T}_4$) $\Rightarrow$ ($\mathbf{T}''_4$) are known as **Uryson's lemma** (*Math. Ann.*, 94 (1925)) and the **Tietze extension theorem** (*J. Reine Angew. Math.*, 145 (1915)), respectively. In addition, there are two more related axioms:

($\mathbf{T}_{3\frac{1}{2}}$) **Tikhonov's separation axiom**. For each closed subset F and each point x not in F, there is a real-valued continuous function f on the space such that $f(x) = 0$ and f is identically 1 on F.

($\mathbf{T}_6$) (N. Vedenisov). For each closed subset F, there is a real-valued continuous function f on the space such that $F = \{x \mid f(x) = 0\}$.

Axioms ($\mathbf{T}_5$) and ($\mathbf{T}_6$) are equivalent to the following ($\mathbf{T}'_5$) and ($\mathbf{T}'_6$), respectively:

($\mathbf{T}'_5$) Each subspace satisfies ($\mathbf{T}_4$)

($\mathbf{T}'_6$) X satisfies ($\mathbf{T}_4$) and each closed set is a $^{\dagger}G_{\delta}$-set.

The following implications are valid: ($\mathbf{T}_{3\frac{1}{2}}$) $\Rightarrow$ ($\mathbf{T}_3$), ($\mathbf{T}_6$) $\Rightarrow$ ($\mathbf{T}_5$), ($\mathbf{T}_4$) and ($\mathbf{T}_1$) $\Rightarrow$ ($\mathbf{T}_{3\frac{1}{2}}$).

Table 1 gives a classification of topological spaces by the separation axioms. Each line represents a special case of the preceding line.

A †metrizable space is perfectly normal, but the converse is false (for metrization theorems → 273 Metric Spaces). Among the spaces satisfying the second countability axiom, regular spaces are normal (**Tikhonov's theorem**, *Math. Ann.*, 95 (1925)) and metrizable (**Tikhonov-Uryson theorem**; P. Uryson, *Math. Ann.*, 94 (1925)).

Table 2 shows whether various topological properties are preserved in subspaces, product spaces, and quotient spaces. The topological properties considered are T_1, T_2 = Hausdorff, T_3 = regular, CR = completely regular, T_4 = normal, T_5 = completely normal, M = metrizable, C_I = first axiom of countability, C_{II} = second axiom of countability, C = compact, S = separable, and L = Lindelöf. Each position is filled with ○ or × according as the property (say, P) listed at the head of the column is preserved or not in the sort of space listed on the left obtained from space(s) all having property P.

R. Coverings

A family $\mathfrak{M} = \{M_\lambda\}_{\lambda \in \Lambda}$ of subsets of a set X is called a **covering** of a subset A of X if $A \subset \bigcup_\lambda M_\lambda$. If $\mathfrak{M}$ is finite (countable), it is called a **finite covering (countable covering)**. An **open (closed) covering** is a covering consisting of open (closed) sets.

A family $\mathfrak{M}$ of subsets of a topological space X is said to be **locally finite** if for each point x of X, there is a neighborhood of x which intersects only a finite number of members of $\mathfrak{M}$. If moreover $\{\bar{M}_\lambda\}_{\lambda \in \Lambda}$ is disjoint, then $\mathfrak{M}$ is called **discrete**. $\mathfrak{M}$ is called **star-finite** if each member of $\mathfrak{M}$ intersects only a finite number of members of $\mathfrak{M}$. A **σ-locally finite** or **σ-discrete** family of subsets of X is respectively the union of a countable number of locally finite or discrete families of subsets of X. A covering $\mathfrak{M}$

Table 1. Separation Axioms

Axioms	Spaces Satisfying the Axioms
(T_0)	**T_0-space (Kolmogorov space)**
(T_1)	**T_1-space (Kuratowski space)**
(T_2)	**T_2-space (Hausdorff space, separated space)**
(T_0) and (T_3)	**T_3-space (regular space)**
(T_1) and ($T_{3\frac{1}{2}}$)	**Completely regular space (Tikhonov space)**
(T_1) and (T_4)	**T_4-space (normal space)**
(T_1) and (T_5)	**T_5-space (completely normal space, hereditarily normal space)**
(T_1) and (T_6)	**T_6-space (perfectly normal space)**

Table 2. Topological Properties and Spaces

Space	T_1	T_2	T_3	CR	T_4	T_5	M	C_I	C_{II}	C	S	L
Subspace	○	○	○	○	×	○	○	○	○	×	×	×
Closed subspace	○	○	○	○	○	○	○	○	○	○	×	○
Open subspace	○	○	○	○	×	○	○	○	○	×	○	×
Product	○	○	○	○	×	×	×	×	×	○	×	×
Countable product	○	○	○	○	×	×	○	○	○	○	○	×
Quotient space	×	×	×	×	×	×	×	×	×	○	○	○

is called **point-finite** if each infinite number of members of $\mathfrak{M}$ has an empty intersection. A covering $\mathfrak{M}$ is a **refinement** of a covering $\mathfrak{N}$ (written $\mathfrak{M} \prec \mathfrak{N}$) if each member of $\mathfrak{M}$ is contained in a member of $\mathfrak{N}$. The **order** of the covering $\mathfrak{M}$ is the least integer r such that any subfamily of $\mathfrak{M}$ consisting of $r+1$ members has an empty intersection.

Let $\mathfrak{M}$ be a covering of X, and let A be a subset of X. The **star** of A relative to $\mathfrak{M}$, denoted by $S(A, \mathfrak{M})$, is the union of all members of $\mathfrak{M}$ whose intersection with A is nonempty. Let $\mathfrak{M}^{\Delta}$ denote the family $\{S(\{x\}, \mathfrak{M})\}_{x \in X}$ and $\mathfrak{M}^*$ the family $\{S(M, \mathfrak{M})\}_{M \in \mathfrak{M}}$. Then $\mathfrak{M}^{\Delta}$ and $\mathfrak{M}^*$ are coverings of X, and $\mathfrak{M} \prec \mathfrak{M}^{\Delta} \prec \mathfrak{M}^* \prec \mathfrak{M}^{\Delta\Delta}$. A covering $\mathfrak{M}$ is a **star refinement** of a covering $\mathfrak{N}$ if $\mathfrak{M}^* \prec \mathfrak{N}$, and $\mathfrak{M}$ is a **barycentric refinement** (or **Δ-refinement**) of $\mathfrak{N}$ if $\mathfrak{M}^{\Delta} \prec \mathfrak{N}$.

A sequence $\mathfrak{M}_1, \mathfrak{M}_2, \dots$ of open coverings of a topological space is called a **normal sequence** if $\mathfrak{M}_{n+1}^{\Delta} \prec \mathfrak{M}_n$ for $n = 1, 2, \dots$, and an open covering $\mathfrak{M}$ is said to be a **normal covering** if there is a normal sequence $\mathfrak{M}_1, \mathfrak{M}_2, \dots$ such that $\mathfrak{M}_1 \prec \mathfrak{M}$. The **support** (or **carrier**) of a real-valued function f on a topological space X is defined to be the closure of the set $\{x \mid f(x) \neq 0\}$. Let $\{f_\alpha\}_{\alpha \in A}$ be a family of continuous nonnegative real-valued functions on a topological space X, and for each α in A, let C_α be the support of f_α. The family $\{f_\alpha\}_{\alpha \in A}$ is called a **partition of unity** if the family $\{C_\alpha\}_{\alpha \in A}$ is locally finite and $\sum_\alpha f_\alpha(x) = 1$ for each x in X. If the covering $\{C_\alpha\}_{\alpha \in A}$ is a refinement of a covering $\mathfrak{M}$, the family $\{f_\alpha\}_{\alpha \in A}$ is called a **partition of unity subordinate to the covering** $\mathfrak{M}$. A partition of unity subordinate to a covering $\mathfrak{M}$ exists only if $\mathfrak{M}$ is a normal covering (→ Section X). If ρ is a continuous †pseudometric on a T_1-space X, then define a covering M_n for each natural number n by $M_n = \{U(x; 2^{-n})\}_{x \in X}$, where $U(x; \varepsilon) = \{y \mid \rho(x, y) < \varepsilon\}$. Then the sequence $\mathfrak{M}_1, \mathfrak{M}_2, \dots$ is a normal sequence of open coverings. Conversely, given a normal sequence $\mathfrak{M}_1, \mathfrak{M}_2, \dots$ of open coverings of X, there exists a continuous pseudometric ρ such that $\rho(x, y) \leqslant 2^{-n}$ whenever $x \in S(y, \mathfrak{M}_n)$, and $\rho(x, y) \geqslant 2^{-n-1}$ whenever $x \notin S(y, \mathfrak{M}_n)$. If in addition for each x the family $\{S(x, \mathfrak{M}_n) \mid n = 1, 2, \dots\}$ is a local base at x, then the metric topology of ρ agrees with the topology of X.

S. Compactness

If each open covering of a topological space X admits a finite open covering as its refinement, the space X is called **compact**; if each open covering of X admits a countable open refinement, X is said to be a **Lindelöf space** (P. Aleksandrov and P. Uryson, *Verh. Akad. Wetensch.*, Amsterdam, 19 (1929)); if each open covering of X admits a locally finite open refinement, X is called **paracompact** (J. Dieudonné, *J. Math. Pures Appl.*, 23 (1944)); and if each open covering of X admits a star-finite open refinement, X is said to be **strongly paracompact** (C. H. Dowker, *Amer. J. Math.*, 69 (1947)) or to have the **star-finite property** (K. Morita, *Math. Japonicae*, 1 (1948)). The space X is compact (Lindelöf) if for each open covering $\mathfrak{M}$ of X, there is a finite (countable) subfamily of $\mathfrak{M}$ whose union is X.

The following properties for a topological space X are equivalent: (1) The space X is compact. (2) If a family $\{F_\lambda\}_{\lambda \in \Lambda}$ of closed subsets of X has the **finite intersection property**, i.e., each finite subfamily of $\{F_\lambda\}_{\lambda \in \Lambda}$ has nonempty intersection, then $\bigcap_\lambda F_\lambda \neq \varnothing$. (3) Each

infinite subset of X has a complete accumulation point. (4) Each †net has a convergent †subnet. (5) Each †universal net and each †ultrafilter converge.

If a subset A of X is compact for the relative topology, A is called a **compact** subset. A subset A of X is said to be **relatively compact** if the closure of A in X is a compact subset. A closed subset of a compact topological space is compact, and a compact subset of a Hausdorff space is closed. A continuous image of a compact space is compact, each continuous mapping of a compact space into a Hausdorff space is a closed mapping, and a continuous bijection of a compact space onto a Hausdorff space is a homeomorphism. The product space of a family $\{X_\lambda\}_{\lambda\in\Lambda}$ of topological spaces is compact if and only if each factor space is compact (**Tikhonov's product theorem**, *Math. Ann.*, 102 (1930)). A compact Hausdorff space is normal. A compact Hausdorff space is metrizable if and only if it satisfies the second countability axiom. A metric space or a †uniform space is compact if and only if it is †totally bounded and †complete. A subset of a Euclidean space is compact if and only if it is closed and bounded. In a discrete space only finite subsets are compact. The cardinality of a compact Hausdorff space with the first countability axiom cannot exceed the power of the continuum (Arkhangel'skiĭ).

There are a number of conditions related to compactness. A topological space is **sequentially compact** if each sequence in X has a convergent subsequence. A space X is **countably compact** (M. Fréchet [1]) if each countable open covering of X contains a finite subfamily that covers X. A space X is **pseudocompact** (E. Hewitt, 1948) if each continuous real-valued function on X is bounded. Some authors use *compact* and *bicompact* for what we call countably compact and compact, respectively. N. Bourbaki [9] uses *compact* and *quasicompact* instead of compact Hausdorff and compact, respectively. A T_1-space is countably compact if and only if each infinite set possesses an accumulation point. If X is countably compact, then X is pseudocompact, and if X is normal, the converse also holds. If a †complete uniform space is pseudocompact, then it is compact. A space satisfying the second countability axiom is compact if and only if it is sequentially compact. If X is sequentially compact, then X is countably compact, and if X satisfies the first countability axiom, the converse is true.

T. Compactification

A **compactification** of a topological space X consists of a compact space Y and a homeomorphism of X onto a dense subspace X_1 of Y. We can always regard X as a dense subspace of a compactification Y. If X is completely regular, then there is a Hausdorff compactification Y such that each bounded real-valued continuous function on X can be extended continuously to Y. Such a compactification is unique up to homeomorphism; it is called the **Stone-Čech compactification** of X (E. Čech, *Ann. Math.*, 38 (1937); M. H. Stone, *Trans. Amer. Math. Soc.*, 41 (1937)) and is denoted by $\beta(X)$. Let $\{f_\lambda\}_{\lambda\in\Lambda}$ be the set of all continuous functions on a completely regular space X into the closed interval $I=[0,1]$. Then a continuous mapping φ of X into a **parallelotope** $I^\Lambda=\prod_\lambda I_\lambda$ $(I_\lambda=I)$ is defined by $\varphi(x)=\{f_\lambda(x)\}_{\lambda\in\Lambda}$, and the mapping φ is a homeomorphism of X onto the subspace $\varphi(X)$ of I^Λ (**Tikhonov's embedding theorem**, *Math. Ann.*, 102 (1930)). The closure $\overline{\varphi(X)}$ of $\varphi(X)$ in I^Λ is the Stone-Čech compactification of X. The natural mapping $\beta(X_1\times X_2)\rightarrow\beta(X_1)\times\beta(X_2)$ is a homeomorphism if and only if $X_1\times X_2$ is pseudocompact (I. Glicksberg, 1959).

For a topological space X, let ∞ be a point not in X, and define a topology on the union $X\cup\{\infty\}$ as follows: A subset U of $X\cup\{\infty\}$ is open if and only if either $\infty\neq U$ and U is open in X, or $\infty\in U$ and $X-U$ is a compact closed subset of X. The topological space $X\cup\{\infty\}$ thus obtained is compact, and if X is not already compact, the space $X\cup\{\infty\}$ is a compactification of X called the **one-point compactification** of X (P. S. Aleksandrov, *C. R. Acad. Sci. Paris*, 178 (1924)). The one-point compactification of a Hausdorff space is not necessarily Hausdorff. The one-point compactification of the n-dimensional Euclidean space $\mathbf{R}^n$ is homeomorphic to the n-dimensional sphere S^n.

A completely regular space X is a †G_δ-set in the Stone-Čech compactification $\beta(X)$ if and only if it is a G_δ-set in any Hausdorff space Y which contains X as a dense subspace. Then X is said to be **Čech-complete**.

U. Absolutely Closed Spaces

A Hausdorff space X is said to be **absolutely closed** (or **H-closed**; P. Aleksandrov and P. Uryson, 1929) if X is closed in each Hausdorff space containing it. A compact Hausdorff space is absolutely closed. A Hausdorff space is absolutely closed if and only if for each open covering $\{N_\lambda\}_{\lambda\in\Lambda}$ of X, there is a finite subfamily of $\{\overline{N_\lambda}\}_{\lambda\in\Lambda}$ that covers X. The product space of a family of absolutely closed spaces is absolutely closed. Each Hausdorff space is a dense subset of an absolutely closed space. Similarly, a regular space X is said to be **r-**

closed if X is closed in each regular space containing it (N. Weinberg, 1941).

V. Locally Compact Spaces

A topological space X is said to be **locally compact** if each point of X has a compact neighborhood (P. Aleksandrov and P. Uryson, 1929). A †uniform space X is said to be **uniformly locally compact** if there is a member U of the †uniformity such that $U(x)$ is compact for each x in X (→ 436 Uniform Spaces). A noncompact space X is locally compact and Hausdorff if and only if the one-point compactification of X is Hausdorff, and this is the case if and only if X is homeomorphic to an open subset of a compact Hausdorff space. A locally compact Hausdorff space is completely regular, and for each point of the space, the family of all of its compact neighborhoods forms a local base at the point. A locally closed, hence open or closed, subset of a locally compact Hausdorff space is also locally compact for the relative topology. If a subspace A of a Hausdorff space X is locally compact, then A is a locally closed subset of X. The Euclidean space $\mathbf{R}^n$ is locally compact, and hence each **locally Euclidean space**, i.e., a space such that each point admits a neighborhood homeomorphic to a Euclidean space, is locally compact. A topological space is called **σ-compact** if it can be expressed as the union of at most countably many compact subsets.

W. Proper (Perfect) Mappings

A mapping f of a topological space X into a topological space Y is said to be **proper** (N. Bourbaki [9]) (or **perfect** [14]) if it is continuous and for each topological space Z, the mapping $f\times 1: X\times Z\rightarrow Y\times Z$ is closed, where $(f\times 1)(x,z)=(f(x),z)$. A continuous mapping $f: X\rightarrow Y$ is proper if and only if it is closed and $f^{-1}(\{y\})$ is compact for each y in Y. Another necessary and sufficient condition is that if $\{x_\alpha\}_{\mathfrak{A}}$ is a †net in X such that its image $\{f(x_\alpha)\}$ converges to $y\in Y$, then a subnet of $\{x_\alpha\}$ converges to an $x\in f^{-1}(y)$ in X. A continuous mapping of a compact space into a Hausdorff space is always proper. For a compact Hausdorff space X, a quotient space Y is Hausdorff if and only if the canonical projection $\varphi: X\rightarrow Y$ is proper.

For a continuous mapping f of a locally compact Hausdorff space X into a locally compact Hausdorff space Y, the following three conditions are equivalent: (1) f is proper. (2) For each compact subset K of Y, the inverse image $f^{-1}(K)$ is compact. (3) If $X\cup\{x_\infty\}$ and $Y\cup\{y_\infty\}$ are the one-point compactifications of X and Y, then the extension f_1 of f such that $f_1(x_\infty)=y_\infty$ is continuous.

The composition of two proper mappings is proper and the direct product of an arbitrary number of proper mappings is proper.

X. Paracompact Hausdorff Spaces

A paracompact Hausdorff space (often called simply a paracompact space) is normal. For a Hausdorff space X, the following five conditions are equivalent: (1) X is paracompact. (2) X is **fully normal** (J. W. Tukey [8]), i.e., each open covering of X admits an open barycentric refinement. (3) Each open covering has a partition of unity subordinate to it. (4) Each open covering is refined by a closed covering $\{F_\alpha \mid \alpha\in A\}$ that is **closure-preserving**, i.e., $\bigcup\{F_\beta \mid \beta\in B\}$ is closed for each $B\subset A$. (5) Each open covering $\{U_\alpha \mid \alpha\in A\}$ has a **cushioned refinement** $\{V_\alpha \mid \alpha\in A\}$, i.e., $\mathrm{Cl}(\bigcup\{V_\beta \mid \beta\in B\})\subset \bigcup\{U_\beta \mid \beta\in B\}$ for each $B\subset A$. The implication (1)–(2) is **Dieudonné's theorem**. The implication (2)→(1) is A. H. **Stone's theorem** (1948), from which it follows that each metric space is paracompact. The implications (5)→(4)→(1) is **Michael's theorem** (1959, 1957).

For normal spaces, the following weaker versions of (2) and (3) hold: A T_1-space X is normal if and only if each finite open covering of X admits a finite open star refinement (or finite open barycentric refinement). For each locally finite open covering of a normal space, there is a partition of unity subordinate to it.

For a regular space X the following three conditions are equivalent: (1) X is paracompact. (2) Each open covering of X is refined by a σ-discrete open covering. (3) Each open covering of X is refined by a σ-locally finite open covering. **Tamano's product theorem**: For a completely regular space X to be paracompact it is necessary and sufficient that $X\times\beta(X)$ be normal (1960).

For a †connected locally compact space X, the following conditions are equivalent: (1) X is paracompact. (2) X is σ-compact. (3) In the one-point compactification $X\cup\{\infty\}$, the point ∞ admits a countable local base. (4) There is a locally finite open covering $\{U_\lambda\}_{\lambda\in\Lambda}$ of X such that $\bar{U}_\lambda$ is compact for each λ. (5) X is the union of a sequence $\{U_n\}$ of open sets such that $\bar{U}_n$ is compact and $\bar{U}_n\subset U_{n+1}$ $(n=1,2,\ldots)$. (6) X is strongly paracompact.

Every †F_σ-set of a paracompact Hausdorff space is paracompact (Michael, 1953). When a T_1-space X has the hereditarily weak topology with respect to a closed covering $\{F_\lambda\}$, then X is paracompact Hausdorff (normal, completely normal or perfectly normal) if and only if each

F_λ is (Morita, 1954; Michael, 1956). In particular, every CW-complex is paracompact and perfectly normal (Morita, 1953).

Y. Normality and Paracompactness of Direct Products

A topological space X is discrete if $X \times Y$ is normal for any normal space Y (M. Atsuji and M. Rudin, 1978). There are a paracompact Lindelöf space X and a separable metric space Y such that the product $X \times Y$ is not normal (Michael, 1963). The following are conditions under which the products are normal or paracompact. Let $\mathfrak{m}$ be an infinite †cardinal number. A topological space X is called $\mathfrak{m}$-paracompact if every open covering consisting of at most $\mathfrak{m}$ open sets admits a locally finite open covering as its refinement. When $\mathfrak{m}$ is countable, it is called **countably paracompact**. If X has an open base of at most $\mathfrak{m}$ members, $\mathfrak{m}$-paracompact means paracompact. The following conditions are equivalent for a topological space X: (1) X is normal and countably paracompact; (2) The product $X \times Y$ is normal and countably paracompact for any compact metric space Y; (3) $X \times I$ is normal, where $I = [0, 1]$ (C. H. Dowker, 1951). Rudin (1971) constructed an example of a collectionwise normal space (→ Section AA) that is not countably paracompact. When $\mathfrak{m}$ is general the following conditions are equivalent: (1) X is normal and $\mathfrak{m}$-paracompact; (2) If Y is a compact Hausdorff space with an open base consisting of at most $\mathfrak{m}$ sets, then $X \times Y$ is normal and $\mathfrak{m}$-paracompact; (3) $X \times I^{\mathfrak{m}}$ is normal; (4) $X \times \{0, 1\}^{\mathfrak{m}}$ is normal (Morita, 1961). In particular, the product $X \times Y$ of a paracompact Hausdorff space X and a compact Hausdorff space Y is paracompact (Dieudonné, 1944).

A topological space X is called a **P-space** if it satisfies the following conditions: Let Ω be an arbitrary set and $\{G(\alpha_1, \ldots, \alpha_i) \mid \alpha_1, \ldots \alpha_i \in \Omega, i = 1, 2, \ldots\}$ be a family of open sets such that $G(\alpha_1, \ldots, \alpha_i) \subset G(\alpha_1, \ldots, \alpha_i, \alpha_{i+1})$. Then there is a family of closed sets $\{F(\alpha_1, \ldots, \alpha_i) \mid \alpha_1, \ldots, \alpha_i \in \Omega, i = 1, 2, \ldots\}$ such that $F(\alpha_1, \ldots, \alpha_i) \subset G(\alpha_1, \ldots, \alpha_i)$ and that if $\bigcup_{i=1}^{\infty} G(\alpha_1, \ldots, \alpha_i) = X$ for a sequence $\{\alpha_i\}$, then $\bigcup_{i=1}^{\infty} F(\alpha_1, \ldots, \alpha_i) = X$. Perfectly normal spaces, countably compact spaces, Čech-complete paracompact spaces and σ-compact regular spaces are P-spaces. Normal P-spaces are countably paracompact. A Hausdorff space X is a normal (resp. paracompact) P-space if and only if the product $X \times Y$ is normal (resp. paracompact) for any metric space Y (Morita, *Math. Ann.*, 154 1964).

The product $X \times Y$ of locally compact Hausdorff spaces X and Y is a locally compact Hausdorff space. If, in this case, X and Y are paracompact, then so is the product. If the direct product space $\prod_\lambda X_\lambda$ of metric spaces is normal, then X_λ are compact except for at most countably many λ, and hence the product space is paracompact (A. H. Stone, 1948).

A class $\mathscr{C}$ of topological spaces is called **countably productive** if for a sequence X_i of members of $\mathscr{C}$ their product $\prod X_i$ is again a member of $\mathscr{C}$. The classes of (complete) (separable) metric spaces form such examples. The class of paracompact and Čech-complete spaces is countably productive (Z. Frolík, 1960). A topological space X is called a **p-space** if it is completely regular and there is a sequence $\mathfrak{M}_i$ of families of open sets in the Stone-Čech compactification $\beta(X)$ such that, for each point $x \in X$, $x \in \bigcap S(x, \mathfrak{M}_i) \subset X$ (Arkhangel'skiĭ, 1963). X is called an **M-space** if there is a normal sequence $\mathfrak{M}_i$ of open coverings of X such that if $K_1 \supset K_2 \supset \ldots$ is a sequence of nonempty closed sets and $K_i \subset S(x, \mathfrak{M}_i)$, $i = 1, 2, \ldots$, for an $x \in X$, then $\bigcap K_i \neq \varnothing$ (Morita, 1963). The class of paracompact p-spaces and that of paracompact Hausdorff M-spaces are the same and are countably productive. For a covering $\mathfrak{F}$ of X and an $x \in X$ we set $C(x, \mathfrak{F}) = \bigcap \{F \mid x \in F \in \mathfrak{F}\}$. X is called a **Σ-space** if X admits a sequence $\mathfrak{F}_i$ of locally finite closed coverings such that if $K_1 \supset K_2 \supset \ldots$ is a sequence of nonempty closed sets and $K_i \subset C(x, \mathfrak{F}_i)$, $i = 1, 2, \ldots$, for an $x \in X$, then $\bigcap K_i \neq \varnothing$ (K. Nagami, 1969). Σ-spaces are P-spaces. The class of all paracompact Σ-spaces is also countably productive. Among the above classes each one is always wider than its predecessors. Yet the product $X \times Y$ of a paracompact Hausdorff P-space X and a paracompact Hausdorff Σ-space Y is paracompact. Other examples of countably productive classes are the Suslin spaces and the Luzin spaces (→ Section CC) introduced by Bourbaki (1958), the **stratifiable spaces** by J. G. Ceder (1961) and C. J. R. Borges (1966), the **$\aleph_0$-spaces** by Michael (1966) and the **σ-spaces** by A. Okuyama (1967).

Z. Strongly Paracompact Spaces

Regular Lindelöf spaces are strongly paracompact. Conversely, if a connected regular space is strongly paracompact, then it is a Lindelöf space (Morita, 1948). Hence a connected nonseparable metric space is not strongly paracompact. Paracompact locally compact Hausdorff spaces and uniformly locally compact Hausdorff spaces are strongly paracompact.

These classes of spaces coincide under suitable †uniform structures.

AA. Collectionwise Normal Spaces

A Hausdorff space X is called a **collectionwise normal space** if for each discrete collection $\{F_\alpha | \alpha \in A\}$ of closed sets of X there exists a disjoint collection $\{U_\alpha | \alpha \in A\}$ of open sets with $F_\alpha \subset U_\alpha$ $(\alpha \in A)$ (R. H. Bing, 1951). If X satisfies an analogous condition for the case where each F_α is a singleton, X is called a **collectionwise Hausdorff space**. Paracompact Hausdorff spaces are collectionwise normal (Bing). Every point-finite open covering of a collectionwise normal space has a locally finite open refinement (Michael, Nagami).

A topological space X is called a **developable space** if it admits a sequence $\mathfrak{U}_i$, $i=1,2,\ldots$, of open coverings such that, for each point $x \in X$, $\{S(x, \mathfrak{U}_i) | i=1,2,\ldots\}$ forms a base for the neighborhood system of x (R. L. Moore, 1916). A regular developable space is called a **Moore space**. The question of whether or not every normal Moore space is metrizable is known as the **normal Moore space problem** (→ 273 Metric Spaces K). Collectionwise normal Moore spaces are metrizable (Bing).

BB. Real-Compact Spaces

A completely regular space X is called **real-compact** if X is complete under the smallest †uniformity such that each continuous real-valued function on X is uniformly continuous (→ 422 Uniform Spaces). This notion was introduced by E. Hewitt (*Trans. Amer. Math. Soc.*, 64 (1948)) under the name of **Q-space**, and independently by L. Nachbin (*Proc. International Congress of Mathematicians*, Cambridge, Mass., 1950).

A Lindelöf space is real-compact. If X_1 and X_2 are real-compact spaces such that the rings $C(X_1)$ and $C(X_2)$ of continuous real-valued functions on X_1 and X_2 are isomorphic, then X_1 and X_2 are homeomorphic (Hewitt). If X is real-compact, then X is homeomorphic to a closed subspace of the product space of copies of the space of real numbers, and conversely.

CC. Images and Inverse Images of Topological Spaces

Each continuous mapping $f: X \to Y$ is decomposed into the product $i \circ h \circ p$ of continuous mappings $p: X \to X/\sim$, $h: X/\sim \to f(X)$ and $i: f(X) \to Y$, where $\sim$ is the equivalence relation such that $x_1 \sim x_2$ if and only if $f(x_1) = f(x_2)$. The mapping f is open (resp. closed) if and only if these mappings are all open (resp. closed). Then h is a homeomorphism. The image of a paracompact Hausdorff space under a closed continuous mapping is paracompact (Michael, 1957).

Let $f: X \to Y$ be a perfect surjection. Then Y is called a **perfect image** of X and X a **perfect inverse image** of Y. If, in this case, one of X and Y satisfies a property such as being compact, locally compact, σ-compact, Lindelöf, or countably compact, then the other also satisfies the property. When X and Y are completely regular, the same is true with regard to Čech completeness. Properties such as regularity, normality, complete normality, perfect normality, and the second countability axiom are preserved in perfect images; but complete regularity and strong paracompactness are not. Perfect images of metric spaces are also metrizable (S. Hanai and Morita, A. H. Stone, 1956). Conversely, perfect inverse images of paracompact spaces are paracompact. If a Hausdorff space is a perfect inverse image of a regular space (resp. k-space; → below), then it is a regular space (resp. k-space). Every paracompact Čech-complete space is a perfect inverse image of a †complete metric space (Z. Frolík, 1961). A completely regular space is a paracompact p-space if and only if it is a perfect inverse image of a metric space (Arkhangel'skiĭ, 1963). A mapping $f: X \to Y$ is called **quasi-perfect** if it is closed and continuous and the inverse image $f^{-1}(y)$ of each point $y \in Y$ is countably compact. A topological space X is an M-space if and only if there is a quasi-perfect mapping from X onto a metric space Y (Morita, 1964). Let $f: X \to Y$ be a quasi-perfect surjection. If one of X and Y is a Σ-space, then the other is also a Σ-space (Nagami, 1969).

A topological space X is called a **Fréchet-Uryson space** (or a **Fréchet space**) if the closure of an arbitrary set $A \subset X$ is the set of all limits of sequences in A (Arkhangel'skiĭ, 1963). X is called a **sequential space** if $A \subset X$ is closed whenever A contains all the limits of sequences in A (S. P. Franklin, 1965). X is called a **k′-space** if the closure of an arbitrary set A is the set of all points adherent to the intersection $A \cap K$ for a compact set K in X (Arkhangel'skiĭ, 1963). X is called a **k-space** if $A \subset X$ is closed whenever $A \cap K$ is closed in K for any compact set K (→ Arkhangel'skiĭ, *Trudy Moskov. Mat. Obshch.*, 13 (1965)). Spaces satisfying the first countability axiom are Fréchet-Uryson spaces. The Fréchet-Uryson spaces (resp. sequential spaces) are characterized as the images under hereditarily quotient (resp. quotient) mappings of metric spaces or locally compact metric

spaces. Similarly the k′-spaces (resp. k-spaces) coincide with the images under hereditarily quotient (resp. quotient) mappings of locally compact spaces. The image of a metric space under a closed continuous mapping is called a **Lashnev space**. Any subspace of a Fréchet-Uryson space is a Fréchet-Uryson space. Conversely, a Hausdorff space is a Fréchet-Uryson space if any of its subspaces is a k-space. Čech-complete spaces are k-spaces. A Hausdorff space is called a **Suslin space** (resp. **Luzin space**) if it is the image under a continuous surjection (resp. continuous bijection) of a complete separable metric space (Bourbaki [9]; also → 22 Analytic Sets).

In Figs. 1, 2, and 3, the relationships between the various properties are indicated by the arrows.

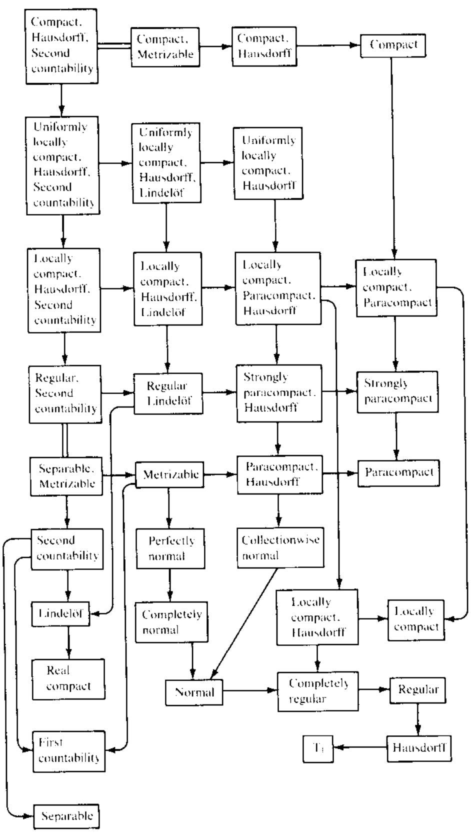


Fig. 1

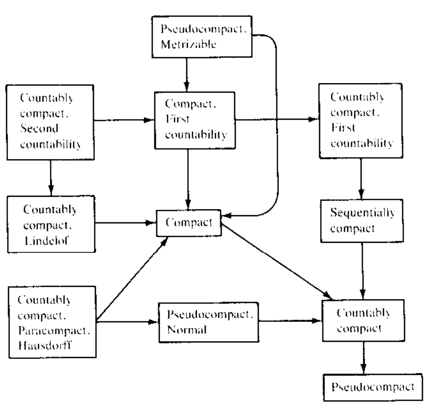


Fig. 2

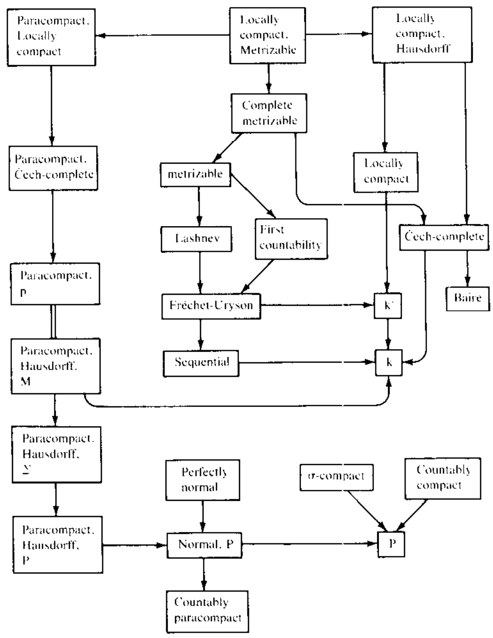


Fig. 3

References

[1] M. Fréchet, Sur quelques points du calcul fonctionnel, Rend. Circ. Mat. Palermo, 22 (1906), 1–74.
[2] F. Riesz, Stetigkeitsbegriff und abstrakte Mengenlehre, Atti del IV Congr. Intern. Mat. Roma 1908, II, 18–24, Rome, 1909.
[3] F. Hausdorff, Grundzüge der Mengenlehre, Teubner, 1914.
[4] M. Fréchet, Les espaces abstraits, Gauthier-Villars, 1928.
[5] C. Kuratowski, Topologie, I, Monograf. Mat., I, 1933, revised edition, 1948; II, 1950.
[6] W. Sierpiński, Introduction to general topology, University of Toronto Press, 1934.
[7] P. S. Aleksandrov (Alexandroff) and H. Hopf, Topologie I, Springer, 1935.
[8] J. W. Tukey, Convergence and uniformity in topology, Ann. Math. Studies, Princeton Univ. Press, 1940.
[9] N. Bourbaki, Eléments de mathématique, III. Topologie générale, 1940–1949; new edition, 1970–1974; Hermann, English translation, General topology, Addison-Wesley, 1966.
[10] J. L. Kelley, General topology, Van Nostrand, 1955.
[11] J. Dugundji, Topology, Allyn and Bacon, 1966.
[12] J. Nagata, Modern general topology, North-Holland, 1968.
[13] R. Engelking, General topology, Polish Scientific Publ., 1977.

[14] P. S. Aleksandrov, Some results in the theory of topological spaces, obtained within the last twenty-five years, Russian Math. Surveys, 16, no. 2 (1960), 23–83.
[15] A. V. Arkhangel'skiĭ, Mappings and spaces, Russian Math. Surveys, 21, no. 4 (1966), 115–162.
[16] Y. Kodama and K. Nagami, Theory of topological spaces (in Japanese), Iwanami, 1974.
[17] L. Steen and J. Seebach, Jr., Counterexamples in topology, Holt, Rinehart and Winston, 1970.
[18] L. Gillman and M. Jerison, Rings of continuous functions, Van Nostrand, 1960.
[19] S. A. Gaal, Point set topology, Academic Press, 1964.
[20] E. Čech, Topological spaces, Interscience, 1966.
[21] R. C. Walker, The Stone-Čech compactification, Springer, 1974.
[22] R. A. Alò and H. L. Shapiro, Normal topological spaces, Cambridge Univ. Press, 1974.
[23] M. D. Weir, Hewitt-Nachbin spaces, North-Holland, 1975.

426 (IX.1) Topology

The term **topology** means a branch of mathematics that deals with topological properties of geometric figures or point sets. A classical result in topology is the Euler relation on polyhedra: Let α_0, α_1, and α_2 be the numbers of vertices, edges, and faces of a polyhedron homeomorphic to the 2-dimensional sphere; then $\alpha_0-\alpha_1+\alpha_2=2$ (†Euler-Poincaré formula for the 2-dimensional case; actually, the formula was known to Descartes). It is one of the earliest results in topology. In 1833, C. F. Gauss used integrals to define the notion of †linking numbers of two closed curves in a space (→ 99 Degree of Mapping). It was in J. B. Listing's classical work *Vorstudien zur Topologie* (1847) that the term *topology* first appeared in print.

In the 19th century, B. Riemann published many works on function theory in which topological methods played an essential role. He solved the homeomorphism problem for compact surfaces (→ 410 Surfaces); his result is basic in the theory of algebraic functions. In the same period, mathematicians began to study topological properties of n-dimensional polyhedra. E. Betti considered the notion of †homology. H. Poincaré, however, was the first to recognize the importance of a topological approach to analysis in general; he defined the †homology groups of a complex [1]. He obtained the famous †Poincaré duality theorem and defined the †fundamental group. He considered †polyhedra as the basic objects in topology, and deduced topological properties utilizing †complexes obtained from polyhedra by †simplicial decompositions. He thus constructed a branch of topology known as **combinatorial topology**.

In its beginning stages combinatorial topology dealt only with polyhedra. In the late 1920s, however, it became possible to apply combinatorial methods to general †compact spaces. P. S. Alexandrov introduced the concept of approximation of a †compact metric space by an inverse sequence of complexes and the definition of homology groups for these spaces. His idea had a precursor in the notion of †simplicial approximations of continuous mappings, which was introduced by L. E. J. Brouwer in 1911. In 1932, E. Čech defined homology groups for arbitrary spaces utilizing the †inductive limit of the homology groups of polyhedra; and †Čech cohomology groups for arbitrary spaces were also defined. S. Eilenberg established †singular (co)homology theory using †singular chain complexes (1944). The axiomatic approach to (co)homology theory is due to Eilenberg and Steenrod, who gave axioms for (co)homology theory in a most comprehensive way and unified various (co)homology theories (1945) (→ 201 Homology Theory).

The approach using algebraic methods has progressed extensively in connection with the development of homology theory. This branch is called **algebraic topology**. In the 1920s and 1930s, a number of remarkable results in algebraic topology, such as the †Alexander duality theorem, the †Lefschetz fixed-point theorem, and the †Hopf invariant, were obtained. In the late 1930s, W. Hurewicz developed the theory of higher-dimensional †homotopy groups (→ 153 Fixed-Point Theorems, 201 Homology Theory, 202 Homotopy Theory). J. H. C. Whitehead introduced the concept of †CW complexes and proved an algebraic characterization of the homotopy equivalence of CW complexes. N. Steenrod developed †obstruction theory utilizing †squaring operations in the cohomology ring (1947). Subsequently, the theory of †cohomology operations was introduced (→ 64 Cohomology Operations, 305 Obstructions). The theory of †spectral sequences for †fiber spaces was originated by J. Leray (1945) and J.-P. Serre (1951) and was successfully applied to cohomology operations and homotopy theory by H. Cartan and Serre (1954) (→ 148 Fiber Spaces, 200 Homological Algebra). The study of the combinatorial

structures of polyhedra and †piecewise linear mappings has flourished since 1940 in the works of Whitehead, S. S. Cairns, and others. S. Smale and, independently, J. Stallings solved the †generalized Poincaré conjecture in 1960. The †Hauptvermutung in combinatorial topology was solved negatively in 1961 by B. Mazur and J. Milnor. E. C. Zeeman proved the unknottedness of codimension 3 (1962). The recent development of the theory in conjunction with progress in †differential topology is notable. The Hauptvermutung for combinatorial manifolds was solved in 1969 by Kirby, Siebenmann, and Wall. In particular, there exist different combinatorial structures on tori of dimension $\geqslant 5$, and there are topological manifolds that do not admit any combinatorial structure (→ 65 Combinatorial Manifolds, 114 Differential Topology, 235 Knot Theory).

The global theory of differentiable manifolds started from the algebraic-topological study of †fiber bundles and †characteristic classes in the 1940s. R. Thom's fundamental theorem of †cobordism (1954) was obtained through extensive use of cohomology operations and homotopy groups. Milnor (1956) showed that the sphere S^7 may have differentiable structures that are essentially distinct from each other by using †Morse theory and the †index theorem of Thom and Hirzebruch. These results led to the creation of a new field, †differential topology (→ 56 Characteristic Classes, 114 Differential Topology).

Since 1959, A. Grothendieck, M. F. Atiyah, F. Hirzebruch, and J. F. Adams have developed †K-theory, which is a generalized cohomology theory constructed using stable classes of †vector bundles (→ 237 K-Theory).

†Knot theory, an interesting branch of topology, was one of the classical branches of topology and is now studied in connection with the theory of low-dimensional manifolds (→ 235 Knot Theory).

On the other hand, G. Cantor established general set theory in the 1870s and introduced such notions as †accumulation points, †open sets, and †closed sets in Euclidean space. The first important generalization of this theory was the concept of †topological space, which was proposed by M. Fréchet and developed by F. Hausdorff at the beginning of the 20th century. The theory subsequently became a new field of study, called **general topology** or **set-theoretic topology**. It deals with the topological properties of point sets in a Euclidean or topological space without reference to polyhedra. There has been a remarkable development of the theory since abount 1920, notably by Polish mathematicians S. Janiszewski, W. Sierpiński, S. Mazurkiewicz, C. Kuratowski, and others. The contributions of R. L. Moore, G. T. Whyburn, and K. Menger are also important (→ 382 Shape Theory, 425 Topological Spaces).

Topology is not only a foundation of various theories, but is also itself one of the most important branches of mathematics. It consists of †homology theory, †homotopy theory, †differential topology, †combinatorial manifolds, †K-theory, †transformation groups, †theory of singularities, †foliations, †dynamical systems, †catastrophe theory, etc. It continues to develop in interaction with other branches of mathematics (→ 51 Catastrophe Theory, 126 Dynamical Systems, 154 Foliations, 418 Theory of Singularities, 431 Transformation Groups).

References

[1] H. Poincaré, Analysis situs, J. Ecole Polytech., (2) 1 (1895), 1–121. (Oeuvres, Gauthier-Villars, 1953, vol. 6, 193–288.)
[2] P. S. Aleksandrov and H. Hopf, Topologie I, Springer, 1935 (Chelsea, 1965).
[3] S. Lefschetz, Algebraic topology, Amer. Math. Soc. Colloq. Publ., 1942.
[4] S. Eilenberg and N. E. Steenrod, Foundations of algebraic topology, Princeton Univ. Press, 1952.

427 (IX.12) Topology of Lie Groups and Homogeneous Spaces

A. General Remarks

Among various topological structures of †Lie groups and †homogeneous spaces, the structures of their †(co)homology groups and †homotopy groups are of special interest. Let G/H be a homogeneous space, where G is a Lie group and H is its closed subgroup. Then $(G, G/H, H)$ is a †fiber bundle, where G/H is the base space and H is the fiber. Thus homology and homotopy theory of fiber bundles (†spectral sequences and †homotopy exact sequences) can be applied. The †cellular decomposition of †Stiefel manifolds, †Grassmann manifolds, and †Kähler homogeneous spaces are known. Concerning †symmetric Riemannian spaces, we have various interesting methods, such as the use of invariant differential forms in connection with real cohomology rings and the use of †Morse theory in order to establish relations between the diagrams of symmetric Riemannian spaces G/H and homological properties of their †loop spaces and

some related homogeneous spaces [4, 5]. Lie groups can be regarded as special cases of homogeneous spaces or symmetric spaces, although their group structures are of particular importance. A connected Lie group is homeomorphic to the product of one of its compact subgroups and a Euclidean space (†Cartan-Mal'tsev-Iwasawa theorem). Hence the topological structure of a connected Lie group is essentially determined by the topological structures of its compact subgroups.

B. Homology of Compact Lie Groups

Let G be a connected compact Lie group. Since G is an †H-space whose multiplication is given by its group multiplication h, $H^*(G;k)$ and $H_*(G;k)$ are dual †Hopf algebras for any coefficient field k. Also, $H^*(G;k)$ is isomorphic as a †graded algebra to the tensor product of †elementary Hopf algebras (→ 203 Hopf Algebras), but no factor of the tensor product is isomorphic to a polynomial ring because G is a finite †polyhedron. In particular, if $k=\mathbf{R}$ (the field of real numbers), then $H^*(G;\mathbf{R})\cong \bigwedge_{\mathbf{R}}(x_1,\dots,x_l)$ (the exterior (Grassmann) algebra over $\mathbf{R}$ with generators $x_1,\dots,x_l$ of odd degrees). Here we can choose generators x_i such that $h^*(x_i)=1\otimes x_i+x_i\otimes 1$, $1\leqslant i\leqslant l$. The x_i that satisfy this property are said to be **primitive**. Since in this case the †comultiplication h^* is commutative, the multiplication h_* is also commutative and the Hopf algebra $H_*(G;R)$ is an exterior algebra generated by elements y_i having the same degree as x_i $(i=1,\dots,l)$. When the characteristic of the coefficient field k is nonzero, h_* need not be commutative.

The dimension of a †maximal torus of a connected compact Lie group G is independent of the choice of the maximal torus and is called the **rank** of G. The rank of G coincides with the number l of generators of $H^*(G;\mathbf{R})$. E. Cartan studied $H^*(G;\mathbf{R})$ by utilizing invariant differential forms. The cohomology theory of Lie algebras originated from the method he used in his study. $H^*(G;\mathbf{R})$ is invariant under †local isomorphisms of groups G. For †classical compact simple Lie groups G, R. Brauer calculated $H^*(G;\mathbf{R})$, while C.-T. Yen and C. Chevalley calculated $H^*(G;\mathbf{R})$ for †exceptional compact simple Lie groups (→ Appendix A, Table 6.IV). The degrees of the generators have group-theoretic meaning. Suppose that the degree of the ith generator is $2m_i-1$, $1\leqslant i\leqslant l$, and that $m_1\leqslant m_2\leqslant\dots\leqslant m_l$. When G is simple, there is a relation $m_i+m_{l-i+1}=$ constant (Chevalley's duality). We have a proof for this property that does not use classification.

The cohomology groups $H^*(G;\mathbf{Z}_p)$ (where p is a prime and $\mathbf{Z}_p=\mathbf{Z}/p\mathbf{Z}$) have been determined as graded algebras for all compact simple Lie groups by A. Borel, S. Araki, and P. Baum and W. Browder (→ Appendix A, Table 6.IV).

C. Cohomology of Classifying Spaces

Let (E_G, B_G, G) be a †universal bundle of a connected compact Lie group G and p a prime or zero. Suppose that the integral cohomology of G has no p-torsion (no torsion when $p=0$). Then $H^*(G;\mathbf{Z}_p)=\bigwedge_{\mathbf{Z}_p}(x'_1,\dots,x'_l)$ $(H^*(G;\mathbf{Z})=\bigwedge_{\mathbf{Z}}(x'_1,\dots,x'_l)$ when $p=0)$, an exterior algebra with $\deg x'_i=2m_i-1$, $1\leqslant i\leqslant l$, and the generators x'_i can be chosen to be †transgressive in the spectral sequence of the universal bundle. Let $y_1,\dots,y_l$ be their transgression images. Then $\deg y_i=2m_i$, $1\leqslant i\leqslant l$, and the cohomology of the †classifying space B_G over $\mathbf{Z}_p$ (resp. $\mathbf{Z}$) is the polynomial algebra with generators $y_1,\dots,y_l$. Let T be a maximal torus of G. Then $B_T=E_G/T$ is a classifying space of T, the †Weyl group $W=N(T)/T$ of G with respect to T operates on B_T by †right translations, and $H^*(T;\mathbf{Z})$ has no torsion and is an exterior algebra with l generators of degree 1. Thus $H^*(B_T;\mathbf{Z})=\mathbf{Z}[u_1,\dots,u_l]$, $\deg u_i=2$. Let I_W be the subalgebra of $H^*(B_T;\mathbf{Z})$ consisting of W-invariant polynomials, and let ρ be the projection of the bundle $(B_T, B_G, G/T)$. Then under the assumption that G has no p-torsion (no torsion), the cohomology mapping ρ^* over $\mathbf{Z}_p(\mathbf{Z})$ is monomorphic, and $\rho^*: H^*(B_G;\mathbf{Z}_p)\cong I_W\otimes\mathbf{Z}_p$ $(H^*(B_G;\mathbf{Z})\cong I_W)$ [1]. In the case of real coefficients, we have $H^*(B_G;\mathbf{R})\cong I_W\otimes\mathbf{R}$ for all G, and $m_1,\dots,m_l$ are the degrees of generators of the ring I_W of W-invariant polynomials.

Example (1) $G=U(n)$: $l=n$ and G has no torsion. W operates on $H^*(B_T;\mathbf{Z})$ as the group of all permutations of generators $u_1,\dots,u_n$. Thus generators of I_W are the †elementary symmetric polynomials $\sigma_1,\dots,\sigma_n$ of $u_1,\dots,u_n$. Let $c_1,\dots,c_n$ be the †universal Chern classes; then $\rho^*(c_i)=\sigma_i$ and $H^*(B_{U(n)};\mathbf{Z})=\mathbf{Z}[c_1,\dots,c_n]$.

Example (2) $G=SO(n)$: $l=[n/2]$ and G has no p-torsion for $p\neq 2$. W operates on $H^*(B_T;\mathbf{Z})$ as the group generated by the permutations of generators $u_1,\dots,u_l$ and by the transformations $\sigma(u_i)=e_iu_i$, $e_i=\pm 1$, where the number of u_i for which $e_i=-1$ is arbitrary for odd n and even for even n. Thus the generators of I_W are the elementary symmetric polynomials $\sigma'_1,\dots,\sigma'_l$ of $u_1^2,\dots,u_l^2$ for odd n and $\sigma'_1,\dots,\sigma'_{l-1}$ and $u_1\dots u_l$ for even n. Let $p_1,\dots,p_l$ be the †universal Pontryagin classes and χ be the †universal Euler-Poincaré class in the case of even n. Then $\rho^*(p_i)=\sigma'_i$ and $\rho^*(\chi)=u_1\dots u_l$ for integral cohomology. Denote the mod p

reduction of p_i and χ by $\bar{p}_i$ and $\bar{\chi}$, respectively. Then $H^*(B_{SO(2l+1)};\mathbf{Z}_p)=\mathbf{Z}_p[\bar{p}_1,\ldots,\bar{p}_l]$ and $H^*(B_{SO(2l)};\mathbf{Z}_p)=\mathbf{Z}_p[\bar{p}_1,\ldots,\bar{p}_{l-1},\bar{\chi}]$ ($p=0$ or >2).

Example (3) $G=O(n)$: If we use the subgroup Q consisting of all diagonal matrices instead of T, then we can make a similar argument for $\mathbf{Z}_2$-cohomology. Since $Q\cong(\mathbf{Z}_2)^n$, $H^*(B_Q;\mathbf{Z}_2)=\mathbf{Z}_2[v_1,\ldots,v_n]$ ($\mathbf{Z}_2[v_1,\ldots,v_n]$ is a polynomial ring with $\deg v_i=1$), and $W_2=N(Q)/Q$ operates on B_Q by right translations and on $H^*(B_Q;\mathbf{Z}_2)$ as the group of all permutations of $v_1,\ldots,v_n$. Let I_{W_2} be the subalgebra of $H^*(B_Q;\mathbf{Z}_2)$ consisting of all W_2-invariant polynomials. Then I_{W_2} is a polynomial ring generated by the elementary symmetric polynomials $\sigma_1'',\ldots,\sigma_n''$ of $v_1,\ldots,v_n$. The projection $\rho_2:B_Q\to B_{O(n)}$ induces a monomorphic cohomology mapping ρ_2^* over $\mathbf{Z}_2$, and ρ_2^*: $H^*(B_{O(n)};\mathbf{Z}_2)\cong I_{W_2}$. Let $w_1,\ldots,w_n$ be the †universal Stiefel-Whitney classes. Then $\rho_2^*(w_i)=\sigma_i''$ and $H^*(B_{O(n)};\mathbf{Z}_2)=\mathbf{Z}_2[w_1,\ldots,w_n]$ [2].

D. Grassmann Manifolds

The following manifolds are called **Grassmann manifolds**: The manifold $M_{n+m,n}(\mathbf{R})$ consisting of all n-subspaces of $\mathbf{R}^{n+m}$; the manifold $\tilde{M}_{n+m,n}(\mathbf{R})$ consisting of all oriented n-subspaces of $\mathbf{R}^{n+m}$; and the manifold $M_{n+m,n}(\mathbf{C})$ consisting of all complex n-subspaces of $\mathbf{C}^{n+m}$. These are expressed as quotient spaces as follows: $M_{n+m,n}(\mathbf{R})=O(n+m)/O(n)\times O(m)$, $\tilde{M}_{n+m,n}(\mathbf{R})=SO(n+m)/SO(n)\times SO(m)$, and $M_{n+m,n}(\mathbf{C})=U(n+m)/U(n)\times U(m)$. They admit cellular decompositions by †Schubert varieties from which their cohomologies can be computed (→ 56 Characteristic Classes). $M_{n+m,n}(\mathbf{R})$ and $\tilde{M}_{n+m,n}(\mathbf{R})$ have no p-torsion for $p\neq 2$, and $M_{n+m,n}(\mathbf{C})$ has no torsion. These spaces are m-, m-, and $(2m+1)$-classifying spaces of $O(n)$, $SO(n)$, and $U(n)$, respectively. Hence their cohomologies are isomorphic to those of B_G ($G=O(n)$, $SO(n)$, $U(n)$) in dimensions $<m$, $<m$, and $\leqslant 2m$, respectively; and they are polynomial rings generated by suitable universal characteristic classes in low dimensions.

E. Cohomologies of Homogeneous Spaces G/U (Rank G = Rank U)

Let G be a compact connected Lie group and U a closed subgroup of G with the same rank as G. Denote the degrees of generators of $H^*(G;\mathbf{R})$ and $H^*(U;\mathbf{R})$ by $2m_1-1,\ldots,2m_l-1$, and $2n_1-1,\ldots,2n_l-1$, respectively. Then the real-coefficient †Poincaré polynomial P_0 of the homogeneous space G/U is given by $P_0(G/U,t)=\prod_i(1-t^{2m_i})/(1-t^{2n_i})$ (G. Hirsch). When G, U, and G/U have no p-torsion, the same formula is valid for the $\mathbf{Z}_p$-coefficient Poincaré polynomial [1]. When U is the †centralizer of a torus, G/U has a complex analytic cellular decomposition [3]. Hence G/U has no torsion in this case. This was proved by R. Bott and H. Samelson by utilizing Morse theory [5] (→ 279 Morse Theory). The case $U=T$ has also been studied.

F. Homotopy Groups of Compact Lie Groups

The †fundamental group $\pi_1(G)$ of a compact Lie group G is Abelian. Furthermore, $\pi_2(G)=0$. If we apply Morse theory to G, the variational completeness of G can be utilized to show that the loop space ΩG has no torsion and that its odd-dimensional cohomologies vanish [4]. Consequently, when G is non-Abelian and simple, we have $\pi_3(G)\cong\mathbf{Z}$. A †periodicity theorem on †stable homotopy groups of classical groups proved by Bott is used in K-theory (→ 202 Homotopy Theory; 237 K-Theory). (For explicit forms of homotopy groups → Appendix A, Table 6.VI).

Homotopy groups of Stiefel manifolds are used to define characteristic classes by †obstruction cocycles (→ 147 Fiber Bundles; Appendix A, Table 6.VI).

References

[1] A. Borel, Sur la cohomologie des espaces fibrés principaux et des espaces homogènes de groupes de Lie compacts, Ann. Math., (2) 57 (1953), 115–207.
[2] A. Borel, La cohomologie mod 2 de certains espaces homogènes, Comment. Math. Helv., 27 (1953), 165–197.
[3] A. Borel, Kählerian coset spaces of semisimple Lie groups, Proc. Nat. Acad. Sci. US, 40 (1954), 1147–1151.
[4] R. Bott, An application of the Morse theory to the topology of Lie-groups, Bull. Soc. Math. France, 84 (1956), 251–281.
[5] R. Bott and H. Samelson, Applications of the theory of Morse to symmetric spaces, Amer. J. Math., 80 (1958), 964–1029.

428 (XIII.17) Total Differential Equations

A. Pfaff's Problem

A **total differential equation** is an equation of the form

$$\omega=0, \qquad (1)$$

where ω is a †differential 1-form $\sum_{i=1}^{n} a_i(x)dx_i$ on a manifold X. A submanifold M of X is called an **integral manifold** of (1) if each vector ξ of the †tangent vector space $T_x(M)$ of M at every point x on M satisfies $\omega(\xi)=0$. We denote the maximal dimension of integral manifolds of (1) by $m(\omega)$. J. F. Pfaff showed that $m(\omega)\geqslant(n-1)/2$ for any ω. The problem of determining $m(\omega)$ for a given form ω is called **Pfaff's problem.** This problem was solved by G. Frobenius, J. G. Darboux, and others as follows: Form an †alternating matrix

$$(a_{ij})_{1\leqslant i,j\leqslant n} \tag{2}$$

from the coefficients of the †exterior derivative of ω,

$$d\omega=\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}(x)dx_i\wedge dx_j,$$

where $a_{ij}=\partial a_j/\partial x_i-\partial a_i/\partial x_j$. Suppose that the rank of (2) is $2t$. Then the rank of the matrix

$$\begin{pmatrix} a_{ij} & -a_i \\ a_j & 0 \end{pmatrix}_{1\leqslant i,j\leqslant n}$$

is $2t$ or $2t+2$. In the former case $m(\omega)=n-t$, and ω can be expressed in the form

$$\sum_{i=1}^{t} u_{2i-1}\,du_{2i}$$

by choosing a suitable coordinate system $(u_1,\ldots,u_n)$. In the latter case $m(\omega)=n-t-1$, and ω can be expressed in the form

$$\sum_{i=1}^{t} u_{2i-1}\,du_{2i}+du_{2t+1}$$

by choosing a suitable coordinate system $(u_1,\ldots,u_n)$. This theorem is called **Darboux's theorem.**

A 1-form ω is called a **Pfaffian form**, and equation (1) is called a **Pfaffian equation.** A system of equations $\omega_i=0$ $(1\leqslant i\leqslant s)$ for 1-form ω_i is called a **system of Pfaffian equations** or a **system of total differential equations** [6, 12, 26].

B. Systems of Differential Forms and Systems of Partial Differential Equations

Let Ω be a system of differential forms ω_i^p, $0\leqslant p\leqslant n$, $1\leqslant i\leqslant v_p$, on X, where ω_i^p is a p-form on X. A submanifold M of X is called an **integral manifold** of $\Omega=0$ if for each p $(0\leqslant p\leqslant \dim M)$, any p-dimensional subspace E_p of $T_x(M)$ satisfies $\omega_i^p(E_p)=0$ $(1\leqslant i\leqslant v_p)$ at every point x on M. Denote the maximal dimension of integral manifolds of $\Omega=0$ by $m(\Omega)$. The problem of determining $m(\Omega)$ for a given system Ω is called the **generalized Pfaff problem,** and will be explained in later sections. By fixing a local coordinate system of X and dividing it into two systems $(x_1,\ldots,x_r)$ and $(y_1,\ldots,y_m)$ $(m=n-r)$, we can consider the problem of finding an integral manifold of $\Omega=0$ defined by

$$y_\alpha=y_\alpha(x_1,\ldots,x_r),\qquad 1\leqslant\alpha\leqslant m.$$

This problem can be reduced to solving a system of partial differential equations of the first order on the submanifold N with the local coordinate system $(x_1,\ldots,x_r)$.

Consider a system of partial differential equations $\Phi=0$ of order l:

$$\varphi_\lambda(x_i,y_\alpha,p_\beta^{j_1\ldots j_r})=0,\qquad 1\leqslant\lambda\leqslant s, \tag{3}$$

with $1\leqslant i\leqslant r$, $1\leqslant\alpha$, $\beta\leqslant m$, $j_1+\ldots+j_r\leqslant l$, where

$$p_\beta^{j_1\ldots j_r}=\frac{\partial^{j_1+\ldots+j_r}y_\beta}{\partial x_1^{j_1}\ldots\partial x_r^{j_r}}. \tag{4}$$

A submanifold defined by $y_\alpha=y_\alpha(x_1,\ldots,x_r)$, $1\leqslant\alpha\leqslant m$, is called a **solution** of $\Phi=0$ if it satisfies (3) identically. The problem of determining whether a given system $\Phi=0$ has a solution was solved by C. Riquier, who showed that any system can be prolonged either to a passive orthonomic system or to an incompatible system by a finite number of steps. A system of partial differential equations is called a **prolongation** of another system if the former contains the latter and they have the same solution. A **passive orthonomic system** is one whose general solution can be parametrized by an infinite number of arbitrary constants. A solution containing parameters is called a **general solution** if by specifying the parameters we can obtain a solution of the †Cauchy problem for any initial data. A system (3) is said to be **incompatible** if it implies a nontrivial relation $f(x_1,\ldots,x_r)=0$ among the x_i.

The problem of solving a system $\Phi=0$ of partial differential equations can be reduced to that of finding integral manifolds of a system of differential forms Σ as follows: Let J^l be a manifold with the local coordinate system

$$(x_i,y_\alpha,p_\beta^{j_1\ldots j_r};\ 1\leqslant i\leqslant r,\ 1\leqslant\alpha,\ \beta\leqslant m,\ j_1+\ldots+j_r\leqslant l),$$

and Σ be a system of 0-forms φ_λ $(1\leqslant\lambda\leqslant s)$ and 1-forms

$$dy_\alpha-\sum_{i=1}^{r} p_\alpha^i dx_i,$$

$$dp_\beta^{j_1\ldots j_r}-\sum_{k=1}^{r} p_\beta^{j_1\ldots j_k+1\ldots j_r}dx_k$$

$(1\leqslant\alpha,\beta\leqslant m, j_1+\ldots+j_r<l)$. Then an integral manifold of $\Sigma=0$ of the form

$$y_\alpha=y_\alpha(x_1,\ldots,x_r),\qquad 1\leqslant\alpha\leqslant m,$$

$$p_\beta^{j_1\ldots j_r}=p_\beta^{j_1\ldots j_r}(x_1,\ldots,x_r),$$

$$1\leqslant\beta\leqslant m,\qquad j_1+\ldots+j_r\leqslant l,$$

gives a solution $y_\alpha = y_\alpha(x_1, \ldots, x_r)$, $1 \leqslant \alpha \leqslant m$, of $\Phi = 0$, and y_β and $p_\beta^{j_1 \ldots j_r}$ satisfy (4).

Conversely, a solution $y_\alpha = y_\alpha(x_1, \ldots, x_r)$, $1 \leqslant \alpha \leqslant m$, of $\Phi = 0$ gives an integral manifold of $\Sigma = 0$ if we define $p_\beta^{j_1 \ldots j_r}(x_1, \ldots, x_r)$ by (4) [23, 24, 26].

C. Systems of Partial Differential Equations of First Order with One Unknown Function

Consider a system of independent †vector fields on N:

$$L_\lambda = \sum_{i=1}^{r} b_{\lambda i}(x) \frac{\partial}{\partial x_i}, \qquad 1 \leqslant \lambda \leqslant s.$$

We solve a system of inhomogeneous equations

$$L_\lambda y - f_\lambda(x) y - g_\lambda(x) = 0, \qquad 1 \leqslant \lambda \leqslant s, \tag{5}$$

for a given system of $f_\lambda(x)$ and $g_\lambda(x)$. The system (5) is called a **complete system** if each of the expressions

$$[L_\lambda, L_\mu] y - (L_\lambda f_\mu - L_\mu f_\lambda) y - (f_\mu g_\lambda - f_\lambda g_\mu) - (L_\lambda g_\mu - L_\mu g_\lambda), \qquad 1 \leqslant \lambda < \mu \leqslant s, \tag{6}$$

is a linear combination of the left-hand sides of (5), where $[L_\lambda L_\mu]$ means the †commutator of L_λ and L_μ. This condition is called the **complete integrability condition** for (5). Suppose that the homogeneous system

$$L_\lambda y = 0, \qquad 1 \leqslant \lambda \leqslant s, \tag{7}$$

is complete. Then it has a system of †functionally independent solutions $y_1, \ldots, y_{r-s}$, and any solution y of (8) is a function of them: $y = \psi(y_1, \ldots, y_{r-s})$. If the inhomogeneous system (5) is complete, then the homogeneous system (7) is complete. This notion of a complete system is due to Lagrange and was extended to a system of nonlinear equations by Jacobi as follows (→ 324 Partial Differential Equations of First Order C).

Consider a system of nonlinear equations

$$F_\lambda(x_1, \ldots, x_r, y, p_1, \ldots, p_r) = 0, \qquad 1 \leqslant \lambda \leqslant s, \tag{8}$$

where $p_i = \partial y / \partial x_i$. The system (8) is called an **involutory system** if each of $[F_\lambda, F_\mu]$, $1 \leqslant \lambda < \mu \leqslant s$, is a linear combination of $F_1, \ldots, F_s$. Here †Lagrange's bracket $[F, G]$ is defined by

$$[F, G] = \sum_{i=1}^{r} \frac{\partial F}{\partial p_i}\left(\frac{\partial G}{\partial x_i} + p_i \frac{\partial G}{\partial y}\right) - \sum_{i=1}^{r} \frac{\partial G}{\partial p_i}\left(\frac{\partial F}{\partial x_i} + p_i \frac{\partial F}{\partial y}\right).$$

Suppose that the system (8) is involutory and $F_1, \ldots, F_s$ are functionally independent. Then, in general, we can solve the following †Cauchy problem for an $(r-s)$-dimensional submanifold N_{r-s} of N: Given a function f on N_{r-s}, find a solution y of (8) satisfying $y = f$ on N_{r-s}. We can construct a solution by integrating a system of ordinary differential equations called a †characteristic system of differential equations. Hence the solution of these problems may be carried out in the C^∞-category (→ 322 Partial Differential Equations (Methods of Integration) B) [7, 11].

D. Frobenius's Theorem

Let X be a †differentiable manifold of class C^∞ and Ω be a system of independent 1-forms ω_i, $1 \leqslant i \leqslant s$, on X. Then the system of Pfaffian equations $\Omega = 0$ is called a **completely integrable system** if at every point x of X,

$$d\omega_i = \sum_{j=1}^{s} \theta_{ij} \wedge \omega_j, \qquad 1 \leqslant i \leqslant s,$$

for 1-forms θ_{ij} on a neighborhood of x. Suppose that $\Omega = 0$ is completely integrable. Then at every point x of X, there exists a local coordinate system $(f_1, \ldots, f_s, x_{s+1}, \ldots, x_n)$ in a neighborhood U of x for which a tangent vector ξ of X at $z \in U$ satisfies $\omega_i(\xi) = 0$, $1 \leqslant i \leqslant s$, if and only if $\xi f_i = 0$, $1 \leqslant i \leqslant s$. In this case, each of the df_i is a linear combination of $\omega_1, \ldots, \omega_s$, and conversely, each of the ω_i is a linear combination of $df_1, \ldots, df_s$. In general, a function f for which df is a linear combination of $\omega_1, \ldots, \omega_s$ is called a **first integral** of $\Omega = 0$.

The theorem of the previous paragraph is called **Frobenius's theorem**, which can be stated in the dual form as follows: Let $D(X)$ be a †subbundle of the †tangent bundle $T(X)$ over X. The mapping $X \ni x \to D_x(X)$ is called a **distribution** on X. It is said to be an **involutive distribution** if at every point x of X we can find a system of independent vector fields L_i $(1 \leqslant i \leqslant s)$ on a neighborhood U of x such that the $L_i(z)$ $(1 \leqslant i \leqslant s)$ form a basis of $D_z(X)$ at every $z \in U$ and satisfy $[L_i, L_j] \equiv 0$ $(L_1, \ldots, L_s)$, $1 \leqslant i < j \leqslant s$, on U. A connected submanifold M of X is called an **integral manifold** of $D(X)$ if $T_x(M) = D_x(X)$ at every point x of M. Suppose that $D(X)$ gives an involutive distribution on X. Then every point x of X is in a maximal integral manifold M that contains any integral manifold including x as a submanifold.

E. Cartan-Kähler Existence Theorems

Let X be a †real analytic manifold. Denote the †sheaf of rings of differential forms on X by $\Lambda(X)$ and its subsheaf of $\mathcal{O}(X)$-modules of p-forms on X by $\Lambda_p(X)$, $1 \leqslant p \leqslant n$, where $\mathcal{O}(X)$ is the sheaf of rings of 0-forms on X. A subsheaf of ideals Σ is called a **differential ideal** if it is generated by Σ_p, $0 \leqslant p \leqslant n$, and contains $d\Sigma$,

where $\Sigma_p = \Sigma \cap \Lambda_p(X)$. Consider a differential ideal Σ on X. Denote the †Grassmann manifold of p-dimensional subspaces of $T_x(X)$ with origin $x \in X$ by $G_p(x)$, and the Grassmann manifold $\bigcup_{x\in X} G(x)$ over X by $G_p(X)$. An element E_p of $G_p(x)$ is called a p-dimensional **contact element** with **origin** x. An element E_p of $G_p(x)$ is called an **integral element** of Σ_p if $\omega(E_p)=0$ at x for any p-form ω in Σ; furthermore, E_p is called an integral element of Σ if any element E_q contained in E_p, $0 \leqslant q \leqslant p$, is an integral element of Σ_q. In particular, 0-dimensional and 1-dimensional integral elements are called **integral points** and **integral vectors**, respectively. It can be proved that an element E_p is an integral element of Σ if and only if it is an integral element of Σ_p. The **polar element** $H(E_p)$ of an integral element E_p with origin x is defined as the subspace of $T_x(X)$ consisting of all vectors that generate with E_p an integral element of Σ. Let $(\Sigma_p)^0$, $0 \leqslant p \leqslant n$, be the subsheaf of $\mathcal{O}(X)$-modules in $\mathcal{O}(G_p(X))$ consisting of all 0-forms

$$\sum_{1\leqslant i_1<\ldots<i_p\leqslant n} a_{i_1\ldots i_p} z_{i_1\ldots i_p}$$

on $G_p(X)$ derived from a p-form

$$\sum_{1\leqslant i_1<\ldots<i_p\leqslant n} a_{i_1\ldots i_p} dx_{i_1}\wedge\ldots\wedge dx_{i_p} \in \Sigma_p,$$

where $z_{i_1\ldots i_p}$ is the †Grassmann coordinate of E_p. An integral element E_p^0 is called a **regular integral element** if the following two conditions are satisfied: (i) $(\Sigma_p)^0$ is a regular local equation of $I\Sigma_p$ at E_p^0, where $I\Sigma_p$ is the set of all integral elements of Σ_p; (ii) $\dim H(E_p) = \text{constant}$ around E_p^0 on $I\Sigma_p$. This definition, due to E. Kähler, is different from that given by E. Cartan [4].

Here, in general, a subsheaf Φ of $\mathcal{O}(X)$ is called a **regular local equation** of $I\Phi$ at an integral point x_0 if there exists a neighborhood U of x_0 and †cross sections $\varphi_1, \ldots, \varphi_s$ of Φ on U that satisfy the following two conditions: (i) $d\varphi_1, \ldots, d\varphi_s$ are linearly independent at every x on U; (ii) a point x of U is an integral point of Φ if and only if $\varphi_1(x) = \ldots = \varphi_s(x) = 0$.

First existence theorem. Suppose that we are given a p-dimensional integral manifold M with a regular integral element $T_x(M)$ at a point x on M. Suppose further that there exists a submanifold F of X containing M such that $\dim F = n - t_{p+1}$, $\dim(T_x(F) \cap H(E_p)) = p+1$, where $E_p = T_x(M)$ and $t_{p+1} = \dim H(E_p) - p - 1$. Then around x there exists a unique integral manifold N such that $\dim N = p+1$ and $F \supset N \supset M$.

This theorem is proved by integrating a system of partial differential equations of Cauchy-Kovalevskaya type. E. Cartan [2–4] also tried to obtain an existence theorem by integrating a system of ordinary differential equations.

A chain of integral elements $E_0 \subset E_1 \subset \ldots \subset E_r$ is called a **regular chain** if each of E_p $(0 \leqslant p < r)$ is a regular integral element. For a regular chain $E_0 \subset E_1 \subset \ldots \subset E_r$, define t_{p+1} by $t_{p+1} = \dim H(E_p) - p - 1$, $0 \leqslant p < r$, and define s_p by $s_p = t_p - t_{p+1} - 1$ $(0 \leqslant p < r)$, $s_r = t_r$, where $t_0 = \dim I\Sigma_0$. Then we have $s_p \geqslant 0$ $(0 \leqslant p \leqslant r)$, $s_0 + \ldots + s_r = t_0 - r$, and we can take a local coordinate system $(x_1, \ldots, x_r, y_1, \ldots, y_m)$, $m = n - r$, around E_0 that satisfies the following four conditions:

(i) $I\Sigma_0$ is defined by $y_{t_0-r+1} = \ldots = y_m = 0$;

(ii) $$H(E_p) = \left\{\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_r}, \frac{\partial}{\partial y_{s_0+\ldots+s_{p-1}+1}}, \ldots, \frac{\partial}{\partial y_{t_0-r}}\right\}, \quad 0 \leqslant p < r;$$

(iii) $$E_p = \left\{\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_p}\right\}, \qquad 1 \leqslant p \leqslant r;$$

(iv) $E_0 = (0, \ldots, 0, 0, \ldots, 0)$.

The integers $s_0, \ldots, s_r$ are called the **characters** of the regular chain $E_0 \subset \ldots \subset E_r$.

Second existence theorem. Suppose that a chain of integral elements $E_0 \subset \ldots \subset E_r$ is regular, and take a local coordinate system satisfying (i)–(iv). Consider a system of initial data

$$f_1, \ldots, f_{s_0},$$

$$f_{s_0+1}(x_1), \ldots, f_{s_0+s_1}(x_1),$$

$$f_{s_0+s_1+1}(x_1, x_2), \ldots, f_{s_0+s_1+s_2}(x_1, x_2),$$

$$\ldots$$

$$f_{s_0+\ldots+s_{r-1}+1}(x_1, \ldots, x_r), \ldots, f_{t_0-r}(x_1, \ldots, x_r).$$

Then if their values and derivatives of the first order are sufficiently small, there exists a unique integral manifold defined by $y_\alpha = y_\alpha(x_1, \ldots, x_r)$, $y_\beta = 0$, $1 \leqslant \alpha \leqslant t_0 - r < \beta \leqslant m$, such that

$$y_\alpha(x_1, \ldots, x_p, 0, \ldots, 0) = f_\alpha(x_1, \ldots, x_p),$$

$$s_0 + \ldots + s_{p-1} < \alpha \leqslant s_0 + \ldots + s_p, \qquad 0 \leqslant p \leqslant r.$$

This theorem is proved by successive application of the first existence theorem. These two theorems are called the **Cartan-Kähler existence theorems.** Σ is said to be **involutive** at an integral element E_r if there exists a regular chain $E_0 \subset \ldots \subset E_r$. An integral manifold possessing a tangent space at which Σ is involutive is called an **ordinary integral manifold** or **ordinary solution** of Σ. An integral manifold that does not possess such a tangent space is called a **singular integral manifold** or **singular solution** of Σ.

Cartan's definition of ordinary and regular integral elements is as follows: An integral point E_0^0 is an ordinary integral point if Σ_0 is a regular local equation of $I\Sigma_0$ at E_0^0. An ordi-

nary integral point E_0^0 is a regular integral point if $\dim H(E_0)$ is constant on $I\Sigma_0$ around E_0^0. Inductively, an integral element E_p^0 is called an **ordinary integral element** if $(\Sigma_p)^0$ is a regular local equation of $I\Sigma_p$ at E_p^0 and E_p^0 contains a regular integral element E_{p-1}^0. An ordinary integral element E_p^0 is a regular integral element (in the sense of Cartan) if $\dim H(E_p)$ is constant on $I\Sigma_p$ around E_p^0. It can be proved that Σ is involutive at an integral element E_r if and only if E_r is an ordinary integral element of Σ. An integral manifold possessing a tangent space that is a regular integral element of Σ is called a **regular integral manifold** or **regular solution** of Σ. Let m_{p+1} be the minimal dimension of $H(E_p)$, where E_p varies over the set of p-dimensional ordinary integral elements, and g be an integer such that $m_p \geqslant p$ $(1 \leqslant p \leqslant g)$ and $m_{g+1} = p$. Then this integer g is called the **genus** of Σ. It is the maximal dimension of ordinary integral manifolds of Σ. However, in general, it is not the maximal dimension of integral manifolds of Σ.

D. C. Spencer and others have been trying to obtain an existence theorem in the C^∞-category analogous to that of Cartan and Kähler. (For a system of linear partial differential equations → [2, 4, 11, 13, 25, 27].)

F. Involutive Systems of Partial Differential Equations

To give a definition of an involutive system of partial differential equations, we define an involutive subspace of $\mathrm{Hom}(V, W)$, where V and W are finite-dimensional vector spaces over the real number field $\mathbf{R}$. Let A be a subspace of $\mathrm{Hom}(V, W)$. For a system of vectors $v_1, \ldots, v_p$ in V, $A(v_1, \ldots, v_p)$ denotes the subspace of A that annihilates $v_1, \ldots, v_p$. Let g_p be the minimal dimension of $A(v_1, \ldots, v_p)$ as $(v_1, \ldots, v_p)$ varies, where $0 \leqslant p \leqslant r = \dim V$. A basis $(v_1, \ldots, v_r)$ of V is called a generic basis if it satisfies $g_p = \dim A(v_1, \ldots, v_p)$ for each p. There exists a generic basis for any A. Let $W \otimes S^2(V^*)$ be the subspace of $\mathrm{Hom}(V, \mathrm{Hom}(V, W))$ consisting of all elements ξ satisfying $\xi(u)v = \xi(v)u$ for any u and v in V. Then the prolongation pA of A is defined by $pA = \mathrm{Hom}(V, A) \cap W \otimes S^2(V^*)$. For any basis $(v_1, \ldots, v_r)$ of V, we have the inequality

$$\dim pA \leqslant \sum_{p=0}^{r} \dim A(v_1, \ldots, v_p).$$

The subspace A is called an **involutive subspace** of $\mathrm{Hom}(V, W)$ if $\dim pA = \sum_{p=0}^{r} g_p$. This notion of an involutive subspace was obtained by V. W. Guillemin and S. Sternberg [13].

A triple $(X, N; \pi)$ consisting of two manifolds X, N and a projection π from X onto N is called a **fibered manifold** if the †differential π_* is surjective at every point of X. Take the set of all mappings f from a domain in N to X satisfying $\pi \circ f = \text{identity}$ for a fibered manifold $(X, N; \pi)$. Then an †l-jet $j_x^l(f)$ is an equivalence class under the equivalence relation defined as follows: $j_x^l(f) = j_u^l(g)$ if and only if $x = u$, $f(x) = g(u)$, and

$$\frac{\partial^{i_1 + \ldots + i_r} f}{\partial x_1^{i_1} \ldots \partial x_r^{i_r}}(x) = \frac{\partial^{i_1 + \ldots + i_r} g}{\partial x_1^{i_1} \ldots \partial x_r^{i_r}}(u),$$

$i_1 + \ldots + i_r \leqslant l$, where $(x_1, \ldots, x_r)$ is a local coordinate system of N around $x = u$ (→ 105 Differentiable Manifolds X).

Denote the space of all l-jets of a fibered manifold $(X, N; \pi)$ by $J^l(X, N; \pi)$ or simply J^l. Then a subsheaf of ideals Φ in $\mathcal{O}(J^l)$ is called a **system of partial differential equations of order** l on N. A point z of J^l is called an integral point of Φ if $\varphi(z) = 0$ for all $\varphi \in \Phi$. The set of all **integral points** of Φ is denoted by $I\Phi$. Let π^l be the natural projection of J^l onto J^{l-1}. Then at a point z of J^l, we can identify $\mathrm{Ker}\, \pi_*^l$ with $\mathrm{Hom}(T_x(N), \mathrm{Ker}\, \pi_*^l)$, where $x = \pi\pi^1 \ldots \pi^l z$. The principal part $C_z(\Phi)$ of Φ is defined as the subspace of $\mathrm{Ker}\, \pi_*^l$ that annihilates Φ. The **prolongation** $p\Phi$ of Φ is defined as the system of order $l+1$ on N generated by Φ and $\partial_k \Phi$, $1 \leqslant k \leqslant \dim N$, where ∂_k is the formal derivative with respect to a coordinate x_k of N:

$$(\partial_k \varphi)(j_x^{l+1}(f)) = \frac{\partial}{\partial x_k} \varphi(j_x^l(f)), \qquad \varphi \in \mathcal{O}(J^l).$$

Let w be an integral point of $p\Phi$ and z be $\pi^{l+1} w$. Then we have the identity

$$pC_z(\Phi) = C_w(p\Phi).$$

The following definition of an involutive system is due to M. Kuranishi [19]: Φ is involutive at an integral point z if the following two conditions are satisfied: (i) Φ is a regular local equation of $I\Phi$ at z; (ii) there exists a neighborhood U of z in J^l such that $(\pi^{l+1})^{-1} U \cap I(p\Phi)$ forms a fibered manifold with base $U \cap I\Phi$ and projection π^{l+1}.

A system of partial differential equations is said to be **involutive** (or **involutory**) if it has an integral point at which it is involutive. Fix a system of independent variables $(y_1, \ldots, y_N)$ in X. Then a system of differential forms is said to be **involutive** (or **involutory**) if it has an integral element at which it is involutive and $dy_1 \wedge \ldots \wedge dy_N \neq 0$. It can be proved that these two definitions of involutive system are equivalent [19, 25].

G. Prolongation Theorems

Cartan gave a method of prolongation by which we can obtain an involutive system from a given system with two independent

variables, if it has a solution. He proposed the following problem: For any $r>2$, construct a method of prolongation by which we can obtain an involutive system from a given system with r independent variables, if it has a solution. To solve this problem, Kuranishi prolonged a given system Φ successively to $p^t\Phi$, $t=1,2,3,\ldots$, and proved the following theorem: Suppose that there exists a sequence of integral points z^t of $p^t\Phi$ with $\pi^{t+l}z^t=z^{t-l}$, $t=1,2,3,\ldots$, that satisfies the following two conditions for each t: (i) $p^t\Phi$ is a regular local equation of $I(p^t\Phi)$ at z^t; (ii) there exists a neighborhood V^t of z^t in $I(p^t\Phi)$ such that $\pi^{t+l}V^t$ contains a neighborhood of z^{t-1} in $I(p^{t-1}\Phi)$ and forms a fibered manifold $(V^t,\pi^{t+l}V^t;\pi^{t+l})$. Then $p^t\Phi$ is involutive at z^t for a sufficiently large integer t.

This prolongation theorem gives a powerful tool to the theory of †infinite Lie groups. However, if we consider a system of partial differential equations of general type, there exist examples of systems that cannot be prolonged to an involutive system by this prolongation, although they have a solution. To improve Kuranishi's prolongation theorem, M. Matsuda [22] defined the prolongation of the same order by $p_0\Phi=p\Phi\cap\mathcal{O}(J^l)$ for a system Φ of order l. This is a generalization of the classical method of completion given by Lagrange and Jacobi. Applying this prolongation successively to a given system Φ, we have $\Psi=\bigcup_{\sigma=1}^{\infty}p_0^{\sigma}\Phi$. Define the p_*-operation by $p_*=\bigcup_{\sigma=1}^{\infty}p_0^{\sigma}p$. Then applying this prolongation successively to Ψ, we have the following theorem: suppose that there exists a sequence of integral points z^t of $p_*^t\Psi$ with $\pi^{l+t}z^t=z^{t-1}$, $t=1,2,3,\ldots$, that satisfies the following two conditions for each t: (i) $p_*^t\Psi$ is a regular local equation of $I(p_*^t\Psi)$ at z^t; (ii) $\dim pC(p_*^t\Psi)$ is constant around z^t on $I(p_*^t\Psi)$. Then $p_*^t\Psi$ is involutive at z^t for a sufficiently large integer t.

To prove this theorem Matsuda applied the following theorem obtained by V. W. Guillemin, S. Sternberg, and J.-P. Serre [25, appendix]: suppose that we are given a subspace A_0 of $\mathrm{Hom}(V,W)$ and subspaces A_t of $\mathrm{Hom}(V,A_{t-1})$ satisfying $A_t\subset pA_{t-1}$, $t=1,2,3,\ldots$. Then A_t is an involutive subspace of $\mathrm{Hom}(V,A_{t-1})$ for a sufficiently large integer t. Thus Cartan's problem was solved affirmatively. To the generalized Pfaff problem these prolongation theorems give another solution, which differs from that obtained by Riquier.

H. Pfaffian Systems in the Complex Domain

Consider a linear system of Pfaffian equations

$$du_i=\sum_{k=1}^{n}\sum_{j=1}^{m}a_{ij}^k(x)u_j\,dx_k,\qquad i=1,\ldots,m,$$

where $x=(x_1,\ldots,x_n)$ is a local coordinate of a complex manifold X and a_{ij}^k are meromorphic functions on X. If we put $u={}^t(u_1,\ldots,u_m)$ and $A^k(x)=(a_{ij}^k(x))$, $k=1,\ldots,n$, the system is written as

$$du=\left(\sum_{k=1}^{n}A^k(x)\,dx_k\right)u. \tag{9}$$

System (9) is completely integrable if and only if

$$\frac{\partial A^j}{\partial x^l}-\frac{\partial A^l}{\partial x^j}=[A^l,A^j],\qquad j,l=1,\ldots,n.$$

Suppose that (9) is completely integrable. If the $A^k(x)$ are holomorphic at $x^0=(x_1^0,\ldots,x_n^0)\in X$, there exists for any $u^0\in\mathbf{C}^m$ one and only one solution of (9) that is holomorphic at x^0 and satisfies $u(x^0)=u^0$. This implies that the solution space of (9) is an m-dimensional vector space; the basis of this space is called a fundamental system of solutions. Therefore any solution is expressible as a linear combination of a fundamental system of solutions and can be continued analytically in a domain where the $A^k(x)$ are holomorphic. A subvariety of X that is the pole set of at least one of the $A^k(x)$ is called a singular locus of (9), and a point on a singular locus is called a singular point.

R. Gérard has given a definition of regular singular points and an analytic expression of a fundamental system of solutions around a regular singular point, and he studied systems of Fuchsian type [8; also 9, 30].

Let $\Omega=\sum_{k=1}^{n}A^k(x)\,dx_k$. Then the system (9) can be rewritten as

$$(d-\Omega)u=0.$$

If we consider a local coordinate (x,u) of a fiber bundle over X, the operator $d-\Omega$ induces a meromorphic linear connection ∇ over X. Starting from this point of view, P. Deligne [5] introduced several important concepts and obtained many results.

The first results for irregular singular points were obtained by Gérard and Y. Sibuya [10], and H. Majima [20] studied irregular singular points of mixed type.

The systems of partial differential equations that are satisfied by the hypergeometric functions of several variables are equivalent to linear systems of Pfaffian equations [1]. This means that such systems of partial differential equations are †holonomic systems. M. Kashiwara and T. Kawai [15] studied holonomic systems with regular singularities from the standpoint of microlocal analysis. Special types of holonomic systems were investigated by T. Terada [28] and M. Yoshida [29].

Consider a system of Pfaffian equations

$$\omega_j=0,\qquad j=1,\ldots,r, \tag{10}$$

where $\omega_j=\sum_{k=1}^{n} a_{jk}(x)dx_k$ and $x=(x_1,\ldots,x_n)$. Suppose that a_{jk} are holomorphic in a domain D of $\mathbf{C}^n$ and that $d\omega_j\wedge\omega_1\wedge\ldots\wedge\omega_r=0$ in D. Denote by S the zero set of $\omega_1\wedge\ldots\wedge\omega_r=0$. A point of S is called a singular point of (10). If the codimension of S is $\geqslant 1$, then system (10) is completely integrable in $D-S$. The following theorem was proved by B. Malgrange [21]: Let $x^0\in S$, and suppose that the codimension of S is $\geqslant 3$ around x^0; then there exist functions $f_j, j=1,\ldots,r$, and $g_{jk}, j, k=1,\ldots,r$, that are holomorphic at x^0 and satisfy $\omega_j=\sum_{k=1}^{r} g_{jk}\,df_k$ and $\det(g_{jk}(x^0))\neq 0$.

References

[1] P. Appell and J. Kampé de Fériet, Fonctions hypergéométriques et hypersphériques, Gauthier-Villars, 1926.
[2] E. Cartan, Sur l'intégration des systèmes d'équations aux différentielles totales, Ann. Sci. Ecole Norm. Sup., 18 (1901), 241–311.
[3] E. Cartan, Leçons sur les invariants intégraux, Hermann, 1922.
[4] E. Cartan, Les systèmes différentielles extérieures et leurs applications géométriques, Hermann, Actualités Sci. Ind., 1945.
[5] P. Deligne, Equations différentielles à points singuliers réguliers, Lecture notes in math. 163, Springer, 1970.
[6] A. R. Forsyth, Theory of differential equations, pt. I. Exact equations and Pfaff's problem, Cambridge Univ. Press, 1890.
[7] A. R. Forsyth, Theory of differential equations, pt. IV. Partial differential equations, Cambridge Univ. Press, 1906.
[8] R. Gérard, Théorie de Fuchs sur une variété analytique complexe, J. Math. Pures Appl., 47 (1968), 321–404.
[9] R. Gérard and A. H. M. Levelt, Sur les connexions à singularités régulières dans le cas de plusieurs variables, Funkcial. Ekvac., 19 (1976), 149–173.
[10] R. Gérard and Y. Sibuya, Etude de certains systèms de Pfaff avec singularités, Lecture notes in math. 712, Springer, 1979.
[11] E. Goursat, Leçons sur l'intégration des équations aux dérivées partielles du premier ordre, Hermann, second edition, 1920.
[12] E. Goursat, Leçons sur le problème de Pfaff, Hermann, 1922.
[13] V. W. Guillemin and S. Sternberg, An algebraic model of transitive differential geometry, Bull. Amer. Math. Soc., 70 (1964), 16–47.
[14] E. Kähler, Einführung in die Theorie der Systeme von Differentialgleichungen, Teubner, 1934.
[15] M. Kashiwara and T. Kawai, On holonomic systems of microdifferential equations IV, systems with regular singularities, Publ. Res. Inst. Math. Sci., 17 (1981), 813–979.
[16] M. Kuranishi, On E. Cartan's prolongation theorem of exterior differential systems, Amer. J. Math., 79 (1957), 1–47.
[17] M. Kuranishi, Lectures on exterior differential systems, Lecture notes, Tata Inst., 1962.
[18] M. Kuranishi, On the local theory of continuous infinite pseudo groups I, II, Nagoya Math. J., 15 (1959), 225–260; 19 (1961), 55–91.
[19] M. Kuranishi, Lectures on involutive systems of partial differential equations, Publ. Soc. Math. São Paulo, 1967.
[20] H. Majima, Asymptotic analysis for integrable connections with irregular singular points, Lecture notes in math. 1075, Springer, 1984.
[21] B. Malgrange, Frobenius avec singularités II, Le cas général, Inventiones Math., 39 (1976), 67–89.
[22] M. Matsuda, Cartan-Kuranishi's prolongation of differential systems combined with that of Lagrange and Jacobi, Publ. Res. Inst. Math. Sci., 3 (1967/68), 69–84.
[23] C. Riquier, Les systèmes d'équations aux dérivées partielles, Gauthier-Villars, 1910.
[24] J. F. Ritt, Differential algebra, Amer. Math. Soc. Colloq. Publ., 1950.
[25] I. M. Singer and S. Sternberg, The infinite groups of Lie and Cartan I. The transitive groups, J. Analyse Math., 15 (1965), 1–114.
[26] J. A. Schouten and W. v. d. Kulk, Pfaff's problem and its generalizations, Clarendon Press, 1949.
[27] D. C. Spencer, Overdetermined systems of linear partial differential equations, Bull. Amer. Math. Soc., 75 (1969), 179–239.
[28] T. Terada, Problème de Rieman et fonctions automorphes provenant des fonctions hypergéométriques de plusieurs variables, J. Math. Kyoto Univ., 13 (1973), 557–578.
[29] M. Yoshida, Local theory of Fuchsian systems with certain discrete monodromy groups I, II, Funkcial. Ekvac., 21 (1978), 105–137, 203–221.
[30] M. Yoshida and K. Takano, On a linear system of Pfaffian equations with regular singular point, Funkcial. Ekvac., 19 (1976), 175–189.

429 (XI.6) Transcendental Entire Functions

A. General Remarks

An **entire function** (or **integral function**) $f(z)$ is a complex-valued function of a complex variable

z that is holomorphic in the finite z-plane, $z \neq \infty$. If $f(z)$ has a pole at ∞, then $f(z)$ is a polynomial in z. A polynomial is called a **rational entire function**. If an entire function is bounded, it is constant (†Liouville's theorem). A **transcendental entire function** is an entire function that is not a polynomial, for example, $\exp z$, $\sin z$, $\cos z$. An entire function can be developed in a power series $\sum_{n=0}^{\infty} a_n z^n$ with infinite radius of convergence. If $f(z)$ is a transcendental entire function, this is actually an infinite series.

B. The Order of an Entire Function

If a transcendental entire function $f(z)$ has a zero of order m ($m \geqslant 0$) at $z=0$ and other zeros at $\alpha_1, \alpha_2, \ldots, \alpha_n, \ldots$ $(0<|\alpha_1| \leqslant |\alpha_2| \leqslant |\alpha_3| \leqslant \ldots \to \infty)$, multiple zeros being repeated, then $f(z)$ can be written in the form

$$f(z)=e^{g(z)} z^m \prod_{k=1}^{\infty}\left(1-\frac{z}{\alpha_k}\right) e^{g_k(z)},$$

where $g(z)$ is an entire function, $g_k(z)=(z/\alpha_k)+(1/2)(z/\alpha_k)^2+(1/3)(z/\alpha_k)^3+\ldots+(1/p_k)(z/\alpha_k)^{p_k}$, and $p_1, p_2, \ldots$ are integers with the property that $\sum_{k=1}^{\infty}|z/\alpha_k|^{p_k+1}$ converges for all z (**Weierstrass's canonical product**).

E. N. Laguerre introduced the concept of the genus of a transcendental entire function $f(z)$. Assume that there exists an integer p for which $\sum_{k=0}^{\infty}|\alpha_k|^{-(p+1)}$ converges, and take the smallest such p. Assume further that in the representation for $f(z)$ in the previous paragraph, when $p_1=p_2=\ldots=p$, the function $g(z)$ reduces to a polynomial of degree q; then $\max(p, q)$ is called the **genus** of $f(z)$. For transcendental entire functions, however, the order is more essential than the genus. The **order** ρ of a transcendental entire function $f(z)$ is defined by

$$\rho=\limsup_{r \to \infty} \log\log M(r)/\log r,$$

where $M(r)$ is the maximum value of $|f(z)|$ on $|z|=r$. By using the coefficients of $f(z)=\sum a_n z^n$, we can write

$$\rho=\limsup_{n \to \infty} n \log n/\log(1/|a_n|).$$

The entire functions of order 0, which were studied by Valiron and others, have properties similar to polynomials, and the entire functions of order less than 1/2 satisfy $\lim_{r_n \to \infty} \min_{|z|=r_n}|f(z)|=\infty$ for some increasing sequence $r_n \uparrow \infty$ (**Wiman's theorem**). Hence entire functions of order less than 1/2 cannot be bounded in any domain extending to infinity. Among the functions of order greater than 1/2 there exist functions bounded in a given angular domain $D: \alpha<\arg z<\alpha+\pi/\mu$. If $|f(z)| < \exp r^{\rho}$ $(\rho<\mu)$ and $f(z)$ is bounded on the boundary of D, then $f(z)$ is bounded in the angular domain (→ 272 Meromorphic Functions). In particular, if the order ρ of $f(z)$ is an integer p, then it is equal to the genus, and $g(z)$ reduces to a polynomial of degree $\leqslant p$ (J. Hadamard). These theorems originated in the study of the zeros of the †Riemann zeta function and constitute the beginning of the theory of entire functions.

There is some difference between the properties of functions of integral order and those of others. Generally, the point z at which $f(z)=w$ is called a **w-point** of $f(z)$. If $\{z_n\}$ consists of w-points different from the origin, the infimum $\rho_1(w)$ of k for which $\sum 1/|z_n|^k$ converges is called the **exponent of convergence** of $f-w$. If the order ρ of an entire function is integral, then $\rho_1(w)=\rho$ for each value w with one possible exception, and if ρ is not integral, then $\rho_1(w)=\rho$ for all w (É. Borel). Therefore any transcendental entire function has an infinite number of w-points for each value w except for at most one value, called an **exceptional value** of $f(z)$ (**Picard's theorem**). In particular, $f(z)$ has no exceptional values if ρ is not integral. For instance, $\sin z$ and $\cos z$ have no exceptional values, while e^z has 0 as an exceptional value. Since transcendental entire functions have no poles, ∞ can be counted as an exceptional value. Then we must change the statement in Picard's theorem to "except for at most two values." Since the theorem was obtained by E. Picard in 1879, problems of this type have been studied intensively (→ 62 Cluster Sets, 272 Meromorphic Functions).

After Picard proved the theorem by using the inverse of a †modular function, several alternative proofs were given. For instance, there is a proof using the Landau-Schottky theorem and †Bloch's theorem and one using †normal families. Picard's theorem was extended to meromorphic functions and has also been studied for analytic functions defined in more general domains. There are many fully quantitative results, too. For instance, Valiron [3] gave such results by performing some calculations on neighborhoods of points where entire functions attain their maximum absolute values.

Thereafter, the distribution of w-points in a neighborhood of an essential singularity was studied by many people, and in 1925 the Nevanlinna theory of meromorphic functions was established. The core of the theory consists of two fundamental theorems, †Nevanlinna's first and second fundamental theorems (→ 272 Meromorphic Functions). Concerning composite entire functions $F(z)=f(g(z))$, Pólya proved the following fact: The finiteness of the order of F implies that the order of f should

be zero unless g is a polynomial. This gives the starting point of the factorization theory, on which several people have been working recently. Several theorems in the theory of meromorphic functions can be applied to the theory. One of the fundamental theorems is the following: Let $F(z)$ be an entire function, which admits the factorizations $F(z)=P_m(f_m(z))$ with a polynomial P_m of degree m and an entire function f_m for all integers m. Then $F(z)=A\cos\sqrt{H(z)}+B$ unless $F(z)=A\exp H(z)+B$. Here, H is a nonconstant entire function and A, B are constant, $A\neq 0$.

C. Julia Directions

Applying the theory of †normal families of holomorphic functions, G. Julia proved the existence of Julia directions as a precise form of Picard's theorem [5]. A transcendental entire function $f(z)$ has at least one direction $\arg z=\theta$ such that for any $\varepsilon>0$, $f(z)$ takes on every (finite) value with one possible exception infinitely often in the angular domain $\theta-\varepsilon<\arg z<\theta+\varepsilon$. This direction $\arg z=\theta$ is called a **Julia direction** of $f(z)$.

D. Asymptotic Values

†Asymptotic values, †asymptotic paths, etc., are defined for entire functions as for meromorphic functions. In relation to †Iversen's theorem and †Gross's theorem for inverse functions and results on †cluster sets, †ordinary singularities of inverse functions hold for entire functions in the same way as for meromorphic functions. Also, as for meromorphic functions, †transcendental singularities of inverse functions are divided into two classes, the †direct and the †indirect transcendental singularities.

The exceptional values in Picard's theorem are asymptotic values of the functions, and ∞ is an asymptotic value of any transcendental entire function. Therefore $f(z)\rightarrow\infty$ along some curve extending to infinity. Between the asymptotic paths corresponding to two distinct asymptotic values, there is always an asymptotic path with asymptotic value ∞. By †Bloch's theorem, A. Bloch showed that the †Riemann surface of the inverse function of a transcendental entire function contains a disk with arbitrarily large radius. Denjoy conjectured in 1907 that $\mu\leqslant 2\rho$, where ρ is the order of an entire function and μ is the number of distinct finite asymptotic values of the function, and L. V. Ahlfors gave the first proof (1929). This result contains Wiman's theorem. There are transcendental entire functions with $\mu=2\rho$. It was shown by W. Gross that among entire functions of infinite order there exists an entire function having every value as its asymptotic value.

References

[1] E. C. Titchmarsh, The theory of functions, Clarendon Press, second edition, 1939.
[2] R. P. Boas, Entire functions, Academic Press, 1954.
[3] G. Valiron, Lectures on the general theory of integral functions, Librairie de l'Université, Deighton, Bell and Co., 1923 (Chelsea, 1949).
[4] L. Bieberbach, Lehrbuch der Funktionentheorie II, Teubner, 1931 (Johnson Reprint Co., 1969).
[5] G. Julia, Sur quelques propriétés nouvelles des fonctions entières ou méromorphes, Ann. Sci. Ecole Norm. Sup., (3) 36 (1919), 93–125.

430 (V.11) Transcendental Numbers

A. History

A complex number α is called a **transcendental number** if α is not †algebraic over the field of rational numbers **Q**. C. Hermite showed in 1873 that e is a transcendental number. Following a similar line of thought as that taken by Hermite, C. L. F. Lindemann showed that π is also transcendental (1882). Among the 23 problems posed by D. Hilbert in 1900 (→ 196 Hilbert), the seventh was the problem of establishing the transcendence of certain numbers (e.g., $2^{\sqrt{2}}$). This stimulated fruitful investigations by A. O. Gel'fond, T. Schneider, C. L. Siegel, and others. The theory of transcendental numbers is, however, far from complete. There is no general criterion that can be utilized to characterize transcendental numbers. For example, neither the transcendence nor even the irrationality of the †Euler constant $C=\lim_{n\to\infty}(1+1/2+\ldots+1/n-\log n)$ has been established. A survey of the development of the theory of transcendental numbers can be found in [18], in which an extensive list of relevant publications up to 1966 is given.

B. Construction of Transcendental Numbers

Let $\bar{\mathbf{Q}}$ be the field of †algebraic numbers. Suppose that α is an element of $\bar{\mathbf{Q}}$ that satisfies the irreducible equation $f(x)=a_0x^n+a_1x^{n-1}+\ldots+a_n=0$, where the a_i are rational integers, $a_0\neq 0$, and $a_0, a_1, \ldots, a_n$ have no common factors. Then we define $H(\alpha)$ to be the maxi-

mum of $|a_i|$ ($i=0,\ldots,n$) and call it the **height** of α. J. Liouville proved the following theorem (1844): Let ξ be a real number ($\xi\notin\mathbf{Q}$). If $\inf\{q^n|\xi-p/q|\,|\,p/q\in\mathbf{Q}\}=0$ for any positive integer n, then ξ is transcendental.

Transcendental numbers having this property are called **Liouville numbers**. Examples are: (i) $\xi=\sum_{\nu=1}^{\infty}g^{-\nu!}$, where g is an integer not smaller than 2. (ii) Suppose that we are given a sequence $\{n_k\}$ of positive integers such that $n_k\to\infty$ ($k\to\infty$). Let ξ be the real number expressed as an †infinite simple continued fraction $b_0+1/b_1\dotplus 1/b_2\dotplus\ldots$. Let B_l be the denominator of the lth †convergent of the continued fraction. If $b_{n_k+1}\geqslant B_{n_k}^{n_k-2}$ for $k\geqslant 1$, then ξ is a Liouville number.

On the other hand, K. Mahler [8, 9] proved the existence of transcendental numbers that are not Liouville numbers. For example, he showed that if $f(x)$ is a nonconstant integral polynomial function mapping the set of positive integers into itself, then a number ξ expressed, e.g., in the decimal system as $0y_1y_2y_3\ldots$ is such a number if we put $y_n=f(n)$, $n=1,2,3,\ldots$. (In particular, from $f(x)=x$ we get the non-Liouville transcendental number $\xi=0.123456789101112\ldots$.) Mahler proved this result by using †Roth's theorem (1955) (→ 182 Geometry of Numbers). Both Liouville and Mahler utilized the theory of †Diophantine approximation to construct transcendental numbers.

On the other hand, Schneider [10–12] and Siegel [3] constructed transcendental numbers using certain functions. Examples are: $\exp\alpha$ ($\alpha\in\bar{\mathbf{Q}}$, $\alpha\neq 0$); α^{β} ($\alpha\in\bar{\mathbf{Q}}$, $\alpha\neq 0,1$; $\beta\in\bar{\mathbf{Q}}-\mathbf{Q}$); $J(\tau)$, where J is the †modular function and τ is an algebraic number that is not contained in any imaginary quadratic number field; $\wp(2\pi i/\alpha)$, where $\wp$ is the Weierstrass $\wp$-function, $\alpha\in\bar{\mathbf{Q}}$, and $\alpha\neq 0$; and $B(p,q)$, where B is the †Beta function and $p,q\in\mathbf{Q}-\mathbf{Z}$.

Since $e=\exp 1$ and $1=\exp 2\pi i$, the transcendence of e and π is directly implied by the transcendence of $\exp\alpha$ ($\alpha\in\bar{\mathbf{Q}}$, $\alpha\neq 0$).

C. Classification of Transcendental Numbers

(1) Mahler's classification: Given a complex number ξ and positive integers n and H, we consider the following:

$$w_n(H,\xi)=\min\left\{\left|\sum_{\nu=0}^{n}a_\nu\xi^\nu\right|\;\middle|\;a_\nu\in\mathbf{Z},\ |a_\nu|\leqslant H,\ \sum_{\nu=0}^{n}a_\nu\xi^\nu\neq 0\right\},$$

$$w_n(\xi)=w_n=\limsup_{H\to\infty}(-\log w_n(H,\xi)/\log H),$$

$$w(\xi)=w=\limsup_{n\to\infty}w_n(\xi)/n,$$

and let $\mu=$ the first number n for which w_n is ∞. Then we have the following four cases: (i) $w=0$, $\mu=\infty$; (ii) $0<w<\infty$, $\mu=\infty$; (iii) $w=\mu=\infty$; (iv) $w=\infty$, $\mu<\infty$, corresponding to which we call ξ an **A-number**, **S-number**, **T-number**, or **U-number**. The set of A-numbers is denoted by **A**, and similarly we have the classes **S**, **T**, and **U**. It is known that $\mathbf{A}=\bar{\mathbf{Q}}$. If two numbers ξ and η are †algebraically dependent over **Q**, then they belong to the same class. If ξ belongs to **S**, the quantity $\theta(\xi)=\sup\{w_n(\xi)/n\,|\,n=1,2,\ldots\}$ is called the **type** of ξ (in the sense of Mahler). Mahler conjectured that almost all transcendental numbers (except a set of Lebesgue measure zero) are S-numbers of the type 1 or 1/2 according as they belong to **R** or not. Various results were obtained concerning this conjecture (W. J. LeVeque, J. F. Koksma, B. Volkmann) until it was proved by V. G. Sprindzhuk in 1965 [14, 15]. The existence of T-numbers was proved by W. M. Schmidt (1968) [16]. All Liouville numbers are U-numbers [7]. On the other hand, $\log\alpha$ ($\alpha\in\mathbf{Q}$, $\alpha>0$, $\alpha\neq 1$) and π are transcendental numbers that do not belong to **U**.

(2) Koksma's classification: For a given transcendental number ξ and positive numbers n and H, we consider the following:

$$w_n^*(H,\xi)=\min\{|\xi-\alpha|\,|\,\alpha\in\bar{\mathbf{Q}},\ H(\alpha)\leqslant H,\ [\mathbf{Q}(\alpha):\mathbf{Q}]\leqslant n\},$$

$$w_n^*(\xi)=w_n^*=\limsup_{H\to\infty}(-\log(Hw_n^*(H,\xi))/\log H),$$

$$w^*(\xi)=w^*=\limsup_{n\to\infty}w_n^*(\xi)/n,$$

and let $\mu^*=$ the first number n for which w_n^* is ∞. Then we have the following three cases: (i) $w^*<\infty$, $\mu^*=\infty$; (ii) $w^*=\mu^*=\infty$; (iii) $w^*=\infty$, $\mu^*<\infty$. We call ξ an **S*-number**, **T*-number**, or **U*-number** according as (i), (ii), or (iii) holds and denote the set of S*-numbers by **S***, etc. If ξ belongs to **S***, we call $\theta^*(\xi)=\sup\{w_n^*(\xi)/n\,|\,n=1,2,\ldots\}$ the **type** of ξ (in the sense of Koksma). It can be shown that $\mathbf{S}=\mathbf{S}^*$, $\mathbf{T}=\mathbf{T}^*$, and $\mathbf{U}=\mathbf{U}^*$, and that if $\xi\in\mathbf{S}$, then $\theta^*(\xi)\leqslant\theta(\xi)\leqslant\theta^*(\xi)+1$.

D. Algebraic Independence

Concerning the algebraic relations of transcendental numbers, we have the following three principal theorems:

(1) Let $\alpha_1,\ldots,\alpha_m$ be elements of $\bar{\mathbf{Q}}$ that are linearly independent over **Q**. Then $\exp\alpha_1,\ldots,\exp\alpha_m$ are transcendental and algebraically independent over $\bar{\mathbf{Q}}$ (**Lindemann-Weierstrass theorem**).

(2) Let $J_0(x)$ be the †Bessel function and α a nonzero algebraic number. Then $J_0(\alpha)$ and $J_0'(\alpha)$ are transcendental and algebraically independent over **Q** (Siegel).

(3) Let $\alpha_1, \ldots, \alpha_n$ be nonzero elements of $\bar{\mathbf{Q}}$ such that $\log\alpha_1, \ldots, \log\alpha_n$ are linearly independent over $\mathbf{Q}$. Then $1, \log\alpha_1, \ldots, \log\alpha_n$ are linearly independent over $\bar{\mathbf{Q}}$ (A. Baker).

Besides these theorems, various related results have been obtained by A. B. Shidlovskiĭ, Gel'fond, N. I. Fel'dman, and others. A quantitative extension of theorem (3), also by Baker, will be discussed later.

First we give more detailed descriptions of theorems (1) and (2). Let $\alpha_1, \ldots, \alpha_m$ be as in theorem (1), $s = [\mathbf{Q}(\alpha_1, \ldots, \alpha_m):\mathbf{Q}]$, $P(X_1, \ldots, X_m)$ be an arbitrary polynomial in $\bar{\mathbf{Q}}[X_1, \ldots, X_m]$ of degree n, and $H(P)$ be the maximum of the absolute values of the coefficients of the polynomial P. Then there exists a positive number C determined only by the numbers $\alpha_1, \ldots, \alpha_m$ and $n(=\deg P)$ such that

$$|P(e^{\alpha_1}, \ldots, e^{\alpha_m})| > CH(P)^{-2s\left(2\left(\frac{2smn+m+n}{m}\right)-1\right)}.$$

In particular, if α is a nonzero algebraic number, then $\exp\alpha$ belongs to $\mathbf{S}$ and $\theta(\exp\alpha) \leqslant 8s^2 + 6s$.

(2′) Let α be a nonzero algebraic number, $s = [\mathbf{Q}(\alpha):\mathbf{Q}]$, $P \in \mathbf{Q}[X_1, X_2]$, $\deg P = n$. Then there exists a positive number C determined only by α and n such that $|P(J_0(\alpha), J_0'(\alpha))| > CH(P)^{-82s^3n^3}$.

Theorems (1) and (2) are actually special cases of a theorem obtained by Siegel. To state this theorem, the following terminology is used: An entire function $f(z) = \sum_{n=0}^{\infty} C_n \cdot z^n/n!$ is called an **E-function** defined over an †algebraic number field K of finite degree if the following three conditions are satisfied: (i) $C_n \in K$ $(n = 0, 1, 2, \ldots)$. (ii) For any positive number ε, $C_n = O(n^{\varepsilon n})$. (iii) Let q_n be the least positive integer such that $C_k q_n$ belongs to the ring $\mathfrak{O}$ of algebraic integers in K $(0 \leqslant n, 0 \leqslant k \leqslant n)$. Then for an arbitrary positive number ε, $q_n = O(n^{\varepsilon n})$.

A system $\{f_1(z), \ldots, f_m(z)\}$ of E-functions defined over K is said to be **normal** if it satisfies the following two conditions: (i) None of the functions $f_i(z)$ is identically zero. (ii) If the functions $w_k = f_k(z)$ $(k = 1, \ldots, m)$ satisfy a system of †homogeneous linear differential equations of the first order, then $w_k' = \sum_{l=1}^{m} Q_{kl}(z) w_l$, where the $Q_{kl}(z)$ are rational functions of z, with coefficients in the ring $\mathfrak{O}$. The matrix (Q_{kl}) can be decomposed by rearranging the order of the indices k, l if necessary into the form

$$\begin{pmatrix} W_1 & & 0 \\ & \ddots & \\ 0 & & W_r \end{pmatrix},$$

where

$$W_t = \begin{bmatrix} Q_{11,t} & \cdots & Q_{1m_t,t} \\ \cdots & \cdots & \cdots \\ Q_{m_t1,t} & \cdots & Q_{m_tm_t,t} \end{bmatrix}, \quad 1 \leqslant t \leqslant r, \quad \sum_{t=1}^{n} m_t = m.$$

The decomposition is unique if we choose r as large as possible, in which case we call $W_1, \ldots, W_r$ the primitive parts of (Q_{kl}). The requirement is that the primitive parts W_i are independent in the following sense: If there are numbers $C_{st} \in K$ and polynomial functions $P_{kt}(z) \in K[z]$ such that

$$\sum_{t=1}^{r} (C_{1t} \ldots C_{m_t t}) W_t \begin{bmatrix} P_{1t}(z) \\ \vdots \\ P_{m_t t}(z) \end{bmatrix} = 0,$$

then $C_{st} = 0$, $P_{kl}(z) = 0$.

Let N be a positive integer. A normal system $\{f_1(z), \ldots, f_m(z)\}$ of E-functions is said to be of degree N if the system $\{F_{n_1, \ldots, n_m}(z) = f_1(z)^{n_1} \ldots f_m(z)^{n_m} \mid n_i \geqslant 0, \sum_{i=1}^{m} n_i \leqslant N\}$ is also a normal system of E-functions. Then the theorem obtained by Siegel [4] is: Let N be an arbitrary positive integer and $\{f_1(z), \ldots, f_m(z)\}$ be a normal system of E-functions of degree N defined over an algebraic number field of finite degree K satisfying the system of differential equations $f_k'(z) = \sum_{l=1}^{m} Q_{kl}(z) f_l(z)$, where $Q_{kl}(z) \in \mathfrak{O}(z)$, $1 \leqslant k \leqslant m$. If α is a nonzero algebraic number that is not a †pole of any one of the functions $Q_{kl}(z)$, then $f_1(\alpha), \ldots, f_m(\alpha)$ are transcendental numbers that are algebraically independent over the field $\bar{\mathbf{Q}}$.

Theorem (3) at the beginning of this section implies, for example, the following: (i) If $\alpha_1, \ldots, \alpha_n$ and $\beta_1, \ldots, \beta_n$ all belong to $\bar{\mathbf{Q}}$ and $\gamma = \alpha_1 \log\beta_1 + \ldots + \alpha_n \log\beta_n \neq 0$, then γ is transcendental. (ii) If $\alpha_1, \ldots, \alpha_n, \beta_0, \beta_1, \ldots, \beta_n$ are nonzero algebraic numbers, then $e^{\beta_0}\alpha_1^{\beta_1} \ldots \alpha_n^{\beta_n}$ is transcendental. (iii) If $\alpha_1, \ldots, \alpha_n$ are algebraic numbers other than 0 and 1, and $\beta_1, \ldots, \beta_n$ also belong to $\bar{\mathbf{Q}}$, with $1, \beta_1, \ldots, \beta_n$ linearly independent over $\mathbf{Q}$, then $\alpha_1^{\beta_1} \ldots \alpha_n^{\beta_n}$ is transcendental.

Baker [17] also obtained a quantitative extension of theorem (3): Suppose that we are given integers $A \geqslant 4$, $d \geqslant 4$ and nonzero algebraic numbers $\alpha_1, \ldots, \alpha_n$ $(n \geqslant 2)$ whose heights and degrees do not exceed A and d, respectively. Suppose further that $0 < \delta \leqslant 1$, and let $\log\alpha_1, \ldots, \log\alpha_n$ be the principal values of the logarithms. If there exist rational integers $b_1, \ldots, b_n$ with absolute value at most H such that

$$0 < |b_1 \log\alpha_1 + \ldots + b_n \log\alpha_n| < e^{-\delta H},$$

then

$$H < (4^{n^2}\delta^{-1} d^{2n} \log A)^{(2n+1)^2}.$$

This theorem has extensive applications in various problems of number theory, including a wide class of †Diophantine problems [19].

A number of new, interesting results on the algebraic independence of values of exponential functions, elliptic functions, and some other special functions have been obtained

recently by D. Masser, G. V. Chudnovskiĭ, M. Waldschmidt, and other writers. In particular, Chudnovskiĭ (1975) obtained the remarkable result that $\Gamma(1/3)$ and $\Gamma(1/4)$ are transcendental numbers. See [20–24].

References

[1] A. O. Gel'fond, Transcendental and algebraic numbers, Dover, 1960. (Original in Russian, 1952.)
[2] T. Schneider, Einführung in die transzendenten Zahlen, Springer, 1957.
[3] C. L. Siegel, Über einige Anwendungen diophantischer Approximationen, Abh. Preuss. Akad. Wiss., 1929 (Gesammelte Abhandlungen, Springer, 1966, vol. 1, 209–266).
[4] C. L. Siegel, Transcendental numbers, Ann. Math. Studies, Princeton Univ. Press, 1949.
[5] N. I. Fel'dman, Approximation of certain transcendental numbers I, Amer. Math. Soc. Transl., 59 (1966), 224–245. (Original in Russian, 1951.)
[6] W. J. LeVeque, Note on S-numbers, Proc. Amer. Math. Soc., 4 (1953), 189–190.
[7] W. J. LeVeque, On Mahler's U-numbers, J. London Math. Soc., 28 (1953), 220–229.
[8] K. Mahler, Arithmetische Eigenschaften einer Klasse von Dezimalbrüchen, Proc. Acad. Amsterdam, 40 (1937), 421–428.
[9] K. Mahler, Über die Dezimalbruchentwicklung gewisser Irrationalzahlen, Mathematica B, Zutphen, 6 (1937), 22–36.
[10] T. Schneider, Transzendenzuntersuchungen periodischer Funktionen I, J. Reine Angew. Math., 172 (1935), 65–69.
[11] T. Schneider, Arithmetische Untersuchungen elliptischer Integrale, Math. Ann., 113 (1936), 1–13.
[12] T. Schneider, Zur Theorie der Abelschen Funktionen und Integrale, J. Reine Angew. Math., 183 (1941), 110–128.
[13] S. Lang, Introduction to transcendental numbers, Addison-Wesley, 1966.
[14] V. G. Sprindzhuk, A proof of Mahler's conjecture on the measure of the set of S-numbers, Amer. Math. Soc. Transl., 51 (1966), 215–272. (Original in Russian, 1965.)
[15] V. G. Sprindzhuk, Mahler's problem in metric number theory, Amer. Math. Soc. Transl. Math. Monographs, 1969. (Original in Russian, 1967.)
[16] W. M. Schmidt, T-numbers do exist, Rend. Convegno di Teoria dei Numeri, Roma, 1968.
[17] A. Baker, Linear forms in the logarithms of algebraic numbers I, II, III, IV, Mathematika, 13 (1966), 204–216; 14 (1967), 102–107; 14 (1967), 220–228; 15 (1968), 204–216.
[18] N. I. Fel'dman and A. B. Shidlovskiĭ, The development and present state of the theory of transcendental numbers, Russian Math. Surveys, 22 (1967), 1–79. (Original in Russian, 1967.)
[19] A. Baker, Transcendental number theory, Cambridge Univ. Press, 1975.
[20] D. Masser, Elliptic functions and transcendence, Lecture notes in math. 437, Springer, 1975.
[21] G. V. Chudnovskiĭ, Algebraic independence of values of exponential and elliptic functions, Proc. Int. Congr. Math., Helsinki, 1978.
[22] A. Baker and D. Masser (eds.), Transcendence theory: Advances and applications, Academic Press, 1977.
[23] M. Waldschmidt, Transcendence methods, Queen's papers in pure and appl. math. 52, Queen's Univ., 1979.
[24] M. Waldschmidt, Nombres transcendants et groups algébriques, Astérisque, 69–70 (1980).

431 (IX.19) Transformation Groups

A. Topological Transformation Groups

Let G be a group, M a set, and f a mapping from $G \times M$ into M. Put $f(g,x)=g(x)$ ($g \in G$, $x \in M$). Then the group G is said to be a **transformation group** of the set M if the following two conditions are satisfied: (i) $e(x)=x$ ($x \in M$), where e is the identity element of G; and (ii) $(gh)(x)=g(h(x))$ ($x \in M$) for any $g, h \in G$. In this case the mapping $x \rightarrow g(x)$ is a one-to-one mapping of M onto itself.

Let G be a transformation group of M. If G is a topological group, M a topological space, and the mapping $(g,x) \rightarrow g(x)$ a continuous mapping from $G \times M$ into M, then G is called a **topological transformation group** of M. In this case $x \rightarrow g(x)$ is a homeomorphism of M onto itself. The mapping $(g,x) \rightarrow g(x)$ is called an **action** of G on M. The space M, together with a given action of G, is called a **G-space**.

For a point x of M, the set $G(x)=\{g(x) \mid g \in G\}$ is called the **orbit** of G passing through the point x. Defining as equivalent two points x and y of M belonging to the same orbit, we get an equivalence relation in M. The quotient space of M by this equivalence relation, denoted by M/G, is called the **orbit space** of G-space M.

If $G(x)=\{x\}$, then x is called a **fixed point**. The set of all fixed points is denoted by M^G. For a point x of M, the set $G_x=\{g \in G \mid g(x)=$

$x\}$ is a subgroup of G called the **isotropy subgroup (stabilizer, stability subgroup)** of G at the point x. A conjugacy class of the subgroup G_x is called an **isotropy type** of the transformation group G on M.

The group G is said to act **nontrivially** (resp. **trivially**) on M if $M \neq M^G$ (resp. $M = M^G$). The group G is said to act **freely** on M if the isotropy subgroup G_x consists only of the identity element for any point x of M.

The group G is said to act **transitively** on M if for any two points x and y of M, there exists an element $g \in G$ such that $g(x) = y$.

Let N be the set of all elements $g \in G$ such that $g(x) = x$ for all points x of M. Then N is a normal subgroup of G. If N consists only of the identity element e, we say that G acts **effectively** on M, and if N is a discrete subgroup of G, we say that G acts **almost effectively** on M. When $N \neq \{e\}$, the quotient topological group G/N acts effectively on M in a natural fashion.

An **equivariant mapping (equivariant map)** (or a **G-mapping, G-map**) $h: X \rightarrow Y$ between G-spaces is a continuous mapping which commutes with the group actions, that is, $h(g(x)) = g(h(x))$ for all $g \in G$ and $x \in X$. An equivariant mapping which is also a homeomorphism is called an equivalence of G-spaces.

For a G-space M, an equivalence class of the G-spaces $G(x)$, $x \in M$, is called an **orbit type** of the G-space M.

B. Cohomological Properties

We consider only †paracompact G-spaces and †Čech cohomology theory in this section. We shall say that a topological space X is **finitistic** if every open covering has a finite-dimensional refinement. The following theorems are useful [1–3].

(1) If G is finite, X a finitistic paracompact G-space, and K a field of characteristic zero or prime to the order of G, then the induced homomorphism $\pi^*: H^*(X/G; K) \rightarrow H^*(X; K)^G$ is an isomorphism. Here, π is a natural projection of X onto X/G. The group G acts naturally on $H^*(X; K)$, and $H^*(X; K)^G$ denotes the fixed-point set of this G-action.

(2) Let X be a finitistic G-space and G cyclic of prime order p. Then, with coefficients in $\mathbf{Z}/p\mathbf{Z}$, we have

(a) for each n $\quad \sum_{i=n}^{\infty} \operatorname{rank} H^i(X^G) \leqslant \sum_{i=n}^{\infty} \operatorname{rank} H^i(X)$,

(b) $\quad \chi(X) + (p-1)\chi(X^G) = p\chi(X/G)$.

Here the †Euler-Poincaré characteristics $\chi(\ )$ are defined in terms of $\bmod p$ cohomology.

(3) **Smith's theorem**: If G is a p-group (p prime) and if x is a finitistic G-space whose $\bmod p$ cohomology is isomorphic to the n-sphere, then the $\bmod p$ cohomology of the fixed-point set X^G is isomorphic to that of the r-sphere for some $-1 \leqslant r \leqslant n$, where (-1)-sphere means the empty set.

(4) Let T^k denote the k-dimensional toral group. Let X be a T^k-space whose rational cohomology is isomorphic to the n-sphere, and assume that there are only a finite number of orbit types and that the orbit spaces of all subtori are finitistic. Let H be a subtorus of T^k. Then by the above theorem the rational cohomology of X^H is isomorphic to that of the $r(H)$-sphere for some $-1 \leqslant r(H) \leqslant n$. Assume further that there is no fixed point of the T^k-action. Then, with H ranging over all subtori of dimension $k-1$, we have

$$n + 1 = \sum_H (r(H) + 1).$$

C. Differentiable Transformation Groups

Suppose that the group G is a transformation group of a †differentiable manifold M, G is a †Lie group, and the mapping $(g, x) \rightarrow g(x)$ of $G \times M$ into M is a differentiable mapping. Then G is called a **differentiable transformation group** (or **Lie transformation group**) of M, and M is called a differentiable **G-manifold**.

The following are basic facts about compact differentiable transformation groups [3, 4]:

(5) **Differentiable slice theorem**: Let G be a compact Lie group acting differentiably on a manifold M. Then, by averaging an arbitrary †Riemannian metric on M, we may have a G-invariant Riemannian metric on M. That is, the mapping $x \rightarrow g(x)$ is an †isometry of this Riemannian manifold M for each $g \in G$. For each point $x \in M$, the orbit $G(x)$ through x is a compact submanifold of M and the mapping $g \mapsto g(x)$ defines a G-equivariant diffeomorphism $G/G_x \cong G(x)$, where G/G_x is the left quotient space by the isotropy subgroup G_x. G_x acts orthogonally on the †tangent space T_xM at x (resp. the †normal vector space N_x of the orbit $G(x)$); we call it the **isotropy representation** (resp. **slice representation**) of G_x at x. Let E be the †normal vector bundle of the orbit $G(x)$. Since G acts naturally on E as a bundle mapping, the bundle E is equivalent to the bundle $(G \times N_x)/G_x$ over G/G_x as a †G-vector bundle, where G_x acts on N_x by means of the slice representation and G_x acts on G by the right translation. We can choose a small positive real number ε such that the †exponential mapping gives an equivariant †diffeomorphism of the ε-disk bundle of E onto an invariant †tubular neighborhood of $G(x)$.

(6) Assume that a compact Lie group G acts differentiably on M with the orbit space $M^* = M/G$ connected. Then there exists a maximum

orbit type G/H for G on M (i.e., H is an isotropy subgroup and H is conjugate to a subgroup of each isotropy group). The union $M_{(H)}$ of the orbits of type G/H is open and dense in M, and its image $M^*_{(H)}$ in M^* is connected.

The maximum orbit type for orbits in M guaranteed by the above theorem is called the **principal orbit type**, and orbits of this type are called **principal orbits**. The corresponding isotropy groups are called **principal isotropy groups**. Let P be a principal orbit and Q any orbit. If $\dim P > \dim Q$, then Q is called a **singular orbit**. If $\dim P = \dim Q$ but P and Q are not equivalent, then Q is called an **exceptional orbit**.

(7) Let G be a compact Lie group and M a compact G-manifold. Then the orbit types are finite in number.

By applying (5) and (6) we have that an isotropy group is principal if and only if its slice representation is trivial.

The situation is quite different in the case of noncompact transformation groups. For example, there exists an analytic action of $G = SL(4, \mathbf{R})$ on an analytic manifold M such that each orbit of G on M is closed and of codimension one and such that, for $x, y \in M$, G_x is not isomorphic to G_y unless x and y lie on the same G-orbit [5].

D. Compact Differentiable Transformation Groups

Many powerful techniques in †differential topology have been applied to the study of differentiable transformation groups. For example, using the techniques of †surgery, we can show that there are infinitely many free differentiable circle actions on †homotopy $(2n+1)$-spheres ($n \geqslant 3$) that are differentiably inequivalent and distinguished by the rational †Pontryagin classes of the orbit manifolds (W. C. Hsiang [6]). Also, using †Brieskorn varieties, we can construct many examples of differentiable transformation groups on homotopy spheres [3, 4, 7]. Differentiable actions of compact connected Lie groups on homology spheres have been studied systematically (Hsiang and W. Y. Hsiang [4]).

The Atiyah-Singer †index theorem has many applications in the study of transformation groups. The following are notable applications:

(8) Let M be a compact connected †oriented differentiable manifold of dimension $4k$ with a †spin-structure. If a compact connected Lie group G acts differentiably and nontrivially on M, then the $\hat{A}$-genus $\langle \hat{\mathscr{A}}(M), [M] \rangle$ of M vanishes (where $\hat{\mathscr{A}}(M)$ denotes the †$\hat{A}$-characteristic class of M) (M. F. Atiyah and F. Hirzebruch [8], K. Kawakubo [9]). For further developments, see A. Hattori [10].

(9) Let M be a closed oriented manifold with a differentiable circle action. Then each connected component F_k of the fixed point set can be oriented canonically, and we have

$$I(M) = \sum_k I(F_k),$$

where $I(\ )$ denotes the †Thom-Hirzebruch index [8, 9].

Let G be a compact Lie group and $G \to EG \to BG$ the †universal G-bundle. Then the †singular cohomology $H^*(EG \times_G X)$ is called **equivariant cohomology** for a G-space X and is an $H^*(BG)$-module. Let $G = U(1)$, M a differentiable $U(1)$-manifold, $F = M^G$, and $i: F \to M$ the inclusion mapping. Then the †localization of the induced homomorphism

$$S^{-1}i^*: S^{-1}H^*(EG \times_G M) \to S^{-1}H^*(BG \times F)$$

is an isomorphism, where S^{-1} denotes the localization with respect to the multiplicative set $S = \{at^k\}$ with a, k ranging over all positive integers and t the generator of $H^2(BG)$. Theorems (8) and (9) can be proved by the above localization isomorphism.

Let M be a differentiable manifold. The upper bound $N(M)$ of the dimension of all the compact Lie groups that acts effectively and differentiably on M is called the **degree of symmetry** of M. It measures, in some crude sense, the symmetry of the differentiable manifold M. The number $N(M)$ depends heavily on the differentiable structure. For example, $N(S^m) = m(m+1)/2$ for the standard m-sphere, but $N(\Sigma^m) < (m+1)^2/16 + 5$ for a †homotopy m-sphere ($m \geqslant 300$) that does not bound a †π-manifold [11]. Also, $N(P_n(\mathbf{C})) = n(n+2)$ for the complex projective n-space $P_n(\mathbf{C})$, but $N(hP_n(C)) < (n+1)(n+2)/2$ for any homotopy complex projective n-space $hP_n(\mathbf{C})$ ($n \geqslant 13$) other than $P_n(\mathbf{C})$ (T. Watabe [12]).

Let X be a differentiable closed manifold and $h: X \to P_n(\mathbf{C})$ be an orientation-preserving †homotopy equivalence. There is a conjecture about the total $\hat{A}$-classes that states: If X admits a nontrivial differentiable circle action, then $\hat{\mathscr{A}}(X) = h^* \hat{\mathscr{A}}(P_n(\mathbf{C}))$ (T. Petrie [13]). It is known that if the action is free outside the fixed-point set, then the conjecture is true (T. Yoshida [14]).

E. Equivariant Bordism

Fix a compact Lie group G; a compact **oriented G-manifold** (ψ, M) consists of a compact †oriented differentiable manifold M and an orientation-preserving differentiable G-action $\psi: G \times M \to M$ on M.

Given families $F \supset F'$ of subgroups of G, a compact oriented G-manifold (ψ, M) is (F, F')-**free** if the following conditions are satisfied: (i) if $x \in M$, then the isotropy group G_x is conjugate to a member of F; (ii) if $x \in \partial M$, then G_x is conjugate to a member of F'.

If F' is the empty family, then necessarily ∂M is empty and M is closed. In this case we say that (ψ, M) is F-**free**.

Given (ψ, M), define $-(\psi, M) = (\psi, -M)$ with the structure precisely the same as (ψ, M) except for †orientation. Also define $\partial(\psi, M) = (\psi, \partial M)$. Note that if (ψ, M) is (F, F')-free, then $(\psi, \partial M)$ is F'-free. Define (ψ, M) and (ψ', M') to be isomorphic if there exists an equivariant orientation-preserving diffeomorphism of M onto M'.

An (F, F')-free compact oriented n-dimensional G-manifold (ψ, M) is said to **bord** if there exists an (F, F)-free compact oriented $(n+1)$-dimensional G-manifold (Φ, W) together with a regularly embedded compact n-dimensional manifold M_1 in ∂W with M_1 invariant under the G-action Φ such that (Φ, M_1) is isomorphic to (ψ, M) and G_x is conjugate to a member of F' for $x \in \partial W - M_1$. Also, M_1 is required to have its orientation induced by that of W.

We say that (ψ_1, M_1) is **bordant** to (ψ_2, M_2) if the disjoint union $(\psi_1, M_1) + (\psi_2, -M_2)$ bords. Bordism is an equivalence relation on the class of (F, F')-free compact oriented n-dimensional G-manifolds. The bordism classes constitute an Abelian group $\mathbf{O}_n^G(F, F')$ under the operation of disjoint union. If F' is empty, denote the above group by $\mathbf{O}_n^G(F)$. The direct sum

$$\mathbf{O}_*^G(F, F') = \bigoplus_n \mathbf{O}_n^G(F, F')$$

is naturally an Ω-module, where Ω is the †oriented cobordism ring. If F consists of all subgroups of G, then $\mathbf{O}_*^G(F)$ is denoted by $\mathbf{O}_*^G$.

Suppose now that $F \supset F'$ are fixed families of subgroups of G. Every F'-free G-manifold is also F-free, and so this inclusion induces a homomorphism $\alpha: \mathbf{O}_n^G(F') \to \mathbf{O}_n^G(F)$. Similarly every F-free G-manifold is also (F, F')-free, inducing a homomorphism $\beta: \mathbf{O}_n^G(F) \to \mathbf{O}_n^G(F, F')$. Finally, there is a homomorphism $\partial: \mathbf{O}_n^G(F, F') \to \mathbf{O}_{n-1}^G(F')$ given by $\partial(\psi, M) = (\psi, \partial M)$. Then the following sequence is exact [15]:

$$\ldots \xrightarrow{\partial} \mathbf{O}_n^G(F') \xrightarrow{\alpha} \mathbf{O}_n^G(F) \xrightarrow{\beta} \mathbf{O}_n^G(F, F') \xrightarrow{\partial} \mathbf{O}_{n-1}^G(F') \xrightarrow{\alpha} \ldots.$$

A weakly almost complex compact G-manifold (ψ, M) consists of a †weakly almost complex compact manifold M and a differentiable G-action $\psi: G \times M \to M$ that preserves the weakly almost complex structure on M. $\mathbf{U}_*^G(F, F')$, $\mathbf{U}_*^G$ are defined similarly, and they are $\mathbf{U}_*$-modules, where $\mathbf{U}_*$ is the †complex cobordism ring of compact weakly almost complex manifolds.

To study $\mathbf{O}_*^G$ and $\mathbf{U}_*^G$, (co)bordism theory is introduced (P. E. Conner and E. E. Floyd [16]), which is one of the †generalized (co)-homology theories. Miscellaneous results are known, in particular, for G a cyclic group of prime period. By means of the equivariant †Thom spectrum, equivariant cobordism theory can be developed (T. tom Dieck [17]); this is a multiplicative generalized cohomology theory with Thom classes (→ 114 Differential Topology; also → 201 Homology Theory, 56 Characteristic Classes).

F. Equivariant Homotopy

Let G be a compact Lie group. On the category of closed G-manifolds, we say that two objects M, N are **χ-equivalent** if $\chi(M^H) = \chi(N^H)$ for all closed subgroups H of G, where $\chi(\ )$ is the †Euler-Poincaré characteristic. On the set of equivalence classes $\mathbf{A}(G)$, a ring structure is imposed by disjoint union and the Cartesian product. We call $\mathbf{A}(G)$ the **Burnside ring** of G. If G is finite, $\mathbf{A}(G)$ is naturally isomorphic to the classical Burnside ring of G [18].

Denote by $S(V)$ the unit sphere of an orthogonal G-representation space V. Let V, W be orthogonal G-representation spaces. The equivariant stable homotopy group $[[S(V), S(W)]]$, which is defined as the direct limit of the equivariant homotopy sets $[S(V+U), S(W+U)]_G$ taken over orthogonal G-representation spaces U and suspensions, is denoted by ω_α for $\alpha = V - W \in RO(G)$. The †smash product of representatives induces a bilinear pairing $\omega_\alpha \times \omega_\beta \to \omega_{\alpha+\beta}$. Then ω_0 is a ring, and ω_α is an ω_0-module. The ring ω_0 is isomorphic to the Burnside ring of G, and ω_α is a †projective ω_0-module of rank one. The ω_0-module ω_α is free if and only if $S(V)$ and $S(W)$ are stably G-homotopy equivalent [18].

Let E be an orthogonal G-vector bundle over a compact G-space X. Denote by $S(E)$ the sphere bundle associated with E. Let E, F be orthogonal G-vector bundles over X. Then E and F have the **same spherical G-fiber homotopy type** if there exist fiber-preserving G-mappings $f: S(E) \to S(F)$, $f': S(F) \to S(E)$ and fiber-preserving G-homotopies $h_t: S(E) \to S(E)$, $h_t': S(F) \to S(F)$ such that $h_0 = f' \circ f$, $h_1 = \text{identity}$, $h_0' = f \circ f'$, $h_1' = \text{identity}$. Let $KO_G(X)$ be the †equivariant K-group of real G-vector bundles over X. Let $T_G(X)$ be the additive subgroup of $KO_G(X)$ generated by elements of the form $[E] - [F]$, where E and F are orthogonal G-vector bundles having the same spherical G-fiber homotopy type. The factor group $J_G(X) = KO_G(X)/T_G(X)$ and the natural projection

$J_G: KO_G(X) \to J_G(X)$ are called an **equivariant** ***J*-group** and an **equivariant *J*-homomorphism**, respectively (→ 237 *K*-Theory).

In particular, $J_G(\{x_0\})$ is a factor group of the real representation ring $RO(G)$. †Adams operations on representation rings are the main tools for studying the group $J_G(\{x_0\})$ [18].

G. Infinitesimal Transformations

Let $f: G \times M \to M$ be a differentiable action of a Lie group G on a differentiable manifold M. Let X be a †left invariant vector field on G. Then we can define a differentiable vector field $f^+(X)$ on M as

$$f^+(X)_q h = \lim_{t \to 0} (h(f(\exp(-tX), q)) - h(q))/t$$

for each $q \in M$ and any differentiable function h defined on a neighborhood of q. It is easy to see that $f^+(X)_q = 0$ if and only if q is a fixed point of the one-parameter subgroup $\{\exp(tX)\}$. A vector field $f^+(X)$ is called an **infinitesimal transformation** of the differentiable transformation group G.

The set $\mathfrak{g}$ of all infinitesimal transformations of G forms a finite-dimensional †Lie algebra (the laws of addition and †bracket product are defined from those for the vector fields on M). If G acts effectively on M, $\mathfrak{g}$ is isomorphic to the Lie algebra of the Lie group G (→ 249 Lie Groups). In fact, the correspondence $X \to f^+(X)$ defines a Lie algebra homomorphism f^+ from the Lie algebra of all left invariant vector fields on G into the Lie algebra of all differentiable vector fields on M [19].

The following fact [20] is useful for the study of noncompact real analytic transformation groups. Let $\mathfrak{g}$ be a real †semisimple Lie algebra and $\rho: \mathfrak{g} \to L(M)$ be a Lie algebra homomorphism of $\mathfrak{g}$ into a Lie algebra of real analytic vector fields on a †real analytic manifold M. Let p be a point at which the vector fields in the image $\rho(\mathfrak{g})$ have common zero. Then there exists an analytic system of coordinates $(U; u_1, \ldots, u_m)$ with origin at p in which all the vector fields in $\rho(\mathfrak{g})$ are linear. Namely, there exists $a_{ij} \in \mathfrak{g}^* = \mathrm{Hom}_{\mathbf{R}}(\mathfrak{g}, \mathbf{R})$ such that

$$\rho(X)_q = -\sum_{i,j} a_{ij}(X) u_j(q) \frac{\partial}{\partial u_i}; \qquad X \in \mathfrak{g},\ q \in U.$$

The correspondence $X \to (a_{ij}(X))$ defines a Lie algebra homomorphism of $\mathfrak{g}$ into $\mathfrak{sl}(m, \mathbf{R})$.

For example, we can show that a real analytic $SL(n, \mathbf{R})$ action on the m-sphere is characterized by a certain real analytic vector field on $(m-n+1)$-sphere $(5 \leqslant n \leqslant m \leqslant 2n-2)$ [21]. In particular, there are infinitely many (at least the cardinality of the real numbers) inequivalent real analytic $SL(n, \mathbf{R})$ actions on the m-sphere $(3 \leqslant n \leqslant m)$.

Conversely, let $\mathfrak{g}$ be a finite-dimensional Lie algebra of vector fields on M. Although there is not always a differentiable transformation group G that admits $\mathfrak{g}$ as its Lie algebra of infinitesimal transformations, the following local result holds. Let $\tilde{G}$ be the †simply connected Lie group corresponding to the Lie algebra $\mathfrak{g}$. Then for each point x of M, there exist a neighborhood $\tilde{U}$ of the identity element e of $\tilde{G}$, neighborhoods V, W $(V \subset W)$ of x, and a differentiable mapping f of $\tilde{U} \times V$ into W with the following properties. Putting $f(g, y) = g(y)$ $(g \in \tilde{U}, y \in V)$, we have: (i) For all $y \in V$, $e(y) = y$. (ii) If $g, h \in \tilde{U}$, $y \in V$, then $(gh)(y) = g(h(y))$, provided that $gh \in \tilde{U}$, $h(y) \in V$. (iii) Let X be an arbitrary element of $\mathfrak{g}$. Put $g_t = \exp(-tX)$, the corresponding one-parameter subgroup of $\tilde{G}$. If $\varepsilon > 0$ is taken small enough, then we have $g_t \in \tilde{U}$ for $|t| < \varepsilon$ so that $g_t(y)(|t| < \varepsilon, y \in V)$ is well defined. Therefore g_t determines a vector field $\tilde{X}$ on V by the formula

$$\tilde{X}_y h = \lim_{t \to 0} (h(g_t(y)) - h(y))/t.$$

The vector field $\tilde{X}$ coincides with the restriction of X to V. This local proposition is often expressed by the statement that $\mathfrak{g}$ generates a **local Lie group of local transformations**, which is called **Lie's fundamental theorem** on local Lie groups of local transformations.

H. Criteria

It is important to know whether a given transformation group is a topological or a Lie transformation group. The following theorems are useful for this purpose [22, 23]:

(10) Let G be a transformation group of a †locally compact Hausdorff space M. If we introduce the †compact-open topology in G, then G is a topological transformation group of M when M is locally connected or M is a †uniform topological space and G acts †equicontinuously on M.

(11) Suppose that M is a †C^1-manifold and G is a topological transformation group of M acting effectively on M. If G is locally compact and the mapping $x \to g(x)$ of M is of class C^1 for each element g of G, then G is a Lie transformation group of M.

(12) Assume that G is a transformation group of a differentiable manifold M and G acts effectively on M. Let $\mathfrak{g}$ be the set of all vector fields on M defined by one-parameter groups of transformations of M contained in G as subgroups. If $\mathfrak{g}$ is a finite-dimensional Lie algebra, then G has a Lie group structure with respect to which G is a Lie transformation group of M, and then $\mathfrak{g}$ coincides with the Lie

algebra formed by the infinitesimal transformations of G.

By applying theorems (10), (11), and (12) we can show that the following groups are Lie transformation groups: the group of all †isometries of a †Riemannian manifold; the group of all affine transformations of a differentiable manifold with a †linear connection (generally, the group of all transformations of a differentiable manifold that leave invariant a given †Cartan connection); the group of all analytic transformations of a compact complex manifold (this group is actually a complex Lie group); and the group of all analytic (holomorphic) transformations of a bounded domain in $\mathbf{C}^n$.

For related topics → 105 Differentiable Manifolds, 114 Differential Topology, 122 Discontinuous Groups, 153 Fixed-Point Theorems, 427 Topology of Lie Groups and Homogeneous Spaces, etc.

References

[1] A. Borel et al., Seminar on transformation groups, Ann. Math. Studies, Princeton Univ. Press, 1960.
[2] W. Y. Hsiang, Cohomology theory of topological transformation groups, Erg. math. 85, Springer, 1975.
[3] G. E. Bredon, Introduction to compact transformation groups, Academic Press, 1972.
[4] W. C. Hsiang and W. Y. Hsiang, Differentiable actions of compact connected classical groups I, II, III, Amer. J. Math. 89 (1967); Ann. Math., (2) 92 (1970); Ann. Math., (2) 99 (1974).
[5] R. W. Richardson, Deformations of Lie subgroups and the variation of isotropy subgroups, Acta Math., 129 (1972).
[6] W. C. Hsiang, A note on free differentiable actions of S^1 and S^3 on homotopy spheres, Ann. Math., (2) 83 (1966).
[7] F. Hirzebruch and K. H. Mayer, $O(n)$-Mannigfaltigkeiten, Exotische Sphären und Singularitäten, Lecture notes in math. 57, Springer, 1968.
[8] M. F. Atiyah and F. Hirzebruch, Spin-manifolds and group actions, Essays on Topology and Related Topics, Memoires dédiés à Georges de Rham, Springer, 1970.
[9] K. Kawakubo, Equivariant Riemann-Roch theorems, localization and formal group law, Osaka J. Math., 17 (1980).
[10] A. Hattori, Spin^c-structures and S^1-actions, Inventiones Math., 48 (1978).
[11] W. C. Hsiang and W. Y. Hsiang, The degree of symmetry of homotopy spheres, Ann. Math., (2) 89 (1969).
[12] T. Watabe, On the degree of symmetry of complex quadric and homotopy complex projective space, Sci. Rep. Niigata Univ., 11 (1974).
[13] T. Petrie, Smooth S^1 actions on homotopy complex projective spaces and related topics, Bull. Amer. Math. Soc., 78 (1972).
[14] T. Yoshida, On smooth semifree S^1 actions on cohomology complex projective spaces, Publ. Res. Inst. Math. Sci., 11 (1976).
[15] P. E. Conner and E. E. Floyd, Maps of odd period, Ann. Math., (2) 84 (1966).
[16] P. E. Conner and E. E. Floyd, Differentiable periodic maps, Erg. Math., 33 (1964).
[17] T. tom Dieck, Bordism of G-manifolds and integrality theorems, Topology, 9 (1970).
[18] T. tom Dieck, Transformation groups and representation theory, Lecture notes in math. 766, Springer, 1979.
[19] R. S. Palais, A global formulation of the Lie theory of transformation groups, Mem. Amer. Math. Soc., 22 (1957).
[20] V. Guillemin and S. Sternberg, Remarks on a paper of Hermann, Trans. Amer. Math. Soc., 130 (1968).
[21] F. Uchida, Real analytic $SL(n, R)$ actions on spheres, Tôhoku Math. J., 33 (1981).
[22] D. Montgomery and L. Zippin, Topological transformation groups, Interscience, 1955.
[23] S. Kobayashi, Transformation groups in differential geometry, Erg. math. 70, Springer, 1972.

432 (VI.8) Trigonometry

A. Plane Trigonometry

Fix an orthogonal frame O-XY in a plane, and take a point P on the plane such that the angle POX is α. Denote by (x, y) the coordinates of P, and put $OP=r$ (Fig. 1). We call the six ratios $\sin\alpha=y/r$, $\cos\alpha=x/r$, $\tan\alpha=y/x$, $\cot\alpha=x/y$, $\sec\alpha=r/x$, $\operatorname{cosec}\alpha=r/y$ the **sine**, **cosine**, **tangent**, **cotangent**, **secant**, and **cosecant** of α, respectively. These functions of the angle α are called **trigonometric functions** or **circular functions** (→ 131 Elementary Functions). They are periodic functions with the fundamental period π for the tangent and cotangent, and 2π for the others. The relation $\sin^2\alpha+\cos^2\alpha=1$ and the **addition formulas** $\sin(\alpha\pm\beta)=\sin\alpha\cos\beta\pm\cos\alpha\sin\beta$, $\cos(\alpha\pm\beta)=\cos\alpha\cos\beta\mp\sin\alpha\sin\beta$ follow from the definitions (→ Appendix A, Table 2). Given a plane triangle ABC (Fig. 2), we have the following three properties: (i) $a=b\cos C+c\cos B$ (**the first law of cosines**); (ii) $a^2=b^2+c^2-2bc\cos A$ (**the second law of cosines**); (iii) $a/\sin A=b/\sin B=c/\sin C=2R$, where R is the radius of the circle circum-

scribed about $\triangle ABC$ (**laws of sines**) (→ Appendix A, Table 2). Thus we obtain relations among the six quantities a, b, c, $\angle A$, $\angle B$, and $\angle C$ associated with the triangle ABC. The study of plane figures by means of trigonometric functions is called **plane trigonometry**. For example, if a suitable combination of three of these six quantities (including a side) associated with a triangle is given, then the other three quantities are uniquely determined. The determination of unknown quantities associated with a triangle by means of these laws is called **solving a triangle**.

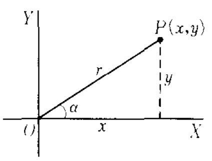


Fig. 1

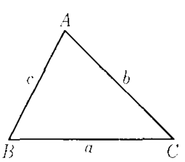


Fig. 2

B. Spherical Trigonometry

The part ABC of a spherical surface bounded by three arcs of great circles is called a **spherical triangle**. Points A, B, C are called the **vertices**; the three arcs a, b, c are called the **sides**; and the angles formed by lines tangent to the sides and intersecting at the vertices are called the **angles** of the spherical triangle (Fig. 3). If we denote the angles by A, B, C, we have the relation $A+B+C-\pi=E>0$, and E is called the **spherical excess**. Spherical triangles have properties similar to those of plane triangles: $\sin a/\sin A=\sin b/\sin B=\sin c/\sin C$ (**laws of sines**), and $\cos a=\cos b\cos c+\sin b\sin c\cos A$ (**law of cosines**). The study of spherical figures by means of trigonometric functions, called **spherical trigonometry**, is widely used in astronomy, geodesy, and navigation (→ Appendix A, Table 2).

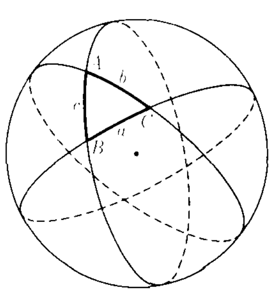


Fig. 3

C. History

Trigonometry originated from practical problems of determining a triangle from three of its elements. The development of spherical trigonometry, which was spurred on by its applications to astronomy, preceded the development of plane trigonometry. In Egypt, Babylon, and China, people had some knowledge of trigonometry, and the founder of trigonometry is believed to have been Hipparchus of Greece (fl. 150 B.C.). In the *Almagest* of Ptolemy (c. 150 A.D.) we find a table for $2\sin\alpha$ for $\alpha=0, 30', 1°, 1°30', \ldots$ that is exact to five decimal places, and the addition formulas. The Greeks calculated $2\sin\alpha$, which is the length of the chord corresponding to the double arc. Indian mathematicians, on the other hand, calculated half of the above quantities, that is, $\sin\alpha$ and $1-\cos\alpha$ for the arc α. In the book by Aryabhatta (c. 500 A.D.) we find laws of cosines. The Arabs, influenced by Indian mathematicians, expressed geometric computations algebraically, a technique also known to the Greeks. Abûl Wafâ (in the latter half of the 10th century A.D.) gave the correct sines of angles for every 30′ to 9 decimal places and studied with Al Battani the projection triangle of the sundial, thereby obtaining the concepts of sine, cosine, secant, and cosecant. Later, a table of sines and cosines for every minute was established by the Arabs. Regiomontanus (d. 1476), a German, elaborated on this table. The form he gave to trigonometry has been maintained nearly intact to the present day. Various theorems in trigonometry were established by G. J. Rhaeticus, J. Napier, J. Kepler, and L. Euler (1748). Euler treated trigonometry as a branch of analysis, generalized it to functions of complex variables, and introduced the abbreviated notations that are still in use (→ 131 Elementary Functions).

References

[1] H. Flanders and J. Price, Trigonometry, Academic Press, 1975.
[2] A. Washington and C. Edmond, Plane trigonometry, Addison-Wesley, 1977.

433 (XX.12) Turbulence and Chaos

Turbulent flow is the irregular motion of fluids, whereas relatively simple types of flows that are either stationary, slowly varying, or periodic in time are called **laminar flow**. When

a laminar flow is stable against external disturbances, it remains laminar, but if the flow is unstable, it usually changes into either another type of laminar flow or a turbulent flow.

A. Stability and Bifurcation of Flows

The velocity field $\mathbf{u}(\mathbf{x}, t)$, $\mathbf{x}$ being the space coordinates and t the time, of a flow of an incompressible viscous fluid in a bounded domain G is determined by the †Navier-Stokes equation of motion,

$$\frac{\partial \mathbf{u}}{\partial t}+(\mathbf{u}\cdot\operatorname{grad})\mathbf{u}-\nu\Delta\mathbf{u}+\frac{1}{\rho}\operatorname{grad} p=0, \qquad (1)$$

and the equation of continuity,

$$\operatorname{div}\mathbf{u}=0, \qquad (2)$$

with the prescribed initial and boundary conditions, where Δ denotes the Laplacian, p the pressure, ρ the density, and ν the kinematic viscosity of the fluid. Suitable extensions must be made in the foregoing system of equations if other field variables, such as the temperature in thermal-convection problems, are to be considered.

The stability of a fluid flow is studied by examining the behavior of the solution of equations (1) and (2) against external disturbances, and, in particular, stability against infinitesimal disturbances constitutes the linear stability problem. The stability characteristics of the solution of equations (1) and (2) depend largely upon the value of the †Reynolds number $R=UL/\nu$, U and L being the representative velocity and length of the flow, respectively. Let a stationary solution of equations (1) and (2) be $\mathbf{u}_0(\mathbf{x}, R)$. If the perturbed flow is given by $\mathbf{u}_0(\mathbf{x}, R)+\mathbf{v}(\mathbf{x}, R)\exp(\sigma t)$, $\mathbf{v}$ being the perturbation velocity, and equation (1) is linearized with respect to $\mathbf{v}$, we obtain a †linear eigenvalue problem for σ. The flow is called linearly stable if $\max(\operatorname{Re}\sigma)$ is negative, and linearly unstable if it is positive. For small values of R, a flow is generally stable, but it becomes unstable if R exceeds a critical value R_c, which is called the critical Reynolds number [1].

The instability of a stationary solution gives rise to the †bifurcation to another solution at a †bifurcation point R_c of the parameter R. If $\operatorname{Im}\sigma=0$ for an eigenvalue σ at $R=R_c$, a stationary solution bifurcates from the solution u_0 at R_c, and if $\operatorname{Im}\sigma\neq 0$, a time-periodic solution bifurcates at R_c. The latter bifurcation is called the Hopf bifurcation. A typical example of stationary bifurcation is the generation of an axially periodic row of Taylor vortices in Couette flow between two rotating coaxial cylinders, which was studied by G. I. Taylor (1923), with excellent agreement between theory and experiment [2]. Hopf bifurcation is exemplified by the generation of Tollmien-Schlichting waves in the laminar †boundary layer along a flat plate, which was predicted theoretically by W. Tollmien (1929) and H. Schlichting (1933) and later confirmed experimentally by G. B. Schubauer and H. K. Skramstad (1947) [3].

In either type of bifurcation ($\operatorname{Im}\sigma=0$ or $\neq 0$) the bifurcation is called supercritical if the bifurcating solution exists only for $R>R_c$, subcritical if it exists only for $R<R_c$, and transcritical if it happens to exist on both sides of R_c. The amplitude of the departure of the bifurcating solution from the unperturbed solution u_0 tends to zero as $R\to R_c$. The behavior of the bifurcating solution around the bifurcation point R_c is dealt with systematically by means of bifurcation analysis. In supercritical bifurcation, the bifurcating solution is stable and represents an equilibrium state to which the perturbed flow approaches just as in the cases of Taylor vortices and Tollmien-Schlichting waves. On the other hand, for subcritical bifurcation the bifurcating solution is unstable and gives a critical amplitude of the disturbance above which the linearly stable basic flow ($R<R_c$) becomes unstable. In this case, the instability of the basic flow gives rise to a sudden change of the flow pattern resulting in either a stationary (or time-periodic) or even turbulent flow. The transition to turbulent flow that takes place in Hagen-Poiseuille flow through a circular tube and is linearly stable at all values of R ($R_c=\infty$) may be attributed to this type of bifurcation.

The concept of bifurcation can be extended to the case where the flow u_0 is nonstationary, but the bifurcation analysis then becomes much more difficult.

B. Onset of Turbulence

The fluctuating flow resulting from an instability does not itself necessarily constitute a turbulent flow. In order that a flow be turbulent, the fluctuations must take on some irregularity. The turbulent flow is usually defined in terms of the long-time behavior of the flow velocity $\mathbf{u}(\mathbf{x}, t)$ at a fixed point $\mathbf{x}$ in space. The flow is expected to be turbulent if the fluctuating velocity

$$\delta\mathbf{u}(\mathbf{x}, t)=\mathbf{u}(\mathbf{x}, t)-\lim_{T\to\infty}\frac{1}{T}\int_0^T \mathbf{u}(\mathbf{x}, t)\,dt \qquad (3)$$

satisfies the condition

$$\lim_{\tau\to\infty}\lim_{T\to\infty}\frac{1}{T}\int_0^T \delta u_i(\mathbf{x}, t)\,\delta u_i(\mathbf{x}, t+\tau)\,dt=0, \qquad (4)$$

where the subscripts label the components. Condition (4) implies that the †dynamical system of a fluid has the mixing property. This condition also states that the velocity fluctuation δu_i has a continuous frequency spectrum. In practical situations the frequency spectrum of a turbulent flow may contain both the line and continuous spectra, in which case the flow is said to be partially turbulent.

L. D. Landau (1959) and E. Hopf (1948) proposed a picture of turbulent flow as one composed of a †quasiperiodic motion, $\mathbf{u}(t)=\mathbf{f}(\omega_1 t, \omega_2 t, \ldots, \omega_n t)$, with a large number of rationally independent frequencies $\omega_1, \ldots, \omega_n$ produced by successive supercritical bifurcations of Hopf type. This picture of turbulence is not compatible with the foregoing definition of turbulence, since it does not satisfy the mixing property (4). The fact that the generation of real turbulence is not necessarily preceded by successive supercritical bifurcations casts another limitation on the validity of this picture.

The concept of turbulence is more clearly exhibited with respect to a dynamical system of finite dimension. Although we are without a general proof, it is expected that the Navier-Stokes equation with nonzero viscosity ν can be approximated within any degree of accuracy by a system of finite-dimensional †first-order ordinary differential equations

$$\frac{dX}{dt}=F(X). \tag{5}$$

Thus the onset and some general properties of turbulence are understood in the context of the theory of †dynamical systems. Turbulence is related to those solutions of equation (5) that tend to a †set in the †phase space that is neither a †fixed point, a †closed orbit, nor a †torus. A set of such complicated structure is called a nonperiodic †attractor or a strange attractor. Historically, the strange attractor originates from the strange Axiom A attractor that was found in a certain class of dynamical systems called the Axiom A systems. However, this term has come to be used in a broader sense, and it now represents a variety of nonperiodic motions exhibited by a system that is not necessarily of Axiom A type. The above-mentioned Landau-Hopf picture of turbulence was criticized by D. Ruelle and F. Takens (1971), who proved for the dynamical system (5) that an arbitrary small perturbation on a quasiperiodic †flow on a k-dimensional torus ($k\geqslant 4$) generically (in the sense of residual sets) produces a flow with a strange Axiom A attractor [4].

There exist a number of examples of first-order ordinary differential equations of relatively low dimension whose solutions exhibit nonperiodic behavior. An important model system related to turbulence is the Lorenz model (1963) of thermal convection in a horizontal fluid layer. This model is obtained by taking only three components out of an infinite number of spatial †Fourier components of the velocity and temperature fields. The model is written as

$$\frac{dX}{dt}=-\sigma X+\sigma Y,$$

$$\frac{dY}{dt}=-XZ+rX-Y,$$

$$\frac{dZ}{dt}=XY-bZ, \tag{6}$$

where σ $(>b+1)$ and b are positive constants and r is a parameter proportional to the Rayleigh number. Obviously, equations (6) have a fixed point $X=Y=Z=0$ representing the state of thermal convection without fluid flow. For $r<1$, this fixed point is stable, but it becomes unstable for $r>1$, and a pair of new fixed points $X=Y=\pm\sqrt{b(r-1)}$, $Z=r-1$ emerges supercritically. This corresponds to the onset of stationary convection at $r=1$. At a still higher value of $r=\sigma(\sigma+b+3)/(\sigma-b-1)$, a subcritical Hopf bifurcation occurs with respect to this pair of fixed points, and for a certain range of r above this threshold the solutions with almost any initial conditions exhibit nonperiodic behavior. This corresponds to the generation of turbulence. The property

$$\frac{\partial \dot{X}}{\partial X}+\frac{\partial \dot{Y}}{\partial Y}+\frac{\partial \dot{Z}}{\partial Z}=-(\sigma+b+1)<0, \tag{7}$$

where the dots denote time derivatives, shows that each volume element of the phase space shrinks asymptotically to zero as the time increases indefinitely. This property is characteristic of dynamical systems with energy dissipation, in sharp contrast to the †measure-preserving character of †Hamiltonian systems [5].

For a certain class of ordinary differential equations, the bifurcation to nonperiodic motion corresponds neither to the bifurcation of tori, just as in the Ruelle-Takens theory, nor to subcritical bifurcation, as in the Lorenz model. Such a bifurcation takes place when nonperiodic motion emerges as the consequence of an infinite sequence of supercritical bifurcations at each of which a periodic orbit of period T bifurcates into one of period $2T$. If we denote the nth bifurcation point by r_n, the distance $r_{n+1}-r_n$ between two successive bifurcation points decreases exponentially with increasing n, and eventually the bifurcation points accumulate at a point r_c, beyond which nonperiodic motion is expected to emerge. It is

not yet clear if any of the above three types of bifurcation leading to nonperiodic behavior is actually responsible for the generation of real turbulence.

Some important properties of a dynamical system with a nonperiodic attractor, which may be either a flow or a †diffeomorphism, can be stated as follows:
(i) The distance between two points in the phase space that are initially close to each other grows exponentially in time, so that the solutions exhibit a sensitive dependence on the initial conditions.
(ii) The nonperiodic attractor has †Lebesgue measure zero, and such a system is expected to have many other †ergodic †invariant measures.

The irregular behavior of a deterministic dynamical system is also called **chaos**, but this concept is more abstract and general than that of turbulence, and covers phenomena exhibited by systems such as nonlinear electric circuits, chemical reactions, and ecological systems.

C. Statistical Theory of Turbulence

The statistical theory of turbulence deals with the statistical behavior of fully developed turbulence. The turbulent field is sometimes idealized for mathematical simplicity to be homogeneous or isotropic. In **homogeneous turbulence** the statistical laws are invariant under all parallel displacements of the coordinates, whereas in **isotropic turbulence** invariance under rotations and reflections of the coordinates is required in addition.

The instantaneous state of the fluid motion is completely determined by specifying the fluid velocity $\mathbf{u}$ at all space points $\mathbf{x}$ and can be expressed as a phase point in the infinite-dimensional †phase space spanned by these velocities. The phase point moves with time along a path uniquely determined by the solution of the Navier-Stokes equation. In the turbulent state the path is unstable to the initial disturbance and describes an irregular line in the phase space. In this situation the deterministic description is no longer useful and should be replaced by a statistical treatment. Abstractly speaking, turbulence is just a view of fluid motion as the random motion of the phase point $\mathbf{u}(\mathbf{x})$ (→ 407 Stochastic Processes). The equation for the †characteristic functional of the random velocity $\mathbf{u}(\mathbf{x})$ was first given by E. Hopf (1952). An exact solution obtained by Hopf represents a †normal distribution associated with a white energy spectrum, but so far no general solution has been obtained [6].

Besides the formulation in terms of the †probability distribution or the characteristic functional, there is another way of describing turbulence by †moments of lower orders. This is the conventional statistical theory originated by G. I. Taylor (1935) and T. von Kármán (1938), which made remarkable progress after World War II. The principal moments in this theory are the **correlation tensor**, whose (i,j)-component is the mean of the product of two velocity components u_i at a point $\mathbf{x}$ and u_j at another point $\mathbf{x}+\mathbf{r}$,

$$B_{ij}(\mathbf{r})=\langle u_i(\mathbf{x})u_j(\mathbf{x}+\mathbf{r})\rangle, \tag{8}$$

and its †Fourier transform, or the **energy spectrum tensor**,

$$\Phi_{ij}(\mathbf{k})=\frac{1}{(2\pi)^3}\int B_{ij}(\mathbf{r})\exp(-\sqrt{-1}\,\mathbf{k}\cdot\mathbf{r})\,d\mathbf{r}. \tag{9}$$

In isotropic turbulence Φ_{ij} is expressed as

$$\Phi_{ij}(\mathbf{k})=\frac{1}{4\pi k^2}E(k)\left(\delta_{ij}-\frac{k_ik_j}{k^2}\right), \quad k=|\mathbf{k}|, \tag{10}$$

where $E(k)$ is the **energy spectrum function**, representing the amount of energy included in a spherical shell of radius k in the wave number space. The energy of turbulence $\mathscr{E}$ per unit mass is expressed as

$$\mathscr{E}=\frac{1}{2}\langle|\mathbf{u}|^2\rangle=\frac{1}{2}B_{ii}(0)=\frac{1}{2}\int\Phi_{ii}(\mathbf{k})\,d\mathbf{k}$$
$$=\int_0^\infty E(k)\,dk. \tag{11}$$

The state of turbulence is characterized by the Reynolds number $R=E_0^{1/2}/(\nu k_0^{1/2})$, where E_0 and k_0 are representative values of $E(k)$ and k, respectively. For weak turbulence of small R, $E(k)$ is governed by a linear equation with the general solution

$$E(k,t)=E(k,0)\exp(-2\nu k^2t), \tag{12}$$

$E(k,0)$ being an arbitrary function. Thus $E(k)$ decays in time due to the viscous dissipation. For strong turbulence of large R, it is difficult to obtain the precise form of $E(k)$, and this is usually done by way of some assumption that allows us to approximate the nonlinear effects [7].

Some of the similarity laws governing the energy spectrum and other statistical functions can be determined rigorously but not necessarily uniquely. For 3-dimensional incompressible turbulence, the energy spectrum satisfies an inviscid similarity law

$$E(k)/E_0=F_e(k/k_0) \tag{13}$$

in the energy-containing region $k=O(k_0)$ characterized by a wave number k_0, and a viscous similarity law

$$E(k)/E_0=R^{-5/4}F_d(k/(R^{3/4}k_0)), \tag{14}$$

in the energy dissipation region $k=O(R^{3/4}k_0)$, where F_e and F_d denote dimensionless functions generally dependent on the initial condition and the time [6].

If an assumption is made to the effect that the statistical state in the energy-dissipation region depends only upon the energy-dissipation rate $\varepsilon=-d\mathscr{E}/dt$ besides the viscosity ν (or R), then (14) becomes Kolmogorov's equilibrium similarity law (1941):

$$E(k)=\varepsilon^{1/4}\nu^{5/4}F(k/(\varepsilon^{1/4}\nu^{-3/4})), \tag{15}$$

where F is a dimensionless function. For extremely large R (or small ν) there exists an inertial subregion between the energy-containing and energy-dissipation regions such that the viscous effect vanishes and (15) takes the form

$$E(k)=K\varepsilon^{2/3}k^{-5/3}, \tag{16}$$

where K is an absolute constant. **Kolmogorov's spectrum** (16) has been observed experimentally several times, and now its consistency with experimental results at large Reynolds numbers is well established [8].

Kolmogorov (1962) and others modified (16) by taking account of the fluctuation of ε due to the **intermittent structure** of the energy-dissipation region as

$$E(k)=K'\varepsilon^{2/3}k^{-5/3}(Lk)^{-\mu/9}, \tag{17}$$

where ε is now the average of the fluctuating ε, μ is the covariance of the log-normal distribution of ε, and L is the length scale of the spatial domain in which the average of ε is taken [8]. A similar modification, with the exponent $-\mu/3$ in place of $-\mu/9$, is obtained using a fractal model of the energy-cascade process. These corrections to $E(k)$, based upon the experimentally estimated μ of 0.3–0.5, are too small to be detected experimentally, but the deviation is expected to appear more clearly in the higher-order moments [8–10]. It should be noted that Kolmogorov's spectrum (16) itself does not contradict the notion of intermittent turbulence and gives one of the possible asymptotic forms in the limit $R\to\infty$.

The 1-dimensional Burgers model of turbulence satisfies the same similarity laws as (13) and (14), but it has an inviscid spectrum $E(k)\propto k^{-2}$ instead of (16). Two-dimensional incompressible turbulence has no energy-dissipation region, and hence Kolmogorov's theory is not valid for this turbulence. It has an inviscid spectrum $E(k)\propto k^{-3}$, first derived by R. H. Kraichnan (1967), C. E. Leith (1968), and G. K. Batchelor (1969). These inviscid spectra for 1- and 2-dimensional turbulence have been confirmed by numerical simulation [11].

References

[1] C. C. Lin, The theory of hydrodynamical stability, Cambridge Univ. Press, 1955.
[2] S. Chandrasekhar, Hydrodynamic and hydromagnetic stability, Clarendon Press, 1961.
[3] H. L. Dryden, Recent advances in the mechanics of boundary layer flow, Adv. Appl. Mech., 1 (1948), 1–40.
[4] D. Ruelle and F. Takens, On the nature of turbulence, Comm. Math. Phys., 20 (1971), 167–192.
[5] E. N. Lorenz, Deterministic nonperiodic flow, J. Atmospheric Sci., 20 (1963), 130–141.
[6] T. Tatsumi, Theory of homogeneous turbulence, Adv. Appl. Mech., 20 (1980), 39–133.
[7] G. K. Batchelor, The theory of homogeneous turbulence, Cambridge Univ. Press, 1953.
[8] A. S. Monin and A. M. Yaglom, Statistical fluid mechanics II, MIT Press, 1975.
[9] B. Mandelbrot, Fractals and turbulence: Attractors and dispersion, Lecture notes in math. 615, Springer, 1977, 83–93.
[10] B. Mandelbrot, Intermittent turbulence and fractal dimension: Kurtosis and the spectral exponent 5/3 + B, Lecture notes in math. 565, Springer, 1977, 121–145.
[11] T. Tatsumi, Analytical theories and numerical experiments on two-dimensional turbulence, Theor. Appl. Mech., 29 (1979), 375–393.

U

434 (XX.22) Unified Field Theory

A. History

Unified field theory is a branch of theoretical physics that arose from the success of †general relativity theory. Its purpose is to discuss in a unified way the fields of gravitation, electromagnetism, and nuclear force from the standpoint of the geometric structure of space and time. Studies have continued since 1918, and many theories of mathematical interest have been published without attaining, however, any conclusive physical theory.

A characteristic feature of relativity theory is that it is based on a completely new concept of space and time. That is, in general relativity theory it is considered that when a gravitational field is generated by matter, the structure of space and time changes, and the flat †Minkowski world becomes a 4-dimensional †Riemannian manifold (with signature (1, 3)) having nonvanishing curvature. The †fundamental tensor g_{ij} of the manifold is interpreted as the gravitational potential, and the basic gravitational equation can be described as a geometric law of the manifold. It is characteristic of general relativity theory that gravitational phenomena are reduced to space-time structure (→ 359 Relativity). The introduction of the Minkowski world in †special relativity theory was a revolutionary advance over the 3-dimensional space of Newtonian mechanics. But the inner structure of the Minkowski world does not reflect gravitational phenomena. The latter shortcoming is overcome by introducing the concept of space-time represented by a Riemannian manifold into general relativity theory.

When a coexisting system of gravitational and electromagnetic fields is discussed in general relativity theory, simultaneous equations (Einstein-Maxwell equations) must be solved for the gravitational potential g_{ij} and the electromagnetic field tensor F_{ij}. Thus the gravitational potential g_{ij} is affected by the existence of an electromagnetic field. As the validity of general relativity began to be accepted, it came to be expected that all physical actions might be attributed to the gravitational and electromagnetic fields. Thus various extensions of general relativity theory have been proposed in order to devise a geometry in which the electromagnetic as well as the gravitational field directly contributes to the space-time structure, and to establish a unified theory of both fields on the basis of the geometry thus obtained. These attempts are illustrated in Fig. 1.

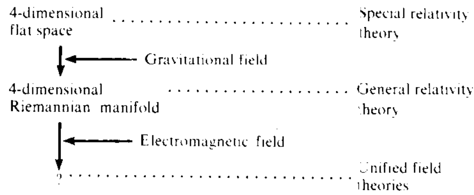


Fig. 1

B. Weyl's Theory

The first unified field theory was proposed by H. Weyl in 1918. In Riemannian geometry, which is the mathematical framework of general relativity theory, the †covariant derivative of the †fundamental tensor g_{ij} vanishes, i.e.,

$$\nabla_i g_{jk} \equiv \partial g_{jk}/\partial x^i - g_{ja}\Gamma^a_{ik} - g_{ak}\Gamma^a_{ij} = 0, \tag{1}$$

where Γ^i_{jk} is the †Christoffel symbol derived from g_{ij}. Conversely, if Γ^i_{jk} is considered as the coefficient of a general †affine connection and (1) is solved with respect to Γ^i_{jk} under the condition $\Gamma^i_{jk} = \Gamma^i_{kj}$, then the Christoffel symbol derived from g_{ij} coincides with Γ^i_{jk}. In this sense, (1) means that the space-time manifold has Riemannian structure. On the other hand, Weyl considered a space whose structure is given by an extension of (1),

$$\nabla_i g_{jk} = 2A_i g_{jk}, \tag{2}$$

and developed a unified field theory by regarding A_i as the electromagnetic potential. This theory has mathematical significance in that it motivated the discovery of Cartan's geometry of connection, but it has some unsatisfactory points concerning the derivation of the field equation and the equation of motion for a charged particle.

The scale transformation given by $\bar{g}_{ij} = \rho^2 g_{ij}$ is important in Weyl's theory. If in addition to this transformation, A_i is changed to

$$\bar{A}_i = A_i - \partial \log \rho/\partial x^i, \tag{3}$$

then (2) is left invariant and the space-time structure in Weyl's theory remains unchanged. We call (3) the **gauge transformation**, corresponding to the fact that the electromagnetic potential A_i is determined by the electromagnetic field tensor F_{ij} up to a gradient vector. In the †field theories known at present, the gauge transformation is generalized to various fields, and the law of charge conservation is derived from the invariance of field equations under generalized gauge transformation.

C. Further Developments

A unified field theory that appeared after Weyl's is **Kaluza's 5-dimensional theory** (Th.

Kaluza, 1921). This theory has been criticized as being artificial, but it is logically consistent, and therefore many of the later unified field theories are improved or generalized versions of it. The underlying space of Kaluza's theory is a 5-dimensional Riemannian manifold with the fundamental form

$$ds^2=(dx^4+A_a dx^a)^2+g_{ab}dx^a dx^b,$$

where A_i and g_{ij} are functions of x^i alone ($a, b, \ldots, i, j=0,1,2,3$). The field equation and the equation of motion of a particle are derived from the variational principle in general relativity theory. The field equation is equivalent to the Einstein-Maxwell equations. The trajectory of a charged particle is given by a geodesic in the manifold, and its equation is reducible to the Lorentz equation in general relativity.

After the introduction of Kaluza's theory, various unifield field theories were proposed, and we give here the underlying manifolds or geometries of some mathematically interesting theories: a manifold with †affine connection admitting absolute parallelism (A. Einstein, 1928); a manifold with †projective connection (O. Veblen, B. Hoffman, 1930 [4]; J. A. Schouten, D. van Dantzig, 1932); **wave geometry** (a theory based on the linearization of the fundamental form; Y. Mimura, 1934 [3]); a nonholonomic geometry (G. Vranceanu, 1936); a manifold with †conformal connection (Hoffman, 1948).

The investigations since 1945 have been motivated by the problem of the representation of matter in general relativity theory. Einstein first represented matter by an energy-momentum tensor T_{ij} of class C^0, which must be determined by information obtained from outside relativity. Afterward he felt that this point was unsatisfactory and tried to develop a theory on the basis of field variables alone, without introducing such a quantity as T_{ij}. This theory is the so-called **unitary field theory**, and a solution without singularities is required from a physical point of view. His first attempt was to remove singularities from an exterior solution in general relativity by changing the topological structure of the space-time manifold. This idea was then extended to a unified field theory by J. A. Wheeler, and an interpretation was given to mass and charge by applying the theory of †harmonic integrals (1957) [2].

Einstein's second attempt was to propose a **nonsymmetric unified field theory** (1945) [1, Appendix II; 6]. The fundamental quantities in this theory are a nonsymmetric tensor g_{ij} and a nonsymmetric affine connection Γ^i_{jk}. The underlying space of the theory can be considered a direct extension of the Riemannian manifold, since (1) is contained in the field equations (notice the order of indices in this equation). E. Schrödinger obtained field equations of almost the same form by taking only Γ^i_{jk} as a fundamental quantity (1947) [5].

References

[1] A. Einstein, The meaning of relativity, Princeton Univ. Press, fifth edition, 1956.
[2] J. A. Wheeler, Geometrodynamics, Academic Press, 1962.
[3] Y. Mimura and H. Takeno, Wave geometry, Sci. Rep. Res. Inst. Theoret. Phys. Hiroshima Univ., no. 2 (1962).
[4] O. Veblen, Projektive Relativitätstheorie, Springer, 1933.
[5] E. Schrödinger, Space-time structure, Cambridge Univ. Press, 1950.
[6] A. Einstein, A generalization of the relativistic theory of gravitation, Ann. Math., (2) 46 (1945), 578–584.
[7] V. Hlavatý, Geometry of Einstein's unified field theory, Noordhoff, 1957.
[8] M. A. Tonnelat, Einstein's unified field theory, Gordon & Breach, 1966. (Original in French, 1955.)

435 (II.23) Uniform Convergence

A. Uniform Convergence of a Sequence of Real-Valued Functions

A sequence of real-valued functions $\{f_n(x)\}$ defined on a set B is said to be **uniformly convergent** (or to **converge uniformly**) to a function $f(x)$ on the set B if it converges with respect to the †norm $\|\varphi\|=\sup\{|\varphi(x)|\,|\,x\in B\}$, i.e., $\lim_{n\to\infty}\|f_n-f\|=0$ (→ 87 Convergence). In other words, $\{f_n(x)\}$ converges uniformly to $f(x)$ on B if for every positive constant ε we can select a number N independent of the point x such that $|f_n(x)-f(x)|<\varepsilon$ holds for all $n>N$ and $x\in B$. By the †completeness of the real numbers, a sequence of functions $\{f_n(x)\}$ converges uniformly on B if and only if we can select for every positive constant ε a number N independent of the point x such that $|f_m(x)-f_n(x)|<\varepsilon$ holds for all $m, n>N$ and $x\in B$.
The **uniform convergence** of a series $\sum_n f_n(x)$ or of an infinite product $\prod_n f_n(x)$ is defined by the uniform convergence of the sequence of its partial sums or products. If the series of the absolute values $\sum_n |f_n(x)|$ converges uniformly, then the series $\sum_n f_n(x)$ also converges uniformly. In this case the series $\sum_n f_n(x)$ is said to

be **uniformly absolutely convergent**. A sequence of (nonnegative) constants M_n satisfying $|f_n(x)| \leqslant M_n$ is called a **dominant** (or **majorant**) of the sequence of functions $\{f_n(x)\}$. A series of functions $\sum_n f_n(x)$ with converging **majorant series** $\sum_n M_n$ is uniformly absolutely convergent (**Weierstrass's criterion for uniform convergence**).

Let $\{\lambda_n(x)\}$ be another sequence of functions on B. The series $\sum_n \lambda_n(x) f_n(x)$ is uniformly convergent if either of the following conditions holds: (i) The series $\sum_n f_n(x)$ converges uniformly and the partial sums of the series $\sum_n |\lambda_n(x) - \lambda_{n+1}(x)|$ are uniformly bounded, i.e., bounded by a constant independent of $x \in B$ and of the number of terms; or (ii) the series $\sum_n |\lambda_n(x) - \lambda_{n+1}(x)|$ converges uniformly, the sequence $\{\lambda_n(x)\}$ converges uniformly to 0, and the partial sums of $\sum_n |f_n(x)|$ are uniformly bounded.

B. Uniform Convergence and Pointwise Convergence

Let $\{f_n(x)\}$ be a sequence of real-valued functions on B, and let $f(x)$ be a real-valued function also defined on B. If the sequence of numbers $\{f_n(x_0)\}$ converges to $f(x_0)$ for every point $x_0 \in B$, we say that $\{f_n(x)\}$ is **pointwise convergent** (or **simply convergent**) to the function $f(x)$. Pointwise convergence is, of course, weaker than uniform convergence. If we represent the function $f(x)$ by the point $\prod_{x \in B} f(x) = [f]$ of the †Cartesian product $\mathbf{R}^B = \prod_{x \in B} \mathbf{R}$, then the pointwise convergence of $\{f_n(x)\}$ to $f(x)$ is equivalent to the convergence of the sequence of points $\{[f_n]\}$ to $[f]$ in the †product topology of $\mathbf{R}^B$.

When B is a †topological space and every $f_n(x)$ is continuous, the pointwise limit $f(x)$ of the sequence $\{f_n(x)\}$ is not necessarily continuous. However, if the sequence of continuous functions $\{f_n(x)\}$ converges uniformly to $f(x)$, then the limit function $f(x)$ is continuous. On the other hand, the continuity of the limit does not imply that the convergence is uniform. If the set B is †compact and the sequence of continuous functions $\{f_n(x)\}$ is monotone (i.e., $f_n(x) \leqslant f_{n+1}(x)$ for all n or $f_n(x) \geqslant f_{n+1}(x)$ for all n) and pointwise convergent to a continuous function $f(x)$, then the convergence is uniform (**Dini's theorem**).

C. Uniform Convergence on a Family of Sets

Let B be a topological space. We say that a sequence of functions $\{f_n(x)\}$ is **uniformly convergent in the wider sense** to the function $f(x)$, depending on circumstances, in either of the following two cases: (i) Every point $x_0 \in B$ has a neighborhood U on which the sequence $\{f_n(x)\}$ converges uniformly to $f(x)$; or (ii) $\{f_n(x)\}$ converges uniformly to $f(x)$ on every compact subset K in B. If B is †locally compact, the two definitions coincide. The term **uniform convergence on compact sets** is also used for (ii).

In general, given a family $\mathscr{P}$ of subsets in B, we may introduce in the space $\mathscr{F}$ of real-valued functions on B a family of †seminorms $\|f\|_K = \sup\{|f(x)| \mid x \in K\}$ for every set $K \in \mathscr{P}$. Let T be the topology of $\mathscr{F}$ defined by this family of seminorms (→ 424 Topological Linear Spaces). A sequence $\{f_n(x)\}$ is called **uniformly convergent on** $\mathscr{P}$ if it is convergent with respect to T. In particular, when $\mathscr{P}$ coincides with $\{B\}$, $\{\{x\} \mid x \in B\}$, or the family of all compact sets in B, then uniform convergence on $\mathscr{P}$ coincides with the usual uniform convergence, pointwise convergence, or uniform convergence on compact sets, respectively. If $\mathscr{P}$ is a countable set, the topology T is †metrizable. Most of these definitions and results may be extended to the case of functions whose values are in the complex number field, in a †normed space, or in any †uniform space.

D. Topology of the Space of Mappings

Let X, Y be two topological spaces. Denote by $C(X, Y)$ the space of all continuous mappings $f: X \rightarrow Y$. This space $C(X, Y)$, or a subspace $\mathscr{F}$ of $C(X, Y)$, is called a **mapping space** (or **function space** or **space of continuous mappings**) from X to Y. A natural mapping $\Phi: \mathscr{F} \times X \rightarrow Y$ is defined by $\Phi(f, x) = f(x)(f \in \mathscr{F}, x \in X)$. We define a topology in $\mathscr{F}$ as follows: for a compact set K in X and an open set U in Y, put $W(K, U) = \{f \in \mathscr{F} \mid f(K) \in U\}$, and introduce a topology in $\mathscr{F}$ such that the base for the topology consists of intersections of finite numbers of $W(K_i, U_i)$. This topology is called the **compact-open topology** (R. H. Fox, *Bull. Amer. Math. Soc.*, 51 (1945)). When X is a †locally compact Hausdorff space and Y is a †Hausdorff space, the compact-open topology is the †weakest topology on $\mathscr{F}$ for which the function Φ is continuous. If, in this case, $\mathscr{F}$ is compact with respect to the compact-open topology, then the compact-open topology coincides with the topology of pointwise convergence.

In particular, when Y is a †metric space (or, in general, a †uniform space with the uniformity $\mathfrak{U}$), the compact-open topology in $\mathscr{F}$ coincides with the topology of uniform convergence on compact sets. A family $\mathscr{F}$ is called

equicontinuous at a point $x \in X$ if for every positive number ε (in the case of uniform space, for every $V \in \mathfrak{U}$) there exists a neighborhood U of x such that $\rho(f(x), f(p)) < \varepsilon$ ($(f(x), f(p)) \in V$) for every point $p \in U$ and for every function $f \in \mathscr{F}$ (G. Ascoli, 1883–1884). If X is a †locally compact Hausdorff space, a necessary and sufficient condition for $\mathscr{F}$ to be relatively compact (i.e., for the closure of $\mathscr{F}$ to be compact) with respect to the compact-open topology (i.e., to the topology of uniform convergence on compact sets) is that $\mathscr{F}$ be equicontinuous at every point $x \in X$ and that the set $\{f(x) \mid f \in \mathscr{F}\}$ be relatively compact in Y for every point $x \in X$ (**Ascoli's theorem**). In particular, when X is a σ-compact locally compact Hausdorff space and Y is the space of real numbers, a family of functions $\mathscr{F}$ that are equicontinuous (at every point $x \in X$) and uniformly bounded is relatively compact. Hence, for any sequence of functions $\{f_n\}$ in $\mathscr{F}$, we can select a subsequence $\{f_{n(v)}\}$ which converges uniformly on compact sets (**Ascoli-Arzelà theorem**).

E. Normal Families

P. Montel (1912) gave the name **normal family** to the family of functions that is relatively compact with respect to the topology of uniform convergence on compact sets. This terminology is used mainly for the family of complex analytic functions. In that case, it is customary to compactify the range space and consider Y to be the †Riemann sphere. Using this notion, Montel succeeded in giving a unified treatment of various results in the theory of complex functions.

A family of analytic functions $\mathscr{F}$ on a finite-dimensional †complex manifold X is a normal family if it is uniformly bounded on each compact set (**Montel's theorem**). Another criterion is that there are three values on the Riemann sphere which no function $f \in \mathscr{F}$ takes. More generally, three exceptional values not taken by $f \in \mathscr{F}$ may depend on f, if there is a positive lower bound for the distances between these three values on the Riemann sphere. This gives an easy proof of the †Picard theorem stating that every †transcendental meromorphic function $f(z)$ in $|z| < \infty$ must take all values except possibly two values. In fact the family of functions $f_n(z) = f(z/2^n)$, $n = 1, 2, 3, \ldots$, in $\{1 < |z| < 2\}$ cannot be normal. Using a similar procedure, G. Julia obtained the results on †Julia's direction.

F. Marty introduced the notion of **spherical derivative** $|f'(z)|/(1 + |f(z)|^2)$ for the analytic or meromorphic function $f(z)$ and proved that for a family $\mathscr{F} = \{f(z)\}$ of analytic functions to be normal, it is necessary and sufficient that the spherical derivatives of $f \in \mathscr{F}$ be uniformly bounded. This theorem implies Montel's theorem and its various extensions, including, for example, quantitative results concerning †Borel's direction.

A family $\mathscr{F}$ of analytic functions of one variable defined on X is said to form a **quasi-normal family** if there exists a subset P of X consisting only of isolated points such that from any sequence $\{f_n\}$ ($f_n \in \mathscr{F}$) we can select a subsequence $\{f_{n(v)}\}$ converging uniformly on $X - P$. If P is finite and consists of p points, the family $\mathscr{F}$ is called a quasinormal family of order p. For example, the family of at most †p-valent functions is quasinormal of order p.

The theory of normal families of complex analytic functions is not only applied to †value distribution theory, as above, but also used to show the existence of a function that gives the extremal of functionals. The extremal function is usually obtained as a limit of a subsequence of a sequence in a normal family. A typical example of this method is seen in the proof of the †Riemann mapping theorem. This is perhaps the only general method known today in the study of the iteration of †holomorphic functions. By this method, Julia (1919) made an exhaustive study of the iteration of meromorphic functions; there are several other investigations on the iteration of elementary transcendental functions. On the other hand, A. Wintner (*Comm. Math. Helv.*, 23 (1949)) gave the implicit function theorem for analytic functions in a precise form using the theory of normal families of analytic functions of several variables.

References

[1] N. Bourbaki, Eléments de mathématique, III., Topologie générale, ch. 10, Espaces fonctionnels, Actualités Sci. Ind., 1084b, Hermann, second edition, 1967; English translation, Theory of sets, Addison-Wesley, 1966.
[2] J. L. Kelley, General topology, Van Nostrand, 1955, ch. 7.
[3] J. Dieudonné, Foundations of modern analysis, Academic Press, 1960, enlarged and corrected printing, 1969.
For normal families of complex functions,
[4] P. Montel, Leçons sur les familles normales de fonctions analytiques et leurs applications, Gauthier-Villars, 1927.
[5] G. Valiron, Familles normales et quasinormales de fonctions méromorphes, Gauthier-Villars, 1929.

436 (II.22) Uniform Spaces

A. Introduction

There are certain properties defined on †metric spaces but not on general †topological spaces, for example, †completeness or †uniform continuity of functions. Generalizing metric spaces, A. Weil introduced the notion of uniform spaces. This notion can be defined in several ways [3,4]. The definition in Section B is that of Weil [1] without the †separation axiom for topology.

We denote by Δ_X the **diagonal** $\{(x,x) \mid x \in X\}$ of the Cartesian product $X \times X$ of a set X with itself. If U and V are subsets of $X \times X$, then the **composite** $V \circ U$ is defined to be the set of all pairs (x, y) such that for some element z of X, the pair (x, z) is in U and the pair (z, y) is in V. The inverse U^{-1} of U is defined to be the set of all pairs (x, y) such that $(y, x) \in U$.

B. Definitions

Let $\mathscr{U}$ be a nonempty family of subsets of $X \times X$ such that (i) if $U \in \mathscr{U}$ and $U \subset V$, then $V \in \mathscr{U}$; (ii) if U, $V \in \mathscr{U}$, then $U \cap V \in \mathscr{U}$; (iii) if $U \in \mathscr{U}$, then $\Delta_X \subset U$; (iv) if $U \in \mathscr{U}$, then $U^{-1} \in \mathscr{U}$; and (v) if $U \in \mathscr{U}$, then $V \circ V \subset U$ for some $V \in \mathscr{U}$. Then we say that a **uniform structure** (or simply a **uniformity**) is defined on X by $\mathscr{U}$. If a uniformity is defined on X by $\mathscr{U}$, then the pair $(X, \mathscr{U})$ or simply the set X itself is called a **uniform space**, and $\mathscr{U}$ is usually called a **uniformity** for X.

A subfamily $\mathscr{B}$ of the uniformity $\mathscr{U}$ is called a **base for the uniformity** $\mathscr{U}$ if every member of $\mathscr{U}$ contains a member of $\mathscr{B}$. If a family $\mathscr{B}$ of subsets of $X \times X$ is a base for a uniformity $\mathscr{U}$, then the following propositions hold: (ii′) if U, $V \in \mathscr{B}$, then there exists a $W \in \mathscr{B}$ such that $W \subset U \cap V$; (iii′) if $U \in \mathscr{B}$, then $\Delta_X \subset U$; (iv′) if $U \in \mathscr{B}$, then there exists a $V \in \mathscr{B}$ such that $V \subset U^{-1}$; (v′) if $U \in \mathscr{B}$, then there exists a $V \in \mathscr{B}$ such that $V \circ V \subset U$. Conversely, if a family $\mathscr{B}$ of subsets of a Cartesian product $X \times X$ satisfies (ii′)–(v′), then the family $\mathscr{U} = \{U \mid U \subset X \times X, V \subset U \text{ for some } V \in \mathscr{B}\}$ defines a uniformity on X and $\mathscr{B}$ is a base for $\mathscr{U}$. Given a uniform space $(X, \mathscr{U})$, a member V of $\mathscr{U}$ is said to be **symmetric** if $V = V^{-1}$. The family of all symmetric members of $\mathscr{U}$ is a base for $\mathscr{U}$.

C. Topology of Uniform Spaces

Given a uniform space $(X, \mathscr{U})$, an element $x \in X$, and $U \in \mathscr{U}$, we put $U(x) = \{y \mid y \in X, (x, y) \in U\}$. Then the family $\mathscr{U}(x) = \{U(x) \mid U \in \mathscr{U}\}$ forms a neighborhood system of $x \in X$, which gives rise to a topology of X (→ 425 Topological Spaces). This topology is called the **uniform topology** (or **topology of the uniformity**). When we refer to a topology of a uniform space $(X, \mathscr{U})$, it is understood to be the uniform topology; thus a uniform space is also called a **uniform topological space**. If $\mathscr{B}$ is a base for the uniformity of a uniform space $(X, \mathscr{U})$, then $\mathscr{B}(x) = \{U(x) \mid U \in \mathfrak{B}\}$ is a base for the neighborhood system at each point $x \in X$. Each member of $\mathscr{U}$ is a subset of the topological space $X \times X$, which is supplied with the product topology. The family of all open (closed) symmetric members of $\mathscr{U}$ forms a base for $\mathscr{U}$. A uniform space $(X, \mathscr{U})$ is a †T_1-topological space if and only if the intersection of all members of $\mathscr{U}$ is the diagonal Δ_X. In this case, the uniformity of $(X, \mathscr{U})$ is called a **T_1-uniformity**, and $(X, \mathscr{U})$ is called a **T_1-uniform space**. A T_1-uniform space is always †regular; a fortiori, it is a T_2-topological space. Hence a T_1-uniform space is also said to be a **Hausdorff uniform space** (or **separated uniform space**). Moreover, a uniform topology satisfies †Tikhonov's separation axiom; in particular, a T_1-uniform space is †completely regular.

D. Examples

(1) Discrete Uniformity. Let X be a nonempty set, and let $\mathscr{U} = \{U \mid \Delta_X \subset U \subset X \times X\}$. Then $(X, \mathscr{U})$ is a T_1-uniform space and $\mathscr{B} = \{\Delta_X\}$ is a base for $\mathscr{U}$. This uniformity is called the **discrete uniformity** for X.

(2) Uniform Family of Neighborhood System. A family $\{U_\alpha(x)\}_{\alpha \in A} (x \in X)$ of subsets of a set X is called a **uniform neighborhood system** in X if it satisfies the following four requirements: (i) $x \in U_\alpha(x)$ for each $\alpha \in A$ and each $x \in X$; (ii) if x and y are distinct elements of X, then $y \notin U_\alpha(x)$ for some $\alpha \in A$; (iii) if α and β are two elements of A, then there is another element $\gamma \in A$ such that $U_\gamma(x) \subset U_\alpha(x) \cap U_\beta(x)$ for all $x \in X$; (iv) if α is an arbitrary element in A, then there is an element β in A such that $y \in U_\alpha(x)$ whenever x, $y \in U_\beta(z)$ for some z in X. If we denote by $U_\alpha (\alpha \in A)$ the subset of $X \times X$ consisting of all elements (x, y) such that $x \in X$ and $y \in U_\alpha(x)$, then the family $\{U_\alpha \mid \alpha \in A\}$ satisfies all the conditions for a base for a uniformity. In particular, it follows from (ii) that $\bigcap_{\alpha \in A} U_\alpha = \Delta_X$, which is a stronger condition than (iii′) in Section B. For instance, if $\{U_\alpha \mid \alpha \in A\}$ is a base for the neighborhood system at the identity element of a T_1-topological group G, then we have two uniform neighborhood systems $\{U_\alpha^l(x)\}$ and $\{U_\alpha^r(x)\}$, where $U_\alpha^l(x) = xU_\alpha$ and

$U_\alpha^r(x)=U_\alpha x$. Two uniformities derived from these uniform neighborhood systems are called a †left uniformity and a †right uniformity, respectively. Generally, these two uniformities do not coincide (→ 423 Topological Groups).

(3) Uniform Covering System [4]. A family $\{\mathfrak{U}_\alpha\}_{\alpha\in A}$ of †coverings of a set X is called a **uniform covering system** if the following three conditions are satisfied: (i) if $\mathfrak{U}$ is a covering of X such that $\mathfrak{U}<\mathfrak{U}_\alpha$ for all $\alpha\in A$, then $\mathfrak{U}$ coincides with the covering $\Delta=\{\{x\}\}_{x\in X}$; (ii) if $\alpha,\beta\in A$, then there is a $\gamma\in A$ such that $\mathfrak{U}_\gamma<\mathfrak{U}_\alpha$ and $\mathfrak{U}_\gamma<\mathfrak{U}_\beta$; (iii) if $\alpha\in A$, then there is a $\beta\in A$ such that $\mathfrak{U}_\beta$ is a †Δ-refinement of $\mathfrak{U}_\alpha$ ($(\mathfrak{U}_\beta)^\Delta<\mathfrak{U}_\alpha$). For an example of a uniform covering system of X, suppose that we are given a uniform neighborhood system $\{U_\alpha(x)\}_{\alpha\in A}$ $(x\in X)$. Let $\mathscr{U}_\alpha=\{U_\alpha(x)\}_{\alpha\in X}$ $(\alpha\in A)$. Then $\{\mathscr{U}_\alpha\}_{\alpha\in A}$ is a uniform covering system. On the other hand, for a covering $\mathfrak{U}=\{U_\lambda\}_{\lambda\in\Lambda}$, let $S(x,\mathfrak{U})$ be the union of all members of $\mathfrak{U}$ that contain x. If $\{\mathfrak{U}_\alpha\}_{\alpha\in A}$ is a uniform covering system and $U_\alpha(x)=S(x,\mathfrak{U}_\alpha)$, then $\{U_\alpha(x)\}_{\alpha\in A}$ $(x\in X)$ is a uniform neighborhood system. Hence defining a uniform covering system of X is equivalent to defining a T_1-uniformity on X.

(4). In a metric space (x,d) the subsets $\mathfrak{U}_r=\{(x,y)\mid d(x,Y)<r\}$, $r>0$, form a base of uniformity. The uniform topology defined by this coincides with the topology defined by the metric.

E. Some Notions on Uniform Spaces

Some of the terminology concerning topological spaces can be restated in the language of uniform structures. A mapping f from a uniform space $(X,\mathscr{U})$ into another $(X',\mathscr{U}')$ is said to be **uniformly continuous** if for each member U' in $\mathscr{U}'$ there is a member U in $\mathscr{U}$ such that $(f(x),f(y))\in U'$ for every $(x,y)\in U$. This condition implies that f is continuous with respect to the uniform topologies of the uniform spaces. Equivalently, the mapping is uniformly continuous with respect to the uniform neighborhood system $\{U_\alpha(x)\}_{\alpha\in A}$ if for any index β there is an index α such that $y\in U_\alpha(x)$ implies $f(y)\in U_\beta(f(x))$. If $f:X\rightarrow X'$ and $g:X'\rightarrow X''$ are uniformly continuous, then the composite $g\circ f:X\rightarrow X''$ is also uniformly continuous. A bijection f of a uniform space $(X,\mathscr{U})$ to another $(X',\mathscr{U}')$ is said to be a **uniform isomorphism** if both f and f^{-1} are uniformly continuous; in this case $(X,\mathscr{U})$ and $(X',\mathscr{U}')$ are said to be **uniformly equivalent**. A uniform isomorphism is a homeomorphism with respect to the uniform topologies, and a uniform equivalence defines an equivalence relation between uniform spaces.

If $\mathscr{U}_1$ and $\mathscr{U}_2$ are uniformities for a set X, we say that the uniformity $\mathscr{U}_1$ is **stronger** than the uniformity $\mathscr{U}_2$ and $\mathscr{U}_2$ is **weaker** than $\mathscr{U}_1$ if the identity mapping of $(X,\mathscr{U}_1)$ to $(X,\mathscr{U}_2)$ is uniformly continuous. The discrete uniformity is the strongest among the uniformities for a set X. The weakest uniformity for X is defined by the single member $X\times X$; this uniformity is not a T_1-uniformity unless X is a singleton. Generally, there is no weakest T_1-uniformity. A uniformity $\mathscr{U}_1$ for X is stronger than another $\mathscr{U}_2$ if and only if every member of $\mathscr{U}_2$ is also a member of $\mathscr{U}_1$.

If f is a mapping from a set X into a uniform space $(Y,\mathscr{V})$ and g is the mapping of $X\times X$ into $Y\times Y$ defined by $g(x,y)=(f(x),f(y))$, then $\mathscr{B}=\{g^{-1}(V)\mid V\in\mathscr{V}\}$ satisfies conditions (ii')–(v') in Section B for a base for a uniformity. The uniformity $\mathscr{U}$ for X determined by $\mathscr{B}$ is called the **inverse image** of the uniformity $\mathscr{V}$ for Y by f; $\mathscr{U}$ is the weakest uniformity for X such that f is uniformly continuous. Hence a mapping f from a uniform space $(X,\mathscr{U})$ into another $(Y,\mathscr{V})$ is uniformly continuous if and only if the inverse image of the uniformity $\mathscr{V}$ under f is weaker than the uniformity $\mathscr{U}$. If A is a subset of a uniform space $(X,\mathscr{U})$, then there is a uniformity $\mathscr{V}$ for A determined as the inverse image of $\mathscr{U}$ by the inclusion mapping of A into X. This uniformity $\mathscr{V}$ for A is called the **relative uniformity** for A induced by $\mathscr{U}$, or the **relativization** of $\mathscr{U}$ to A, and the uniform space $(A,\mathscr{V})$ is called a **uniform subspace** of $(X,\mathscr{U})$. The uniform topology for $(A,\mathscr{V})$ is the relative topology for A induced by the uniform topology for $(X,\mathscr{U})$.

If $\{(X_\lambda,\mathscr{U}_\lambda)\}_{\lambda\in\Lambda}$ is a family of uniform spaces, then the **product uniformity** for $X=\prod_{\lambda\in\Lambda}X_\lambda$ is defined to be the weakest uniformity $\mathscr{U}$ such that the projection of X onto each X_λ is uniformly continuous, and $(X,\mathscr{U})$ is called the **product uniform space** of $\{(X_\lambda,\mathscr{U}_\lambda)\}_{\lambda\in\Lambda}$. The topology for $(X,\mathscr{U})$ is the product of the topologies for $(X_\lambda,\mathscr{U}_\lambda)$ $(\lambda\in\Lambda)$.

F. Metrization

Each †pseudometric d for a set X generates a uniformity in the following way. For each positive number r, let $V_{d,r}=\{(x,y)\in X\times X\mid d(x,y)<r\}$. Then the family $\{V_{d,r}\mid r>0\}$ satisfies conditions (ii')–(v') in Section B for a base for a uniformity $\mathscr{U}$. This uniformity is called the **pseudometric uniformity** or **uniformity generated by** d. The uniform topology for $(X,\mathscr{U})$ is the pseudometric topology. A uniform space $(X,\mathscr{U})$ is said to be **pseudometrizable** (**metrizable**) if there is a pseudometric (metric)

d such that the uniformity $\mathscr{U}$ is identical with the uniformity generated by d. A uniform space is pseudometrizable if and only if its uniformity has a countable base. Consequently, a uniform space is metrizable if and only if its uniformity is a T_1-uniformity and has a countable base. For a family P of pseudometrics on a set X, let $V_{d,r}=\{(x,y)\in X\times X \mid d(x,y)<r\}$ for $d\in P$ and positive r. The weakest uniformity containing every $V_{d,r}$ $(d\in P, r>0)$ is called the **uniformity generated by** P. This uniformity may also be described as the weakest one such that each pseudometric in P is uniformly continuous on $X\times X$ with respect to the product uniformity.

Each uniformity $\mathscr{U}$ on a set X coincides with the uniformity generated by the family P_X of all pseudometrics that are uniformly continuous on $X\times X$ with respect to the product uniformity of $\mathscr{U}$ with itself. It follows that each uniform space is uniformly isomorphic to a subspace of a product of pseudometric spaces (in which the number of components is equal to the cardinal number of P_X) and that each T_1-uniform space is uniformly isomorphic to a subspace of a product of metric spaces. A topology τ for a set X is the uniform topology for some uniformity for X if and only if the topological space (X,τ) satisfies †Tikhonov's separation axiom; in particular, the uniformity is a T_1-uniformity if and only if (X,τ) is †completely regular.

G. Completeness

If $(X,\mathscr{U})$ is a uniform space, a subset A of X is called a **small set of order** U $(U\in\mathscr{U})$ if $A\times A\subset U$. A †filter on X is called a **Cauchy filter** (with respect to the uniformity $\mathscr{U}$) if it contains a small set of order U for each U in $\mathscr{U}$. If a filter on X converges to some point in X, then it is a Cauchy filter. If f is a uniformly continuous mapping from a uniform space X into another X', then the image of a base for a Cauchy filter on X under f is a base for a Cauchy filter on X'. A point contained in the closure of every set in a Cauchy filter $\mathfrak{F}$ is the limit point of $\mathfrak{F}$. Hence if a filter converges to x, a Cauchy filter contained in the filter also converges to x.

A †net $x(\mathfrak{A})=\{x_\alpha\}_{\alpha\in\mathfrak{A}}$ (where $\mathfrak{A}$ is a directed set with a preordering $\leqslant$) in a uniform space $(X,\mathscr{U})$ is called a **Cauchy net** if for each U in $\mathscr{U}$ there is a γ in $\mathfrak{A}$ such that $(x_\alpha,x_\beta)\in U$ for every α and β such that $\gamma\leqslant\alpha$, $\gamma\leqslant\beta$. If $\mathfrak{A}$ is the set $\mathbf{N}$ of all natural numbers, a Cauchy net $\{x_n\}_{n\in\mathbf{N}}$ is called a **Cauchy sequence** (or **fundamental sequence**). Given a Cauchy net $\{x_\alpha\}_{\alpha\in\mathfrak{A}}$, let $A_\alpha=\{x_\beta\mid\beta\geqslant\alpha\}$. Then $\mathfrak{B}=\{A_\alpha\mid\alpha\in\mathfrak{A}\}$ is a base for a filter, and the filter is a Cauchy filter. On the other hand, let $\mathfrak{B}$ be a base for a Cauchy filter $\mathfrak{F}$. For $U,V\in\mathfrak{B}$, we put $U\leqslant V$ if and only if $U\supset V$. Then $\mathfrak{B}$ is a directed set with respect to $\leqslant$. The net $\{x_U\}_{U\in\mathfrak{B}}$, where x_U is an arbitrary point in U, is a Cauchy net. A proposition concerning convergence of a Cauchy filter is always equivalent to a proposition concerning convergence of the corresponding Cauchy net.

A Cauchy filter (or Cauchy net) in a uniform space X does not always converge to a point of X. A uniform space is said to be **complete** (with respect to the uniformity) if every Cauchy filter (or Cauchy net) converges to a point of that space. A complete uniform space is called for brevity a **complete space**. A closed subspace of a complete space is complete with respect to the relative uniformity. A pseudometrizable uniform space is complete if and only if every Cauchy sequence in the space converges to a point. Hence in the case of a metric space, our definition of completeness coincides with the usual one (→ 273 Metric Spaces).

A mapping f from a uniform space X to another X' is said to be **uniformly continuous on a subset** A of X if the restriction of f to A is uniformly continuous with respect to the relative uniformity for A. If f is a uniformly continuous mapping from a subset A of a uniform space into a complete T_1-uniform space, then there is a unique uniformly continuous extension $\bar{f}$ of f on the closure $\bar{A}$.

Each T_1-uniform space is uniformly equivalent to a dense subspace of a complete T_1-uniform space; this property is a generalization of the fact that each metric space can be mapped by an isometry onto a dense subset of a complete metric space. **A completion** of a uniform space $(X,\mathscr{U})$ is a pair $(f,(X^*,\mathscr{U}^*))$, where $(X^*,\mathscr{U}^*)$ is a complete space and f is a uniform isomorphism of X onto a dense subspace of X^*. The T_1-completion of a T_1-uniform space is unique up to uniform equivalence.

H. Compact Spaces

A uniformity $\mathscr{U}$ for a topological space (X,τ) is said to be **compatible** with the topology τ if the uniform topology for $(X,\mathscr{U})$ coincides with τ. A topological space (X,τ) is said to be **uniformizable** if there is a uniformity compatible with τ. If (X,τ) is a compact Hausdorff space, then there is a unique uniformity $\mathscr{U}$ compatible with τ; in fact, $\mathscr{U}$ consists of all neighborhoods of the diagonal Δ_X in $X\times X$; and the compact Hausdorff space is complete with this uniformity. Hence every subspace of a compact Hausdorff space is uniformizable, and every †locally compact Hausdorff space is

uniformizable. Any continuous mapping from a compact Hausdorff space to a uniform space is uniformly continuous. A uniform space $(X, \mathscr{U})$ is said to be **totally bounded** (or **precompact**) if for each $U \in \mathscr{U}$ there is a finite covering consisting of small sets of order U; a subset of a uniform space is called **totally bounded** if it is totally bounded with respect to the relative uniformity. A uniform space X is said to be **locally totally bounded** if for each point of X there is a base for a neighborhood system consisting of totally bounded open subsets. A uniform space is compact if and only if it is totally bounded and complete. If f is a uniformly continuous mapping from a uniform space X to another, then the image $f(A)$ of a totally bounded subset A of X is totally bounded.

I. Topologically Complete Spaces

A topological space (X, τ) is said to be **topologically complete** (or **Dieudonné complete**) if it admits a uniformity compatible with τ with respect to which X is complete. Each †paracompact Hausdorff space is topologically complete. Actually such a space is complete with respect to its strongest uniformity. A Hausdorff space which is homeomorphic to a †G_δ-set in a compact Hausdorff space is said to be **Čech-complete**; A metric space is homeomorphic to a complete metric space if and only if it is Čech-complete. A Hausdorff space X is paracompact and Čech-complete if and only if there is a †perfect mapping from X onto a complete metric space.

References

[1] A. Weil, Sur les espaces à structure uniforme et sur la topologie générale, Actualités Sci. Ind., Hermann, 1938.
[2] J. W. Tukey, Convergence and uniformity in topology, Ann. Math. Studies, Princeton Univ. Press, 1940.
[3] J. L. Kelley, General topology, Van Nostrand, 1955.
[4] N. Bourbaki, Eléments de mathématique, III. Topologie générale, ch. 2, Actualités Sci. Ind., 1142d, Hermann, fourth edition 1965; English translation, General topology, Addision-Wesley, 1966.
[5] J. R. Isbell, Uniform spaces, Amer. Math. Soc. Math. Surveys, 1964.
[6] H. Nakano, Uniform spaces and transformation groups, Wayne State Univ. Press, 1968.
[7] R. Emgelking, General topology, Polish Scientific Publishers, 1977.
[8] Z. Frolík, Generalization of the G_δ-property of complete metric spaces, Czech. Math. J., 10 (1960), 359–379.

437 (IV.17) Unitary Representations

A. Definitions

A homomorphism U of a †topological group G into the group of †unitary operators on a †Hilbert space $\mathfrak{H}$ $(\neq \{0\})$ is called a **unitary representation** of G if U is **strongly continuous** in the following sense: For any element $x \in \mathfrak{H}$, the mapping $g \rightarrow U_g x$ is a continuous mapping from G into $\mathfrak{H}$. The Hilbert space $\mathfrak{H}$ is called the **representation space** of U and is denoted by $\mathfrak{H}(U)$. Two unitary representations U and U' are said to be **equivalent** (**similar** or **isomorphic**), denoted by $U \cong U'$, if there exists an †isometry T from $\mathfrak{H}(U)$ onto $\mathfrak{H}(U')$ that satisfies the equality $T \circ U_g = U'_g \circ T$ for every g in G. If the representation space $\mathfrak{H}(U)$ contains no closed subspace other than $\mathfrak{H}$ and $\{0\}$ that is invariant under every U_g, the unitary representation U is said to be **irreducible**. An element x in $\mathfrak{H}(U)$ is called a **cyclic vector** if the set of all finite linear combinations of the elements $U_g x (g \in G)$ is dense in $\mathfrak{H}(U)$. A representation U having a cyclic vector is called a **cyclic representation**. Every nonzero element of the representation space of an irreducible representation is a cyclic vector.

Examples. Let G be a †topological transformation group acting on a †locally compact Hausdorff space X from the right. Suppose that there exists a †Radon measure μ that is invariant under the group G. Then a unitary representation R^μ is defined on the Hilbert space $\mathfrak{H} = L^2(X, \mu)$ by the formula $(R^\mu_g f)(x) = f(xg)$ $(f \in \mathfrak{H}, x \in X, g \in G)$. The representation R^μ is called the **regular representation** of G on (X, μ). If G acts on X from the left, then the regular representation L^μ is defined by $(L^\mu_g f)(x) = f(g^{-1}x)$. In particular, when X is the †quotient space $H \backslash G$ of a †locally compact group G by a closed subgroup H, any two invariant measures μ, μ' (if they exist) coincide up to a constant factor. Hence the regular representation R^μ on (X, μ) and the regular representation $R^{\mu'}$ on (X, μ') are equivalent. In this case, the representation R^μ is called the regular representation on X. When $H = \{e\}$, a locally compact group G has a Radon measure $\mu \neq 0$ that is invariant under every right (left) translation $h \rightarrow hg$ $(h \rightarrow gh)$ and is called a right (left) †Haar measure on G. So G has the regu-

lar representation R (L) on G. R (L) is called the right (left) regular representation of G.

B. Positive Definite Functions and Existence of Representations

A complex-valued continuous function φ on a topological group G is called **positive definite** if the matrix having $\varphi(g_i^{-1}g_j)$ as the (i,j)-component is a †positive semidefinite Hermitian matrix for any finite number of elements $g_1, \ldots, g_n$ in G. If U is a unitary representation of G, then the function $\varphi(g)=(U_g x, x)$ is positive definite for every element x in $\mathfrak{H}(U)$. Conversely, any positive definite function $\varphi(g)$ on a topological group G can be expressed as $\varphi(g)=(U_g x, x)$ for some unitary representation U and x in $\mathfrak{H}(U)$. Using this fact and the †Kreĭn-Milman theorem, it can be proved that every locally compact group G has **sufficiently many irreducible unitary representations** in the following sense: For every element g in G other than the identity element e, there exists an irreducible unitary representation U, generally depending on g, that satisfies the inequality $U_g \neq 1$. The groups having sufficiently many finite-dimensional (irreducible) unitary representations are called †maximally almost periodic. If a connected locally compact group G is maximally almost periodic, then G is the direct product of a compact group and a vector group $\mathbf{R}^m$. On the other hand, any noncompact connected †simple Lie group has no finite-dimensional irreducible unitary representation other than the unit representation $g \rightarrow 1$ (→ 18 Almost Periodic Functions).

C. Subrepresentations

Let U be a unitary representation of a topological group G. A closed subspace $\mathfrak{N}$ of $\mathfrak{H}(U)$ is called **U-invariant** if $\mathfrak{N}$ is invariant under every U_g $(g \in G)$. Let $\mathfrak{N} \neq \{0\}$ be a closed invariant subspace of $\mathfrak{H}(U)$ and V_g be the restriction of U_g on $\mathfrak{N}$. Then V is a unitary representation of G on the representation space $\mathfrak{N}$ and is called a **subrepresentation** of U. Two unitary representations L and M are called **disjoint** if no subrepresentation of L is equivalent to a subrepresentation of M; they are called **quasi-equivalent** if no subrepresentation of L is disjoint from M and no subrepresentation of M is disjoint from L.

D. Irreducible Representations

Let U be a unitary representation of G, $\mathbf{M}$ be the †von Neumann algebra generated by $\{U_g \mid g \in G\}$, and $\mathbf{M}'$ be the †commutant of $\mathbf{M}$. Then a closed subspace $\mathfrak{N}$ of $\mathfrak{H}(U)$ is invariant under U if and only if the †projection operator P corresponding to $\mathfrak{N}$ belongs to $\mathbf{M}'$. Therefore U is irreducible if and only if $\mathbf{M}'$ consists of scalar operators $\{\alpha 1 \mid \alpha \in \mathbf{C}\}$ **(Schur's lemma)**. A representation space of a cyclic or irreducible representation of a †separable topological group is †separable.

E. Factor Representations

A unitary representation U of G is called a **factor representation** if the von Neumann algebra $\mathbf{M}=\{U_g \mid g \in G\}$ is a †factor, that is, $\mathbf{M} \cap \mathbf{M}'=\{\alpha 1 \mid \alpha \in \mathbf{C}\}$. Two factor representations are quasi-equivalent if and only if they are not disjoint. U is called a **factor representation of type I, II,** or **III** if the von Neumann algebra $\mathbf{M}$ is a factor of †type I, II, or III, respectively (→ 308 Operator Algebras). A topological group G is called a **group of type I** (or **type I group**) if every factor representation of G is of type I. Compact groups, locally compact Abelian groups, connected †nilpotent Lie groups, connected †semisimple Lie groups, and real or complex †linear algebraic groups are examples of groups of type I. There exists a connected solvable Lie group that is not of type I (→ Section U), but a connected solvable Lie group is of type I if the exponential mapping is surjective (O. Takenouchi). A discrete group G with countably many elements is a type I group if and only if G has an Abelian normal subgroup with finite index (E. Thoma).

F. Representation of Direct Products

Let G_1 and G_2 be topological groups, G the †direct product of G_1 and G_2 $(G=G_1 \times G_2)$, and U_i an irreducible unitary representation of G_i $(i=1,2)$. Then the †tensor product representation $U_1 \otimes U_2 : (g_1, g_2) \rightarrow U_{g_1} \otimes U_{g_2}$ is an irreducible unitary representation of G. Conversely, if one of the groups G_1 and G_2 is of type I, then every irreducible unitary representation of G is equivalent to the tensor product $U_1 \otimes U_2$ of some irreducible representations U_i of G_i $(i=1,2)$.

G. Direct Sums

If the representation space $\mathfrak{H}$ of a unitary representation U is the †direct sum $\bigoplus_{\alpha \in I} \mathfrak{H}(\alpha)$ of mutually orthogonal closed invariant subspaces $\{\mathfrak{H}(\alpha)\}_{\alpha \in I}$, then U is called the **direct sum** of the subrepresentations $U(\alpha)$ induced on $\mathfrak{H}(\alpha)$ by U, and is denoted by $U=\bigoplus_{\alpha \in I} U(\alpha)$. Any unitary representation is the direct sum of cyclic representations. A unitary representa-

tion U is called a **representation without multiplicity** if U cannot be decomposed as a direct sum $U_1 \oplus U_2$ unless U_1 and U_2 are disjoint. If U is the direct sum of $\{U(\alpha)\}_{\alpha\in I}$ and every $U(\alpha)$ is irreducible, then U is said to be **decomposed into the direct sum of irreducible representations**. Decomposition into direct sums of irreducible representations is essentially unique if it exists; that is, if $U = \bigoplus_{\alpha\in I} U(\alpha) = \bigoplus_{\beta\in J} V(\beta)$ are two decompositions of U into direct sums of irreducible representations, then there exists a bijection φ from I onto J such that $U(\alpha)$ is equivalent to $V(\varphi(\alpha))$ for every α in I. A factor representation U of type I can be decomposed as the direct sum $U = \bigoplus_{\alpha\in I} U(\alpha)$ of equivalent irreducible representations $U(\alpha)$. In general, a unitary representation U cannot be decomposed as the direct sum of irreducible representations even if U is not irreducible. Thus it becomes necessary to use direct integrals to obtain an irreducible decomposition.

H. Direct Integrals

Let U be a unitary representation of a group G and (X,μ) be a †measure space. Assume that the following two conditions are satisfied by U: (i) There exists a unitary representation $U(x)$ of G corresponding to every element x of X, and $\mathfrak{H}(U)$ is a †direct integral (→ 308 Operator Algebras) of $\mathfrak{H}(U(x))$ $(x\in X)$ (written $\mathfrak{H}(U)=\int_X \mathfrak{H}(U(x))\,d\mu(x)$); (ii) for every g in G, the operator U_g is a decomposable operator and can be written as $U_g=\int_X U_g(x)\,d\mu(x)$. Then the unitary representation U is called the **direct integral** of the family $\{U(x)\}_{x\in X}$ of unitary representations and is denoted by $U=\int_X U(x)\,d\mu(x)$. If every point of X has measure 1, then a direct integral is reduced to a direct sum.

I. Decomposition into Factor Representations

We assume that G is a locally compact group satisfying the †second countability axiom, and also that a Hilbert space is separable. Every unitary representation U of G can be decomposed as a direct integral $U=\int_X U(x)\,d\mu(x)$ in such a way that the center $\mathbf{A}$ of the von Neumann algebra $\mathbf{M}''=\{U_g \mid g\in G\}''$ is the set of all †diagonalizable operators. In this case almost all the $U(x)$ are factor representations. Such a decomposition of U is essentially unique. There exists a †null set N in X such that for every x and x' in $X-N$ $(x\neq x')$, $U(x)$ and $U(x')$ are mutually disjoint factor representations. Hence the space X can be identified with the set G^* of all quasi-equivalence classes of factor representations of G endowed with a suitable structure of a measure space. The space G^* is called the **quasidual** of G. The measure μ is determined by U up to †equivalence of measures.

J. Duals

A topology is introduced on the set $\hat{G}$ of all equivalence classes of irreducible unitary representations of a locally compact group G in the following way. Let H_n be the n-dimensional Hilbert space $l_2(n)$ and I_n the set of all irreducible unitary representations of G realized on H_n $(1\leqslant n\leqslant\infty)$. We topologize I_n in such a way that a †net $\{U^\lambda\}_{\lambda\in L}$ in I_n converges to U if and only if $(U_g^\lambda x, y)$ converges uniformly to $(U_g x, y)$ on every compact subset of G for any x and y in H_n. Equivalence between representations in I_n is an open relation. Let $\hat{G}_n$ be the set of all equivalence classes of n-dimensional irreducible unitary representations of G with the topology of a quotient space of I_n and $\hat{G}=\bigcup_n \hat{G}_n$ be the direct sum of topological spaces $\hat{G}_n$. Then the topological space $\hat{G}$ is called the **dual** of G. $\hat{G}$ is a locally compact †Baire space with countable open base, but it does not satisfy the †Hausdorff separation axiom in general. If G is a compact Hausdorff topological group, then $\hat{G}$ is discrete. If G is a locally compact Abelian group, then $\hat{G}$ coincides with the †character group of G in the sense of Pontryagin. If G is a type I group, then there exists a dense open subset of $\hat{G}$ that is a locally compact Hausdorff space. The †σ-additive family generated by closed sets in $\hat{G}$ is denoted by $\mathfrak{B}$. In the following sections, a measure on $\hat{G}$ means a measure defined on $\mathfrak{B}$.

K. Irreducible Decompositions

In this section G is assumed to be a locally compact group of type I with countable open base. For any equivalence class x in $\hat{G}$, we choose a representative $U(x)\in x$ with the representation space $H(U(x))=l_2(n)$ if x is n-dimensional. For any measure μ on $\hat{G}$, the representation $U^\mu=\int_{\hat{G}} U(x)\,d\mu(x)$ is a unitary representation without multiplicity. Conversely, any unitary representation of G without multiplicity is equivalent to a U^μ for some measure μ on $\hat{G}$. Moreover, U^μ is equivalent to U^ν if and only if the two measures μ and ν are equivalent (that is, μ is absolutely continuous with respect to ν, and vice versa). A unitary representation U with multiplicity on a separable Hilbert space $\mathfrak{H}$ can be decomposed as follows: There exists a countable set of measures $\mu_1, \mu_2, \dots, \mu_\infty$ whose supports are mutu-

ally disjoint such that $U \cong \int_{\hat{G}} U(x) d\mu_1(x) \oplus 2\int_{\hat{G}} U(x) d\mu_2(x) \oplus \ldots \oplus \infty \int_{\hat{G}} U(x) d\mu_\infty(x)$. The measures $\mu_1, \mu_2, \ldots, \mu_\infty$ are uniquely determined by U up to equivalence of measures. Any unitary representation U on a separable Hilbert space $\mathfrak{H}$ of an arbitrary locally compact group with countable open base (even if not of type I) can be decomposed as a direct integral of irreducible representations. In order to obtain such a decomposition, it is sufficient to decompose $\mathfrak{H}$ as a direct integral in such a way that a maximal Abelian von Neumann subalgebra $\mathbf{A}$ of $\mathbf{M}' = \{U_g | g \in G\}'$ is the set of all diagonalizable operators. In this case, however, a different choice of $\mathbf{A}$ induces in general an essentially different decomposition, and uniqueness of the decomposition does not hold. For a group of type I, the irreducible representations are the "atoms" of representations, as in the case of compact groups. For a group not of type I, it is more natural to take the factor representations for the irreducible representations, quasi-equivalence for the equivalence, and the quasidual for the dual of G. Therefore the theory of unitary representations for a group not of type I has different features from the one for a type I group. The theory of unitary representation for groups not of type I has not yet been successfully developed, but some important results have been obtained (e.g., L. Pukanszky, *Ann. Sci. Ecole Norm. Sup.*, 4 (1971)).

Tatsuuma [1] proved a duality theorem for general locally compact groups which is an extension of both Pontryagin's and Tannaka's duality theorems considering the direct integral decomposition of tensor product representations.

L. The Plancherel Formula

Let G be a unimodular locally compact group with countable open base, $R(L)$ be the right (left) regular representation of G, and $\mathbf{M}$, $\mathbf{N}$, and $\mathbf{P}$ be the von Neumann algebras generated by $\{R_g\}$, $\{L_g\}$, and $\{R_g, L_g\}$, respectively. Then $\mathbf{M}' = \mathbf{N}$, $\mathbf{N}' = \mathbf{M}$, and $\mathbf{P}' = \mathbf{M} \cap \mathbf{N}$. If we decompose $\mathfrak{H}$ into a direct integral in such a way that $\mathbf{P}'$ is the algebra of all diagonalizable operators, then $\mathbf{M}(x)$ and $\mathbf{N}(x)$ are factors for almost all x. This decomposition of $\mathfrak{H}$ produces a decomposition of the two-sided regular representation $\{R_g, L_g\}$ into irreducible representations and a decomposition of the regular representation $R(L)$ into factor representations. Hence the decomposition is realized as the direct integral over the quasidual G^* of G. Moreover, the factors $\mathbf{M}(x)$ and $\mathbf{N}(x)$ are of type I or II for almost all x in G^*, and there exists a †trace t in the factor $\mathbf{M}(x)$. For any f and g in $L_1(G) \cap L_2(G)$, the **Plancherel formula**

$$\int_G f(s)\overline{g(s)}\, ds = \int_{G^*} t(U_g^*(x) U_f(x))\, d\mu(x) \tag{1}$$

holds, where $U_f(x) = \int_G f(s) U_s(x) ds$ and U^* is the †adjoint of U. The **inversion formula**

$$h(s) = \int_{G^*} t(U_h(x) U_s^*(x))\, d\mu(x) \tag{2}$$

is derived from (1) for a function $h = f * g$ ($f, g \in L_1(G) \cap L_2(G)$). In (1) and (2), because of the impossibility of normalization of the trace t in a factor of type II_∞, the measure μ cannot in general be determined uniquely. However, if G is a type I group, then (1) and (2) can be rewritten as similar formulas, where the representation $U(x)$ in (1) and (2) is irreducible, the trace t is the usual trace, and the domain of integration is not the quasidual G^* but the dual $\hat{G}$ of G. The revised formula (1) is also called the Plancherel formula. In this case the measure μ on $\hat{G}$ in formulas (1) and (2) is uniquely determined by the given Haar measure on G. The measure μ is called the **Plancherel measure** of G. The support $\hat{G}_r$ of the Plancherel measure μ is called the **reduced dual** of G. The Plancherel formula gives the direct integral decomposition of the regular representation into the irreducible representations belonging to $\hat{G}_r$. Each U in $\hat{G}_r$ is contained in this decomposition, with the multiplicity equal to $\dim \mathfrak{H}(U)$.

M. Square Integrable Representations

An irreducible unitary representation U of a unimodular locally compact group G is said to be **square integrable** when for some element $x \neq 0$, in $\mathfrak{H}(U)$, the function $\varphi(g) =: (U_g x, x)$ belongs to $L^2(G, dg)$, where dg is the Haar measure of G. If U is square integrable, then $\varphi_{x,y}(g) = (U_g x, y)$ belongs to $L^2(G, dg)$ for any x and y in $\mathfrak{H}(U)$. Let U and U' be the two square integrable representations of G. Then the following **orthogonality relations** hold:

$$\int_G (U_g x, y)\overline{(U'_g u, v)}\, dg = \begin{cases} 0 & \text{if } U \text{ is not equivalent to } U', \\ d_U^{-1}(x,u)(v,y) & \text{if } U = U'. \end{cases} \tag{3}$$

When G is compact, every irreducible unitary representation U is square integrable and finite-dimensional. Moreover, the scalar d_U in (3) is the degree of U if the total measure of G is normalized to 1. In the general case, the scalar d_U in (3) is called the **formal degree** of U and is determined uniquely by the given Haar

measure dg. Let y be an element in $\mathfrak{H}(U)$ with norm 1 and V be the subspace $\overline{\{\varphi_{x,y} \mid x \in \mathfrak{H}(U)\}}$ of $L^2(G)$. Then the linear mapping $T: x \to \sqrt{d_U}\, \varphi_{x,y}$ is an isometry of $\mathfrak{H}(U)$ onto V. Hence U is equivalent to a subrepresentation of the right regular representation R of G. Conversely, every irreducible subrepresentation of R is square integrable. Thus a square integrable representation is an irreducible subrepresentation of R ($\cong L$). Therefore, in the irreducible decomposition of R, the square integrable representations appear as discrete direct summands. Hence every square integrable representation U has a positive Plancherel measure $\mu(U)$ that is equal to the formal degree d_U. There exist noncompact groups that have square integrable representations. An example of such a group is $SL(2, \mathbf{R})$ (→ Section X).

N. Representations of $L_1(G)$

Let G be a locally compact group and $L_1(G)$ be the space of all complex-valued integrable functions on G. Then $L_1(G)$ is an algebra over $\mathbf{C}$, where the convolution

$$(f*g)(s) = \int_G f(st^{-1})g(t)\,dt$$

is defined to be the product of f and g. Let Δ be the †modular function of G. Then the mapping $f(s) \to f^*(s) = \Delta(s^{-1})\overline{f(s^{-1})}$ is an †involution of the algebra $L_1(G)$. Let U be a unitary representation of G, and put $U'_f = \int_G U_s f(s)\,ds$. Then the mapping $f \to U'_f$ gives a **nondegenerate representation** of the Banach algebra $L_1(G)$ with an involution, where nondegenerate means that $\{U'_f x \mid f \in L_1(G), x \in \mathfrak{H}(U)\}^{\perp}$ reduces to $\{0\}$. The mapping $U \to U'$ gives a bijection between the set of equivalence classes of unitary representations of G and the set of equivalence classes of nondegenerate representations of the Banach algebra $L_1(G)$ with an involution on Hilbert spaces. U is an irreducible (factor) representation if and only if U' is an irreducible (factor) representation. Therefore the study of unitary representations of G reduces to that of representations of $L_1(G)$. If U'_f is a †compact operator for every f in $L_1(G)$, then U is the discrete direct sum of irreducible representations, and the multiplicity of every irreducible component is finite. (See [2] for Sections A–N.)

O. Induced Representations

Induced representation is the method of constructing a representation of a group G in a canonical way from a representation of a subgroup H of G. It is a fundamental method of obtaining a unitary representation of G. Let G be a locally compact group satisfying the second countability axiom, L be a unitary representation on a separable Hilbert space $\mathfrak{H}(L)$ of a closed subgroup H of G, and m, n, Δ, and δ be the right Haar measures and the modular functions of the groups G and H, respectively. Then there exists a continuous positive function ρ on G satisfying $\rho(hg) = \delta(h)\Delta(h)^{-1}\rho(g)$ for every h in H and g in G. The †quotient measure $\mu = (\rho m)/n$ is a quasi-invariant measure on the coset space $H \backslash G$ (→ 225 Invariant Measures). Let $\mathfrak{H}$ be the vector space of weakly measurable functions f on G with values in $\mathfrak{H}(L)$ satisfying the following two conditions: (i) $f(hg) = L_h f(g)$ for every h in H and g in G; and (ii) $\|f\|^2 = \int_{H\backslash G} \|f(g)\|^2\, d\mu(\dot{g}) < +\infty$, where $\dot{g}$ represents the coset Hg. By condition (i), the norm $\|f(g)\|$ is constant on a coset $Hg = \dot{g}$ and is a function on $H \backslash G$, so the integral in condition (ii) is well defined. Then $\mathfrak{H}$ is a Hilbert space with the norm defined in (ii). A unitary representation U of G on the Hilbert space $\mathfrak{H}$ is defined by the formula

$$(U_s f)(g) = \sqrt{\rho(gs)/\rho(g)}\; f(gs).$$

U is called the **unitary representation induced by the representation L of a subgroup** H and is denoted by $U = U^L$ or $\mathrm{Ind}_H^G L$. **Induced representations** have the following properties.

(1) $U^{L_1 \oplus L_2} \cong U^{L_1} \oplus U^{L_2}$ or more generally, $U^{\int U(x) d\mu(x)} \cong \int U^{L(x)} d\mu(x)$. Therefore if U^L is irreducible, L is also irreducible (the converse does not hold in general).

(2) Let H, K be two subgroups of G such that $H \subset K$, L be a unitary representation of H, and M be the representation of K induced by L. Then two unitary representations U^M and U^L of G are equivalent.

An induced representation U^L is the representation on the space of square integrable sections of the †vector bundle with fiber $H(L)$ †associated with the principal bundle $(G, H\backslash G, H)$ (→ G. W. Mackey [3], F. Bruhat [4]).

P. Unitary Representations of Special Groups

In the following sections we describe the fundamental results on the unitary representations of certain special groups.

Q. Compact Groups

Irreducible unitary representations of a compact group are always finite-dimensional. Every unitary representation of a compact group is decomposed into the direct sum of irreducible representations. Irreducible unitary representations of a compact connected Lie

group are completely classified. The characters of irreducible representations are calculated in an explicit form (→ 69 Compact Groups; 249 Lie Groups). Every irreducible unitary representation U of a connected compact Lie group G can be extended uniquely to an irreducible holomorphic representation U^C of the complexification G^C of G. U^C is holomorphically induced from a 1-dimensional representation of a Borel subgroup B of G^C (**Borel-Weil theorem**; → R. Bott [5]).

R. Abelian Groups

Every irreducible unitary representation of an Abelian group G is 1-dimensional. †Stone's theorem concerning one-parameter groups of unitary operators, $U_t=\int_{-\infty}^{\infty} e^{i\lambda t}dE_\lambda$, gives irreducible decompositions of unitary representations of the additive group $\mathbf{R}$ of real numbers. †Bochner's theorem on †positive definite functions on $\mathbf{R}$ is a restatement of Stone's theorem in terms of positive definite functions. The theory of the †Fourier transform on $\mathbf{R}$, in particular †Plancherel's theorem, gives the irreducible decomposition of the regular representation of $\mathbf{R}$. The theorems of Stone, Bochner, and Plancherel have been extended to an arbitrary locally compact Abelian group (→ 192 Harmonic Analysis).

S. Representations of Lie Groups and Lie Algebras

Let U be a unitary representation of a Lie group G with the Lie algebra $\mathfrak{g}$. An element x in $\mathfrak{H}(U)$ is called an **analytic vector** with respect to U if the mapping $g\rightarrow U_g x$ is a real analytic function on G with values in $\mathfrak{H}(U)$. The set of all analytic vectors with respect to U forms a dense subspace $\mathfrak{A}=\mathfrak{A}(U)$ of $\mathfrak{H}(U)$. For any elements X in $\mathfrak{g}$ and x in $\mathfrak{A}(U)$, the derivative at $t=0$ of a real analytic function $U_{\exp tX}x$ is denoted by $V(X)x$. Then $V(X)$ is a linear transformation on $\mathfrak{A}$, and the mapping $V:X\rightarrow V(X)$ is a representation of $\mathfrak{g}$ on $\mathfrak{A}$. We call V the **differential representation** of U. The representation V of $\mathfrak{g}$ can be extended uniquely to a representation of the †universal enveloping algebra $\mathfrak{B}$ of $\mathfrak{g}$. Two unitary representations $U^{(1)}$ and $U^{(2)}$ of a connected Lie group G are equivalent if and only if there exists a bijective bounded linear mapping T from $\mathfrak{H}(U^{(1)})$ onto $\mathfrak{H}(U^{(2)})$ such that T maps $\mathfrak{A}(U^{(1)})$ onto $\mathfrak{A}(U^{(2)})$ and satisfies the equality

$$(T\circ V^{(1)}(X))x=(V^{(2)}(X)\circ T)x$$

for all X in $\mathfrak{g}$ and x in $\mathfrak{A}(U^{(1)})$. Let $X_1,\ldots,X_n$ be a basis of $\mathfrak{g}$ and U be a unitary representation of G. Then the element $\Delta=X_1^2+\ldots+X_n^2$ in the universal enveloping algebra $\mathfrak{B}$ of $\mathfrak{g}$ is represented in the differential representation V of U by an †essentially self-adjoint operator $V(\Delta)$. Conversely, if to each element X in $\mathfrak{g}$ there corresponds a (not necessarily bounded) †skew-Hermitian operator $\rho(x)$ that satisfies the following three conditions, then there exists a unique unitary representation U of the simply connected Lie group G with the Lie algebra $\mathfrak{g}$ such that the †closure of $V(X)$ coincides with the closure of $\rho(X)$ for every X in $\mathfrak{g}$: (i) There exists a dense subspace $\mathfrak{D}$ contained in the domain of $\rho(X)\rho(Y)$ for every X and Y in $\mathfrak{g}$; (ii) for each X and Y in $\mathfrak{g}$, a and b in $\mathbf{R}$, and x in $\mathfrak{D}$, $\rho(aX+bY)x=a\rho(X)x+b\rho(Y)x$, $\rho([X,Y])x=(\rho(X)\rho(Y)-\rho(Y)\rho(X))x$; (iii) the restriction of $\rho(X_1)^2+\ldots+\rho(X_n)^2$ to $\mathfrak{D}$ is an essentially self-adjoint operator if $X_1,\ldots,X_n$ is a basis of $\mathfrak{g}$ (E. Nelson [6]).

T. Nilpotent Lie Groups

For every irreducible unitary representation of a connected nilpotent Lie group G, there is some 1-dimensional unitary representation of some subgroup of G that induces it. Let G be a simply connected nilpotent Lie group, $\mathfrak{g}$ be the Lie algebra of G, and ρ be the contragredient representation of the adjoint representation of G. The representation space of ρ is the dual space $\mathfrak{g}^*$ of $\mathfrak{g}$. A subalgebra $\mathfrak{h}$ of $\mathfrak{g}$ is called **subordinate** to an element f in $\mathfrak{g}^*$ if f annihilates each bracket $[X,Y]$ for every X and Y in $\mathfrak{h}$: $(f,[X,Y])=0$. When $\mathfrak{h}$ is subordinate to f, a 1-dimensional unitary representation L of the analytic subgroup H of G with the Lie algebra $\mathfrak{h}$ is defined by the formula $\lambda_f(\exp X)=e^{2\pi i(f,X)}$ $(X\in\mathfrak{h})$. Every 1-dimensional unitary representation λ_f of H is defined as in this formula by an element f in $\mathfrak{g}^*$ to which $\mathfrak{h}$ is subordinate. The unitary representation of G induced by such a λ_f is denoted by $U(f,\mathfrak{h})$. The representation $U(f,\mathfrak{h})$ is irreducible if and only if $\mathfrak{h}$ has maximal dimension among the subalgebras subordinate to f. Two irreducible representations $U(f,\mathfrak{h})$ and $U(f,\mathfrak{h}')$ are equivalent if and only if f and f' are conjugate under the group $\rho(G)$. Therefore there exists a bijection between the set of equivalence classes of the irreducible unitary representations of a simply connected nilpotent Lie group G and the set of orbits of $\rho(G)$ on $\mathfrak{g}^*$ (A. A. Kirillov [7]).

U. Solvable Lie Groups

Let G be a simply connected solvable Lie group. If the exponential mapping is bijective, G is called an **exponential group**. All results stated above for nilpotent Lie groups hold for exponential groups except the irreducibility

criterion. In this case the representation $U(f,\mathfrak{h})$ is irreducible if and only if $\mathfrak{h}$ is of maximal dimension among subordinate subalgebras and the orbit $O=\rho(G)f$ contains the affine subspace $f+\mathfrak{h}^{\perp}=f+\{g\mid g(\mathfrak{h})=0\}$ (Pukanszky condition).

The situation is more complicated for general solvable Lie groups. The isotropy subgroup $G_f=\{g\in G\mid\rho(g)f=f\}$ at $f\in\mathfrak{g}^*$ is, in general, not connected. A linear form f is called integral if there exists a unitary character η_f of G_f whose differential is the restriction of $2\pi i f$ to $\mathfrak{g}_f$ (the Lie algebra of G_f). Using the notion of "polarization," an irreducible unitary representation of G is constructed from a pair (f,η_f) of an integral form $f\in\mathfrak{g}^*$ and a character η_f. If G is of type I, then every irreducible unitary representation of G is obtained in this way. A simply connected solvable Lie group G is of type I if and only if (i) every $f\in\mathfrak{g}^*$ is integral and (ii) every G-orbit $\rho(G)f$ in $\mathfrak{g}^*$ is locally closed (Auslander and Kostant [8]).

As an example, let α be an irrational real number. Then the following Lie group G is not of type I: $G=\left\{\begin{pmatrix} e^{it} & 0 & z \\ 0 & e^{i\alpha t} & w \\ 0 & 0 & 1\end{pmatrix} \,\middle|\, t\in\mathbf{R}, z,w\in\mathbf{C}\right\}$.

V. Semisimple Lie Groups

A connected semisimple Lie group is of type I. The character $\chi=\chi_U$ of an irreducible unitary representation U of G is defined as follows: Let $C_0^\infty(G)$ be the set of all complex-valued C^∞-functions with compact support on G. Then for any function f in $C_0^\infty(G)$, the operator $U_f=\int_G U_g f(g)dg$ belongs to the †trace class, and the linear form $\chi: f\rightarrow T_r U_f$ is a †distribution in the sense of Schwartz. The distribution χ is called the **character** of an irreducible unitary representation U. A character χ is invariant under any inner automorphism of G and is a simultaneous eigendistribution of the algebra of all two-sided invariant linear differential operators on G. Two irreducible unitary representations of G are equivalent if and only if their characters coincide. The distribution χ is a †locally summable function on G and coincides with a real analytic function on each connected component of the dense open submanifold G' consisting of regular elements in G. In general, χ is not real analytic on all of G (Harish-Chandra [9, III; 10].

W. Complex Semisimple Lie Groups

There are four series of irreducible representations of a complex semisimple Lie group G.

(1) A **principal series** consists of unitary representations of G induced from 1-dimensional unitary representations L of a †Borel subgroup B of G. L is uniquely determined by a unitary character $\nu\in\mathrm{Hom}(A,U(1))=A^*$ of the †Cartan subgroup A of G contained in B. Hence the representations in the principal series are parametrized by the elements in the character group A^* of the Cartan subgroup A. If we denote U^L by U^ν, two representations U^ν and $U^{\nu'}(\nu,\nu'\in A^*)$ are equivalent if and only if ν and ν' are conjugate under the †Weyl group W of G with respect to A.

(2) A **degenerate series** consists of unitary representations induced by 1-dimensional unitary representations of a †parabolic subgroup P of G other than B. (A parabolic subgroup P is any subgroup of G containing a Borel subgroup B.)

(3) A **complementary series** consists of irreducible unitary representations U^L induced by nonunitary 1-dimensional representations L of a Borel subgroup B. In this case, condition (ii) in the definition of U^L (→ Section O) must be changed. When L is a nonunitary representation, then the operator U_g^L is not a unitary operator with respect to the usual L_2-inner product (ii). However, if L satisfies a certain condition, then U_g^L leaves invariant some positive definite Hermitian form on the space of sufficiently nice functions. Completing this space, we get a unitary representation U^L. The representations thus obtained form the complementary series.

(4) A **complementary degenerate series** consists of irreducible unitary representations induced by nonunitary 1-dimensional representations of a parabolic subgroup $P\neq B$.

Representations belonging to different series are never equivalent. It seems certain that any irreducible unitary representation of a connected complex semisimple Lie group is equivalent to a representation belonging to one of the above four series, but this conjecture has not yet been proved. Moreover, E. M. Stein [11] constructed irreducible unitary representations different from any in the list obtained by I. M. Gel'fand and M. A. Naĭmark (Neumark) [12]. These representations belong to the complementary degenerate series. The characters of the representations in these four series are computed in explicit form. For example, the character χ_ν of the representation U^ν in the principal series can be calculated as follows: Let λ be a linear form on a Cartan subalgebra $\mathfrak{a}$ such that $\nu(\exp H)=e^{\lambda(H)}$ for every H in $\mathfrak{a}$, let D be the function on A defined by $D(\exp H)=\prod_\alpha|e^{\alpha(H)/2}-e^{-\alpha(H)/2}|^2$, where α runs over all positive roots. Then the character χ_ν of a representation U^ν in the principal series is given by the formula

$$\chi_\nu(\exp H)=D(\exp H)^{-1}\sum_{s\in W}e^{s\lambda(H)}.$$

In the irreducible decomposition of the regular representation of G, only irreducible representations belonging to the principal series arise. Hence the right-hand side in the Plancherel formula is an integral over the character group A^* of a Cartan subgroup A. Under a suitable normalization of the Haar measures in G and A^*, the Plancherel measure μ of G can be expressed by using the Haar measure dv of A^*:

$$d\mu(v)=w^{-1}\prod_{\alpha}|(\lambda,\alpha)/(\rho,\alpha)|^2 dv,$$

where w is the order of the Weyl group, ρ is the half-sum of all †positive roots, and α runs over all positive roots (Gel'fand and Naĭmark [12]).

X. Real Semisimple Lie Groups

As in the case of a complex semisimple Lie group, a connected real semisimple Lie group G has four series of irreducible unitary representations. However, if G has no parabolic subgroup other than a minimal parabolic subgroup B and G itself, then G has no representation in the degenerate or complementary degenerate series. Examples of such groups are $SL(2, \mathbf{R})$ and higher-dimensional †Lorentz groups. In general, the classification of irreducible unitary representations in the real semisimple case is more complicated than in the complex semisimple case. Irreducible unitary representations arising from the irreducible decomposition of the regular representation are called representations in the **principal series**. The principal series of G are divided into a finite number of subseries corresponding bijectively to the conjugate classes of the †Cartan subgroups of G.

A connected semisimple Lie group G has a square integrable representation if and only if G has a compact Cartan subgroup H. The set of all square integrable representations of G is called the **discrete series** of irreducible unitary representations. The discrete series is the subseries in the principal series corresponding to a compact Cartan subgroup H. The representations in the discrete series were classified by Harish-Chandra. Let $\mathfrak{h}$ be the Lie algebra of H, P the set of all positive roots in $\mathfrak{h}$ for a fixed linear order, π the polynomial $\prod_{\alpha\in P}H_\alpha$, and $\mathscr{F}$ the set of all real-valued linear forms on $\sqrt{-1}\mathfrak{h}$. Moreover, let L be the set of all linear forms λ in $\mathscr{F}$ such that a single-valued character ξ_λ of the group H is defined by the formula $\xi_\lambda(\exp X)=e^{\lambda(X)}$, and let L' be the set of all λ in L such that $\pi(\lambda)\neq 0$. Then for each λ in L', there exists a representation $\omega(\lambda)$ of G in the discrete series, and conversely, every representation in the discrete series is equivalent to $\omega(\lambda)$ for some λ in L'. Two representations $\omega(\lambda_1)$ and $\omega(\lambda_2)$ $(\lambda_1,\lambda_2\in L')$ are equivalent if and only if there exists an element s in $W_G=N(H)/H$ such that $\lambda_2=s\lambda_1$, where $N(H)$ is the normalizer of H in G (W_G can act on $\mathscr{F}$ as a linear transformation group in the natural way). The value of the character χ_λ on the subgroup H of the representation $\omega(\lambda)$ $(\lambda\in L')$ is given as follows: Let $\varepsilon(\lambda)$ be the signature of $\pi(\lambda)=\prod_{\alpha\in P}\lambda(H_\alpha)$, and define q and Δ by $q=(\dim G/K)/2$ and $\Delta(\exp H)=\prod_{\alpha\in P}(e^{\alpha(H)/2}-e^{-\alpha(H)/2})$. Then the character χ_λ of the representation $\omega(\lambda)$ has the value $(-1)^q\varepsilon(\lambda)\chi_\lambda(h)=\Delta(h)^{-1}\sum_{s\in W_G}(\det s)\xi_{s\lambda}(h)$ on a regular element h in H. The formal degree $d(\omega(\lambda))$ of the representation $\omega(\lambda)$ is given by the formula $d(\omega(\lambda))=C^{-1}[W_G]|\pi(\lambda)|$, where C is a positive constant (not depending on λ) and $[W_G]$ is the order of the finite group W_G (Harish-Chandra [13]). A formula expressing the character χ_λ on the whole set of regular elements in G has been given by T. Hirai [14]. The representations in discrete series are realized on L^2-cohomology spaces of homogeneous holomorphic line bundles over G/H (W. Schmid [15]). They are also realized on the spaces of harmonic spinors on the †Riemannian symmetric space G/K (M. Atiyah and Schmid [16]). They are also realized on the eigenspaces of a Casimir operator acting on the sections of vector bundles on G/K (R. Hotta, *J. Math. Soc. Japan*, 23; N. Wallach [17]). An irreducible unitary representation is called **integrable** if at least one of its matrix coefficients belongs to $L^1(G)$. Integrable representations belong to the discrete series. They have been characterized by H. Hecht and Schmid (*Math. Ann.*, 220 (1976)). The theory of the discrete series is easily extended to reductive Lie groups.

The general principal series representations of a connected semisimple Lie group G with finite center are constructed as follows. Let K be a maximal compact subgroup of G. Then there exists a unique involutive automorphism θ of G whose fixed point set coincides with K. θ is called a **Cartan involution** of G. Let H be a θ-stable Cartan subgroup of G. Then H is the direct product of a compact group $T=H\cap K$ and a vector group A. The centralizer $Z(A)$ of A in G is the direct product of a reductive Lie group $M=\theta(M)$ and A. M has a compact Cartan subgroup T. Hence the set $\hat{M}_d$ of the discrete series representations of M is not empty. Let α be an element of the dual space $\mathfrak{a}^*$ of the Lie algebra $\mathfrak{a}$ of A and put $\mathfrak{g}_\alpha=\{X\in\mathfrak{g}\,|\,[H,X]=\alpha(H)X(\forall H\in\mathfrak{a})\}$ and $\Delta=\{\alpha\in\mathfrak{a}^*\,|\,\mathfrak{g}_\alpha\neq\{0\}\}$. Let Δ^+ be the set of positive elements of Δ in a certain order of $\mathfrak{a}^*$ and put $\mathfrak{n}=\sum_{\alpha\in\Delta^+}\mathfrak{g}_\alpha$ and $N=\exp\mathfrak{n}$. Then $P=MAN$ is a closed subgroup of G. P is called a **cuspidal parabolic subgroup** of G. Let $D\in\hat{M}_d$ and $v\in\mathfrak{a}^*$. Then a unitary representation $D\otimes e^{iv}$ of P

is defined by $(D \otimes e^{i\nu})(man) = D(m)e^{i\nu(\log a)}$ $(m \in M, a \in A, n \in N)$. The unitary representation $\pi_{D,\nu}$ of G induced by $D \otimes e^{i\nu}$ is independent of the choice of Δ^+ up to equivalence. Thus $\pi_{D,\nu}$ depends only on (H, D, ν). The set of representations $\{\pi_{D,\nu} | D \in \hat{M}_d, \nu \in \mathfrak{a}^*\}$ is called the **principal** H**-series**. If ν is regular in $\mathfrak{a}^*$ (i.e., $(\nu, \alpha) \neq 0$ for all $\alpha \in \Delta$), then $\pi_{D,\nu}$ is irreducible. Every $\pi_{D,\nu}$ is a finite sum of irreducible representations. The character $\theta_{D,\nu}$ of $\pi_{D,\nu}$ is a locally summable function which is supported in the closure of $\bigcup_{g \in G} g(MA)g^{-1}$. If two Cartan subgroups H_1 and H_2 are not conjugate in G, then every H_1-series representation is disjoint from every H_2-series representation. Choose a complete system $\{H_1, \ldots, H_r\}$ of conjugacy classes of Cartan subgroups of G. Then every H_i can be chosen as θ-stable. The union of the principal H_i-series $(1 \leqslant i \leqslant r)$ is the principal series of G. The right (or left) regular representation of G is decomposed as the direct integral of the principal series representations. Every complex-valued C^∞-function on G with compact support has an expansion in terms of the matrix coefficients of the principal series representations. Harish-Chandra [18] proved these theorems and determined explicitly the Plancherel measure by studying the asymptotic behavior of the Eisenstein integral [19, 20].

Y. Spherical Functions

Let G be a locally compact †unimodular group and K a compact subgroup of G. The set of all complex-valued continuous functions on G that are invariant under every left translation L_k by elements k in K is denoted by $C(K \backslash G)$. The subset of $C(K \backslash G)$ that consists of all two-sided K-invariant functions is denoted by $C(G, K)$. The subset of $C(G, K)$ consisting of all functions with compact support is denoted by $\mathbf{L} = L(G, K)$. $\mathbf{L}$ is an algebra over $\mathbf{C}$ if the product of two elements f and g in $\mathbf{L}$ is defined by the convolution.

Let λ be an algebra homomorphism from $\mathbf{L}$ into $\mathbf{C}$. Then an element of the eigenspace $F(\lambda) = \{\psi \in C(K, G) | f * \psi = \lambda(f)\psi \ (\forall f \in \mathbf{L})\}$ is called a **spherical function** on $K \backslash G$. If $F(\lambda)$ contains a nonzero element, then $F(\lambda)$ contains a unique two-sided K-invariant element ω normalized by $\omega(e) = 1$, where e is the identity element in G. This function ω is called the **zonal spherical function** associated with λ. In this case, the homomorphism λ is defined by $\lambda(f) = \int_G f(g)\omega(g^{-1})dg$. Hence the eigenspace $F(\lambda)$ is uniquely determined by the zonal spherical function ω. A function $\omega \neq 0$ in $C(G, K)$ is a zonal spherical function on $K \backslash G$ if and only if ω satisfies either of the following two conditions: (i) The mapping $f \mapsto \int f(g)\omega(g^{-1})dg$ is an algebra homomorphism of $\mathbf{L}$ into $\mathbf{C}$; (ii) ω satisfies the functional equation

$$\int_K \omega(gkh)\,dk = \omega(g)\omega(h).$$

When G is a Lie group, every spherical function is a real analytic function on $K \backslash G$.

Z. Expansion by Spherical Functions

In this section, we assume that the algebra $\mathbf{L}$ of two-sided K-invariant functions is commutative. In this case there are sufficiently many spherical functions of $K \backslash G$, and two-sided K-invariant functions are expanded by spherical functions. An irreducible unitary representation U of G is called a **spherical representation** with respect to K if the representation space $\mathfrak{H}(U)$ contains a nonzero vector invariant under every U_k, where k runs over K. By the commutativity of $\mathbf{L}$, the K-invariant vectors in $\mathfrak{H}(U)$ form a 1-dimensional subspace. Let x be a K-invariant vector in $\mathfrak{H}(U)$ with the norm $\|x\| = 1$. Then $\omega(g) = (U_g x, x)$ is a zonal spherical function on $K \backslash G$, and for every y in $\mathfrak{H}(U)$, the function $\varphi_y(g) = (U_g x, y)$ is a spherical function associated with ω. Moreover, in this case the zonal spherical function ω is a positive definite function on G. Conversely, every positive definite zonal spherical function ω can be expressed as $\omega(g) = (U_g x, x)$ for some spherical representation U and some K-invariant vector x in $\mathfrak{H}(U)$.

The set of all positive definite zonal spherical functions becomes a locally compact space Ω by the topology of compact convergence. The **spherical Fourier transform** $\hat{f}$ of a function f in $L_1(K \backslash G)$ is defined by

$$\hat{f}(\omega) = \int_G f(g)\omega(g^{-1})\,dg.$$

There exists a unique †Radon measure μ on Ω such that for every f in $\mathbf{L}$, $\hat{f}$ belongs to $L_2(\Omega, \mu)$. Also, the Plancherel formula

$$\int_G f(s)\overline{g(s)}\,ds = \int_\Omega \hat{f}(\omega)\overline{\hat{g}(\omega)}\,d\mu(\omega) \qquad (4)$$

holds for every f and g in $\mathbf{L}$, and an inversion formula $f(s) = \int_\Omega \hat{f}(\omega)\omega(s)\,d\mu(\omega)$ holds for a sufficiently nice two-sided K-invariant function f [21]. Identifying a positive definite zonal spherical function with the corresponding spherical representation, we can regard Ω as a subset of the dual $\hat{G}$ of G. The Plancherel formula for two-sided K-invariant functions is obtained from the general Plancherel formula on G by restricting the domain of the integral from $\hat{G}$ to Ω. When G is a Lie group and $\mathbf{L}$ is commutative, a spherical function on $K \backslash G$ can be characterized as a simultaneous eigenfunc-

tion of G-invariant linear differential operators on $K\backslash G$.

AA. Spherical Function on Symmetric Spaces

The most important case where the algebra $\mathbf{L}=L(G,K)$ is commutative is when $K\backslash G$ is a †weakly symmetric Riemannian space or, in particular, a †symmetric Riemannian space When $K\backslash G$ is a compact symmetric Riemannian space, a spherical representation with respect to K is the irreducible component of the regular representation T on $K\backslash G$, and a spherical function on $K\backslash G$ is a function that belongs to the irreducible subspaces in $L_2(K\backslash G)$. In particular, if G is a compact connected semisimple Lie group, the highest weights of spherical representations of G with respect to K are explicitly given by using the **Satake diagram** of $K\backslash G$. The Satake diagram of $K\backslash G$ is the †Satake diagram of the noncompact symmetric Riemannian space $K\backslash G_0$ dual to $K\backslash G$ or the Satake diagram of the Lie algebra of G_0. If a symmetric space is the underlying manifold of a compact Lie group G, then G can be expressed as $G=K\backslash(G\times G)$, where K is the diagonal subgroup of $G\times G$. In this case, a zonal spherical function ω on $G=K\backslash(G\times G)$ is the normalized character of an irreducible unitary representation U of $G:\omega(g)=(\deg U)^{-1}T_rU_g$. The explicit form of ω is given by †Weyl's character formula (→ 249 Lie Groups).

The zonal spherical functions on a symmetric Riemannian space $K\backslash G$ of noncompact type are obtained in the following way: Let G be a connected semisimple Lie group with finite center, K be a maximal compact subgroup of G, and $G=NA_+K$ be an †Iwasawa decomposition. Then for any g in G there exists a unique element $H(g)$ in the Lie algebra $\mathfrak{a}_+$ of A_+ such that g belongs to $N\exp H(g)K$. Let $\mathfrak{a}$ be a Cartan subalgebra containing $\mathfrak{a}_+$, P be the set of all positive roots in $\mathfrak{a}$, and $\rho=(\sum_{\alpha\in P}\alpha)/2$. Then for any complex-valued linear form ν on $\mathfrak{a}_+$, the function

$$\omega_\nu(g)=\int_K e^{(i\nu-\rho)(H(kg))}dk$$

is a zonal spherical function on the symmetric Riemannian space $K\backslash G$. Conversely, every zonal spherical function ω on $K\backslash G$ is equal to ω_ν for some ν. Two zonal spherical functions ω_ν and $\omega_{\nu'}$ coincide if and only if ν and ν' are conjugate under the operation of the Weyl group $W_0=N_K(A)/Z_K(A)$ of $K\backslash G$ (Harish-Chandra [22], S. Helgason [23]). If ν is real-valued, then ω_ν is positive definite. Such a zonal spherical function ω_ν is obtained from a spherical representation belonging to the principal A-series. Let Ω_0 be the set of all zonal spherical functions ω_ν associated with the real-valued linear form ν. Then the support of the Plancherel measure μ on $K\backslash G$ is contained in Ω_0. We can choose ν as a parameter on the space Ω_0. Then the right-hand side of the Plancherel formula can be expressed as an integral over the dual space L of $\mathfrak{a}_+$. Moreover, the Plancherel measure μ is absolutely continuous with respect to the Lebesgue measure $d\nu$ on the Euclidean space L and can be expressed as

$$d\mu(\omega_\nu)=\omega_0^{-1}|c(\nu)|^{-2}d\nu$$

under suitable normalization of μ and $d\nu$. The problem of calculating the function $c(\nu)$ can be reduced to the case of symmetric spaces of rank 1 and can be solved explicitly. Let p_α be the multiplicity of a restricted root α and $I(\nu)$ be the product

$$I(\nu)=\prod_\alpha B(\tfrac{1}{2}p_\alpha,\tfrac{1}{4}p_\alpha+(\nu,\alpha)(\alpha,\alpha)^{-1}),$$

where α runs over all positive restricted roots and B is the †beta function. Then $c(\nu)=I(i\nu)/I(\rho)$ [20, 24]. Every spherical function f on $K\backslash G$ is expressed as the Poisson integral of its "boundary values" on the Martin boundary $P\backslash G$ of $K\backslash G$, where $P=MA_+N$ is a minimal parabolic subgroup of G. The boundary values of f form a hyperfunction with values in a line bundle over $P\backslash G$ (K. Okamoto et al. [25]).

BB. Spherical Functions and Special Functions

Some important special functions are obtained as the zonal spherical functions on a certain symmetric Riemannian space $M=K\backslash G$ (G is the motion group of M). In particular when M is of rank 1, then the zonal spherical functions are essentially the functions of a single variable. For example, the zonal spherical functions on an n-dimensional Euclidean space can be expressed as

$$\omega_\nu(r)=2^m\Gamma(m+1)(\nu r)^{-m}J_m(\nu r),$$

where $2m=n-2$ and J_m is the †Bessel function of the mth order. The zonal spherical function on an $(n-1)$-dimensional sphere $S^{n-1}=SO(n-1)\backslash SO(n)$ is given by

$$\omega_\nu(\theta)=\Gamma(\nu+1)\Gamma(n-2)\Gamma(\nu+n-2)^{-1}C_\nu^m(\cos\theta) \qquad (\nu=0,1,2,\ldots),$$

where $C_\nu^m(z)$ is the †Gegenbauer polynomial. The zonal spherical functions on an $(n-1)$-dimensional Lobachevskiĭ space can be expressed as

$$\omega_\nu(t)=2^{m-1/2}\Gamma(m+1/2)\sinh^{-m+1/2}t \times \mathfrak{P}^{1/2-m}_{-1/2-m+\nu}(\cosh t)$$

using a generalized †associated Legendre function $\mathfrak{P}_\nu^\mu$. Many properties of special functions can be proved from a group-theoretic point of view. For example, the addition theorem is merely the homomorphism property $U_{gh}=U_gU_h$ expressed in terms of the matrix components of U. The differential equation satisfied by these special functions is derived from the fact that a zonal spherical function ω is an eigenfunction of an invariant differential operator. The integral expression of such a special function can be obtained by constructing a spherical representation U in a certain function space and calculating explicitly the inner product in the expression $\omega(g)=(U_gx,x)$ (N. Ya. Vilenkin [26]).

CC. Generalization of the Theory of Spherical Functions

The theory of spherical functions described in Sections Y–BB can be generalized in several ways. First, spherical functions are related to the trivial representation of K. A generalization is obtained if the trivial representation of K is replaced by an irreducible representation of K. The theory of such zonal spherical functions is useful for representation theory [20]. For example, the Plancherel formula for $SL(2,\mathbf{R})$ can be obtained using such spherical functions (R. Takahashi, *Japan. J. Math.*, 31 (1961)). Harish-Chandra's Eisenstein integral is such a spherical function on a general semisimple Lie group G. He used it successfully to obtain the Plancherel measure of G. Another generalization can be obtained by removing the condition that K is compact. In particular, when $K\backslash G$ is a symmetric homogeneous space of a Lie group G, the algebra $\mathscr{D}$ of all G-invariant linear differential operators is commutative if the space $K\backslash G$ has an invariant volume element. In this case, a spherical function on $K\backslash G$ can be defined as a simultaneous eigenfunction of $\mathscr{D}$. The character of a semisimple Lie group is a zonal spherical function (distribution) in this sense. The spherical functions and harmonic analysis on symmetric homogeneous space have been studied by T. Oshima and others. T. Oshima and J. Sekiguchi [27] proved the Poisson integral theorem (→ Section AA) for a certain kind of symmetric homogeneous spaces.

The spherical functions and unitary representations of topological groups that are not locally compact are studied in connection with probability theory and physics. For example, the zonal spherical functions of the rotation group of a real Hilbert space are expressed by Hermite polynomials.

DD. Discontinuous Subgroups and Representations

Let G be a connected semisimple Lie group and Γ be a discrete subgroup of G. Then the regular representation T of G on $\Gamma\backslash G$ is defined by $(T_gf)(x)=f(xg)$ $(f\in L^2(\Gamma\backslash G))$. The problem of decomposing the representation T into irreducible components is important in connection with the theory of †automorphic forms and number theory. First assume that the quotient space $\Gamma\backslash G$ is compact. Then for every function f in $L_1(G)$, the operator $T(f)$ is a compact operator. Hence the regular representation T on $\Gamma\backslash G$ can be decomposed into the discrete sum $T=\sum_{k=1}^\infty T^{(k)}$ of irreducible unitary representations $T^{(k)}$, and the multiplicity of every irreducible component is finite. The irreducible unitary representation U of G is related to the automorphic forms of Γ in the following way: Let x be a nonzero element in the representation space $\mathfrak{H}=\mathfrak{H}(U)$ of U. $\mathfrak{H}$ is topologized into a †locally convex topological vector space $\mathfrak{H}_x$ by the set N_x of †seminorms: $N_x=\{P_C(y)=\max_{g\in C}|(U_gx,y)|\}$, where C runs over all compact subsets in G. The topology $\mathscr{T}_x$ of $\mathfrak{H}_x$ is independent of the choice of x provided that $\dim\{T_kx\,|\,k\in K\}<\infty$, where K is a maximal compact subgroup of G. Let $\mathfrak{H}^*$ be the completion of $\mathfrak{H}_x$ with respect to the topology $\mathscr{T}_x$ (the completion is independent of the choice of x). $\mathfrak{H}^*$ contains the original Hilbert space $\mathfrak{H}$ as a subspace. Then the representation U of G on $\mathfrak{H}$ can be extended to a representation U^* of G on the space $\mathfrak{H}^*$. An element f in $\mathfrak{H}^*$ invariant under U_γ^* for every γ in Γ is called an **automorphic form** of Γ of type U. Then the multiplicity of an irreducible representation U in the regular representation T on $\Gamma\backslash G$ is equal to the dimension of the vector space consisting of all automorphic forms of type U. This theorem is called the **Gel'fand–Pyatetskiĭ-Shapiro reciprocity law** [28]. Let $T=\sum_{k=1}^\infty T^{(k)}$ be the irreducible decomposition of T and χ_k be the character of the irreducible unitary representation $T^{(k)}$. Then for a suitable function f on G, the integral operator K_f on $\mathfrak{H}(T)=L^2(T\backslash G)$ with kernel $k_f(x,y)=\sum_{\gamma\in\Gamma}f(x^{-1}\gamma y)$ belongs to the trace class. By calculating the trace of K_f in two ways, the following **trace formula** is obtained:

$$\sum_{k=1}^\infty\int_G f(g)\chi_k(g)\,dg=\sum_{\{\gamma\}}\int_{\mathfrak{D}_\gamma}f(x^{-1}\gamma x)\,dx,$$

where $\{\gamma\}$ is the conjugate class of γ in Γ and $\mathfrak{D}_\gamma$ is the quotient space of the centralizer G_γ of γ in G by the centralizer Γ_γ of γ in Γ.

When the groups G and Γ are given explicitly, the right-hand side of the trace formula can be expressed in a more explicit form,

and the trace formula leads to useful consequences. A similar trace formula holds for the unitary representation U^L induced by a finite-dimensional unitary representation L of Γ instead of the regular representation T on $\Gamma\backslash G$. When the quotient space $\Gamma\backslash G$ is not compact, the irreducible decomposition of the regular representation T on $\Gamma\backslash G$ contains not only the discrete direct sum but also the direct integral (continuous spectrum). A. Selberg showed that even in this case, there are explicit examples for which the trace formula holds for the part with discrete spectrum. Also, the part with continuous spectrum can be described by the †generalized Eisenstein series. Analytic properties and the functional equation of the generalized Eisenstein series have been studied by R. Langlands [30]. Recent developments are surveyed in [31].

EE. History

Finite-dimensional unitary representations of a finite group were studied by Frobenius and Schur (1896–1905). In 1925, †Weyl studied the finite-dimensional unitary representation of compact Lie groups. The theory of infinite-dimensional unitary representation was initiated in 1939 by E. P. Wigner in his work on the inhomogeneous Lorentz group, motivated by problems of quantum mechanics.

In 1943, Gel'fand and D. A. Raikov proved the existence of sufficiently many irreducible unitary representations for an arbitrary locally compact group. The first systematic studies of unitary representations appeared in 1947 in the work of V. Bargmann on $SL(2,\mathbf{R})$ [31] and the work of Gel'fand and Neumark on $SL(2,\mathbf{C})$. Gel'fand and Naĭmark established the theory of unitary representation for complex semisimple Lie groups [12].

Harish-Chandra proved theorems concerning the unitary representations of a general semisimple Lie group; for instance, he proved that a semisimple Lie group G is of type I [7] and defined the character of a unitary representation of G and proved its basic properties [9, III; 10]. Harish-Chandra also determined the discrete series of G and their characters. Harish-Chandra [18] proved the Plancherel formula for an arbitrary connected semisimple Lie group G with finite center. Hence harmonic analysis of square integrable functions on G is established.

Further studies on harmonic analysis on semisimple Lie groups have been carried out. In particular, Paley-Wiener–type theorems, which determine the Fourier transform image of the space $C_C^\infty(G)$ of C^∞-functions with compact support, have been proved for the group $PSL(2,\mathbf{R})$ (L. Ehrenpreis and F. Mautner [33]), complex semisimple Lie groups (Zhelobenko [34]), and two-sided K-invariant functions on general semisimple Lie groups (R. Gangolli [35]). A. W. Knapp and E. M. Stein [36] studied the intertwining operators.

Concerning the construction of irreducible representations, G. W. Mackey [3] and Bruhat [4] developed the theory of induced representations of locally compact groups and Lie groups, respectively. B. Kostant [37] (see Blattner's article in [38]) noticed a relation between homogeneous †symplectic manifolds and unitary representations and proposed a method of constructing irreducible unitary representations of a Lie group. Selberg's research [29] revealed a connection between unitary representations (or spherical functions) and the theory of automorphic forms and number theory. A number of papers along these lines have since appeared [31]. In connection with number-theoretic investigations of an †algebraic group defined over an algebraic number field, unitary representations of the †adele group of G or an algebraic group over a †p-adic number field have been studied (→ [31, 38], Gel'fand, M. I. Grayev, and I. I. Pyatetskiĭ-Shapiro [39], and H. M. Jacquet and R. P. Langlands [40]).

For the algebraic approach to the infinite-dimensional representations of semisimple Lie groups and Lie algebras → [41].

For surveys of the theory of unitary representations → [2, 19, 20, 31, 38].

References

[1] N. Tatsuuma, A duality theorem for locally compact groups, J. Math. Kyoto Univ., 6 (1967), 187–293.
[2] J. Dixmier, Les C^*-algèbres et leurs représentations, Gauthier-Villars, 1964.
[3] G. W. Mackey, Induced representations of locally compact groups I, Ann. Math., 55 (1952), 101–139.
[4] F. Bruhat, Sur les representations induites des groupes de Lie, Bull. Soc. Math. France, 84 (1956), 97–205.
[5] R. Bott, Homogeneous vector bundles, Ann. Math., 66 (1957), 203–248.
[6] E. Nelson, Analytic vectors, Ann. Math., 70 (1959), 572–615.
[7] A. A. Kirillov, Unitary representations of nilpotent Lie groups, Russian Math. Surveys, 17, no. 4 (1962), 53–104. (Original in Russian, 1962.)
[8] L. Auslander and B. Kostant, Polarization

and unitary representations of solvable Lie groups, Inventiones Math., 14 (1971), 255–354.
[9] Harish-Chandra, Representations of a semisimple Lie group on a Banach space I–VI, Trans. Amer. Math. Soc., 75 (1953), 185–243; 76 (1954), 26–65; 76 (1954), 234–253; Amer. J. Math., 77 (1955), 743–777; 78 (1956), 1–41; 78 (1956), 564–628.
[10] Harish-Chandra, Invariant eigendistributions on a semisimple Lie group, Trans. Amer. Math. Soc., 119 (1965), 457–508.
[11] E. M. Stein, Analysis in matrix spaces and some new representations of $SL(N, C)$, Ann. Math., (2) 86 (1967), 461–490.
[12] I. M. Gel'fand and M. A. Naĭmark (Neumark), Unitäre Darstellungen der klassischen Gruppen, Akademische Verlag., 1957. (Original in Russian, 1950.)
[13] Harish-Chandra, Discrete series for semisimple Lie groups I, II, Acta Math., 113 (1965), 241–318; 116 (1966), 1–111.
[14] T. Hirai, The characters of the discrete series for semisimple Lie groups, J. Math. Kyoto Univ., 21 (1981), 417–500.
[15] W. Schmid, L^2-cohomology and the discrete series, Ann. Math., 102 (1976), 375–394.
[16] M. F. Atiyah and W. Schmid, A geometric construction of the discrete series for semisimple Lie groups, Inventiones Math., 42 (1977), 1–62.
[17] N. Wallach, On the Enright-Varadarajan modules: A construction of the discrete series, Ann. Sci. Ecole Norm. Sup., 9 (1976), 81–102.
[18] Harish-Chandra, Harmonic analysis on real reductive groups I–III, J. Functional Anal., 19 (1975), 104–204; Inventiones Math., 36 (1976), 1–55; Ann. Math., 104 (1976), 117–201.
[19] J. A. Wolf, M. Cahen, and M. De Wilde (eds.), Harmonic analysis and representations of semisimple Lie groups, Reidel, 1980.
[20] G. Warner, Harmonic analysis on semisimple Lie groups I, II, Springer, 1972.
[21] H. Yoshizawa, A proof of the Plancherel theorem, Proc. Japan Acad., 30 (1954), 276–281.
[22] Harish-Chandra, Spherical functions on a semisimple Lie group I, II, Amer. J. Math., 80 (1958), 241–310, 553–613.
[23] S. Helgason, Differential geometry and symmetric spaces, Academic Press, 1962.
[24] J. Rosenberg, A quick proof of Harish-Chandra's Plancherel theorem for spherical functions on a semisimple Lie groups, Proc. Amer. Math. Soc., 63 (1977), 143–149.
[25] M. Kashiwara, A. Kowata, K. Minemura, K. Okamoto, T. Oshima, and M. Tanaka, Eigenfunctions of invariant differential operators on a symmetric space, Ann. Math., 107 (1978), 1–39.
[26] N. Ya Vilenkin, Special functions and theory of group representations, Amer. Math. Soc. Transl. of Math. Monographs, 22, 1968. (Original in Russian, 1965.)
[27] T. Oshima and J. Sekiguchi, Eigenspaces of invariant differential operators on an affine symmetric space, Inventiones Math., 57 (1980), 1–81.
[28] I. M. Gel'fand and I. I. Pyatetskiĭ-Shapiro, Theory of representations and theory of automorphic functions, Amer. Math. Soc. Transl., (2) 26 (1963), 173–200. (Original in Russian, 1959.)
[29] A. Selberg, Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series, J. Indian Math. Soc., 20 (1956), 47–87.
[30] R. P. Langlands, On the functional equations satisfied by Eisenstein series, Lecture notes in math. 544, Springer, 1976.
[31] Automorphic forms, representations and L-functions, Amer. Math. Soc. Proc. Symposia in Pure Math., 33, 1979.
[32] V. Bargmann, Irreducible unitary representations of the Lorentz group, Ann. Math., (2) 48 (1947), 568–640.
[33] L. Ehrenpreis and F. I. Mautner, Some properties of the Fourier transform on semisimple Lie groups I–III, Ann. Math., (2) 61 (1955), 406–439; Trans. Amer. Math. Soc., 84 (1957), 1–55; 90 (1959), 431–484.
[34] D. P. Zhelobenko, Harmonic analysis of functions on semisimple Lie groups I, II, Amer. Math. Soc. Transl., (2) 54 (1966), 177–230. (Original in Russian, 1963, 1969.)
[35] R. Gangolli, On the Plancherel formula and the Paley-Wiener theorem for spherical functions on a semisimple Lie group, Ann. Math., 93 (1971), 150–165.
[36] A. W. Knapp and E. M. Stein, Intertwining operators for semisimple groups, Ann. Math., 93 (1971), 489–578.
[37] B. Kostant, Quantization and unitary representations, Lecture notes in math. 170, Springer, 1970, 87–207.
[38] Harmonic analysis on homogeneous spaces, Amer. Math. Soc. Proc. Symposia in Pure Math., 26 (1973).
[39] I. M. Gel'fand, M. I. Grayev, and I. I. Pyatetskiĭ-Shapiro, Representation theory and automorphic functions, Saunders, 1969. (Original in Russian, 1966.)
[40] H. Jacquet and R. P. Langlands, Automorphic forms on $GL(2)$, Lecture notes in math. 114, Springer, 1970.
[41] D. A. Vogan, Representations of real reductive Lie groups, Birkhäuser, 1981.

438 (XI.5) Univalent and Multivalent Functions

A. General Remarks

A single-valued †analytic function $f(z)$ defined in a domain D of the complex plane is said to be **univalent** (or **simple** or **schlicht**) if it is injective, i.e., if $f(z_1) \neq f(z_2)$ for all distinct points z_1, z_2 in D. A multiple-valued function $f(z)$ is also said to be univalent if its distinct function elements always attain distinct values at their centers. The derivative of a univalent function is never zero. The limit function of a †uniformly convergent sequence of univalent functions is univalent unless it reduces to a constant. When $f(z)$ is single-valued, the univalent function $w = f(z)$ gives rise to a one-to-one †conformal mapping between D and its image $f(D)$.

B. Univalent Functions in the Unit Disk

A systematic theory of the family of functions †holomorphic and univalent in the unit disk originates from a **distortion theorem** obtained by P. Koebe (1909) in connection with the uniformization of analytic functions. In general, distortion theorems are theorems for determining bounds of functionals, such as $|f(z)|$, $|f'(z)|$, $\arg f'(z)$, within the family under consideration. In particular, distortion theorems concerning the bounds of the arguments of $f(z)$ and $f'(z)$ are also called **rotation theorems**. Though results were at first qualitative, they were made quantitative subsequently by L. Bieberbach (1916), G. Faber (1916), and others. Any univalent function $f(z)$ holomorphic in the unit disk and normalized by $f(0)=0$ and $f'(0)=1$ satisfies the **distortion inequalities**

$$\frac{|z|}{(1+|z|)^2} \leqslant |f(z)| \leqslant \frac{|z|}{(1-|z|)^2},$$

$$\frac{1-|z|}{(1+|z|)^3} \leqslant |f'(z)| \leqslant \frac{1+|z|}{(1-|z|)^3}.$$

Here the equality holds only if $f(z)$ is of the form $z/(1-\varepsilon z)^2$ ($|\varepsilon|=1$). In deriving these inequalities, Bieberbach centered his attention on the family of †meromorphic functions $g(\zeta) = \zeta + \sum_{\nu=0}^{\infty} b_\nu \zeta^{-\nu}$ univalent in $|\zeta|>1$. He established the **area theorem** $\sum_{\nu=1}^{\infty} \nu |b_\nu|^2 \leqslant 1$, which illustrates the fact that the area of the complementary set of the image domain is nonnegative. Bieberbach, R. Nevanlinna (1919–1920), and others constructed a systematic theory of univalent functions in the unit disk based on this theorem.

After the area theorem, the chief tools in the theory of univalent functions have been Löwner's method, the method of contour integration, the variational method, and the method of the extremal metric. In contrast to the theory of univalent functions based on Bieberbach's area theorem, K. Löwner (1923) introduced a new method. In view of a theorem on the domain kernel (C. Carathéodory, 1912), it suffices to consider an everywhere dense subfamily in order to estimate a continuous functional within the family of univalent functions holomorphic in the unit disk. Löwner used the subfamily of functions mapping the unit disk onto the so-called bounded slit domains. Namely, the range of a member of this subfamily consists of the unit disk slit along a Jordan arc that starts at a periphery point and does not pass through the origin. A mapping function of this nature is determined as the integral $f(z,t_0)$ of **Löwner's differential equation**

$$\frac{\partial f(z,t)}{\partial t} = -f(z,t)\frac{1+\kappa(t)f(z,t)}{1-\kappa(t)f(z,t)}, \qquad 0 \leqslant t \leqslant t_0,$$

with the initial condition $f(z,0)=z$, where $\kappa(t)$ is a continuous function with absolute value equal to 1. Any univalent function $f(z)$ holomorphic in the unit disk and satisfying $f(0)=0$, $f'(0)=1$ has an arbitrarily close approximation by functions of the form $e^{t_0} f(z,t_0)$. By means of this differential equation Löwner proved that $|a_3| \leqslant 3$ for any univalent function $f(z) = z + \sum_{n=2}^{\infty} a_n z^n$ ($|z|<1$) and also derived a decisive estimate concerning a coefficient problem for the inverse function [2].

G. M. Golusin (1935) and I. E. Bazilevich (1936) first noticed that Löwner's method is also a powerful tool for deriving several distortion theorems. They showed that classical distortion theorems can be derived in more detailed form (Golusin, *Mat. Sb.*, 2 (1937), 685); in particular, Golusin (1938) obtained a precise estimate concerning the rotation theorem, i.e.,

$$|\arg f'(z)| \leqslant \begin{cases} 4\arcsin|z|, & |z|<1/\sqrt{2}, \\ \pi + \log(|z|^2/(1-|z|^2)), & 1/\sqrt{2} \leqslant |z| < 1. \end{cases}$$

Löwner's method was also investigated by A. C. Schaeffer and D. C. Spencer (1945) [8].

The method of contour integration was introduced by H. Grunsky. It starts with some 2-dimensional integral which can be shown to be positive. Transforming it into a boundary integral and using the †residue theorem, we obtain an appropriate inequality by means of this integral. By this method Grunsky established the following useful inequality (*Math.*

Z., 45 (1939)). For $g(\zeta)=\zeta+\sum_{\nu=0}^{\infty} b_\nu \zeta^{-\nu}$, which is univalent in $|\zeta|>1$, let

$$\log\frac{g(z)-g(\zeta)}{z-\zeta}=\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} c_{mn} z^{-m}\zeta^{-n} \qquad (|z|>1, |\zeta|>1).$$

The coefficients c_{mn} are polynomials in the coefficients b_ν of g. Then **Grunsky's inequality** is: For each integer N and for all complex numbers $\lambda_1, \ldots, \lambda_N$,

$$\left|\sum_{m=1}^{N}\sum_{n=1}^{N} c_{mn}\lambda_m\lambda_n\right| \leqslant \sum_{n=1}^{N}\frac{1}{n}|\lambda_n|^2.$$

It is known that if this inequality holds for an arbitrary integer N and for all complex numbers $\lambda_1, \ldots, \lambda_N$, then $g(\zeta)$ is univalent in $|\zeta|>1$. There are several generalizations of Grunsky's inequality [13].

The variational method was first developed by M. Schiffer for application to the theory of univalent functions. He first used boundary variations (*Proc. London Math. Soc.*, 44 (1938)) and later interior variations (*Amer. J. Math.*, 65 (1943)). The problem of maximizing a given real-valued functional on a family of univalent functions is called an extremal problem, and a function for which the functional attains its maximum is called an extremal function. The **variational method** is used to uncover characteristic properties of an extremal function by comparing it with nearby functions. Typical results are the qualitative information that the extremal function maps the disk $|z|<1$ onto the complement of a system of analytic arcs satisfying a differential equation and that the extremal function satisfies a differential equation. Following Schiffer, Schaeffer and Spencer [8] and Golusin (*Math. Sb.*, 19 (1946)) gave variants of the method of interior variations.

H. Grötzsch (1928–1934) treated the theory of univalent functions in a unified manner by the method of the †extremal metric. The idea of this method is to estimate the length of curves and the area of some region swept out by them together with an application of †Schwarz's inequality (→ 143 Extremal Length). After Grötzsch, the method of the extremal metric has been used by many authors. In particular, O. Teichmüller, in connection with this method, formulated the principle that the solution of a certain type of extremal problem is in general associated with a †quadratic differential, although he did not prove any general result realizing this principle in concrete form. J. A. Jenkins gave a concrete expression of the Teichmüller principle; namely, he established the general coefficient theorem and showed that this theorem contains as special cases a great many of the known results on univalent functions [11].

Univalence criteria have been given by various authors. In particular, Z. Nehari (*Bull. Amer. Math. Soc.*, 55 (1949)) proved that if $|\{f(z),z\}|\leqslant 2(1-|z|^2)^{-2}$ in $|z|<1$, then $f(z)$ is univalent in $|z|<1$, and E. Hille (*Bull. Amer. Math. Soc.*, 55 (1949)) proved that 2 is the best possible constant in the above inequality. Here, $\{f(z),z\}$ denotes the †Schwarzian derivative of $f(z)$ with respect to z:

$$\{f(z),z\}=\left(\frac{f''(z)}{f'(z)}\right)'-\frac{1}{2}\left(\frac{f''(z)}{f'(z)}\right)^2.$$

C. Coefficient Problems

In several distortion theorems **Koebe's extremal function** $z/(1-\varepsilon z)^2=\sum_{n=1}^{\infty} n\varepsilon^{n-1}z^n$ ($|\varepsilon|=1$) is extensively utilized. Concerning this, Bieberbach stated the following conjecture. If $f(z)=z+\sum_{n=2}^{\infty} a_n z^n$ is holomorphic and univalent in $|z|<1$, then $|a_n|\leqslant n$ ($n=2,3,\ldots$), with equality holding only for Koebe's extremal function $z/(1-\varepsilon z)^2$ ($|\varepsilon|=1$). This conjecture was solved affirmatively by L. de Branges in 1985 after enormous effort by many mathematicians, as described below.

Bieberbach (1916, [1]) proved $|a_2|\leqslant 2$ as a corollary to the area theorem. This result can be proved easily by most of the methods. In 1923 Löwner [2] proved $|a_3|\leqslant 3$, introducing his own method. Schaeffer and Spencer gave a proof of $|a_3|\leqslant 3$ by the variational method (*Duke Math. J.*, 10 (1943)). Furthermore, Jenkins used the method of the extremal metric to prove a coefficient inequality that implies $|a_3|\leqslant 3$ (*Analytic Functions*, Princeton Univ. Press, 1960). The problem of the fourth coefficient remained open until 1955, when P. R. Garabedian and Schiffer [3] proved $|a_4|\leqslant 4$ by the variational method. Their proof was extremely complicated. Subsequently, Z. Charzynski and Schiffer gave an alternative brief proof of $|a_4|\leqslant 4$ by using the Grunsky inequality (*Arch. Rational Mech. Anal.*, 5 (1960)). M. Ozawa (1969, [4]) and R. N. Pederson (1968, [5]) also used the Grunsky inequality to prove $|a_6|\leqslant 6$. In 1972, Pederson and Schiffer [6] proved $|a_5|\leqslant 5$. They applied the Garabedian-Schiffer inequality, a generalization of the Grunsky inequality which Garabedian and Schiffer had derived by the variational method.

On the other hand, W. K. Hayman [7] showed that for each fixed $f(z)=z+\sum_{n=2}^{\infty} a_n z^n$,

$$\lim_{n\to\infty}\frac{|a_n|}{n}=\alpha\leqslant 1,$$

with the equality holding only for Koebe's extremal function $z/(1-\varepsilon z)^2$ ($|\varepsilon|=1$). Further, it was shown that Koebe's extremal function $z/(1-z)^2$ gives a local maximum for the nth

coefficient in the sense that $\operatorname{Re}\{a_n\} \leqslant n$ whenever $|a_2-2|<\delta_n$ for some $\delta_n>0$ (Garabedian, G. G. Ross, and Schiffer, *J. Math. Mech.*, 14 (1964); E. Bombieri, *Inventiones Math.*, 4 (1967); Garabedian and Schiffer, *Arch. Rational Math. Anal.*, 26 (1967)).

In the most general form, the coefficient problem is to determine the region occupied by the points $(a_2, \ldots, a_n)$ for all functions $f(z)= z+\sum_{n=2}^{\infty} a_n z^n$ univalent in $|z|<1$. Schaeffer and Spencer [8] found explicitly the region for (a_2, a_3).

For the coefficients of functions $g(\zeta)=\zeta+\sum_{v=0}^{\infty} b_v \zeta^{-v}$ univalent in $|\zeta|>1$, the following results are known: $|b_1| \leqslant 1$ (Bieberbach [1]), $|b_2| \leqslant 2/3$ (Schiffer, *Bull. Soc. Math. France*, 66 (1938); Golusin, *Mat. Sb.*, 3 (1938)), $|b_3| \leqslant 1/2+e^{-6}$ (Garabedian and Schiffer, *Ann. Math.*, (2) 61 (1955)).

D. Other Classes of Univalent Functions

We have discussed the general family of functions univalent in the unit disk. There are also several results on distortion theorems and coefficient problems for subfamilies determined by conditions such as that the images are bounded, †starlike with respect to the origin, or †convex. For instance, if $f(z)=z+\sum_{n=2}^{\infty} a_n z^n$ is holomorphic and univalent in $|z|<1$ and its image is starlike with respect to the origin, then $|a_n| \leqslant n$ $(n=2,3, \ldots)$. If the image of $f(z)$ is convex, then $f(z)$ satisfies $|a_n| \leqslant 1$ $(n=2,3, \ldots)$ and the distortion inequalities

$$\frac{|z|}{1+|z|} \leqslant|f(z)| \leqslant \frac{|z|}{1-|z|},$$

$$\frac{1}{(1+|z|)^2} \leqslant|f'(z)| \leqslant \frac{1}{(1-|z|)^2}.$$

Here the equality sign appears at z_0 $(0<|z_0|<1)$ if and only if $f(z)$ is of the form $z/(1+\varepsilon z)$ with $\varepsilon= \pm|z_0|/z_0$.

On the other hand, problems on conformal mappings of multiply connected domains involve essential difficulties in comparison with the simply connected case. Although Bieberbach's method is unsuitable for multiply connected domains, Löwner's method, the method of contour integration, the variational method, and the method of the extremal metric remain useful (→ 77 Conformal Mappings).

E. Multivalent Functions

Multivalent functions are a natural generalization of univalent functions. There are several results that generalize classical results on univalent functions.

A function $f(z)$ that attains every value at most p times and some values exactly p times in a domain D is said to be p-**valent** in D and is called a **multivalent function** provided that $p>1$. In order for $f(z)=\sum_{n=0}^{\infty} a_n z^n$, holomorphic in $|z| \leqslant 1$, to be p-valent there, it is sufficient that it satisfies

$$p-1<\operatorname{Re}(zf'(z)/f(z))<p+1$$

on $|z|=1$. Hence it suffices to have

$$|a_p|-\sum_{n=2}^{p} n|a_{p+1-n}|>\sum_{n=2}^{\infty} n|a_{p-1+n}|.$$

If $f(z)=(1+a_1 z+a_2 z^2+\ldots)/z^p$ is holomorphic and p-valent in $0<|z| \leqslant 1$, then

$$\frac{d}{dr} \int_0^{2\pi} F(|f(re^{i\theta})|)\, d\theta \leqslant 0$$

for any increasing function $F(\rho)$ in $\rho \geqslant 0$. In particular, if $F(\rho)=\rho^2$, this becomes an area theorem from which follow coefficient estimates, etc., for p-valent functions.

Various subfamilies and generalized families of multivalent functions have been considered. Let $f(z)$ be p-valent in D, and $c_0+c_1 z+\ldots+c_{p-1} z^{p-1}+c_p f(z)$ be at most p-valent in D for any constants $c_0, c_1, \ldots, c_p$. Then $f(z)$ is said to be **absolutely** p-**valent** in D. If a function $f(z)$ holomorphic in a convex domain K satisfies $\operatorname{Re}(e^{i\alpha} f^{(p)}(z))>0$ for a real constant α, then $f(z)$ is absolutely p-valent in K. If $f(z)$ is absolutely p-valent in D, then

$$\left(\sum_{k=0}^{p-1} b_k z^k+b_p f(z)\right) \Bigg/ \left(\sum_{k=0}^{p-1} c_k z^k+c_p f(z)\right)$$

is at most p-value in D for any constants b_k and c_k.

If $f(z)$ is p-valent in the common part of a domain D and the disk centered at each point of D with a fixed radius ρ, then $f(z)$ is said to be **locally** p-**valent** in D, and ρ is called its **modulus**. A necessary and sufficient condition for $f(z)$, holomorphic in D, to be at most locally p-valent is that $f'(z), \ldots, f^{(p)}(z)$ not vanish simultaneously. In order for $f(z)$, holomorphic in D, to be **locally absolutely** p-**valent** it is necessary and sufficient that $f^{(p)}(z) \neq 0$. Let the number of $Re^{i\varphi}$-points of $f(z)$ in D be $n(D, Re^{i\varphi})$. If $f(z)$ satisfies

$$\frac{1}{2\pi} \int_0^{2\pi} n(D, Re^{i\varphi})\, d\varphi \leqslant p,$$

for any $R>0$, it is said to be **circumferentially mean** p-**valent** in D. If $f(z)$ satisfies

$$\int_0^R \int_0^{2\pi} n(D, Re^{i\varphi}) R\, dR\, d\varphi \leqslant p\pi R^2,$$

it is said to be **areally mean** p-**valent** in D. If $f(z)^q$ with $q>1$ is areally mean p-valent in D, then $f(z)$ is areally mean p/q-valent in D. For

$f(z)=(1+a_1z+a_2z^2+\ldots)/z^\lambda$ holomorphic and areally mean λ-valent in $0<|z|\leqslant 1$, the following area theorem holds:

$$\sum_{n=1}^{\infty}(n-1)|a_n|^2\leqslant\lambda.$$

Let E be a set containing at least three points. If $f(z)$ in D attains every value of E at most p times and a certain value of E exactly p times (it may attain values outside E more than p times), then $f(z)$ is said to be **quasi-p-valent** in D. If $w=f(z)$ is p-valent in D and $g(w)$ is quasi-q-valent in $f(D)$, then $g(f(z))$ is at most quasi-pq-valent in D.

The first success in obtaining sharp inequalities for multivalent functions was attained by Hayman. In his work, an essential role was played by the method of †symmetrization. For instance, he obtained the following result. If $f(z)=z^p+a_{p+1}z^{p+1}+\ldots$ is holomorphic and circumferentially mean p-valent in $|z|<1$, then $|a_{p+1}|\leqslant 2p$, and for $|z|=r$, $0<r<1$,

$$\frac{r^p}{(1+r)^{2p}}\leqslant|f(z)|\leqslant\frac{r^p}{(1-r)^{2p}},$$

$$|f'(z)|\leqslant\frac{p(1+r)}{r(1-r)}|f(z)|\leqslant\frac{pr^{p-1}(1+r)}{(1-r)^{2p+1}}.$$

References

[1] L. Bieberbach, Über die Koeffizienten derjenigen Potenzreihen, welche eine schlichte Abbildung des Einheitskreises vermitteln, S.-B. Preuss., Akad. Wiss., (1916), 940–955.

[2] K. Löwner, Untersuchungen über schlichte konforme Abbildungen des Einheitskreises I, Math. Ann., 89 (1923), 103–121.

[3] P. R. Garabedian and M. Schiffer, A proof of the Bieberbach conjecture for the fourth coefficient, J. Rational Mech. Anal., 4 (1955), 427–465.

[4] M. Ozawa, On the Bieberbach conjecture for the sixth coefficient, Kôdai Math. Sem. Rep., 21 (1969), 97–128.

[5] R. N. Pederson, A proof of the Bieberbach conjecture for the sixth coefficient, Arch. Rational Mech. Anal., 31 (1968–69), 331–351.

[6] R. N. Pederson and M. Schiffer, A proof of the Bieberbach conjecture for the fifth coefficient, Arch. Rational Mech. Anal., 45 (1972), 161–193.

[7] W. K. Hayman, The asymptotic behavior of p-valent functions, Proc. London Math. Soc., (3) 5 (1955), 257–284.

[8] A. C. Schaeffer and D. C. Spencer, Coefficient regions for schlicht functions, Amer. Math. Soc. Colloq. Publ., 35 (1950).

[9] G. M. Golusin, Geometric theory of functions of a complex variable, Amer. Math. Soc. Transl. of Math. Monographs, 26 (1969). (Original in Russian, 1952.)

[10] W. K. Hayman, Multivalent functions, Cambridge Univ. Press, 1958.

[11] J. A. Jenkins, Univalent functions and conformal mapping, Springer, second edition, 1965.

[12] L. V. Ahlfors, Conformal invariants: Topics in geometric function theory, McGraw-Hill, 1973.

[13] C. Pommerenke, Univalent functions (with a chapter on quadratic differentials by G. Jensen), Vandenhoeck & Ruprecht, 1975.

[14] G. Schober, Univalent functions—Selected topics, Lecture notes in math. 478, Springer, 1975.

[15] P. L. Duren, Coefficients of univalent functions, Bull. Amer. Math. Soc., 83 (1977), 891–911.

V

439 (III.19) Valuations

A. Introduction

There are two related kinds of valuations, additive (→ Section B) and multiplicative (→ Section C). The notion of valuations, originally defined on (commutative) †fields, has been extended to more general cases (→ Section K); however, we first consider the case of fields.

B. Additive Valuations

In this article, we mean by an **ordered additive group** a **totally ordered additive group**, namely, a commutative group whose operation is addition, which is a †totally ordered set satisfying the condition that $a \geqslant b$ and $c \geqslant d$ imply $a+c \geqslant b+d$ and $-a \leqslant -b$. Suppose that we are given a field K, an ordered additive group G, and an element ∞ defined to be greater than any element of G. Then a mapping $v: K \rightarrow G \cup \{\infty\}$ is called an **additive valuation** (or simply a **valuation**) of the field K if v satisfies the following three conditions: (i) $v(a)=\infty$ if and only if $a=0$; (ii) $v(ab)=v(a)+v(b)$ for all a, $b \neq 0$; and (iii) $v(a+b) \geqslant \min\{v(a), v(b)\}$.

The set $\{v(a) \mid a \in K-\{0\}\}$ is a submodule of G and is called the **value group** of v, while the set $R_v=\{a \in K \mid v(a) \geqslant 0\}$ is a subring of K and is called the **valuation ring** of v. The ring R_v has only one †maximal ideal $\{a \mid v(a)>0\}$, called the **valuation ideal** of v (or of R_v), and the †residue class field of R_v modulo the maximal ideal is called the **residue class field** of the valuation v. We have $v(a) \leqslant v(b)$ if and only if $aR_v \supset bR_v$. Two valuations v and v' of the field K are said to be **equivalent** when $v(a) \leqslant v(b)$ if and only if $v'(a) \leqslant v'(b)$; hence v and v' are equivalent if and only if $R_v=R_{v'}$. The **rank** of v is defined to be the †Krull dimension of the valuation ring R_v, and the **rational rank** of v to be the maximum (or supremum) of the numbers of linearly independent elements in the value group. An **extension** (or **prolongation**) of v in a field K' containing K is a valuation v' of K' whose restriction on K is v; such an extension exists for any given v and K'. Sometimes a valuation of rank 1 is called a **special valuation** (or **exponential valuation**), and a valuation of a general rank is called a **generalized valuation**. On the other hand, if k is a subfield of K such that $v(a)=0$ for every nonzero element a of k, then v is called a **valuation over the subfield** k.

C. Multiplicative Valuations

A **multiplicative valuation** (or **valuation**) of a field K is a mapping $w: K \rightarrow \Gamma \cup \{0\}$ that satisfies the following three conditions, where Γ is the multiplicative group of positive real numbers: (i) $w(a)=0$ if and only if $a=0$; (ii) $w(ab)=w(a)w(b)$; and (iii) $w(a+b) \leqslant C(w(a)+w(b))$, where C is a constant (independent of the choice of a and b, but dependent on the choice of w).

The **value group** of w is defined to be $\{w(a) \mid a \in K-\{0\}\}$. Extensions of a valuation and equivalence of valuations are defined as in the case of additive valuations. Thus w' is equivalent to w if and only if there is a positive r such that for all $a \in K$, $w(a)=w'(a)^r$. In each equivalence class of valuations of a field, there exists a valuation for which the constant C in condition (iii) can be taken to be 1. A valuation w is said to be a **valuation over a subfield** k if $w(a)=1$ for any nonzero element a of k.

We call w an **Archimedean valuation** if for any elements $a, b \in K$, $a \neq 0$, there exists a natural number n such that $w(na)>w(b)$; otherwise, w is said to be a **non-Archimedean valuation.** If w is an Archimedean valuation of a field K, then there is an injection σ from K into the complex number field **C** such that w is equivalent to the valuation w' defined by $w'(a)=|\sigma(a)|$. If w is a non-Archimedean valuation of a field K, then $w(a+b) \leqslant \max\{w(a), w(b)\}$. Hence in this case we get an additive valuation v of K when we define $v(a)=-\log w(a)$ $(a \in K)$, and either v is of rank 1 or $v(K)=\{1,0\}$ (in the latter case, v is called **trivial**). Conversely, every additive valuation of rank 1 of K is equivalent to an additive valuation obtained in this way from a non-Archimedean valuation. (This is why an additive valuation of rank 1 is called an exponential valuation.) Therefore a non-Archimedean valuation determines a valuation ring and valuation ideal in a natural manner. Thus we can identify a non-Archimedean valuation with an additive valuation of rank 1.

D. Topology Defined by a Valuation

Let w be a multiplicative valuation of a field K. When the †distance between two elements a, b of K is defined by $w(a-b)$, K becomes a †topological field. (Although this distance may not make K into a †metric space, there exists a valuation w' equivalent to the valuation w such that K becomes a metric space with respect to the distance $w'(a-b)$ between a and b $(a, b \in K)$.) If K is †complete under the topology, then we say that K is **complete** with respect to w and w is **complete** on K. On the other hand, suppose that w' is an extension of w in a field K' containing K. If w' is complete and K is †dense in K' under the topology defined by w', then we say that the valuation w' is a **completion** of w and that the field K' is a

completion of K with respect to w. For any w, a completion exists and is unique up to isomorphism. When w is a non-Archimedean valuation, the valuation ring of the completion of w is called the **completion** of the valuation ring of w.

When v is an additive valuation of a field K, we can introduce a topology on K by taking the set of all nonzero ideals of the valuation ring R_v of v as a †base for the neighborhood system of zero. Important cases are given by valuations of rank 1, which are the same as those given by non-Archimedean valuations.

If w is a complete non-Archimedean valuation of a field K, then the valuation ring R_w of w is a †Hensel ring, which implies that if K' is a finite algebraic extension of K such that $[K':K]=n$, then w is uniquely extendable to a valuation w' of K' and $w'(a)^n=w(N(a))$, where N is the †norm $N_{K'/K}$.

E. Discrete Valuations

For a non-Archimedean valuation (or an additive valuation of rank 1) w, if the valuation ideal of w is a nonzero †principal ideal generated by an element p, then we say that p is a **prime element** for w, w is a **discrete valuation**, and the valuation ring for w is a **discrete valuation ring**. The condition on the valuation ideal of w holds if and only if the value group of w is a discrete subgroup of the (multiplicative) group Γ of positive real numbers: In the terminology of additive valuations, a valuation w is discrete if and only if it is equivalent to a valuation w' whose value group is the additive group of integers. Such a valuation w' is called a **normalized valuation** (or **normal valuation**). However, we usually mean normalization of a discrete non-Archimedean valuation as in Section H. Sometimes an additive valuation whose value group is isomorphic to the direct sum of a finite number of copies of $\mathbf{Z}$ (the additive group of integers) with a natural †lexicographic order is called a discrete valuation. Concerning a complete discrete valuation w, it is known that if the valuation ring of w contains a field, then it is isomorphic to the ring of †formal power series in one variable over a field (for other cases → 449 Witt Vectors A).

F. Examples

(1) **Trivial valuations** of a field K are the additive valuation v of K such that $v(a)=0$ for all $a\in K-\{0\}$ and the multiplicative valuation w of K such that $w(a)=1$ for all $a\in K-\{0\}$.

(2) If K is isomorphic to a subfield of the complex number field, then we get an Archimedean valuation using the absolute value, and as stated in Section C, every Archimedean valuation of K is equivalent to a valuation obtained in this way.

(3) Let $\mathfrak{p}$ be a †prime ideal of a †Dedekind domain R, $\pi\in\mathfrak{p}$ be such that $\pi\notin\mathfrak{p}^2$, and K be the field of quotients of R. Then each nonzero element α of K can be expressed in the form $\pi^r ab^{-1}$ $(r\in\mathbf{Z}; a,b\in R; a,b\notin\mathfrak{p})$, where r, the **degree** of α with respect to $\mathfrak{p}$, is uniquely determined by α. Hence, letting c be a constant greater than 1, we obtain a non-Archimedean valuation w defined by $w(\alpha)=c^{-r}$. This valuation w is called a **$\mathfrak{p}$-adic valuation**. We also get an additive valuation v defined by $v(\alpha)=r$, called a **$\mathfrak{p}$-adic exponential valuation**. The completion $K_\mathfrak{p}$ of K with respect to v is called the **$\mathfrak{p}$-adic extension** of K. If K is a finite †algebraic number field, the $K_\mathfrak{p}$ is called a **$\mathfrak{p}$-adic number field**. If $\mathfrak{p}$ is generated by an element p, then "$\mathfrak{p}$-adic" is replaced by "p-adic." For instance, given a rational prime number p, we have a **p-adic valuation** of the rational number field $\mathbf{Q}$, and we obtain the p-adic extension $\mathbf{Q}_p$ of $\mathbf{Q}$, which is called the **p-adic number field**. Every nonzero element α of $\mathbf{Q}_p$ can be written as a uniquely determined expansion $\sum_{n=r}^{\infty}a_np^n$ $(a_r\neq 0, r\in\mathbf{Z}, a_n\in Z, 0\leqslant a_n<p)$. Then we obtain a valuation v of $\mathbf{Q}_p$ defined by $v(\alpha)=r$. This valuation v is a discrete additive valuation, and $\mathbf{Q}_p$ is complete with respect to v. The valuation ring of v is usually denoted by $\mathbf{Z}_p$, which is called the **ring of p-adic integers**. Each element of $\mathbf{Q}_p$ $(\mathbf{Z}_p)$ is called a **p-adic number** (**p-adic integer**).

(4) Consider the field of †power series $k((t))$ in one variable t over a field k. For $0\neq\alpha\in k((t))$, we define $v(\alpha)=r$ if $\alpha=\sum_{n=r}^{\infty}a_nt^n$ $(a_n\in k, a_r\neq 0)$. Then v is a discrete valuation of $k((t))$, and $k((t))$ is complete with respect to this valuation.

(5) Let v be an additive valuation of a field K with the valuation ring R_v and the valuation ideal $\mathfrak{m}_v$. Let v' be an additive valuation of the field $R_v/\mathfrak{m}_v$ with the valuation ring $R_{v'}$. Then $R''=\{a\in R_v\mid (a \bmod \mathfrak{m}_v)\in R_{v'}\}$ is a valuation ring of K. A valuation v'' whose valuation ring coincides with R'' is called the **composite** of v and v'.

G. The Approximation Theorem and the Independence Theorem

The **approximation theorem** states: Let $w_1,\ldots,w_n$ be mutually nonequivalent and nontrivial multiplicative valuations of a field K. Then for any given n elements $a_1,\ldots,a_n$ of K and a positive number ε, there exists an element a of K such that $w_i(a-a_i)<\varepsilon$ $(i=1,2,\ldots,n)$.

From this follows the **independence theorem**: Let $e_1, \ldots, e_n$ be real numbers, and let w_i and K be as in the approximation theorem. If $\prod_i w_i(a)^{e_i}=1$ for all $a \in K-\{0\}$, then $e_1=\ldots=e_n=0$.

Similar theorems hold for additive valuations. The following independence theorem is basic: Let $v_1, \ldots, v_n$ be additive valuations of a field K, $R_1, \ldots, R_n$ their valuation rings, and $\mathfrak{m}_1, \ldots, \mathfrak{m}_n$ their maximal ideals. Let $D=\bigcap_i R_i$, $\mathfrak{p}_i=\mathfrak{m}_i \cap D$, and consider the rings of quotients $D_{\mathfrak{p}_i}$. Then $D_{\mathfrak{p}_i}=R_i$. If $R_i \not\subset R_j$ (for $i \neq j$), then D has exactly n maximal ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$.

H. Prime Divisors

Let K be an †algebraic number field (algebraic function field of one variable over a field k). An equivalence class of nontrivial multiplicative valuations (over k) is called a **prime divisor (prime spot)** of K.

If K is an algebraic number field of degree n, there are exactly n mutually distinct injections $\sigma_1, \ldots, \sigma_n$ of K into the complex number field $\mathbf{C}$. We may assume that $\sigma_i(K)$ is contained in the real number field if and only if $i \leqslant r_1$ and $\sigma_{n-i+1}(a)$ and $\sigma_{r_1+i}(a)$ are conjugate complex numbers ($n-r_1 \geqslant i>0$, $a \in K$). For $i \leqslant r_1$, let $v_i(a)=|\sigma_i(a)|$, and for $1 \leqslant i \leqslant (n-r_1)/2$, let $v_{r_1+i}(a)=|\sigma_{r_1+i}(a)|^2$. Then $v_1, \ldots, v_{r_1+r_2}$ ($r_2=(n-r_1)/2$) is a maximal set of mutually nonequivalent Archimedean valuations of K. Equivalence classes of $v_1, \ldots, v_{r_1}$ are called **real (infinite) prime divisors**, and those of $v_{r_1+1}, \ldots, v_{r_1+r_2}$ are called **imaginary (infinite) prime divisors**; all of them are called **infinite prime divisors**. An equivalence class of non-Archimedean valuations of K is called a **finite prime divisor**.

An Archimedean valuation of K is said to be **normal** if it is one of the valuations v_i. If v is non-Archimedean, then v is a $\mathfrak{p}$-adic valuation, where $\mathfrak{p}$ is a prime ideal of the principal order $\mathfrak{o}$ of K (→ Section F, example (3)). Hence if a is an element of K, there exists a constant c ($c>1$) such that $v(a)=c^{-r}$, where r is the degree of a with respect to $\mathfrak{p}$. In particular, if c is the norm of $\mathfrak{p}$ (i.e., c is the cardinality of the set $\mathfrak{o}/\mathfrak{p}$), then the valuation v is called **normal**. Any finite prime divisor is represented by a normal valuation. Then we have the **product formula** $\prod_w w(a)=1$ for all $a \in K-\{0\}$, where w ranges over all normal valuations of K.

For a function field, a **normal valuation** is defined similarly, using e^f instead of the norm of $\mathfrak{p}$, where e is a fixed real number greater than 1 and f is the degree of the residue class field of the valuation over k. In this case we also have the product formula.

I. Extending Valuations to an Algebraic Extension of Finite Degree

Assume that a field K' is a finite algebraic extension of a field K. Let v be an additive valuation of K and v' be an extension of v to K'. We denote the valuation rings, valuation ideals, and value groups of v and v' by R_v, $R_{v'}$, $\mathfrak{m}_v$, $\mathfrak{m}_{v'}$, and G, G', respectively. Then the degree of the extension $f_{v'}=[R_{v'}/\mathfrak{m}_{v'}:R_v/\mathfrak{m}_v]$ is called the **degree** of v' over v. The group index $e_{v'}=[G':G]$ is called the **ramification index** of v' over v. If v' ranges over all extensions of v in K', then the sum $\sum f_{v'}e_{v'}$ is not greater than $[K':K]$ and the equality holds when v is a discrete valuation and either K' is †separable over K or v is complete.

J. Places

Let k, K, and L be fields, and suppose that $k \subset K$. Let f be a mapping of K onto $L \cup \{\infty\}$ such that $f(ab)=f(a)f(b)$ and $f(a+b)=f(a)+f(b)$, whenever the right member is meaningful, and such that the restriction of f to k is an injection. Here ∞ is an element adjoined to L and satisfying $\infty+a=a+\infty=\infty$, $\infty a=a\infty=\infty$ (for any nonzero element a of K), $1/\infty=0$, and $1/0=\infty$. Then f is called a **place** of K over k. In this case $R=\{x \in K \mid f(x) \neq \infty\}$ is a valuation ring of K containing k. Let $\mathfrak{m}$ be the maximal ideal of R. Then f can be identified with the mapping $g: K \to R/\mathfrak{m} \cup \{\infty\}$ defined as follows: If $a \in R$, then $g(a)=(a \bmod \mathfrak{m})$; otherwise, $g(a)=\infty$. Places of K over k can be classified in a natural way, and there exists a one-to-one correspondence between the set of classes of places of K over k and the set of equivalence classes of additive valuations over k. When K is an †algebraic function field, we usually consider the case where k is the †ground field. Then if $a_1, \ldots, a_n \in R$, $(a_1, \ldots, a_n) \to (g(a_1), \ldots, g(a_n))$ gives a †specialization of points over k. Conversely, if $a_i, b_j \in K$ are such that $(a_1, \ldots, a_n) \to (b_1, \ldots, b_n)$ is a specialization over k, then there is a place f of K over k such that $(b_1, \ldots, b_n)$ is isomorphic to $(f(a_1), \ldots, f(a_n))$ (usually there are infinitely many such f's).

K. Pseudovaluations

A **pseudovaluation** φ of a ring A (not necessarily commutative) is a mapping of A into the set of nonnegative real numbers satisfying the following four conditions: (i) $\varphi(a)=0$ if and only if $a=0$; (ii) $\varphi(ab) \leqslant \varphi(a)\varphi(b)$; (iii) $\varphi(a+b) \leqslant \varphi(a)+\varphi(b)$; and (iv) $\varphi(-a)=\varphi(a)$. These conditions are weaker than those for multiplicative valuations, but with them a topology

can be introduced into A as in Section D, with respect to which A becomes a topological ring.

L. History

The theory of valuations was originated by K. Hensel when he introduced p-adic numbers and applied them to number theory [1]. J. Kürschák (*J. Reine Angew. Math.*, 142 (1913)) first treated the theory of multiplicative valuations axiomatically; it was then developed remarkably by A. Ostrowski (*Acta Math.*, 41 (1918)). However, in their theory condition (iii) (→ Section C) was given only in the case $C=1$, thus excluding the normal valuation of an imaginary prime divisor in an algebraic number field. A valuation with general C was introduced by E. Artin [3]. The theory of additive valuations was originated by W. Krull (*J. Reine Angew. Math.*, 167 (1932)), although the concept of exponential valuations existed before. The theory of valuations is used to simplify †class field theory and the theory of algebraic function fields in one variable. For these purposes, the notion of multiplicative valuations is sufficient (→ 9 Algebraic Curves; 59 Class Field Theory). The idea is also used in the theory of normal rings and in algebraic geometry, for both of which the concept of additive valuations is also necessary. Pseudovaluations were used by M. Deuring (*Erg. Math.*, Springer, 1935) in the arithmetic of algebras.

References

[1] K. Hensel, Theorie der algebraischen Zahlen, Teubner, 1908.
[2] O. F. G. Schilling, The theory of valuations, Amer. Math. Soc. Math. Surveys, 1950.
[3] E. Artin, Algebraic numbers and algebraic functions, Gordon & Breach, 1967.
[4] O. Zariski and P. Samuel, Commutative algebra II, Van Nostrand, 1960.
Also → references to 67 Commutative Rings.

440 (X.35) Variational Inequalities

A. Introduction

Variational inequalities arise when we consider extremal problems of functionals under **unilateral constraints**. Some problems in physics and engineering are studied by formulating them as elliptic, parabolic, and hyperbolic variational inequalities [1–8].

B. Stationary Variational Inequality

Let D be a bounded domain in m-dimensional Euclidean space and $f \in L_2(D)$ be a given real-valued function. Consider the variational problem of minimizing the following functional J with the argument function v:

$$J[v]=\int_D |\operatorname{grad} v|^2 dx - 2\int_D fv\,dx.$$

Here, we suppose the set of admissible functions to be the closed convex subset

$$K=\{v \in H_0^1(D) \mid v \leqslant 0 \text{ a.e. in } D\}$$

of the Hilbert space $H_0^1(D)$ (→ 168 Function Spaces). It can be shown by choosing a minimizing sequence that there exists a minimum value of J which is realized by a unique $u \in K$. Since the stationary function u belongs to $H_0^1(D)$, it can be shown that the boundary condition $u|_{\partial D}=0$ is satisfied in the sense that the †trace $\gamma_0 u \in H^{1/2}(\partial D)$ (→ 224 Interpolation of Operators) of u on ∂D vanishes a.e. on ∂D. In view of the fact that $J[u] \leqslant J[v]$ is valid for any $v \in K$, it can be verified that the **stationary variational inequality**

$$\left.\begin{aligned} &-\Delta u - f \leqslant 0 \\ &u \leqslant 0 \\ &(-\Delta u - f)\cdot u = 0 \end{aligned}\right\} \qquad (1)$$

is satisfied in D in the sense of differentiation of distributions (→ 125 Distributions and Hyperfunctions). The problem (1) is a **Dirichlet problem with obstacle**. Moreover, we can prove the regularity of $u \in H^2(D)$ under an assumption of suitable smoothness for ∂D by establishing the boundedness of the solutions u_ε in $H^2(D)$ of the **penalized problems** associated with (1):

$$-\Delta u_\varepsilon + \frac{1}{\varepsilon} u_\varepsilon^+ = f \qquad (\varepsilon > 0),$$

$$u_\varepsilon|_{\partial D} = 0.$$

Here we note that the u_ε are the stationary functions of the ordinary variational problems of minimization in $H_0^1(D)$ of the functionals

$$J_\varepsilon[v]=\int_D |\operatorname{grad} v|^2 dx - 2\int_D fv\,dx + \frac{1}{\varepsilon}\int |v^+|^2 dx$$

with the **penalty term** (the third term of the right-hand side of the equality above). We have thus found that the stationary variational inequality (1) is the Euler equation of a conditional problem of variation (→ 46 Calculus of Variations).

C. Variational Inequality of Evolution

Let $\psi \in H^1(D)$ be a given function on D such that $\psi|_{\partial D} \geqslant 0$ and $\Delta\psi \in L_2(D)$. The **variational**

inequality of evolution

$$\frac{\partial u}{\partial t}-\Delta u\leqslant 0,$$

$$u\leqslant\psi,$$

$$\left(\frac{\partial u}{\partial t}-\Delta u\right)\cdot(u-\psi)=0\quad (t>0, x\in D),$$

$$u(0,x)=a(x)\qquad (x\in D),$$

$$u(t,x)|_{\partial D}=0\qquad (t>0)$$

can be formulated as an abstract Cauchy problem (→ 286 Nonlinear Functional Analysis X)

$$\frac{du}{dt}\in Au\qquad (t>0),$$

$$u(+0)=a$$

in a Hilbert space with a multivalued operator $A=-\partial\varphi$, where $\partial\varphi$ is the subdifferential of the following lower semicontinuous proper convex function on the Hilbert space $L_2(D)$:

$$\varphi(v)=\begin{cases}\dfrac{1}{2}\displaystyle\int_D|\operatorname{grad}v|^2dx & \text{if } v\in H_0^1(D) \text{ and } v\leqslant\psi,\\ +\infty & \text{otherwise.}\end{cases}$$

Thus the solution u is given by the vector-valued function

$$u(t)=e^{tA}a.$$

Here e^{tA} is the †nonlinear semigroup generated by A (→ 88 Convex Analysis, 378 Semigroups of Operators and Evolution Equations).

D. Optimal Stopping Time Problem and Variational Inequalities

Let $\{X_t\}_{t\geqslant 0}$ be an m-dimensional Brownian motion (→ 45 Brownian Motion) and consider the problem of finding a †stopping time σ that minimizes

$$J_x[\sigma]=\mathrm{E}_x\left(\int_0^\sigma f(X_t)\,dt\right)\qquad (x\in\mathbf{R}^m)$$

under the restriction that $0\leqslant\sigma\leqslant\sigma_{\partial D}$, where $\sigma_{\partial D}$ is the †hitting time for the boundary ∂D. Let us define

$$u(x)=\min_\sigma J_x[\sigma].$$

Then the †principle of optimality in dynamic programming gives the stationary variational inequality (1) with Δ replaced by $\frac{1}{2}\Delta$, and we can show by the †Dynkin formula that an optimal stopping time $\hat\sigma$ is the hitting time for the set $\{x\in\bar\Omega\,|\,u(x)=0\}$ (→ 127 Dynamic Programming). We can systematically discuss problems in mathematical programming and operations research by introducing quasivariational inequalities, which are slight generalizations of variational inequalities (→ 227 Inventory Control, 408 Stochastic Programming). The above-mentioned facts are applicable to general †diffusion processes described by †stochastic differential equations (→ 115 Diffusion Processes, 406 Stochastic Differential Equations). We have thus found the relation

free boundary problem ↔ variational inequality
↘ optimal stopping time problem ↙

(→ 405 Stochastic Control and Stochastic Filtering).

E. Numerical Solution of Variational Inequalities

Since the solution u of the variational inequality (1) is the stationary function for the variational problem, we can apply to the evaluation of the function u numerical methods based on the direct method of the calculus of variations (→ 300 Numerical Methods). The †finite element method, which can be regarded as a type of Ritz-Galerkin method, is extensively employed to calculate numerical solutions. In view of the unilateral constraint $u\leqslant 0$, iteration methods, such as the Gauss-Seidel iteration method, are used with modifications. An algorithm of relaxation with projection is proposed in [3] (→ 304 Numerical Solution of Partial Differential Equations).

References

[1] H. Brézis, Problèmes unilatéraux, J. Math. Pures Appl., 51 (1972), 1–168.
[2] V. Barbu, Nonlinear semigroups and differential equations in Banach spaces, Noordhoff, 1976.
[3] J.-L. Lions, Sur quelques questions d'analyse, de mécanique et de contrôle optimal, Les Presses de l'Université de Montréal, 1976.
[4] J.-L. Lions and G. Stampacchia, Variational inequalities, Comm. Pure Appl. Math., 20 (1967), 493–519.
[5] C. Baiocchi, Su un problema di frontiera libera connesso a questioni di idraulica, Ann. Mat. Pura Appl. 92 (1972), 107–127.
[6] G. Duvaut and J.-L. Lions, Les inéquations en mécanique et en physique, Dunod, 1972; English translation, Springer, 1975.
[7] A. Friedman and D. Kinderlehrer, A one phase Stefan problem, Indiana Univ. Math. J., 24 (1975), 1005–1035.
[8] R. Glowinski, J.-L. Lions, and R. Trémolières, Analyse numérique des inéquations variationelles I, II, Dunod, 1976; English translation, North-Holland, 1981.

441 (XX.3) Variational Principles

A. General Remarks

Among the principles that appear in physics are those expressed not in terms of differential forms but in terms of variational forms. These principles, describing the conditions under which certain quantities attain extremal values, are generally called **variational principles**. Besides Hamilton's principle in classical mechanics (→ Section B) and Fermat's principle in geometric optics (→ Section C), examples are found in †electromagnetism, †relativity theory, †quantum mechanics, †field theory, etc. Independence of the choice of coordinate system is an important characteristic of variational principles. Originally these principles had theological and metaphysical connotations, but a variational principle is now regarded simply as a postulate that precedes a theory and furnishes its foundation. Thus a variational principle is considered to be the supreme form of a law of physics.

B. Mechanics

In 1744 P. L. Maupertuis published an almost theological thesis, dealing with the **principle of least action**. This was the beginning of the search for a single, universal principle of mechanics, contributions to which were made successively by L. Euler, C. F. Gauss, W. R. Hamilton, H. R. Hertz, and others.

Let $\{q_r\}$ be the †generalized coordinates of a system of particles, and consider the integral of a function $L(q_r, \dot{q}_r, t)$ taken from time t_0 to t_1. If we compare the values of the integral taken along any arbitrary path starting from a fixed point P_0 in the coordinate space at time t_0 and arriving at another fixed point P_1 at time t_1, then the actual motion $q_r(t)$ (which obeys the laws of mechanics) is given by the condition that the integral is an †extremum (†stationary value), that is, $\delta\int_{t_0}^{t_1} L\,dt=0$, provided that the function L is properly chosen. This is **Hamilton's principle**, and L is the †Lagrangian function. In †Newtonian mechanics, the †kinetic energy T of a system of particles is expressed as a †quadratic form in $\dot{q}_r$. Furthermore, if the forces acting on the particles can be given by $-\operatorname{grad} V$, where the potential V does not depend explicitly on $\dot{q}_r$, we can choose $L=T-V$. Also, for a charged particle in †special relativity, we can take $L=-m_0c^2(1-v^2/c^2)^{1/2}-e\varphi+e(\mathbf{v}\cdot\mathbf{A})$, where m_0 is the rest mass of the particle, e is the charge, $\mathbf{v}$ is the velocity (with v its magnitude), c is the speed of light in vacuum, and φ and $\mathbf{A}$ are the scalar and vector potentials of the electromagnetic field, respectively.

In general relativity theory, the motion of a particle can be derived from the variational principle $\delta\int ds=0$ (ds is the Riemannian line element). Hence, geometrically, the particle moves along a †geodesic curve in 4-dimensional space-time.

C. Geometric Optics

The path of a light ray between two points P_0 and P_1 (subject to reflection and refraction) is such that the time of transit along the path among all neighboring virtual paths is an extremum (stationary value). This is called **Fermat's principle**. If the index of refraction is n, Fermat's principle can be expressed as $\delta\int_{P_0}^{P_1} n\,ds=0$ (ds is the Euclidean line element). The laws of reflection and refraction of light, as well as the law of rectilinear propagation of light in homogeneous media, can be derived from this principle.

D. Field Theory

Not only the equations of motion of a system of particles, but also various field equations (†Maxwell's equations of the electromagnetic field, †Dirac's equation of the electron field, the meson field equation, the gravitational field equation, etc.) can be derived from variational principles in terms of appropriate Lagrangian functions. In †field theory the essential virtue of the variational principle appears in the fact that the properties of various possible fields as well as conservation laws can be systematically discussed by assuming relativistic invariance and gauge invariance of the Lagrangian functions adopted. In particular, for an electromagnetic field in vacuum, the Lagrangian function density is $L=(\mathbf{H}^2-\mathbf{E}^2)/2$, and the integration is carried out over a certain 4-dimensional domain.

E. Quantum Mechanics

If H is the †Hamiltonian operator for any quantum-mechanical system, the eigenfunction ψ can be determined by the variational principle

$$\delta\int\bar{\psi}H\psi\,d\tau=0,\qquad\text{with}\qquad\int\bar{\psi}\psi\,d\tau=1,$$

where $\bar{\psi}$ is the complex conjugate of ψ and $d\tau$ is the volume element. Based on this variational principle, the †direct method of the calculus of variations is often employed for

an approximate numerical calculation of the energy eigenvalues and eigenfunctions. In particular, by restricting the functional form of ψ to the product of one-body wave functions, we can obtain Hartree's equation. A further suitable symmetrization of ψ leads to Fock's equation.

F. Statistical Mechanics

Let φ be a statistical-mechanical state of a system, and let $S(\varphi)$ and $E(\varphi)$ be the state's entropy and energy (mean entropy and mean energy for an infinitely extended system); T is the thermodynamical temperature, and $f(\varphi)=E(\varphi)-TS(\varphi)$ is the free energy. Then the equilibrium state for $T\geqslant 0$ is determined as the state φ that gives the minimum value of $f(\varphi)$ (maximum for $T<0$).

References

[1] R. Courant and D. Hilbert, Methods of mathematical physics, Interscience, I, 1953; II, 1962.
[2] B. L. Moiseiwitsch, Variational principles, Interscience, 1966.
[3] R. Weinstock, Calculus of variations, with applications to physics and engineering, McGraw-Hill, 1952.
Also → references to 46 Calculus of Variations.

442 (VI.12) Vectors

A. Definitions

The **vector** concept originated in physics from such well-known notions as velocity, acceleration, and force. These physical quantities are supplied with length and direction; they can be added or multiplied by scalars. In the Euclidean space E^n (or, in general, an †affine space), a vector **a** is represented by an **oriented segment** $\overrightarrow{pq}$. Two oriented segments $\overrightarrow{p_1q_1}$ and $\overrightarrow{p_2q_2}$ are considered to represent the same vector **a** if and only if the following two conditions are satisfied: (1) The four points p_1, q_1, p_2, q_2 lie in the same plane π. (2) $p_1q_1//p_2q_2$ and $p_1q_2//q_1q_2$. Hence a vector in E^n is an equivalence class of oriented segments $\overrightarrow{pq}$, where the equivalence relation $\overrightarrow{p_1q_1}\sim\overrightarrow{p_2q_2}$ is defined by the two conditions just given. Hereafter, we denote the vector by $[\overrightarrow{pq}]$, or simply $\overrightarrow{pq}$. The points p and q are called the **initial point** and **terminal point** of the vector $\overrightarrow{pq}$.

Given a vector $\mathbf{a}=\overrightarrow{pq}$ and a real number λ, we define the **scalar multiple** $\lambda\mathbf{a}$ as the vector $\overrightarrow{pr}$, where r is the point on the straight line containing both p and q such that the ratio $[\overrightarrow{pr}:\overrightarrow{pq}]$ is equal to λ (if $p=q$, then we put $r=p$). The operation $(\lambda,\mathbf{a})\rightarrow\lambda\mathbf{a}$ is called **scalar multiplication**. Given two vectors $\mathbf{a}=\overrightarrow{pq}$ and $\mathbf{b}=\overrightarrow{qs}$, the vector $\mathbf{c}=\overrightarrow{ps}$ is called the **sum** of **a** and **b** and is denoted by $\mathbf{c}=\mathbf{a}+\mathbf{b}$. The vector $\overrightarrow{pp}=\mathbf{0}$ is called the **zero vector**. If $\mathbf{a}=\overrightarrow{pq}$, we put $-\mathbf{a}=\overrightarrow{qp}$.

Scalar multiplication and addition of vectors satisfy the following seven conditions: (1) $\mathbf{a}+\mathbf{b}=\mathbf{b}+\mathbf{a}$ (commutative law); (2) $\mathbf{a}+(\mathbf{b}+\mathbf{c})=(\mathbf{a}+\mathbf{b})+\mathbf{c}$ (associative law); (3) $\mathbf{a}+\mathbf{0}=\mathbf{a}$; (4) for each **a** there is $-\mathbf{a}$ such that $\mathbf{a}+(-\mathbf{a})=\mathbf{0}$; (5) $\lambda(\mathbf{a}+\mathbf{b})=\lambda\mathbf{a}+\lambda\mathbf{b}$, $(\lambda+\mu)\mathbf{a}=\lambda\mathbf{a}+\mu\mathbf{a}$ (distributive laws); (6) $\lambda(\mu\mathbf{a})=(\lambda\mu)\mathbf{a}$ (associative law for scalar multiplication); and (7) $1\mathbf{a}=\mathbf{a}$. Hence the set V of all vectors in E^n forms a †real linear space. Sometimes, a set satisfying (1)–(7), that is, by definition, a linear space, is called a vector space, and its elements are called vectors.

The pair consisting of a vector $\overrightarrow{pq}$ and a specific initial point p of $\overrightarrow{pq}$ is sometimes called a **fixed vector**. An illustration of this is given by the force vector with its initial point being where the force is applied. By contrast, a vector $\overrightarrow{pq}$ is sometimes called a **free vector**. If we fix the origin o in E^n, then for any point p in E^n, the vector $\overrightarrow{op}$ is called the **position vector** of p.

If two vectors $\mathbf{a}=\overrightarrow{op}$ and $\mathbf{b}=\overrightarrow{oq}$ are †linearly dependent, they are sometimes said to be **collinear**. If there vectors $\mathbf{a}=\overrightarrow{op}$, $\mathbf{b}=\overrightarrow{oq}$, and $\mathbf{c}=\overrightarrow{or}$ are linearly dependent, they are sometimes said to be **coplanar**.

If a set of vectors $\mathbf{e}_1,\ldots,\mathbf{e}_n$ forms a †basis of a vector space V, then the vectors $\mathbf{e}_i$ are called **fundamental vectors** in V. Each vector $\mathbf{a}\in V$ is uniquely expressed as $\mathbf{a}=\sum\alpha_i\mathbf{e}_i$ ($\alpha_i\in\mathbf{R}$). We call $(\alpha_1,\ldots,\alpha_n)$ the **components** of the vector **a** with respect to the fundamental vectors $\mathbf{e}_1,\ldots,\mathbf{e}_n$.

B. Inner Product

In the Euclidean space E^n, the length of the line segment $\overline{pq}$ is called the **absolute value** (or **magnitude**) of the vector $\mathbf{a}=\overrightarrow{pq}$ and is denoted by $|\mathbf{a}|$. A vector of length one is called a **unit vector**. For two vectors $\mathbf{a}=\overrightarrow{op}$ and $\mathbf{b}=\overrightarrow{oq}$, the value $(\mathbf{a},\mathbf{b})=|\mathbf{a}||\mathbf{b}|\cos\theta$ is called the **inner product** (or **scalar product**) of **a** and **b**, where θ is the angle $\angle poq$. Instead of $(\mathbf{a},\mathbf{b})$, the notations $\mathbf{a}\cdot\mathbf{b}$, or $\mathbf{ab}$ are also used. If neither vector **a** nor vector **b** is equal to **0**, then $(\mathbf{a},\mathbf{b})=0$ implies $\angle poq=\pi/2$, that is, the orthogonality of the two vectors $\overrightarrow{op}$ and $\overrightarrow{oq}$. If we take an

†orthonormal basis $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ in E^n (i.e., a set of fundamental vectors with $|\mathbf{e}_i| = 1$, $(\mathbf{e}_i, \mathbf{e}_j) = 0$ $(i \neq j)$), the inner product of vectors $\mathbf{a} = \sum \alpha_i \mathbf{e}_i$, $\mathbf{b} = \sum \beta_i \mathbf{e}_i$ is equal to $\sum_{i=1}^n \alpha_i \beta_i$. The inner product has the following three properties (i) $(\mathbf{x}, \mathbf{x}) \geqslant 0$ and is zero if and only if $\mathbf{x} = \mathbf{0}$; (ii) $(\mathbf{x}, \mathbf{y}) = (\mathbf{y}, \mathbf{x})$; (iii) $(\mathbf{x}_1 + \mathbf{x}_2, \mathbf{y}) = (\mathbf{x}_1, \mathbf{y}) + (\mathbf{x}_2, \mathbf{y})$, $(\alpha \mathbf{x}, \mathbf{y}) = \alpha(\mathbf{x}, \mathbf{y})$ $(\alpha \in \mathbf{R})$. Similar linearity holds for $\mathbf{y}$.

Generally, an $\mathbf{R}$-valued †bilinear form $(\mathbf{x}, \mathbf{y})$ on a linear space V satisfying the previous three conditions is also called an inner product. If a linear space V is equipped with an inner product, the space is called an **inner product space** (→ 256 Linear Spaces H; 197 Hilbert Spaces). If V is an inner product space, the absolute value $|\mathbf{x}|$ of $\mathbf{x} \in V$ is defined to be $\sqrt{(\mathbf{x}, \mathbf{x})}$.

C. Vector Product

In the 3-dimensional Euclidean space E^3, we take an orthonormal basis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. Let $\mathbf{a}$ and $\mathbf{b}$ be vectors in E^3 whose components with respect to $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ are $(\alpha_1, \alpha_2, \alpha_3)$, $(\beta_1, \beta_2, \beta_3)$. The vector

$$\begin{vmatrix} \alpha_2 & \alpha_3 \\ \beta_2 & \beta_3 \end{vmatrix} \mathbf{e}_1 + \begin{vmatrix} \alpha_3 & \alpha_1 \\ \beta_3 & \beta_1 \end{vmatrix} \mathbf{e}_2 + \begin{vmatrix} \alpha_1 & \alpha_2 \\ \beta_1 & \beta_2 \end{vmatrix} \mathbf{e}_3,$$

which is symbolically written as

$$\begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ \alpha_1 & \alpha_2 & \alpha_3 \\ \beta_1 & \beta_2 & \beta_3 \end{vmatrix},$$

is called the **exterior product** or **vector product** of $\mathbf{a}$ and $\mathbf{b}$ and is denoted by $[\mathbf{a}, \mathbf{b}]$ or $\mathbf{a} \times \mathbf{b}$. The vector $[\mathbf{a}, \mathbf{b}]$ is determined uniquely up to its sign by $\mathbf{a}$ and $\mathbf{b}$ and is independent of the choice of the orthonormal basis.

Suppose that we have $\mathbf{a} = \overrightarrow{op}$, $\mathbf{b} = \overrightarrow{oq}$. Then $|[\mathbf{a}, \mathbf{b}]| = |\mathbf{a}| \cdot |\mathbf{b}| \sin \theta$, where $\theta = \angle poq$. Also $|[\mathbf{a}, \mathbf{b}]|$ is equal to the area of the parallelogram determined by $\mathbf{a}$ and $\mathbf{b}$. To illustrate the orientation of $[\mathbf{a}, \mathbf{b}]$, we sometimes use the idea of a turning screw. That is, the direction of a right-handed screw advancing while turning at o from p to q (within the angle less than 180°) coincides with the direction of $[\mathbf{a}, \mathbf{b}]$ (Fig. 1). The exterior product has the following three properties: (1) $[\mathbf{a}, \mathbf{b}] = -[\mathbf{b}, \mathbf{a}]$ (antisymmetric law); (2) $[\lambda \mathbf{a}, \mathbf{b}] = \lambda [\mathbf{a}, \mathbf{b}]$ (associative law for scalar multiplication); (3) $[\mathbf{a}, \mathbf{b} + \mathbf{c}] = [\mathbf{a}, \mathbf{b}] + [\mathbf{a}, \mathbf{c}]$ (distributive law). The vector product does not satisfy the associative law, but it does satisfy the †Jacobi identity $[\mathbf{a}, [\mathbf{b}, \mathbf{c}]] + [\mathbf{b}, [\mathbf{c}, \mathbf{a}]] + [\mathbf{c}, [\mathbf{a}, \mathbf{b}]] = \mathbf{0}$. The vector $[\mathbf{a}, [\mathbf{b}, \mathbf{c}]]$ is sometimes called the **vector triple product**, and for this we have **Lagrange's formula** $[\mathbf{a}, [\mathbf{b}, \mathbf{c}]] = (\mathbf{a}, \mathbf{c})\mathbf{b} - (\mathbf{a}, \mathbf{b})\mathbf{c}$.

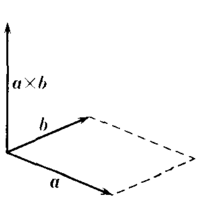


Fig. 1

Let $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ be vectors in E^3 whose components with respect to an orthonormal fundamental basis are $(\alpha_1, \alpha_2, \alpha_3)$, $(\beta_1, \beta_2, \beta_3)$, and $(\gamma_1, \gamma_2, \gamma_3)$. Then $(\mathbf{a}, [\mathbf{b}, \mathbf{c}]) = (\mathbf{b}, [\mathbf{c}, \mathbf{a}]) = (\mathbf{c}, [\mathbf{a}, \mathbf{b}]) = [\mathbf{a}, \mathbf{b}, \mathbf{c}]$, and the common value is equal to the determinant of the 3×3 matrix

$$\begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \beta_1 & \beta_2 & \beta_3 \\ \gamma_1 & \gamma_2 & \gamma_3 \end{pmatrix}.$$

The value denoted by $[\mathbf{a}, \mathbf{b}, \mathbf{c}]$ is called the **scalar triple product** of $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ and is equal to the volume of the parallelotope whose three edges are $\mathbf{a} = \overrightarrow{op}$, $\mathbf{b} = \overrightarrow{oq}$, and $\mathbf{c} = \overrightarrow{or}$ with common initial point o. The triple $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ is called a right-hand system or a left-hand system according as $[\mathbf{a}, \mathbf{b}, \mathbf{c}]$ is positive or negative. We have $[\mathbf{a}, \mathbf{b}, \mathbf{c}] = 0$ if and only if $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ are coplanar. (For the †exterior product of vectors in E^n and the concept of †p-vectors → 256 Linear Spaces O.)

D. Vector Fields

In this section we consider the case of a 3-dimensional Euclidean space E^3 (for the general case → 105 Differentiable Manifolds). A scalar-valued or a vector-valued function defined on a set D in E^3 is called a **scalar field** or a **vector field**, respectively. The continuity or the differentiability of a vector field is defined by the continuity or the differentiability of its components.

For a differentiable scalar field $f(x, y, z)$, the vector field with the components $(\partial f/\partial x, \partial f/\partial y, \partial f/\partial z)$ is called the **gradient** of f and is denoted by $\mathbf{grad} f$. For a differentiable vector field $\mathbf{V}(x, y, z)$ whose components are $(u(x, y, z), v(x, y, z), w(x, y, z))$, the vector field with components

$$\left(\frac{\partial w}{\partial y} - \frac{\partial v}{\partial z}, \frac{\partial u}{\partial z} - \frac{\partial w}{\partial x}, \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right)$$

is called the **rotation** (or **curl**) of $\mathbf{V}$ and is denoted by $\mathbf{rot}\,\mathbf{V}$ (or $\mathbf{curl}\,\mathbf{V}$). Also, for a differentiable vector field $\mathbf{V}$, the scalar field defined by $\partial u/\partial x + \partial v/\partial y + \partial w/\partial z$ is called the **divergence** of $\mathbf{V}$ and is denoted by $\mathbf{div}\,\mathbf{V}$. Utilizing the vector operator ∇ having differential operators $(\partial/\partial x, \partial/\partial y, \partial/\partial z)$ as its components, we may write simply $\operatorname{grad} f = \nabla f$, $\operatorname{div} \mathbf{V} = (\nabla, \mathbf{V})$, $\operatorname{rot} \mathbf{V} =$

$[\nabla, \mathbf{V}]$. The symbol ∇ is called **nabla**, **atled** (inverse of delta), or **Hamiltonian**.

A vector field $\mathbf{V}$ with $\operatorname{rot}\mathbf{V}=0$ is said to be **irrotational**, (**lamellar**, or **without vortex**). A vector field $\mathbf{V}$ with $\operatorname{div}\mathbf{V}=0$ is said to be **solenoidal** (or **without source**). Thus grad f is irrotational and rot $\mathbf{V}$ is solenoidal. In a small neighborhood or in a †simply-connected domain, an irrotational field is a gradient, a solenoidal field is a rotation, and an arbitrary vector field $\mathbf{V}$ is the sum of these two kinds of vector fields: $V=\operatorname{grad}\varphi+\operatorname{rot}\mathbf{u}$ (**Helmholtz theorem**); the function φ is called the **scalar potential** of $\mathbf{V}$, and the vector field $\mathbf{u}$ is called the **vector potential** of $\mathbf{V}$. Furthermore, the operator $\nabla^2=\nabla\nabla=\operatorname{div}\operatorname{grad}=\partial^2/\partial x^2+\partial^2/\partial y^2+\partial^2/\partial z^2$ is called the **Laplace operator** (or **Laplacian**) and is denoted by Δ. A function that satisfies $\Delta\varphi=0$ is called a †harmonic function. Locally, an irrotational and solenoidal vector field is the gradient of a harmonic function. If $\mathbf{A}$ is a vector field whose components are $(\varphi_1, \varphi_2, \varphi_3)$ (i.e., $\mathbf{A}(\mathbf{v})=(\varphi_1(\mathbf{v}), \varphi_2(\mathbf{v}), \varphi_3(\mathbf{v}))$), we can let Δ operate on $\mathbf{A}$ by setting $\Delta\mathbf{A}=(\Delta\varphi_1, \Delta\varphi_2, \Delta\varphi_3)$. We then have $\Delta\mathbf{A}=\nabla^2\mathbf{A}=\operatorname{grad}\operatorname{div}\mathbf{A}-\operatorname{rot}\operatorname{rot}\mathbf{A}$.

Suppose that we are given a vector field $\mathbf{V}$ and a curve C such that the vector $\mathbf{V}(p)$ is tangent to the curve at each point $p\in C$. The curve C is the †integral curve of the vector field $\mathbf{V}$ and is called the **vector line** of the vector field $\mathbf{V}$. The set of all vector lines intersecting with a given closed curve C is called a **vector tube**. Given a closed curve C and a vector field $\mathbf{V}$, the †curvilinear integral $\int(\mathbf{V}, d\mathbf{s})$ (where $d\mathbf{s}$ is the line element of C) is called the **circulation** (of $\mathbf{V}$) along the closed curve C. A vector field is irrotational if its circulation along every closed curve vanishes; the converse is true in a simply connected domain. Further, let v_n be the †normal component of a vector field $\mathbf{V}$ with respect to a surace S, and let dS be the volume element of the surface. We put $\mathbf{n}dS=d\mathbf{S}$, where $\mathbf{n}$ is the unit normal vector in the positive direction of the surface S. Then the †surface integral $\int v_n dS=\int(\mathbf{V}, d\mathbf{S})$ is called the **vector flux** through the surface S. A vector field whose vector flux vanishes for every closed surface is solenoidal. (For the corresponding formulas → 94 Curvilinear Integrals and Surface Integrals. For generalizations to higher-dimensional manifolds → 105 Differentiable Manifolds; 194 Harmonic Integrals; Appendix A, Table 3.)

References

[1] H. Weyl, Raum, Zeit, Materie, Springer, fifth edition, 1923; English translation, Space, time, matter, Dover, 1952.

[2] S. Banach, Mechanics, Warsaw, 1951.

[3] H. Flanders, Differential forms, with applications to the physical sciences, Academic Press, 1963.

[4] H. K. Nickerson, D. C. Spencer, and N. E. Steenrod, Advanced calculus, Van Nostrand, 1959.

443 (XII.8) Vector-Valued Integrals

A. General Remarks

Integrals whose values are elements (or subsets) of †topological linear spaces are generally called **vector-valued integrals** or **vector integrals**. As in the scalar case, there are vector-valued integrals of Riemann type (→ 37 Banach Spaces K) and of Lebesgue type. In this article we consider only the latter. There are cases where integrands are vector-valued, where measures are vector-valued, and where both are vector-valued. The methods of integration are also divided into the strong type, in which the integrals are defined by means of the original topology of the topological linear space X, and the weak type, in which they are reduced to numerical integrals by applying continuous linear functionals on X. Combining these we can define many kinds of integrals.

Historically, D. Hilbert's †spectral resolution is the first example of vector-valued integrals, but the general theory of vector-valued integrals started only after S. Bochner [1] defined in 1933 an integral of strong type for functions with values in a Banach space with respect to numerical measures. Then G. Birkhoff [2] defined a more general integral by replacing absolutely convergent sums with unconditionally convergent sums. At approximately the same time, N. Dunford introduced integrals equivalent to these. Later, R. S. Phillips (*Trans. Amer. Math. Soc.*, 47 (1940)) generalized the definition to the case where values of functions are in a †locally convex topological linear space, and C. E. Rickart (*Trans. Amer. Math. Soc.*, 52 (1942)) to the case where functions take subsets of a locally convex topological linear space as their values. The theory of integrals of weak type for functions with values in a Banach space and numerical measures was constructed by I. M. Gel'fand [3], Dunford [4], B. J. Pettis [5], and others (1936–1938). N. Bourbaki [6] dealt with the case where integrands take values in a locally convex topological linear space. As for integrals of numerical functions by vector-valued

measures, a representative of strong type integrals is the integral of R. G. Bartle, Dunford, and J. T. Schwartz [7] (1955). Weak type integrals have been discussed by Bourbaki [6], D. R. Lewis (*Pacific J. Math.*, 33 (1970)), and I. Kluvánek (*Studia Math.*, 37 (1970)). The bilinear integral of Bartle (*Studia Math.*, 15 (1956)) is typical of integrals in the case where both integrands and measures are vector-valued. For interrelations of these integrals → the papers by Pettis and Bartle cited above and T. H. Hildebrandt's report in the *Bulletin of the American Mathematical Society*, 59 (1953).

Since the earliest investigations [1–3] the main aim of the theory of vector-valued integrals has been to obtain integral representations of vector-valued (set) functions and various linear operators [8]. However, there is the fundamental difficulty of the nonvalidity of the Radon-Nikodým theorem. Whatever definition of integrals we take, the theorem does not hold for vector-valued set functions unconditionally. Many works sought conditions for functions, operators, or spaces such that the conclusion of the theorem would be restored; the works of Dunford and Pettis [9] and Phillips (*Amer. J. Math.*, 65 (1943)) marked a summit of these attempts. Later, after A. Grothendieck's investigations (1953–1956), this problem began to be studied again, beginning in the late 1960s, by many mathematicians (→ J. Diestel and J. J. Uhl, Jr., *Rocky Mountain J. Math.*, 6 (1976); [10]).

Recently, integrals of multivalued vector-valued functions have also been employed in mathematical statistics, economics, control theory, and many other fields. Some contributions are, besides Rickart cited above, G. B. Price (*Trans. Amer. Math. Soc.*, 47 (1940), H. Kudo (*Sci. Rep. Ochanomizu Univ.*, 4 (1953)), H. Richter (*Math. Ann.*, 150 (1963)), R. J. Aumann [11], G. Debreu [12], and M. Hukuhara (*Funkcial. Ekvac.*, 10 (1967)). Furthermore, C. Castaing and M. Varadier [13] have defined weak type integrals of multivalued functions and introduced many results concerning them. In the following we shall give explanations of typical vector-valued integrals with values in a Banach space only.

B. Measurable Vector-Valued Functions

Let $x(s)$ be a function defined on a †σ-finite measure space $(S, \mathfrak{S}, \mu)$ with values in a Banach space X. This is called a **simple function** or **finite-valued function** if there exists a partition of S into a finite number of mutually disjoint measurable sets $A_1, A_2, \ldots, A_n$ in each of which $x(s)$ takes a contant value c_j. Then $x(s)$ can be written as $\sum_{j=1}^{n} c_j\chi_{A_j}(s)$, where $\chi_{A_j}(s)$ is the †characteristic function of A_j. A function $x(s)$ is said to be **measurable** or **strongly measurable** if it is the strong limit of a sequence of simple functions almost everywhere, that is, $\lim_{n\to\infty}\|x_n(s)-x(s)\|=0$ a.e. Then the numerical function $\|x(s)\|$ is measurable. If μ is a †Radon measure on a compact Hausdorff space S, then the measurable functions can be characterized by †Luzin's property (→ 270 Measure Theory I).

A function $x(s)$ is said to be **scalarly measurable** or **weakly measurable** if the numerical function $\langle x(s), x'\rangle$ is measurable for any †continuous linear functional $x'\in X'$. A function $x(s)$ is measurable if and only if it is scalarly measurable and there are a †null set $E_0\subset S$ and a †separable closed subspace $Y\subset X$ such that $x(s)\in Y$ whenever $s\notin E_0$ (**Pettis measurability theorem**).

C. Bochner Integrals

A measurable vector-valued function $x(s)$ is said to be **Bochner integrable** if the norm $\|x(s)\|$ is †integrable. If $x(s)$ is a Bochner integrable simple function $\sum c_j\chi_{A_j}(s)$, then its Bochner integral is defined by

$$\int_S x(s)\,d\mu=\sum \mu(A_j)c_j.$$

For a general Bochner integrable function $x(s)$ there exists a sequence of simple functions satisfying the following conditions: (i) $\lim_{n\to\infty}\|x_n(s)-x(s)\|=0$ a.e. (ii) $\lim_{n\to\infty}\int_S\|x_n(s)-x(s)\|\,d\mu=0$. Then $\int_S x_n(s)\,d\mu$ converges strongly and its limit does not depend on the choice of the sequence $\{x_n(s)\}$. We call the limit the **Bochner integral** of $x(s)$ and denote it by $\int_S x(s)\,d\mu$ or by $(\mathrm{Bn})\int_S x(s)\,d\mu$ to distinguish it from other kinds of integrals. A Bochner integrable function on S is Bochner integrable on every measurable subset of S. The Bochner integral has the basic properties of Lebesgue integrals, such as linearity, †complete additivity, and †absolute continuity, with absolute values replaced by norms. †Lebesgue's convergence theorem and †Fubini's theorem also hold. However, the Radon-Nikodým theorem does not hold in general (→ Section H). Let T be a †closed linear operator from X to another Banach space Y. If both $x(s)$ and $Tx(s)$ are Bochner integrable, then the integral of $x(s)$ belongs to the domain of T and

$$T\left(\int_S x(s)\,d\mu\right)=\int_S Tx(s)\,d\mu.$$

If, in particular, T is bounded, then the assumption is always satisfied. If μ is the †Le-

besgue measure on the Euclidean space $\mathbf{R}^n$, then Lebesgue's differentiability theorem holds for the Bochner integrals regarded as a set function on the regular closed sets (→ 380 Set Functions D).

D. Unconditionally Convergent Series

Let $\sum_{j=1}^{\infty} x_j$ be a series of elements x_j of a Banach space X. It is said to be **absolutely convergent** if $\sum \|x_j\| < \infty$. It is called **unconditionally convergent** if for any rearrangement α the resulting series $\sum x_{\alpha(j)}$ converges strongly. Then the sum does not depend on α. Clearly, an absolutely convergent series is unconditionally convergent. If X is the number space or is finite-dimensional, then the converse holds. However, if X is infinite-dimensional, there is always an unconditionally convergent series which is not absolutely convergent (**Dvoretzky-Rogers theorem**).

A series $\sum x_j$ is unconditionally convergent if and only if each subseries converges weakly (**Orlicz-Pettis theorem**). If $\sum x_j$ is an unconditionally convergent series, then $\sum \langle x_j, x' \rangle$ converges absolutely for any continuous linear functional $x' \in X'$. If X is a Banach space containing no closed linear subspace isomorphic to the †sequence space c_0, then conversely a series $\sum x_j$ converges unconditionally whenever $\sum |\langle x_j, x' \rangle| < \infty$ for any $x' \in X'$ (**Bessaga-Pełczyński theorem**). A Banach space that is †sequentially complete relative to the weak topology, such as a †reflexive Banach space, and a separable Banach space that is the dual of another Banach space, such as l_1 and the †Hardy space $H_1(\mathbf{R}^n)$, satisfy the assumption, while c_0, l_∞, and $L_\infty(\Omega)$ for an infinitely divisible Ω do not. The totality of absolutely convergent series (resp. unconditionally convergent series) in X is identified with the †topological tensor product $l_1 \hat{\otimes} X$ (resp. $l_1 \hat{\hat{\otimes}} X$) (Grothendieck).

E. Birkhoff Integrals

We say that a series $\sum B_j$ of subsets of X converges unconditionally if for any $x_j \in B_j$ the series $\sum x_j$ converges unconditionally. Then $\sum B_j$ denotes the set of such sums. A vector-valued function $x(s)$ is said to be **Birkhoff integrable** if there is a countable partition Δ: $S = \bigcup_{j=1}^{\infty} A_j$ ($A_j \in \mathfrak{S}$, $A_j \cap A_k = \varnothing$ $(j \neq k)$, $\mu(A_j) < \infty$) such that the set $x(A_j)$ of values on A_j are bounded and $\sum \mu(A_j) x(A_j)$ converges unconditionally and if the sum converges to an element of X as the partition is subdivided. The limit is called the **Birkhoff integral** of $x(s)$ and is denoted by $(\mathrm{Bk})\int_S x(s)\,d\mu$ or simply by $\int_S x(s)\,d\mu$. A Birkhoff integrable function is Birkhoff integrable on any measurable set. The Birkhoff integral has, as a set function, complete additivity and absolute continuity in μ. It is linear in the integrand but Fubini's theorem and the Radon-Nikodým theorem do not hold. A Bochner integrable function is Birkhoff integrable, and the integrals coincide. The converse does not hold.

F. Gel'fand-Pettis Integrals

A scalarly measurable function $x(s)$ is said to be **scalarly integrable** or **weakly integrable** if for each $x' \in X'$, $\langle x(s), x' \rangle$ is integrable. Then the linear functional x^* on X' defined by

$$\int_S \langle x(s), x' \rangle \, d\mu = \langle x', x^* \rangle$$

is called the **scalar integral** of $x(s)$. Gel'fand [3] and Dunford [4] proved that x^* belongs to the bidual X''. Hence scalarly integrable functions are often called **Dunford integrable** and the integrals x^* the **Dunford integrals**. More generally, Gel'fand [3] showed that if $x'(s)$ is a function with values in the dual X' of a Banach space X such that $\langle x, x'(s) \rangle$ is integrable for any $x \in X$, then there is an $x' \in X'$ satisfying

$$\int_S \langle x, x'(s) \rangle \, d\mu = \langle x, x' \rangle.$$

This element is sometimes called the **Gel'fand integral** of $x'(s)$. A scalarly integrable function $x(s)$ is scalarly integrable on any measurable subset A. If the scalar integral is always in X, i.e., for each A there is an $x_A \in X$ such that

$$\int_A \langle x, x'(s) \rangle \, d\mu = \langle x_A, x' \rangle, \; x' \in X',$$

then $x(s)$ is said to be **Pettis integrable** or **Gel'fand-Pettis integrable** and x_A is called the **Pettis integral** or **Gel'fand-Pettis integral** on A and is denoted by $(\mathrm{P})\int_A x(s)\,d\mu$ or simply by $\int_A x(s)\,d\mu$. The Pettis integral has complete additivity and absolute continuity as a set function, similarly to the Birkhoff integral. Again, Fubini's theorem and the Radon-Nikodým theorem do not hold. The scalar integral on measurable sets of a scalarly integrable function $x(s)$ is completely additive and absolutely continuous with respect to the †weak* topology of X'' as the dual to X'. It is completely additive or absolutely continuous in the norm topology if and only if $x(s)$ is Pettis integrable (Pettis [5]; [10]). If $x(s)$ is Pettis integrable and $f(s)$ is a numerical function in $L_\infty(S)$, then the product $f(s)x(s)$ is Pettis integrable. Birkhoff integrable functions are Pettis integrable, and the integrals coin-

cide. Conversely, if a measurable function is Pettis integrable, then it is Birkhoff integrable. When X satisfies the Bessaga-Pełczyński condition (→ Section D), a measurable scalarly integrable function is Pettis integrable.

G. Vector Measures

Let Φ be a set function defined on a completely additive class $\mathfrak{S}$ of subsets of the space S and with values in a Banach space X. It is called a **finitely additive vector measure** (resp. a **completely additive vector measure** or simply a **vector measure**) if $\Phi(A_1 \cup A_2) = \Phi(A_1) + \Phi(A_2)$ whenever A_1 and $A_2 \in \mathfrak{S}$ are disjoint (resp. $\Phi(\bigcup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} \Phi(A_j)$ in the norm topology for all $A_j \in \mathfrak{S}$ such that $A_j \cap A_k = \varnothing$ $(j \neq k)$). We remark that the latter sum always converges unconditionally. A set function Φ is completely additive if and only if $\langle \Phi(A), x' \rangle$ is completely additive for all $x' \in X'$ (**Pettis complete additivity theorem**).

Let Φ be a finitely additive vector measure and E be a measurable set. The **total variation** of Φ on E and the **semivariation** of Φ on E are defined by

$$V(\Phi)(E) = \sup \sum_{j=1}^{n} \|\Phi(A_j)\| \tag{1}$$

and

$$\|\Phi\|(E) = \sup \left\| \sum_{j=1}^{n} \alpha_j \Phi(A_j) \right\|, \tag{2}$$

respectively, where the suprema are taken over all finite partitions of $E: E = \bigcup A_j$ $(A_j \in \mathfrak{S}, A_j \cap A_k = \varnothing (j \neq k))$ and all numbers α_j with $|\alpha_j| \leq 1$. If $V(\Phi)(S) < \infty$, then Φ is called a measure **of bounded variation**. $\|\Phi\|(S) < \infty$ if and only if $\sup\{\|\Phi(A)\| \mid A \in \mathfrak{S}\} < \infty$. Then Φ is said to be **bounded**. The function $V(\Phi)(E)$ of E is finitely additive but $\|\Phi\|(E)$ is only subadditive: $\|\Phi\|(A \cup B) \leq \|\Phi\|(A) + \|\Phi\|(B)$. If Φ is a vector measure of bounded variation, then $V(\Phi)$ is a positive measure. Every vector measure is bounded. A completely additive vector measure on a †finitely additive class $\mathfrak{L}$ can uniquely be extended to a vector measure on the completely additive class $\mathfrak{S}$ generated by $\mathfrak{L}$ (Kluvánek).

Let μ be a positive measure and Φ be a vector measure. Then we have $\Phi(A) \to 0$ as $\mu(A) \to 0$ if and only if Φ vanishes on every A with $\mu(A) = 0$. Then Φ is said to be **absolutely continuous** with respect to μ. For every vector measure Φ there is a measure μ such that $\|\Phi\|(A) \to 0$ as $\mu(A) \to 0$ and that $0 \leq \mu(A) \leq \|\Phi\|(A)$ (Bartle, Dunford, and Schwartz). As a set function, the Bochner integral is a vector measure of bounded variation and the Pettis integral is a bounded vector measure. Both are absolutely continuous with respect to the integrating measure. Let X be $L_p(0,1)$ for $1 \leq p \leq \infty$, and define $\Phi(E)$ for a Lebesgue measurable set E to be the characteristic function of E. If $p = 1$, Φ is a vector measure of bounded variation. If $1 < p < \infty$, Φ is a bounded vector measure, but it is not of bounded variation on any set E with $\mu(E) > 0$. If $p = \infty$, then Φ is no longer completely additive. These vector measures are absolutely continuous with respect to the Lebesgue measure, but they cannot be represented as the Bochner integral or the Pettis integral.

Let Φ be a vector measure on $\mathfrak{S}$. An $\mathfrak{S}$-measurable numerical function $f(s)$ is said to be Φ-integrable if there exists a sequence of simple functions $f_n(s)$ such that $f_n(s) \to f(s)$ a.e. and that for each $E \in \mathfrak{S}$, $\int_E f_n(s)\, d\Phi$ converges in the norm of X. Then the limit is independent of the choice of f_n. It is called the **Bartle-Dunford-Schwartz integral** and is denoted by $\int_E f(s)\, d\Phi$. Lebesgue's convergence theorem holds for this integral. If Φ is absolutely continuous with respect to the measure μ, then every $f \in L_\infty(\mu)$ is Φ-integrable, and the operator that maps f to $\int_S f\, d\Phi$ is continuous with respect to the weak* topology in $L_\infty(\mu)$ and the weak topology of X. Hence it is a †weakly compact operator. In particular, the range of a vector measure is relatively compact in the weak topology [7]. If Φ is the vector measure of the Pettis integral of a vector-valued function $x(s)$, then the above integral is equal to the Pettis integral of $f(s)x(s)$.

A vector measure Φ is said to be **nonatomic** if for each set A with $\Phi(A) \neq 0$ there is a subset B of A such that $\Phi(B) \neq 0$ and $\Phi(A \setminus B) \neq 0$. If X is finite-dimensional, then the range of a nonatomic vector measure is a compact convex set (**Lyapunov convexity theorem**). This has been generalized to the infinite-dimensional case in many ways, but the conclusion does not hold in the original form (→ Kluvánek and G. Knowles [15]; [10]).

H. The Radon-Nikodým Theorem

As the above examples show, the †Radon-Nikodým theorem does not hold for vector measures in the original form. From 1967 to 1971, M. Metivier, M. A. Rieffel, and S. Moedomo and Uhl improved the classical result of Phillips (1943) and proved the following theorem.

Radon-Nikodým theorem for vector measures. The following conditions are equivalent for μ-absolutely continuous vector measures Φ defined on a finite measure space $(S, \mathfrak{S}, \mu)$: (i) There is a Pettis integrable measurable func-

tion $x(s)$ such that

$$\Phi(A)=(\mathrm{P})\int_A x(s)\,d\mu.$$

(ii) For each $\varepsilon>0$ there is an $E\in\mathfrak{S}$ such that $\mu(S\setminus E)<\varepsilon$ and such that $\{\Phi(A)/\mu(A)\mid A\in\mathfrak{S}, A\subset E\}$ is relatively compact. (iii) For each $E\in\mathfrak{S}$ with $\mu(E)>0$ there is a subset F of E with $\mu(F)>0$ such that $\{\Phi(A)/\mu(A)\mid A\in\mathfrak{S}, A\subset F\}$ is relatively weakly compact. Then Φ is of bounded variation if and only if $x(s)$ is Bochner integrable.

On the other hand, since Birkhoff and Gel'fand it has been known that for special Banach spaces X every μ-absolutely continuous vector measure of bounded variation with values in X can be represented as a Bochner integral with respect to μ. Such spaces are said to have the **Radon-Nikodým property**. Separable dual spaces (Gel'fand, Pettis; Dunford and Pettis), reflexive spaces (Gel'fand, Pettis, Phillips), and $l_1(\Omega)$, Ω arbitrary, etc., have the Radon-Nikodým property, while $L_\infty(0,1)$ (Bochner), c_0 (J. A. Clarkson), $L_1(\Omega)$ on a nonatomic Ω (Clarkson, Gel'fand), and $C(\Omega)$ on an infinite compact Hausdorff space Ω, etc., do not. Gel'fand proved that $L_1(0,1)$ (and c_0) is not a dual by means of this fact. From 1967 to 1974, Riefell, H. B. Maynard, R. E. Huff, and W. J. Davis and R. P. Phelps succeeded in characterizing geometrically the Banach spaces with the Radon-Nikodým property. We know today that the following conditions for Banach spaces X are equivalent [10]: (i) X has the Radon-Nikodým property. (ii) Every separable closed linear subspace of X has the Radon-Nikodým property. (iii) Every function $f:[0,1]\to X$ of bounded variation is (strongly or weakly) differentiable a.e. (iv) For any finite measure space $(S,\mathfrak{S},\mu)$ and bounded linear operator $T:L_1(S)\to X$, there is an essentially bounded measurable function $x(s)$ with values in X such that

$$Tf=\int_S f(s)x(s)\,d\mu,\qquad f\in L_1(S).$$

(v) Each nonvoid bounded closed convex set K in X is the †closed convex hull of the set of its strongly exposed points, where a point $x_0\in K$ is called a **strongly exposed point** of K if there is an $x'\in X'$ such that $\langle x_0,x'\rangle>\langle x,x'\rangle$ for all $x\in K\setminus\{x_0\}$ and that any sequence $x_n\in K$ with $\lim\langle x_n,x'\rangle=\langle x_0,x'\rangle$ converges to x_0 strongly.

A Banach space X is said to have the **Kreĭn-Mil'man property** if each bounded closed convex set in X is the closed convex hull of its †extreme points. A Banach space X with the Radon-Nikodým property has the Kreĭn-Mil'man property (J. Lindenstrauss). If X is a dual space, then the converse holds (Huff and P. D. Morris). A Banach space with the Kreĭn-Mil'man property clearly has no closed linear space isomorphic to c_0, but there are Banach spaces that do not contain c_0 and do not have the Kreĭn-Mil'man property. The dual X of a Banach space Y has the Radon-Nikodým property if and only if the dual of every separable closed linear subspace of Y is separable (Uhl, C. Stegall).

I. Integrals of Multivalued Vector Functions

Let $\Gamma(s)$ be a multivalued function defined on a σ-finite complete measure space $(S,\mathfrak{S},\mu)$ with values that are nonempty closed subsets of a separable Banach space X. The inverse image of a subset E of X under $\Gamma(s)$ is, by definition, the set of all s such that $\Gamma(s)\cap E\neq\varnothing$. $\Gamma(s)$ is said to be **measurable** or **strongly measurable** if the inverse image of each open set in X under $\Gamma(s)$ belongs to $\mathfrak{S}$. Let $\mathfrak{B}(X)$ be the †Borel field of X, and $\mathfrak{S}\times\mathfrak{B}(X)$ be the product completely additive class, that is, the smallest completely additive class containing all direct products $A\times B$ of $A\in\mathfrak{S}$ and $B\in\mathfrak{B}(X)$. Then the measurability of $\Gamma(s)$ is equivalent to each of the following: (i) The graph $\{(s,x)\mid x\in\Gamma(s), s\in\mathfrak{S}\}$ of $\Gamma(s)$ belongs to $\mathfrak{S}\times\mathfrak{B}(X)$. (ii) The inverse image of every Borel set in X under $\Gamma(s)$ belongs to $\mathfrak{S}$. (iii) For each $x\in X$, the distance $d(x,\Gamma(s))=\inf\{\|x-y\|\mid y\in\Gamma(s)\}$ between x and $\Gamma(s)$ is measurable as a function on S.

A measurable function $x(s)$ on S with values in X is called a **measurable selection** of $\Gamma(s)$ if $x(s)$ is in $\Gamma(s)$ for all s. (X being separable, we need not discriminate between strong and weak measurability.) The measurability of $\Gamma(s)$ is also equivalent to the following important statement on the existence of measurable selections of $\Gamma(s)$: (iv) There are a countable number of measurable selections $\{x_n(s)\}$ of $\Gamma(s)$ such that the closure of the set $\{x_n(s)\mid n=1,2,\dots\}$ coincides with $\Gamma(s)$ for all $s\in S$. $\Gamma(s)$ is said to be **scalarly measurable** or **weakly measurable** if the support function $\delta'(x',\Gamma(s))=\sup\{\langle x,x'\rangle\mid x\in\Gamma(s)\}$ is measurable on S for all $x'\in X'$. The strong measurability of $\Gamma(s)$ clearly implies the weak one. If the values of $\Gamma(s)$ are nonempty weakly compact convex sets, then the measurabilities are equivalent. Hereafter we shall assume that $\Gamma(s)$ takes the values in the weakly compact convex sets. If the support function $\delta'(x',\Gamma(s))$ is integrable on S for all $x'\in X'$, then $\Gamma(s)$ is said to be **scalarly integrable**. Then the **scalar integral** of $\Gamma(s)$ is defined to be the set in X'' of all scalar integrals of its measurable selections, i.e.,

$$\int_S \Gamma(s)\,d\mu=\left\{\int_S x(s)\,d\mu\;\middle|\; x(s)\text{ is a measurable selection of }\Gamma(s)\right\}.$$

If $\|\Gamma(s)\| = \sup\{\|x\| \mid x \in \Gamma(s)\}$ is integrable, then every measurable selection is Bochner integrable and the integral $\int_S \Gamma(s)\,d\mu$ becomes a nonempty weakly compact convex set in X. When the values of $\Gamma(s)$ are nonempty compact convex sets, there is another method, by G. Debreu, of defining the integral. Let $\mathfrak{L}$ be the class of all nonempty compact convex sets in X and δ be the Hausdorff metric, i.e., for K_1 and $K_2 \in \mathfrak{L}$ define $\delta(K_1, K_2) = \max[\sup\{d(x, K_2) \mid x \in K_1\}, \sup\{d(x, K_1) \mid x \in K_2\}]$. Further, for $K_1, K_2 \in \mathfrak{L}$ and $\alpha \geqslant 0$ define the sum and the nonnegative scalar multiple by $K_1 + K_2 = \{x_1 + x_2 \mid x_1 \in K_1, x_2 \in K_2\}$ and $\alpha \cdot K_1 = \{\alpha x \mid x \in K_1\}$, respectively. Then $\mathfrak{L}$ endowed with the Hausdorff metric and the above addition and scalar multiplication is isometrically embedded in a closed convex cone in a separable Banach space Y by the Rådsröm embedding theorem (*Proc. Amer. Math. Soc.*, 3 (1952)). Let φ be this isometry. Then the **(strong) measurability** and the **(strong) integrability** of $\Gamma(s)$ are defined by the measurability and the Bochner integrability of the Y-valued function $\varphi(\Gamma(s))$, respectively, and its (strong) integral as the inverse image of the Bochner integral of $\varphi(\Gamma(s))$ under φ:

$$\int_S \Gamma(s)\,d\mu = \varphi^{-1}\left(\int_S \varphi(\Gamma(s))\,d\mu\right).$$

This definition of integral for strongly measurable $\Gamma(s)$ is shown to be compatible with that mentioned before. It is clear by the definition that the integral value in this case is a nonempty compact convex set and that most properties of Bochner integrals also hold for this integral.

References

[1] S. Bochner, Integration von Funktionen, deren Werte die Elemente eines Vektorraumes sind, Fund. Math., 20 (1933), 262–276.
[2] G. Birkhoff, Integration of functions with values in a Banach space, Trans. Amer. Math. Soc., 38 (1935), 357–378.
[3] I. Gel'fand, Abstrakte Funktionen und lineare Operatoren, Mat. Sb., 4 (46) (1938), 235–286.
[4] N. Dunford, Uniformity in linear spaces, Trans. Amer. Math. Soc., 44 (1938), 305–356.
[5] B. J. Pettis, On integration in vector spaces, Trans. Amer. Math. Soc., 44 (1938), 277–304.
[6] N. Bourbaki, Eléments de mathématique, Intégration, Hermann, ch. 6, 1959.
[7] R. G. Bartle, N. Dunford, and J. Schwartz, Weak compactness and vector measures, Canad. J. Math., 7 (1955), 289–305.
[8] N. Dunford and J. T. Schwartz, Linear operators I, Interscience, 1958.
[9] N. Dunford and B. J. Pettis, Linear operations on summable functions, Trans. Amer. Math. Soc., 47 (1940), 323–392.
[10] J. Diestel and J. J. Uhl, Jr., Vector measures, Amer. Math. Soc. Math. Surveys 13 (1977).
[11] R. J. Aumann, Integrals of set-valued functions, J. Math. Anal. Appl., 12 (1965), 1–22.
[12] G. Debreu, Integration of correspondences, Proc. Fifth Berkeley Symp. Math. Statist. Probab., II, pt. I (1967), 351–372.
[13] C. Castaing and M. Valadier, Convex analysis and measurable multifunctions, Springer, 1977.
[14] N. Dinculeanu, Vector measures, Pergamon, 1967.
[15] I. Kluvánek and G. Knowles, Vector measures and control systems, North-Holland, 1975.

444 (XXI.42)
Viète, François

François Viète (1540–December 13, 1603) was born in Fontenay-le-Comte, Poitou, in western France. He served under Henri IV, first as a lawyer and later as a political advisor. His mathematics was done in his leisure time. He used symbols for known variables for the first time and established the methodology and principles of symbolic algebra. He also systematized the algebra of the time and used it as a method of discovery. He is often called the father of algebra. He improved the methods of solving equations of the third and fourth degrees obtained by G. Cardano and L. Ferrari. Realizing that solving the algebraic equation of the 45th degree proposed by the Belgian mathematician A. van Roomen can be reduced to searching for $\sin(\alpha/45)$ knowing $\sin\alpha$, he was able to solve it almost immediately. However, he would not acknowledge negative roots and refused to add terms of different degrees because of his belief in the Greek principle of homogeneity of magnitudes. He also contributed to trigonometry and represented the number π as an infinite product.

References

[1] Francisci Vietae, Opera mathematica, F. van Schooten (ed.), Leyden, 1646 (Georg Olms, 1970).
[2] Jacob Klain, Die griechische Logistik und die Entstehung der Algebra I, II, Quellen und

Studien zur Gesch. Math., (B) 3 (1934), 18–105; (B) 3 (1936), 122–235.

445 (XXI.43) Von Neumann, John

John von Neumann (December 28, 1903–February 8, 1957) was born in Budapest, Hungary, the son of a banker. By the time he graduated from the university there in 1921, he had already published a paper with M. Fekete. He was later influenced by H. Weyl and E. Schmidt at the universities of Zürich and Berlin, respectively, and he became a lecturer at the universities of Berlin and Hamburg. He moved to the United States in 1930 and in 1933 became professor at the Institute for Advanced Study at Princeton. In 1954 he was appointed a member of the US Atomic Energy Commission. The fields in which he was first interested were †set theory, theory of †functions of real variables, and †foundations of mathematics. He made important contributions to the axiomatization of set theory. At the same time, however, he was deeply interested in theoretical physics, especially in the mathematical foundations of quantum mechanics. From this field, he was led into research on the theory of †Hilbert spaces, and he obtained basic results in the theory of †operator rings of Hilbert spaces. To extend the theory of operator rings, he introduced †continuous geometry. Among his many famous works are the theory of †almost periodic functions on a group and the solving of †Hilbert's fifth problem for compact groups. In his later years, he contributed to †game theory and to the design of computers, thus playing a major role in all fields of applied mathematics.

References

[1] J. von Neumann, Collected works I–VI, Pergamon, 1961–1963.
[2] J. von Neumann, 1903–1957, J. C. Oxtoby, B. J. Pettis, and G. B. Price (eds.), Bull. Amer. Math. Soc., 64 (1958), 1–129.
[3] J. von Neumann, Mathematische Grundlagen der Quantenmechanik, Springer, 1932.
[4] J. von Neumann, Functional operators I, II, Ann. Math. Studies, Princeton Univ. Press, 1950.
[5] J. von Neumann, Continuous geometry, Princeton Univ. Press, 1960.
[6] J. von Neumann and O. Morgenstern, Theory of games and economic behavior, Princeton Univ. Press, third edition, 1953.

W

446 (XX.13) Wave Propagation

A disturbance originating at a point in a medium and propagating at a finite speed in the medium is called a **wave**. For example, a sound wave propagates a change of density or stress in a gas, liquid, or solid. A wave in an elastic solid body is called an elastic wave. **Surface waves** appear near the surface of a medium, such as water or the earth. When electromagnetic disturbances are propagated in a gas, liquid, or solid or in a vacuum, they are called **electromagnetic waves**. Light is a kind of electromagnetic wave. According to †general relativity theory, gravitational action can also be propagated as a wave.

It many cases waves can be described by the **wave equation**:

$$\frac{\partial^2\psi}{\partial t^2}=c^2\left(\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+\frac{\partial^2\psi}{\partial z^2}\right).$$

Here t is time, x, y, z are the Cartesian coordinates of points in the space, c is the propagation velocity, and ψ represents the state of the medium.

If we take a closed surface surrounding the origin of the coordinate system, the state $\psi(0,t)$ at the origin at time t can be determined by the state at the points on the closed surface at time $t-r/c$, with r the distance of the point from the origin. More precisely, we have

$$\psi(0,t)=\frac{1}{4\pi}\int\left(\psi\frac{\partial}{\partial n}\left(\frac{1}{r}\right)-\frac{1}{r}\frac{\partial\psi}{\partial n}-\frac{1}{cr}\frac{\partial\psi}{\partial t}\frac{\partial r}{\partial n}\right)_{t-r/c}df.$$

Here n is the inward normal at any point of the closed surface, and the integral is taken over the surface, while the value of the integrand is taken at time $t-r/c$. This relation is a mathematical representation of **Huygens's principle**, which is valid for the 3-dimensional case but does not hold for the 2-dimensional case (→ 325 Partial Differential Equations of Hyperbolic Type).

A **plane wave** propagating in the direction of a unit vector $\mathbf{n}$ can be represented by $\psi=F(t-\mathbf{n}\cdot\mathbf{r}/c)$, where F is an arbitrary function and $\mathbf{r}(x,y,z)$ is the position vector. The simplest case is given by a **sine wave (sinusoidal wave)**: $\psi=A\sin(\omega t-\mathbf{k}\cdot\mathbf{r}+\delta)$. Here A **(amplitude)** and δ **(phase constant)** are arbitrary constants, $\mathbf{k}$ is in the direction of wave propagation and satisfies the relation $|\mathbf{k}|c=\omega$. ω is the **angular frequency**, $\omega/2\pi$ the **frequency**, $\mathbf{k}$ the **wave number vector**, $|\mathbf{k}|$ the **wave number**, $2\pi/\omega$ the **period**, and $2\pi/|\mathbf{k}|$ the **wavelength**. The velocity with which the crest of the wave advances is equal to $\omega/|\mathbf{k}|=c$ and is called the **phase velocity**.

A **spherical wave** radiating from the origin can generally be represented by

$$\psi=\sum_n\varphi_n\left(\frac{d}{rdr}\right)^n\frac{1}{r}F\left(t-\frac{r}{c}\right),$$

where φ_n is the †solid harmonic of order n.

Waves are not restricted to those governed by the wave equation. In general, ψ is not a scalar, but has several components (e.g., ψ may be a vector), which satisfy a set of simultaneous differential equations of various kinds. Usually they have solutions in the form of sinusoidal waves, but the phase velocity $c=\omega/|\mathbf{k}|$ is generally a function of the wavelength λ. Such a wave, called a **dispersive wave**, has a propagation velocity (velocity of propagation of the disturbance through the medium) that is not equal to the phase velocity. A disturbance of finite extent that can be approximately represented by a plane wave is propagated with a velocity $c-\lambda dc/d\lambda$, called the **group velocity**. Often there exists a definite relationship between the amplitude vector $\mathbf{A}$ (and the corresponding phase constant δ) and wave number vector $\mathbf{k}$, in which case the wave is said to be **polarized**. In particular, when $\mathbf{A}$ and $\mathbf{k}$ are parallel (perpendicular), the wave is called a **longitudinal (transverse) wave**. Usually equations governing the wave are linear, and therefore superposition of two solutions gives a new solution (†principle of superposition). Superposition of two sinusoidal waves traveling in opposite directions gives rise to a wave whose crests do not move (e.g., $\psi=A\sin\omega t\sin\mathbf{k}\cdot\mathbf{r}$). Such a wave is called a **stationary wave**. Since the energy of a wave is proportional to the square of ψ, the energy of the resultant wave formed by superposition of two waves is not equal to the sum of the energies of the component waves. This phenomenon is called **interference**. When a wave reaches an obstacle it propagates into the shadow region of the obstacle, where there is formed a special distribution of energy dependent on the shape and size of the obtacle. This phenomenon is called **diffraction**.

For aerial sound waves and water waves, if the amplitude is so large that the wave equation is no longer valid, we are faced with †nonlinear problems. For instance, **shock waves** appear in the air when surfaces of discontinuity of density and pressure exist. They appear in explosions and for bodies traveling at high speeds. Concerning wave mechanics dealing with atomic phenomena → 351 Quantum Mechanics.

References

[1] H. Lamb, Hydrodynamics, Cambridge Univ. Press, sixth edition, 1932.
[2] Lord Rayleigh, The theory of sound, Macmillan, second revised edition, I, 1937; II, 1929.
[3] M. Born and E. Wolf, Principles of optics, Pergamon, fourth edition, 1970.
[4] F. S. Crawford, Jr., Waves, Berkeley phys. course III, McGraw-Hill, 1968.
[5] C. A. Coulson, Waves; A mathematical theory of the common type of wave motion, Oliver & Boyd, seventh edition, 1955.
[6] L. Brillouin, Wave propagation and group velocity, Academic Press, 1960.
[7] I. Tolstoy, Wave propagation, McGraw-Hill, 1973.
[8] J. D. Achenbach, Wave propagation in elastic solids, North-Holland, 1973.
[9] K. F. Graff, Wave motion in elastic solids, Ohio State Univ. Press, 1975.
[10] J. Lighthill, Waves in fluids, Cambridge Univ. Press, 1978.
[11] R. Courant and D. Hilbert, Methods of mathematical physics II, Interscience, 1962.

447 (XXI.44) Weierstrass, Karl

Karl Weierstrass (October 31, 1815–February 19, 1897) was born into a Catholic family in Ostenfelde, in Westfalen, Germany. From 1834 to 1838 he studied law at the University of Bonn. In 1839 he moved to Münster, where he came under the influence of C. Gudermann, who was then studying the theory of elliptic functions. From this time until 1855, he taught in a parochial junior high school; during this period he published an important paper on the theory of analytic functions. Invited to the University of Berlin in 1856, he worked there with L. Kronecker and E. E. Kummer. In 1864, he was appointed to a full professorship, which he held until his death.

His foundation of the theory of analytic functions of a complex variable at about the same time as Riemann is his most fundamental work. In contrast to Riemann, who utilized geometric and physical intuition, Weierstrass stressed the importance of rigorous analytic formulation. Aside from the theory of analytic functions, he contributed to the theory of functions of real variables by giving examples of continuous functions that were nowhere differentiable. With his theory of †minimal surfaces, he also contributed to geometry. His lectures at the University of Berlin drew many listeners, and in his later years he was a respected authority in the mathematical world.

References

[1] K. Weierstrass, Mathematische Werke I–VII, Mayer & Miller, 1894–1927.
[2] F. Klein, Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert I, Springer, 1926 (Chelsea, 1956).

448 (XXI.45) Weyl, Hermann

Hermann Weyl (November 9, 1885–December 8, 1955) was born in Elmshorn in the state of Schleswig-Holstein in Germany. Entering the University of Göttingen in 1904, he also audited courses for a time at the University of Münich. In 1908, he obtained his doctorate from the University of Göttingen with a paper on the theory of integral equations, and by 1910 he was a lecturer at the same university. In 1913, he became a professor at the Federal Technological Institute at Zürich; in 1928–1929, a visiting professor at Princeton University; in 1930, a professor at the University of Göttingen; and in 1933, a professor at the Institute for Advanced Study at Princeton. He retired from his professorship there in 1951, when he became professor emeritus. He died in Zürich in 1955.

Weyl contributed fresh and fundamental works covering all aspects of mathematics and theoretical physics. Among the most notable are results on problems in †integral equations, †Riemann surfaces, the theory of †Diophantine approximation, the representation of groups, in particular compact groups and †semisimple Lie groups (whose structure he elucidated), the space-time problem, the introduction of †affine connections in differential geometry, †quantum mechanics, and the foundations of mathematics. In his later years, with his son Joachim he studied meromorphic functions. In addition to his many mathematical works he left works in philosophy, history, and criticism.

References

[1] H. Weyl, Gesammelte Abhandlungen I–IV, Springer, 1968.
[2] H. Weyl, Die Idee der Riemannschen Fläche, Teubner, 1913, revised edition, 1955; English translation, The concept of a Riemann surface, Addison-Wesley, 1964.

[3] H. Weyl, Raum, Zeit, Materie, Springer, 1918, fifth edition, 1923; English translation, Space, time, matter, Dover, 1952.
[4] H. Weyl, Das Kontinuum, Veit, 1918.
[5] H. Weyl, Gruppentheorie und Quantenmechanik, Hirzel, 1928.
[6] H. Weyl, Classical groups, Princeton Univ. Press, 1939, revised edition, 1946.
[7] H. Weyl and F. J. Weyl, Meromorphic functions and analytic curves, Princeton Univ. Press, 1943.
[8] H. Weyl, Philosophie der Mathematik und Naturwissenschaften, Oldenbourg, 1926; English translation, Philosophy of mathematics and natural science, Princeton Univ. Press, 1949.
[9] H. Weyl, Symmetry, Princeton Univ. Press, 1952.

449 (III.18) Witt Vectors

A. General Remarks

Let Γ be an †integral domain of characteristic 0, and p a fixed prime number. For each infinite-dimensional vector $x=(x_0, x_1, \dots)$ with components in Γ, we define its **ghost components** $x^{(0)}, x^{(1)}, \dots$ by $x^{(0)}=x_0, x^{(n)}=x_0^{p^n}+px_1^{p^{n-1}}+\dots+p^n x_n$. We define the sum of the vectors x and $y=(y_0, y_1, \dots)$ to be the vector with ghost components $x^{(0)}+y^{(0)}, x^{(1)}+y^{(1)}, \dots$, and their product to be the vector with ghost components $x^{(0)}y^{(0)}, x^{(1)}y^{(1)}, \dots$. The sum and product are uniquely determined vectors with components in Γ. Writing their first two terms explicitly, we have

$$x+y$$
$$=\left[x_0+y_0, x_1+y_1-\sum_{\nu=1}^{p-1}\frac{1}{p}\binom{p}{\nu}x_0^{\nu}y_0^{p-\nu}, \dots\right],$$

$$xy=(x_0y_0, x_1y_0^p+y_1x_0^p+px_1y_1, \dots).$$

In general, it can be proved that the nth components $\sigma_n(x,y)$ and $\pi_n(x,y)$ of the sum and product are polynomials in $x_0, y_0, x_1, y_1, \dots, x_n, y_n$ whose coefficients are rational integers. With these operations of addition and multiplication, the set of these vectors forms a †commutative ring, of which the zero element is $(0,0,\dots)$ and the unity element is $(1,0,\dots)$. Let k be a field of characteristic p. For vectors $(\xi_0, \xi_1, \dots)$, and $(\eta_0, \eta_1, \dots)$ with components in k, we define their sum and product by $(\xi_0, \xi_1, \dots)+(\eta_0, \eta_1, \dots)=(\dots, \sigma_n(\xi, \eta), \dots)$ and $(\xi_0, \xi_1, \dots)(\eta_0, \eta_1, \dots)=(\dots, \pi_n(\xi, \eta), \dots)$. Since the coefficients of σ_n and π_n are rational integers, these operations are well defined. With these operations, the set of such vectors becomes an integral domain $W(k)$ of characteristic 0. Elements of $W(k)$ are called **Witt vectors** over k.

If we put $V(\xi_0, \xi_1, \dots)=(0, \xi_0, \xi_1, \dots)$ and $(\xi_0, \xi_1, \dots)^p=(\xi_0^p, \xi_1^p, \dots)$, we get the formula $p\xi=V\xi^p$. (Note that this ξ^p is not the pth power of ξ in $W(k)$ in the usual sense.) Therefore, if we put $|\xi|=p^{-n}$ for a vector ξ whose first nonzero component is ξ_n, then this absolute value $|\ |$ gives a †valuation of $W(k)$. In particular, when k is a †perfect field, denoting the vector $(\xi_0, 0, \dots)$ by $\{\xi_0\}$ we get $(\xi_0, \xi_1, \dots)=\sum p^i\{\xi_i^{p^{-i}}\}$, and $W(k)$ is a †complete valuation ring with respect to this valuation. Therefore the †field of quotients of $W(k)$ is a complete valuation field of which p is a prime element and k is the †residue class field. Conversely, let K be a field of characteristic 0 that is complete under a †discrete valuation v, $\mathfrak{o}$ be the valuation ring of v, and k be the residue class field of v. Assume that k is a perfect field of characteristic p. If p is a prime element of v, then $\mathfrak{o}=W(k)$. If $v(p)=v(\pi^e)$ $(e>1)$ with a prime element π of $\mathfrak{o}$, we have $\mathfrak{o}=W(k)[\pi]$, and π is a root of an †Eisenstein polynomial $X^e+a_1X^{e-1}+\dots+a_e$ $(a_i \in pW(k), a_e \notin p^2W(k))$. In this way we can determine explicitly the structure of a †p-adic number field (→ 257 Local Fields).

B. Applications to Abelian p-Extensions and Cyclic Algebras of Characteristic p

Next we consider $W_n(k)=W(k)/V^nW(k)$. The elements of $W_n(k)$ can be viewed as the n-dimensional vectors $(\xi_0, \dots, \xi_{n-1})$, but their laws of composition are defined as in the previous section. They are called **Witt vectors of length** n. We define an operator $\wp$ by $\wp\xi=\xi^p-\xi$. Using it, we can generalize the theory of †Artin-Schreier extensions (→ 172 Galois Theory) to the case of Abelian extensions of exponent p^n over a field of characteristic p. Indeed, let k be a field of characteristic p and $\xi=(\xi_0, \dots, \xi_{n-1})$ an element of $W(k)$. If $\eta=(\eta_0, \dots, \eta_{n-1})$ is a root of the vector equation $\wp X-\xi=0$, then the other roots are of the form $\eta+\alpha(\alpha=(\alpha_0, \dots, \alpha_{n-1}), \alpha_i \in \mathbf{F}_p)$. In particular, if $\xi_0 \notin \wp k=\{\alpha^p-\alpha \mid \alpha \in k\}$, the field $K=k(\eta_0, \dots, \eta_{n-1})$ is a cyclic extension of degree p^n over k, and conversely, every cyclic extension of k of degree p^n is obtained in this way. Let $(1/\wp)\xi$ denote the set of all roots of $\wp X-\xi=0$. Then more generally, any finite Abelian extension of exponent p^n of k can be obtained as $K=k((1/\wp)\xi \mid \xi \in H)$ with a suitable finite subgroup $H/\wp W_n(k)$ of $W_n(k)/\wp W_n(k)$, and

the Galois group of K/k is isomorphic to $H/\wp W_n(k)$.

Moreover, for a †cyclic extension $K = k((1/\wp)\beta)$ of exponent p^n over k and for $\alpha \in k(\alpha \neq 0)$, we can define a †cyclic algebra $(\alpha, \beta]$ generated by an element u over K by the fundamental relations $u^{p^n} = \alpha$, $\wp\theta = \beta$, $u\theta u^{-1} = \theta + (1, 0, \ldots, 0)$ (where $\theta = (\theta_0, \ldots, \theta_{n-1})$, $u\theta u^{-1} = (u\theta_0 u^{-1}, \ldots, u\theta_{n-1} u^{-1})$), and $(\alpha, \beta]$ is a central simple algebra over k.

Using these results, we can develop the structure theory of the †Brauer group of exponent p^n of a †field of power series in one variable with coefficients in a finite field $\mathbf{F}_q$ (of a †field of algebraic functions in one variable over $\mathbf{F}_q$) exactly as in the case of a p-adic field (of an algebraic number field) (E. Witt [1]; → 29 Associative Algebras G).

On the other hand, $W_n(k)$ is a commutative †algebraic group over k and is important in the theories of algebraic groups and †formal groups (→ 13 Algebraic Groups).

References

[1] E. Witt, Zyklische Körper und Algebren der Charakteristik p vom Grad p^n, J. Reine Angew. Math., 176 (1937), 126–140.

[2] H. Hasse, Zahlentheorie, Akademie-Verlag., 1949.

[3] N. Jacobson, Lectures in abstract algebra III, Van Nostrand, 1964.

N

450 (V.19) Zeta Functions

A. Introduction

Since the 19th century, many special functions called ζ-functions (zeta functions) have been defined and investigated. The four main problems concerning ζ-functions are: (1) Methods of defining ζ-functions. (2) Investigation of the properties of ζ-functions. Generally, ζ-functions have the following four properties in common: (i) They are meromorphic on the whole complex plane; (ii) they have †Dirichlet series expansions; (iii) they have Euler product expansions; and (iv) they satisfy certain functional equations. Also, it is an important problem to find the poles, residues, and zeros of ζ-functions. (3) Application to number theory, in particular to the theory of decomposition of prime ideals in finite extensions of algebraic number fields (→ 59 Class Field Theory). (4) Study of the relations between different ζ-functions.

Most of the functions called ζ-functions or L-functions have the four properties of problem (2). The following is a classification of the important types of ζ-functions that are already known, which will be discussed later in this article:

(1) The ζ- and L-functions of algebraic number fields: the Riemann ζ-function, Dirichlet L-functions (study of these functions gave impetus to the theory of ζ-functions), Dedekind ζ-functions, Hecke L-functions, Hecke L-functions with †Grössencharakters, Artin L-functions, and Weil L-functions. (2) The p-adic L-functions related to the works of H. W. Leopoldt, T. Kubota, K. Iwasawa, etc. (3) The ζ-functions of quadratic forms: Epstein ζ-functions, ζ-functions of indefinite quadratic forms (C. L. Siegel), etc. (4) The ζ- and L-functions of algebras: Hey ζ-functions and the ζ-functions given by R. Godement, T. Tamagawa, etc. (5) The ζ-functions associated with Hecke operators, related to the work of E. Hecke, M. Eichler, G. Shimura, H. Jacquet, R. P. Langlands, etc. (6) The congruence ζ- and L-functions attached to algebraic varieties defined over finite fields (E. Artin, A. Weil, A. Grothendieck, P. Deligne), ζ- and L-functions of schemes. (7) Hasse ζ-functions attached to the algebraic varieties defined over algebraic number fields. (8) The ζ-functions attached to discontinuous groups: Selberg ζ-functions, the Eisenstein series defined by A. Selberg, Godement, and I. M. Gel'fand, etc. (9) Y. Ihara's ζ-function related to non-Abelian class field theory over a function field over a finite field. (10) ζ-functions associated with prehomogeneous vector spaces (M. Sato, T. Shintani).

B. The Riemann ζ-Function

Consider the series

$$\zeta(s)=1+\frac{1}{2^s}+\frac{1}{3^s}+\ldots+\frac{1}{n^s}+\ldots,$$

which converges for all real numbers $s>1$. It was already recognized by L. Euler that $\zeta(s)$ can also be expressed by a convergent infinite product $\prod_p(1-p^{-s})^{-1}$, where p runs over all prime numbers (*Werke*, ser. I, vol. VII, ch. XV, § 274). This expansion is called **Euler's infinite product expansion** or simply the **Euler product**. However, Riemann was the first to treat $\zeta(s)$ successfully as a function of a complex variable s (1859) [R1]; for this reason, it is called the **Riemann ζ-function**. As can be seen from its Euler product expansion, $\zeta(s)$ is holomorphic and has no zeros in the domain $\mathrm{Re}\, s>1$. Riemann proved, moreover, that it has an analytic continuation to the whole complex plane, is meromorphic everywhere, and has a unique pole $s=1$. The functions $(s-1)\zeta(s)$ and $\zeta(s)-1/(s-1)$ are †integral functions of s. This can be seen by considering the integral expression

$$\pi^{-s/2}\Gamma\left(\frac{s}{2}\right)\zeta(s)=\int_0^\infty x^{s/2-1}\left(\sum_{n=1}^\infty e^{-n^2\pi x}\right)dx$$

$$=-\frac{1}{s}-\frac{1}{1-s}+\int_1^\infty (x^{(1-s)/2-1}+x^{(s/2)-1})\times\left(\sum_{n=1}^\infty e^{-n^2\pi x}\right)dx.$$

From this last formula, we also obtain an equality

$$\xi(s)=\xi(1-s),$$

where

$$\xi(s)=\pi^{-s/2}\Gamma(s/2)\zeta(s).$$

This equality is called the **functional equation** for the ζ-function. The residue of $\zeta(s)$ at $s=1$ is 1, and around $s=1$,

$$\zeta(s)=\frac{1}{s-1}+C+O(|s-1|),$$

where C is †Euler's constant. This is called the **Kronecker limit formula** for $\zeta(s)$.

The function $\zeta(s)$ has no zeros in $\mathrm{Re}\, s\geqslant 1$, and its only zeros in $\mathrm{Re}\, s\leqslant 0$ are simple zeros at $s=-2,\ -4,\ldots,\ -2n,\ldots$. But $\zeta(s)$ has infinitely many zeros in $0<\mathrm{Re}\, s<1$, which are called the nontrivial zeros. B. Riemann conjec-

tured that all nontrivial zeros lie on the line $\mathrm{Re}\, s = 1/2$ (1859). This is called the **Riemann hypothesis**, which has been neither proved nor disproved (→ Section I).

If $N(T)$ denotes the number of zeros of $\zeta(s)$ in the rectangle $0 < \mathrm{Re}\, s < 1$, $0 < \mathrm{Im}\, s < T$, we have an asymptotic formula

$$N(T) = \frac{1}{2\pi} T \log T - \frac{1 + \log 2\pi}{2\pi} T + O(\log T)$$

(H. von Mangoldt, 1905). Also, $\zeta(s)$ has the following infinite product expansion:

$$(s-1)\zeta(s) = \frac{1}{2} e^{bs} \frac{1}{\Gamma\left(\frac{s}{2}+1\right)} \prod_{\rho} \left(1 - \frac{s}{\rho}\right) e^{s/\rho},$$

where b is a constant and ρ runs over all nontrivial zeros of $\zeta(s)$ (J. Hadamard, 1893). Hadamard and C. de La Vallée-Poussin proved the †prime number theorem, almost simultaneously, by using some properties of $\zeta(s)$ (→ 123 Distribution of Prime Numbers B).

The following **approximate functional equation** is important in investigating the values of $\zeta(s)$:

$$\zeta(s) = \sum_{n<x} \frac{1}{n^s} + \varphi(s) \sum_{n<y} \frac{1}{n^{1-s}} + O(x^{-\sigma}) + O(y^{\sigma-1} |t|^{(1/2)-\sigma}),$$

where φ is †Euler's function, and $\zeta(s) = \varphi(s)\zeta(1-s)$, $s = \sigma + it$, $2\pi xy = |t|$, and the approximation is uniform for $-h \leqslant \sigma \leqslant h$, $x > k$, $y > k$ with h and k positive constants (G. H. Hardy and J. E. Littlewood, 1921).

Euler obtained the values of $\zeta(s)$ for positive even integers s:

$$\zeta(2m) = \frac{2^{2m-1} \pi^{2m} B_{2m}}{(2m)!}$$

($m = 1, 2, 3, \ldots$, and the B_{2m} are †Bernoulli numbers). The values of $\zeta(s)$ for positive odd integers s, however, have not been expressed in such a simple form. The values of $\zeta(s)$ for negative integers s are given by $\zeta(0) = B_1(0) = -\frac{1}{2}$, $\zeta(1-n) = -\frac{B_n(0)}{n}$, $n = 2, 3, \ldots$, where the $B_n(x)$ are †Bernoulli polynomials.

As a slight generalization of $\zeta(s)$, A. Hurwitz (1862) considered

$$\zeta(s, a) = \sum_{n=0}^{\infty} \frac{1}{(n+a)^s}, \quad 0 < a \leqslant 1.$$

This is called the **Hurwitz ζ-function**. Thus $\zeta(s, 1) = \zeta(s)$, and $\zeta(s, 1/2) = (2^s - 1)\zeta(s)$. This function $\zeta(s, a)$ can also be continued analytically to the whole complex plane and satisfies a certain functional equation. But in general it has no Euler product expansion.

C. Dirichlet L-Functions

Let m be a positive integer, and classify all rational integers modulo m. The set of all classes coprime to m forms a multiplicative Abelian group of order $h = \varphi(m)$. Let χ be a †character of this group. Call (n) the residue class of $n \bmod m$, and put $\chi(n) = \chi((n))$ when $(n, m) = 1$ and $\chi(n) = 0$ when $(n, m) \neq 1$. Now, the function of a complex variable s defined by

$$L(s) = L(s, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}$$

is called a **Dirichlet L-function**. This function converges absolutely for $\mathrm{Re}\, s > 1$ and has an Euler product expansion

$$L(s, \chi) = \prod_{p} \frac{1}{1 - \chi(p) p^{-s}}.$$

If there exist a divisor f of m ($f \neq m$) and a character χ^0 modulo f such that $\chi(n) = \chi^0(n)$ for all n with $(n, m) = 1$, we call χ a **nonprimitive character**. Otherwise, χ is called a **primitive character**. If χ is nonprimitive, there exists such a unique primitive χ^0. In this situation, the divisor f of m associated with χ^0 is called the **conductor** of χ (and of χ^0). We have

$$L(s, \chi) = L(s, \chi^0) \prod_{p|m} (1 - \chi^0(p) p^{-s}).$$

Let χ be primitive. If the conductor $f = 1$, then χ is a trivial character ($\chi = 1$), and $L(s)$ is equal to the Riemann ζ-function $\zeta(s)$. On the other hand, if $f > 1$, then $L(s)$ is an entire function of s. In particular, if χ is a nontrivial primitive character, $L(1) = L(1, \chi)$ is finite and nonzero. P. G. L. Dirichlet proved the theorem of existence of prime numbers in arithmetic progressions using this fact (→ 123 Distribution of Prime Numbers D).

$L(s, \chi)$ has a functional equation similar to that of $\zeta(s)$; namely, if χ is a primitive character with conductor f and we put

$$\xi(s, \chi) = (f/\pi)^{s/2} \Gamma((s+a)/2) L(s, \chi),$$

where $a = 0$ for $\chi(-1) = 1$ and $a = 1$ for $\chi(-1) = -1$, then we have

$$\xi(s, \chi) = W(\chi) \xi(1-s, \bar{\chi}),$$

where

$$W(\chi) = (-i)^a f^{-1/2} \tau(\chi), \quad \tau(\chi) = \sum_{n \bmod f} \chi(n) \zeta_f^n$$

($\zeta_f = \exp(2\pi i/f)$). The latter sum is called the **Gaussian sum**. Note that $|W(\chi)| = 1$.

The values of $L(s)$ for negative integers s are given by $L(1-m, \chi) = -B_{\chi,m}/m$ ($m = 1, 2, 3, \ldots$), where the $B_{\chi,m}$ are defined by

$$\sum_{\mu=1}^{f} \frac{\chi(\mu) t e^{\mu t}}{e^{ft} - 1} = \sum_{m=0}^{\infty} B_{\chi,m} t^m.$$

Moreover, if $\chi(-1)=-1$, we have

$$L(1,\chi)=\frac{\pi}{i}\frac{\tau(\chi)}{f}\sum_{x=1}^{f}(-\overline{\chi(x)}\cdot x)$$

$$=\pi i\tau(\chi)B_{\bar{\chi},1},$$

and if $\chi(-1)=1$, $\chi\neq 1$, we have

$$L(1,\chi)=2\frac{\tau(\chi)}{f}\sum_{x=1}^{[f/2]}(-\chi(x)\log|1-\zeta_f^x|).$$

In certain cases, the functional equation can be utilized to obtain the values of $L(m,\chi)$ from those of $L(1-m,\chi)$. Actually, if $\chi(-1)=1$, $m=2n=2,4,6,\ldots$, we have

$$L(2n,\chi)=\frac{(-1)^n}{(2n)!}\left(\frac{2\pi}{f}\right)^{2n}\tau(\chi)(-\mathrm{B}_{\chi,2n}),$$

and if $\chi(-1)=-1$, $m=2n+1=3,5,7,\ldots$, we have

$$L(2n+1,\chi)$$

$$=(-i)\frac{(-1)^n}{(2n+1)!}\left(\frac{2\pi}{f}\right)^{2n+1}\tau(\chi)(-B_{\chi,2n+1}).$$

Dirichlet L-functions are important not only in the arithmetic of rational number fields but also in the arithmetic of quadratic or cyclotomic fields.

D. ζ-Functions of Algebraic Number Fields (Dedekind ζ-Functions)

The Riemann ζ-function can be generalized to ζ-functions of algebraic number fields (→ 14 Algebraic Number Fields). Let k be an algebraic number field of degree n, and let $\mathfrak{a}$ run over all integral ideals of k. Consider the sequence $\zeta_k(s)=\sum_{\mathfrak{a}}N(\mathfrak{a})^{-s}$. This sequence converges for $\mathrm{Re}\,s>1$ and has an Euler product expansion $\zeta_k(s)=\prod_{\mathfrak{p}}(1-N(\mathfrak{p})^{-s})^{-1}$, where $\mathfrak{p}$ runs over all prime ideals of k. This function, which is continued analytically to the whole complex plane as a meromorphic function, is called a **Dedekind ζ-function**. Its only pole is a simple pole at $s=1$, with the residue $h_k\kappa_k$. Here h_k is the †class number of k, and $\kappa_k=2^{r_1+r_2}\pi^{r_2}R/(w|d|^{1/2})$, where r_1 $(2r_2)$ is the number of isomorphisms of k into the real (complex) number field, w is the number of roots of unity in k, d is the †discriminant of k, and R is the †regulator of k (R. Dedekind, 1877) [D1].

The function $\zeta_k(s)$ has no zeros in $\mathrm{Re}\,s\geqslant 1$, while in $\mathrm{Re}\,s\leqslant 0$ it has zeros of order r_2 at -1, -3, $-5,\ldots$, zeros of order r_1+r_2 at -2, -4, $-6,\ldots$, and a zero of order r_1+r_2-1 at $s=0$. All other zeros lie in the open strip $0<\mathrm{Re}\,s<1$, which actually contains infinitely many zeros. It is conjectured that all these zeros lie on the line $\mathrm{Re}\,s=1/2$ (the Riemann hypothesis for Dedekind ζ-functions). To obtain a generalization of the functional equation for the Riemann $\zeta(s)$ to the case of $\zeta_k(s)$, we put

$$\Xi_k(s)=\left(\frac{\sqrt{|d|}}{2^{r_2}\pi^{n/2}}\right)^s\Gamma\left(\frac{s}{2}\right)^{r_1}\Gamma(s)^{r_2}\zeta_k(s).$$

Then $\Xi_k(s)=\Xi_k(1-s)$ (Hecke, 1917). If K is a Galois extension of k, then $\zeta_K(s)/\zeta_k(s)$ is an integral function (H. Aramata, 1933; R. Brauer, 1947).

E. Hecke L-Functions

As a generalization of Dirichlet L-functions to algebraic number fields, Hecke (1917) defined the following L-function $L_k(s,\chi)$: Let k be an algebraic number field of finite degree, and let $\tilde{\mathfrak{m}}=\mathfrak{m}\prod\mathfrak{p}_\infty$ be an †integral divisor ($\mathfrak{m}$ the finite part, $\prod\mathfrak{p}_\infty$ the †infinite part). Consider the †ideal class group of k modulo $\tilde{\mathfrak{m}}$ and its character χ (here we put $\chi(\mathfrak{a})=0$ for $(\mathfrak{a},\mathfrak{m})\neq 1$). Then the L-functions are defined by

$$L_k(s,\chi)=\sum_{\mathfrak{a}}\chi(\mathfrak{a})/N(\mathfrak{a})^s$$

[H2], where $\mathfrak{a}$ runs over all integral ideals of k. $L_k(s,\chi)$ is called a **Hecke L-function**. It converges for $\mathrm{Re}\,s>1$ and has an Euler product expansion

$$L_k(s,\chi)=\prod_{\mathfrak{p}}\frac{1}{1-\chi(\mathfrak{p})N(\mathfrak{p})^{-s}}.$$

Here $\mathfrak{p}$ runs over all prime ideals of k. If there is a divisor $\tilde{\mathfrak{f}}\,|\,\tilde{\mathfrak{m}}$ $(\tilde{\mathfrak{f}}\neq\tilde{\mathfrak{m}})$ and a character χ^0 modulo $\tilde{\mathfrak{f}}$ such that $\chi^0(\mathfrak{a})=\chi(\mathfrak{a})$ for all $\mathfrak{a}$ with $(\mathfrak{a},\mathfrak{m})=1$, then χ is called **nonprimitive**: otherwise, χ is called a **primitive character**. In general, there exist unique such $\tilde{\mathfrak{f}}$ and χ^0. In this situation, $\tilde{\mathfrak{f}}$ is called the **conductor** of χ. If χ is primitive and the conductor $\tilde{\mathfrak{f}}$ is (1), then χ is a trivial character and $L_k(s,\chi)$ coincides with $\zeta_k(s)$. If χ is primitive and $\chi\neq 1$, then $L_k(s,\chi)$ is an integral function of s, and $L_k(1,\chi)\neq 0$. Utilizing this fact, it can be proved that there exist infinitely many prime ideals in each class of the ideal class group modulo an integral divisor $\tilde{\mathfrak{m}}$ of k.

Let χ be a primitive character with the conductor $\tilde{\mathfrak{f}}$, d be the discriminant of k, $\sigma_1,\ldots,\sigma_{r_1}$ be all distinct isomorphisms of k into the real number field $\mathbf{R}$, and $\mathfrak{f}$ be the finite part of $\tilde{\mathfrak{f}}$. Then if ξ is an integer of k such that $\xi\equiv 1$ (mod $\mathfrak{f}$), we have

$$\chi((\xi))=(\mathrm{sgn}\,\xi^{\sigma_1})^{a_1}\cdot\ldots\cdot(\mathrm{sgn}\,\xi^{\sigma_{r_1}})^{a_{r_1}},$$

where a_m $(m=1,\ldots,r_1)$ is either 0 or 1, depending on χ. By putting

$$\xi_k(s,\chi)=$$

$$\left(\frac{\sqrt{|d|N(\mathfrak{f})}}{2^{r_2}\pi^{n/2}}\right)^s\cdot\prod_{m=1}^{r_1}\Gamma\left(\frac{s+a_m}{2}\right)\Gamma(s)^{r_2}\cdot L_k(s,\chi),$$

we have the following functional equation for the Hecke L-function:

$$\xi_k(s,\chi)=W(\chi)\xi_k(1-s,\bar{\chi}),$$

where $W(\chi)$ is a complex number with absolute value 1 and the exact value of $W(\chi)$ is given as a Gaussian sum. Just as some properties concerning the distribution of prime numbers can be proved using the Riemann ζ-function and Dirichlet L-functions, some properties concerning the distribution of prime ideals can be proved using the Hecke L-functions (→ 123 Distribution of Prime Numbers F).

T. Takagi used Hecke L-functions in founding his †class field theory. In the other direction, this theory implies $L(1,\chi)\neq 0$ $(\chi\neq 1)$.

Let K be a †class field over k that corresponds to an ideal class group H of k with index h. By using class field theory, we obtain $\zeta_K(s)=\prod_\chi L_k(s,\chi)$, where the product is over all characters χ of ideal class groups of k, such that $\chi(H)=1$. This formula can be regarded as an alternative formulation of the decomposition theorem of class field theory (→ 59 Class Field Theory). By taking the residues of both sides of the formula at $s=1$, we obtain $h_K\kappa_K=h_k\kappa_k\prod_{\chi\neq 1}L_k(1,\chi)$.

In particular, if $k=\mathbf{Q}$ (the rational number field) and K is a quadratic number field $\mathbf{Q}(\sqrt{d})$ (d is the discriminant of K), then we have

$$\zeta_K(s)=\zeta(s)\cdot L(s),\quad L(s)=\sum_{n=1}^{\infty}\left(\frac{d}{n}\right)n^{-s},$$

where (d/n) is the †Kronecker symbol, and we put $(d/n)=0$ when $(n,d)\neq 1$. From this, we obtain the class number formula for quadratic number fields (→ 347 Quadratic Fields). A similar method is used for computation of class numbers of cyclotomic fields K (→ 14 Algebraic Number Fields L).

In general, the computation of the relative class number h_K/h_k when K/k is an Abelian extension is reduced to the evaluation of $L(1,\chi)$. This computation has been made successfully for the following cases (besides for the examples in the previous paragraph): k is imaginary quadratic and K is the absolute class field of k or the class field corresponding to †ray $S(\mathfrak{m})$; k is totally real and K is a totally imaginary quadratic extension of k. H. M. Stark and T. Shintani made conjectures about the values of $L(1,\chi)$ [S25, S19].

Let $L(s,\chi)$ be a Hecke L-function for the character χ. Then it follows from the functional equation that the values of $L(s,\chi)$ at $s=0,-1,-2,-3,\ldots$ are zero if k is not totally real. Furthermore, if k is a totally real finite algebraic number field, then these values of $L(s,\chi)$ are algebraic numbers (C. L. Siegel, H. Klingen, T. Shintani).

F. Hecke L-Functions with Grössencharakters

E. Hecke (1918, 1920) extended the notion of characters by introducing the †Grössencharakter χ and defined L-functions with such characters:

$$L_k(s,\chi)=\sum_{\mathfrak{a}}\frac{\chi(\mathfrak{a})}{N(\mathfrak{a})^s}.$$

He also proved the existence of their Euler product expansions and showed that they satisfy certain functional equations [H3]. Moreover, by estimating the sum $\sum_{N(\mathfrak{p})<x}\chi(\mathfrak{p})$, he obtained some results on the distribution of prime ideals.

Later, Iwasawa and J. Tate independently gave clearer definitions of the Grössencharakter χ and $L_k(s,\chi)$ by using harmonic analysis on the adele and idele groups of k (→ 6 Adeles and Ideles) [L3].

Let $\mathbf{J}_k$ be the idele group of k, $\mathbf{P}_k$ be the group of †principal ideles, and $\mathbf{C}_k=\mathbf{J}_k/\mathbf{P}_k$ be the idele class group. Then a Grössencharakter is a continuous character χ of $\mathbf{C}_k$, and χ induces a character of $\mathbf{J}_k$, which is also denoted by χ. Let $\mathbf{J}_k=\mathbf{J}_\infty\times\mathbf{J}_0$ be the decomposition of $\mathbf{J}_k$ into the infinite part $\mathbf{J}_\infty$ and the finite part $\mathbf{J}_0$. Let $\mathbf{U}_0$ be the unit group of $\mathbf{J}_0$, and for each integral ideal $\mathfrak{m}$ of k, put $\mathbf{U}_{\mathfrak{m},0}=\{\mathfrak{u}\in\mathbf{U}_0\mid\mathfrak{u}\equiv 1\ (\mathrm{mod}\,\mathfrak{m})\}$, so that $\{\mathbf{U}_{\mathfrak{m},0}\}$ forms a base for the neighborhood system of 1 in $\mathbf{J}_0$. Put $\mathbf{J}_{\mathfrak{m},0}=\{\mathfrak{a}\in\mathbf{J}_0\mid\mathfrak{a}_\mathfrak{p}=1$ for all $\mathfrak{p}\mid\mathfrak{m}\}$, and with each $\mathfrak{a}\in\mathbf{J}_{\mathfrak{m},0}$, associate an ideal $\tilde{\mathfrak{a}}=\prod_\mathfrak{p}\mathfrak{p}^{v_\mathfrak{p}(\mathfrak{a})}$, where $\mathfrak{a}=(\mathfrak{a}_\mathfrak{p})$ and the ideal in $k_\mathfrak{p}$ generated by $\mathfrak{a}_\mathfrak{p}$ is equal to $\mathfrak{p}^{v_\mathfrak{p}(\mathfrak{a})}$. Then the mapping $\mathfrak{a}\to\tilde{\mathfrak{a}}$ gives a homomorphism of $\mathbf{J}_{\mathfrak{m},0}$ into the group $G(\mathfrak{m})=\{\tilde{\mathfrak{a}}\mid(\mathfrak{a},\mathfrak{m})=1\}$, and its kernel is contained in $\mathbf{U}_{\mathfrak{m},0}$. Since χ is continuous, $\chi(\mathbf{U}_{\mathfrak{m},0})=1$ for some $\mathfrak{m}$. The greatest common divisor $\mathfrak{f}$ of all such ideals $\mathfrak{m}$ is called the **conductor** of χ. For each $\mathfrak{a}\in\mathbf{J}_{\mathfrak{f},0}$, $\chi(\mathfrak{a})$ depends only on the ideal $\tilde{\mathfrak{a}}$ $(\in G(\mathfrak{f}))$; hence by putting $\chi(\mathfrak{a})=\tilde{\chi}(\tilde{\mathfrak{a}})$, we obtain a character $\tilde{\chi}$ of $G(\mathfrak{f})$. Now put $L_k(s,\chi)=\sum\tilde{\chi}(\tilde{\mathfrak{a}})/N(\tilde{\mathfrak{a}})^s$, where the sum is over all integral ideals $\tilde{\mathfrak{a}}\in G(\mathfrak{f})$. This is called a **Hecke L-function with Grössencharakter** χ. For $\chi\neq 1$, it is an entire function. On the other hand, if we restrict χ to $\mathbf{J}_\infty=\mathbf{R}^{*r_1}\times\mathbf{C}^{*r_2}$, then for $u=(a_1,\ldots,a_{r_1},a_{r_1+1},\ldots,a_{r_1+r_2})\in\mathbf{J}_\infty$, we have

$$\chi(u)=\prod_{j=1}^{r_1+r_2}|a_j|^{\lambda_j\sqrt{-1}}\cdot\prod_{j=1}^{r_1}(\operatorname{sgn}a_j)^{e_j}\cdot\prod_{l=r_1+1}^{r_1+r_2}\left(\frac{a_l}{|a_l|}\right)^{e_l},$$

where $e_j=0$ or 1, $e_l\in\mathbf{Z}$, $\lambda_j\in\mathbf{R}$. The numbers e_j, e_l, λ_j are determined uniquely by χ. Putting

$$\xi_k(s,\chi)=\left(\frac{\sqrt{|d|N(\mathfrak{f})}}{2^{r_2}\pi^{n/2}}\right)^s\cdot\prod_{j=1}^{r_1}\Gamma\left(\frac{s+e_j+\lambda_j\sqrt{-1}}{2}\right)\times\prod_{l=r_1+1}^{r_1+r_2}\Gamma\left(s+\frac{|e_l|+\lambda_l\sqrt{-1}}{2}\right)\cdot L(s,\chi),$$

we have a functional equation

$$\xi_k(s,\chi)=W(\chi)\xi_k(1-s,\bar{\chi}),$$

where $W(\chi)$ is a complex number with absolute value 1.

We can express $\xi_k(s,\chi)$ by an integral form on $\mathbf{J}_k$ as

$$\xi_k(s,\chi)=c\int_{\mathbf{J}_k}\varphi(\mathfrak{r})\chi(\mathfrak{r})V(\mathfrak{r})^s d^*\mathfrak{r},$$

where $V(\mathfrak{r})$ is the †total volume of the idele $\mathfrak{r}$, c is a constant that depends on the †Haar measure $d^*\mathfrak{r}$ of $\mathbf{J}_k$, and $\varphi(\mathfrak{r})$ is defined by

$$\varphi(\mathfrak{r})=\prod_{\mathfrak{p}}\varphi_{\mathfrak{p}}(x_{\mathfrak{p}}),\qquad \mathfrak{r}=(\ldots x_{\mathfrak{p}}\ldots),$$

$$\begin{aligned}&\varphi_{\mathfrak{p}_{\infty,i}}(x)\\ &=x^{e_i}e^{-\pi x^2},\quad i\leqslant r_1,\quad k_{\mathfrak{p}_{\infty,i}}=\mathbf{R},\\ &\left.\begin{aligned}&=\frac{1}{2\pi}\bar{x}^{e_i}e^{-2\pi|x|^2},\quad e_i\geqslant 0,\\ &=\frac{1}{2\pi}x^{-e_i}e^{-2\pi|x|^2},\quad e_i<0,\end{aligned}\right\}\ i>r_1,\quad k_{\mathfrak{p}_{\infty,i}}=\mathbf{C},\end{aligned}$$

$$\left.\begin{aligned}\varphi_{\mathfrak{p}}(x)&=e^{2\pi i\lambda(x)},\quad x\in(\mathfrak{b}\mathfrak{f})_{\mathfrak{p}}^{-1},\\ &=0,\qquad x\notin(\mathfrak{b}\mathfrak{f})_{\mathfrak{p}}^{-1}\end{aligned}\right\}\ \mathfrak{p}\text{ finite.}$$

Hence $(\mathfrak{b}\mathfrak{f})_{\mathfrak{p}}^{-1}$ is the $\mathfrak{p}$-component (→ 6 Adeles and Ideles B) of the ideal $(\mathfrak{b}\mathfrak{f})^{-1}$ ($\mathfrak{b}$ is the †different of $k/\mathbf{Q}$) and $\lambda(x)$ is an additive character of $k_{\mathfrak{p}}$ defined as follows. $\mathbf{Q}_p$ is the †p-adic field, $\mathbf{Z}_p$ is the ring of p-adic integers, λ_0 is the mapping $\mathbf{Q}_p\to\mathbf{Q}_p/\mathbf{Z}_p\subset\mathbf{Q}/\mathbf{Z}\subset\mathbf{R}/\mathbf{Z}$, and $\lambda=\lambda_0\circ Tr_{k_{\mathfrak{p}}/\mathbf{Q}_p}$. By putting $\chi(\mathfrak{r})=\prod_{\mathfrak{p}}\chi_{\mathfrak{p}}(x_{\mathfrak{p}})$, $\mathfrak{r}=(\ldots x_{\mathfrak{p}}\ldots)$, we have

$$\xi_k(s,\chi)=c\prod_{\mathfrak{p}}\int_{k_{\mathfrak{p}}}\varphi_{\mathfrak{p}}(x)\chi_{\mathfrak{p}}(x)V_{\mathfrak{p}}(x)^{-s}d^*x,$$

where $\mathfrak{p}$ runs over all prime divisors of k, finite or infinite. Moreover, with a constant $C_{\mathfrak{p}}$, we have

$$\begin{aligned}&\int_{k_{\mathfrak{p}_{\infty,i}}}\varphi_{\mathfrak{p}_{\infty,i}}(x)\chi_{\mathfrak{p}_{\infty,i}}(x)V_{\mathfrak{p}_{\infty,i}}(x)^{-s}d^*x\\ &=C_{\mathfrak{p}_{\infty,i}}\cdot\pi^{-(s+\sqrt{-1}\lambda_i+e_i)/2}\\ &\qquad\times\Gamma((s+\sqrt{-1}\lambda_i+e_i)/2),\quad k_{\mathfrak{p}_{\infty,i}}=\mathbf{R},\\ &=C_{\mathfrak{p}_{\infty,i}}\cdot(2\pi)^{-(s+(\sqrt{-1}\lambda_i+|e_i|)/2)}\\ &\qquad\times\Gamma(s+(\sqrt{-1}\lambda_i+|e_i|)/2),\quad k_{\mathfrak{p}_{\infty,i}}=\mathbf{C},\end{aligned}$$

$$\begin{aligned}&\int_{k_{\mathfrak{p}}}\varphi(x)\chi_{\mathfrak{p}}(x)V_{\mathfrak{p}}(x)^{-s}d^*x\\ &=C_{\mathfrak{p}}N(\mathfrak{b}_{\mathfrak{p}})^{s-1/2}\tilde{\chi}(\mathfrak{b}_{\mathfrak{p}}^{-1})\frac{1}{1-\tilde{\chi}(\mathfrak{p})/N(\mathfrak{p})^s},\\ &=C_{\mathfrak{p}}N((\mathfrak{b}\mathfrak{f})_{\mathfrak{p}})^s\tau_{\mathfrak{p}}(\chi_{\mathfrak{p}})\cdot\mu(U_{\mathfrak{f},\mathfrak{p}}),\quad \mathfrak{p}\,|\,\mathfrak{f}.\end{aligned}$$

Here $\tau_{\mathfrak{p}}(\chi_{\mathfrak{p}})$ is a constant called the **local Gaussian sum**, and $\mu(U_{\mathfrak{f},\mathfrak{p}})$ is the volume of $\{u\in k_{\mathfrak{p}}\,|\,u\equiv 1\ (\mathrm{mod}\ \mathfrak{f})\}$. These integrals over $k_{\mathfrak{p}}$ are the Γ-factors and Euler factors of $\xi_k(s,\chi)$, according as $\mathfrak{p}$ is infinite or finite. The functional equation is obtained by applying the †Poisson summation formula for $\varphi(x)$ and its †Fourier transform on the adele group $\mathbf{A}_k$ (→ 6 Adeles and Ideles).

Let $\mathbf{D}_k$ be the connected component of 1 in $\mathbf{C}_k$. If $\chi(\mathbf{D}_k)=1$, the corresponding $\tilde{\chi}$ is a character of an ideal class group of k with a conductor $\mathfrak{f}$. Conversely, all such characters can be obtained in this manner.

As stated in Section E, the Hecke L-functions with characters (of ideal class groups) can be used to describe the decomposition law of prime divisors in class field theory. However, for L-functions with Grössencharakter, such arithmetic implications have not been found yet, except that in the case of Grössencharakters of A_0 type, Y. Taniyama discovered, following the suggestion of A. Weil, that the L-function has a deep connection with the arithmetic of a certain infinite Abelian extension of k [T2, W7]. In particular, when $L(s,\chi)$ is a factor of the †Hasse ζ-function of an Abelian variety A with †complex multiplication, it describes the arithmetic of the field generated by the coordinates of the division points of A.

G. Artin L-Functions

Let K be a finite Galois extension of an algebraic number field k (of degree n), $G=G(K/k)$ be its Galois group, $\sigma\to A(\sigma)$ be a matrix representation (characteristic 0) of G, and χ be its character. Let $\mathfrak{p}$ be a prime ideal of k, and define $L_{\mathfrak{p}}(s,\chi)$ by

$$\log L_{\mathfrak{p}}(s,\chi)=\sum_{m=1}^{\infty}\frac{\chi(\mathfrak{p}^m)}{mN(\mathfrak{p}^m)^s},\qquad \mathrm{Re}\,s>1,$$

with $\chi(\mathfrak{p}^m)=(1/e)\sum_{\tau\in T}\chi(\sigma^m\tau)$, where T is the †inertia group of $\mathfrak{p}$, $|T|=e$, and σ is a †Frobenius automorphism of $\mathfrak{p}$. Then we have

$$L_{\mathfrak{p}}(s,\chi)=\det(E-A_{\mathfrak{p}}\cdot N(\mathfrak{p})^{-s})^{-1},$$

$$A_{\mathfrak{p}}=\frac{1}{e}\sum_{\tau\in T}A(\sigma\tau).$$

In particular, if $T=\{1\}$ (i.e., $\mathfrak{p}$ is †unramified in K/k), then

$$L_{\mathfrak{p}}(s,\chi)=\det(E-A(\sigma)\cdot N(\mathfrak{p})^{-s})^{-1}.$$

Now put

$$L(s,\chi,K/k)=\prod_{\mathfrak{p}}L_{\mathfrak{p}}(s,\chi),\qquad \mathrm{Re}\,s>1,$$

and call $L(s,\chi,K/k)$ an **Artin L-function** [A2].

(1) The most important property of $L(s,\chi,K/k)$ is that if K/k is an Abelian extension and χ is a linear character, it follows from class field theory that $\chi(\mathfrak{p})$ is the character of the ideal class group of k (modulo the †conductor of K/k) and that the Artin L-function equals

a Hecke L-function. This equality is equivalent to Artin's †reciprocity law, and in fact Artin obtained his reciprocity law after he conjectured the equality.

(2) If $K' \supset K \supset k$ and K'/k is a Galois extension, then $L(s, \chi, K/k) = L(s, \chi, K'/k)$.

(3) If $K \supset \Omega \supset k$ and ψ is a character of $G(K/\Omega)$, then $L(s, \psi, K/\Omega) = L(s, \chi_\psi, K/k)$, where χ_ψ is the character of $G(K/k)$ †induced from ψ.

(4) If $\chi_1 = 1$, then $L(s, \chi_1, K/k) = \zeta_k(s)$.

(5) $L(s, \chi_1 + \chi_2, K/k) = L(s, \chi_1, K/k) \cdot L(s, \chi_2, K/k)$.

Conversely, the Artin L-function $L(s, \chi, K/k)$ is characterized by properties (1)–(5).

(6) If χ_R is the †regular representation of G, then $L(s, \chi_R, K/k) = \zeta_K(s)$; hence

$$\zeta_K(s) = \zeta_k(s) \prod_{\chi \neq 1} L(s, \chi, K/k)^{\chi(1)},$$

where χ runs over all irreducible characters $\neq 1$ of G.

(7) Every character of a finite group G can be expressed as $\chi = \sum m_i \chi_{\psi_i}$ $(m_i \in \mathbf{Z})$, where each χ_{ψ_i} is an †induced character from a certain linear character ψ_i of an elementary subgroup of G (**Brauer's theorem**). (Here an elementary subgroup is a subgroup that is the direct product of a cyclic group and a p-group for some prime p.) Hence (3) and (5) imply that an Artin L-function is the product of integral powers (positive or negative) of Hecke L-functions $L_{\Omega_i}(s, \psi_i)$:

$$L(s, \chi, K/k) = \prod_i L_{\Omega_i}(s, \psi_i)^{m_i}.$$

Hence an Artin L-function is a univalent meromorphic function defined over the whole complex plane. Artin made the still open conjecture that if χ is irreducible and $\chi \neq 1$, then $L(s, \chi, K/k)$ is an entire function (**Artin's conjecture**).

This conjecture holds obviously if all m_i are nonnegative. Except for such a case, Artin's conjecture had no affirmative examples until 1974, when Deligne and Serre [D9] proved that each "new cusp form" of weight 1 gives rise to an entire Artin L-function $L(s, \chi, K/k)$ with $\chi(1) = 2$ and $\chi(\rho) = 0$ (ρ is the complex conjugation); by this method, some nontrivial examples were computed by J. Tate and J. Buhler (*Lecture notes in math.* 654 (1978)). Then R. P. Langlands [L5] constructed nontrivial examples of Artin's conjecture for certain 2-dimensional representations

$$\mathrm{Gal}(K/k) \ni \sigma \mapsto A(\sigma) \in GL(2, \mathbf{C})$$

by using ideas of H. Saito and T. Shintani [S1, S20]. This method works for all representations for which the image of the $A(\sigma)$ in $PGL(2, \mathbf{C})$ is the †tetrahedral group. It also works for some †octahedral cases, but a new idea is needed in the †icosahedral case.

(8) Let $\mathfrak{p}_{\infty,i}$ $(i = 1, \ldots, r_1 + r_2)$ be the infinite primes of k. Put

$$\gamma(s, \chi, \mathfrak{p}_{\infty,i}, K/k)$$
$$= (\Gamma(s/2)\Gamma((s+1)/2))^{\chi(1)}$$

for complex $\mathfrak{p}_{\infty,i}$,

$$= \Gamma(s/2)^{(\chi(1)+\chi(\sigma))/2}\Gamma((s+1)/2)^{(\chi(1)-\chi(\sigma))/2}$$

for real $\mathfrak{p}_{\infty,i}$,

where $\sigma \in G$ is the complex conjugation determined by a prime factor of $\mathfrak{p}_{\infty,i}$ in K. Next we introduce the notion of the **conductor** $\mathfrak{f}_\chi$ **with the group character** χ defined by Artin (*J. Reine Angew. Math.*, 164 (1931)). First, for any subset $\mathfrak{m} \subset G$, we put $\chi(\mathfrak{m}) = \sum_{m \in \mathfrak{m}} \chi(m)$; then $\mathfrak{f}_\chi$ is given by

$$\mathfrak{f}_\chi = \mathfrak{f}(\chi, K/k) = \prod_{\mathfrak{p}} \mathfrak{p}^{f(\mathfrak{p})},$$

where

$$f(\mathfrak{p}) = \frac{1}{e}[\{e\chi(1) - \chi(T)\} + \{p^{e_1}\chi(1) - \chi(V_1)\} + \{p^{e_2}\chi(1) - \chi(V_2)\} + \ldots],$$

and where $V_1, V_2, \ldots,$ are the higher †ramification groups of prime factors of $\mathfrak{p}$ in K (in lower numbering) and $p^{e_i} = |V_i|$ (→ 14 Algebraic Number Fields I).

Now put

$$\xi(s, \chi, K/k) = \left(\frac{|d|^{\chi(1)} N_k(\mathfrak{f}_\chi)}{\pi^{n\chi(1)}}\right)^{s/2} \times \prod_{\mathfrak{p}_{\infty,i}} \gamma(s, \chi, \mathfrak{p}_{\infty,i}, K/k) \cdot L(s, \chi, K/k).$$

Then the functional equation is written

$$\xi(1-s, \bar{\chi}, K/k) = W(\chi)\xi(s, \chi, K/k), \quad |W(\chi)| = 1.$$

The known proof of this functional equation depends on (7) and the functional equations of Hecke L-functions discussed in Section E. As for the constants $W(\chi)$, there are significant results by B. Dwork, Langlands, and Deligne [D6].

(9) There are some applications to the theory of the distribution of prime ideals.

H. Weil L-Functions

Weil defined a new L-function that is a generalization of both Artin L-functions and Hecke L-functions with Grössencharakter [W5]. Let K be a finite Galois extension of an algebraic number field k, let C_K be the idele class group $K_A^\times/K^\times$ of K, and let $\alpha_{K/k} \in H^2(\mathrm{Gal}(K/k), C_K)$ be the †canonical cohomology class of †class field theory. Then this $\alpha_{K/k}$ determines an extension $W_{K/k}$ of $\mathrm{Gal}(K/k)$ by C_K: $1 \to C_K \to W_{K/k} \to \mathrm{Gal}(K/k) \to 1$ (exact), and

the transfer induces an isomorphism $W_{K/k}^{ab} \simeq C_k$, where ab denotes the topological commutator quotient. If L is a Galois extension of k containing K, then there is a canonical homomorphism $W_{L/k} \to W_{K/k}$. Hence we define the **Weil group** W_k for $\bar{k}/k$ as the †projective limit group $\mathrm{proj}_K \lim W_{K/k}$ of the $W_{K/k}$. It is obvious that we have a surjective homomorphism $\varphi: W_k \to \mathrm{Gal}(\bar{k}/k)$ and an isomorphism $r_k: C_k \to W_k^{ab}$, where W_k^{ab} is the maximal Abelian Hausdorff quotient of W_k. For $w \in W_k$, let $\|w\|$ be the adelic norm of $r_k^{-1}(w)$.

If k_v is a †local field, then we define the Weil group W_{k_v} for $\bar{k}_v/k_v$ by replacing the idele class group C_K with the multiplicative group $K_w^\times$ in the above definition, where K_w denotes a Galois extension of k_v. If k_v is the completion of a finite algebraic number field k at a place v, then we have natural homomorphisms $k_v^\times \to C_k$ and $\mathrm{Gal}(\bar{k}_v/k_v) \to \mathrm{Gal}(\bar{k}/k)$. Accordingly, we have a homomorphism $W_{k_v} \to W_k$ that commutes with these homomorphisms.

Let W_k be the Weil group of an algebraic number field k, and let $\rho: W_k \to GL(V)$ be a continuous representation of W_k on a complex vector space V. Let $v = \mathfrak{p}$ be a finite prime of k, and let ρ_v be the representation of W_{k_v} induced from ρ. Let Φ be an element of W_{k_v} such that $\varphi(\Phi)$ is the inverse Frobenius element of $\mathfrak{p}$ in $\mathrm{Gal}(\bar{k}_v/k_v)$, and let I be the subgroup of W_{k_v} consisting of elements w such that $\varphi(w)$ belongs to the †inertia group of $\mathfrak{p}$ in $\mathrm{Gal}(\bar{k}_v/k_v)$. Let V^I be the subspace of elements in V fixed by $\rho_v(I)$, let $N\mathfrak{p}$ be the norm of $\mathfrak{p}$, and let

$$L_{\mathfrak{p}}(V,s) = \det(1-(N\mathfrak{p})^{-s}\rho_v(\Phi) \mid V^I)^{-1}.$$

We can define $L_v(V,s)$ for each Archimedean prime v also, and let

$$L(V,s) = \prod_v L_v(V,s).$$

Then this product converges for s in some right half-plane and defines a function $L(V,s)$. We call $L(V,s)$ the **Weil L-function** for the representation $\rho: W_k \to GL(V)$. This function $L(V,s)$ can be extended to a meromorphic function on the complex plane and satisfies the functional equation

$$L(V,s) = \varepsilon(V,s) L(V^*, 1-s)$$

(T. Tamagawa), where V^* is the dual of V, and $\varepsilon(V,s)$ is an exponential function of s of the form ab^s [T6].

P. Deligne generalized these results in the following manner: Let W_k' be a †group scheme over $\mathbf{Q}$ which is the †semidirect product of W_k by the additive group $\mathbf{G}_a$, on which W_k acts by the rule $wxw^{-1} = \|w\| x$. We can define the notion of representations of W_k' and the L-functions of them in the natural manner [T6].

I. The Riemann Hypothesis

As mentioned in Section B, the Riemann hypothesis asserts that all zeros of the Riemann ζ-function in $0 < \mathrm{Re}\, s < 1$ lie on the line $\mathrm{Re}\, s = 1/2$. In his celebrated paper [R1], Riemann gave six conjectures (including this), and assuming these conjectures, proved the †prime number theorem:

$$\pi(x) \sim \frac{x}{\log x} \sim \mathrm{Li}(x) = \int_2^x \frac{dx}{\log x}, \quad x \to \infty.$$

Here $\pi(x)$ denotes the number of prime numbers smaller than x. Among his six conjectures, all except the Riemann hypothesis have been proved (a detailed discussion is given in [L1]). The prime number theorem was proved independently by Hadamard and de La Vallée-Poussin without using the Riemann hypothesis (→ Section B; 123 Distribution of Prime Numbers B).

R. S. Lehman showed that there are exactly 2,500,000 zeros of $\zeta(\sigma + it)$ for which $0 < t < 170{,}571.35$, all of which lie on the critical line $\sigma = 1/2$ and are simple (*Math. Comp.*, 20 (1966)). Later R. P. Brent extended this computation up to 75,000,000 first zeros (1979).

Hardy proved that there are infinitely many zeros of $\zeta(s)$ on the line $\mathrm{Re}\, s = 1/2$ (1914). Furthermore, A. Selberg [S6] proved that if $N_0(T)$ is the number of zeros of $\zeta(s)$ on the line with $0 < \mathrm{Im}\, s < T$, then $N_0(T) > AT \log T$ (A is a positive constant) (1942). Thus if $N(T)$ is the number of zeros of $\zeta(s)$ in the rectangle $0 < \mathrm{Re}\, s < 1$, $0 < \mathrm{Im}\, s < T$, then $\liminf_{T\to\infty} N_0(T)/N(T) > 0$. N. Levinson proved $\liminf_{T\to\infty} N_0(T)/N(T) > 1/3$ (*Advances in Math.*, 13 (1974)). If $N_\varepsilon(T)$ is the number of zeros of $\zeta(s)$ in $1/2 - \varepsilon < \mathrm{Re}\, s < 1/2 + \varepsilon$, $0 < \mathrm{Im}\, s < T$, then $\lim_{T\to\infty} N_\varepsilon(T)/N(T) = 1$ for any positive number ε (H. Bohr and E. Landau, 1914). Bohr studied the distribution of the values of $\zeta(s)$ in detail and initiated the theory of †almost periodic functions (1925).

D. Hilbert remarked in his lecture at the Paris Congress that the Riemann hypothesis is equivalent to

$$\pi(x) = \mathrm{Li}(x) + O(\sqrt{x} \log x), \quad x \to \infty$$

(H. von Koch, 1901). It is also equivalent to

$$\sum_{n=1}^{N} \mu(n) = O(N^{1/2+\varepsilon}), \quad N \to \infty,$$

for any $\varepsilon > 0$, where $\mu(n)$ is the Möbius function. Assuming the Riemann hypothesis, we get

$$N(T) = \frac{1}{2\pi} T \log T - \frac{1 + \log 2\pi}{2\pi} T + o(\log T)$$

(Littlewood, 1924).

The computation of the zeros of the ζ-functions and the L-functions of general algebraic number fields is more difficult, but conjectures similar to the Riemann hypothesis have been proposed.

Weil showed that a necessary and sufficient condition for the validity of the Riemann hypothesis for all Hecke L-functions $L(s,\chi)$ is that a certain †distribution on the idele group $\mathbf{J}_k$ be positive definite [W1 (1952b)].

It is not known whether the general ζ- and L-functions of algebraic number fields have any zeros in the interval (0, 1) on the real axis (see the works of A. Selberg and S. Chowla). Similar problems are considered for the various ζ-functions given in Sections P, Q, and T.

J. p-Adic L-Functions

Let χ be a †primitive Dirichlet character with conductor f, and let $L(s,\chi)$ be the †Dirichlet L-function for χ. Then the values $L(1-n,\chi)$ of $L(s,\chi)$ at nonpositive integers $1-n$ $(n=1,2,\dots)$ are algebraic numbers (→ Section E). Let p be a prime number, let $\mathbf{Q}_p$ be the †p-adic number field, and let $\mathbf{C}_p$ be the completion of the algebraic closure $\bar{\mathbf{Q}}_p$ of $\mathbf{Q}_p$. It is known that $\mathbf{C}_p$ is also algebraically closed. Since $\mathbf{Q}\subset\mathbf{Q}_p$, we fix an embedding $\bar{\mathbf{Q}}\subset\bar{\mathbf{Q}}_p$ and consider $\{L(1-n,\chi)\}_{n=1}^{\infty}$ as a sequence in $\mathbf{C}_p$.

Let $|\ |_p$ be the extension to $\mathbf{C}_p$ of the standard p-adic valuation of $\mathbf{Q}_p$. Let q be p or 4 according as $p\neq 2$ or $p=2$, and let ω be the primitive Dirichlet character with conductor q satisfying $\omega(n)\equiv n \pmod q$ for any integer n prime to p. Then T. Kubota and H. W. Leopoldt proved that there exists a unique function $L_p(s,\chi)$ satisfying the conditions [K5]:

(1) $L_p(s,\chi)=\dfrac{a_{-1}}{s-1}+\sum_{n=0}^{\infty}a_n(s-1)^n \quad (a_n\in\mathbf{C}_p)$;

(2) $a_{-1}=0$ if $\chi\neq 1$ and the series $\sum_{n=0}^{\infty}a_n(s-1)^n$ converges for $|s-1|_p<|q^{-1}p^{1/(p-1)}|_p$;

(3) $L_p(1-n,\chi)=(1-\chi\omega^{-n}p^{1-n})L(1-n,\chi\omega^{-n})$ holds for $n=1,2,3,\dots$.

The function $L_p(s,\chi)$ satisfying these three conditions is called the **p-adic L-function** for the character χ. It is easy to see that $L_p(s,\chi)$ is identically zero if $\chi(-1)=-1$, but $L_p(s,\chi)$ is nontrivial if $\chi(-1)=1$.

Let B_n be the Bernoulli number. Then B_n satisfies the conditions: (1) B_n/n is p-integral if $(p-1)\nmid n$ (von Staudt) and (2) $(1/n)B_n\equiv(1/(n+p-1))B_{n+p-1} \pmod p$ holds in this case (Kummer). The generalization of these results for the generalized Bermoulli number $B_{\chi,n}$ was obtained by Leopoldt. Since $L(1-n,\chi)=-(1/n)\mathrm{B}_{\chi,n}$, such p-integrabilities and congruences can be naturally interpreted and generalized in terms of the p-adic L-functions $L(s,\chi)$.

We assume $\chi(-1)=1$. Then $L_p(0,\chi)=(1-\chi\omega^{-1}(p))L(0,\chi\omega^{-1})$ and $\chi\omega^{-1}(-1)=-1$. Hence we can express the first factor h_N^- of the class number of a cyclotomic field $\mathbf{Q}(\exp(2\pi i/N))$ as a product of some $L_p(0,\chi)$'s. By using this fact, K. Iwasawa proved [I7] that the p-part $p^{e_n^-}$ of the †first factor $h_{Np^n}^-$ $(N\in\mathbf{N})$ satisfies

$$e_n^-=\lambda n+\mu p^n+\nu \qquad (\lambda,\mu,\nu\in\mathbf{Z};\lambda,\mu\geqslant 0)$$

for any sufficiently large n. Here Iwasawa conjectured $\mu=0$, which was proved by B. Ferrero and L. Washington [F1]. Also, we can obtain some congruences involving the first factor h_N^- of $\mathbf{Q}(\exp(2\pi i/N))$ from this formula.

Let χ be a nontrivial primitive Dirichlet character with conductor f, let

$$\tau(\chi)=\sum_{a=1}^{f}\chi(a)e^{2\pi i a/f}$$

be the †Gaussian sum for χ, and let $\log_p$ be the p-adic logarithmic function. Then Leopoldt [L6] calculated the value $L(1,\chi)$ and obtained

$$L_p(1,\chi)$$
$$=-\left(1-\frac{\chi(p)}{p}\right)\frac{\tau(\chi)}{f}\sum_{a=1}^{f}\chi(a)\log_p(1-e^{-2\pi i a/f}).$$

As an application of this formula, Leopoldt obtained a p-adic †class number formula for the maximal real subfield $F=\mathbf{Q}(\cos(2\pi/N))$ of $\mathbf{Q}(\exp(2\pi i/N))$: Let $\zeta_p(s,F)$ be the product of the $L_p(s,\chi)$ for all primitive Dirichlet characters χ such that (1) $\chi(-1)=1$ and (2) the conductor of χ is a divisor of N. We define the **p-adic regulator** R_p by replacing the usual log by the p-adic logarithmic function $\log_p$. Let h be the class number of F, $m=[F:\mathbf{Q}]$, and let d be the discriminant of F. Then the residue of $\zeta_p(s,F)$ at $s=1$ is

$$\prod_{\chi}\left(1-\frac{\chi(p)}{p}\right)\frac{2^{m-1}hR_p}{\sqrt{d}}.$$

Hence $\zeta_p(s,F)$ has a simple pole at $s=1$ if and only if $R_p\neq 0$. In general, for any totally real finite algebraic number field F, Leopoldt conjectured that the p-adic regulator R_p of F is not zero (**Leopoldt's conjecture**). This conjecture was proved by J. Ax and A. Brumer for the case when F is an Abelian extension of $\mathbf{Q}$ [A4, B7].

By making use of the Stickelberger element, Iwasawa gave another proof of the existence of the p-adic L-function [I7]. In particular, he obtained the following result: Let χ be a primitive Dirichlet character with conductor f. Then there exists a primitive Dirichlet character θ such that the p-part of the conductor of θ is

either 1 or q and such that the conductor and the order of $\chi\theta^{-1}$ are both powers of p. Let $\mathfrak{o}_\theta$ be the ring generated over the ring $\mathbf{Z}_p$ of p-adic integers by the values of θ. Then there exists a unique element $f(x,\theta)$ of the quotient field of $\mathfrak{o}_\theta[[x]]$ depending only on θ and satisfying

$$L_p(s,\chi)=2f(\zeta(1+q_0)^s-1,\theta),$$

where q_0 is the least common multiple of q and the conductor of θ, and $\zeta=\chi(1+q_0)^{-1}$. Furthermore, Iwasawa proved that $f(x,\theta)$ belongs to $\mathfrak{o}_\theta[[x]]$ if θ is not trivial.

Let $P=\mathbf{Q}(\exp(2\pi i/q))$ and, for any $n\geqslant 1$, let $P_n=\mathbf{Q}(\exp(2\pi i/qp^n))$. Let $P_\infty=\bigcup_{n\geqslant 1}P_n$. Then P_∞ is a Galois extension of $\mathbf{Q}$ satisfying $\mathrm{Gal}(P_\infty/\mathbf{Q})\cong\mathbf{Z}_p^\times$ (the multiplicative group of p-adic units), and P is the subfield of $P_\infty/\mathbf{Q}$ corresponding to the subgroup $1+q\mathbf{Z}_p$ of $\mathbf{Z}_p^\times$.

Let ψ be a $\mathbf{C}_p$-valued primitive Dirichlet character such that (1) $\psi(-1)=-1$ and (2) the p-part of the conductor f_ψ of ψ is either 1 or q. Let K_ψ be the cyclic extension of $\mathbf{Q}$ corresponding to ψ by class field theory. Let $K=K_\psi\cdot P$, $K_n=K\cdot P_n$ and $K_\infty=K\cdot P_\infty$. Let A_n be the p-primary part of the ideal class group of K_n, let $A_n\to A_m$ $(n\geqslant m)$ be the mapping induced by the †relative norm N_{k_n/K_m}, and let $X_K=\varprojlim A_n$. Since each A_n is a finite p-group, X_K is a $\mathbf{Z}_p$-module. Let $V_K=X_k\otimes_{\mathbf{Z}_p}C_p$, and let

$$V_\psi=\{v\in V_K\mid\delta(v)=\psi(\delta)v\text{ for all }\delta\in\mathrm{Gal}(K/\mathbf{Q})\}.$$

Let q_0 be the least common multiple of f_ψ and q, and let γ_0 be the element of $\mathrm{Gal}(K_\infty/K)$ that corresponds to

$$1+q_0\in 1+q\mathbf{Z}_p=\mathrm{Gal}(P_\infty/P)$$

by the restriction mapping $\mathrm{Gal}(K_\infty/K)\hookrightarrow\mathrm{Gal}(P_\infty/P)$. Let $f_\psi(x)$ be the characteristic polynomial of γ_0-1 acting on V_ψ. Hence $f_\psi(x)$ is an element of $\mathfrak{o}_\psi[x]$.

We assume that $\omega\psi^{-1}$ is not trivial. Let $f(x,\omega\psi^{-1})$ be as before. Then $f(x,\omega\psi^{-1})$ is an element of $\mathfrak{o}_\psi[[x]]$. Iwasawa conjectured that $f_\psi(x)$ and $f(x,\omega\psi^{-1})$ coincide up to a unit of $\mathfrak{o}_\psi[[x]]$ (**Iwasawa's main conjecture**). This conjecture was proved recently by B. Mazur and A. Wiles in the case where ψ is a power of ω.

Let F be a totally real finite algebraic number field, let K be a totally real Abelian extension of F, and let χ be a character of $\mathrm{Gal}(K/F)$. Let $L(s,\chi)$ be the †Artin L-function for χ. Then we can construct the p-adic analog $L_p(s,\chi)$ of $L(s,\chi)$ (J.-P. Serre, J. Coates, W. Sinnott, P. Deligne, K. Ribet, P. Cassou-Noguès). But, at present, we have no formula for $L_p(1,\chi)$. Coates generalized Iwasawa's main conjecture to this case, but it has not yet been proven.

P-adic L-functions have been defined in some other cases (e.g. → [K3, M1, M3]).

K. ζ-Functions of Quadratic Forms

Dirichlet defined a Dirichlet series associated with a binary quadratic form and also considered a sum of such Dirichlet series extended over all classes of binary quadratic forms with a given discriminant D, which is actually equivalent to the Dedekind ζ-function of a quadratic field. Dirichlet obtained a formula for the class numbers of binary quadratic forms. The formula is interpreted nowadays as a formula for the class numbers of quadratic fields in the narrow sense.

According as the binary quadratic form is definite or indefinite, we apply different methods to obtain its class number.

Epstein ζ-functions: P. Epstein generalized the definition of the ζ-function of a positive definite binary quadratic form to the case of n variables (*Math. Ann.*, 56 (1903), 63 (1907)). Let V be a real vector space of dimension m with a positive definite quadratic form Q. Let M be a †lattice in V, and put

$$\zeta_Q(s,M)=\sum_{\substack{x\in M\\ x\neq 0}}\frac{1}{Q(x)^s},\quad \mathrm{Re}\,s>\frac{m}{2}.$$

This series is absolutely convergent in $\mathrm{Re}\,s>m/2$, and

$$\lim_{s\to m/2}\left(s-\frac{m}{2}\right)\zeta_Q(s,M)=D(M)^{-1/2}\pi^{m/2}\Gamma\left(\frac{m}{2}\right)^{-1},$$

$$D(M)=\det|Q(x_i,x_j)|,$$

where $x_1,\ldots,x_m$ is a basis of M and $Q(x,y)=(Q(x+y)-Q(x)-Q(y))/2$. If the $Q(x)$ $(x\in M, x\neq 0)$ are all positive integers, we can write

$$\zeta_Q(s,M)=\sum_{n=1}^{\infty}\frac{a(n)}{n^s},$$

where $a(n)$ is the number of distinct $x\in M$ with $Q(x)=n$. In general, let $x_1,\ldots,x_m$ be a basis of the lattice M and $x_1^*,\ldots,x_m^*$ be its dual basis $(Q(x_i,x_j^*)=\delta_{ij})$. Call $M^*=\sum_i x_i^*\mathbf{Z}$ the **dual lattice** of M. If we consider the ϑ-series (†theta series)

$$\vartheta_Q(u,M)=\sum_{x\in M}\exp(-\pi uQ(x))\quad(\mathrm{Re}\,u>0),$$

then

$$\vartheta_Q(u,M)=(u^{-m/2}D(M)^{-1/2})\vartheta_Q(u^{-1},M^*).$$

With $\xi_Q(s,M)=\pi^{-s}\Gamma(s)\zeta_Q(s,M)$, the displayed equality leads to the functional equation

$$\xi_Q(s,M)=D(M)^{-1/2}\cdot\xi_Q\left(\frac{m}{2}-s,M^*\right).$$

In general, $\zeta_Q(s,M)$ has no Euler product expansion.

Consider the case where $M=\sum\mathbf{Z}x_i$ $(x_i=$

$(0, \ldots, 0, 1, 0, \ldots, 0))$, $Q(x) = \sum_{i=1}^{m} u_i^2$, for $x = (u_1, \ldots, u_m)$. If we put $\zeta_m(s) = \zeta_Q(s, M)$, $L(s) = \sum_{n=1}^{\infty} (-4/n) n^{-s}$, then we have

$$\zeta_1(s) = 2\zeta(2s),$$

$$\zeta_2(s) = 4\zeta(s) \cdot L(s) = 4 \times (\text{the Dedekind } \zeta\text{-function of } \mathbf{Q}(\sqrt{-1})),$$

$$\zeta_4(s) = 8(1 - 2^{2-2s})\zeta(s)\zeta(s-1),$$

$$\zeta_6(s) = -4(\zeta(s)L(s-2) - 4\zeta(s-2)L(s)),$$

$$\zeta_8(s) = 16(1 - 2^{1-s} + 2^{4-2s})\zeta(s)\zeta(s-3),$$

$$\zeta_{10}(s) = (4/5)(\zeta(s)L(s-4) + 4^2\zeta(s-4)L(s)) - 2\sum_{\substack{\mu \in \mathbf{Z}[\sqrt{-1}] \\ \mu \neq 0}} \frac{\mu^4}{(\mu\bar{\mu})^s},$$

$$\zeta_{12}(s) = c_1 2^{-s}\zeta(s)\zeta(s-5)(2^6 - 2^{6-s}) + c_2\varphi\{\sqrt{\Delta(\tau)}\},$$

where $\varphi\{\sqrt{\Delta(\tau)}\}$ is the Dirichlet series corresponding to $\sqrt{\Delta(\tau)}$ by the †Mellin transform and $\Delta(\tau) = z\{\prod_{n=1}^{\infty}(1 - z^n)\}^{24}$ with $z = e^{2\pi i \tau}$. $\zeta_m(s)$ has zeros on the line $\mathrm{Re}\, s = \sigma = m/4$, given explicitly for $m = 4, 8$ as follows:

$$m = 4: \quad s = 1 + l\pi i/\log 2, \quad l = 1, 2, \ldots,$$

$$m = 8: \quad s = 2 + (i/\log 2)(2l\pi \pm \arctan\sqrt{15}), \quad l = 0, \pm 1, \ldots.$$

Regarding the Epstein ζ-function of binary quadratic forms

$$\zeta_Q(s) = \sum_{m,n}{}' Q(m,n)^{-s},$$

with

$$Q(x, y) = ax^2 + bxy + cy^2, \quad a, b, c \in \mathbf{R}, a > 0, c > 0, \Delta = 4ac - b^2 > 0,$$

we have the Chowla-Selberg formula (1949):

$$\zeta_Q(s) = \left(2\zeta(2s)a^{-s} + \frac{2^{2s}a^{s-1}\sqrt{\pi}}{\Gamma(s)\Delta^{s-1/2}} \times \zeta(2s-1)\Gamma\left(s - \frac{1}{2}\right)\right) + \left(\frac{\pi^s 2^{s+3/2}}{a^{1/2}\Gamma(s)\Delta^{s/2-1/4}} \times \sum_{n=1}^{\infty} n^{s-1/2}\sigma_{1-2s}(n)\cos\frac{n\pi b}{a}\right) \times \int_0^{\infty} \varphi^{s-3/2}\exp\left\{-\frac{\pi n \Delta^{1/2}}{2a}(\varphi + \varphi^{-1})\right\} d\varphi,$$

where $\sigma_k(n) = \sum_{d|n} d^k$ and $\zeta(s)$ is the Riemann ζ-function. By using this formula, we can give another proof of the following result of H. Heilbronn: Let $h(-\Delta)$ be the class number of the imaginary quadratic field with discriminant $-\Delta$. Then $h(-\Delta) \to \infty$ $(\Delta \to \infty)$.

The following generalization of this result was obtained by C. L. Siegel [S22]: Let k be a fixed finite algebraic number field. Let K be a finite Galois extension of k, and let $d = d(K)$, $h = h(K)$, and $R = R(K)$ be the discriminant of K, the class number of K, and the regulator of K, respectively. We assume that K runs over extensions of k such that $[K:k]/\log d \to 0$; then we have

$$\log(hR) \sim \log\sqrt{|d|}.$$

Siegel ζ-functions of indefinite quadratic forms: Siegel defined and investigated some ζ-functions attached to nondegenerate indefinite quadratic forms, which are also meromorphic on the whole complex plane and satisfy certain functional equations [S24].

The case of quadratic forms with irrational algebraic coefficients was treated by Tamagawa and K. G. Ramanathan.

L. ζ-Functions of Algebras

K. Hey defined the ζ-function of a †simple algebra A over the rational number field $\mathbf{Q}$ (M. Deuring [D10]) (→ 27 Arithmetic of Associative Algebras). Consider an arbitrary †maximal order $\mathfrak{o}$ of A, and let

$$\zeta_A(s) = \sum_{\mathfrak{a}} \frac{1}{N(\mathfrak{a})^s}, \quad \mathrm{Re}\, s > 1,$$

with the summation taken over all left integral ideals $\mathfrak{a}$ of $\mathfrak{o}$. Then ζ_A is independent of the choice of a maximal order $\mathfrak{o}$. Let k be the †center of A, and put $[A:k] = n^2$. First, ζ_A is decomposed into Euler's infinite product expansion $\zeta_A(s) = \prod_{\mathfrak{p}} \mathbf{Z}_{\mathfrak{p}}(s)$ ($\mathfrak{p}$ runs over the prime ideals of k). For $\mathfrak{p}$ not dividing the discriminant $\mathfrak{d}$ of A, $Z_{\mathfrak{p}}(s)$ coincides with the $\mathfrak{p}$-component of $\prod_{j=0}^{n-1}\zeta_k(ns - j)$. Hence $\zeta_A(s)$ coincides with $\prod_{j=0}^{n-1}\zeta_k(ns - j)$ up to a product of $\mathfrak{p}$-factors for $\mathfrak{p} | \mathfrak{d}$ which are explicit rational functions of $N(\mathfrak{p})^{-ns}$.

Moreover, if A is the total matrix algebra of degree r over the division algebra $\mathfrak{D}$, then we have $\zeta_A(s) = \prod_{j=0}^{r-1}\zeta_{\mathfrak{D}}(rs - j)$, and $\zeta_{\mathfrak{D}}(s)$ satisfies a functional equation similar to that of $\zeta_k(s)$ (Hey). Also, $\zeta_A(s)$ is meromorphic over the whole complex plane, and at $s = 1, (n-1)/n, \ldots, 1/n$, it has poles of order 1. Using analytic methods, M. Zorn (1931) showed that the simple algebra A with center k such that $A_{\mathfrak{p}}$ is a matrix algebra over $k_{\mathfrak{p}}$ for every finite or infinite prime divisor $\mathfrak{p}$ of k is itself a matrix algebra over k (→ 27 Arithmetic of Associative Algebras D). A purely algebraic proof of this was given by Brauer, H. Hasse, and E. Noether. G. Fujisaki (1958) gave another proof using the Iwasawa-Tate method. As a direct

application of the ζ-function, the computation of the residue at $s=1$ of ζ_A leads to the formula containing the class number of maximal order $\mathfrak{O}$.

Godement defined the ζ-function of fairly general algebras [G1], and Tamagawa investigated in detail the explicit ζ-functions of division algebras, deriving their functional equations [T1].

Let $\tilde{A}=\prod'_{\mathfrak{p}} A_{\mathfrak{p}}$ be the adele ring of A, and let $G=\prod'_{\mathfrak{p}} G_{\mathfrak{p}}$ be the idele group (of A). We take a maximal order $\mathfrak{O}_{\mathfrak{p}}$ of $A_{\mathfrak{p}}$ and a maximal compact subgroup $U_{\mathfrak{p}}$ of $G_{\mathfrak{p}}$. Let $\omega_{\mathfrak{p}}$ be a †zonal spherical function of $G_{\mathfrak{p}}$ with respect to $U_{\mathfrak{p}}$; that is, $\omega_{\mathfrak{p}}$ is a function in $G_{\mathfrak{p}}$ and satisfies

$$\omega_{\mathfrak{p}}(ugv)=\omega_{\mathfrak{p}}(g) \quad (u,v\in U_{\mathfrak{p}}), \quad \omega_{\mathfrak{p}}(1)=1,$$

$$\int_{U_{\mathfrak{p}}} \omega_{\mathfrak{p}}(guh)\,du=\omega_{\mathfrak{p}}(g)\omega_{\mathfrak{p}}(h).$$

In addition, we define the weight function $\varphi_{\mathfrak{p}}$ on $A_{\mathfrak{p}}$ by

$$\varphi_{\mathfrak{p}}(x)=\begin{cases}\text{the characteristic function of } \mathfrak{O}_{\mathfrak{p}} \\ \text{when } \mathfrak{p} \text{ is finite,} \\ \\ \exp(-\pi T_{\mathfrak{p}}(xx^*)) \\ \text{when } \mathfrak{p} \text{ is infinite,}\end{cases}$$

where $T_{\mathfrak{p}}$ is the †reduced trace of $A_{\mathfrak{p}}/\mathbf{R}$ and $*$ is a positive †involution. Tamagawa gave an explicit form of the local ζ-function with the character $\omega_{\mathfrak{p}}$ defined by

$$\zeta_{\mathfrak{p}}(s,\omega_{\mathfrak{p}})=\int_{G_{\mathfrak{p}}} \varphi_{\mathfrak{p}}(g)\omega_{\mathfrak{p}}(g^{-1})|N_{\mathfrak{p}}(g)|_{\mathfrak{p}}^{s}\,dg,$$

where $N_{\mathfrak{p}}$ is the †reduced norm of $A_{\mathfrak{p}}/k_{\mathfrak{p}}$, and $|\ |_{\mathfrak{p}}$ is the valuation of $k_{\mathfrak{p}}$. Then $\omega=\prod_{\mathfrak{p}}\omega_{\mathfrak{p}}$ is the zonal spherical function of G with respect to $\prod U_{\mathfrak{p}}=U$. In particular, if ω is a positive definite zonal spherical function belonging to the spectrum of the discrete subgroup $\Gamma=A^*$ $=\{$all the invertible elements of $A\}$ of G, then the **Tamagawa ζ-function** with character ω is given by

$$\zeta(s,\omega)=\prod_{\mathfrak{p}}\zeta_{\mathfrak{p}}(s,\omega_{\mathfrak{p}})=\int_G \varphi(g)\omega(g^{-1})\|g\|^s\,dg,$$

where $\varphi(g)=\prod\varphi_{\mathfrak{p}}(g_{\mathfrak{p}})$ and $\|\ \|$ is the volume of the element g of G. When A is a division algebra, $\zeta(s,\omega)$ is analytically continued to a meromorphic function over the whole complex plane and satisfies the functional equation. The Tamagawa ζ-function may also be considered as one type of ζ-function of the Hecke operator. When A is an indefinite quaternion algebra over a totally real algebraic number field Φ, the groups of units of various orders of A operate discontinuously on the product of complex upper half-planes. Thus the spaces of holomorphic forms are naturally associated with A. The investigation of ζ-functions associated with these holomorphic automorphic forms was initiated by M. Eichler and extended by G. Shimura, H. Shimizu, and others. Eichler investigated the case $\Phi=\mathbf{Q}$, and Shimura and Shimizu investigated the case for an arbitrary totally real field Φ by defining general holomorphic automorphic forms, Hecke operators, and corresponding ζ-functions. The functional equations of these ζ-functions were proved by Shimizu. Shimizu generalized Eichler's work and found relations among ζ-functions of orders of various quaternion algebras belonging to different discriminants and levels [S10]. For the related results, see, e.g., the work of K. Doi and H. Naganuma [D12].

M. ζ-Functions Defined by Hecke Operators

The ζ-functions of algebraic number fields, algebras, or quadratic forms, and the L-functions are all defined by Dirichlet series, are analytically continued to univalent functions on the complex plane, and satisfy functional equations. One problem is to characterize the functions having such properties.

(1) H. Hamburger (1921–1922) characterized the Riemann ζ-function (up to constant multiples) by the following three properties: (i) It can be expanded as $\zeta(s)=\sum_{n=1}^{\infty}a_n/n^s$ ($\operatorname{Re}s\gg 0$); (ii) it is holomorphic on the complex plane except as $s=1$, and $(s-1)\zeta(s)$ is an entire function of finite †genus; (iii) $G(s)=G(1-s)$, where $G(s)=\pi^{-s/2}\Gamma(s/2)\zeta(s)$.

(2) E. Hecke's theory [H4]: Fixing $\lambda>0$, $k>0$, $\gamma=\pm 1$, and putting

$$R(s)=(2\pi/\lambda)^{-s}\Gamma(s)\varphi(s)$$

for an analytic function $\varphi(s)$, we make the following three assumptions: (i) $(s-k)\varphi(s)$ is an entire function of finite genus; (ii) $R(s)=\gamma R(k-s)$; (iii) $\varphi(s)$ can be expanded as $\varphi(s)=\sum_{n=1}^{\infty}a_n/n^s$ ($\operatorname{Re}s>\sigma_0$). Then we call $\varphi(s)$ a function belonging to the sign (λ,k,γ).

The functions $\zeta(2s)$, $L(2s)$, and $L(2s-1)$ satisfy assumptions (i)–(iii), where L may be either a Dirichlet L-function, an L-function with Grössencharakter of an imaginary quadratic field, or an L-function with class character of a real quadratic form whose Γ-factors are of the form $\Gamma(s/2)\Gamma((s+1)/2)\sim\Gamma(s)$. If $\varphi(s)$ belongs to the sign (λ,k,γ), then $n^{-s}\varphi(s)$ belongs to the sign $(n\lambda,k,\gamma)$. To each Dirichlet series $\varphi(s)=\sum_{n=1}^{\infty}a_n/n^s$ with the sign (λ,k,γ), we attach the series $f(\tau)=a_0+\sum_{n=1}^{\infty}a_n e^{2\pi i n\tau/\lambda}$, where

$$a_0=\gamma(2\pi/\lambda)^{-k}\Gamma(k)\operatorname{Res}_{s=k}(\varphi(s))$$
$$=\gamma\operatorname{Res}_{s=k}(R(s)).$$

This correspondence $\varphi(s)\to f(\tau)$ may also be

realized by the †Mellin transform

$$R(s)=\int_0^{\infty}\left(\sum_{n=1}^{\infty}a_n e^{-2\pi ny/\lambda}\right)y^{s-1}\,dy$$
$$=\int_0^{\infty}(f(iy)-a_0)y^{s-1}\,dy,$$
$$f(iy)-a_0=\frac{1}{2\pi i}\int_{\mathrm{Re}\,s=\sigma_0}R(s)y^{-s}\,ds.$$

In this case, (i) $f(\tau)$ is holomorphic in the upper half-plane and $f(\tau+\lambda)=f(\tau)$, (ii) $f(-1/\tau)/(-i\tau)^k=\gamma f(\tau)$, and (iii) $f(x+iy)=O(y^{\mathrm{const}})$ $(y\rightarrow+0)$ uniformly for all x.

Conversely, the Dirichlet series $\varphi(s)=\sum_{n=1}^{\infty}a_n n^{-s}$ formed by the transformation in the previous paragraph from $f(\tau)$ satisfying (i)–(iii) belongs to the sign (λ,k,γ). We also say that the function $f(\tau)$ belongs to the sign (λ,k,γ).

If k is an even integer, then the functions $f(\tau)$ belonging to $(1,k,(-1)^{k/2})$ are the †modular forms of level 1 and weight k. A necessary and sufficient condition for a function $\varphi(s)$ belonging to $(1,k,(-1)^{k/2})$ to have an Euler product is that the corresponding modular form $f(\tau)$ be a simultaneous eigenfunction of the ring formed by the †Hecke operators T_n $(n=1,2,\dots)$. In this case, the coefficient a_n of $\varphi(s)=\sum a_n n^{-s}$ coincides with the eigenvalue of T_n. Namely, if $f\,|\,T_n=t_n f$, we have $\varphi(s)=a_1(\sum_{n=1}^{\infty}t_n n^{-s})$, and this is decomposed into the Euler product $\varphi(s)=a_1\prod_p(1-t_p p^{-s}+p^{k-1-2s})^{-1}$. We call $\varphi(s)/a_1$ a **ζ-function defined by Hecke operators** (Hecke [H5]). For example, $\zeta(s)\cdot\zeta(s-k+1)$ and the Ramanujan function

$$\sum_{n=1}^{\infty}\tau(n)n^{-s}=\prod_p(1-\tau(p)p^{-s}+p^{11-2s})^{-1}$$

are ζ-functions defined by Hecke operators. Hecke applied the theory of Hecke operators to study the group $\Gamma(N)$ [H5]; the situation is more complicated than the case of $\Gamma(1)=SL(2,\mathbf{Z})$. The space of automorphic forms of weight k belonging to the †congruence subgroup

$$\Gamma_0(N)=\left\{\begin{pmatrix}a & b\\ c & d\end{pmatrix}\in SL(2,\mathbf{Z})\,\middle|\,c\equiv 0\pmod N\right\}$$

is denoted by $\mathfrak{M}_k(\Gamma_0(N))$. The essential part of $\mathfrak{M}_k(\Gamma_0(N))$ is spanned by the functions $f(\tau)=\sum a_n e^{2\pi in\tau}$ satisfying the conditions: (1) $\varphi(s)=\sum a_n n^{-s}$ has the Euler product expansion

$$\varphi(s)=\prod_{p|N}(1-a_p p^{-s})^{-1}$$
$$\times\prod_{p\nmid N}(1-a_p p^{-s}+p^{k-1-2s})^{-1}.$$

(2) The functional equation $R(s)=\gamma R(k-s)$ holds, where $R(s)=(2\pi/\sqrt{N})^{-s}\Gamma(s)\varphi(s)$. (3) When χ is an arbitrary primitive character of $\mathbf{Z}$ such that the conductor f is coprime to N, then

$$R(s,\chi)=(2\pi/\sqrt{N}\,f)^{-s}\Gamma(s)\sum a_n\chi(n)n^{-s}$$

extends to an entire function satisfying the functional equation $R(s,\chi)=\omega R(k-s,\bar\chi)$ $(|\omega|=1)$ (Shimura). Conversely, (2) and (3) characterize the Dirichlet series $\varphi(s)$ corresponding to $f(\tau)\in\mathfrak{M}_k(\Gamma_0(N))$ (Weil [W1 (1967a)]).

Considering the correspondence $f(\tau)=\sum a_n q^n\rightarrow\varphi(s)=\sum a_n n^{-s}$ not as a Mellin transformation but rather as a correspondence effected through Hecke operators, we can derive the ζ-function defined by Hecke operators. When the Hecke operator T_n is defined with respect to a discontinuous group Γ and we have a representation space $\mathfrak{M}$ of the Hecke operator ring $\mathscr{H}$, we denote the matrix of the operation of $T_n\in\mathscr{H}$ on $\mathfrak{M}$ by $(T_n)=(T_n)_{\mathfrak{M}}$ and call the matrix-valued function $\sum_n(T_n)_{\mathfrak{M}}n^{-s}$ the ζ-function defined by Hecke operators. The equation $\varphi(s)=\sum a_n n^{-s}$ is a specific instance of the correspondence in the first sentence, where $\Gamma=\Gamma(N)$, $\mathfrak{M}\subset\mathfrak{M}_k(\Gamma_0(N))$, $\dim\mathfrak{M}=1$. One advantage of this definition is that it may be applied whenever the concept of Hecke operators can be defined with respect to the group Γ (for instance, even for the Fuchsian group without a †cusp). Thus when Γ is a Fuchsian group given by the unit group of a quaternion algebra $\mathbf{\Phi}$ over the rational number field $\mathbf{Q}$ and $\mathfrak{M}$ is the space of automorphic forms with respect to Γ, the ζ-function $\sum(T_n)n^{-s}$ is defined (Eichler). Moreover, by using its integral expression over the idele group $\mathbf{J}_{\Phi}$ of Φ, we can obtain its functional equation following the Iwasawa-Tate method (Shimura). Furthermore, by algebrogeometric consideration of T_n, it can be shown that

$$\zeta(s)\zeta(s-1)\det(\textstyle\sum(T_n)_{\mathfrak{S}_2}n^{-s})$$
$$=\zeta(s)\zeta(s-1)\det\left(\prod_p(1-(T_p)_{\mathfrak{S}_2}p^{-s}\right.$$
$$\left.+(R_p)_{\mathfrak{S}_2}p^{1-2s})^{-1}\right)$$

coincides (up to a trivial factor) with the Hasse ζ-function of some model of the Riemann surface defined by Γ when $\mathfrak{M}$ is the space $\mathfrak{S}_2$ of all †cusp forms of weight 2 (Eichler [E1], Shimura [S12]).

The algebrogeometric meaning of $\det(\sum(T_n)_{\mathfrak{S}_k}n^{-s})$, when $\mathfrak{M}$ is the space $\mathfrak{S}_k$ of all cusp forms of weight k, has been made clear for the case where Γ is obtained from $\Gamma_0(N)$, $\Gamma(N)$, and the quaternion algebra (M. Kuga, M. Sato, Shimura, Y. Ihara). From these facts, it becomes possible to express $(T_p)_{\mathfrak{S}_k}$, the decomposition of the prime number p in some type of Galois extension (Shimura [S14], Kuga), in terms of Hecke operators. These works gave the first examples of non-Abelian class field

theory. Note that this type of ζ-function may be regarded as the analog (or generalization) of L-functions of algebraic number fields, as can be seen from the comparison in Table 1.

Table 1

Algebraic number field	k	Ideal group	Character χ	$\Sigma\chi(n)n^{-s}$
	↕	↕	↕	↕
Algebraic group	G	Hecke ring	Representation space $\mathfrak{M}$	$\Sigma(T_n)_{\mathfrak{M}}n^{-s}$

As for special values of ζ-functions defined by Hecke operators, the following fact is known: Let $f(\tau)=\sum a_n q^n \in \mathfrak{M}_k(SL(2,\mathbf{Z}))$ be a common eigenfunction of the Hecke operators, and let $\varphi(s)=\sum a_n n^{-s}$ be the corresponding Dirichlet series. Let K_f be the field generated over the rational number field **Q** by the coefficients a_n of f. Then, for any two integers m and m' satisfying $0<m, m'<k$ and $m\equiv m'\ (\mathrm{mod}\,2)$, the ratio $(R(m):R(m'))$ of the special values of

$$R(s)=\frac{\Gamma(s)}{(2\pi)^s}\varphi(s)=\int_0^{\infty}(f(iy)-a_0)y^{s-1}\,dy$$

at m and m' belongs to the field K_f.

G. Shimura discovered this fact for Ramanujan's function $\Delta(\tau)$ (*J. Math. Soc. Japan*, 11 (1959)), and then Yu. I. Manin generalized it to the above case and, by constructing a p-adic analog of $\varphi(s)$ from it, pointed out the importance of such results [M1]. R. M. Damerell also used such results to study special values of Hecke's L-function with Grössencharakter of an imaginary quadratic field (*Acta Arith.*, 17 (1970), 19 (1971)). Furthermore, Shimura generalized these results to congruence subgroups of $SL(2,\mathbf{Z})$ (*Comm. Pure Appl. Math.*, 29 (1976)), and to Hilbert modular groups (*Ann. Math.*, 102 (1975)). The connection between special values of ζ-functions and the periods of integrals has been studied further by Shimura, Deligne, and others.

In addition, in connection with nonholomorphic automorphic forms H. Maass considered L-functions of real quadratic fields (with class characters) having $\Gamma(s/2)^2$ or $\Gamma((s+1)/2)^2$ as Γ-factors. Furthermore, T. Kubota studied the relation of ζ-functions $\zeta_k(s)$ of an arbitrary algebraic field k or ζ-functions of simple rings to (nonanalytic) automorphic forms of several variables and considered the reciprocity law for the Gaussian sum from a new viewpoint.

N. L-Functions of Automorphic Representations (I)

R. P. Langlands reconstructed the theory of †Hecke operators from the viewpoint of representation theory and defined very general L-functions. He proposed many conjectures about them in [L4], and he and H. Jacquet proved most of them in [J1] for the case $G=GL_2$.

First Langlands defined the L-group ${}^LG$ for any connected reductive algebraic group G defined over a field k in the following manner [B6].

There is a canonical bijection between isomorphism classes of connected †reductive algebraic groups defined over a fixed algebraically closed field $\bar{k}$ and isomorphism classes of †root systems. It is defined by associating to G the root data $\Psi(G)=(X^*(T),\Phi,X_*(T),\Phi^\vee)$, where T is a †maximal torus of G, $X^*(T)$ ($X_*(T)$) the group of characters (†1-parameter subgroups) of T, Φ ($\Phi^\vee$) the set of roots (coroots) of G with respect to T.

Since the choice of a †Borel subgroup B of G containing T is equivalent to that of a basis Δ of Φ, the aforementioned bijection yields one between isomorphism classes of triples (G,B,T) and isomorphism classes of based root data $\Psi_0(G)=(X^*(T),\Delta,X_*(T),\Delta^\vee)$. There is a split exact sequence

$$1\to\mathrm{Int}\,G\to\mathrm{Aut}\,G\to\mathrm{Aut}\,\Psi_0(G)\to 1.$$

and this mapping induces a canonical bijection $\mathrm{Aut}\,\Psi_0(G)\cong\mathrm{Aut}(G,B,T,\{x_\alpha\}_{\alpha\in\Delta})$ if $x_\alpha\in G_\alpha$ $(\alpha\in\Delta)$ are fixed.

Let G be a connected reductive algebraic group defined over $\bar{k}$. Let T be a maximal torus of G, and let B be a Borel subgroup of G containing T. Let $\Psi_0(G)=(X^*(T),\Delta,X_*(T),\Delta^\vee)$ be as before. Then there is a connected reductive algebraic group ${}^LG^0$ over **C** such that $\Psi_0(G)^\vee=(X_*(T),\Delta^\vee,X^*(T),\Delta)$ corresponds to the triple $({}^LG^0,{}^LB^0,{}^LT^0)$, where ${}^LB^0$ and ${}^LT^0$ are a Borel subgroup of ${}^LG^0$ and the maximal torus of ${}^LB^0$. For example, (1) if $G=GL_n$, then ${}^LG^0=GL_n(\mathbf{C})$; (2) if $G=Sp_{2n}$, then ${}^LG^0=SO_{2n+1}(\mathbf{C})$.

We assume that $\bar{k}$ is the algebraic closure of k and G is defined over k. Then $\gamma\in\mathrm{Gal}(\bar{k}/k)$ induces an automorphism of the $\bar{k}$-group $G\times_k\bar{k}$. Hence γ defines an element of $\mathrm{Aut}({}^LG^0,{}^LB^0,{}^LT^0)$ because it is a holomorphic image of $\mathrm{Aut}\,\Psi_0(G\times_k\bar{k})=\mathrm{Aut}\,\Psi_0(G\times_k\bar{k})^\vee$. Hence we can define the †semidirect product ${}^LG={}^LG^0\rtimes\mathrm{Gal}(\bar{k}/k)$, and call it the **L-group** of G.

Let k be a †local field, and let G be a connected reductive algebraic group defined over k. We identify G with the group of its k-rational points. Let W_k' be the Weil-Deligne group of k (→ Section H), and let $\Phi(G)$ be the set of homomorphisms $\varphi: W_k'\to{}^LG$ over $\mathrm{Gal}(\bar{k}/k)$. Let $\Pi(G)$ be the set of infinitesimal equivalence classes of irreducible **admissible** representations of G. If k is a non-Archimedean field, then

$\Pi(G)$ is the set consisting of equivalence classes of irreducible representations $\pi: G \to \mathrm{Aut}\, V$ on complex vector spaces V such that the space V^K of vectors invariant by K is finite dimensional for every compact open subgroup K of G and such that $V = \bigcup V^K$, where K runs over the compact open subgroups of G. If k is an Archimedean field, then $\Pi(G)$ is the set consisting of equivalence classes of representations π of the pair $(\mathfrak{g}, K)$ of the Lie algebra $\mathfrak{g}$ of G and a maximal compact subgroup K satisfying similar conditions [B6]. Then Langlands conjectured that we can parametrize $\Pi(G)$ by $\Phi(G)$ as $\Pi(G) = \bigcup_\varphi \Pi(G)_\varphi$. Let $\pi \in \Pi(G)_\varphi$ $(\varphi \in \Phi(G))$, and let r be a representation of ${}^L G$. Then we can define the **L-function** $L(s, \pi, r)$ and the **ε-factor** $\varepsilon(s, \pi, r)$ of π by

$$L(s, \pi, r) = L(s, r \circ \varphi), \quad \varepsilon(s, \pi, r) = \varepsilon(s, r \circ \varphi, \psi),$$

where the right-hand sides are those of the Weil-Deligne group (→ Section H) and ψ is a nontrivial character of k.

Let G be a connected reductive group over a global field k (i.e., an algebraic number field of finite degree or an algebraic function field of one variable over a finite field), let π be an irreducible admissible representation of $G_{\mathbf{A}}$, where $G_{\mathbf{A}}$ is the group of rational points of G over the †adele ring $k_{\mathbf{A}}$ of k, and let r be a finite-dimensional representation of ${}^L G$. Let ψ be a nontrivial character of $k_{\mathbf{A}}$ which is trivial on k. For any place v of k, let r_v be the representation of the L-group of $G_v = G \times_k k_v$ induced by r, and let ψ_v be the additive character of k_v associated with ψ. It is known that π is decomposed into the tensor product $\bigotimes \pi_v$ of $\pi_v \in \Pi(G(k_v))$ [B6]. Hence we put

$$L(s, \pi, r) = \prod_v L(s, \pi_v, r_v),$$

$$\varepsilon(s, \pi, r) = \prod_v \varepsilon(s, \pi_v, r_v).$$

The local factor $L(s, \pi_v, r_v)$ is in fact defined if v is Archimedean, or G is a †torus, or φ is unramified (i.e., G_v is quasisplit and splits over an unramified extension of k_v, and $G(\mathfrak{o}_v)$ is a special maximal compact subgroup of $G(k_v)$, and π_v is of class one with respect to $G(\mathfrak{o}_v)$, where $\mathfrak{o}_v$ is the integer ring of k_v). It follows that the right-hand side $\prod L(s, \pi_v, r_v)$ is defined up to a finite number of non-Archimedean places v. Furthermore, Langlands proved that $\prod \varepsilon(s, \pi_v, r_v)$ is in fact a finite product, and the infinite product $\prod L(s, \pi_v, r_v)$ converges in some right half-plane if π is automorphic (i.e., if π is a subquotient of the right regular representation of $G_{\mathbf{A}}$ in $G_k \backslash G_{\mathbf{A}}$). It is conjectured that $L(s, \pi, r)$ admits a meromorphic continuation to the whole complex plane and satisfies a functional equation

$$L(s, \pi, r) = \varepsilon(s, \pi, r) L(1 - s, \tilde{\pi}, r)$$

if π is automorphic, where $\tilde{\pi}$ is the †contragredient representation of π. Furthermore, if $G = GL_n$ and r is the standard representation of GL_n, then we can construct $L(s, \pi, r)$ and $\varepsilon(s, \pi, r)$ by generalizing the Iwasawa-Tate method. We can also show in this case that $L(s, \pi, r)$ is entire if π is cuspidal. The conjectures are studied in some other cases [B6].

O. L-Functions of Automorphic Representations (II)

A. Weil generalized the theory of †Hecke operators and the corresponding L-functions to the case of †automorphic forms (for holomorphic and nonholomorphic cases together) of GL_2 over a global field [W9]. Then H. Jacquet and Langlands developed a theory from the viewpoint of †representation theory [J1, J2]. They attached L-functions not to automorphic forms but to †automorphic representations of $GL_2(k)$.

Let k be a non-Archimedean local field, and let $\mathfrak{o}_k$ be the maximal order of k. Let $\mathscr{H}_k$ be the space of functions on $G_k = GL_2(k)$ that are locally constant and compactly supported. Then $\mathscr{H}_k$ becomes an algebra with the convolution product

$$(f_1 * f_2)(h) = \int_{G_k} f_1(g) f_2(g^{-1} h)\, dg,$$

where dg is the †Haar measure of G_k that assigns 1 to the maximal compact subgroup $K_k = GL_2(\mathfrak{o}_k)$. Let π be a representation of $\mathscr{H}_k$ on a complex vector space V. Then we say that π is **admissible** if and only if π satisfies the following two conditions: (1) For every v in V, there is an f in $\mathscr{H}_k$ so that $\pi(f)v = v$; (2) Let σ_i $(i = 1, \ldots, r)$ be a family of inequivalent irreducible finite-dimensional representations of K_k, and let

$$\xi(g) = \sum_{i=1}^{r} \dim(\sigma_i)^{-1} \operatorname{tr} \sigma_i(g^{-1}).$$

Then ξ is an idempotent of $\mathscr{H}_k$. We call such a ξ an **elementary idempotent** of $\mathscr{H}_k$. Then for every elementary idempotent ξ of $\mathscr{H}_k$, the operator $\pi(\xi)$ has a finite-dimensional range. If π is an admissible representation of $GL_2(k)$ (→ Section N), then

$$\pi(f) = \int_{G_k} f(g) \pi(g)\, dg \qquad (f \in \mathscr{H}_k)$$

gives an admissible representation of $\mathscr{H}_k$ in this sense. Furthermore, any admissible representation of $\mathscr{H}_k$ can be obtained from an admissible representation of $GL_2(k)$.

Let k be the real number field. Let $\mathscr{H}_1$ be the

space of infinitely differentiable compactly supported functions on $G_k(=GL_2(k))$ that are $K_k(=O(2,k))$ finite on both sides, let $\mathscr{H}_2$ be the space of functions on K_k that are finite sums of matrix elements of irreducible representations of K_k, and let $\mathscr{H}_k=\mathscr{H}_1\oplus\mathscr{H}_2$. Then $\mathscr{H}_1$, $\mathscr{H}_2$, and $\mathscr{H}_k$ become algebras with the convolution product. Let π be a representation of $\mathscr{H}_k$ on a complex vector space V. Then π is **admissible** if and only if the following three conditions are satisfied: (1) Every vector v in V is of the form $v=\sum_{i=1}^r \pi(f_i)v_i$ with $f_i\in\mathscr{H}_1$ and $v_i\in V$; (2) for every elementary idempotent $\xi(g)=\sum_{i=1}^r \dim(\sigma_i)^{-1}\operatorname{tr}\sigma_i(g^{-1})$, where the σ_i are a family of inequivalent irreducible representations of K_k, the range of $\pi(\xi)$ is finite-dimensional; (3) for every elementary idempotent ξ of $\mathscr{H}_k$ and for every vector v in $\pi(\xi)V$, the mapping $f\mapsto\pi(f)v$ of $\xi\mathscr{H}_1\xi$ into the finite-dimensional space $\pi(\xi)V$ is continuous. We can define the Hecke algebra $\mathscr{H}_k$ and the notion of admissible representations also in the case $k=\mathbf{C}$. In these cases, an admissible representation of $\mathscr{H}_k$ comes from a representation of the †universal enveloping algebra of $GL_2(k)$ but may not come from a representation of $GL_2(k)$. It is known that for any local field k, the †character of each irreducible representation is a locally integrable function.

Let k be a global field, $G_k=GL_2(k)$, and let $G_{\mathbf{A}}=GL_2(k_{\mathbf{A}})$ be the group of rational points of G_k over the adele ring $k_{\mathbf{A}}$ of k. For any place v of k, let k_v be the completion of k at v, let $G_v=GL_2(k_v)$, and let k_v be the standard maximal compact subgroup of G_v. Let $\mathscr{H}_v$ be the Hecke algebra $\mathscr{H}_{k_v}$ of G_v, and let ε_v be the normalized Haar measure of K_v. Then ε_v is an elementary idempotent of $\mathscr{H}_v$. Let $\mathscr{H}=\bigotimes_{\varepsilon_v}\mathscr{H}_v$ be the restricted tensor product of the local Hecke algebra $\mathscr{H}_v$ with respect to the family $\{\varepsilon_v\}$. We call $\mathscr{H}$ the **global Hecke algebra** of G_A.

Let π be a representation of $\mathscr{H}$ on a complex vector space V. We define the notion of admissibility of π as before. Then we can show that, for any irreducible admissible representation π of $\mathscr{H}$ and for any place v of k, there exists an irreducible admissible representation π_v of $\mathscr{H}_v$ on a complex vector space V_v such that (1) for almost all v, $\dim V_v^{K_v}=1$ and (2) π is equivalent to the restricted tensor product $\bigotimes\pi_v$ of the π_v with respect to a family of nonzero $x_v\in V_v^{K_v}$. Furthermore, the factors $\{\pi_v\}$ are unique up to equivalence.

Let k be a local field, let ψ be a nontrivial character of k, and let $\mathscr{H}_k$ be the Hecke algebra of $G_k=GL_2(k)$. Let π be an infinite-dimensional admissible irreducible representation of $\mathscr{H}_k$. Then there is exactly one space $W(\pi,\psi)$ of continuous functions on G_k with the following three properties: (1) If W is in $W(\pi,\psi)$, then for all g in G_k and for all x in k,

$$W(\begin{pmatrix}1 & x\\ 0 & 1\end{pmatrix}g)=\psi(x)W(g);$$

(2) $W(\pi,\psi)$ is invariant under the right translations of $\mathscr{H}_k$, and the representation on $W(\pi,\psi)$ is equivalent to π; (3) if k is Archimedean and if W is in $W(\pi,\psi)$, then there is a positive number N such that

$$W(\begin{pmatrix}t & 0\\ 0 & 1\end{pmatrix})=O(|t|^N)$$

as $|t|\to\infty$. We call $W(\pi,\psi)$ the **Whittaker model** of π. The Whittaker model exists in the global case if and only if each factor π_v of $\pi=\bigotimes\pi_v$ is infinite-dimensional.

Let k be a local field, and let π be as before. Then the L-function $L(s,\pi)$ and the ε-factor $\varepsilon(s,\pi,\psi)$ are defined in the following manner: Let ω be the quasicharacter of $k^\times$ (i.e., the continuous homomorphism $k^\times\to\mathbf{C}^\times$) defined by

$$\pi(\begin{pmatrix}a & 0\\ 0 & a\end{pmatrix})=\omega(a)\,id_V.$$

Then the †contragredient representation $\tilde{\pi}$ of π is equivalent to $\omega^{-1}\otimes\pi$. For any g in G_k and W in $W(\pi,\psi)$, let

$$\Psi(g,s,W)=\int_{k^\times}W(\begin{pmatrix}a & 0\\ 0 & 1\end{pmatrix}g)|a|^{s-1/2}\,d^\times a,$$

$$\tilde{\Psi}(g,s,W)=\int_{k^\times}W(\begin{pmatrix}a & 0\\ 0 & 1\end{pmatrix}g)|a|^{s-1/2}\omega^{-1}(a)\,d^\times a.$$

Then there is a real number s_0 such that these integrals converge for $\mathrm{Re}(s)>s_0$ for any $g\in G_k$ and $W\in W(\pi,\psi)$. If k is a non-Archimedean local field with $\mathbf{F}_q$ as its residue field, then there is a unique factor $L(s,\pi)$ such that $L(s,\pi)^{-1}$ is a polynomial of q^{-s} with constant term 1,

$$\Phi(g,s,W)=\Psi(g,s,W)/L(s,\pi)$$

is a holomorphic function of s for all g and W, and there is at least one W in $W(\pi,\psi)$ so that $\Phi(e,s,W)=a^s$ with a positive constant a. If k is an Archimedean local field, then we can define the gamma factor $L(s,\pi)$ in the same manner. Furthermore, for any local field k, if

$$\tilde{\Phi}(g,s,W)=\tilde{\Psi}(g,s,W)/L(s,\tilde{\pi}),$$

then there is a unique factor $\varepsilon(s,\psi,\pi)$ which, as a function of s, is an exponential such that

$$\tilde{\Phi}(\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}g,1-s,W)=\varepsilon(s,\pi,\psi)\Phi(g,s,W)$$

for all $g\in G_k$ and $W\in W(\pi,\psi)$.

Let π and π' be two infinite-dimensional irreducible admissible representations of G_k. Then π and π' are equivalent if and only if the

quasicharacters ω and ω' are equal and

$$\frac{L(1-s,\chi^{-1}\otimes\tilde{\pi})\varepsilon(s,\chi\otimes\pi,\psi)}{L(s,\chi\otimes\pi)} = \frac{L(1-s,\chi^{-1}\otimes\tilde{\pi}')\varepsilon(s,\chi\otimes\pi',\psi)}{L(s,\chi\otimes\pi')}$$

holds for any quasicharacter χ. In particular, the set $\{L(s,\chi\otimes\pi)$ and $\varepsilon(s,\chi\otimes\pi,\psi)$ for all $\chi\}$ characterizes the representation π.

Let k be a global field, $G_k=GL_2(k)$, $G_{\mathbf{A}}=GL_2(k_{\mathbf{A}})$, and let $K_{\mathbf{A}}=\prod K_v$ be the standard maximal compact subgroup of $G_{\mathbf{A}}$. Then the †global Hecke algebra $\mathscr{H}$ acts on the space of continuous functions on $G_k\backslash G_{\mathbf{A}}$ by the right translations. Let φ be a continuous function on $G_k\backslash G_{\mathbf{A}}$. Then φ is an **automorphic form** if and only if (1) φ is $K_{\mathbf{A}}$-finite on the right, (2) for every †elementary idempotent ξ in $\mathscr{H}$, the space $(\xi\mathscr{H})\varphi$ is finite-dimensional, and (3) φ is slowly increasing if k is an algebraic number field. An automorphic form φ is a **cusp form** if and only if

$$\int_{k\backslash k_{\mathbf{A}}}\varphi(\begin{pmatrix}1 & x\\ 0 & 1\end{pmatrix}g)\,dx=0$$

for all g in $G_{\mathbf{A}}$. Let $\mathscr{A}$ be the space of automorphic forms on $G_k\backslash G_{\mathbf{A}}$, and let $\mathscr{A}_0$ be the space of cusp forms on $G_k\backslash G_{\mathbf{A}}$. They are $\mathscr{H}$-modules. Let $\psi=\prod\psi_v$ be a nontrivial character of $k\backslash k_{\mathbf{A}}$, and let π be an irreducible admissible representation $\pi=\bigotimes_v\pi_v$ of the global Hecke algebra $\mathscr{H}=\bigotimes_{\varepsilon_v}\mathscr{H}_v$. If π is a †constituent of the $\mathscr{H}$-module $\mathscr{A}$, then we can define the local factors $L(s,\pi_v)$ and $\varepsilon(s,\pi_v,\psi_v)$ for all v, although π_v may not be infinite-dimensional. Further, the infinite products

$$L(s,\pi)=\prod L(s,\pi_v) \text{ and } L(s,\tilde{\pi})=\prod L(s,\tilde{\pi}_v)$$

converge absolutely in a right half-plane, and the functions $L(s,\pi)$ and $L(s,\tilde{\pi})$ can be analytically continued to the whole complex plane as meromorphic functions of s. If π is a constituent of $\mathscr{A}_0$, then all π_v are infinite-dimensional, $L(s,\pi)$ and $L(s,\tilde{\pi})$ are entire functions, and π is contained in $\mathscr{A}_0$ with multiplicity one. If k is an algebraic number field, then they have only a finite number of poles and are bounded at infinity in any vertical strip of finite width. If k is an algebraic function field of one variable with field of constant $\mathbf{F}_q$, then they are rational functions of q^{-s}. In either case, $\varepsilon(s,\pi_v,\psi_v)=1$ for almost all v, and hence

$$\varepsilon(s,\pi)=\prod\varepsilon(s,\pi_v,\psi_v)$$

is well defined. Furthermore, the functional equation

$$L(s,\pi)=\varepsilon(s,\pi)\,L(1-s,\tilde{\pi})$$

is satisfied.

As for the condition for π being a constituent of $\mathscr{A}_0$, we have the following: Let $\pi=\bigotimes\pi_v$ be an irreducible admissible representation of $\mathscr{H}$. Then π is a constituent of $\mathscr{A}_0$ if and only if (1) for every v, π_v is infinite-dimensional; (2) the quasicharacter η defined by

$$\pi(\begin{pmatrix}a & 0\\ 0 & a\end{pmatrix})=\eta(a)id.$$

is trivial on $k^\times$; (3) π satisfies a certain condition so that, for any quasicharacter ω of $k^\times\backslash k_{\mathbf{A}}^\times$, $L(s,\omega\otimes\pi)=\prod L(s,\omega_v\otimes\pi_v)$ and $L(s,\omega^{-1}\otimes\tilde{\pi}_v)=\prod L(s,\omega_v^{-1}\otimes\tilde{\pi}_v)$ converge on a right half-plane; and (4) for any quasicharacter ω of $k^\times\backslash k_{\mathbf{A}}^\times$, $L(s,\omega\otimes\pi)$ and $L(s,\omega^{-1}\otimes\tilde{\pi})$ are entire functions of s which are bounded in vertical strips and satisfy the functional equation

$$L(s,\omega\otimes\pi)=\varepsilon(s,\omega\otimes\pi)\,L(1-s,\omega^{-1}\otimes\tilde{\pi}).$$

P. Congruence ζ-Functions of Algebraic Function Fields of One Variable or of Algebraic Curves

Let K be an †algebraic function field of one variable over $k=\mathbf{F}_q$ (finite field with q elements). The **ζ-function of the algebraic function field** K/k, denoted by $\zeta_K(s)$, is defined by the infinite sum $\sum_{\mathfrak{A}}N(\mathfrak{A})^{-s}$, where the summation is over all integral divisors $\mathfrak{A}$ of K/k and where the norm $N(\mathfrak{A})$ equals $q^{\deg(\mathfrak{A})}$. Equivalently, $\zeta_K(s)$ is defined by the infinite product $\prod_{\mathfrak{p}}(1-N(\mathfrak{p})^{-s})^{-1}$, where $\mathfrak{p}$ runs over all prime divisors of K/k. By the change of variable $u=q^{-s}$, $\zeta_K(s)=Z_K(u)$ becomes a formal power series in u. $\zeta_K(s)$ and $Z_K(u)$ are sometimes called the **congruence ζ-functions** of K/k.

The fundamental theorem states that (i) (Rationality) $Z_K(u)$ is a rational function of u of the form $Z_K(u)=P(u)/(1-u)(1-qu)$, where $P(u)\in\mathbf{Z}[u]$ is a polynomial of degree $2g$, g being the genus of K; (ii) (Functional equation) $Z_K(u)$ satisfies the functional equation

$$Z_K(1/u)=q^{g-1}u^{2-2g}Z_K(u/q);$$

and (iii) if $P(u)$ is decomposed into linear factors in $\mathbf{C}[u]$: $P(u)=\prod_{i=1}^{2g}(1-\alpha_iu)$, then all the reciprocal roots α_i are complex numbers of absolute value $\sqrt{q}$. Statement (iii) is the analog of the **Riemann hypothesis** because it is equivalent to saying that all the zeros of $\zeta_K(s)=Z_K(q^{-s})$ lie on the line $\mathrm{Re}\,s=1/2$.

The congruence ζ-function was introduced by E. Artin [A1 (1924)] as an analog of the Riemann or Dedekind ζ-functions. Of its fundamental properties, the rationality (i) and the functional equation (ii) were proven by

F. K. Schmidt (1931), using the †Riemann-Roch theorem for the function field K/k. The Riemann hypothesis (iii) was verified first in the elliptic case ($g=1$) by H. Hasse [H1] and then in the general case by A. Weil [W2 (1948)]. For the proof of (iii), it was essential to consider the geometry of algebraic curves that correspond to given function fields.

Let C be a nonsingular complete curve over k with function field K. Then $Z_K(u)$ coincides with the ζ-function of C/k, denoted by $Z(u, C)$, which is defined by the formal power series $\exp(\sum_{m=1}^{\infty} N_m u^m/m)$. Here N_m is the number of rational points of C over the extension k_m of k of degree m. The rationality of $Z_K(u)$ is then equivalent to the formula

$$N_m = 1 + q^m - \sum_{i=1}^{2g} \alpha_i^m \qquad (m \in \mathbf{N}),$$

and the Riemann hypothesis for $Z_K(u)$ is equivalent to the estimate

$$(*) \quad |N_m - 1 - q^m| \leqslant 2g\, q^{m/2} \qquad (m \in \mathbf{N}).$$

Now if F is the qth power morphism of C to itself (the **Frobenius morphism** of C relative to k), then an important observation is that N_m is the number of fixed points of the mth iterate F^m of F. In other words, N_m is equal to the intersection number of the graph of F^m with the diagonal on the surface $C \times C$, and is related to the "trace" of the Frobenius correspondence. Then (*) follows from †Castelnuovo's lemma in the theory of correspondences on a curve. This is Weil's proof of the Riemann hypothesis in [W2]; compare the proof by A. Mattuck and J. Tate (*Abh. Math. Sem. Hamburg* 22 (1958)) and A. Grothendieck (*J. Reine Angew. Math.*, 200 (1958)) using the Riemann-Roch theorem for an algebraic surface.

On the other hand, let J be the †Jacobian variety of C over k. For each prime number l different from the characteristic of k, let $M_l(\alpha)$ denote the †l-adic representation of an endomorphism α of J obtained from its action on points of J of order l^n ($n = 1, 2, \ldots$). Letting π be the endomorphism of J induced from F (which is the same as the Frobenius morphism of J), we have $P(u) = \det(1 - M_l(\pi)u)$, i.e., the numerator of the ζ-function coincides with the characteristic polynomial of $M_l(\pi)$. In this setting, the Riemann hypothesis is a consequence of the positivity of the Rosati antiautomorphism [E1]. This is the second proof given by Weil [W2], and applies to arbitary Abelian varieties.

Recently E. Bombieri, inspired by Stepanov's idea, gave an elementary proof of (*) using only the Riemann-Roch theorem for a curve (*Sém. Bourbaki*, no. 430 (1973)).

Q. ζ-Functions of Algebraic Varieties over Finite Fields

Let V be an algebraic variety over the finite field with q elements $\mathbf{F}_q$, and let N_m be the number of $\mathbf{F}_q{}^m$-rational points of V. Then the **ζ-function of V over $\mathbf{F}_q$** is the formal power series in $\mathbf{Z}[[u]]$ defined by

$$Z(u, V) = \exp\left(\sum_{m=1}^{\infty} N_m u^m/m\right);$$

alternatively it can be defined by the infinite product $\prod_P (1 - u^{\deg P})^{-1}$, where P runs over the set of prime divisors of V and $\deg P$ is the degree of the residue field of P over $\mathbf{F}_q$ (in other words, P runs over prime rational 0-cycles of V over $\mathbf{F}_q$).

Weil Conjecture. In 1949, the following properties of the ζ-function were conjectured by Weil [W3]. Let V be an n-dimensional complete nonsingular (absolutely irreducible) variety over $\mathbf{F}_q$. Then (1) $Z(u, V)$ is a rational function of u. (2) $Z(u, V)$ satisfies the functional equation

$$Z((q^n u)^{-1}, V) = \pm q^{n\chi/2} u^{\chi} Z(u, V),$$

where the integer χ is the intersection number (the degree of $\Delta_V \cdot \Delta_V$) of the diagonal subvariety Δ_V with itself in the product $V \times V$, which is called the Euler-Poincaré characteristic of V. (3) Moreover, we have

$$Z(u, V) = \frac{P_1(u) \cdot P_3(u) \cdot \ldots \cdot P_{2n-1}(u)}{P_0(u) \cdot P_2(u) \cdot \ldots \cdot P_{2n}(u)},$$

where $P_h(u) = \prod_{j=1}^{B_h} (1 - \alpha_j^{(h)} u)$ is a polynomial with $\mathbf{Z}$-coefficients such that $\alpha_j^{(h)}$ are algebraic integers of absolute value $q^{h/2}$ ($0 \leqslant h \leqslant 2n$); the latter statement is the **Riemann hypothesis for** $V/\mathbf{F}_q$. (4) When V is the reduction $\bmod\, p$ of a complete nonsingular variety V^* of characteristic 0, then the degree B_h of $P_h(u)$ is the hth Betti number of V^* considered as a complex manifold.

This conjecture, called the **Weil conjecture**, has been completely proven. To give a brief history, first B. Dwork [D13] proved the rationality of the ζ-function for any (not necessarily complete or nonsingular) variety over $\mathbf{F}_q$. Then A. Grothendieck [A3, G2, G3] developed the l-adic étale cohomology theory with M. Artin and others, and proved the above statements (1)–(4) (except for the Riemann hypothesis) with $P_h(u)$ replaced by some $P_{h,l}(u) \in \mathbf{Q}_l[u]$; and S. Lubkin [L7] obtained similar results for liftable varieties. Finally Deligne [D4] proved the Riemann hypothesis and the independence of l of $P_{h,l}(u)$. More details will be given below. Before the final solution for the general case was obtained, the

conjecture had been verified for some special types of varieties. For curves and Abelian varieties, its truth was previously shown by Weil (→ Section P). In the paper [W3] in which he proposed the above conjecture, Weil verified it for Fermat hypersurfaces, i.e., those defined by the equation $a_0x_0^m+\ldots+a_{n+1}x_{n+1}^m=0$ ($a_i\in\mathbf{F}_q^\times$); in this case, the ζ-function is of the form $P(u)^{(-1)^{n+1}}/\prod_{j=0}^{n}(1-q^ju)$ with a polynomial $P(u)$ that can be explicitly described in terms of Jacobi sums. Dwork [D14] studied by p-adic analysis the case of hypersurfaces in a projective space, verifying the conjecture for them except for the Riemann hypothesis. Further nontrivial examples were provided by †$K3$ surfaces (Deligne [D2], Pyatetskiĭ-Shapiro, Shafarevich [P1]) and cubic 3-folds (E. Bombieri, H. Swinnerton-Dyer [B5]); in these cases the proof of the Riemann hypothesis was reduced to that of certain Abelian varieties naturally attached to these varieties. It can be said that the Weil conjecture has greatly influenced the development of algebraic geometry, as regards both the foundations and the methods of proof of the conjecture itself; see the expositions by N. Katz [K2] or B. Mazur [M2].

Weil Cohomology, l-Adic Cohomology. The Weil conjecture suggested the possibility of a good cohomology theory for algebraic varieties over a field of arbitrary characteristic. We first formulate the desired properties of a good cohomology (S. Kleiman [K4]). Let $\bar{k}$ be an algebraically closed field and K a field of characteristic 0, which is called the coefficient field. A contravariant functor $V\to H^*(V)$ from the category of complete connected smooth varieties over $\bar{k}$ to the category of augmented $\mathbf{Z}^+$-graded finite-dimensional anticommutative K-algebras (cup product as multiplication) is called a **Weil cohomology** with coefficients in K if it has the following three properties. (1) **Poincaré duality**: If $n=\dim V$, then a canonical isomorphism $H^{2n}(V)\cong K$ exists and the cup product $H^j(V)\times H^{2n-j}(V)\to H^{2n}(V)\cong K$ induces a perfect pairing. (2) **Künneth formula**: For any V_1 and V_2 the mapping $H^*(V_1)\otimes H^*(V_2)\to H^*(V_1\times V_2)$ defined by $a\otimes b\to \mathrm{Proj}_1^*(a)\cdot\mathrm{Proj}_2^*(b)$ is an isomorphism. (3) Good relation with **algebraic cycles**: Let $C^j(V)$ be the group of algebraic cycles of codimension j on V. There exists a fundamental-class homomorphism FUND: $C^j(V)\to H^{2j}(V)$ for all j, which is functorial in V, compatible with products via Künneth's formula, has compatibility of the intersection with the cup product, and maps 0-cycle$\in C^n(V)$ to its degree as an element of $K\cong H^{2n}(V)$. If a Weil cohomology theory H exists for the V's over $\bar{k}$, we can prove the **Lefschetz fixed-point formula**:

$$((\text{graph of } F)\cdot(\text{diagonal}))_{V\times V}$$

$$=\sum_{j=0}^{2n}(-1)^j\mathrm{tr}(F^*\mid H^j(V))$$

for a morphism $F: V\to V$.

If $\bar{k}=\mathbf{C}$ (the field of complex numbers), the classical cohomology $V\to H^*(V^{an},\mathbf{Q})$, where V^{an} denotes the complex manifold associated with V, gives a Weil cohomology. If $\bar{k}$ is an arbitrary algebraically closed field and if l is a prime number different from the characteristic of $\bar{k}$, then the principal results in the theory of the étale cohomology state that the l-adic cohomology $V\to H_{\mathrm{et}}^*(V,\mathbf{Q}_l)$ is a Weil cohomology with coefficient field $\mathbf{Q}_l$ (the field of l-adic numbers) [A3, D5, G3, M4]. In defining this, Grothendieck introduced a new concept of topology, which is now called Grothendieck topology. In the étale topology of a variety V, for example, any étale covering of a Zariski open subset is regarded as an "open set." With respect to the étale topology, the cohomology group $H^*(V,\mathbf{Z}/n)$ of V with coefficients in $\mathbf{Z}/n$ is defined in the usual manner and is a finite $\mathbf{Z}/n$-module. If l is a prime number as above, $\varprojlim_\nu H^*(V,\mathbf{Z}/l^\nu)$ is a module over $\mathbf{Z}_l=\varprojlim_\nu\mathbf{Z}/l^\nu$ of finite rank, and

$$H_{\mathrm{et}}^*(V,\mathbf{Q}_l)=(\varprojlim_\nu H^*(V,\mathbf{Z}/l^\nu))\otimes_{\mathbf{Z}_l}\mathbf{Q}_l$$

defines the l-adic cohomology group, giving rise to a Weil cohomology.

For the characteristic p of k, p-adic étale cohomology does not give Weil cohomology; but the crystalline cohomology (Grothendieck and P. Berthelot [B2, B3]) takes the place of p-adic cohomology and is almost a Weil cohomology: in this theory the fundamental class is defined only for smooth subvarieties.

Now fix a Weil cohomology for $\bar{k}=\bar{\mathbf{F}}_q$, an algebraic closure of a finite field $\mathbf{F}_q$. Given an algebraic variety V over $\mathbf{F}_q$, let $\bar{V}=V\otimes\bar{k}$ denote the base extension of V to $\bar{k}$; then $\mathbf{F}_{q^m}$-rational points of V can be identified with the fixed points of the mth iterate of the Frobenius morphism F of V relative to $\mathbf{F}_q$. Then the Lefschetz fixed-point formula implies the rationality of $Z(u,V)$; more precisely, letting $P_j(u)=\det(1-uF^*\mid H^j(\bar{V}))$ be the characteristic polynomial of the automorphism F^* of $H^j(\bar{V})$ induced by F, we have

$$Z(u,V)=\prod_{j=0}^{2n}P_j(u)^{(-1)^{j+1}}.$$

The functional equation of the ζ-function then follows from the Poincaré duality. This proves (1), (2), and a part of (3) in the statement of the Weil conjecture. Further, in the case of l-adic cohomology, (4) means that $\deg P_j(u)=$

$\dim_{\mathbf{Q}_l} H^j(\bar{V}, \mathbf{Q}_l)$ is equal to the jth Betti number of a lifting of V to characteristic 0; this follows from the comparison theorem of M. Artin for the l-adic cohomology and the classical cohomology, combined with the invariance of l-adic cohomology under specialization.

Proof of the Riemann Hypothesis. In 1974, Deligne [D4, I] completed the proof of the Weil conjecture for projective nonsingular varieties by proving that, given such a V over $\mathbf{F}_q$, any eigenvalue of F^* on $H^j_{\text{ét}}(\bar{V}, \mathbf{Q}_l)$ is an algebraic integer, all the conjugates of which are of absolute value $q^{j/2}$. (This implies that $P_j(u) = \det(1 - uF^* \mid H^j_{\text{ét}}(\bar{V}, \mathbf{Q}_l))$ is in $\mathbf{Z}[u]$ and is independent of l.) The proof is done by induction on $n = \dim V$; by the general results in l-adic cohomology (the weak Lefschetz theorem on a hyperplane section, the Poincaré duality, and the Künneth formula), the proof is reduced to the assertion that (*) any eigenvalue α of F^* on $H^n_{\text{ét}}(\bar{V}, \mathbf{Q}_l)$ is an algebraic integer such that $|\alpha'| \leqslant q^{(n+1)/2}$ for all conjugates α' of α. The main ingredients in proving (*) are (1) Grothendieck's theory of L-functions, based on the étale cohomology with compact support and with coefficients in a $\mathbf{Q}_l$-sheaf [G2, G3]; (2) the theory of Lefschetz pencils (Deligne and Katz [D7]), and the Kajdan-Margulis theorem on the monodromy of a Lefschetz pencil (J. L. Verdier, *Sém. Bourbaki*, no. 423 (1972)); and (3) Rankin's methods to estimate the coefficients of modular forms, as adapted to the Grothendieck's L-series. By means of these geometric and arithmetic techniques, Deligne achieved the proof of the Riemann hypothesis for projective nonsingular varieties. For the generalization to complete varieties, see Deligne [D4, II].

Applications of the (Verified) Weil Conjecture. (1) The Ramanujan conjecture (→ 32 Automorphic Functions D): The connection of this conjecture and the Weil conjecture for certain fiber varieties over a modular curve was observed by M. Sato and partially verified by Y. Ihara [I1] and then established by Deligne [D3]. The Weil conjecture as proven above implies the truth of the Ramanujan conjecture and its generalization by H. Petersson.

(2) Estimation of trigonometric sums: Let q be the power of a prime number p. Then

$$\left|\sum_{(x_1,\ldots,x_n)\in\mathbf{F}_q^n} \exp\frac{2\pi i}{p}\operatorname{tr}_{\mathbf{F}_q/\mathbf{F}_p}(F(x_1,\ldots,x_n))\right| \leqslant (d-1)^n q^{n/2},$$

where $F(X_1, \ldots, X_n) \in \mathbf{F}_q[X_1, \ldots, X_n]$ is a polynomial of degree d that is not divisible by p, and the homogeneous part of the highest degree of F defines a smooth irreducible hypersurface in $\mathbf{P}^{n-1}$. This is a generalization of the Weil estimation of the Kloosterman sum ([D4, W1 (1948c)]; → 4 Additive Number Theory D).

(3) The **hard Lefschetz theorem**: Let $L \in H^2(V)$ be the class of a hyperplane section of an n-dimensional projective nonsingular variety V over an algebraically closed field. Then the cup product by $L^i : H^{n-i}(V) \to H^{n+i}(V)$ is an isomorphism for all $i \leqslant n$. Deligne [D4, II] proved this for l-adic cohomology, from which N. Katz and W. Messing [K1] deduced its validity in any Weil cohomology or in the crystalline cohomology.

Also some geometric properties of an algebraic variety V are reflected in the properties of $Z(u, V)$. The ζ-function $Z(u, A)$ of an Abelian variety A determines the isogeny class of A [T4]. For any algebraic integer α, every conjugate of which has absolute value $q^{1/2}$, there exists an Abelian variety $A/\mathbf{F}_q$ such that α is a root of $\det(1 - uF^* \mid H^1(A)) = 0$ [H6]. J. Tate [T3] conjectured that the rank of the space cohomology classes of algebraic cycles of codimension r is equal to the order of the pole at $u = 1/q^r$ of $Z(u, V)$. This conjecture is still open but has been verified in certain nontrivial cases, e.g., (1) products of curves and Abelian varieties, $r = 1$ (Tate [T4]), (2) Fermat hypersurfaces of dimension $2r$ with some condition on the degree and the characteristic (Tate [T3], T. Shioda, *Proc. Japan Acad.* 55 (1979)), and (3) elliptic $K3$ surfaces, $r = 1$ (M. Artin and Swinnerton-Dyer, *Inventiones Math.* 20 (1973)).

R. ζ- and L-Functions of Schemes

Let X be a †scheme of finite type over $\mathbf{Z}$, and let $|X|$ denote the set of closed points of X; for each $x \in |X|$, the residue field $k(x)$ is finite, and its cardinality is called the norm $N(x)$ of x. The **ζ-function of a scheme** X is defined by the product $\zeta(s, X) = \prod_{x \in |X|}(1 - N(x)^{-s})^{-1}$. This converges absolutely for $\operatorname{Re} s > \dim X$, and it is conjectured to have an analytic continuation in the entire s-plane (Serre [S7]). It reduces to the Riemann (resp. Dedekind) ζ-function if $X = \operatorname{Spec}(\mathbf{Z})$ (resp. $\operatorname{Spec}(\mathfrak{o})$, $\mathfrak{o}$ being the ring of integers of an algebraic number field), and to the ζ-function $Z(q^{-s}, X)$ (→ Section Q) if X is a variety over a finite field $\mathbf{F}_q$. The case of varieties defined over an algebraic number field is discussed in Section S.

Let G be a finite group of automorphisms of a scheme X, and assume that the quotient $Y = X/G$ exists (e.g., X is quasiprojective). For an element x in $|X|$, let y be its image in $|Y|$, and let $D(x) = \{g \in G \mid g(x) = x\}$, the decomposition group of x over y. The natural mapping $D(x) \to \operatorname{Gal}(k(x)/k(y))$ is surjective, and its

kernel $I(x)$ is called the inertia group at x. An element of $D(x)$ is called a Frobenius element at x if its image in $\mathrm{Gal}(k(x)/k(y))$ corresponds to the $N(y)$th-power automorphism of $k(x)$. Now let R be a representation of G with character χ. The **Artin L-function** $L(s, X, \chi)$ is defined by

$$L(s,X,\chi)=\exp\left(\sum_{y\in|Y|}\sum_{n=1}^{\infty}\chi(y^n)N(y)^{-ns}/n\right)$$
$$=\prod_{y\in|Y|}\det(1-R(F_y)N(y)^{-s})^{-1},$$

where $\chi(y^n)$ denotes the mean value of χ on the nth power of Frobenius elements F_x at x (x any point of $|X|$ over y), and similarly $R(F_y)$ denotes the mean value of $R(F_x)$; it converges absolutely for $\mathrm{Re}\, s > \dim X$. Again this is reduced to the usual Artin L-function (→ Section G) if X is the spectrum of the ring of integers of an algebraic number field. The Artin L-functions of a scheme have many formal properties analogous to those of Artin L-functions of a number field (Serre [S7]).

Let us consider the case where X is an algebraic variety over a finite field $\mathbf{F}_q$ and elements of G are automorphisms of X over $\mathbf{F}_2$; in this case, $L(s, X, \chi)$ is a formal power series in $u = q^{-s}$, which is called a congruence Artin L-function. For the case where X is a complete nonsingular algebraic curve and χ is an irreducible character of G different from the trivial one, Weil [W2] proved that $L(s, X, \chi)$ is a polynomial in $u = q^{-s}$; thus the analog of †Artin's conjecture holds here. More generally, for any algebraic variety X over $\mathbf{F}_q$, Grothendieck [G2, G3] proved the rationality of L-functions together with the alternating product expression by polynomials in u, as in the case of ζ-functions, by the methods of l-adic cohomology. Actually, Grothendieck treated a more general type of L-function associated with l-adic sheaves on X, which also play an important role in Deligne's proof of the Riemann hypothesis (→ Section Q).

S. Hasse ζ-Functions

For a nonsingular complete algebraic variety V defined over a finite algebraic number field K, let $V_\mathfrak{p}$ be the reduction of V modulo a prime ideal $\mathfrak{p}$ of K, $K_\mathfrak{p}$ be the residue field of $\mathfrak{p}$, and $Z(u, V_\mathfrak{p})$ be the ζ-function of $V_\mathfrak{p}$ over $K_\mathfrak{p}$. The ζ-function $\zeta(s, V)$ of the complex variable s, determined by the infinite product (excluding the finite number of $\mathfrak{p}$'s where $V_\mathfrak{p}$ is not defined),

$$\zeta(s,V)=\prod_{\mathfrak{p}}{}' Z(N(\mathfrak{p})^{-s},V_\mathfrak{p}),$$

is called the **Hasse ζ-function** of V over the algebraic number field K. For this function, we have **Hasse's conjecture** [W4]: $\zeta(s, V)$ is a meromorphic function over the whole complex plane of s and satisfies the functional equation of ordinary type. Sometimes it is more natural to consider

$$\zeta_j(s,V)=\prod_{\mathfrak{p}}{}' P_j(N(\mathfrak{p})^{-s},V_\mathfrak{p})^{-1}\quad(0\leqslant j\leqslant 2\dim V),$$

where $P_j(u, V_\mathfrak{p})$ is the jth factor of $Z(u, V_\mathfrak{p})$, and we have a similar conjecture for them. For the definition of $\zeta_j(s, V)$ taking into account the factors for bad primes and the precise form of the conjectural functional equation, see Serre [S8]. Note that $\zeta_j(s, V)$ converges absolutely for $\mathrm{Re}\, s > j/2+1$ as a consequence of the Weil conjecture.

Hasse's conjecture remains unsolved for the general case, but has been verified when V is one of the following varieties:

($\mathrm{I_a}$) Algebraic curves defined by the equation $y^e = \gamma x^f + \delta$ and Fermat hypersurfaces (Weil [W6]).

($\mathrm{I_b}$) Elliptic curves with complex multiplication (Deuring [D11]).

($\mathrm{I_c}$) Abelian varieties with complex multiplication (Taniyama [T2], Shimura and Taniyama [S11], Shimura, H. Yoshida).

($\mathrm{I_d}$) Singular $K3$ surfaces, i.e., $K3$ surfaces with 20 Picard numbers (Shafarevich and Pyatetskiĭ-Shapiro [P1], Deligne [D2], T. Shioda and H. Inose [S21]).

($\mathrm{II_a}$) Algebraic curves that are suitable models of the elliptic modular function fields (Eichler [E1], Shimura [S12]).

($\mathrm{II_b}$) Algebraic curves that are suitable models of the automorphic function fields obtained from a quaternion algebra (Shimura [S13, S15]).

($\mathrm{II_c}$) Certain fiber varieties of which the base is a curve of type ($\mathrm{II_a}$) or ($\mathrm{II_b}$) and the fibers are Abelian varieties (Kuga and Shimura [K6], Ihara [I1], Deligne [D3]).

($\mathrm{II_d}$) Certain Shimura varieties of higher dimension (Langlands and others; → [B6]).

In these cases, $\zeta(s, V)$ can be expressed by known functions, i.e., by Hecke L-functions with Grössencharakters of algebraic number fields in cases (I) or by Dirichlet series corresponding to modular forms in cases (II). This fact has an essential meaning for the arithmetic properties of these functions. For example, the extended †Ramanujan conjecture concerning the Hecke operator of the automorphic form reduces to Weil's conjecture on varieties related to those in cases II. Moreover, for ($\mathrm{II_a}$)–($\mathrm{II_c}$) the essential point is the congruence relation $\tilde{T}_p = \Pi + \Pi^*$ (Kronecker, Eichler [E1], Shimura). In particular, for ($\mathrm{II_b}$) this formula is related to the problem of constructing class fields over totally imaginary quadratic extensions of a totally real field F utilizing special

values of automorphic functions and class fields over F. Actually, the formula is equivalent to the reciprocity law for class fields (Shimura).

One of the facts that makes the Hasse ζ-function important is that it describes the decomposition law of prime ideals of algebraic number fields when V is an algebraic curve or an Abelian variety (Weil, Shimura [S14], Taniyama [T2], T. Honda [H6]). In that case, its Hasse ζ-function has the following arithmetic meaning.

Let C be a complete, nonsingular algebraic curve defined over an algebraic number field K, and let J be the Jacobian variety of C defined over K. For a prime number l, fix an l-adic coordinate system Σ_l on J, and let $K(J, l^\infty)$ be the extension field of K obtained by adjoining to K all the coordinates of the l^νth division points ($\nu = 1, 2, \ldots$) of J. Then $K(J, l^\infty)/K$ is an infinite Galois extension of K. The corresponding Galois group $\mathfrak{G}(J, l^\infty)$ has the l-adic representation $\sigma \to M_l^*(\sigma)$ by the l-adic coordinates Σ_l. Almost all prime ideals $\mathfrak{p}$ of K are unramified in $K(J, l^\infty)/K$. Thus when we take an arbitrary prime factor $\mathfrak{P}$ of $\mathfrak{p}$ in $K(J, l^\infty)$, the Frobenius substitution of $\mathfrak{P}$,

$$\sigma_{\mathfrak{P}} = \left[\frac{K(J, l^\infty)/K}{\mathfrak{P}}\right],$$

is uniquely determined. Furthermore, the characteristic polynomial $\det(1 - M_l^*(\sigma_{\mathfrak{P}})u)$ is determined only by $\mathfrak{p}$ and does not depend on the choice of the prime factor $\mathfrak{P}$; we denote this polynomial by $P_{\mathfrak{p}}(u, C)$. In this case, for almost all $\mathfrak{p}$, $P_{\mathfrak{p}}(u, C)$ is a polynomial with rational integral coefficients independent of l; namely, the numerator of the ζ-function of the reduction of $C \bmod p$. Thus

$$\zeta_1(s, C) = \prod_{\mathfrak{p}}{}' P_{\mathfrak{p}}(N(\mathfrak{p})^{-s}, C)^{-1}$$
$$\sim \prod_{\mathfrak{p}}{}' \det(1 - M_l^*(\sigma_{\mathfrak{P}})N(\mathfrak{p})^{-s})^{-1}.$$

Here the product $\prod' \det(1 - M_l^*(\sigma_{\mathfrak{P}})N(\mathfrak{p})^{-s})^{-1}$ has the same expression as the Artin L-function if we ignore the fact that M_l^* is the l-adic representation and $K(J, l^\infty)$ is the infinite extension. Thus if we can describe $\zeta(s, C)$ explicitly, then the decomposition process of the prime ideal for intermediate fields between $K(J, l^\infty)$ and K can be made fairly clear. In fact, this is the case for examples ($\mathrm{I_a}$)–($\mathrm{I_c}$) and ($\mathrm{II_a}$)–($\mathrm{II_c}$), from which the relations between the arithmetic of the field of division points $K(J, l^\infty)/K$ and the eigenvalues of the Hecke operator have been obtained. Thus for curves and Abelian varieties, $\zeta(s, V)$ is related to the arithmetic of some number fields; but it is not known whether similar arithmetical relations exist for other kinds of varieties except in a few cases.

Tate's Conjecture. For a projective nonsingular variety V over a finite algebraic number field K, let $\mathfrak{A}^r(\bar{V})$ denote the group of algebraic cycles of codimension r on $\bar{V} = V \otimes_K \mathbf{C}$ modulo homological equivalence and let $\mathfrak{A}^r(V)$ be the subgroup of $\mathfrak{A}^r(\bar{V})$ generated by algebraic cycles rational over K. Then Tate [T3] conjectured that the rank of $\mathfrak{A}^r(V)$ is equal to the order of the pole of $\zeta_{2r}(s, V)$ at $s = r + 1$. This conjecture is closely connected with **Hodge's conjecture** that the space of rational cohomology classes of type (r, r) on $\bar{V}$ is spanned by $\mathfrak{A}^r(\bar{V})$; in fact, the equivalence of these conjectures is known for Abelian varieties of †CM type (H. Pohlmann, *Ann. Math.*, 88 (1968)) and for Fermat hypersurfaces of dimension $2r$ (Tate [T3], Weil [W6]). Thus, when $r = 1$, Tate's conjecture for these varieties holds by Lefschetz's theorem, and when $r > 1$, it holds in certain cases where the Hodge conjecture is verified (Shioda, *Math. Ann.*, 245 (1979); Z. Ran, *Compositio Math.*, 42 (1981)). Further examples are given by $K3$ surfaces with large Picard numbers (Shioda and Inose [S21]; T. Oda, *Proc. Japan Acad.*, 56 (1980)).

L-Functions of Elliptic Curves. Let E be an elliptic curve (with a rational point) over the rational number field $\mathbf{Q}$, and let N be its conductor; a prime number p divides N if and only if E has bad reduction mod p (Tate [T5]). The L-function of E over $\mathbf{Q}$ is defined as follows:

$$L(s, E)$$
$$= \prod_{p \mid N} (1 - \varepsilon_p p^{-s})^{-1} \prod_{p \nmid N} (1 - a_p p^{-s} + p^{1-2s})^{-1},$$

where $\varepsilon_p = 0$ or ± 1 and $1 - a_p u + p u^2 = P_1(u, E \bmod p)$. There are many interesting results and conjectures concerning $L(s, E)$ [T5]:

(1) Functional equation. Let

$$\xi(s, E) = N^{s/2} (2\pi)^{-s} \Gamma(s) L(s, E).$$

Then it is conjectured that $\xi(s, E)$ is holomorphic in the entire s-plane and satisfies the functional equation $\xi(s, E) = \pm \xi(2 - s, E)$. This is true if E has complex multiplication (Deuring) or E is a certain modular curve (Eichler, Shimura).

(2) **Taniyama-Weil conjecture**. Weil [W1 (1967a)] conjectured that, if $L(s, E) = \sum_{n=1}^{\infty} a_n n^{-s}$, then $f(\tau) = \sum_{n=1}^{\infty} a_n e^{2\pi i n \tau}$ is a cusp form of weight 2 for the congruence subgroup $\Gamma_0(N)$ which is an eigenfunction for Hecke operators; moreover E is isogenous to a factor of the Jacobian variety of the modular curve for $\Gamma_0(N)$ in such a way that $f(\tau) d\tau$ corresponds to the differential of the first kind on E. If this conjecture is true, then the statements in (1) follow.

(3) **Birch–Swinnerton-Dyer conjecture**. Assuming analytic continuation of $L(s, E)$, B. Birch and H. Swinnerton-Dyer [B4] conjectured that the order of the zero of $L(s, E)$ at $s=1$ is equal to the rank r of the group $E(\mathbf{Q})$ of rational points of E which is finitely generated by the Mordell-Weil theorem. They verified this for many examples, especially for curves of the type $y^2=x^3-ax$. J. Coates and A. Wiles (*Inventiones Math.*, 39 (1977)) proved that if E has complex multiplication and if $r>0$ then $L(s, E)$ vanishes at $s=1$. This conjecture has a refinement which extends also to Abelian varieties over a global field (Tate, *Sém. Bourbaki*, no. 306 (1966)).

(4) **Sato's conjecture**. Let

$$1-a_p u+pu^2=(1-\pi_p u)(1-\bar{\pi}_p u),$$

with $\pi_p=\sqrt{p}\,e^{i\theta_p}$ $(0<\theta_p<\pi)$. When E has complex multiplication, the distribution of θ_p for half of p is uniform in the interval $[0, \pi]$, and θ_p is $\pi/2$ for the remaining half of p. Suppose that E does not have complex multiplication. Then Sato conjectured that

$$\lim_{x\to\infty}\frac{\text{(the number of prime numbers } p \text{ less than } x \text{ such that } \theta_p\in[\alpha,\beta])}{\text{(the number of prime numbers less than } x)}$$

$$=\frac{2}{\pi}\int_\alpha^\beta \sin^2\theta\, d\theta \quad (0<\alpha<\beta<\pi)$$

(Tate [T3]).

H. Yoshida [Y1] posed an analog of Sato's conjecture for elliptic curves defined over function fields with finite constant fields and proved it in certain cases.

(5) Formal groups. Letting $L(s, E)=\sum a_n n^{-s}$ as before, set $f(x)=\sum_{n=1}^{\infty} a_n x^n/n$. Honda [H6] showed that $f^{-1}(f(x)+f(y))$ is a †formal (Lie) group with coefficients in $\mathbf{Z}$ and that this group is isomorphic over $\mathbf{Z}$ to a formal group obtained by power series expansion of the group law of E with respect to suitable †local uniformizing coordinates at the origin. Such an interpretation of the ζ-function also applies to other cases in which ζ-functions of †group varieties may be characterized as Dirichlet series whose coefficients give a normal form of the group law; e.g., the case of algebraic tori (T. Ibukiyama, *J. Fac. Sci. Univ. Tokyo*, (IA) 21 (1974)).

T. Selberg ζ-Functions and ζ-Functions Associated with Discontinuous Groups

Let $\Gamma\subset SL(2,\mathbf{R})$ be a †Fuchsian group operating on the complex upper half-plane $H=\{z=x+iy \mid y>0\}$. When the two eigenvalues of an element $\gamma\in\Gamma$ are distinct real numbers ξ_1, ξ_2 $(\xi_1\xi_2=1, \xi_1<\xi_2)$, we call γ †hyperbolic. Then the number ξ_2^2 is denoted by $N(\gamma)$ and is called the norm of γ. When γ is hyperbolic, γ^n $(n=1,2,3,\ldots)$ is also hyperbolic. When $\pm\gamma$ is not a positive power of other hyperbolic elements, γ is called a primitive hyperbolic element. The elements conjugate to primitive hyperbolic elements are also primitive hyperbolic elements and have the same norm as γ. Let P_1, $P_2, \ldots$ be the conjugacy classes of primitive hyperbolic elements of Γ, and let $\gamma_i\in P_i$ be their representatives. Suppose that a matrix representation $\gamma\to M(\gamma)$ of Γ is given. Then the analytic function given by

$$Z_\Gamma(s, M)=\prod_i\prod_{n=0}^{\infty}\det(I-M(\gamma_i)N(\gamma_i)^{-s-n})$$

is called the **Selberg ζ-function** (Selberg [S5]). When $\Gamma\backslash H$ is compact and Γ is torsion-free, then $Z_\Gamma(s, M)$ has the following properties.

(1) It can be analytically continued to the whole complex plane of s and gives an †integral function of genus at most 2.

(2) It has zeros of order $(2n+1)(2g-2)v$ at $-n$ $(n=0,1,2,3,\ldots)$. Here g is the genus of the Riemann surface $\Gamma\backslash H$ and v is the degree of the representation M. All other zeros lie on the line $\operatorname{Re}s=1/2$, except for a finite number that lie on the interval $(0,1)$ of the real axis.

(3) It satisfies the functional equation

$$Z_\Gamma(1-s, M)=Z_\Gamma(s, M)\exp\left(-vA(\Gamma\backslash H)\times\int_0^{s-1/2} v\tan(\pi v)\,dv\right),$$

where

$$A(\Gamma\backslash H)=\iint_{\Gamma\backslash H}\frac{dx\,dy}{y^2}=2\pi(2g-2), \qquad x+iy\in H.$$

Property (2) shows that the Riemann hypothesis is almost valid for $Z_\Gamma(s, M)$. The proof is based on the following fact concerning the eigenvalue problem for the variety $\Gamma\backslash H$: The eigenvalue λ of the equation

$$y^2(\partial^2/\partial x^2+\partial^2/\partial y^2)u+\lambda u=0, \quad u\in L^2(\Gamma\backslash H)$$

cannot be a negative number.

Using this function, T. Yamada (1965) investigated the unit distribution of real quadratic fields.

Selberg ζ-functions are defined similarly when $\Gamma\backslash G$ has finite volume but is noncompact. In this case, however, the decomposition of $L_2(\Gamma\backslash G)$ into irreducible representation spaces has a continuous spectrum; hence the properties of the Selberg ζ-function of Γ are quite different from the case when $\Gamma\backslash G$ is compact. Selberg defined the **generalized Eisenstein series** to give the eigenfunctions of this continuous spectrum explicitly. When $\Gamma=$

$SL(2, \mathbf{Z})$, the series is given by $\sum_{(c,d)=1} \frac{y^s}{|c\tau+d|^{2s}}$.

This type of generalized Eisenstein series is also defined for the general semisimple algebraic group G and its arithmetic subgroup. It has been studied by Selberg, Godement, Gel'fand, Harish-Chandra, Langlands, D. Zagier, and others.

U. Ihara ζ-Functions

Let k_p be a p-adic field, $\mathfrak{o}_p$ the ring of integers in k_p, and $G = PSL_2(\mathbf{R}) \times PSL_2(k_p)$. Suppose that Γ is a subgroup of G such that (1) Γ is discrete, (2) $\Gamma\backslash G$ is compact, (3) Γ has no torsion element except the identity, (4) $\Gamma_{\mathbf{R}}$ (the projection of Γ in $PSL_2(\mathbf{R})$) is dense in $PSL_2(\mathbf{R})$, and (5) Γ_p (the projection of Γ in $PSL_2(k_p)$) is dense in $PSL_2(k_p)$. Then $\Gamma \cong \Gamma_{\mathbf{R}} \cong \Gamma_p$. Let $X = \{x + iy \mid y > 0\}$ be the upper half-plane, and let Γ act on X via $\Gamma_{\mathbf{R}}$. The action of Γ on X is not discontinuous, but the subgroup $\Gamma_0 = \{\gamma \in \Gamma \mid \text{projection of } \gamma \text{ to } \Gamma_p \in PSL_2(\mathfrak{o}_p)\}$ operates on X properly discontinuously. For each $z \in X$, define $\Gamma_z = \{\gamma \in \Gamma \mid \gamma(z) = z\}$. Then Γ_z is isomorphic to $\mathbf{Z}$ or $\{1\}$. Let $\tilde{\mathbf{P}}(\Gamma) = \{z \in X \mid \Gamma_z \cong \mathbf{Z}\}$. The group Γ acts on $\tilde{\mathbf{P}}(\Gamma)$, since Γ_z and $\Gamma_{\gamma z}$ are conjugate in Γ. Let $\mathbf{P}(\Gamma) = \tilde{\mathbf{P}}(\Gamma)/\Gamma$. Suppose that $P \in \mathbf{P}(\Gamma)$ is represented by $z \in X$. Choose a generator γ of Γ_z and project γ to Γ_p. Then γ is equivalent to a diagonal matrix $\begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}$ with $\lambda \in k_p$. We denote the valuation of k_p by ord_p and consider $|\mathrm{ord}_p(\lambda)|$. This value depends only on P and we denote it by $\deg(P)$. The **Ihara ζ-function** of Γ is defined by

$$Z_\Gamma(u) = \prod_{P \in \mathbf{P}(\Gamma)} (1 - u^{\deg(P)})^{-1}.$$

Ihara proved that

$$Z_\Gamma(u) = \frac{\prod_{i=1}^{g} (1 - \pi_i u)(1 - \pi_i' u)}{(1-u)(1-q^2 u)} (1-u)^H,$$

where q is the number of elements in the residue class field of $\mathfrak{p}$, and g is the genus of the Riemann surface $\Gamma_0\backslash X$ and $H = (g-1)q(q-1)$. Similar results hold even if Γ has torsion elements and the quotient $\Gamma\backslash G$ is only assumed to have finite volume.

Aside from the factor $(1-u)^H$, this looks like Weil's formula for the congruence ζ-function of an algebraic curve defined over $\mathbf{F}_{q^2}$. Ihara conjectured that the first factor of $Z_\Gamma(u)$ is always the congruence ζ-function of some algebraic curve over $\mathbf{F}_{q^2}$, and furthermore that Γ could be regarded as the fundamental group of a certain Galois covering of this curve which describes the decomposition law of prime divisors in this covering [12, 13, 14]. He verified the conjecture in the case $\Gamma = PGL_2(\mathbf{Z}[1/p])$ by using the †moduli of elliptic curves. Related results have been obtained by Shimura, Ihara, Y. Morita, and others.

V. ζ-Functions Associated with Prehomogeneous Vector Spaces

M. Sato posed a notion of prehomogeneous vector spaces and defined ζ-functions associated with them. Sato's program has been carried on by himself and T. Shintani [S2, S3, S17, S18]. Let G be a linear algebraic group, V a finite-dimensional linear space of dimension n, and ρ a rational representation $G \to GL(V)$, where G, V, and ρ are defined over $\mathbf{Q}$. The triple (G, ρ, V) is called a **prehomogeneous vector space** if there exists a proper algebraic subset S of $V_{\mathbf{C}}$ such that $V_{\mathbf{C}} - S$ is a single $G_{\mathbf{C}}$-orbit. The algebraic set S is called the set of singular points of V. We also assume that G is reductive and S is an irreducible hypersurface of V. Let V^* be the dual vector space of V, and ρ^* the dual (contragredient) representation of G. Then (G, ρ^*, V^*) is again a prehomogeneous vector space, and we denote its set of singular points by S^*. There are homogeneous polynomials P and Q of the same degree d on V and V^*, respectively, such that $S = \{x \in X \mid P(x) = 0\}$ and $S^* = \{x^* \in V^* \mid Q(x^*) = 0\}$. P and Q are relative invariants of G, i.e., $P(\rho(g)x) = \chi(g)P(x)$ and $Q(\rho^*(g)x^*) = \chi(g)^{-1} Q(x^*)$ (for $g \in G$, $x \in V$, and $x^* \in V^*$) hold with a rational character χ of G. Put $G^1 = \ker \chi = \{g \in G \mid \chi(g) = 1\}$. Denote by $G_{\mathbf{R}}^+$ the connected component of 1 of the Lie group $G_{\mathbf{R}}$. Let $V_{\mathbf{R}} - S = V_1 \cup \ldots \cup V_l$, $V_{\mathbf{R}}^* - S^* = V_1^* \cup \ldots \cup V_l^*$ be the decompositions of $V_{\mathbf{R}} - S$ and $V_{\mathbf{R}}^* - S^*$ into their topologically connected components. Then V_i and V_j^* are $G_{\mathbf{R}}^+$-orbits. We further assume that $V_{\mathbf{R}} \cap S$ decomposes into the union of a finite number of $G_{\mathbf{R}}^1$-orbits. Set $\Gamma = G_{\mathbf{R}}^+ \cap G_{\mathbf{Z}}^1$, and take Γ-invariant lattices L and L^* in $V_{\mathbf{Q}}$ and $V_{\mathbf{Q}}^*$, respectively. Consider the following functions in s:

$$\Phi_i(f, s) = \int_{V_i} f(x) |P(x)|^s dx,$$

$$\Phi_j^*(f, s) = \int_{V_j^*} f^*(x^*) |Q(x^*)|^s dx^*,$$

and

$$Z_i(f, L, s) = \int_{G_{\mathbf{R}}^+/\Gamma} \chi(g)^s \sum_{x \in L \cap V_i} f(\rho(g)x)\, dg,$$

$$Z_j^*(f^*, L^*, s)$$

$$= \int_{G_{\mathbf{R}}^+/\Gamma} \chi(g)^{-s} \sum_{x^* \in L^* \cap V_j^*} f^*(\rho^*(g)x^*)\, dg,$$

where f and f^* are †rapidly decreasing functions on $V_{\mathbf{R}}$ and $V_{\mathbf{R}}^*$, respectively, dx and dx^* are Haar measures of $V_{\mathbf{R}}$ and $V_{\mathbf{R}}^*$, respectively, and dg is a Haar measure of G. Then the ratios

$$\frac{Z_i(f, L, s)}{\Phi_i(f, s-n/d)} = \xi_i(s, L),$$

$$\frac{Z_j^*(f^*, L^*, s)}{\Phi_j^*(f^*, s-n/d)} = \xi_j^*(s, L^*)$$

are independent of the choice of f and f^* and are Dirichlet series in s. These Dirichlet series $\xi_i(s, L)$ and $\xi_j^*(s, L^*)$ are called **ξ-functions associated with the prehomogeneous space**. Considering Fourier transforms of $|P(x)|^s$ and $|Q(x^*)|^s$, we obtain functional equations for ξ_i and ξ_j^* under some additional (but mild) conditions on (G, ρ, V) as follows. The Dirichlet series ξ_i and ξ_j^* are analytically continuable to meromorphic functions on the whole s-plane, and they satisfy

$$v(L^*)\xi_j^*(n/d-s, L^*)$$
$$=\gamma(s-n/d)(2\pi)^{-ds}|b_0|^s \exp(\pi d\sqrt{-1}\, s/2)$$
$$\times \sum_{j=1}^{l} u_{ij}(s)\xi_i(s, L),$$

with a Γ-factor $\gamma(s)=\prod_{i=1}^{d}\Gamma(s-c_i+1)$. Here $u_{ij}(s)$ $(1 \leqslant i, j \leqslant l)$ are polynomials in $\exp(-\pi\sqrt{-1}\, s)$ with degree $\leqslant d$, and b_0 and c_i are constants depending only on (G, ρ, V).

Epstein's ζ-functions and Siegel's Dirichlet series associated with indefinite quadratic forms are examples of the above-defined ζ-functions. Shintani defined such ζ-functions related to integral binary cubic forms and obtained asymptotic formulas concerning the class numbers of irreducible integral binary cubic forms with discriminant n, which are improvements on the results of Davenport [S17].

Recently M. Sato studied ζ-functions of prehomogeneous vector spaces without assuming the conditions that G is reductive and S is irreducible. In this case, ζ-functions of several complex variables are obtained. For examples and classification of prehomogeneous vector spaces → [S4].

References

[A1] E. Artin, Collected papers, S. Lang and J. T. Tate (eds.), Addison-Wesley, 1965.
[A2] E. Artin, Zur Theorie der L-Reihen mit allgemeinen Gruppencharakteren, Abh. Math. Sem. Univ. Hamburg, 8 (1930), 292–306 [A1, 165–179].
[A3] M. Artin, A. Grothendieck, and J. L. Verdier, Théorie des topos et cohomologie étale des schémas (SGA 4), Lecture notes in math. 269, 270, 305, Springer, 1972–1973.
[A4] J. Ax, On the units of an algebraic number fields, Illinois J. Math., 9 (1965), 584–589.
[B1] P. Bayer and J. Neukirch, On values of zeta functions and l-adic Euler characteristics, Inventiones Math., 50 (1978), 35–64.
[B2] P. Berthelot, Cohomogie cristalline des schéma de caracteristique $p>0$, Lecture notes in math. 407, Springer, 1974.
[B3] P. Berthelot and A. Ogus, Notes on crystalline cohomology, Math. notes 21, Princeton Univ. Press, 1978.
[B4] B. Birch and H. Swinnerton-Dyer, Notes on elliptic curves II, J. Reine Angew. Math., 218 (1965), 79–108.
[B5] E. Bombieri and H. P. F. Swinnerton-Dyer, On the local zeta function of a cubic threefold, Ann. Scuola Norm. Sup. Pisa, (3) 21 (1967), 1–29.
[B6] A. Borel and W. Casselman (eds.), Automorphic forms, representations, and L-functions, Amer. Math. Soc. Proc. Symp. Pure Math., 33 (pts. 1 and 2) (1979).
[B7] A. Brumer, On the units of algebraic number fields, Mathematika, 14 (1967), 121–124.
[C1] K. Chandrasekharan, Lectures on the Riemann zeta-function, Tata Inst., 1953.
[C2] S. Chowla and A. Selberg, On Epstein's zeta function I, Proc. Nat. Acad. Sci. US, 35 (1949), 371–374.
[C3] J. Coates, P-adic L-functions and Iwasawa's theory, Algebraic Number Fields (L-functions and Galois Properties), A. Fröhlich (ed.), Academic Press, 1977.
[D1] R. Dedekind, Über die Theorie der ganzen algebraischen Zahlen, Dirichlet's Vorlesungen über Zahlentheorie, Supplement XI, fourth edition, 1894, secs. 184, 185, 186 (Gesammelte mathematische Werke III, Braunschweig, 1932, 297–314; Chelsea, 1968).
[D2] P. Deligne, La conjecture de Weil pour les surfaces $K3$, Inventiones Math., 75 (1972), 206–226.
[D3] P. Deligne, Formes modulaires et représentations l-adiques, Sém. Bourbaki, exp. 355, Lecture notes in math. 349, Springer, 1973.
[D4] P. Deligne, La conjecture de Weil, I, II, Publ. Math. Inst. HES, 43 (1974), 273–307; 52 (1980), 137–252.
[D5] P. Deligne, Cohomologie étale (SGA $4\frac{1}{2}$), Lecture notes in math. 569, Springer, 1977.
[D6] P. Deligne, Les constantes des équations fonctionnelles des fonctions L, Lecture notes in math. 349, Springer, 1973.
[D7] P. Deligne and N. Katz, Groupes de monodromie en géométrie algébrique (SGA

7II), Lecture notes in math. 340, Springer, 1973.
[D8] P. Deligne and K. Ribet, Values of Abelian L-functions at negative integers over totally real fields, Inventiones Math., 59 (1980), 227–286.
[D9] P. Deligne and J-P. Serre, Formes modulaires de poids 1, Ann. Sci. Ecole Norm. Sup. (4), 7 (1974), 507–530.
[D10] M. Deuring, Algebren, Erg. Math., Springer, second edition, 1968.
[D11] M. Deuring, Die Zetafunktionen einer algebraischen Kurve vom Geschlechte Eins I, II, III, IV, Nachr. Akad. Wiss. Göttingen (1953), 85–94; (1955), 13–42; (1956), 37–76; (1957), 55–80.
[D12] K. Doi and H. Naganuma, On the functional equation of certain Dirichlet series, Inventiones Math., 9 (1969–1970), 1–14.
[D13] B. Dwork, On the rationality of the zeta-function of an algebraic variety, Amer. J. Math., 82 (1960), 631–648.
[D14] B. Dwork, On the zeta function of a hypersurface I, II, III, Publ. Math. Inst. HES, 1962, 5–68; Ann. Math., (2) 80 (1964), 227–299; (2) 83 (1966), 457–519.
[E1] M. Eichler, Einführung in die Theorie der algebraischen Zahlen und Funktionen, Birkhäuser, 1963; English translation, Introduction to the theory of algebraic numbers and functions, Academic Press, 1966.
[F1] B. Ferrero and L. C. Washington, The Iwasawa invariant μ_p vanishes for Abelian number fields, Ann. Math., (2) 109 (1979), 377–395.
[G1] R. Godement, Les fonctions ζ des algèbres simples I, II, Sém. Bourbaki, exp. 171, 176, 1958–1959.
[G2] A. Grothendieck, Formule de Lefschetz et rationalité des fonction L, Sém. Bourbaki, exp. 379, 1964–1965 (Benjamin, 1966).
[G3] A. Grothendieck, Cohomologie l-adique et fonctions L (SGA 5), Lecture notes in math. 589, Springer, 1977.
[H1] H. Hasse, Zur Theorie der abstrakten elliptischen Funktionenkörper I, II, III, J. Reine Angew. Math., 175 (1936), 55–62, 69–88, 193–208.
[H2] E. Hecke, Mathematische Werke, Vandenhoeck & Ruprecht, 1959.
[H3] E. Hecke, Eine neue Art von Zetafunctionen und ihre Beziehung zur Verteilung der Primzahlen I, II, Math. Z., 1 (1918), 357–376; 6 (1920), 11–51 [H2, 215–234, 249–289].
[H4] E. Hecke, Über die Bestimmung Dirichletscher Reihen durch ihre Funktionalgleichung, Math. Ann., 112 (1936), 664–699 (H2, 591–626].
[H5] E. Hecke, Über Modulfunktionen und die Dirichletschen Reihen mit Eulerscher Produktentwicklung I, II, Math. Ann., 114 (1937), 1–28; 316–351 [H2, 644–671, 672–707].
[H6] T. Honda, Formal groups and zeta functions, Osaka J. Math., 5 (1968), 199–213.
[H7] T. Honda, Isogeny classes cf Abelian varieties over finite fields, J. Math. Soc. Japan, 20 (1968), 83–95.
[I1] Y. Ihara, Hecke polynomials as congruence ζ functions in elliptic modular case (To validate Sato's identity), Ann. Math., (2) 85 (1967), 267–295.
[I2] Y. Ihara, On congruence monodromy problems, I, II, Lecture notes, Univ. of Tokyo, 1968–1969.
[I3] Y. Ihara, Non-Abelian class fields over function fields in special cases, Actes Congr. Intern. Math., 1970, Nice, Gauthier-Villars, p. 381–389.
[I4] Y. Ihara, Congruence relations and Shimura curves. Amer. Math. Soc. Proc. Symp. Pure Math., 33 (1979), pt. 2, 291–311.
[I5] K. Iwasawa, On Γ-extensions of algebraic number fields, Bull. Amer. Math. Soc., 65 (1959), 183–226.
[I6] K. Iwasawa, On p-adic L-functions, Ann. Math., (2) 89 (1969), 198–205.
[I7] K. Iwasawa, Lectures on p-adic L-functions, Ann. math. studies 74, Princeton Univ. Press, 1972.
[J1] H. Jacquet and R. P. Langlands, Automorphic forms on $GL(2)$, Lecture notes in math. 114, Springer, 1970.
[J2] H. Jacquet, Automorphic forms on $GL(2)$, II, Lecture notes in math. 278, Springer, 1972.
[K1] N. Katz and W. Messing, Some consequences of the Riemann hypothesis for varieties over finite fields, Inventiones Math., 23 (1974), 73–77.
[K2] N. Katz, An overview of Deligne's proof of the Riemann hypothesis for varieties over finite fields, Amer. Math. Soc. Proc. Symp. Pure Math., 28 (1976), 275–305.
[K3] N. Katz, P-adic L-functions for CM-fields, Inventiones Math., 49 (1978), 199–297.
[K4] S. Kleiman, Algebraic cycles and the Weil conjectures, Dix exposés sur la cohomologies des schémas, North-Holland, 1968, 359–386.
[K5] T. Kubota and H. W. Leopoldt, Eine p-adischer Theorie der Zetawerte, J. Reine Angew. Math., 214/215 (1964), 328–339.
[K6] M. Kuga and G. Shimura, On the zeta function of a fiber variety whose fibers are Abelian varieties, Ann. Math., (2) 82 (1965), 478–539.
[L1] E. G. H. Landau, Handbuch der Lehre von der Verteilung der Primzahlen I, II, Teubner, 1909 (Chelsea, 1953).
[L2] S. Lang, Sur les series L d'une variété algébrique, Bull. Soc. Math. France, 84 (1956), 385–407.

[L3] S. Lang, Algebraic number theory, Addison-Wesley, 1970.
[L4] R. P. Langlands, Problems in the theory of automorphic forms, Lecture notes in math. 170, Springer, 1970, 18–86.
[L5] R. P. Langlands, Base change for $GL(2)$, Ann. math. studies 96, Princeton Univ. Press, 1980.
[L6] H. W. Leopoldt, Eine p-adische Theorie der Zetawerte II, J. Reine Angew. Math., 274/275 (1975), 224–239.
[L7] S. Lubkin, A p-adic proof of the Weil conjecture, Ann. Math., (2) 87 (1968), 105–194; (2) (1968), 195–225.
[M1] Yu. I. Manin, Periods of parabolic forms and p-adic Hecke series, Math. USSR-Sb., 21 (1973), 371–393.
[M2] B. Mazur, Eigenvalues of Frobenius acting on algebraic varieties over finite fields, Amer. Math. Soc. Proc. Symp. Pure Math., 29 (1975).
[M3] B. Mazur and H. Swinnerton-Dyer, On the p-adic L-series of an elliptic curve, Inventiones Math., 25 (1974), 1–61.
[M4] J. S. Milne, Etale cohomology, Princeton series in mathematics 33, Princeton Univ. Press, 1980.
[P1] I. I. Pyatetskiĭ-Shapiro and I. R. Shafarevich, The arithmetic of $K3$ surfaces, Proc. Steklov Inst. Math., 132 (1973), 45–57.
[P2] K. Prachar, Primzahlverteilung, Springer, 1957.
[R1] B. Riemann, Über die Anzahl der Primzahlen unter einer gegebenen Grösse, Gesammelte mathematische Werke, 1859, 145–153 (Dover, 1953).
[S1] H. Saito, Automorphic forms and algebraic extensions of number fields, Lectures in math. 8, Kinokuniya, Tokyo, 1975.
[S2] M. Sato, Theory of prehomogeneous vector spaces (notes by T. Shintani) (in Japanese), Sûgaku no Ayumi, 15 (1970), 85–157.
[S3] M. Sato and T. Shintani, On zeta functions associated with prehomogeneous vector spaces, Ann. Math., (2) 100 (1974), 131–170.
[S4] M. Sato and T. Kimura, A classification of irreducible prehomogeneous vector spaces and their invariants, Nagoya Math. J., 65 (1977), 1–155.
[S5] A. Selberg, Harmonic analysis and discontinuous groups on weakly symmetric Riemannian spaces with applications to Dirichlet series, J. Indian Math. Soc., 20 (1956), 47–87.
[S6] A. Selberg, On the zeros of Riemann's zeta-function, Skr. Norske Vid. Akad. Oslo, no. 10 (1942), 1–59.
[S7] J.-P. Serre, Zeta and L functions, Arithmetical Algebraic Geometry, Harper & Row, 1965, 82–92.
[S8] J.-P. Serre, Facteurs locaux des fonctions zêta des variétés algébriques (Définitions et conjectures), Sém. Delange-Pisot-Poitou, 11 (1969/70), no. 19.
[S9] J.-P. Serre, Formes modulaires et fonctions zêta p-adiques, Lecture notes in math. 317, Springer, 1973, 319–338.
[S10] H. Shimizu, On zeta functions of quaternion algebras, Ann. Math., (2) 81 (1965), 166–193.
[S11] G. Shimura and Y. Taniyama, Complex multiplication of Abelian varieties and its applications to number theory, Publ. Math. Soc. Japan, no. 6, 1961.
[S12] G. Shimura, Introduction to the arithmetic theory of automorphic functions, Publ. Math. Soc. Japan, no. 11, 1971.
[S13] G. Shimura, On the zeta-functions of the algebraic curves uniformized by certain automorphic functions, J. Math. Soc. Japan, 13 (1961), 275–331.
[S14] G. Shimura, A reciprocity law in non-solvable extensions, J. Reine Angew. Math., 221 (1966), 209–220.
[S15] G. Shimura, Construction of class fields and zeta functions of algebraic curves, Ann. Math., (2) 85 (1967), 58–159.
[S16] G. Shimura, On canonical models of arithmetic quotients of bounded symmetric domains, I, II, Ann. Math., (2) 91 (1970) 144–222; 92 (1970), 528–549.
[S17] T. Shintani, On Dirichlet series whose coefficients are class numbers of integral binary cubic forms, J. Math. Soc. Japan, 24 (1972), 132–188.
[S18] T. Shintani, On zeta functions associated with the vector space of quadratic forms, J. Fac. Sci. Univ. Tokyo, (IA) 22 (1975), 25–65.
[S19] T. Shintani, On a Kronecker limit formula for real quadratic fields, J. Fac. Sci. Univ. Tokyo, (IA) 24 (1977), 167–199.
[S20] T. Shintani, On liftings of holomorphic automorphic forms, Amer. Math. Soc. Proc. Symp. Pure Math., 33 (1979), pt. 2, 97–110.
[S21] T. Shioda and H. Inose, On singular $K3$ surfaces, Complex Analysis and Algebraic Geometry, Iwanami and Cambridge Univ. Press, 1977, 119–136.
[S22] C. L. Siegel, Gesammelte Abhandlungen I, II, III, IV, Springer, 1966, 1979.
[S23] C. L. Siegel, Über die Classenzahl quadratischer Zahlkörper, Acta Arith., 1 (1935), 83–86 [S22, I. 406–409].
[S24] C. L. Siegel, Über die Zetafunktionen indefiniter quadratischer Formen I, II, Math. Z., 43 (1938), 682–708; 44 (1939), 398–426 [S22, II, 41–67, 68–96].
[S25] H. M. Stark, L-functions at $s=1$, II, III, Advances in Math., 17 (1975), 60–92; 22 (1976), 64–84.
[T1] T. Tamagawa, On the ζ-functions of a

division algebra, Ann. Math., (2) 77 (1963), 387–405.
[T2] Y. Taniyama, L-functions of number fields and zeta functions of Abelian varieties, J. Math. Soc. Japan, 9 (1957), 330–366.
[T3] J. Tate, Algebraic cycles and poles of zeta functions, Arithmetical Algebraic Geometry, Harper & Row, 1965, 93–110.
[T4] J. Tate, Endomorphisms of Abelian varieties over finite fields, Inventiones Math., 2 (1966), 134–144.
[T5] J. Tate, The arithmetic of elliptic curves, Inventiones Math., 23 (1974), 179–206.
[T6] J. Tate, Number theoretic background, Amer. Math. Soc. Proc. Symp. Pure Math., 33 (1979), pt. 2, 3–26.
[T7] T. Tatuzawa, On the Hecke-Landau L-series, Nagoya Math. J., 16 (1960), 11–20.
[T8] E. C. Titchmarch, The theory of the Riemann zeta-function, Clarendon Press, 1951.
[W1] A. Weil, Collected papers, I–III, Springer, 1980.
[W2] A. Weil, Courbes algébriques et variétés abéliennes, Hermann, 1971 (orig., 1948).
[W3] A. Weil, Numbers of solutions of equations in finite fields, Bull. Amer. Math. Soc., 55 (1949), 497–508 [W1, I, 399–400].
[W4] A. Weil, Number theory and algebraic geometry, Proc. Intern. Congr. Math., 1950, Cambridge, vol. 2, 90–100 [W1, I, 442–449].
[W5] A. Weil, Sur la théorie du corps de classes, J. Math. Soc. Japan, 3 (1951), 1–35 [W1, I, 483–517].
[W6] A. Weil, Jacobi sums as "Grössencharaktere," Trans. Amer. Math. Soc., 73 (1952), 487–495 [W1, II, 63–71].
[W7] A. Weil, On a certain type of characters of the idele-class group of an algebraic number field, Proc. Intern. Symposium on Algebraic Number Theory, Tokyo-Nikko, 1955, 1–7 [W1, II, 255–261].
[W8] A. Weil, Basic number theory, Springer, 1967.
[W9] A. Weil, Dirichlet series and automorphic forms, Lecture notes in math. 189, Springer, 1971.
[Y1] H. Yoshida, On an analogue of the Sato conjecture, Inventiones Math., 19 (1973), 261–277.

Appendix A
Tables of Formulas

1. Algebraic Equations (→ 10 Algebraic Equations)

(I) Quadratic Equation $ax^2+bx+c=0 \quad (a\neq 0)$

The roots are

$$x=\frac{-b\pm\sqrt{b^2-4ac}}{2a}=\frac{-b'\pm\sqrt{b'^2-ac}}{a} \qquad (b\equiv 2b').$$

The discriminant is b^2-4ac.

(II) Cubic Equation $ax^3+bx^2+cx+d=0 \quad (a\neq 0)$

By the translation $\xi=x+b/3a$, the equation is transformed into $\xi^3+3p\xi+q=0$, where

$p\equiv(3ac-b^2)/9a^2, \qquad q\equiv(2b^3-9abc+27a^2d)/27a^3.$

Its discriminant is $-27(q^2+4p^3)$. The roots of the latter equation are

$$\xi=\sqrt[3]{\alpha}+\sqrt[3]{\beta}, \quad \omega\sqrt[3]{\alpha}+\omega^2\sqrt[3]{\beta}, \quad \omega^2\sqrt[3]{\alpha}+\omega\sqrt[3]{\beta},$$

where

$$\omega=e^{2\pi i/3}=\frac{-1+\sqrt{3}\,i}{2}, \quad \left.\begin{matrix}\alpha\\ \beta\end{matrix}\right\}=\frac{-q\pm\sqrt{q^2+4p^3}}{2} \qquad \text{(Cardano's formula).}$$

Casus irreducibilis (the case when $q^2+4p^3<0$). Putting $\alpha\equiv re^{i\theta}$ ($\beta=\bar{\alpha}$), the roots are

$$\xi=2\sqrt[3]{r}\cos(\theta/3), \quad 2\sqrt[3]{r}\cos[(\theta+2\pi)/3], \quad 2\sqrt[3]{r}\cos[(\theta+4\pi)/3].$$

(III) Quartic Equation (Biquadratic Equation) $ax^4+bx^3+cx^2+dx+e=0 \quad (a\neq 0)$

By the translation $\xi=x+b/4a$, the equation is transformed into

$\xi^4+p\xi^2+q\xi+r=0.$

The cubic resolvent of the latter is $t^3-pt^2-4rt+(4pr-q^2)=0$. If t_0 is one of the roots of the cubic resolvent, the roots ξ of the above equation are the solutions of two quadratic equations

$$\xi^2\pm\sqrt{t_0-p}\,\left[\xi-q/2(t_0-p)\right]+t_0/2=0 \qquad \text{(Ferrari's formula).}$$

2. Trigonometry

(I) Trigonometric Functions (→ 432 Trigonometry)

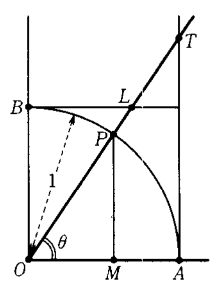


Fig. 1

(1) In Fig. 1, $OA=OB=OP=1$, and

$MP=\sin\theta, \qquad OM=\cos\theta, \qquad AT=\tan\theta,$

$BL=\cot\theta, \qquad OT=\sec\theta, \qquad OL=\operatorname{cosec}\theta.$

(2) $\sin^2\theta+\cos^2\theta=1,$

$\tan\theta=\sin\theta/\cos\theta, \qquad \cot\theta=1/\tan\theta, \qquad \sec\theta=1/\cos\theta,$

$\operatorname{cosec}\theta=1/\sin\theta, \qquad 1+\tan^2\theta=\sec^2\theta, \qquad 1+\cot^2\theta=\operatorname{cosec}^2\theta.$

(3)

	θ	$-\theta$	$\pi/2\pm\theta$	$\pi\pm\theta$	$n\pi\pm\theta$
sin	s	$-s$	c	$\mp s$	$\pm(-1)^n s$
cos	c	c	$\mp s$	$-c$	$(-1)^n c$
tan	t	$-t$	$\mp 1/t$	$\pm t$	$\pm t$

(4)

α	$0°$	$15°$	$18°$	$22.5°$	$30°$	$36°$	$45°$	
↗	0	$\pi/12$	$\pi/10$	$\pi/8$	$\pi/6$	$\pi/5$	$\pi/4$	
$\sin\alpha$	0	$\frac{\sqrt{3}-1}{2\sqrt{2}}$	$\frac{\sqrt{5}-1}{4}$	$\frac{\sqrt{2-\sqrt{2}}}{2}$	$\frac{1}{2}$	$\frac{\sqrt{10-2\sqrt{5}}}{4}$	$\frac{1}{\sqrt{2}}$	$\cos\alpha$
$\cos\alpha$	1	$\frac{\sqrt{3}+1}{2\sqrt{2}}$	$\frac{\sqrt{10+2\sqrt{5}}}{4}$	$\frac{\sqrt{2+\sqrt{2}}}{2}$	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{5}+1}{4}$	$\frac{1}{\sqrt{2}}$	$\sin\alpha$
	$\pi/2$	$5\pi/12$	$2\pi/5$	$3\pi/8$	$\pi/3$	$3\pi/10$	$\pi/4$	↙
	$90°$	$75°$	$72°$	$67.5°$	$60°$	$54°$	$45°$	α

(5) Addition Formulas

$\sin(\alpha\pm\beta)=\sin\alpha\cos\beta\pm\cos\alpha\sin\beta,\qquad \cos(\alpha\pm\beta)=\cos\alpha\cos\beta\mp\sin\alpha\sin\beta,$

$\tan(\alpha\pm\beta)=(\tan\alpha\pm\tan\beta)/(1\mp\tan\alpha\tan\beta).$

(6) $\sin 2\alpha=2\sin\alpha\cos\alpha,\qquad \cos 2\alpha=\cos^2\alpha-\sin^2\alpha=2\cos^2\alpha-1=1-2\sin^2\alpha,$

$\tan 2\alpha=2\tan\alpha/(1-\tan^2\alpha).$

$\sin 3\alpha=3\sin\alpha-4\sin^3\alpha,\qquad \cos 3\alpha=4\cos^3\alpha-3\cos\alpha,$

$\tan 3\alpha=(3\tan\alpha-\tan^3\alpha)/(1-3\tan^2\alpha).$

$$\sin n\alpha=\sum_{i=0}^{[(n-1)/2]}\binom{n}{2i+1}(-1)^i\sin^{2i+1}\alpha\cos^{n-(2i+1)}\alpha,$$

$$\cos n\alpha=\sum_{i=0}^{[n/2]}\binom{n}{2i}(-1)^i\sin^{2i}\alpha\cos^{n-2i}\alpha.$$

(7) $\sin^2(\alpha/2)=(1-\cos\alpha)/2,\qquad \cos^2(\alpha/2)=(1+\cos\alpha)/2,$

$\tan^2(\alpha/2)=(1-\cos\alpha)/(1+\cos\alpha).$

(8) $2\sin\alpha\cos\beta=\sin(\alpha+\beta)+\sin(\alpha-\beta),\qquad 2\cos\alpha\sin\beta=\sin(\alpha+\beta)-\sin(\alpha-\beta),$

$2\cos\alpha\cos\beta=\cos(\alpha+\beta)+\cos(\alpha-\beta),\qquad -2\sin\alpha\sin\beta=\cos(\alpha+\beta)-\cos(\alpha-\beta).$

$\sin\alpha+\sin\beta=2\sin[(\alpha+\beta)/2]\cos[(\alpha-\beta)/2],$

$\sin\alpha-\sin\beta=2\cos[(\alpha+\beta)/2]\sin[(\alpha-\beta)/2],$

$\cos\alpha+\cos\beta=2\cos[(\alpha+\beta)/2]\cos[(\alpha-\beta)/2],$

$\cos\alpha-\cos\beta=-2\sin[(\alpha+\beta)/2]\sin[(\alpha-\beta)/2].$

(II) Plane Triangles

As shown in Fig. 2, we denote the interior angles of a triangle ABC by α, β, γ; the corresponding side lengths by a, b, c; the area by S; the radii of inscribed, circumscribed, and escribed circles by r, R, r_A, respectively; the perpendicular line from the vertex A to the side BC by AH; the midpoint of the side BC by M; bisector of the angle A by AD; and the lengths of AH, AM, AD by h_A, m_A, f_A, respectively. Similar notations are used for B and C. Put $s\equiv(a+b+c)/2$. The symbol ... means similar formulas by the cyclic permutation of the letters A, B, C, and corresponding quantities.

$\frac{a}{\sin\alpha}=\frac{b}{\sin\beta}=\frac{c}{\sin\gamma}=2R$ (law of sines).

$a=b\cos\gamma+c\cos\beta,\ \ldots$ (the first law of cosines).

$a^2=b^2+c^2-2bc\cos\alpha,\ \ldots$ (the second law of cosines).

$\sin^2(\alpha/2)=(s-b)(s-c)/bc,\ \ldots;\qquad \cos^2(\alpha/2)=s(s-a)/bc,\ \ldots.$

$(b+c)\sin(\alpha/2)=a\cos[(\beta-\gamma)/2],\ \ldots;\qquad (b-c)\cos(\alpha/2)=a\sin[(\beta-\gamma)/2],\ \ldots.$

$\frac{a+b}{a-b}=\frac{\tan[(\alpha+\beta)/2]}{\tan[(\alpha-\beta)/2]},\ \ldots$ (Napier's rule).

$$S = ah_A/2 = (1/2)bc\sin\alpha = (1/2)a^2\sin\beta\sin\gamma/\sin\alpha = abc/4R = 2R^2\sin\alpha\sin\beta\sin\gamma$$
$$= rs = r_A(s-a) = \sqrt{rr_Ar_Br_C}$$
$$= \sqrt{s(s-a)(s-b)(s-c)} \qquad \text{(Heron's formula).}$$
$$r = (s-a)\tan(\alpha/2) = 4R\sin(\alpha/2)\sin(\beta/2)\sin(\gamma/2).$$
$$r_A = s\tan(\alpha/2) = (s-b)\cot(\gamma/2) = 4R\sin(\alpha/2)\cos(\beta/2)\cos(\gamma/2).$$
$$1/r = (1/h_A) + (1/h_B) + (1/h_C).$$
$$m_A^2 = (2b^2 + 2c^2 - a^2)/4 = (b^2 + c^2 + 2bc\cos\alpha)/4.$$
$$f_A = 2bc\cos(\alpha/2)/(b+c) = 2\sqrt{bcs(s-a)}\,/(b+c).$$
$$f_Af_Bf_C = 8abcrs^2/(b+c)(c+a)(a+b).$$

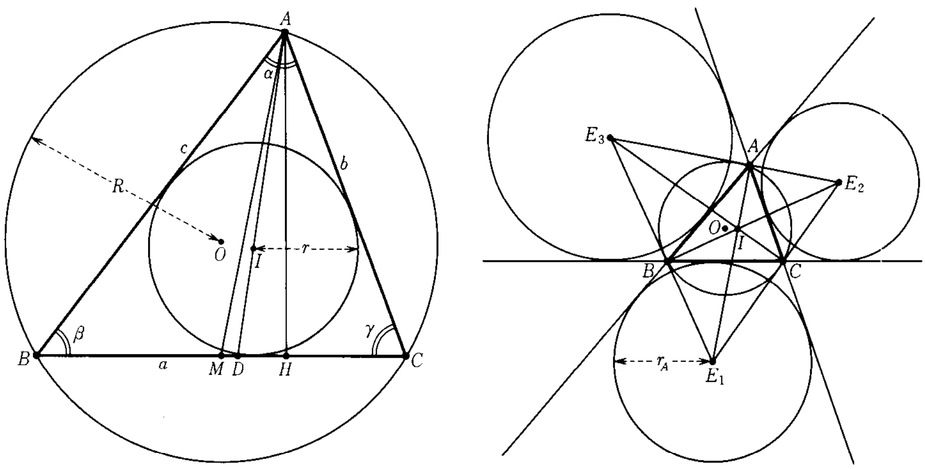


Fig. 2

(III) Spherical Triangles

We denote the interior angles of a spherical triangle by α, β, γ; the corresponding sides by a, b, c; the area by S; and the radius of the supporting sphere by ρ. We have

$$\sin a : \sin b : \sin c = \sin\alpha : \sin\beta : \sin\gamma \qquad \text{(law of sines).}$$

$$\cos a = \cos b\cos c + \sin b\sin c\cos\alpha, \quad \dots; \qquad \cos\alpha = -\cos\beta\cos\gamma + \sin\beta\sin\gamma\cos a, \quad \dots$$

(law of cosines).

$$\sin a\cos\beta = \cos b\sin c - \sin b\cos c\cos\alpha, \quad \dots \qquad \text{(law of sines and cosines).}$$

$$\cot a\sin b = \cos b\cos\gamma + \cot\alpha\sin\gamma, \quad \dots \qquad \text{(law of cotangents).}$$

$$\tan[(a+b)/2]/\tan[(a-b)/2] = \tan[(\alpha+\beta)/2]/\tan[(\alpha-\beta)/2], \quad \dots \qquad \text{(law of tangents).}$$

$$\tan[(\alpha+\beta)/2]\tan(\gamma/2) = \cos[(a-b)/2]/\cos[(a+b)/2], \quad \dots;$$
$$\tan[(\alpha-\beta)/2]\tan(\gamma/2) = \sin[(a-b)/2]/\sin[(a+b)/2], \quad \dots;$$
$$\tan[(a+b)/2]\cot(c/2) = \cos[(\alpha-\beta)/2]/\cos[(\alpha+\beta)/2], \quad \dots;$$
$$\tan[(a-b)/2]\cot(c/2) = \sin[(\alpha-\beta)/2]/\sin[(\alpha+\beta)/2], \quad \dots \qquad \text{(Napier's analogies).}$$

$$S = (\alpha+\beta+\gamma-\pi)\rho^2 = 2\rho^2\arccos\frac{\cos^2(a/2R)+\cos^2(b/2R)+\cos^2(c/2R)}{2\cos(a/2R)\cos(b/2R)\cos(c/2R)} \qquad \text{(Heron's formula).}$$

For a right triangle ($\gamma = \pi/2$), we have Napier's rule of circular parts: taking the subscripts modulo 5 in Fig. 3,

$$\sin\theta_i = \tan\theta_{i+1}\tan\theta_{i-1} = \cos\theta_{i+2}\cos\theta_{i-2}.$$

For example, we have

$$\cos c = \cos a \cos b = \cot\alpha \cot\beta,$$

$$\cos\beta = \tan a \cot c = \cos b \sin\alpha,$$

$$\sin a = \tan b \cot\beta = \sin c \sin\alpha.$$

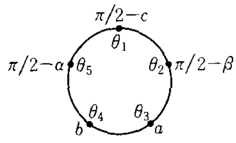


Fig. 3

3. Vector Analysis and Coordinate Systems

We denote a 3-dimensional vector by $\mathbf{A} \equiv (A_x, A_y, A_z) = A_x\boldsymbol{i} + A_y\boldsymbol{j} + A_z\boldsymbol{k}$, $|\mathbf{A}| = \sqrt{A_x^2 + A_y^2 + A_z^2}$.

(I) Vector Algebra (→ 442 Vectors)

Scalar product $\mathbf{A}\cdot\mathbf{B} \equiv \mathbf{AB} \equiv (\mathbf{A},\mathbf{B}) = A_xB_x + A_yB_y + A_zB_z = |\mathbf{A}|\,|\mathbf{B}|\cos\theta$ (where θ is the angle between $\mathbf{A}$ and $\mathbf{B}$).

Vector product

$$\mathbf{A}\times\mathbf{B} \equiv [\mathbf{A},\mathbf{B}] = (A_yB_z - A_zB_y)\mathbf{i} + (A_zB_x - A_xB_z)\mathbf{j} + (A_xB_y - A_yB_x)\mathbf{k} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ A_x & A_y & A_z \\ B_x & B_y & B_z \end{vmatrix}.$$

$$|\mathbf{A}\times\mathbf{B}| = |\mathbf{A}|\,|\mathbf{B}|\sin\theta.$$

$$\mathbf{A}\cdot\mathbf{B} = \mathbf{B}\cdot\mathbf{A}. \quad \mathbf{A}\cdot\mathbf{A} \equiv \mathbf{A}^2 = |\mathbf{A}|^2. \quad \mathbf{A}\times\mathbf{A} = 0. \quad \mathbf{A}\cdot(\mathbf{A}\times\mathbf{B}) = 0. \quad (\mathbf{A}\times\mathbf{B})^2 = |\mathbf{A}|^2|\mathbf{B}|^2 - (\mathbf{A}\cdot\mathbf{B})^2.$$

$$\mathbf{A}\times(\mathbf{B}\times\mathbf{C}) = (\mathbf{A}\cdot\mathbf{C})\mathbf{B} - (\mathbf{A}\cdot\mathbf{B})\mathbf{C}. \quad \mathbf{A}\times(\mathbf{B}\times\mathbf{C}) + \mathbf{B}\times(\mathbf{C}\times\mathbf{A}) + \mathbf{C}\times(\mathbf{A}\times\mathbf{B}) = 0.$$

$$(\mathbf{A}\times\mathbf{B})\cdot(\mathbf{C}\times\mathbf{D}) = \mathbf{A}\cdot\{\mathbf{B}\times(\mathbf{C}\times\mathbf{D})\} = (\mathbf{A}\cdot\mathbf{C})(\mathbf{B}\cdot\mathbf{D}) - (\mathbf{B}\cdot\mathbf{C})(\mathbf{A}\cdot\mathbf{D}).$$

Scalar triple product $[\mathbf{ABC}] \equiv \mathbf{A}\cdot(\mathbf{B}\times\mathbf{C}) = \mathbf{B}\cdot(\mathbf{C}\times\mathbf{A}) = \mathbf{C}\cdot(\mathbf{A}\times\mathbf{B}) = \begin{vmatrix} A_x & A_y & A_z \\ B_x & B_y & B_z \\ C_x & C_y & C_z \end{vmatrix}.$

$$[\mathbf{BCD}]\mathbf{A} + [\mathbf{ACD}]\mathbf{B} + [\mathbf{ABD}]\mathbf{C} = [\mathbf{ABC}]\mathbf{D}. \quad [\mathbf{ABC}][\mathbf{EFG}] = \begin{vmatrix} \mathbf{A}\cdot\mathbf{E} & \mathbf{A}\cdot\mathbf{F} & \mathbf{A}\cdot\mathbf{G} \\ \mathbf{B}\cdot\mathbf{E} & \mathbf{B}\cdot\mathbf{F} & \mathbf{B}\cdot\mathbf{G} \\ \mathbf{C}\cdot\mathbf{E} & \mathbf{C}\cdot\mathbf{F} & \mathbf{C}\cdot\mathbf{G} \end{vmatrix}.$$

(II) Differentiation of a Vector Field (→ 442 Vectors)

$$\nabla \equiv \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z} \qquad \text{(Nabla)},$$

$$\operatorname{grad}\varphi \equiv \nabla\varphi = \frac{\partial\varphi}{\partial x}\mathbf{i} + \frac{\partial\varphi}{\partial y}\mathbf{j} + \frac{\partial\varphi}{\partial z}\mathbf{k} \qquad \text{(gradient of } \varphi),$$

$$\operatorname{rot}\mathbf{A} \equiv \nabla\times\mathbf{A} = \left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}\right)\mathbf{i} + \left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}\right)\mathbf{j} + \left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right)\mathbf{k} \qquad \text{(rotation of } \mathbf{A}),$$

$$\operatorname{div}\mathbf{A} \equiv \nabla\cdot\mathbf{A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} \qquad \text{(divergence of } \mathbf{A}),$$

$$\Delta\varphi \equiv \nabla^2\varphi \equiv \operatorname{div}\operatorname{grad}\varphi = \frac{\partial^2\varphi}{\partial x^2} + \frac{\partial^2\varphi}{\partial y^2} + \frac{\partial^2\varphi}{\partial z^2} \qquad \text{(Laplacian of } \varphi).$$

$$\operatorname{grad}(\varphi\psi) = \varphi\operatorname{grad}\psi + \psi\operatorname{grad}\varphi,$$

$$\operatorname{grad}(\mathbf{A}\cdot\mathbf{B}) = (\mathbf{B}\cdot\operatorname{grad})\mathbf{A} + (\mathbf{A}\cdot\operatorname{grad})\mathbf{B} + \mathbf{A}\times\operatorname{rot}\mathbf{B} + \mathbf{B}\times\operatorname{rot}\mathbf{A},$$

$$\operatorname{rot}(\varphi\mathbf{A}) = \varphi\operatorname{rot}\mathbf{A} - \mathbf{A}\times\operatorname{grad}\varphi, \quad \operatorname{rot}(\mathbf{A}\times\mathbf{B}) = (\mathbf{B}\cdot\operatorname{grad})\mathbf{A} - (\mathbf{A}\cdot\operatorname{grad})\mathbf{B} + \mathbf{A}\operatorname{div}\mathbf{B} - \mathbf{B}\operatorname{div}\mathbf{A},$$

$$\operatorname{div}(\varphi\mathbf{A}) = \varphi\operatorname{div}\mathbf{A} + \mathbf{A}\cdot\operatorname{grad}\varphi, \quad \operatorname{div}(\mathbf{A}\times\mathbf{B}) = \mathbf{B}\cdot\operatorname{rot}\mathbf{A} - \mathbf{A}\cdot\operatorname{rot}\mathbf{B}.$$

$$\operatorname{rot}\operatorname{grad}\varphi = 0, \quad \operatorname{div}\operatorname{rot}\mathbf{A} = 0. \quad \Delta\mathbf{A} = \operatorname{grad}\operatorname{div}\mathbf{A} - \operatorname{rot}\operatorname{rot}\mathbf{A}.$$

$$\Delta(f\circ\varphi) = (df/d\varphi)\Delta\varphi + (d^2f/d\varphi^2)(\operatorname{grad}\varphi)^2, \quad \Delta(\varphi\psi) = \varphi\Delta\psi + \psi\Delta\varphi + 2(\operatorname{grad}\varphi\cdot\operatorname{grad}\psi).$$

(III) Integration of a Vector Field (→ 94 Curvilinear Integrals and Surface Integrals, 442 Vectors)

Let D be a 3-dimensional domain, B its boundary, dV the volume element of D, dS the surface element of B, and $d\mathbf{S}=\mathbf{n}\,dS$, where $\mathbf{n}$ is the outer normal vector of the surface B. We have

Gauss's formula
$$\iiint_D \operatorname{div}\mathbf{A}\,dV=\iint_B d\mathbf{S}\cdot\mathbf{A}=\iint_B(\mathbf{n}\cdot\mathbf{A})\,dS,$$
$$\iiint_D \operatorname{rot}\mathbf{A}\,dV=\iint_B d\mathbf{S}\times\mathbf{A}=\iint_B(\mathbf{n}\times\mathbf{A})\,dS,$$
$$\iiint_D \operatorname{grad}\varphi\,dV=\iint_B \varphi\,d\mathbf{S};$$

Green's formula
$$\iint_B \varphi\frac{\partial\psi}{\partial n}\,dS=\iiint_D(\varphi\Delta\psi+\operatorname{grad}\varphi\cdot\operatorname{grad}\psi)\,dV,$$
$$\iint_B\left(\varphi\frac{\partial\psi}{\partial n}-\psi\frac{\partial\varphi}{\partial n}\right)dS=\iiint_D(\varphi\Delta\psi-\psi\Delta\varphi)\,dV,$$
$$4\pi\varphi(x_0)=-\iiint_D\frac{\Delta\varphi}{r}\,dV+\iint_B\left\{\frac{1}{r}\frac{\partial\varphi}{\partial n}-\varphi\frac{\partial}{\partial n}\left(\frac{1}{r}\right)\right\}dS,$$

where r is the distance from the point x_0.

Let B be a bordered surface with a boundary curve Γ, ds the line element of Γ, dS the surface element of B, and $d\mathbf{s}=\mathbf{t}\,ds$, $d\mathbf{S}=\mathbf{n}\,dS$, for $\mathbf{t}$ the unit tangent vector of Γ and under the proper choice of the positive direction for the surface normal $\mathbf{n}$. We have

Stokes's formula
$$\iint_B d\mathbf{S}\cdot\operatorname{rot}\mathbf{A}=\oint_\Gamma \mathbf{A}\cdot d\mathbf{s}=\oint_\Gamma(\mathbf{t}\cdot\mathbf{A})\,ds,\qquad \iint_B d\mathbf{S}\times\operatorname{grad}\varphi=\oint_\Gamma\varphi\,d\mathbf{s}.$$

If the domain D is simply connected, and the vector field $\mathbf{V}$ tends sufficiently rapidly to 0 near the boundary of D and at infinity, we have

Helmholtz's theorem
$$\mathbf{V}=\operatorname{grad}\varphi+\operatorname{rot}\mathbf{A},\quad \varphi=-\iiint_D\frac{\operatorname{div}\mathbf{V}}{4\pi r}\,dV,\quad \mathbf{A}=\iiint_D\frac{\operatorname{rot}\mathbf{V}}{4\pi r}\,dV.$$

(IV) Moving Coordinate System

Denote differentiation with respect to the rest and the moving systems by d/dt, d^*/dt, respectively. Let the relative velocity of the systems be $\mathbf{v}$. Then we have

$$\frac{d\varphi}{dt}=\frac{d^*\varphi}{dt}-\mathbf{v}\cdot\operatorname{grad}\varphi,\quad \frac{d\mathbf{A}}{dt}=\frac{d^*\mathbf{A}}{dt}-[\mathbf{v}\cdot\operatorname{grad}\mathbf{A}-(\mathbf{A}\cdot\operatorname{grad})\mathbf{v}].$$

With respect to rotating coordinates we have

$$\mathbf{v}=\mathbf{w}\times\mathbf{r}.$$
$$\frac{d\mathbf{A}}{dt}=\frac{d^*\mathbf{A}}{dt}+[\mathbf{w}+\mathbf{A}-((\mathbf{w}\times\mathbf{r})\cdot\operatorname{grad})\mathbf{A}]$$

When the domain of integration is also a function of t,

$$\frac{d}{dt}\int\mathbf{A}\cdot d\mathbf{s}=\int\left\{\frac{\partial\mathbf{A}}{\partial t}+\operatorname{grad}(\mathbf{v}\cdot\mathbf{A})-\mathbf{v}\times\operatorname{rot}\mathbf{A}\right\}\cdot d\mathbf{s},$$
$$\frac{d}{dt}\iint\mathbf{A}\cdot d\mathbf{S}=\iint\left\{\frac{\partial\mathbf{A}}{\partial t}+\operatorname{rot}(\mathbf{A}\times\mathbf{v})+\mathbf{v}\operatorname{div}\mathbf{A}\right\}\cdot d\mathbf{S},$$
$$\frac{d}{dt}\iiint\varphi\,dV=\iiint\left\{\frac{\partial\varphi}{\partial t}+(\mathbf{v}\cdot\operatorname{grad}\varphi)+\varphi\operatorname{div}\mathbf{v}\right\}dV=\iiint\frac{\partial\varphi}{\partial t}\,dV+\iint\varphi\mathbf{v}\cdot d\mathbf{S}.$$

(V) Curvilinear Coordinates (→ 90 Coordinates)

Let $(x_1,\ldots,x_n)$ be rectangular coordinates in an n-dimensional Euclidean space. If

$$x_j=\varphi_j(u_1,\ldots,u_n)\quad(j=1,\ldots,n),\qquad J\equiv\det(\partial\varphi_j/\partial u_k)\neq0,$$

the system $(u_1,\ldots,u_n)$ may be taken as a coordinate system of an n-dimensional space, and the

original space is a Riemannian manifold with the first fundamental form

$$g_{jk}=\sum_{i=1}^{n}\frac{\partial\varphi_i}{\partial u_j}\frac{\partial\varphi_i}{\partial u_k}\qquad (j,k=1,\dots,n),$$

$$g\equiv\det(g_{jk})=J^2.$$

When the metric is of the diagonal form $g_{jk}=g_j^2\delta_{jk}$, the coordinate system $(u_1,\dots,u_n)$ is called an orthogonal curvilinear coordinate system or an isothermal curvilinear coordinate system. In such a case we have $J=g_1\dots g_n$, and the line element is given by $ds^2=\sum_{j=1}^n g_j^2 du_j^2$.

For a scalar f and a vector $\xi=(\xi_1,\dots,\xi_n)$, we have

$$(\operatorname{grad} f)_j=\frac{1}{g_j}\frac{\partial f}{\partial u_j}\quad (j=1,\dots,n),\qquad \Delta f=\frac{1}{J}\sum_{j=1}^{n}\frac{\partial}{\partial u_j}\left(\frac{J}{g_j^2}\frac{\partial f}{\partial u_j}\right),$$

$$\operatorname{div}\xi=\frac{1}{J}\sum_{j=1}^{n}\frac{\partial}{\partial u_j}\left(\frac{J}{g_j}\xi_j\right),\quad (\operatorname{rot}\xi)_{jk}=\frac{1}{g_jg_k}\left[\frac{\partial(g_k\xi_k)}{\partial u_j}-\frac{\partial(g_j\xi_j)}{\partial u_k}\right]\qquad (j,k=1,\dots,n).$$

When $n=2$, the rot may be considered a scalar, rot $\xi=(\operatorname{rot}\xi)_{12}$, and when $n=3$, the rot may be considered a vector, with components

$$\operatorname{rot}\xi=((\operatorname{rot}\xi)_{23},(\operatorname{rot}\xi)_{31},(\operatorname{rot}\xi)_{12}).$$

The following are examples of orthogonal coordinates.

(1) Planar Curvilinear Coordinates. In the present Section (1), we put

$$x_1=x,\quad x_2=y,\quad u_1=u,\quad u_2=v,\quad g_1=p,\quad g_2=q.$$

$$ds^2=p^2dx^2+q^2dy^2,\qquad J\equiv\partial(x,y)/\partial(u,v)=\sqrt{pq}\ .$$

Planar orthogonal curvilinear coordinates may be represented in the form $x+iy=F(U+iV)$, F being a complex analytic function, by suitable choice of the functions $U=U(u)$, $V=V(v)$.

(i) Polar Coordinates (r,θ) (Fig. 4).

$$x=r\cos\theta,\quad y=r\sin\theta;\quad x+iy=\exp(\log r+i\theta).$$

$$r=\sqrt{x^2+y^2}\ ,\quad \theta=\arctan(y/x).$$

$$p=1,\quad q=r,\quad J=r,\quad ds^2=dr^2+r^2d\theta^2.$$

$$\Delta f=\frac{\partial^2 f}{\partial r^2}+\frac{1}{r}\frac{\partial f}{\partial r}+\frac{1}{r^2}\frac{\partial^2 f}{\partial\theta^2}.$$

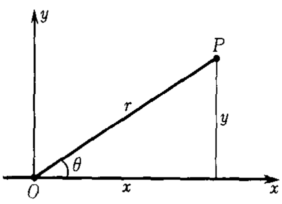


Fig. 4

(ii) Elliptic Coordinates (μ,ν) (Fig. 5). Among the family of confocal conics

$$\frac{x^2}{a^2+\rho}+\frac{y^2}{b^2+\rho}=1\quad (a>b),$$

there are two values of ρ for which the curve passes through a given point $P(x,y)$. Denote the two values of ρ by μ and ν, where $\mu>-b^2>\nu>-a^2$. The curve corresponding to $\rho=\mu$ or $\rho=\nu$ is an ellipse or a hyperbola, respectively. Then we have the relations

$$x^2=(\mu+a^2)(\nu+a^2)/(a^2-b^2),\qquad y^2=(\mu+b^2)(\nu+b^2)/(b^2-a^2).$$

Let the common foci be $(\pm c,0)$ $(c^2=a^2-b^2)$. Then we have

$$r_1=\sqrt{(x-c)^2+y^2}\ ,\qquad r_2=\sqrt{(x+c)^2+y^2}$$

where r_1, r_2 are the distances from the two foci as in Fig. 5, and

$$4(a^2+\mu)=(r_1+r_2)^2,\qquad 4(a^2+\nu)=(r_1-r_2)^2.$$

$$p=\frac{1}{2}\sqrt{\frac{\mu-\nu}{(\mu+a^2)(\mu+b^2)}}\ ,\qquad q=\frac{1}{2}\sqrt{\frac{\nu-\mu}{(\nu+a^2)(\nu+b^2)}}$$

(iii) Parabolic Coordinates (α,β) (Fig. 6). Among the family of parabolas $y^2=4\rho(x+\rho)$ with the focus at the origin and having the x-axis as the principal axis, there are two values of ρ for which the curve passes through a given point $P(x,y)$. Denote the two values of ρ by α, β $(\alpha>0>\beta)$. We have $x=-(\alpha+\beta)$, $y=\sqrt{-4\alpha\beta}$.

(iv) Equilateral (or Rectangular) Hyperbolic Coordinates (u,v) (Fig. 7). This is a system that

replaces $x/2$, $y/2$ in (iii) by $-y$ and x, respectively, with $\sqrt{\alpha}=u$, $\sqrt{-\beta}=v$. The relations are

$$x=uv,\quad y=(u^2-v^2)/2;\quad x+iy=i(u-iv)^2/2,\quad u^2,v^2=\sqrt{x^2+y^2}\pm y,\quad p=q=\sqrt{u^2+v^2}.$$

The curves $x=$ constant or $y=$ constant are equilateral hyperbolas.

(v) Bipolar Coordinates (ξ,η) (Fig. 8). These coordinates represent a point $P(x,y)$ on a plane as the intersection of the family of circles passing through two fixed points $(\pm a,0)$ and the family of loci on which the ratio of distances from the same two fixed points $(\pm a,0)$ is constant. The latter is the set of Apollonius' circles. The relations are

$$x=\frac{a\sinh\xi}{\cosh\xi+\cos\eta},\qquad y=\frac{a\sin\eta}{\cosh\xi+\cos\eta}\qquad(-\infty<\xi<\infty,\ 0\leqslant\eta\leqslant 2\pi).$$

$$p=q=\frac{a}{\cosh\xi+\cos\eta}.$$

(2) Curvilinear Coordinates in 3-Dimensional Space. In the present Section (2), we put $x_1=x$, $x_2=y$, $x_3=z$.

(i) Circular Cylindrical Coordinates (Cylindrical Coordinates) (ρ,φ,z) (Fig. 9).

$$x=\rho\cos\varphi,\quad y=\rho\sin\varphi,\quad z=z.$$

$$ds^2=d\rho^2+\rho^2d\varphi^2+dz^2,\qquad J=\rho.\qquad \Delta f=\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial f}{\partial\rho}\right)+\frac{1}{\rho^2}\frac{\partial^2 f}{\partial\varphi^2}+\frac{\partial^2 f}{\partial z^2}.$$

(ii) Polar Coordinates (Spherical Coordinates) (Fig. 9).

$$x=r\sin\theta\cos\varphi,\quad y=r\sin\theta\sin\varphi,\quad z=r\cos\theta.$$

$$r=\sqrt{x^2+y^2+z^2},\quad \varphi=\arctan(y/x),\quad \theta=\arctan\left(\sqrt{x^2+y^2}/z\right).$$

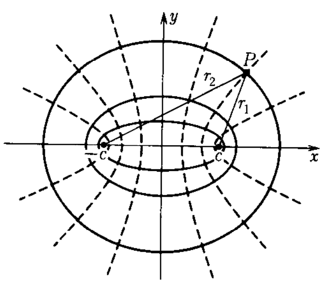


Fig. 5

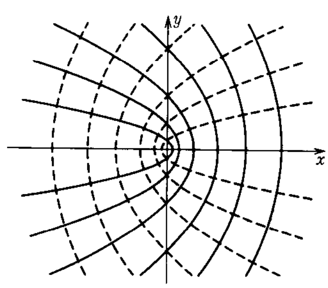


Fig. 6

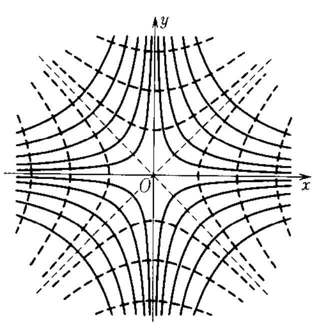


Fig. 7

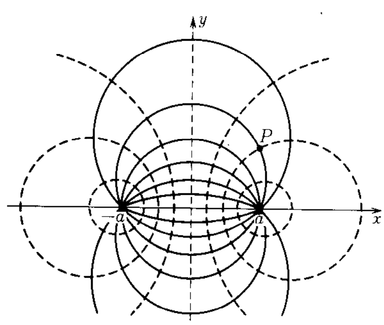


Fig. 8

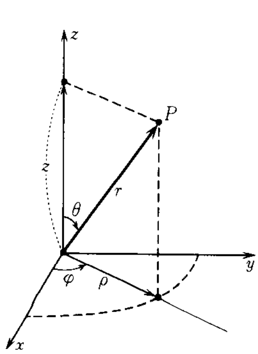


Fig. 9

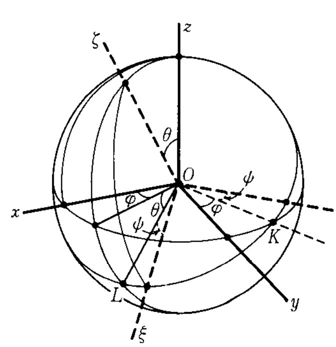


Fig. 10

The angles φ and θ are called azimuth and zenith angle, respectively. We further have

$$ds^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta \, d\varphi^2, \qquad J = r^2 \sin\theta.$$

$$\Delta f = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial f}{\partial r}\right) + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta \frac{\partial f}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 f}{\partial\varphi^2}.$$

(iii) Euler's Angles (Fig. 10). Let (x,y,z) and (ξ,η,ζ) be two linear orthogonal coordinate systems with common origin O. Denote the angle between the z-axis and the ζ-axis by θ; the angle between the zx-plane and the $z\zeta$-plane by φ; and the angle between the η-axis and the intersection OK of the xy-plane and the $\xi\eta$-plane (or the angle between the ξ-axis and the intersection OL of the $z\zeta$-plane and the $\xi\eta$-plane) by ψ. The angles θ, φ, and ψ are called Euler's angles. The direction cosines of one coordinate axis with respect to the other coordinate system are as follows:

	x	y	z
ξ	$\cos\varphi\cos\theta\cos\psi - \sin\varphi\sin\psi$	$\sin\varphi\cos\theta\cos\psi + \cos\varphi\sin\psi$	$-\sin\theta\cos\psi$
η	$-\cos\varphi\cos\theta\sin\psi - \sin\varphi\cos\psi$	$-\sin\varphi\cos\theta\sin\psi + \cos\varphi\cos\psi$	$\sin\theta\sin\psi$
ζ	$\cos\varphi\sin\theta$	$\sin\varphi\sin\theta$	$\cos\theta$

(iv) Rotational (or Revolutional) Coordinates (u,v,ρ). Let (u,v) be curvilinear coordinates (Section (1)) on the $z\rho$-plane. The rotational coordinates (u,v,ρ) are given by the combination of $x=\rho\cos\varphi$, $y=\rho\sin\varphi$ with the coordinates on the $z\rho$-plane. We have

$$ds^2 = p^2 du^2 + q^2 dv^2 + \rho^2 d\varphi^2,$$

where p, q are the corresponding values for the coordinates (u,v).

(v) Generalized Cylindrical Coordinates (u,v,z). These are a combination of curvilinear coordinates (u,v) on the xy-plane with z. We have

$$ds^2 = p^2 du^2 + q^2 dv^2 + dz^2.$$

For various selections of (u,v) we have coordinates as follows:

(u,v)	Rotational Coordinate System	Generalized Cylindrical Coordinate System
Linear rectangular coordinates	Circular cylindrical coordinates	Linear rectangular coordinates
Polar coordinates ((1)(i))	Spherical coordinates	Circular cylindrical coordinates
Elliptic coordinates ((1)(ii))	Spheroidal coordinates[(1)]	Elliptic cylindrical coordinates
Parabolic coordinates ((1)(iii))	Rotational parabolic coordinates[(2)]	Parabolic cylindrical coordinates
Equilateral hyperbolic coordinates ((1)(iv))	Rotational hyperbolic coordinates	Hyperbolic cylindrical coordinates
Bipolar coordinates ((1)(v))	Toroidal coordinates[(3)] Bipolar coordinates[(4)]	Bipolar cylindrical coordinates

Notes

(1) When the ρ-axis is a minor or major axis, we have prolate or oblate spheroidal coordinates, respectively.
(2) We take the z-axis as the common principal axis of the parabolas.
(3) Where the line passing through two fixed points is the ρ-axis.
(4) Where the line passing through two fixed points is the z-axis.

(vi) Ellipsoidal Coordinates (λ,μ,ν) (Fig. 11). Among the family of confocal quadrics

$$\frac{x^2}{a^2+\rho} + \frac{y^2}{b^2+\rho} + \frac{z^2}{c^2+\rho} = 1 \quad (a>b>c>0),$$

there are three values of ρ for which the surface passes through a given point $P(x,y,z)$. Denote the three values of ρ by λ, μ, ν, where $\lambda > -c^2 > \mu > -b^2 > \nu > -a^2$. The surfaces corresponding to $\rho=\lambda$, $\rho=\mu$, and $\rho=\nu$ are an ellipsoid, a hyperboloid of one sheet, and a hyperboloid of two sheets, respectively. We have

$$x^2=\frac{h(a)}{(a^2-b^2)(a^2-c^2)},\qquad y^2=\frac{h(b)}{(b^2-c^2)(b^2-a^2)},$$

$$z^2=\frac{h(c)}{(c^2-a^2)(c^2-b^2)};\qquad h(\alpha)\equiv(\lambda+\alpha^2)(\mu+\alpha^2)(\nu+\alpha^2).$$

$$g_1=\frac{\sqrt{(\lambda-\mu)(\lambda-\nu)}}{2\rho(\lambda)},\qquad g_2=\frac{\sqrt{(\mu-\nu)(\mu-\lambda)}}{2\rho(\mu)},$$

$$g_3=\frac{\sqrt{(\nu-\lambda)(\nu-\mu)}}{2\rho(\nu)};\qquad \rho(t)\equiv\sqrt{(t+a^2)(t+b^2)(t+c^2)}\ .$$

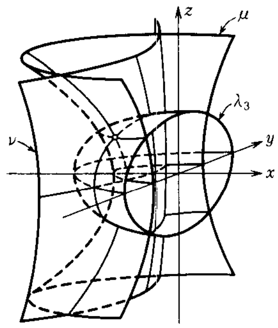


Fig. 11

4. Differential Geometry

(I) Classical Differential Geometry (→ 111 Differential Geometry of Curves and Surfaces)

(1) Plane Curves (Fig. 12). At a point $P(x_0,y_0)$ on a curve $y=f(x)$, the equation of the tangent line is $y-y_0=f'(x_0)(x-x_0)$,

$$PT=\left|y_0\sqrt{1+y_0'^2}\,/y_0'\right|,$$

and the tangential shadow $TM=y_0/y_0'$. The equation of the normal line is $f'(x_0)(y-y_0)+(x-x_0)=0$,

$$PN=\left|y_0\sqrt{1+y_0'^2}\,\right|,$$

and the normal shadow $MN=y_0y_0'$. The slope of the tangent is $\tan\alpha=f'(x_0)=y_0'$. The curvature at P is

$$\kappa=1/PQ=f''(x_0)/\left[1+f'(x_0)^2\right]^{3/2}$$

The coordinates of the center of curvature Q are

$$\left(x_0-f'(x_0)\left[1+f'(x_0)^2\right]/f''(x_0),\ \ f(x_0)+[1+f'(x_0)^2]/f''(x_0)\right).$$

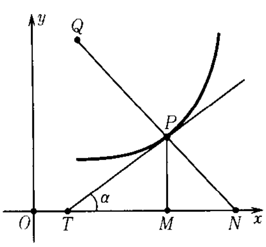


Fig. 12

(2) Space Curves $x_i=x_i(t)$ $(i=1,2,3)$, or $\mathbf{x}=\mathbf{x}(t)$. The line element of a curve $\mathbf{x}=\mathbf{x}(t)$ is

$$ds=\sqrt{(dx_1)^2+(dx_2)^2+(dx_3)^2}=\sqrt{\sum_{\alpha=1}^{3}\dot{x}_\alpha^2}\ dt\qquad\left(\cdot=\frac{d}{dt}\right).$$

The curvature is

$$\kappa=\sqrt{\sum\ddot{x}_\alpha^2-\ddot{s}^2/\dot{s}^2}$$

For $t=s$ (arc length), the curvature is $\kappa=\sqrt{(\sum x_\alpha''^2)}$, and the torsion is $\tau=[\det(x_\alpha',x_\alpha'',x_\alpha''')_{\alpha=1,2,3}]/\kappa^2$, where $'=d/ds$. When we denote Frenet's frame by $(\boldsymbol{\xi},\boldsymbol{\eta},\boldsymbol{\zeta})$, we have $\boldsymbol{\xi}=\mathbf{x}'$, $\boldsymbol{\eta}=\boldsymbol{\xi}'/\kappa$, $\boldsymbol{\zeta}=\boldsymbol{\xi}\times\boldsymbol{\eta}$ (vector product).

The Frenet-Serret formulas are

$$\xi' = \kappa\eta, \quad \eta' = -\kappa\xi + \tau\zeta, \quad \zeta' = -\tau\eta.$$

(3) Surface in 3-Dimensional Space $x_\alpha = x_\alpha(u_1, u_2)$ $(\alpha = 1,2,3)$. The first fundamental form of the surface is

$$g_{jk} = \sum_{\alpha=1}^{3} \frac{\partial x_\alpha}{\partial u_j}\frac{\partial x_\alpha}{\partial u_k} \quad (j,k = 1,2). \quad g = \det(g_{jk}) > 0.$$

Let (g^{jk}) be the inverse matrix of (g_{jk}). The tangent plane at the point $x_\alpha^{(0)}$ is given by

$$\det\left(x_\alpha - x_\alpha^{(0)}, \quad (\partial x_\alpha/\partial u_1)^{(0)}, \quad (\partial x_\alpha/\partial u_2)^{(0)}\right) = 0.$$

The normal line at the point $x_\alpha^{(0)}$ is given by $x_\alpha - x_\alpha^{(0)} = t\nu_\alpha^{(0)}$, where t is a parameter, and ν_α is the unit normal vector, given by

$$\nu_\alpha = \frac{1}{\sqrt{g}}\frac{\partial(x_\beta, x_\gamma)}{\partial(u_1, u_2)} \quad \left(\delta_{\alpha\beta\gamma}^{123} = +1\right).$$

The second fundamental form is

$$h_{jk} \equiv \sum_{\alpha=1}^{3} \nu_\alpha \frac{\partial^2 x_\alpha}{\partial u_j \partial u_k} = -\sum_{\alpha=1}^{3} \frac{\partial \nu_\alpha}{\partial u_j}\frac{\partial x_\alpha}{\partial u_k}. \quad h \equiv \det(h_{jk}).$$

The principal radii of curvature R_1, R_2 are the roots of the quadratic equation

$$\frac{1}{R^2} - \sum_{j,k} g^{jk} h_{jk} \frac{1}{R} + \frac{h}{g} = 0.$$

The mean curvature (or Germain's curvature) is

$$H \equiv \frac{1}{2}\left(\frac{1}{R_1} + \frac{1}{R_2}\right) = \frac{1}{2}\sum_{j,k} g^{jk} h_{jk},$$

and $H = 0$ is the condition for the given surface to be a minimal surface. The Gaussian curvature (or total curvature) is

$$K \equiv \frac{1}{R_1 R_2} = \frac{h}{g},$$

and $K = 0$ is the condition for the surface to be developable.

We use the notations of Riemannian geometry, with g_{jk} the fundamental tensor:

$$\frac{\partial^2 x_\alpha}{\partial u_j \partial u_k} = \sum_{a=1}^{3} \left\{ {a \atop jk} \right\} \frac{\partial x_\alpha}{\partial u_a} + h_{jk}\nu_\alpha \qquad \text{(Gauss's formula)}.$$

$$R_{ijkl} = h_{jl}h_{ik} - h_{jk}h_{il} \quad \text{(Gauss's equation)}. \qquad h_{jk;l} = h_{jl;k} \quad \text{(Codazzi-Mainardi equation)}.$$

$$K = \frac{R_{1212}}{g} = \frac{R}{2} = \frac{1}{\sqrt{g}}\left[\frac{\partial}{\partial u_2}\left(\frac{\sqrt{g}}{g_{11}}\left\{ {2 \atop 11} \right\}\right) - \frac{\partial}{\partial u_1}\left(\frac{\sqrt{g}}{g_{11}}\left\{ {2 \atop 12} \right\}\right)\right]$$

$$= \frac{1}{\sqrt{g}}\left[\frac{\partial}{\partial u_1}\left(\frac{\sqrt{g}}{g_{22}}\left\{ {1 \atop 22} \right\}\right) - \frac{\partial}{\partial u_2}\left(\frac{\sqrt{g}}{g_{22}}\left\{ {1 \atop 12} \right\}\right)\right].$$

$$\frac{\partial \nu_\alpha}{\partial u_j} = -\sum_{k,l} h_{jk} g^{kl} \frac{\partial x_\alpha}{\partial u_l} \qquad \text{(Weingarten's formula)}.$$

The third fundamental form is given by

$$l_{jk} \equiv \sum_{\alpha=1}^{3} \frac{\partial \nu_\alpha}{\partial u_j}\frac{\partial \nu_\alpha}{\partial u_k} = \sum_{s,t} g^{st} h_{js} h_{kt} = 2Hh_{jk} - Kg_{jk}. \quad \det(l_{jk}) = K^2 g = Kh.$$

(4) Geodesic Curvature. Let $C: u_i = u_i(s)$ be a curve on a surface S and ρ be the curvature of C at a point P. Let θ be the angle between the osculating plane of C and the plane tangent to S. The geodesic curvature ρ_g of C at P is given by

$$\rho_g = \rho\cos\theta = \sqrt{g}\,\det\left(\frac{du_i}{ds}, \; \frac{d^2u_i}{ds^2} + \sum_{j,k}\left\{ {i \atop jk} \right\}\frac{du_j}{ds}\frac{du_k}{ds}\right)_{i=1,2}$$

$\rho_g = 0$ is the condition for C to be a geodesic. Let D be a simply connected domain on the surface S, whose boundary Γ consists of n smooth curves. Let θ_α be the outer angle at the intersection of two consecutive curves ($\alpha = 1, \ldots, n$). Then we have the Gauss-Bonnet formula:

$$\int_\Gamma \rho_g ds + \iint_D K dS = 2\pi - \sum_{\alpha=1}^{n} \theta_\alpha.$$

(II) Riemannian Geometry, Tensor Calculus (→ 417 Tensor Calculus)

In the present section, we use Einstein's convention (omission of the summation symbol).

(1) Numerical Tensor.

$$\text{Kronecker's } \delta \quad \delta_{jk},\ \delta^{jk},\ \delta^j_k = \begin{cases} 1 & (j=k) \\ 0 & (j \neq k). \end{cases}$$

$$\delta^{j_1 \cdots j_p}_{k_1 \ldots k_p} = \det\left(\delta^{j_\mu}_{k_\nu}\right)_{\mu,\nu=1,\ldots,p} = \begin{cases} 0 & (\{j_\mu\} \neq \{k_\nu\}), \\ +1 & (\{j_\mu\} = \{k_\nu\} \text{ and } (j_\mu) \text{ is an even permutation of } (k_\nu)), \\ -1 & (\{j_\mu\} = \{k_\nu\}) \text{ and } (j_\mu) \text{ is an odd permutation of } (k_\nu). \end{cases}$$

$$\text{Eddington's } \varepsilon \quad \varepsilon_{j_1 \ldots j_n} = \delta^{1 \ldots n}_{j_1 \ldots j_n}, \quad \varepsilon^{j_1 \cdots j_n} = \delta^{j_1 \cdots j_n}_{1 \ldots n}.$$

$$\delta^{j_1 \cdots j_p j_{p+1} \cdots j_n}_{k_1 \cdots k_p j_{p+1} \cdots j_n} = (n-p)!\,\delta^{j_1 \cdots j_p}_{k_1 \ldots k_p}. \quad \det(a^u_v)_{\mu,\nu=1,\ldots,n} = \varepsilon^{j_1 \cdots j_n} a^1_{j_1} a^2_{j_2} \ldots a^n_{j_n} = \varepsilon_{j_1 \ldots j_n} a^{j_1}_1 a^{j_2}_2 \ldots a^{j_n}_n.$$

(2) Fundamental Objects in Riemannian Geometry. Let g_{jk} be the fundamental tensor, and (g^{jk}) be the inverse matrix of (g_{jk}). We put $g \equiv \det(g_{jk})$.

The Christoffel symbol is

$$\left\{ {i \atop jk} \right\} = \frac{1}{2} g^{ia} \left[\frac{\partial g_{aj}}{\partial x^k} + \frac{\partial g_{ak}}{\partial x^j} - \frac{\partial g_{jk}}{\partial x^a} \right] = \left\{ {i \atop kj} \right\}, \quad \left\{ {a \atop ak} \right\} = \frac{\partial \log \sqrt{g}}{\partial x^k},$$

which has the transformation rule

$$\left\{ {\bar{i} \atop \bar{j}\bar{k}} \right\} = \frac{\partial \bar{x}^{\bar{i}}}{\partial x^i} \left(\frac{\partial x^j}{\partial \bar{x}^{\bar{j}}} \frac{\partial x^k}{\partial \bar{x}^{\bar{k}}} \left\{ {i \atop jk} \right\} + \frac{\partial^2 x^i}{\partial \bar{x}^{\bar{j}} \partial \bar{x}^{\bar{k}}} \right)$$

under a coordinate transformation.

A geometrical object Γ^i_{jk} with a similar transformation rule is called the coefficient of the affine connection. The torsion tensor is

$$S^i_{jk} \equiv \Gamma^i_{jk} - \Gamma^i_{kj}.$$

The equation of a geodesic is

$$\frac{d^2 x^i}{ds^2} + \left\{ {i \atop jk} \right\} \frac{dx^j}{ds} \frac{dx^k}{ds} = 0.$$

The covariant derivative of a tensor of weight W with respect to a coefficient of affine connection Γ^i_{jk} is given by

$$T^{j_1 \cdots j_p}_{k_1 \ldots k_q | l} \equiv \partial T^{j_1 \cdots j_p}_{k_1 \cdots k_q} / \partial x^l + \sum_{\nu=1}^{p} T^{j_1 \cdots j_{\nu-1} a j_{\nu+1} \cdots j_p}_{k_1 \cdots\cdots\cdots k_q} \Gamma^{j_\nu}_{al} - \sum_{\mu=1}^{q} T^{j_1 \cdots\cdots\cdots j_p}_{k_1 \ldots k_{\mu-1} a k_{\mu+1} \ldots k_q} \Gamma^a_{lk_\mu} - W T^{j_1 \cdots j_p}_{k_1 \cdots k_q} \Gamma^a_{al}.$$

For the Christoffel symbol, we denote the covariant derivative by $;l$. Then we have the following formulas:

$$g_{jk;l} = 0, \quad g^{jk}{}_{;l} = 0, \quad \delta^j_{k;l} = 0, \quad \sqrt{g}\,\varepsilon_{j_1 \ldots j_n;l} = 0, \quad (1/\sqrt{g})\varepsilon^{j_1 \cdots j_n}{}_{;l} = 0.$$

For a scalar f $\quad \operatorname{grad} f = (f_{;j})$,

for a covariant vector v_j $\quad \operatorname{rot} v = (v_{j;k} - v_{k;j}) = (\partial v_j / \partial x^k - \partial v_k / \partial x^j)$,

and for a contravariant vector v^j $\quad \operatorname{div} v = v^j{}_{;j} = \dfrac{1}{\sqrt{g}} \dfrac{\partial(\sqrt{g} v^j}{\partial x^j}$.

Beltrami's differential operator of the first kind is

$$\Delta_1 f \equiv g^{jk} f_{;j} f_{;k}.$$

Beltrami's differential operator of the second kind is

$$\Delta_2 f = \operatorname{div}\operatorname{grad} f = \frac{1}{\sqrt{g}} \frac{\partial\left(\sqrt{g}\; g^{jk}\left(\partial f/\partial x^k\right)\right)}{\partial x^j}.$$

For a domain D with sufficiently smooth boundary Γ, we denote the directional derivative along the inner normal by $\partial/\partial n$, the volume element by dV, and the surface element on Γ by dS. Then we have Green's formulas,

$$\int_D (\Delta_1(\varphi\psi) + \psi\Delta_2\varphi)\, dV = -\int_\Gamma \psi \frac{\partial\varphi}{\partial n}\, dS, \qquad \int_D (\varphi\Delta_2\psi - \psi\Delta_2\varphi)\, dV = \int_\Gamma \left(\psi\frac{\partial\varphi}{\partial n} - \varphi\frac{\partial\psi}{\partial n}\right) dS.$$

We denote the curvature tensor with respect to the coefficients of a general affine connection Γ^i_{jk} by B^i_{jkl}, and by R^i_{jkl} when $\Gamma^i_{jk} = \left\{ {i \atop jk} \right\}$. We have the following formulas:

$$B^i_{jkl} \equiv \frac{\partial \Gamma^i_{jl}}{\partial x^k} - \frac{\partial \Gamma^i_{jk}}{\partial x^l} + \Gamma^a_{jl}\Gamma^i_{ak} - \Gamma^a_{jk}\Gamma^i_{al};$$

$$R_{ijkl} \equiv g_{ai} R^a_{jkl} = -R_{ijlk} = R_{klij} = \frac{1}{2}\left[\frac{\partial^2 g_{ik}}{\partial x^j \partial x^l} + \frac{\partial^2 g_{jl}}{\partial x^i \partial x^k} - \frac{\partial^2 g_{jk}}{\partial x^i \partial x^l} - \frac{\partial^2 g_{il}}{\partial x^j \partial x^k}\right]$$

$$+ g_{ab}\left(\left\{ {b \atop ik} \right\}\left\{ {a \atop jl} \right\} - \left\{ {b \atop il} \right\}\left\{ {a \atop jk} \right\}\right);$$

Bianchi's first identity $\quad R^i_{jkl} + R^i_{klj} + R^i_{ljk} = 0,$

$$-\left(B^i_{jkl} + B^i_{klj} + B^i_{ljk}\right) = 2\left(S^i_{jk|l} + S^i_{kl|j} + S^i_{lj|k}\right) + 4\left(S^i_{ja}S^a_{kl} + S^i_{ka}S^a_{lj} + S^i_{la}S^a_{jk}\right);$$

Bianchi's second identity $\quad R^i_{jkl;m} + R^i_{jlm;k} + R^i_{jmk;l} = 0,$

$$B^i_{jkl|m} + B^i_{jlm|k} + B^i_{jmk|l} = -2\left(B^i_{jma}S^a_{kl} + B^i_{jka}S^a_{lm} + B^i_{jla}S^a_{mk}\right);$$

Ricci's tensor $\quad R_{jk} \equiv -R^i_{jki} = R_{kj};$

scalar curvature $\quad R \equiv g^{jk} R_{jk};$

Ricci's formula $\quad T^{j_1 \cdots j_p}_{k_1 \cdots k_q|s|t} - T^{j_1 \cdots j_p}_{k_1 \cdots k_q|t|s}$

$$= -\sum_{\nu=1}^{p} T^{j_1 \cdots j_{\nu-1} a j_{\nu+1} \cdots j_p}_{k_1 \cdots\cdots k_q} B^{j_\nu}_{ast} + \sum_{\mu=1}^{q} T^{j_1 \cdots\cdots j_p}_{k_1 \ldots k_{\mu-1} a k_{\mu+1}, \ldots, k_q} B^a_{k_\mu st} + 2T^{j_1 \cdots j_p}_{k_1 \cdots k_q|l} S^l_{st} + W T^{j_1 \cdots j_p}_{k_1 \cdots k_q} B^a_{ast},$$

where S and B are the torsion and curvature tensors given above, respectively, and W is the weight of the tensor T.

(3) Special Riemannian Spaces (→ 364 Riemannian Manifolds). In the present Section (3), n means the dimension of the space.

(i) Space of Constant Curvature $R^i_{jkl} = \rho(g_{jl}\delta^i_k - g_{jk}\delta^i_l)$; $\rho = R/n(n-1)$,

(ii) Einstein Space $R_{jk} = \rho g_{jk}$, $\rho = R/n$, for $n \geq 3$, where R is a constant.

(iii) Locally Symmetric Riemannian Space $R^i_{jkl;m} = 0$.

(iv) Projectively Flat Space. Weyl's projective curvature tensor is defined by

$$W^i_{jkl} \equiv R^i_{jkl} + \frac{1}{n-1}(R_{jk}\delta^i_l - R_{jl}\delta^i_k).$$

The condition for the space to be projectively flat is given by $W^i_{jkl} = 0$, $R_{jk;l} = R_{jl;k}$.

If $n \geq 3$, the latter condition follows from the former condition, and the space reduces to a space of constant curvature. If $n = 2$, the former condition $W = 0$ always holds.

(v) Concircularly Flat Space $Z^i_{jkl} \equiv R^i_{jkl} + \dfrac{R}{n(n-1)}(g_{jk}\delta^i_l - g_{jl}\delta^i_k) = 0$. This space reduces to a space of constant curvature.

(vi) Conformally Flat Space. Weyl's conformal curvature tensor is defined by

$$C^i_{jkl} \equiv R^i_{jkl} + \frac{1}{n-2}(R_{jk}\delta^i_l - R_{jl}\delta^i_k + g_{jk}R^i_l - g_{jl}R^i_k) - \frac{R(g_{jk}\delta^i_l - g_{jl}\delta^i_k)}{(n-1)(n-2)},$$

$$\Pi_{jk} \equiv -\frac{R_{jk}}{(n-2)} + \frac{R g_{jk}}{2(n-1)(n-2)}.$$

The condition for the space to be conformally flat is given by $C^i_{jkl} = 0$, $\Pi_{jk;l} = \Pi_{jl;k}$.

If $n \geq 4$, the latter condition follows from the former condition, and if $n = 3$, the former condition $C = 0$ always holds.

5. Lie Algebras, Symmetric Riemannian Spaces, and Singularities

(I) The Classification of Complex Simple Lie Algebras and Compact Real Simple Lie Algebras
(→ 248 Lie Algebras)

(1) Lie Algebra. The unitary restriction of a noncommutative finite-dimensional complex simple Lie algebra $\mathfrak{g}$ is a compact real simple Lie algebra $\mathfrak{g}_u$, and $\mathfrak{g}$ is given by the complexification $\mathfrak{g}_u^c$ of $\mathfrak{g}_u$. There exists a bijective correspondence between the classifications of these two kinds of Lie algebras. Using Dynkin diagrams, the classification is done as in Fig. 14 (→ 248 Lie Algebras). The system of fundamental roots $\{\alpha_1, \ldots, \alpha_l\}$ of a simple Lie algebra $\mathfrak{g}$ is in one-to-one correspondence with the vertices of a Dynkin diagram shown by simple circles in Fig. 14. The number of simple circles coincides with the rank l of $\mathfrak{g}$. The double circle in Fig. 14 means -1 times the highest root θ. Sometimes we mean by the term "Dynkin diagram" the diagram without the double circle and the lines issuing from it. Here we call the diagram with double circle representing $-\theta$ the extended Dynkin diagram. Corresponding to the value of the inner product with respect to the Killing form $-2(\alpha_i, \alpha_j)/(\alpha_j, \alpha_j)$ $(i \neq j)$ (which must be 0, 1, 2, or 3), we connect two vertices representing α_i and α_j as in Fig. 13. When the value is 0, we do not connect α_i and α_j. In Fig. 13, the left circle corresponds to α_i and the right circle to α_j.

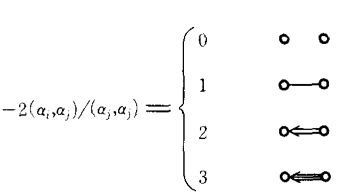


Fig. 13

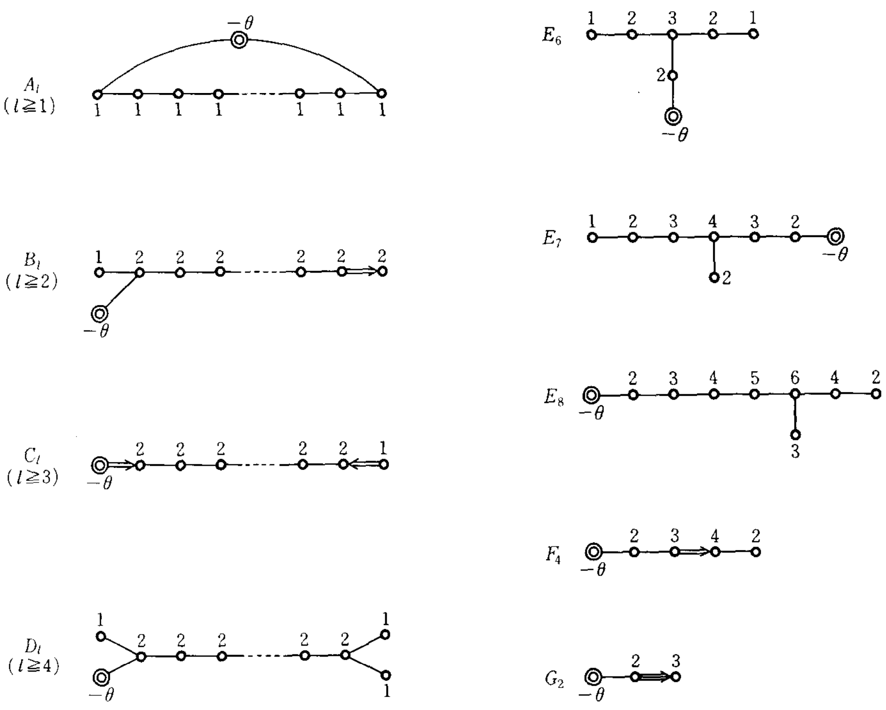


Fig. 14 We have relations $B_1 = C_1 = A_1$, $C_2 = B_2$, and $D_3 = A_3$. ($D_2 = A_1 + A_1$, which is not simple.) In this figure, the number at each vertex means the coefficient m_i in $\theta = \sum m_i \alpha_i$.

From Fig. 14, we have the following information.
(i) The quotient group of the automorphism group $A(\mathfrak{g})$ of $\mathfrak{g}$ with respect to the inner automorphism group $I(\mathfrak{g})$ is isomorphic to the automorphism group of the corresponding Dynkin diagram. The order of the latter group is 2 for A_l $(l \geqslant 2)$ since the diagram is symmetric. It is also 2 for D_l $(l \geqslant 5)$ and for E_6, and it is 6 $(=3!)$ for D_4. For all other cases, the order is 1.

(ii) The order of the center of the simply connected Lie group associated with $\mathfrak{g}$ is equal to the index of the subgroup consisting of elements stabilizing $-\theta$ in the group of automorphisms of the extended Dynkin diagram of $\mathfrak{g}$ (S. Murakami). This index is equal to the order of the fundamental group of the adjoint group of $\mathfrak{g}$ and the number of connected Lie groups, whose Lie algebra is $\mathfrak{g}$.
(iii) Any parabolic Lie subalgebra of $\mathfrak{g}$ is isomorphic to a subalgebra generated by the root vector X_α (and elements of the Cartan subalgebra) such that $\alpha = \sum n_i \alpha_i$, where $\{\alpha_1, \ldots, \alpha_l\}$ is a system of fundamental roots, $n_i \geqslant 0$ $(i = 1, \ldots, l)$ or $n_i \leqslant 0$ $(i = 1, \ldots, l)$, and $n_j = 0$ for α_j belonging to a fixed subset S of $\{\alpha_1, \ldots, \alpha_l\}$.
Hence, isomorphism classes of parabolic Lie subalgebras are in one-to-one correspondence with the set of subsets S of $\{\alpha_1, \ldots, \alpha_l\}$.
(iv) Maximal Lie subalgebra $\mathfrak{k}$ of $\mathfrak{g}$ with the same rank l as $\mathfrak{g}$. The Lie subalgebra $\mathfrak{k}$ is classified by the following rule. First we remove a vertex α_i from the Dynkin diagram. If the number m_i attached to the vertex is 1, $\mathfrak{k}$ is given by the product of the simple Lie algebra corresponding to the Dynkin diagram after removing the vertex α_i and a one-dimensional Lie subalgebra. If $m_i > 1$, $\mathfrak{k}$ is given by the diagram after removing α_i from the extended Dynkin diagram.

(2) Lie Groups. The classical complex simple Lie groups of rank n represented by A, B, C, D (in Cartan's symbolism) are the complex special linear group $SL(n+1, \mathbf{C})$, the complex special orthogonal group $SO(2n+1, \mathbf{C})$, the complex symplectic group $Sp(n, \mathbf{C})$, and the complex special orthogonal group $SO(2n, \mathbf{C})$, respectively. The classical compact simple Lie groups of rank n represented by A, B, C, D are the special unitary group $SU(n+1)$, the special orthogonal group $SO(2n+1)$, the unitary-symplectic group $Sp(n)$, and the special orthogonal group $SO(2n)$, respectively (→ 60 Classical Groups).

Cartan's Symbol	Complex Form	Compact Form	Dimension	Rank
A_n	$SL(n+1,\mathbf{C})$	$SU(n+1)$	$(n+1)^2-1$	n
B_n	$SO(2n+1,\mathbf{C})$	$SO(2n+1)$	$2n^2+n$	n
C_n	$Sp(n,\mathbf{C})$	$Sp(n)$	$2n^2+n$	n
D_n	$SO(2n,\mathbf{C})$	$SO(2n)$	$2n^2-n$	n
G_2	$\mathrm{Aut}\,\mathfrak{C}^c$	$\mathrm{Aut}\,\mathfrak{C}$	14	2
F_4	$\mathrm{Aut}\,\mathfrak{J}^c$	$\mathrm{Aut}\,\mathfrak{J}$	52	4
E_6			78	6
E_7			133	7
E_8			248	8

Here $\mathfrak{C}$ is the Cayley algebra over $\mathbf{R}$, $\mathfrak{C}^c$ is the complexification of $\mathfrak{C}$, $\mathfrak{J}$ is the Jordan algebra of Hermitian matrices of order 3 over $\mathfrak{C}$, $\mathfrak{J}^c$ is the complexification of $\mathfrak{J}$, and $\mathrm{Aut}\,A$ is the automorphism group of A.

(II) Classification of Noncompact Real Simple Lie Algebras

Classical Cases

Cartan's Symbol	Noncompact Real Simple Lie Algebra $\mathfrak{g}$	Maximal Compact Lie Algebra of $\mathfrak{g}$
AI	$\mathfrak{sl}(p+1;\mathbf{R})$	$\mathfrak{su}(p+1)$
AII	$\mathfrak{sl}(n;\mathbf{H})$	$\mathfrak{sp}(n)$
AIII	$\mathfrak{su}(p,q;\mathbf{C})$	$\mathfrak{su}(p)+\mathfrak{su}(q)$
BI	$\mathfrak{so}(p,q;\mathbf{R})$	$\mathfrak{so}(p)+\mathfrak{so}(q)$ $(p+q=2m+1)$
BII	$\mathfrak{so}(1,n-1;\mathbf{R})$	$\mathfrak{so}(n-1)$ $(n=2m+1)$
CI	$\mathfrak{sp}(p;\mathbf{R})$	$\mathfrak{u}(p)$
CII	$\mathfrak{u}(p,q;\mathbf{H})$	$\mathfrak{sp}(p)+\mathfrak{sp}(q)$
DI	$\mathfrak{so}(p,q;\mathbf{R})$	$\mathfrak{so}(p)+\mathfrak{so}(q)$ $(p+q=2m)$
DII	$\mathfrak{so}(1,n-1;\mathbf{R})$	$\mathfrak{so}(n-1)$ $(n=2m)$
DIII	$\mathfrak{so}(p;\mathbf{H})$	$\mathfrak{u}(2p)$

Here the field F is the real field $\mathbf{R}$, the complex field $\mathbf{C}$, or the quaternion field $\mathbf{H}$ ($\mathbf{R}\subset\mathbf{C}\subset\mathbf{H}$). $\mathbf{H}$ is an algebra over $\mathbf{R}$. For a quaternion $x=x_0+x_1i+x_2j+x_3k\,(x_0,x_1,x_2,x_3\in\mathbf{R})$, we put

$$\bar{x}=x_0-x_1i-x_2j-x_3k,$$
$$x^*=x_0+x_1i-x_2j+x_3k.$$

Then $\mathfrak{gl}(n;F)=\{$set of all square matrices over F of order $n\}$,

$$\mathfrak{sl}(n;F)=\{A\in\mathfrak{gl}(n;F)\mid \operatorname{tr}A=0\},$$
$$\mathfrak{so}(p,q;F)=\left\{A\in\mathfrak{gl}(p+q;F)\mid {}^tA^*I_{p,q}+I_{p,q}A=0\right\},$$

where $I_{p,q}$ is the symmetric transformation of the Euclidean space $\mathbf{R}^{p+q}$ with respect to $\mathbf{R}^p$, i.e.,

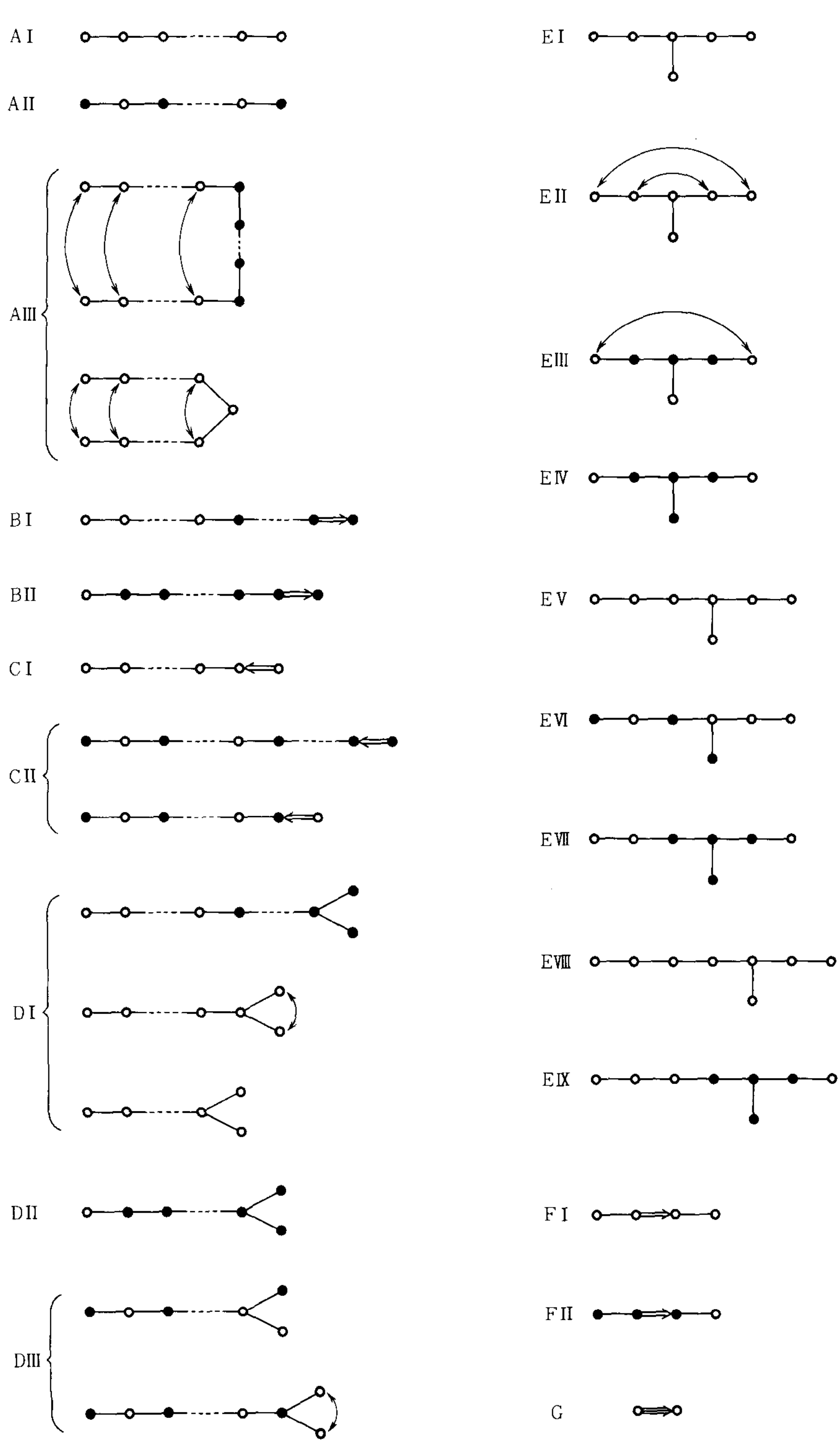


Fig. 15

$I_{p,q}$ is the diagonal sum of the unit matrix I_p of order p and $-I_q$. We have

$$\mathfrak{so}(n;F)=\mathfrak{so}(n,0;F),$$

$$\mathfrak{u}(p,q;F)=\{A\in\mathfrak{gl}(p+q;F)\mid {}^t\bar{A}I_{p,q}+I_{p,q}A=0\},$$

$$\mathfrak{u}(n;F)=\mathfrak{u}(n,0;F),$$

$$\mathfrak{sp}(n;F)=\{A\in\mathfrak{gl}(2n;F)\mid {}^t\bar{A}J+JA=0\},$$

where J is the matrix of an alternating form $\Sigma_{i=1}^{n}(x_i y_{i+n}-x_{i+n}y_i)$ of order $2n$.

A noncompact real simple Lie algebra $\mathfrak{g}$ is classified by the relation of the complex conjugation operator σ with respect to the complexification $\mathfrak{g}^c$ of $\mathfrak{g}$. The results are given by Satake's diagram (Fig. 15).

In the diagram, the fundamental root corresponding to a black circle is multiplied by -1 under σ for a suitable choice of Cartan subalgebra, and the arc with an arrow means that two elements corresponding to both ends of the arc are mutually transformed by a special transformation p such that $\sigma=pw$ ($w\in W$).

(III) Classification of Irreducible Symmetric Riemannian Spaces (→ 412 Symmetric Riemannian Spaces and Real Forms)

A simply connected irreducible symmetric Riemannian space $M=G/K$ is either a space in the following table or a simply connected compact simple Lie group mentioned in (I). The noncompact forms uniquely corresponding to the compact symmetric Riemannian space are in one-to-one correspondence with the noncompact real simple Lie algebras mentioned in (II).

Cartan's Symbol	$G/K=M$	Dimension	Rank
AI	$SU(n)/SO(n)$ $(n>2)$	$(n-1)(n+2)/2$	$n-1$
AII	$SU(2n)/Sp(n)$ $(n>1)$	$(n-1)(2n+1)$	$n-1$
AIII	$U(p+q)/U(p)\times U(q)$ $(p\geqslant q\geqslant 1)$	$2pq$	q
BDI	$SO(p+q)/SO(p)\times SO(q)$ $(p\geqslant q\geqslant 2, p+q\neq 4)$	pq	q
BDII	$SO(n+1)/SO(n)$ $(n\geqslant 2)$	n	1
DIII	$SO(2l)/U(l)$ $(l\geqslant 4)$	$l(l-1)$	$[l/2]$
CI	$Sp(n)/U(n)$ $(n\geqslant 3)$	$n(n+1)$	n
CII	$Sp(p+q)/Sp(p)\times Sp(q)$ $(p\geqslant q\geqslant 1)$	$4pq$	q
EI	$E_6/Sp(4)$	42	6
EII	$E_6/SU(2)\cdot SU(6)$	40	4
EIII	$E_6/Spin(10)\cdot SO(2)$	32	2
EIV	E_6/F_4	26	2
EV	$E_7/SU(8)$	70	7
EVI	$E_7/Spin(12)\cdot SU(2)$	64	4
EVII	$E_7/E_6\cdot SO(2)$	54	3
EVIII	$E_8/Spin(16)$	128	8
EIX	$E_8/E_7\cdot SU(2)$	112	4
FI	$F_4/Sp(3)\cdot SU(2)$	28	4
FII	$F_4/Spin(9)$	16	1
G	$G_2/SO(4)$	8	2

Notes

The group $G=U(p+q)$ in AIII is not effective, unless it is replaced by $SU(p+q)$. To be precise, $K=Sp(4)$ in EI should be replaced by its quotient group factored by a subgroup of order 2 of its center. K in EII is not a direct product of simple groups; the order of its fundamental group $\pi_1(K)$ is 2. To be precise, K in EV or EVIII should be replaced by its quotient group factored by a subgroup of order 2 of its center. The K's in EII, EIII, EVI, EVII, EIX, and FI are not direct products. The fundamental group $\pi_1(K)$ of K is the infinite cyclic group $\mathbf{Z}$ for EIII, EVII; for all other cases, the order of $\pi_1(K)$ is 2.

In EIII, EVII, the groups E_6, E_7 are adjoint groups of compact simple Lie algebras. In other cases, E_6 and E_7 (E_8, F_4 and G_2 also) are simply connected Lie groups.

The compact symmetric Riemannian space M is a complex Grassmann manifold for AIII, a real Grassmann manifold for BDI, a sphere for BDII, a quaternion Grassmann manifold for CII, and a Cayley projective plane for FII.

(IV) Isomorphic Relations among Classical Lie Algebras

The isomorphic relations among the classical Lie algebras over **R** or **C** are all given in the following table. In the table, we denote, for example, the real form of type AI of the complex Lie algebra with rank 3 by A_3I in Cartan's symbolism. When there are nonisomorphic real forms of the same type and same rank (e.g., in the case of D_3I) we distinguish them by the rank of the corresponding symmetric Riemannian space and denote them by, e.g., D_3I_p, where p is the index of total isotropy of the sesquilinear form which is invariant under the corresponding Lie algebra.

Cartan's Symbol	Isomorphisms among Classical Lie Algebras
$A_1=B_1=C_1$	$\mathfrak{sl}(2,\mathbf{C})\cong\mathfrak{so}(3,\mathbf{C})\cong\mathfrak{sp}(1,\mathbf{C})$; $\mathfrak{su}(2)\cong\mathfrak{so}(3)\cong\mathfrak{sp}(1)$
$B_2=C_2$	$\mathfrak{so}(5,\mathbf{C})\cong\mathfrak{sp}(2,\mathbf{C})$; $\mathfrak{so}(5)\cong\mathfrak{sp}(2)$
$A_3=D_3$	$\mathfrak{sl}(4,\mathbf{C})\cong\mathfrak{so}(6,\mathbf{C})$; $\mathfrak{su}(4)\cong\mathfrak{so}(6)$
$A_1I=A_1III=B_1I=C_1I$	$\mathfrak{sl}(2,\mathbf{R})\cong\mathfrak{su}(1,1;\mathbf{C})\cong\mathfrak{so}(2,1;\mathbf{R})\cong\mathfrak{sp}(1;\mathbf{R})$
$B_2I_2=C_2I$	$\mathfrak{so}(3,2;\mathbf{R})\cong\mathfrak{sp}(2,\mathbf{R})$
$B_2I_1=C_2II$	$\mathfrak{so}(4,1;\mathbf{R})\cong\mathfrak{u}(1,1;\mathbf{H})$
$A_3I=D_3I_3$	$\mathfrak{sl}(4,\mathbf{R})\cong\mathfrak{so}(3,3;\mathbf{R})$
$A_3II=D_3I_1$	$\mathfrak{sl}(2,\mathbf{H})\cong\mathfrak{so}(5,1;\mathbf{R})$
$A_3III_2=D_3I_2$	$\mathfrak{su}(2,2;\mathbf{C})\cong\mathfrak{so}(4,2;\mathbf{R})$
$A_3III_1=D_3III$	$\mathfrak{su}(3,1;\mathbf{C})\cong\mathfrak{so}(3;\mathbf{H})$
$D_4I_2=D_4III$	$\mathfrak{so}(6,2;\mathbf{R})\cong\mathfrak{so}(4,\mathbf{H})$
$D_2=A_1\times A_1$*	$\mathfrak{so}(4,\mathbf{C})\cong\mathfrak{sl}(2,\mathbf{C})\times\mathfrak{sl}(2,\mathbf{C})$; $\mathfrak{so}(4)\cong\mathfrak{su}(2)\times\mathfrak{su}(2)$
$D_2I_2=A_1I\times A_1I$	$\mathfrak{so}(2,2;\mathbf{R})\cong\mathfrak{sl}(2,\mathbf{R})\times\mathfrak{sl}(2;\mathbf{R})$
$D_2III=A_1\times A_1I$*	$\mathfrak{so}(2;\mathbf{H})\cong\mathfrak{su}(2)\times\mathfrak{sl}(2;\mathbf{R})$
$D_2I_1=A_1$*	$\mathfrak{so}(3,1;\mathbf{R})\cong\mathfrak{sl}(2,\mathbf{C})$

Note

(*) In these 3 cases, there are isomorphisms given by the replacement of $\mathfrak{sl}(2,\mathbf{C})$ or $\mathfrak{su}(2)$ by isomorphic Lie algebras of type B_1 or type C_1 due to the isomorphism $A_1\cong B_1\cong C_1$.

(V) Lists of Normal Forms of Singularities with Modulus Number $m=0$, 1, and 2 (→ 418 Theory of Singularities)

Letters $A,\dots,Z$ stand here for stable equivalence classes of function germs (or families of function germs).

(1) Simple Singularities ($m=0$). There are 2 infinite series A, D, and 3 "exceptional" singularities E_6, E_7, E_8:

Notation	Normal form	Restrictions
A_n	$x^{n+1}+y^2+z^2$	$n\geqslant 1$
D_n	$x^{n-1}+xy^2+z^2$	$n\geqslant 4$
E_6	$x^4+y^3+z^2$	
E_7	$x^3y+y^3+z^2$	
E_8	$x^5+y^3+z^2$	

(2) Unimodular Singularities ($m=1$). There are 3 families of parabolic singularities, one series of hyperbolic singularities (with 3 subscripts), and 14 families of exceptional singularities.

The parabolic singularities

Notation	Normal form	Restrictions
$P_8=\tilde{E}_6$	$x^3+y^3+z^3+axyz$	$a^3+27\neq 0$
$X_9=\tilde{E}_7$	$x^4+y^4+z^2+axyz$	$a^4-64\neq 0$
$J_{10}=\tilde{E}_8$	$x^6+y^3+z^2+axyz$	$a^6-432\neq 0$

The hyperbolic singularities

Notation	Normal form	Restrictions
T_{pqr}	$x^p+y^q+z^r+axyz$	$a\neq 0, \frac{1}{p}+\frac{1}{q}+\frac{1}{r}<1$

The 14 exceptional families

Notation	Normal form	Gabrielov numbers	Dolgachëv numbers	Notation	Normal form	Gabrielov numbers	Dolgachëv numbers
K_{12}	$x^3+y^7+z^2+axy^5$	2 3 7	2 3 7	W_{13}	$x^4+xy^4+z^2+ay^6$	2 5 6	3 4 4
K_{13}	$x^3+xy^5+z^2+ay^8$	2 3 8	2 4 5	Q_{10}	$x^3+y^4+yz^2+axy^3$	3 3 4	2 3 9
K_{14}	$x^3+y^8+z^2+axy^6$	2 3 9	3 3 4	Q_{11}	$x^3+y^2z+xz^3+az^5$	3 3 5	2 4 7
Z_{11}	$x^3y+y^5+z^2+axy^4$	2 4 5	2 3 8	Q_{12}	$x^3+y^5+yz^2+axy^4$	3 3 6	3 3 6
Z_{12}	$x^3y+xy^4+z^2+ax^2y^3$	2 4 6	2 4 6	S_{11}	$x^4+y^2z+xz^2+ax^3z$	3 4 4	2 5 6
Z_{13}	$x^3y+y^6+z^2+axy^5$	2 4 7	3 3 5	S_{12}	$x^2y+y^2z+xz^3+az^5$	3 4 5	3 4 5
W_{12}	$x^4+y^5+z^2+ax^2y^3$	2 5 5	2 5 5	U_{12}	$x^3+y^3+z^4+axyz^2$	4 4 4	4 4 4

(3) Bimodular Singularities ($m=2$). There are 8 infinite series and 14 exceptional families. In all the formulas, $\mathbf{a}=a_0+a_1y$.

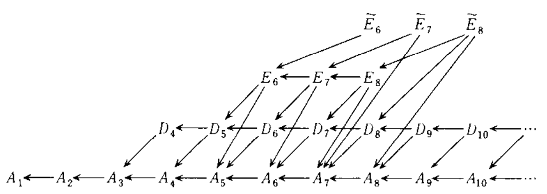
The 8 infinite series of bimodular singularities

Notation	Normal form	Restriction	Milnor number
$J_{3,0}$	$x^3+bx^2y^3+y^9+z^2+cxy^7$	$4b^3+27\neq 0$	16
$J_{3,p}$	$x^3+x^2y^3+z^2+ay^{9+p}$	$p>0, a_0\neq 0$	$16+p$
$Z_{1,0}$	$y(x^3+dx^2y^2+cxy^5+y^6)+z^2$	$4d^3+27\neq 0$	15
$Z_{1,p}$	$y(x^3+x^2y^2+ay^{6+p})+z^2$	$p>0, a_0\neq 0$	$15+p$
$W_{1,0}$	$x^4+ax^2y^3+y^6+z^2$	$a_0^2\neq 4$	15
$W_{1,p}$	$x^4+x^2y^3+ay^{6+p}+z^2$	$p>0, a_0\neq 0$	$15+p$
$W^{\#}_{1,2q-1}$	$(x^2+y^3)^2+axy^{4+q}+z^2$	$q>0, a_0\neq 0$	$15+2q-1$
$W^{\#}_{1,2q}$	$(x^2+y^3)^2+ax^2y^{3+q}+z^2$	$q>0, a_0\neq 0$	$15+2q$
$Q_{2,0}$	$x^3+yz^2+ax^2y^2+xy^4$	$a_0^2\neq 4$	14
$Q_{2,p}$	$x^3+yz^2+x^2y^2+az^{6+p}$	$p>0, a_0\neq 0$	$14+p$
$S_{1,0}$	$x^2z+yz^2+y^5+azy^3$	$a_0^2\neq 4$	14
$S_{1,p}$	$x^2z+yz^2+x^2y^2+ay^{5+p}$	$p>0, a_0\neq 0$	$14+p$
$S^{\#}_{1,2q-1}$	$x^2z+yz^2+zy^3+axy^{2+q}$	$q>0, a_0\neq 0$	$14+2q-1$
$S^{\#}_{1,2q}$	$x^2z+yz^2+zy^3+ax^2y^{2+q}$	$q>0, a_0\neq 0$	$14+2q$
$U_{1,0}$	$x^3+xz^2+xy^3+ay^3z$	$a_0(a_0^2+1)\neq 0$	14
$U_{1,2q-1}$	$x^3+xz^2+xy^3+ay^{1+q}z^2$	$q>0, a_0\neq 0$	$14+2q-1$
$U_{1,2q}$	$x^3+xz^2+xy^3+ay^{3+q}z$	$q>0, a_0\neq 0$	$14+2q$

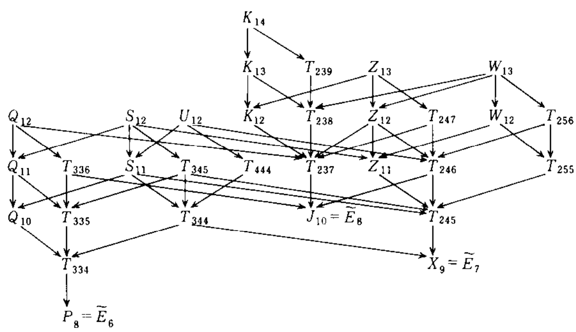
The 14 exceptional families

Notation	Normal form	Notation	Normal form
E_{18}	$x^3+y^{10}+z^2+axy^7$	W_{18}	$x^4+y^7+z^2+ax^2y^4$
E_{19}	$x^3+xy^7+z^2+ay^{11}$	Q_{16}	$x^3+yz^2+y^7+z^2+axy^5$
E_{20}	$x^3+y^{11}+z^2+axy^8$	Q_{17}	$x^3+yz^2+xy^5+z^2+ay^8$
Z_{17}	$x^3y+y^8+z^2+axy^6$	Q_{18}	$x^3+yz^2+y^8+z^2+axy^6$
Z_{18}	$x^3y+xy^6+z^2+ay^9$	S_{16}	$x^2z+yz^2+xy^4+z^2+ay^6$
Z_{19}	$x^3y+y^9+z^2+axy^7$	S_{17}	$x^2z+yz^2+y^6+z^2+azy^4$
W_{17}	$x^4+xy^5+z^2+ay^7$	U_{16}	$x^3+xz^2+y^5+z^2+ax^2y^2$

Adjacency relations between simple and simply elliptic singularities



Adjacency relations among unimodular singularities



References

[1] V. I. Arnol'd, Singularity theory, Lecture note ser. 53, London Math. Soc., 1981.
[2] E. Breiskorn, Die Hierarchie der 1-modularen Singularitäten, Manuscripta Math., 27 (1979), 183–219.
[3] K. Saito, Einfach elliptische Singularitäten, Inventiones Math., 23 (1974), 289–325.

6. Topology

(I) h-Cobordism Groups of Homotopy Spheres and Groups of Differentiable Structures on Combinatorial Spheres

(1) The Structure of the h-Cobordism Group θ_n of n-Dimensional Homotopy Spheres. In the following table, values of θ_n have the following meanings: 0 means that the group consists only of the identity element, an integer l means that the group is isomorphic to the cyclic group of order l, 2^l means that the group is the direct sum of l groups of order 2, + means the direct sum, and ? means that the structure of the group is unknown.

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
$\theta_n \cong$	0	0	?	0	0	0	28	2	2^3 or $4+2$	6	992	0	3	2	$8128+2$	2	2^4 or $4+2^2$	$8+2$

(2) The Group Γ_n of Differentiable Structures on the n-Dimensional Combinatorial Sphere.

$$\Gamma_n \cong \theta_n \quad (n \neq 3), \quad \Gamma_3 = 0.$$

(II) Adem's Formula Concerning Steenrod Operators Sq and $\mathscr{P}$ (→ 64 Cohomology Operations)

For the cohomology operators Sq and $\mathscr{P}$, we have

$$Sq^a Sq^b = \sum_{c=0}^{[a/2]} \binom{b-c-1}{a-2c} Sq^{a+b-c} Sq^c \quad (a < 2b).$$

$$\mathscr{P}^a\mathscr{P}^b=\sum_{c=0}^{[a/p]}(-1)^{a+c}\binom{(b-c)(p-1)-1}{a-pc}\mathscr{P}^{a+b-c}\mathscr{P}^c \quad (a<pb),$$

$$\mathscr{P}^a\delta\mathscr{P}^b=\sum_{c=0}^{[a/p]}(-1)^{a+c}\binom{(b-c)(p-1)}{a-pc}\delta\mathscr{P}^{a+b-c}\mathscr{P}^c$$

$$+\sum_{c=0}^{[(a-1)/p]}(-1)^{a+c-1}\binom{(b-c)(p-1)-1}{a-pc-1}\mathscr{P}^{a+b-c}\delta\mathscr{P}^c \quad (a\leqslant pb).$$

Several simple cases of the formula above are as follows.

$Sq^1Sq^{2n}=Sq^{2n+1}$, $\quad Sq^1Sq^{2n+1}=0$,

$Sq^2Sq^{4n}=Sq^{4n+2}+Sq^{4n+1}Sq^1$, $\quad Sq^2Sq^{4n+1}=Sq^{4n+2}Sq^1$,

$Sq^2Sq^{4n+2}=Sq^{4n+3}Sq^1$, $\quad Sq^2Sq^{4n+3}=Sq^{4n+5}+Sq^{4n+4}Sq^1$,

$Sq^4Sq^{8n}=Sq^{8n+4}+Sq^{8n+3}Sq^1+Sq^{8n+2}Sq^2$, $\quad Sq^4Sq^{8n+1}=Sq^{8n+4}Sq^1+Sq^{8n+3}Sq^2$,

$Sq^4Sq^{8n+2}=Sq^{8n+4}Sq^2$, $\quad Sq^4Sq^{8n+3}=Sq^{8n+5}Sq^2$,

$Sq^4Sq^{8n+4}=Sq^{8n+7}Sq^1+Sq^{8n+6}Sq^2$, $\quad Sq^4Sq^{8n+5}=Sq^{8n+9}+Sq^{8n+8}Sq^1+Sq^{8n+7}Sq^2$,

$Sq^4Sq^{8n+6}=Sq^{8n+10}+Sq^{8n+8}Sq^2$, $\quad Sq^4Sq^{8n+7}=Sq^{8n+11}+Sq^{8n+9}Sq^2$,

$\mathscr{P}^1\mathscr{P}^n=(n+1)\mathscr{P}^{n+1}$,

$\mathscr{P}^1\delta\mathscr{P}^n=n\cdot\delta\mathscr{P}^{n+1}+\mathscr{P}^{n+1}\delta$.

(III) Cohomology Ring $H^*(\pi,n;\mathbf{Z}_p)$ of Eilenberg-MacLane Complex (→ 70 Complexes)

$\mathbf{Z}$ means the set of integers, and $\mathbf{Z}_p=\mathbf{Z}/p\mathbf{Z}$, where p is a prime number.

(1) The case $p=2$, $\pi=\mathbf{Z}$ or $\mathbf{Z}_{2^f}$ $(f\geqslant 1)$. The degree of a finite sequence $I=(i_r,i_{r-1},\ldots,i_1)$ of positive integers is defined by $d(I)=i_1+i_2+\ldots+i_r$. If such a sequence satisfies $i_{k+1}\geqslant 2i_k$ $(k=1,\ldots,r-1)$, it is called admissible, and we define its excess by

$$e(I)=(i_r-2i_{r-1})+\ldots+(i_2-2r_1)+i_1=2i_r-d(I).$$

Further, we put $Sq^I=Sq^{i_r}Sq^{i_{r-1}}\ldots Sq^{i_1}$. Then we have $H^*(\mathbf{Z}_{2^f},n;\mathbf{Z}_2)=\mathbf{Z}_2[Sq^Iu_n\,|\,I$ is admissible, $e(I)<n]$, $H^*(\mathbf{Z},n;\mathbf{Z}_2)=\mathbf{Z}_2[Sq^Iu_n\,|\,I$ is admissible, $e(I)<n$, $i_1>1]$.

Here, $u_n\in H^n(\pi,n;\mathbf{Z}_2)$ is the fundamental cohomology class. $I=\varnothing$ (empty) is also admissible, and for this case we put $n(I)=e(I)=0$, $Sq^I=1$. Due to the Künneth theorem, we have

$$H^*(\pi+\pi',n;\mathbf{Z}_p)=H^*(\pi,n;\mathbf{Z}_p)\otimes H^*(\pi',n;\mathbf{Z}_p)$$

if π is finitely generated. In particular, we have

$$H^*(\mathbf{Z}_2,1;\mathbf{Z}_2)=\mathbf{Z}_2[u_1],$$

$$H^*(\mathbf{Z}_2,2;\mathbf{Z}_2)=\mathbf{Z}_2\left[u_2,Sq^1u_2,Sq^2Sq^1u_2,\ldots,Sq^{2^r}Sq^{2^r-1}\ldots Sq^1u_2,\ldots\right],$$

$$H^*(\mathbf{Z}_2,3;\mathbf{Z}_2)=\mathbf{Z}_2\left[u_3,Sq^{2^r}Sq^{2^r-1}\ldots Sq^1u_3,Sq^{(2^r+1)2^s}Sq^{(2^r+1)2^s-1}\ldots Sq^{2^r+1}Sq^{2^r-1}\ldots Sq^1u_3\,\middle|\,r\geqslant 0,\ \ s\geqslant 0\right].$$

(2) The case $p\neq 2$, $\pi=\mathbf{Z}$ or $\mathbf{Z}_{p^f}$ $(f\geqslant 1)$. We define the degree of a finite sequence $I=(i_r,i_{r-1},\ldots,i_1,i_0)$ of nonnegative integers by $d(I)=i_r+\ldots+i_1+i_0$. The sequence I is called admissible if it satisfies the following conditions:

$$i_k=2\lambda_k(p-1)+\varepsilon_k \quad (\lambda_k \text{ is a nonnegative integer}, \varepsilon_k=0 \text{ or } 1\ (0\leqslant k\leqslant r)), \text{ and}$$

$$i_0=0 \text{ or } 1,\quad i_1\geqslant 2p-2,\quad i_{k+1}\geqslant pi_k \qquad (1\leqslant k\leqslant r-1).$$

We define its excess by $e(I)=pi_r-(p-1)d(I)$. Further, we put $\mathscr{P}^I=\delta^{\varepsilon_k}\mathscr{P}^{\lambda_k}\ldots\delta^{\varepsilon_1}\mathscr{P}^{\lambda_1}\delta^{\varepsilon_0}$, and assume that $u_n\in H^n(\pi,n;\mathbf{Z})$ is the fundamental cohomology class. Then we have

$$H^*(\mathbf{Z}_{p^f},n;\mathbf{Z}_p)=\mathbf{Z}_p[\mathscr{P}^Iu_n\,|\,I \text{ is admissible}, e(I)<n(p-1), n+d(I) \text{ is even}]$$

$$\otimes\wedge_{\mathbf{Z}_p}(\mathscr{P}^Iu_n\,|\,I \text{ is admissible}, e(I)<n(p-1), n+d(I) \text{ is odd}).$$

$H^*(\mathbf{Z},n;\mathbf{Z})$ is given by the above formula when the admissible sequence is I with $i_0=0$.

(IV) Cohomology Ring of Compact Connected Lie Groups (→ 427 Topology of Lie Groups and Homogeneous Spaces)

(1) General Remarks. Let G be a compact connected Lie group with rank l and dimension n. We have $H^*(G;\mathbf{R})\cong\bigwedge_{\mathbf{R}}(x_1,\ldots,x_l)$, where $\bigwedge_K(x_1,\ldots,x_l)$ means the exterior algebra over K of a linear space $V=Kx_1+\ldots+Kx_l$ with the basis $\{x_1,\ldots,x_l\}$ over K. We define a new degree in $\bigwedge_K(x_1,\ldots,x_l)$ by putting $\deg x_i=m_i$ (m_i is odd) $(1\leqslant i\leqslant l)$, where $m_1+\ldots+m_l=n$. The $\cong$ means isomorphism as graded rings.

(2) Classical Compact Simple Lie Groups. We set $\deg x_i=i$.

$$H^*(U(n);\mathbf{R})\cong\bigwedge_{\mathbf{R}}(x_1,x_3,\ldots,x_{2n-1}),$$

$$H^*(SU(n);\mathbf{R})\cong\bigwedge_{\mathbf{R}}(x_3,x_5,\ldots,x_{2n-1}),$$

$$H^*(Sp(n);\mathbf{R})\cong\bigwedge_{\mathbf{R}}(x_3,x_7,\ldots,x_{4n-1}),$$

$H^*(SO(n);\mathbf{Z}_2)\cong$(Having $x_1,x_2,\ldots,x_{n-1}$ as a simple system of generators)

$$\cong\mathbf{Z}_2[x_1,x_3,\ldots,x_{2n'-1}]/(x_i^{2^{s(i)}}\mid i=1,\ldots,n')$$

($n'=[n/2]$, $s(i)$ is the least integer satisfying $2^{s(i)}(2i-1)\geqslant n$)

$$H^*(SO(2n);K)\cong\bigwedge_K(x_3,x_7,\ldots,x_{4n-5},x_{2n-1}),$$

$H^*(SO(2n-1);K)\cong\bigwedge_K(x_3,x_7,\ldots,x_{4n-5})$, where K is a commutative field whose characteristic is not 2.

For $SO(n)$, $Sq^a(x_i)=\binom{i}{a}x_{i+a}$. For $SU(n)$, $p^a(x_{2i-1})=\binom{i-1}{a}x_{2i-1+2a(p-1)}$.

For $Sp(n)$, $\mathfrak{P}^a(x_{4i-1})=(-1)^{a(p-1)/2}\binom{2i-1}{a}x_{4i-1+2a(p-1)}$.

(3) Exceptional Compact Simple Lie Groups. n and m_i $(1\leqslant i\leqslant l)$ given in (1) are as follows.

G_2:	$n=14$,	$m_i=3$,	11.							
F_4:	$n=52$,	$m_i=3$,	11,	15,	23.					
E_6:	$n=78$,	$m_i=3$,	9,	11,	15,	17,	23.			
E_7:	$n=133$,	$m_i=3$,	11,	15,	19,	23,	27,	35.		
E_8:	$n=248$,	$m_i=3$,	15,	23,	27,	35,	39,	47,	59.	

(4) p-Torsion Groups of Exceptional Groups. The p-torsion groups of exceptional Lie groups are unit groups except when $p=2$ for G_2; $p=2$, 3 for F_4, E_6, E_7; and $p=2$, 3, 5 for E_8. The cohomology ring of $\mathbf{Z}_p$ as a coefficient group in these exceptional cases is given as follows. Here we put $\deg x_i=i$.

$$H^*(G_2;\mathbf{Z}_2)=\mathbf{Z}_2[x_3]/(x_3^4)\otimes\bigwedge_{\mathbf{Z}_2}(Sq^2x_3);$$

$$H^*(F_4;\mathbf{Z}_2)=\mathbf{Z}_2[x_3]/(x_3^4)\otimes\bigwedge_{\mathbf{Z}_2}(Sq^2x_3,x_{15},Sq^8x_{15}),$$

$$H^*(F_4;\mathbf{Z}_3)=\mathbf{Z}_3[\delta\mathfrak{P}^1x_3]/((\delta\mathfrak{P}^1x_3)^3)\otimes\bigwedge_{\mathbf{Z}_3}(x_3,\mathfrak{P}^1x_3,x_{11},\mathfrak{P}^1x_{11});$$

$$H^*(E_6;\mathbf{Z}_2)=\mathbf{Z}_2[x_3]/(x_3^4)\otimes\bigwedge_{\mathbf{Z}_2}(Sq^2x_3,Sq^4Sq^2x_3,x_{15},Sq^8Sq^4Sq^2x_3,Sq^8x_{15}),$$

$$H^*(E_6;\mathbf{Z}_3)=\mathbf{Z}_3[\delta\mathfrak{P}^1x_3]/((\delta\mathfrak{P}^{\prime}x_3)^3)\otimes\bigwedge_{\mathbf{Z}_3}(x_3,\mathfrak{P}^1x_3,x_9,x_{11},\mathfrak{P}^1x_{11},x_{17});$$

$$H^*(E_7;\mathbf{Z}_2)=\mathbf{Z}_2[x_3,Sq^2x_3,Sq^4Sq^2x_3]/(x_3^4,(Sq^2x_3)^4,(Sq^4Sq^2x_3)^4)$$
$$\otimes\bigwedge_{\mathbf{Z}_2}(x_{15},Sq^8Sq^4Sq^2x_3,Sq^8x_{15},Sq^4Sq^8x_{15}),$$

$$H^*(E_7;\mathbf{Z}_3)=\mathbf{Z}_3[\delta\mathfrak{P}^1x_3]/((\delta\mathfrak{P}^1x_3)^3)\otimes\bigwedge_{\mathbf{Z}_3}(x_3,\mathfrak{P}^1x_3,x_{11},\mathfrak{P}^1x_{11},\mathfrak{P}^3\mathfrak{P}^1x_3,x_{27},x_{35});$$

$$H^*(E_8;\mathbf{Z}_2)=\mathbf{Z}_2[x_3,Sq^2x_3,Sq^4Sq^2x_3,x_{15}]/(x_3^{16},(Sq^2x_3)^8,(Sq^4Sq^2x_3)^4,x_{15}^4)$$
$$\otimes\bigwedge_{\mathbf{Z}_2}(Sq^8Sq^4Sq^2x_3,Sq^8x_{15},Sq^4Sq^8x_{15},Sq^2Sq^4Sq^8x_{15}),$$

$$H^*(E_8;\mathbf{Z}_3)=\mathbf{Z}_3[\delta\mathfrak{P}^1x_3,\delta\mathfrak{P}^3\mathfrak{P}^1x_3]/((\delta\mathfrak{P}^1x_3)^3,(\delta\mathfrak{P}^3\mathfrak{P}^1x_3)^3)$$
$$\otimes\bigwedge_{\mathbf{Z}_3}(x_3,\mathfrak{P}^1x_3,x_{15},\mathfrak{P}^3\mathfrak{P}^1x_3,\mathfrak{P}^3x_{15},x_{35},x_{39},x_{47}),$$

$$H^*(E_8;\mathbf{Z}_5)=\mathbf{Z}_5[\delta\mathfrak{P}^1x_3]/((\delta\mathfrak{P}^1x_3)^5)\otimes\bigwedge_{\mathbf{Z}_5}(x_3,\mathfrak{P}^1x_3,x_{15},\mathfrak{P}^1x_{15},x_{27},x_{35},x_{39},x_{47}).$$

(V) Cohomology Rings of Classifying Spaces (→ 56 Characteristic Classes, 427 Topology of Lie Groups and Homogeneous Spaces C)

(1) Let $H^*(G;K)=\bigwedge_K(x_1,x_2,\ldots,x_n)$. Then the $\deg x_i$ are odd and the x_i may be assumed to be transgressive. y_i being its image, the following formula holds:

$$H^*(BG;K)=K[y_1,y_2,\ldots,y_n] \qquad \text{(Borel's theorem)}.$$

(2) $H^*(BU(n))=H^*(BGL(n,\mathbf{C}))=\mathbf{Z}[c_1,c_2,\ldots,c_n]$,

$H^*(BSU(n))=H^*(BSL(n,\mathbf{C}))=\mathbf{Z}[c_2,\ldots,c_n]$,

$H^*(BSp(n))=\mathbf{Z}[q_1,q_2,\ldots,q_n]$,

$H^*(BO(n);K_2)=H^*(BGL(n,\mathbf{R});\mathbf{Z}_2)=K_2[w_1,w_2,\ldots,w_n]$,

$H^*(BSO(n);K_2)=H^*(BSL(n,\mathbf{R});\mathbf{Z}_2)=K_2[w_2,\ldots,w_n]$,

$H^*(BSO(2m+1);K)=K[p_1,p_2,\ldots,p_m]$,

$H^*(BSO(2m);K)=K[p_1,p_2,\ldots,p_{m-1},\chi]$.

Here, K denotes a field of characteristic $\neq 2$, and K_2 is the field of characteristic 2. The c_i denote the ith Chern classes and the q_i the ith symplectic Pontryagin classes, the w_i the ith Stiefel-Whitney classes. Moreover, the p_i denote the ith Pontryagin classes, and χ the Euler class. Their degrees are given as follows: $\deg c_i=2i$, $\deg q_i=\deg p_i=4i$, $\deg w_i=i$, and $\deg\chi=2m$.

(3) Wu's Formula. Let $H^2(BSO(n);\mathbf{Z}_2)=\mathbf{Z}_2[w_2,\ldots,w_n]$ and $H^*(BU(n),\mathbf{Z}_2)=\mathbf{Z}_2[c_1,\ c_2,\ldots,c_n]$. We have

$$Sq^j w_i=\sum_{0\leqslant t\leqslant j}\binom{i-j+t-1}{t}w_{j-t}w_{i+t} \qquad (w_0=1),$$

$$Sq^{2j} c_i=\sum_{0\leqslant t\leqslant j}\binom{i-j+t-1}{t}c_{j-t}c_{i+t} \qquad (c_0=1).$$

Here the symbol $\binom{a}{b}$ denotes the binomial coefficient for $a\geqslant b$; $\binom{a}{0}=1$, and $\binom{a}{b}=0$ otherwise.

(VI) Homotopy Groups of Spheres (→ 202 Homotopy Theory)

Table of the $(n+k)$th Homotopy Group $\pi_{n+k}(S^n)$ of the n-Dimensional Sphere S^n. The table represents Abelian groups. 0 stands for the unit group; integer l the cyclic group of order l; ∞ the infinite cyclic group; 2^l the direct sum of l groups of order 2; and $+$ means the direct sum.

n \ k	<0	0	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0	∞	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	∞	∞	2	2	12	2	2	3	15	2	2^2	$12+2$	$84+2^2$	2^2
3	0	∞	2	2	12	2	2	3	15	2	2^2	$12+2$	$84+2^2$	2^2	6
4	0	∞	2	2	$\infty+12$	2^2	2^2	$24+3$	15	2	2^3	$120+12+2$	$84+2^5$	2^6	$24+6+2$
5	0	∞	2	2	24	2	2	2	30	2	2^3	$72+2$	$504+2^2$	2^3	$6+2$
6	0	∞	2	2	24	0	∞	2	60	$24+2$	2^3	$72+2$	$504+4$	240	6
7	0	∞	2	2	24	0	0	2	120	2^3	2^4	$24+2$	$504+2$	0	6
8	0	∞	2	2	24	0	0	2	$\infty+120$	2^4	2^5	$24+24+2$	$504+2$	0	$6+2$
9	0	∞	2	2	24	0	0	2	240	2^3	2^4	$24+2$	$504+2$	0	6
10	0	∞	2	2	24	0	0	2	240	2^2	$\infty+2^3$	$12+2$	504	12	6
11	0	∞	2	2	24	0	0	2	240	2^2	2^3	$6+2$	504	2	$6+2$
12	0	∞	2	2	24	0	0	2	240	2^2	2^3	6	$\infty+504$	2^2	$6+2$
13	0	∞	2	2	24	0	0	2	240	2^2	2^3	6	504	2	6
14	0	∞	2	2	24	0	0	2	240	2^2	2^3	6	504	0	$\infty+3$
$\geqslant 15$	0	∞	2	2	24	0	0	2	240	2^2	2^3	6	504	0	3

Table of the $(n+k)$th Homotopy Group $\pi_{n+k}(S^n)$ of the n-Dimensional Sphere S^n (Continued)

n \ k	14	15	16	17	18	19	20	21	22
1	0	0	0	0	0	0	0	0	0
2	6	30	30	$6+2$	$12+2^2$	$12+2^2$	$132+2$	2^2	2
3	30	30	$6+2$	$12+2^2$	$12+2^2$	$132+2$	2^2	2	210
4	$2520+6+2$	30	$6+6+2$	$24+12+4+2^2$	$120+12+2^5$	$132+2^5$	2^6	$24+2^2$	$9240+6+2$
5	$6+2$	$30+2$	2^2	$4+2^2$	$24+2^2$	$264+2$	$6+2^2$	$6+2$	$90+2^2$
6	$12+2$	$60+2$	$504+2^2$	2^4	$24+6+2$	$1056+8$	$480+12$	6	$180+2^2$
7	$24+4$	$120+2^3$	2^4	2^4	$24+2$	$264+2$	24	$6+2$	$72+2^3$
8	$240+24+4$	$120+2^5$	2^7	$6+2^4$	$504+24+2$	$264+2$	$24+3$	$12+2^3$	$1440+24+2^4$
9	$16+4$	$240+2^3$	2^4	2^4	$24+2$	$264+2$	24	$6+2^2$	$144+2^3$
10	$16+2$	$240+2^2$	$240+2$	2^3	$24+2^2$	$264+6$	$504+24$	$6+2^2$	$144+6+2$
11	$16+2$	$240+2$	2	2^3	$8+4+2$	$264+2^3$	$24+2^2$	2^4	$48+2^2$
12	$48+4+2$	$240+2$	2	2^4	$480+4+4+2$	$264+2^5$	$24+2^5$	$6+2^4$	$2016+12+2^2$
13	$16+2$	$480+2$	2	2^4	$8+8+2$	$264+2^3$	$24+2^3$	$4+2^3$	$16+2^2$
14	$8+2$	$480+2$	$24+2$	2^4	$8+8+2$	$264+4+2$	$240+24$	$4+2^2$	$16+2^2$
15	$4+2$	$480+2$	2^3	2^5	$8+8+2$	$264+2^2$	2^4	2^3	$16+2^3$
16	$2+2$	$\infty+480+2$	2^4	2^6	$24+8+8+2$	$264+2^2$	2^4	2^4	$240+16+2^3$
17	$2+2$	$480+2$	2^3	2^5	$8+8+2$	$264+2^2$	24	2^3	$16+2^3$
18	$2+2$	$480+2$	2^2	$\infty+2^4$	$8+4+2$	$264+2$	$24+12$	2^3	$16+2^2$
19	$2+2$	$480+2$	2^2	2^4	$8+2^2$	$264+2$	$24+2$	2^4	$16+2^2$
20	$2+2$	$480+2$	2^2	2^4	$8+2$	$\infty+264+2$	$24+2^2$	2^4	$16+2^2$
21	$2+2$	$480+2$	2^2	2^4	$8+2$	$264+2$	$24+2$	2^3	$8+2^2$
22	$2+2$	$480+2$	2^2	2^4	$8+2$	$264+2$	24	$\infty+2^2$	$4+2^2$
23	$2+2$	$480+2$	2^2	2^4	$8+2$	$264+2$	24	2^2	2^3
$\geqslant 24$	$2+2$	$480+2$	2^2	2^4	$8+2$	$264+2$	24	2^2	2^2

Remarks

(1) When $n>k+1$ (below the broken line in the table), $\pi_{n+k}(S^n)$ is independent of n and is isomorphic with the kth stable homotopy group G_k.

(2) Let $\iota_n \in \pi_n(S^n)$ be the identity on S^n; $\eta_2 \in \pi_3(S^2)$, $\nu_4 \in \pi_7(S^4)$, $\sigma_8 \in \pi_{15}(S^8)$ be the Hopf mapping $S^3 \to S^2$, $S^7 \to S^4$, $S^{15} \to S^8$ (induced mapping in the homotopy class), respectively; and $[\iota_{2m}, \iota_{2m}] \in \pi_{4m-1}(S^{2m})$ $(m \neq 1,2,4)$ be the Whitehead product of ι_{2m}. These objects generate infinite cyclic groups which are direct factors of $\pi_{n+k}(S^n)$ corresponding to the original mappings.

(3) $\eta_{n+2} = E^n \eta_2$, $\nu_{n+4} = E^n \nu_4$, $\sigma_{n+8} = E^n \sigma_8$ $(n \geqslant 1)$ (E is the suspension) are the generator for $\pi_{n+k}(S^n)$, which contains the mappings.

(4) The orders of the following compositions are 2:gs_{n+7}

$$\eta_n \circ \eta_{n+1} \ (n \geqslant 2), \quad \nu_n \circ \nu_{n+3} \ (n \geqslant 5), \quad \sigma_n \circ \sigma_{n+7} \ (n \geqslant 16), \quad \eta_n \circ \nu_{n+1} \ (n=3,4),$$

$$\nu_n \circ \eta_{n+3} \ (n=4,5), \quad \eta_n \circ \sigma_{n+1} \ (n \geqslant 7), \quad \sigma_n \circ \eta_{n+7} \ (n \geqslant 8), \quad \nu_{10} \circ \sigma_{13}, \quad \sigma_{11} \circ \nu_{18},$$

$$\eta_n \circ \eta_{n+1} \circ \eta_{n+2} \ (n \geqslant 2), \quad \nu_n \circ \nu_{n+3} \circ \nu_{n+6} (n \geqslant 4), \quad \sigma_n \circ \sigma_{n+7} \circ \sigma_{n+14} \ (n \geqslant 9).$$

(VII) The Homotopy Groups $\pi_k(G)$ of Compact Connected Lie Groups G

Here the group G is one of the following:

$$SO(n) \ (n \geqslant 2), \quad Spin(n) \ (n \geqslant 3), \quad U(n) \ (n \geqslant 1), \quad SU(n) \ (n \geqslant 2),$$
$$Sp(n) \ (n \geqslant 1), \quad G_2, \quad F_4, \quad E_6, \quad E_7, \quad E_8.$$

(1) The Fundamental Group $\pi_1(G)$.

$$\pi_1(G) \cong \begin{cases} \infty & (G = U(n) \ (n \geqslant 1), \ SO(2)), \\ 2 & (G = SO(n) \ (n \geqslant 3)), \\ 0 & (\text{for all other groups } G). \end{cases}$$

(2) Isomorphic Relations $(k \geqslant 2)$.

$\pi_k(U(n)) \cong \pi_k(SU(n))$ $(n \geqslant 2)$,

$\pi_k(U(1)) \cong \pi_k(SO(2)) \cong 0.$

$\pi_k(Spin(n)) \cong \pi_k(SO(n))\ (n \geqslant 3),$

$\pi_k(Spin(3)) \cong \pi_k(Sp(1)) \cong \pi_k(SU(2)) \cong \pi_k(S^3),$

$\pi_k(Spin(4)) \cong \pi_k(Spin(3)) + \pi_k(S^3),$

$\pi_k(Spin(5)) \cong \pi_k(Sp(2)),$

$\pi_k(Spin(6)) \cong \pi_k(SU(4)).$

(3) The Homotopy Group $\pi_k(G)$ $(k \geqslant 2)$.

$\pi_2(G) \cong 0.$

$\pi_3(G) \cong \infty \quad (G \neq SO(2),\ U(1),\ SO(4),\ Spin(4)), \quad \pi_3(SO(4)) \cong \infty + \infty.$

$$\pi_4(G) \cong \begin{cases} 2+2 & (G = SO(4),\ Spin(4)), \\ 2 & (G = Sp(n),\ SU(2),\ SO(3),\ SO(5),\ Spin(3),\ Spin(5)), \\ 0 & (G = SU(n)\ (n \geqslant 3),\ SO(n)\ (n \geqslant 6),\ G_2,\ F_4,\ E_6,\ E_7,\ E_8). \end{cases}$$

$$\pi_5(G) \cong \begin{cases} 2+2 & (G = SO(4),\ Spin(4)), \\ 2 & (G = Sp(n),\ SU(2),\ SO(3),\ SO(5)\ Spin(3),\ Spin(5)), \\ \infty & (G = SU(n)\ (n \geqslant 3),\ SO(6),\ Spin(6)), \\ 0 & (G = SO(n),\ Spin(n)\ (n \geqslant 7),\ G_2,\ F_4,\ E_6,\ E_7,\ E_8). \end{cases}$$

$\pi_k(G), \quad k \geqslant 6.$

$G \quad k$	6	7	8	9	10	11	12	13	14	15
$Sp(1)$	12	2	2	3	15	2	2^2	$12+2$	$84+2^2$	2^2
$Sp(2)$	0	∞	0	0	120	2	2^2	$4+2$	1680	2
$Sp(3)$	0	∞	0	0	0	∞	2	2	10080	2
$Sp(4)$	0	∞	0	0	0	∞	2	2	0	∞
$SU(2)$	12	2	2	3	15	2	2^2	$12+3$	$84+2^2$	2^2
$SU(3)$	6	0	12	3	30	4	60	6	$84+2$	36
$SU(4)$	0	∞	24	2	$120+2$	4	60	4	$1680+2$	$72+2$
$SU(5)$	0	∞	0	∞	120	0	360	4	1680	6
$SU(6)$	0	∞	0	∞	0	∞	720	2	$5040+2$	6
$SU(7)$	0	∞	0	∞	0	∞	0	∞	5040	0
$SU(8)$	0	∞	0	∞	0	∞	0	∞	0	∞
$SO(5)$	0	∞	0	0	120	2	2^2	$4+2$	1680	2
$SO(6)$	0	∞	24	2	$120+2$	4	60	4	$1680+2$	$72+2$
$SO(7)$	0	∞	2^2	2^2	8	$\infty+2$	0	2	$2520+8+2$	2^4
$SO(8)$	0	$\infty+\infty$	2^3	2^3	$24+8$	$\infty+2$	0	2^2	$2520+120+8+2$	2^7
$SO(9)$	0	∞	2^2	2^2	8	$\infty+2$	0	2	$8+2$	$\infty+2^3$
$SO(10)$	0	∞	2	$\infty+2$	4	∞	12	2	8	$\infty+2^2$
$SO(11)$	0	∞	2	2	2	∞	2	2^2	8	$\infty+2$
$SO(12)$	0	∞	2	2	0	$\infty+\infty$	2^2	2^2	$24+4$	$\infty+2$
$SO(13)$	0	∞	2	2	0	∞	2	2	8	$\infty+2$
$SO(14)$	0	∞	2	2	0	∞	0	∞	4	∞
$SO(15)$	0	∞	2	2	0	∞	0	0	2	∞
$SO(16)$	0	∞	2	2	0	∞	0	0	0	$\infty+\infty$
$SO(17)$	0	∞	2	2	0	∞	0	0	0	∞
G_2	3	0	2	6	0	$\infty+2$	0	0	$168+2$	2
F_4	0	0	2	2	0	$\infty+2$	0	0	2	∞
E_6	0	0	0	∞	0	∞	12	0	0	∞
E_7	0	0	0	0	0	∞	2	2	0	∞
E_8	0	0	0	0	0	0	0	0	0	∞

(4) Stable Homotopy Groups. For sufficiently large n for fixed k, the homotopy groups for classical compact simple Lie groups $G = Sp(n)$, $SU(n)$, $SO(n)$ become stable. We denote them by the following notations. Here we assume $k \geqslant 2$.

$$\pi_k(Sp) = \pi_k(Sp(n)) \quad (n \geqslant (k-1)/4),$$
$$\pi_k(U) = \pi_k(U(n)) \cong \pi_k(SU(n)) \quad (n \geqslant (k+1)/2),$$
$$\pi_k(O) = \pi_k(SO(n)) \quad (n \geqslant k+2).$$

Bott periodicity theorem

$$\pi_k(Sp) \cong \begin{cases} \infty & (k \equiv 3,7 \ (\mathrm{mod}\, 8)), \\ 2 & (k \equiv 4,5 \ (\mathrm{mod}\, 8)), \\ 0 & (k \equiv 0,1,2,6 \ (\mathrm{mod}\, 8)). \end{cases}$$

$$\pi_k(O) \cong \begin{cases} \infty & (k \equiv 3,7 \ (\mathrm{mod}\, 8)), \\ 2 & (k \equiv 0,1 \ (\mathrm{mod}\, 8)), \\ 0 & (k \equiv 2,4,5,6 \ (\mathrm{mod}\, 8)). \end{cases}$$

$$\pi_k(U) \cong \begin{cases} \infty & (k \equiv 1 \ (\mathrm{mod}\, 2)), \\ 0 & (k \equiv 0 \ (\mathrm{mod}\, 2)). \end{cases}$$

(5) Metastable Homotopy Groups.

(a,b) means the greatest common divisor of two integers a and b.

$$\pi_{2n}(SU(n)) \cong n!.$$

$$\pi_{2n+1}(SU(n)) \cong \begin{cases} 2 & (n \text{ even}), \\ 0 & (n \text{ odd}). \end{cases}$$

$$\pi_{2n+2}(SU(n)) \cong \begin{cases} (n+1)!+2 & (n \text{ even}, \geqslant 4), \\ (n+1)!/2 & (n \text{ odd}). \end{cases}$$

$$\pi_{2n+3}(SU(n)) \cong \begin{cases} (24,n) & (n \text{ even}), \\ (24,\pi+3)/2 & (n \text{ odd}). \end{cases}$$

$$\pi_{2n+4}(SU(n)) \cong \begin{cases} (n+2)!(24,n)/48 & (n \text{ even}, \geqslant 4), \\ (n+2)!(24,n+3)/24 & (n \text{ odd}). \end{cases}$$

$$\pi_{2n+5}(SU(n)) \cong \pi_{2n+5}(U(n+1)).$$

$$\pi_{2n+6}(SU(n)) \cong \begin{cases} \pi_{2n+6}(U(n+1)) & (n \equiv 2,3 \ (\mathrm{mod}\, 4),\ n \geqslant 3), \\ \pi_{2n+6}(U(n+1))+2 & (n \equiv 0,1 \ (\mathrm{mod}\, 4)). \end{cases}$$

$$\pi_{4n+2}(Sp(n)) \cong \begin{cases} (2n+1)! & (n \text{ even}), \\ 2(2n+1)! & (n \text{ odd}). \end{cases}$$

$$\pi_{4n+3}(Sp(n)) \cong 2.$$

$$\pi_{4n+4}(Sp(n)) \cong \begin{cases} 2+2 & (n \text{ even}), \\ 2 & (n \text{ odd}). \end{cases}$$

$$\pi_{4n+5}(Sp(n)) \cong \begin{cases} (24,n+2)+2 & (n \text{ even}), \\ (24,n+2) & (n \text{ odd}). \end{cases}$$

$$\pi_{4n+6}(Sp(n)) \cong \begin{cases} (2n+3)!(24,n+2)/12 & (n \text{ even}), \\ (2n+3)!(24,n+2)/24 & (n \text{ odd}). \end{cases}$$

$$\pi_{4n+7}(Sp(n)) \cong 2.$$
$$\pi_{4n+8}(Sp(n)) \cong 2+2.$$

The homotopy groups $\pi_{n+i}(SO(n))$ for $n \geqslant 16$, $3 \geqslant i \geqslant -1$ are determined by the isomorphism

$$\pi_{n+i}(SO(n)) \cong \pi_{n+i}(O) + \pi_{n+i+1}(V_{i+3+n,i+3}(\mathbf{R}))$$

and the homotopy groups of $V_{m+n,m}(\mathbf{R})$ given below.

(6) Homotopy Groups of Real Stiefel Manifolds $V_{m+n,m}(\mathbf{R}) = O(m+n)/I_m \times O(n)$.

$$\pi_{n+k}(V_{n+1,1}) \cong \pi_{n+k}(S^n).$$

$$\pi_{n-k}(V_{m+n,m}) \cong 0 \quad (k \geqslant 1).$$

$$\pi_n(V_{m+n,m}) \cong \begin{cases} 2 & (n=2s-1,\ m \geqslant 2), \\ \infty & (n=2s). \end{cases}$$

$\pi_{n+k}(V_{m+n,m})$ $(k=1,2,3,4,5)$ are given in the following table.

	$n \backslash m$	1	2	3	4	5	6	$8s-1$	$8s$	$8s+1$	$8s+2$	$8s+3$	$8s+4$	$8s+5$	$8s+6$
π_{n+1}	2	0	∞^2	2	$2+\infty$	2	$2+\infty$	2	$2+\infty$	2	$2+\infty$	2	$2+\infty$	2	$2+\infty$
	$\geqslant 3$	0	∞	0	2^2	2	4	0	2^2	2	4	0	2^2	2	4
π_{n+2}	2	∞	2^2	4	2^2	4	2^2	4	2^2	4	2^2	4	2^2	4	2^2
	3	∞^2	2	$2+\infty$	2^2	$4+\infty$	2	$2+\infty$	2^2	$4+\infty$	2	$2+\infty$	2^2	$4+\infty$	2
	$\geqslant 4$	∞	0	2	2^2	8	0	2	2^2	8	0	2	2^2	8	0
π_{n+3}	2	2	2^2	2	$\infty+12+2$	2^2	$24+2$	2^2	$24+2$	2^2	$24+4$	2^2	$24+2$	2^2	$24+2$
	3	2^2	2	2	$\infty+12+4$	2^3	$12+2$	2^2	$24+4$	2^3	$12+2$	2^2	$24+4$	2^3	$12+2$
	4	2	∞	2	∞^2+12+4	2^2	$12+\infty$	2^2	$24+4+\infty$	2^2	$12+\infty$	2^2	$24+4+\infty$	2^2	$12+\infty$
	$\geqslant 5$	0	0	2	$12+4+\infty$	2	12	2	$24+8$	2	12	2^2	$4+48$	2	12
π_{n+4}	2	2	12^2	$\infty+2$	2^2+24	2^2	24	2	24	2	24	2	24	2	24
	3	2^2	0	$\infty+4$	2^4	2^3	2	4	2^2	2^2	2	4	2^2	2^2	2
	4	2	0	$4+\infty$	2^5	2^2	2	8	2^3	2	2	8	2^3	2	2
	5	∞	0	$4+\infty^2$	2^4	$2+\infty$	2	$8+\infty$	2^2	∞	2	$8+\infty$	2^2	∞	2
	$\geqslant 6$	0	0	$4+\infty$	2^3	2	0	8	2	0	2	16	2	0	0
π_{n+5}	2	12	2^2	2	2^3	0	∞	0	0	0	0	0	0	0	0
	3	12^2	∞	$2+24$	2^4	24	$\infty+2$	24	2	24	2	24	2	24	2
	4	0	∞	2^3	2^5	2	$\infty+4$	2^2	2^2	2	4	2^2	2^2	2	4

(VIII) Immersion and Embedding of Projective Spaces (→ 114 Differential Topology)

We denote immersion by $\subset$, and embedding by $\subseteq$. $\mathbf{P}^n(A)$ is an n-dimensional real or complex projective space where $A=\mathbf{R}$ or $\mathbf{C}$, $k\{\mathbf{P}^n(A)\}$ is the integer k such that $\mathbf{P}^n(A) \subset \mathbf{R}^k$ and $\mathbf{P}^n(A) \not\subset \mathbf{R}^{k-1}$, and $\tilde{k}\{\mathbf{P}^n(A)\}$ is the integer k such that $\mathbf{P}^n(A) \subseteq \mathbf{R}^k$ and $\mathbf{P}^n(A) \not\subseteq \mathbf{R}^{k-1}$.

In the table, for example, numbers 9–11 in the row $k\{\mathbf{P}^n(\mathbf{R})\}$ for $n=6$ mean $\mathbf{P}^6(\mathbf{R}) \not\subset \mathbf{R}^8$, $P^6(\mathbf{R}) \subset \mathbf{R}^{11}$.

n	1	2	3	4	5	6	7	8	9	10	11	12
$k\{\mathbf{P}^n(\mathbf{R})\}$	2	4	5	8	9	9~11	9~12	16	17	17~19	...	...
$\tilde{k}\{\mathbf{P}^n(\mathbf{R})\}$	2	3	4	7	7	7	8	15	15	16	16	17~19
$k\{\mathbf{P}^n(\mathbf{C})\}$	3	7	9	15	17	22	22~25	31	33	38	38~41	...
$\tilde{k}\{\mathbf{P}^n(\mathbf{C})\}$	3	7	8~9	15	16~17	22	22~25	31	32~33	38	38~41	...

	2^r	2^r+1	2^r+2	2^r+3	2^r+2^s $(r>s>0)$
$k\{\mathbf{P}^n(\mathbf{R})\}$	$2n$	$2n-1$	$2n-3$~$2n-1$	...	...
$\tilde{k}\{\mathbf{P}^n(\mathbf{R})\}$	$2n-1$	$2n-3$	$2n-4$	$2n-6$	...
$k\{\mathbf{P}^n(\mathbf{C})\}$	$4n-1$	$4n-3$	$4n-2$	...	$4n-2$
$\tilde{k}\{\mathbf{P}^n(\mathbf{C})\}$	$4n-1$	$4n-4$~$4n-3$	$4n-2$	...	$4n-2$

7. Knot Theory (→ 235 Knot Theory)

Let k be a projection on a plane of a knot K. We color the domains separated by k, white and black alternatively. The outermost (unbounded) domain determined by k is colored white. In Fig. 16, hatching means black. Take a point (a black point in Fig. 16) in each black domain. The self-intersections of k are represented by white points (Fig. 16). Through each white point we draw a line segment connecting the black points in the black regions meeting at the white point. In Fig. 16, we show this as a broken line. We assign the signature + if the torsion of K at the intersection of k has the orientation of a right-hand screw (as in Fig. 17, left), and the signature − if the orientation is opposite (as in Fig. 17, right). The picture of the line segments with signatures is called the graph corresponding to the projection k of the knot K. Given such a graph, we can reconstruct the original knot K.

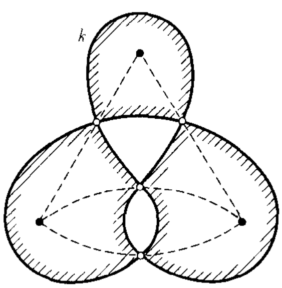


Fig. 16

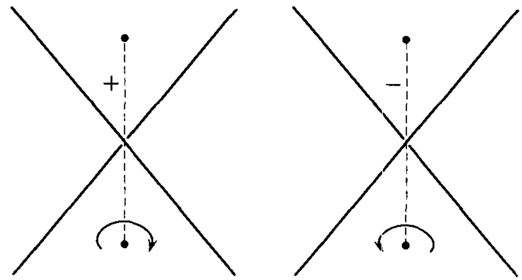


Fig. 17

Fig. 18 shows the classification table of knots for which the numbers of intersections of k are 3 to 8 when we minimize the intersections. The projection of k is described by a solid line, and the graph by broken lines. We omit the signatures since for each graph from 3_1 to 8_{18} they are all + or all −. Such knots are called alternating knots.

8. Inequalities (→ 88 Convex Analysis, 211 Inequalities)

(1) $|a+b| \leqslant |a|+|b|$,

$|a-b| \geqslant ||a|-|b||$.

For real a_ν, we have $\Sigma a_\nu^2 \geqslant 0$, and the equality holds only if all $a_\nu = 0$.

(2) $n! < n^n < (n!)^2 \quad (n \geqslant 3)$.

$e^n \geqslant n^n/n!$.

$n^{1/n} < 3^{1/3} \quad (n \neq 3)$.

(3) $2/\pi < (\sin x)/x < 1 \quad (0 < x < \pi/2)$ (Jordan's inequality).

(4) Denote the elementary symmetric polynomials of positive numbers $a_1, \ldots, a_n > 0$ by S_r $(r = 1, \ldots, n)$. Then

$$S_1 \Big/ \binom{n}{1} \geqslant \left[S_2 \Big/ \binom{n}{2} \right]^{1/2} \geqslant \ldots \geqslant \left[S_r \Big/ \binom{n}{r} \right]^{1/r} \geqslant \ldots \geqslant \left[S_n \Big/ \binom{n}{n} \right]^{1/n}.$$

If at least one equality holds, then $a_1 = \ldots = a_n$. In particular, from the two external terms, we have the following inequalities concerning mean values:

$$\frac{1}{n} \sum_{\nu=1}^{n} a_\nu \geqslant \left(\prod_{\nu=1}^{n} a_\nu \right)^{1/n} \geqslant n \Big/ \sum_{\nu=1}^{n} \frac{1}{a_\nu}.$$

For weighted means, we have

$$\sum_{\nu=1}^{n} \lambda_\nu a_\nu \geqslant \prod_{\nu=1}^{n} a_\nu^{\lambda_\nu} \qquad \left(\sum \lambda_\nu = 1, \quad \lambda_\nu > 0 \right).$$

(5) When $a_\nu > 0$, $b_\nu > 0$, $p > 1$, $q > 1$, $(1/p) + (1/q) = 1$,

$$\left[\sum_{\nu=1}^{n} (a_\nu)^p \right]^{1/p} \left[\sum_{\nu=1}^{n} (b_\nu)^q \right]^{1/q} \geqslant \sum_{\nu=1}^{n} a_\nu b_\nu \qquad \text{(Hölder's inequality).}$$

The equality holds only if $(a_\nu)^p = c(b_\nu)^q$ (c is a constant).

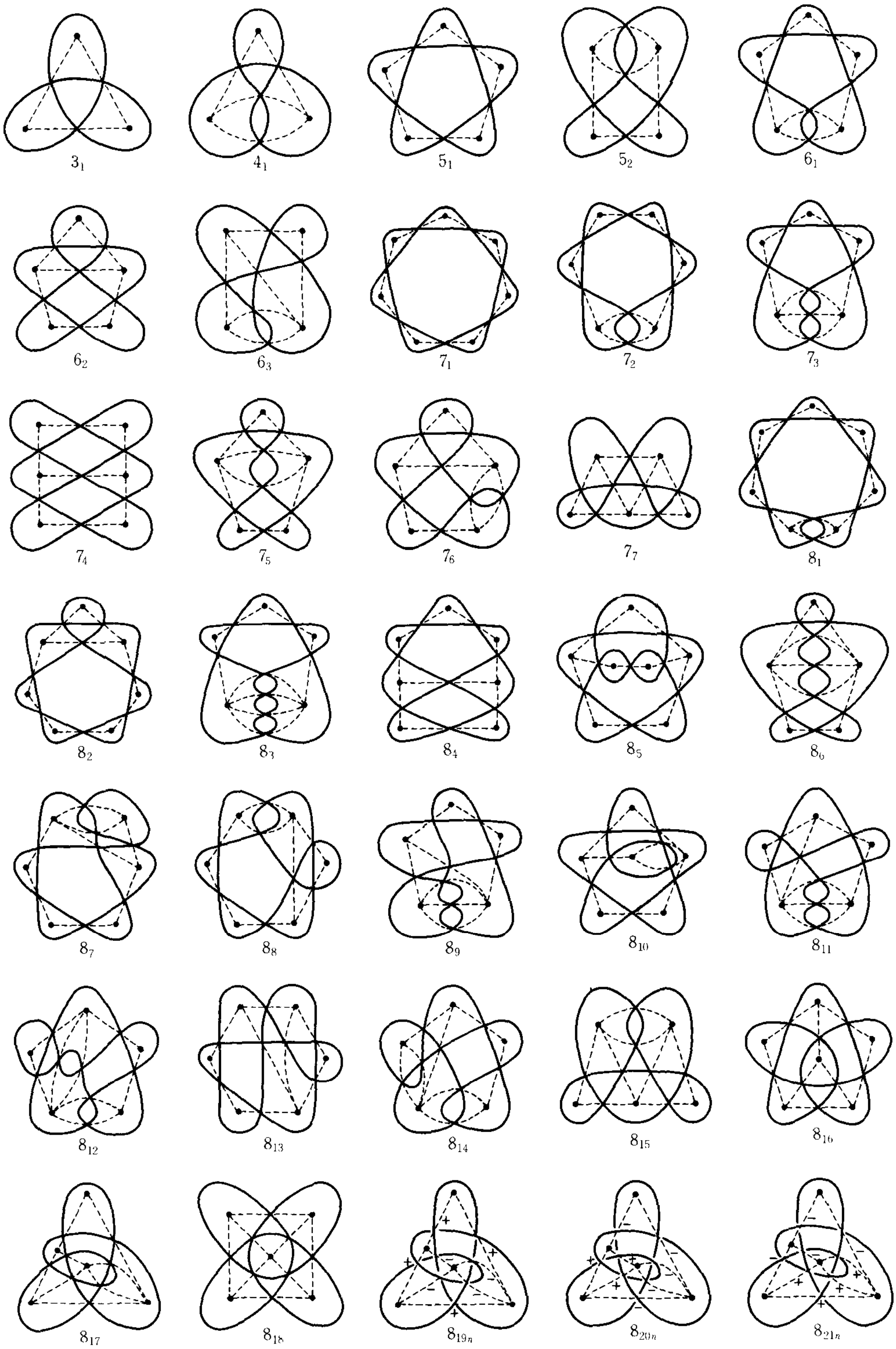


Fig. 18
Classification table of knots. The signatures from 3_1 to 8_{18} are all + or all −.

When $p=q=2$, the inequality is called Cauchy's inequality, the Cauchy-Schwarz inequality, or Bunyakovskiĭ's inequality. As special cases, we have

$$\left(\sum_{\nu=1}^{n} a_\nu\right)\left(\sum_{\nu=1}^{n} \frac{1}{a_\nu}\right) \geqslant n^2 \quad (a_\nu>0),$$

$$\left(\sum_{\nu=1}^{n} a_\nu\right)^2 \leqslant n\left(\sum_{\nu=1}^{n} a_\nu^2\right) \quad (a_\nu>0).$$

When $0<p<1$, we have an inequality by reversing the inequality sign in Hölder's inequality.

(6) When $a_\nu > 0$, $b_\nu > 0$, $p > 0$, and $\{a_\nu\}$ and $\{b_\nu\}$ are not proportional, we have

$$\left[\sum_{\nu=1}^{n}(a_\nu+b_\nu)^p\right]^{1/p} \lesseqgtr \left[\sum_{\nu=1}^{n}(a_\nu)^p\right]^{1/p}+\left[\sum_{\nu=1}^{n}(b_\nu)^p\right]^{1/p} \quad (p \gtrless 1) \qquad \text{(Minkowski's inequality).}$$

The integral inequality corresponding to (5) or (6) has the same name.

(7) If $a_{\mu\nu} > 0$, $\sum_{\mu=1}^{n} a_{\mu\nu} = \sum_{\nu=1}^{n} a_{\mu\nu} = 1$, $b_\nu \geqslant 0$,

$$\prod_{\nu=1}^{n} b_\nu \leqslant \prod_{\mu=1}^{n}\left(\sum_{\nu=1}^{n} a_{\mu\nu} b_\nu\right).$$

In particular, for the determinant $\Delta = \det(a_{\mu\nu})$,

$$|\Delta|^2 \leqslant \prod_{\nu=1}^{n}\left(\sum_{\mu=1}^{n}|a_{\mu\nu}|^2\right).$$

The equality in this holds only if all rows are mutually orthogonal. If all $|a_{\mu\nu}| \leqslant M$, we have

$$|\Delta| \leqslant n^{n/2}M^n \qquad \text{(Hadamard's estimation).}$$

(8) Suppose that a function $f(x)$ is continuous, strictly monotone increasing in $x \geqslant 0$, and $f(0)=0$. Denote the inverse function of f by f^{-1}. For $a,b>0$, we have

$$ab \leqslant \int_0^a f(x)\,dx + \int_0^b f^{-1}(x)\,dx \qquad \text{(Young's inequality),}$$

and the equality holds only if $b=f(a)$.

In particular, for $f(x)=x^{p-1}$ $(p>1)$, we have

$$\frac{a^p}{p}+\frac{b^q}{q} \geqslant ab,$$

where $(1/p)+(1/q)=1$.

(9) If $p,q>1$, $(1/p)+(1/q)=1$, $a_\mu \geqslant 0$, $b_\nu \geqslant 0$,

$$\sum_{\mu,\nu=0}^{\infty}\frac{a_\mu b_\nu}{\mu+\nu+1} \leqslant \frac{\pi}{\sin(\pi/p)}\left[\sum_{\mu=0}^{\infty}(a_\mu)^p\right]^{1/p}\left[\sum_{\nu=0}^{\infty}(b_\nu)^q\right]^{1/q} \qquad \text{(Hilbert's inequality),}$$

and the equality holds only when the right-hand side vanishes.

(10) For a continuous function $f(x) \geqslant 0$ $(0 \leqslant x < \infty)$, we put

$$F(x)=\int_0^x f(t)\,dt,$$

and assume that $p>1$. Then

$$\int_0^\infty\left[\frac{F(x)}{x}\right]^p dx \leqslant \left(\frac{p}{p-1}\right)^p\int_0^\infty [f(x)]^p\,dx \qquad \text{(Hardy's inequality),}$$

and the equality holds only if $f(x)$ is identically 0.

Further, if $f(x)>0$,

$$\int_0^\infty \exp\left[\frac{1}{x}\int_0^x \log f(t)\,dt\right]dx < e\int_0^\infty f(x)\,dx \qquad \text{(Carleman's inequality).}$$

(11) Let $a \leqslant x < \xi \leqslant b$, $p>1$, and

$$\sup_{\xi}\frac{1}{\xi-x}\int_x^\xi f(t)\,dt = \theta(x).$$

Then

$$\int_a^b [\theta(x)]^p\,dx \leqslant 2\left(\frac{p}{p-1}\right)^p\int_a^b |f(x)|^p\,dx \qquad \text{(Hardy-Littlewood supremum theorem).}$$

(12) If $f(x)$ is piecewise smooth in $0 \leqslant x \leqslant \pi$ and $f(0)=f(\pi)=0$,

$$\int_0^\pi [f'(x)]^2\,dx \geqslant \int_0^\pi [f(x)]^2\,dx \qquad \text{(Wirtinger's inequality),}$$

and the equality holds only if $f(x)$ is a constant multiple of $\sin x$.

9. Differential and Integral Calculus

(I) Derivatives and Primitive Functions (→ 106 Differential Calculus, 216 Integral Calculus)

$F(x)\equiv\int f(x)dx$	$f(x)\equiv F'(x)$
$\alpha\varphi+\beta\psi$ (α,β constants)	$\alpha\varphi'+\beta\psi'$
$\varphi\cdot\psi$	$\varphi'\psi+\varphi\psi'$
φ/ψ ($\psi\neq 0$)	$(\varphi'\psi-\varphi\psi')/\psi^2$
$\log\lvert\varphi\rvert$ ($\varphi\neq 0$)	φ'/φ (logarithmic differentiation)
$\Phi(\varphi)$ (composite)	$(d\Phi/d\varphi)\cdot\varphi'$
c (constant)	0
x^n	nx^{n-1}
$x^{n+1}/(n+1)$	x^n ($n\neq -1$)
$\log\lvert x\rvert$	$1/x$
$\log_a\lvert x\rvert$	$(\log_a e)/x$
$x(\log x-1)$	$\log x$
$\exp x=e^x$	$\exp x=e^x$
a^x ($a>0$)	$a^x\log a$
x^x	$x^x(1+\log x)$
$(x-1)e^x$	xe^x
$\sin x$	$\cos x$
$\cos x$	$-\sin x$
$\tan x$	$\sec^2 x$
$\cot x$	$-\operatorname{cosec}^2 x$
$\sec x$	$\sec x\tan x$
$\operatorname{cosec} x$	$-\operatorname{cosec} x\cot x$
$\sinh x=(e^x-e^{-x})/2$	$\cosh x$
$\cosh x=(e^x+e^{-x})/2$	$\sinh x$
$\tanh x=\sinh x/\cosh x$	$\operatorname{sech}^2 x$
$\coth x=\cosh x/\sinh x$	$-\operatorname{cosech}^2 x$
$\operatorname{sech} x=1/\cosh x$	$-\operatorname{sech} x\tanh x$
$\operatorname{cosech} x=1/\sinh x$	$-\operatorname{cosech} x\coth x$
$\arcsin x$ ($\lvert F\rvert<\pi/2$)	$1/\sqrt{1-x^2}$
$\arccos x$ ($0<F<\pi$)	$-1/\sqrt{1-x^2}$
$\arctan x$ ($\lvert F\rvert<\pi/2$)	$1/(1+x^2)$
$\operatorname{arccot} x$ ($\lvert F\rvert<\pi/2$)	$-1/(1+x^2)$
$\operatorname{arcsec} x$ ($0<F<\pi$)	$1/\lvert x\rvert\sqrt{x^2-1}$
$\operatorname{arccosec} x$ ($\lvert F\rvert<\pi/2$)	$-1/\lvert x\rvert\sqrt{x^2-1}$
$\operatorname{arcsinh} x=\log(x+\sqrt{x^2+1})$	$1/\sqrt{x^2+1}$
$\operatorname{arccosh} x=\log(x+\sqrt{x^2-1})$	$1/\sqrt{x^2-1}$
$\frac{1}{2}\log\left\lvert\frac{1+x}{1-x}\right\rvert=\begin{cases}\operatorname{arctanh} x & (\lvert x\rvert<1)\\ \operatorname{arccoth} x & (\lvert x\rvert>1)\end{cases}$	$\frac{1}{1-x^2}$
$\operatorname{arcsech} x$	$-1/x\sqrt{1-x^2}$
$\operatorname{arccosech} x$	$-1/\lvert x\rvert\sqrt{1+x^2}$

$F(x)\equiv\int f(x)\,dx$	$f(x)\equiv F'(x)$
$\frac{1}{2a}\log\left\lvert\frac{a-x}{a+x}\right\rvert \quad (a>0)$	$\frac{1}{x^2-a^2}$
$(1/a)\arctan(x/a)$	$1/(x^2+a^2)$
$(x\sqrt{1-x^2}+\arcsin x)/2$	$\sqrt{1-x^2}$
$[x\sqrt{x^2\pm 1}\pm\log(x+\sqrt{x^2\pm 1})]/2$	$\sqrt{x^2\pm 1}$
$-\log\lvert\cos x\rvert$	$\tan x$
$\log\lvert\sin x\rvert$	$\cot x$
$\log\lvert\tan x\rvert$	$1/\sin x\cos x$
$\log\lvert\tan[(\pi/4)+(x/2)]\rvert$	$\sec x$
$\log\lvert\tan(x/2)\rvert$	$\operatorname{cosec} x$
$(x/2)-(1/4)\sin 2x$	$\sin^2 x$
$\sin x - x\cos x$	$x\sin x$
$\cos x + x\sin x$	$x\cos x$
$\frac{n\sin mx\sin nx + m\cos mx\cos nx}{n^2-m^2} \quad (n^2\neq m^2)$	$\sin mx\cos nx$
$e^{bx}\frac{b\sin ax - a\cos ax}{a^2+b^2}$	$e^{bx}\sin ax$
$e^{bx}\frac{b\cos ax + a\sin ax}{a^2+b^2}$	$e^{bx}\cos ax$
$x\arcsin x+\sqrt{1-x^2}$	$\arcsin x$
$x\arctan x-(1/2)\log(1+x^2)$	$\arctan x$
$\det(\varphi_{jk})_{j,k=1,\ldots,n}$	$\sum\det(\varphi_{j1}\cdots\varphi_{j\nu-1}\varphi'_{j\nu}\varphi_{j\nu+1}\cdots\varphi_{jn})_{j=1,\ldots,n}$

(II) Recurrence Formulas for Indefinite Integrals

(1) $I_m\equiv\int\frac{dx}{(1+x^2)^m}$ (m is a positive integer).

$$I_m=\frac{1}{2m-2}\,\frac{x}{(1+x^2)^{m-1}}+\frac{2m-3}{2m-2}I_{m-1}\quad (m\geqslant 2);\quad I_1=\arctan x.$$

(2) $I_m\equiv\int\frac{x^m}{\sqrt{ax^2+bx+c}}\,dx$ (m is an integer, $a\neq 0$).

The case $m<0$ is reduced to the case $m\geqslant 0$ by the change of variable $1/x=t$.

$$I_m=\frac{1}{ma}x^{m-1}\sqrt{ax^2+bx+c}-\frac{(2m-1)b}{2ma}I_{m-1}-\frac{(m-1)c}{ma}I_{m-2}\quad (m\geqslant 1);$$

$$I_0=\begin{cases}(1/\sqrt{a})\log\lvert 2ax+b+2\sqrt{a}\sqrt{ax^2+bx+c}\rvert & (a>0),\\ \frac{-1}{\sqrt{-a}}\arcsin\frac{2ax+b}{\sqrt{b^2-4ac}} & (a<0);\end{cases}$$

In this case, for the integrand to be a real function it is necessary that $b^2-4ac>0$.

(3) $I_m\equiv\int x^m e^x\,dx$ (m is an integer).

$$I_m=x^m e^x-mI_{m-1};\ I_0=e^x,\ I_{-1}=\operatorname{Ei}x,$$

where Ei is the exponential integral function (⟶ Table 19.II.3, this Appendix).

(4) $I_{m,n}\equiv\int x^m(\log x)^n\,dx$ (m,n are integers, $n\geqslant 0$).

$$I_{m,n}=\frac{x^{m+1}}{m+1}(\log x)^n-\frac{n}{m+1}I_{m,n-1};\; I_{m,0}=\frac{x^{m+1}}{m+1}\quad (m\neq-1),\; I_{-1,n}=(\log x)^{n+1}/(n+1).$$

(5) $I_m\equiv\int x^m\sin x\,dx$, $J_m\equiv\int x^m\cos x\,dx$ (m is a nonnegative integer).

$$I_m=-x^m\cos x+mJ_{m-1}=x^{m-1}(m\sin x-x\cos x)-m(m-1)I_{m-2},$$

$$J_m=x^m\sin x-mI_{m-1}=x^{m-1}(x\sin x+m\cos x)-m(m-1)J_{m-2};$$

$$I_0=-\cos x,$$

$$J_0=\sin x,$$

(6) $I_{m,n}\equiv\int\sin^m x\cdot\cos^n x\,dx$ (m,n are integers).

$$\left.\begin{aligned}I_{m,n}&=\frac{\sin^{m+1}x\cos^{n-1}x}{m+n}+\frac{n-1}{m+n}I_{m,n-2}\\ I_{m,n}&=\frac{-\sin^{m-1}x\cos^{n+1}x}{m+n}+\frac{m-1}{m+n}I_{m-2,n}\end{aligned}\right\}(m+n\neq 0),$$

$$I_{m,n}=\frac{-\sin^{m+1}x\cos^{n+1}x}{n+1}+\frac{m+n+2}{n+1}I_{m,n+2}\;(n\neq-1),$$

$$I_{m,n}=\frac{\sin^{m+1}x\cos^{n+1}x}{m+1}+\frac{m+n+2}{m+1}I_{m+2,n}\;(m\neq-1);$$

$$I_{1,1}=(\sin^2 x)/2,\quad I_{1,0}=-\cos x,\quad I_{1,-1}=-\log|\cos x|,\quad I_{0,1}=\sin x,\quad I_{0,0}=x,$$

$$I_{0,-1}=\log|\tan[(x/2)+(\pi/4)]|,\quad I_{-1,1}=\log|\sin x|,$$

$$I_{-1,0}=\log|\tan(x/2)|,\quad I_{-1,-1}=\log|\tan x|.$$

$$I_{m,-m}\equiv\int\tan^m x\,dx=\frac{\tan^{m-1}x}{m-1}-I_{m-2,-(m-2)}(m\neq 1).$$

(III) Derivatives of Higher Order

$f(x)$	$f^{(n)}(x)$
$\varphi\cdot\psi$	$\sum_{\nu=0}^{n}\binom{n}{\nu}\varphi^{(\nu)}\psi^{(n-\nu)}$ (Leibniz's formula)
x^k	$\prod_{\nu=0}^{n-1}(k-\nu)x^{k-n}$
$(x+a)^n$	$n!$
$\exp x$	$\exp x$
a^x $(a>0)$	$a^x(\log a)^n$
$\log x$	$(-1)^{n-1}(n-1)!/x^n$
$\sin x$	$\sin[x+(n\pi/2)]$
$\cos x$	$\cos[x+(n\pi/2)]$
$e^{ax}\cos bx$	$r^n e^{ax}\cos(bx+n\theta)$ (where $a=r\cos\theta$, $b=r\sin\theta$)
$\arcsin x$	$\frac{1}{2^{n-1}}\sum_{\nu=0}^{n-1}(-1)^\nu\binom{n-1}{\nu}(2\nu-1)!!(2n-2\nu-3)!!(1+x)^{-(1/2)-\nu}(1-x)^{(1/2)-n+\nu}$ (where $(2\nu-1)!!\equiv 1\cdot3\cdot5\cdot\ldots\cdot(2\nu-1)$, $(-1)!!\equiv 1$)
$\arctan x$	$(-1)^{n-1}(n-1)!\sin^n\theta\sin n\theta$ (where $x=\cot\theta$)

$$\left(\frac{1}{f}\right)'=-\frac{f'}{f^2},\quad\left(\frac{1}{f}\right)''=\frac{2f'^2-ff''}{f^3},\quad\left(\frac{1}{f}\right)'''=\frac{6ff'f''-6f'^3-f^2f'''}{f^4}.$$

Higher-order derivatives of a composite function $g(t)\equiv f(x_1(t),\ldots,x_n(t))$

$$\frac{dg}{dt}=\sum_{i=1}^{n}\frac{\partial f}{\partial x_i}\frac{dx_i}{dt},\quad \frac{d^2g}{dt^2}=\sum_{i=1}^{n}\frac{\partial f}{\partial x_i}\frac{d^2x_i}{dt^2}+\sum_{i,j=1}^{n}\frac{\partial^2 f}{\partial x_i\partial x_j}\frac{dx_i}{dt}\frac{dx_j}{dt},$$

$$\frac{d^3g}{dt^3}=\sum_{i=1}^{n}\frac{\partial f}{\partial x_i}\frac{d^3x_i}{dt^3}+2\sum_{i,j=1}^{n}\frac{\partial^2 f}{\partial x_i\partial x_j}\frac{d^2x_i}{dt^2}\frac{dx_j}{dt}+\sum_{i,j,k=1}^{n}\frac{\partial^3 f}{\partial x_i\partial x_j\partial x_k}\frac{dx_i}{dt}\frac{dx_j}{dt}\frac{dx_k}{dt}.$$

For a function $z=z\,(x_1,\ldots,x_n)$ determined implicitly by $F(z;x_1,\ldots,x_n)=0$, we have

$$\frac{\partial z}{\partial x_i}=-\frac{F_{x_i}}{F_z},\quad \frac{\partial^2 z}{\partial x_i\partial x_j}=-\frac{F_{x_ix_j}}{F_z}+\frac{F_{x_i}F_{x_jz}+F_{x_j}F_{x_iz}}{F_z^2}-\frac{F_{x_i}F_{x_j}F_{zz}}{F_z^3}.$$

Schwarzian derivative:

$$\{y;x\}\equiv\left(\frac{d^3y}{dx^3}\right)\Big/\left(\frac{dy}{dx}\right)-\frac{3}{2}\left[\left(\frac{d^2y}{dx^2}\right)\Big/\left(\frac{dy}{dx}\right)\right]^2,\quad \{y;x\}=0\Leftrightarrow y=(ax+b)/(cx+d),$$

$$\{y;x\}=\left(\frac{dz}{dx}\right)^2[\{y;z\}-\{x;z\}]=-\left(\frac{dy}{dx}\right)^2\{x;y\},\ \{(ay+b)/(cy+d);x\}=\{y;x\}.$$

(IV) The Taylor Expansion and Remainder

If $f(x)$ is n times continuously differentiable in the interval $[a,b]$ (i.e., of class C^n),

$$f(b)=\sum_{\nu=0}^{n-1}\frac{(b-a)^\nu}{\nu!}f^{(\nu)}(a)+R_n \qquad \text{(Taylor's formula)}.$$

R_n is called the remainder, and is represented as follows:

$$R_n=\frac{1}{(n-1)!}\int_a^b(b-x)^{n-1}f^{(n)}(x)\,dx=\frac{(b-a)^p(b-\xi)^{n-p}}{(n-1)!p}f^{(n)}(\xi)$$

$$=\frac{(b-a)^n}{(n-1)!p}(1-\theta)^{n-p}f^{(n)}(a+\theta(b-a))\quad (n\geqslant p>0,\ 0<\theta<1,\ a<\xi<b,\ \xi=a+\theta(b-a))$$

(Roche-Schlömilch remainder);

$$=\frac{1}{n!}(b-a)^n f^{(n)}(\xi) \qquad \text{(Lagrange's remainder)};$$

$$=\frac{1}{(n-1)!}(b-a)(b-\xi)^{n-1}f^{(n)}(\xi) \qquad \text{(Cauchy's remainder)}.$$

If $f(x,y)$ is m times continuously differentiable in a neighborhood of a point (x_0,y_0),

$$f(x_0+h,y_0+k)=\sum_{\lambda=0}^{m-1}\frac{1}{\lambda!}\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)^\lambda f(x_0,y_0)+R_m$$

$$=\sum_{0\leqslant\mu+\nu\leqslant m-1;\,\mu,\nu\geqslant0}\frac{1}{\mu!\nu!}h^\mu k^\nu\frac{\partial^{\mu+\nu}f(x_0,y_0)}{\partial x^\mu\partial y^\nu}+R_m,$$

$$R_m=\frac{1}{m!}\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)^m f(x_0+\theta h,y_0+\theta k)\quad (0<\theta<1).$$

If all partial derivatives up to order $m-1$ are totally differentiable,

$$f(x_1+h_1,\ldots,x_n+h_n)=\sum_{\nu=0}^{m-1}\frac{1}{\nu!}\left(\sum_{\mu=1}^{n}h_\mu\frac{\partial}{\partial x_\mu}\right)^\nu f(x_1,\ldots,x_n)+R_m$$

$$=\sum\frac{1}{\nu_1!\ldots\nu_n!}h_1^{\nu_1}\ldots h_n^{\nu_n}\frac{\partial^{\nu_1+\ldots+\nu_n}f(x_1,\ldots,x_n)}{\partial x_1^{\nu_1}\ldots\partial x_n^{\nu_n}}+R_m,$$

where the Σ means the sum for $\nu_1,\ldots,\nu_n$ in the domain $0\leqslant\nu_1+\ldots+\nu_n\leqslant m-1;\ \nu_1,\ldots,\nu_m\geqslant0$. The remainder R_m is expressed as

$$R_m=\frac{1}{m!}\left(\sum_{\mu=1}^{n}h_\mu\frac{\partial}{\partial x_\mu}\right)^m f(x_1+\theta h_1,\ldots,x_n+\theta h_n)\quad (0<\theta<1).$$

(V) Definite Integrals [4]

In the following formulas, we assume that m, n are positive integers. δ_{mn} is Kronecker's delta ($\delta_{mn}=0$ or 1 for $m\neq n$ or $m=n$), Γ is the gamma function, B is the beta function, and C is the Euler constant.

For simplicity, we put

$$m!!\equiv\begin{cases}1\cdot3\cdot5\cdot\ldots\cdot(m-2)\cdot m=2^{(m+1)/2}\Gamma[(m/2)+1]/\sqrt{\pi}=m!/2^{(m-1)/2}[(m-1)/2]! & (m \text{ is odd}),\\ 2\cdot4\cdot6\cdot\ldots\cdot(m-2)\cdot m=2^{m/2}\Gamma[(m/2)+1]=2^{m/2}(m/2)! & (m \text{ is even}).\end{cases}$$

In an n-dimensional real space, the volume of the domain

$$|x_1|^p+\ldots+|x_n|^p\leqslant 1 \quad (p>0) \qquad \text{is} \qquad \frac{2^n[\Gamma(1/p)]^n}{p^{n-1}n\Gamma(n/p)}.$$

For $p=2$, this is the volume of the unit hypersphere, which is

$$\frac{\pi^{n/2}}{\Gamma[(n/2)+1]}=\begin{cases}(2\pi)^{n/2}/n!! & (n \text{ is even}),\\ 2(2\pi)^{(n-1)/2}/n!! & (n \text{ is odd}).\end{cases}$$

The surface area of the $(n-1)$-dimensional unit hypersphere

$$|x_1|^2+\ldots+|x_n|^2=1 \qquad \text{is} \qquad \frac{2\pi^{n/2}}{\Gamma(n/2)}=\begin{cases}(2\pi)^{n/2}/(n-2)!! & (n \text{ is even}),\\ 2(2\pi)^{(n-1)/2}/(n-2)!! & (n \text{ is odd}).\end{cases}$$

$$\int_0^\infty x^{p-1}e^{-x}dx=\int_0^1\left(\log\frac{1}{x}\right)^{p-1}dx\equiv\Gamma(p).$$

$$\int_0^1 x^p\left(\log\frac{1}{x}\right)^q dx=\frac{\Gamma(q+1)}{(p+1)^{q+1}} \quad (\mathrm{Re}\,p,\mathrm{Re}\,q>-1).$$

$$\int_0^1 x^{p-1}(1-x)^{q-1}dx=\int_0^\infty\frac{x^{p-1}}{(1+x)^{p+q}}dx\equiv B(p,q)=\frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}.$$

$$\int_0^\infty\frac{x^a}{(1+x^c)^{1+b}}dx=\frac{1}{c}\,\frac{\Gamma[(a+1)/c]\Gamma[b-\{(a-c+1)/c\}]}{\Gamma(1+b)}$$

$$\left(\mathrm{Re}\,c>0;\mathrm{Re}\,a,\mathrm{Re}\,b>-1;\mathrm{Re}\,b>\mathrm{Re}\frac{a-c+1}{c}\right).$$

$$\int_0^a\frac{(a-x)^{p-1}(x-b)^{q-1}}{|x-c|^{p+q}}dx=\frac{(a-b)^{p+q-1}}{|a-c|^q|b-c|^p}B(p,q)$$

$$(0<c<b<a \quad\text{or}\quad 0<b<a<c;\mathrm{Re}\,p,\mathrm{Re}\,q>0).$$

$$\int_{-\infty}^\infty\frac{1}{(1+x^2)^{n+1}}dx=\frac{\pi(2n)!}{2^{2n}(n!)^2}=\pi\frac{(2n-1)!!}{(2n)!!}.$$

$$\int_{-\infty}^\infty\frac{x^{2m}}{1+x^{2n}}dx=\frac{\pi}{n\sin[(2m+1)\pi/2n]} \quad (2m+1<2n).$$

$$\int_0^\infty\frac{x^{a-1}}{1+x}dx=\Gamma(a)\Gamma(1-a)=\frac{\pi}{\sin a\pi} \quad (0<a<1).$$

$$\int_0^\infty e^{-a^2x^2}dx=\frac{\sqrt{\pi}}{2|a|}.\qquad \int_0^\infty x^pe^{-x^2}dx=\frac{1}{2}\Gamma\left(\frac{p+1}{2}\right)=\begin{cases}(2n-1)!!\sqrt{\pi}/2^{n+1} & (p=2n),\\ n!/2 & (p=2n+1).\end{cases}$$

$$\int_0^\infty(e^{-a^2/x^2}-e^{-b^2/x^2})dx=(b-a)\sqrt{\pi} \quad (a,b\geqslant0).\qquad \int_0^\infty e^{-[x-(1/x)]^2}dx=\frac{\sqrt{\pi}}{2}.$$

$$\int_0^\infty e^{-x^2-(a^2/x^2)}dx=\frac{e^{-2a}\sqrt{\pi}}{2} \quad (a\geqslant0).\qquad \int_0^\infty\frac{1}{e^{ax}+e^{-ax}}dx=\frac{\pi}{4a}.$$

$$\int_0^\infty\frac{x}{e^{ax}-e^{-ax}}dx=\frac{\pi^2}{8a^2} \quad (a>0).$$

$$\int_0^\infty\frac{x}{e^x-e^{-x}}dx=\int_0^1\frac{\log(1/x)}{1-x^2}dx=\frac{\pi^2}{8}.$$

$$\int_0^\infty \frac{x}{e^x+1}\,dx = \int_0^1 \frac{\log(1/x)}{1+x}\,dx = \frac{\pi^2}{12}. \quad \int_0^\infty \frac{x}{e^x-1}\,dx = \int_0^1 \frac{\log(1/x)}{1-x}\,dx = \frac{\pi^2}{6}.$$

$$\int_0^\infty \log\left(\frac{e^x+1}{e^x-1}\right)dx = \int_0^1 \log\left(\frac{1+x}{1-x}\right)\frac{1}{x}\,dx = \frac{\pi^2}{4}.$$

$$\int_0^1 \frac{\log x}{\sqrt{1-x^2}}\,dx = \int_0^{\pi/2} \log\sin x\,dx = -\frac{\pi}{2}\log 2.$$

$$\int_0^1 \frac{\log(1+x)}{1+x^2}\,dx = \frac{\pi}{8}\log 2. \quad \int_0^1 \frac{\log x}{1+x^2}\,dx = -\int_1^\infty \frac{\log x}{1+x^2}\,dx = \sum_{n=0}^\infty \frac{(-1)^{n-1}}{(2n+1)^2} = -0.91596\ldots.$$

$$\int_0^\infty \frac{(\log x)^2}{1+x+x^2}\,dx = \frac{16\pi^3}{81\sqrt{3}}. \quad \int_0^1 \frac{x^p-x^q}{\log x}\,dx = \log\frac{p+1}{q+1} \quad (p,q>-1).$$

$$\int_0^1 \log|\log x|\,dx = \int_0^\infty e^{-t}\log t\,dt = -C = -0.57721\ldots.$$

$$\int_0^\pi \sin mx\sin nx\,dx = \int_0^\pi \cos mx\cos nx\,dx = \delta_{mn}\frac{\pi}{2}.$$

$$\int_0^\pi \sin mx\cos nx\,dx = \begin{cases} \left[1-(-1)^{m+n}\right]\dfrac{m}{m^2-n^2} & (m\neq n); \\ 0 & (m=n). \end{cases}$$

$$\int_0^{\pi/2} \sin^p x\cos^q x\,dx = \frac{1}{2}B\left(\frac{p+1}{2},\frac{q+1}{2}\right) \quad \left(\operatorname{Re}p, \operatorname{Re}q > -\frac{1}{2}\right);$$

$$= \begin{cases} (\pi/2)(p-1)!!(q-1)!!/(p+q)!! & (p,q \text{ are even positive integers}), \\ (p-1)!!(q-1)!!/(p+q)!! & (p,q \text{ are positive integers not both even}). \end{cases}$$

$$\int_0^{\pi/2} \sin^p x\,dx = \int_0^{\pi/2} \cos^p x\,dx = \frac{\sqrt{\pi}}{2}\frac{\Gamma[(p+1)/2]}{\Gamma[(p/2)+1]} \quad (\operatorname{Re}p > -1);$$

$$= \begin{cases} (\pi/2)(2n-1)!!/(2n)!! & (p=2n), \\ (2n)!!/(2n+1)!! & (p=2n+1). \end{cases}$$

$$\int_{-\infty}^\infty \sin(x^2)\,dx = \int_{-\infty}^\infty \cos(x^2)\,dx = \int_0^\infty \frac{\sin x}{\sqrt{x}}\,dx = \int_0^\infty \frac{\cos x}{\sqrt{x}}\,dx = \sqrt{\frac{\pi}{2}} \quad \text{(Fresnel integral)}.$$

$$\int_0^\infty \frac{\sin ax}{x}\,dx = \frac{\pi}{2} \quad (a>0). \quad \int_0^\infty \frac{\tan x}{x}\,dx = \frac{\pi}{2}$$

$$\left(\text{take Cauchy's principal value at } x = \left(n+\frac{1}{2}\right)\pi\right).$$

$$\int_0^\infty \frac{\sin^{2n+1}x}{x}\,dx = \frac{\pi}{2}\frac{(2n-1)!!}{(2n)!!}. \quad \int_0^\infty \frac{\sin^2 x}{x^2}\,dx = \frac{\pi}{2}. \quad \int_0^\infty \frac{\sin(x^2)}{x}\,dx = \frac{\pi}{4}.$$

$$\int_0^\infty \frac{\sin qx}{x^p}\,dx = \frac{\pi q^{p-1}}{2\Gamma(p)\sin(p\pi/2)} \quad (0<p<2).$$

$$\int_0^\infty \frac{\cos px}{1+x^2}\,dx = \frac{\pi}{2}e^{-|p|}. \quad \int_0^\infty \frac{\cos^2 ax}{1+x^2}\,dx = \frac{\pi}{4}(1+e^{-2a}) \quad (a>0).$$

$$\int_0^\infty \frac{\sin ax}{x(1+x^2)}\,dx = \frac{\pi}{2}(1-e^{-a}) \quad (a>0). \quad \int_0^\infty \frac{x\sin ax}{1+x^2}\,dx = \frac{\pi}{2}e^{-a} \quad (a>0).$$

$$\int_0^\infty \frac{\sin^{2m+1}x\cos^{2n}x}{x}\,dx = \int_0^\infty \frac{\sin^{2m+1}x\cos^{2n-1}x}{x}\,dx = \frac{\pi}{2}\frac{(2m-1)!!(2n-1)!!}{(2m+2n)!!}.$$

$$\int_0^\infty \frac{\sin ax\cos bx}{x}\,dx = \begin{cases} \pi/2 & (a>b>0), \\ \pi/4 & (a=b>0), \\ 0 & (b>a>0) \end{cases} \quad \text{(Dirichlet's discontinuous factor)}.$$

$$\int_0^{2\pi} \frac{1}{1+a\cos x}\,dx = \frac{2\pi}{\sqrt{1-a^2}} \quad (|a|<1). \quad \int_0^{\pi/2} \frac{1}{a^2\cos^2 x+b^2\sin^2 x}\,dx = \frac{\pi}{2ab} \quad (ab\neq 0).$$

$$\int_0^\pi \frac{x\sin x}{1+\cos^2 x}\,dx = \frac{\pi^2}{4}.$$

$$\int_0^{\pi}\frac{\cos nx}{1-2a\cos x+a^2}\,dx=\begin{cases}\pi a^n/(1-a^2) & (|a|<1),\\ \pi/a^n(a^2-1) & (|a|>1).\end{cases}$$

References

[1] B. O. Peirce, A short table of integrals, Ginn, Boston, second revised edition, 1910.
[2] D. Bierens de Haan, Nouvelles tables d'intégrales définies, Leiden, 1867.
There are several mistakes in this table. For the errata, see
[3] C. F. Lindmann, Examen des nouvelles tables de M. Bierens de Haan, Handlingar Svenska Vetenskaps-Akad., 1891.
[4] E. W. Sheldon, Critical revision of de Haan's tables of definite integrals, Amer. J. Math., 34 (1912), 39–114.

10. Series (→ 379 Series)

(I) Finite Series

(1) $S_k \equiv 1^k+2^k+\ldots+n^k$ (k is an integer). For $k\geqslant 0$, we have

$$S_k=\frac{B_{k+1}(n+1)-B_{k+1}(1)}{k+1}=\sum_{i=0}^{k}(-1)^i\binom{k+1}{i}\frac{B_{2i}(n+1)^{k+1-i}}{k+1},$$

where B_l is a Bernoulli number and $B_l(x)$ is a Bernoulli polynomial. In particular,

$$S_0=n,\quad S_1=n(n+1)/2,\quad S_2=n(n+1)(2n+1)/6,\quad S_3=n^2(n+1)^2/4,$$
$$S_4=n(n+1)(2n+1)(3n^2+3n-1)/30.$$

For $k<0$ and $k=-l$,

$$S_{-l}=c_l-\left[(-1)^l/(l-1)!\right]\left[d^l\log\Gamma(x)/dx^l\right]_{x=n+1}$$
$$=c_l-\frac{1}{(l-1)(n+1)^{l-1}}-\frac{1}{2(n+1)^l}+\sum_{i=1}^{\infty}(-1)^i\frac{B_{2(i+1)}}{(i+1)!}\frac{(l+i-1)!}{(l-1)!}\frac{1}{(n+1)^{l+i}}.$$

For $l=1$, the second term in the latter formula is replaced by $\log[1/(n+1)]$. Here Γ is the gamma function, and the constants c_l are

$$c_l=\begin{cases}C \quad (\text{Euler constant}) \quad (l=1),\\ \zeta(l) \quad (\zeta \text{ is the Riemann zeta function}) \quad (l\geqslant 2).\end{cases}$$

(2) $$\sum_{i=1}^{n} i(i+1)\ldots(i+m-1)\equiv\sum_{i=1}^{n}\frac{(i+m-1)!}{(i-1)!}=\frac{1}{m+1}\frac{(n+m)!}{(n-1)!},$$

$$\sum_{i=1}^{n}\frac{(i-1)!}{(i+m-1)!}=\frac{1}{m-1}\left[\frac{1}{(m-1)!}-\frac{n!}{(n+m-1)!}\right] \quad (m\geqslant 2),$$

$$\sum_{i=1}^{n} i!i=(n+1)!-1,\quad \sum_{i=1}^{n} i\binom{n}{i}=n2^{n-1},$$

$$\sum_{i=m}^{n}\binom{i}{m}\binom{n+s-i-1}{n-i}=\binom{n+s}{m+s} \quad (m\leqslant n),$$

$$\sum_{i=0}^{n}\binom{n}{i}\binom{m}{r-i}=\binom{n+m}{r}.$$

$$\sum_{i=1}^{n} a^i=\begin{cases}a(a^n-1)/(a-1) & (a\neq 1)\\ n & (a=1)\end{cases} \quad \text{(geometric progression)}$$

$$\sum_{j=0}^{n}(a+jd)=(n+1)a+\frac{n(n+1)}{2}d=\frac{n+1}{2}(a+a+nd) \quad \text{(arithmetic progression)}$$

$$\sum_{j=0}^{n}\sin(\alpha+j\beta)=\sin\left(\alpha+\frac{n}{2}\beta\right)\sin\frac{(n+1)\beta}{2}\Big/\sin\frac{\beta}{2},$$

$$\sum_{j=0}^{n} \cos(\alpha+j\beta)=\cos\left(\alpha+\frac{n}{2}\beta\right)\sin\frac{(n+1)\beta}{2} \Big/ \sin\frac{\beta}{2},$$

$$\sum_{j=0}^{n} \operatorname{cosec} 2^j\alpha=\cot(\alpha/2)-\cot 2^n\alpha.$$

(II) Convergence Criteria for Positive Series $\sum a_n$

In the present Section II, we assume that $a_n \geqslant 0$.
Cauchy's criterion: The series converges when $\limsup \sqrt[n]{a_n}<1$ and it diverges when $\limsup \sqrt[n]{a_n}>1$.
d'Alembert's criterion: The series converges when $\limsup a_{n+1}/a_n<1$ and diverges when $\liminf a_{n+1}/a_n>1$.
Raabe's criterion: The series converges when $\liminf n[(a_n/a_{n+1})-1]>1$ and diverges when $\limsup n[(a_n/a_{n+1})-1]<1$.
Kummer's criterion: For a positive divergent series $\sum(1/b_n)$, the series $\sum a_n$ converges when $\liminf[(b_n a_n/a_{n+1})-b_{n+1}]>0$ and diverges when $\limsup[(b_n a_n/a_{n+1})-b_{n+1}]<0$ diverges.
Gauss's criterion: Suppose $a_n/a_{n+1}=1+(k/n)+(\theta_n/n^{\lambda})$, where $\lambda>1$ and $\{\theta_n\}$ is bounded. Then the series $\sum a_n$ converges when $k>1$; and diverges when $k\leqslant 1$.
Schlömilch's criterion: For a decreasing positive sequence $a_\nu \downarrow 0$, let n_ν be an increasing sequence of positive integers and suppose that $(n_{\nu+2}-n_{\nu+1})/(n_{\nu+1}-n_\nu)$ is bounded. Then the two series $\sum a_n$ and $\sum(n_{\nu+1}-n_\nu)a_{n_\nu}$ converge or diverge simultaneously.
Logarithmic criterion: For a positive integer k, we put

$$\log_k x \equiv \log(\log_{k-1} x), \quad \log_1 x=\log x.$$

Then for sufficiently large n we have
The first logarithmic criterion: If

$$a_n-1/(n\log_1 n \ldots \log_{k-1} n(\log_k n)^p)\begin{cases}\leqslant 0, & p>1 \quad \text{then } \sum a_n \text{ converges},\\ \geqslant 0, & p\leqslant 1 \quad \text{then } \sum a_n \text{ diverges}.\end{cases}$$

The second logarithmic criterion: If

$$\frac{a_{n+1}}{a_n}-\frac{n}{n+1}\frac{\log_1 n}{\log_1(n+1)}\cdots\frac{\log_{k-1} n}{\log_{k-1}(n+1)}\left(\frac{\log_k n}{\log_k(n+1)}\right)^p$$

$$\begin{cases}\leqslant 0, & p>1 \quad \text{then } \sum a_n \text{ converges},\\ \geqslant 0, & p\leqslant 1 \quad \text{then } \sum a_n \text{ diverges}.\end{cases}$$

(III) Infinite Series

$$\sum_{i=1}^{\infty}\frac{(-1)^{i-1}}{i}=\log 2, \quad \sum_{i=1}^{\infty}\frac{(-1)^{i-1}}{2i-1}=\frac{\pi}{4} \qquad \text{(Leibniz's formula)},$$

$$\sum_{i=1}^{\infty}\frac{(-1)^{i-1}}{2i(2i+1)(2i+2)}=\frac{\pi-3}{4},$$

$$\sum_{i=0}^{\infty}\frac{(2i)!}{2^{2i}(i!)^2}\frac{1}{2i+1}=\frac{\pi}{2}, \quad \sum_{i=1}^{\infty}\left(\frac{1}{i}-\log\left(1+\frac{1}{i}\right)\right)=C \quad (C \text{ is Euler's constant}).$$

Putting

$$\zeta(n)=\sum_{i=1}^{\infty}\frac{1}{i^n}, \quad \alpha(n)=\sum_{i=1}^{\infty}\frac{1}{(2i-1)^n}, \quad \beta(n)=\sum_{i=1}^{\infty}\frac{(-1)^{i-1}}{i^n}, \quad \varepsilon(n)=\sum_{i=1}^{\infty}\frac{(-1)^{i-1}}{(2i-1)^n},$$

we have

$$\zeta(2n)=\frac{(2\pi)^{2n}}{2(2n)!}B_{2n},\quad \alpha(2n)=\frac{(2^{2n}-1)\pi^{2n}}{2(2n)!}B_{2n},$$

$$\beta(2n)=\frac{(2^{2n-1}-1)\pi^{2n}}{(2n)!}B_{2n},\quad \varepsilon(2n+1)=\frac{\pi^{2n+1}}{2^{2n+2}(2n)!}E_{2n}.$$

where B_n is a Bernoulli number, and E_n is an Euler number.

$\zeta(2)=\pi^2/6,\quad \zeta(4)=\pi^4/90,\quad \zeta(6)=\pi^6/945.$

$\alpha(2)=\pi^2/8,\quad \alpha(4)=\pi^4/96,\quad \alpha(6)=\pi^6/960.$

$\beta(2)=\pi^2/12,\quad \beta(4)=7\pi^4/720,\quad \beta(6)=31\pi^6/30240.$

$\varepsilon(1)=\pi/4,\quad \varepsilon(2)=0.91596\,55941\,77219\,01505\,46035\,14932\ldots$ (Catalan's constant),

$\varepsilon(3)=\pi^3/32,\quad \varepsilon(5)=5\pi^5/1536,\quad \varepsilon(7)=61\pi^7/92160.$

(IV) Power Series (→ 339 Power Series)

(1) Binomial Series $(1+x)^\alpha=\sum_{i=0}^{\infty}\binom{\alpha}{i}x^i$. This converges always in $|x|<1$. If $\alpha>0$, it converges in $-1\leqslant x\leqslant 1$, and if $-1<\alpha<0$, it converges in $-1<x\leqslant 1$. When α is 0 or a positive integer, it reduces to a polynomial and converges in $|x|<\infty$.

$$\frac{1}{1+x}=\sum_{i=0}^{\infty}(-1)^i x^i\quad (|x|<1),$$

$$\sqrt{1+x}=\sum_{i=0}^{\infty}\frac{(-1)^{i-1}(2i)!}{(2i-1)2^{2i}(i!)^2}x^i\quad (|x|<1),\qquad \frac{1}{\sqrt{1+x}}=\sum_{i=0}^{\infty}\frac{(-1)^i(2i)!}{2^{2i}(i!)^2}x^i\quad (|x|<1).$$

(2) Elementary Transcendental Functions (→ 131 Elementary Functions).

$$e^x\equiv\exp x=\sum_{i=0}^{\infty}\frac{x^i}{i!}=\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n,\quad a^x=\exp(x\log a)\quad (|x|<\infty).$$

$$\log(1+x)=\sum_{i=1}^{\infty}\frac{(-1)^{i-1}}{i}x^i\quad (-1<x\leqslant 1),\quad \log x=2\sum_{i=0}^{\infty}\frac{1}{2i+1}\left(\frac{x-1}{x+1}\right)^{2i+1}\quad (0<x<\infty).$$

$$\sin x=\sum_{i=0}^{\infty}\frac{(-1)^i}{(2i+1)!}x^{2i+1},\quad \cos x=\sum_{i=0}^{\infty}\frac{(-1)^i}{(2i)!}x^{2i}\quad (|x|<\infty).$$

$$\tan x=\sum_{i=1}^{\infty}\frac{2^{2i}(2^{2i}-1)B_{2i}}{(2i)!}x^{2i-1}\quad \left(|x|<\frac{\pi}{2}\right)\quad (B_i \text{ is a Bernoulli number}),$$

$$\cot x=\frac{1}{x}-\sum_{i=1}^{\infty}\frac{2^{2i}B_{2i}}{(2i)!}x^{2i-1}\quad \left(0<|x|<\frac{\pi}{2}\right),$$

$$\sec x=\sum_{i=0}^{\infty}\frac{E_{2i}}{(2i)!}x^{2i}\quad \left(|x|<\frac{\pi}{2}\right)\quad (E_i \text{ is an Euler number}),$$

$$\operatorname{cosec} x=\frac{1}{x}+\sum_{i=1}^{\infty}\frac{(2^{2i}-2)B_{2i}}{(2i)!}x^{2i-1}\quad (0<|x|<\pi).$$

$$\arcsin x=\sum_{i=0}^{\infty}\frac{(2i)!}{2^{2i}(i!)^2}\frac{x^{2i+1}}{2i+1}\quad (|x|\leqslant 1),\quad \arctan x=\sum_{i=0}^{\infty}\frac{(-1)^i}{2i+1}x^{2i+1}\quad (|x|\leqslant 1).$$

(V) Partial Fractions for Elementary Functions

$$\tan x = \sum_{n=0}^{\infty} \frac{8x}{(2n+1)^2\pi^2-4x^2}, \quad \cot x = \frac{1}{x} + 2x \sum_{n=1}^{\infty} \frac{1}{x^2-n^2\pi^2},$$

$$\sec x = 4 \sum_{n=1}^{\infty} \frac{(-1)^n(2n-1)\pi}{(2n-1)^2\pi^2-4x^2}, \quad \operatorname{cosec} x = \frac{1}{x} + 2x \sum_{n=1}^{\infty} \frac{(-1)^n}{x^2-n^2\pi^2},$$

$$\sec^2 x = \sum_{n=-\infty}^{\infty} \frac{1}{\left[x+\{(2n+1)\pi/2\}\right]^2}, \quad \operatorname{cosec}^2 x = \sum_{n=-\infty}^{\infty} \frac{1}{(x+n\pi)^2}.$$

(VI) Infinite Products (→ 379 Series F)

$$\prod_{n=1}^{\infty} \frac{4n^2}{4n^2-1} = \frac{\pi}{2} \quad \text{(Wallis formula)}, \quad \prod_{n=1}^{\infty} \left(1+\frac{x}{a+n}\right)e^{-x/n} = e^{-Cx}\frac{\Gamma(1+a)}{\Gamma(1+a+x)}$$

(C is Euler's constant).

$$\prod_{n=1}^{\infty} \left(1-\frac{x}{2n-1}\right)\left(1+\frac{x}{2n}\right) = \sqrt{\pi} \Big/ \Gamma\left(1+\frac{x}{2}\right)\Gamma\left(\frac{1}{2}-\frac{x}{2}\right).$$

$$\prod_p 1/(1-p^{-s}) = \zeta(s) \quad (p \text{ ranges over all prime numbers}, s>1),$$

$$\prod_{n=1}^{\infty} \left(1-\frac{x^2}{n^2\pi^2}\right) = \frac{\sin x}{x}, \quad \prod_{n=1}^{\infty} \cos\frac{x}{2^n} = \frac{\sin x}{x}, \quad \prod_{n=1}^{\infty} \left(1-\frac{4x^2}{(2n-1)^2\pi^2}\right) = \cos x.$$

For $|q|<1$, putting $q_1 \equiv \prod_{n=1}^{\infty}(1+q^{2n})$, $q_2 \equiv \prod_{n=1}^{\infty}(1+q^{2n-1})$, $q_3 \equiv \prod_{n=1}^{\infty}(1-q^{2n-1})$,

$q_4 \equiv \prod_{n=1}^{\infty}(1-q^{2n})$ we have $q_1q_2q_3=1$.

Further, putting $q=e^{i\pi\tau}$, we have the following formulas concerning ϑ-functions (→134 Elliptic Functions):

$$\vartheta_4(0,\tau)=q_4q_3^2, \quad \vartheta_2(0,\tau)=2q^{1/4}q_4q_1^2, \quad \vartheta_3(0,\tau)=q_4q_2^2, \quad \vartheta_1'(0,\tau)=2\pi q^{1/4}q_4^3.$$

11. Fourier Analysis

(I) Fourier Series (→ 159 Fourier Series)

(1) Fourier coefficients $\quad a_0=\frac{1}{a}\int_0^a f(x)\,dx, \quad a_n=\frac{2}{a}\int_0^a f(x)\cos\frac{n\pi x}{a}\,dx$

$$b_n=\frac{2}{a}\int_0^a f(x)\sin\frac{n\pi x}{a}\,dx.$$

Fourier cosine series $\quad a_0+\sum_{n=1}^{\infty} a_n\cos\frac{n\pi x}{a} = \begin{cases} f(x) & (0<x<a), \\ f(-x) & (-a<x<0). \end{cases}$

Fourier sine series $\quad \sum_{n=1}^{\infty} b_n\sin\frac{n\pi x}{a} = \begin{cases} f(x) & (0<x<a), \\ -f(-x) & (-a<x<0). \end{cases}$

The next table shows the Fourier coefficients of the functions $F(x)$ directly in the following manner from a given function $f(x)$ on the interval $[0,a]$. For x in $[-a,0]$ and when the cosine series $\{a_n\}$ is in question, we set $f(x)=f(-x)$, and when the sine series $\{b_n\}$ is in question we set $f(x)=-f(-x)$. Thus $f(x)$ is extended in two ways to functions on $[-a,a]$. The functions $F(x)$ are the periodic continuations of such functions. We remark that the sum of the Fourier series given by the Fourier coefficients in the right hand side has, in general, some singularities (discontinuity of the function or its higher derivatives, for example) at the points given by the integral multiples of a. We assume that μ is not an integer.

$f(x)$	a_0	$a_n \quad (n=1,2,\ldots)$	$b_n \quad (n=1,2,\ldots)$
1	1	0	$[1+(-1)^{n+1}]2a/n\pi$
x	$\dfrac{a}{2}$	$[1+(-1)^{n+1}]\dfrac{-2a}{n^2\pi^2}$	$(-1)^{n+1}\dfrac{2a}{n\pi}$
x^2	$\dfrac{a^2}{3}$	$(-1)^n\dfrac{4a^2}{n^2\pi^2}$	$(-1)^{n-1}\dfrac{2a^2}{n\pi}-[1+(-1)^{n+1}]\dfrac{4a^2}{n^3\pi^3}$
e^{kx}	$\dfrac{e^{ka}-1}{ka}$	$\dfrac{2ka\left[(-1)^n e^{ka}-1\right]}{k^2a^2+n^2\pi^2}$	$\dfrac{2n\pi\left[1-(-1)^n\right]e^{ka}}{k^2a^2+n^2\pi^2}$
$\cos\dfrac{\mu\pi x}{a}$	$\dfrac{\sin\mu\pi}{\mu\pi}$	$(-1)^n\dfrac{2}{\pi}\dfrac{\mu\sin\mu\pi}{\mu^2-n^2}$	$\dfrac{2}{\pi}\dfrac{\left[(-1)^n\cos\mu\pi-1\right]}{\mu^2-n^2}$
$\sin\dfrac{\mu\pi x}{a}$	$\dfrac{1-\cos\mu\pi}{\mu\pi}$	$\dfrac{2}{\pi}\dfrac{\mu\left[1-(-1)^n\cos\mu\pi\right]}{\mu^2-n^2}$	$(-1)^n\dfrac{2}{\pi}\dfrac{n\sin\mu\pi}{\mu^2-n^2}$
$\dfrac{1-\lambda^2}{1-2\lambda\cos(\pi x/a)+\lambda^2}$	1	$2\lambda^n \quad (\lvert\lambda\rvert<1)$	
$\dfrac{\lambda\sin(\pi x/a)}{1-2\lambda\cos(\pi x/a)+\lambda^2}$			$\lambda^n \quad (\lvert\lambda\rvert<1)$
$B_{2m}(x/2a)$	0	$(-1)^{m+1}2(2m)!/(2n\pi)^{2m}$	
$B_{2m+1}(x/a)$			$(-1)^{m+1}2(2m+1)!/(2n\pi)^{2m+1}$
$\log\sin(\pi x/2a)$	$-\log 2$	$-1/n$	
$(1/2)\cot(\pi x/2a)$			1[(1)]

Note

(1)The Fourier series does not converge in the sense of Cauchy, but it is summable, for example, by the Cesàro summation of the first order.

(2) $\displaystyle\sum_{n=1}^{\infty}(-1)^{n-1}\frac{\cos nx}{n}=\log\left(2\cos\frac{x}{2}\right) \quad (-\pi<x<\pi), \qquad \sum_{n=1}^{\infty}\frac{\sin nx}{n}=\frac{1}{2}(\pi-x) \quad (0<x<2\pi).$

$$\sum_{n=1}^{\infty}\frac{\cos(2n-1)x}{2n-1}=\frac{1}{2}\log\left|\cot\frac{x}{2}\right| \quad (0<x<2\pi,\ \ x\neq\pi),$$

$$\sum_{n=1}^{\infty}\frac{\sin(2n-1)x}{2n-1}=\begin{cases}\pi/4 & (0<x<\pi),\\ -\pi/4 & (\pi<x<2\pi).\end{cases}$$

$$\sum_{n=1}^{\infty}\frac{\cos nx}{n^2}=\frac{1}{4}(x-\pi)^2-\frac{\pi^2}{12} \quad (0\leqslant x\leqslant 2\pi),$$

$$\sum_{n=1}^{\infty}\frac{\sin nx}{n^2}=-x\log 2-\int_0^x\log\left(\sin\frac{t}{2}\right)dt \quad (0\leqslant x<2\pi).$$

$$\sum_{n=1}^{\infty}\frac{a^n}{n!}\cos nx=e^{a\cos x}\cos(a\sin x)-1, \qquad \sum_{n=1}^{\infty}\frac{a^n}{n!}\sin nx=e^{a\cos x}\sin(a\sin x).$$

$$\sum_{n=1}^{\infty}(-1)^{n-1}\frac{\cos nx}{n^2-a^2}=\frac{\pi\cos ax}{2a\sin a\pi}-\frac{1}{2a^2} \quad (-\pi\leqslant x\leqslant\pi),$$

$$\sum_{n=1}^{\infty}(-1)^{n-1}\frac{n\sin nx}{n^2-a^2}=\frac{\pi\sin ax}{2\sin a\pi} \quad (-\pi<x<\pi).$$

In the final two formulas, we assume that a is not an integer.

(II) Fourier Transforms (→ 160 Fourier Transform)

The Fourier transform $\mathscr{F}[f]$ and the inverse Fourier transform $\bar{\mathscr{F}}[g]$ for integrable functions f and g are defined as

$$\mathscr{F}[f(x)]=\mathscr{F}[f](\xi)=(2\pi)^{-n/2}\int_{\mathbf{R}^n}f(x)e^{-ix\xi}dx,$$

$$\bar{\mathscr{F}}[g(\xi)]=\bar{\mathscr{F}}[g](x)=(2\pi)^{-n/2}\int_{\mathbf{R}^n}g(\xi)e^{ix\xi}d\xi, \qquad x\xi=x_1\xi_1+x_2\xi_2+\ldots+x_n\xi_n.$$

In some textbooks the factor $(2\pi)^{-n/2}$ is deleted or the symbols i and $-i$ are switched when defining $\mathscr{F}$ and $\bar{\mathscr{F}}$. However, conversion of the formulas above to ones due to other definitions is straightforward. These transforms are also defined for some nonintegrable functions, or even more generally for †tempered distributions. The Fourier transform $\mathscr{F}$ and the inverse Fourier transform $\bar{\mathscr{F}}$ defined on the space of tempered distributions $\mathscr{S}'=\mathscr{S}'(\mathbf{R}^n)$ are linear homeomorphic mappings from $\mathscr{S}'$ to itself. Useful formulas of these transforms are given in the table below, where $\alpha=(\alpha_1,\alpha_2,\ldots,\alpha_n)$ $(\alpha_j=0,1,2,\ldots)$, $|\alpha|=\alpha_1+\alpha_2+\ldots+\alpha_n$, C is †Euler's constant, $\lambda\in\mathbf{C}$, and $\mathbf{Z}_+=\{m\in\mathbf{Z}\mid m\geqslant 0\}$.

Case 1. $n=1$. First we explain the meaning of the symbols appearing in the table:

$x_+=\max(x,0)$ (the positive part of x),

$x_-=\max(-x,0)$ (the negative part of x),

x_+^λ and x_-^λ are understood in the sense of finite parts (→ 125 Distributions and Hyperfunctions),

$(x+i\varepsilon)^\lambda=\exp[\lambda\operatorname{Log}(x+i\varepsilon)]$ ($\varepsilon\neq 0$; Log is the principal value of log)

$=(x^2+\varepsilon^2)^{\lambda/2}\exp[i\lambda\operatorname{Arg}(x+i\varepsilon)]$ $(-\pi<\operatorname{Arg}z\leqslant\pi)$,

$(x\pm i0)^\lambda=\lim_{\varepsilon\downarrow 0}(x\pm i\varepsilon)^\lambda$ (limit in the sense of distributions).

Then the following formula holds:

$(x\pm i0)^\lambda=x_+^\lambda+e^{\pm i\lambda\pi}x_-^\lambda$

$\operatorname{Pf}x^m=x_+^m+(-1)^m x_-^m$ $(m\in Z)$ (Pf is the finite part).

In the special case $m=-1$, $\operatorname{Pf}x^{-1}$ coincides with Cauchy's principal value p.v.x^{-1}.

$T\,(\in\mathscr{S}')$	$\mathscr{F}[T]\,(\in\mathscr{S}')$
$\delta(x)$	$\sqrt{2\pi}$
$P(x)$ (polynomial)	$\sqrt{2\pi}\,P(id/d\xi)\delta(\xi)$
p.v. $1/x$	$\sqrt{\pi/2}\,i\operatorname{sgn}\xi$
$\operatorname{Pf}x^{-m}$	$\sqrt{\pi/2}\,[(-i)^m/(m-1)!]\xi^{m-1}\operatorname{sgn}\xi$ $(m\in\mathbf{N},\mathbf{Z})$
x_+^λ	$\dfrac{\Gamma(\lambda+1)}{\sqrt{2\pi}}[e^{-i\pi(\lambda+1)/2}\xi_+^{-\lambda-1}+e^{i\pi(\lambda+1)/2}\xi_-^{-\lambda-1}]$
	$\left[=\dfrac{\Gamma(\lambda+1)}{\sqrt{2\pi}}e^{-i\pi(\lambda+1)/2}(\xi+i0)^{-\lambda-1}\right]$ $(\lambda\notin\mathbf{Z})$
x_+^m	$(i^m/\sqrt{2\pi})[\pi\delta^{(m)}-i(-1)^m m!\operatorname{Pf}\xi^{-m-1}]$ $(m\in\mathbf{Z}_+)$
x_+^{-m}	$\dfrac{(-i)^{m-1}}{\sqrt{2\pi}(m-1)!}\left[\left(\sum_{j=1}^{m-1}\dfrac{1}{j}-C\right)\xi^{m-1}-\dfrac{i\pi}{2}\xi^{m-1}\operatorname{sgn}\xi-\xi^{m-1}\log\lvert\xi\rvert\right]$ $(m\in\mathbf{N},\mathbf{Z})$
x_-^λ	$\dfrac{\Gamma(\lambda+1)}{\sqrt{2\pi}}[e^{i\pi(\lambda+1)/2}\xi_+^{-\lambda-1}+e^{-i\pi(\lambda+1)/2}\xi_-^{-\lambda-1}]$
	$\left[=\dfrac{\Gamma(\lambda+1)}{\sqrt{2\pi}}e^{i\pi(\lambda+1)/2}(\xi-i0)^{-\lambda-1}\right]$ $(\lambda\notin\mathbf{Z})$
x_-^m	$\dfrac{(-i)^m}{\sqrt{2\pi}}[\pi\delta^{(m)}+i(-1)^m m!\operatorname{Pf}\xi^{-m-1}]$ $(m\in\mathbf{Z}_+)$
x_-^{-m}	$\dfrac{i^{m-1}}{\sqrt{2\pi}(m-1)!}\left[\left(\sum_{j=1}^{m-1}\dfrac{1}{j}-C\right)\xi^{m-1}+\dfrac{i\pi}{2}\xi^{m-1}\operatorname{sgn}\xi-\xi^{m-1}\log\lvert\xi\rvert\right]$ $(m\in\mathbf{N},\mathbf{Z})$
$(x+i0)^\lambda$	$[\sqrt{2\pi}\,e^{i\pi\lambda/2}/\Gamma(-\lambda)]\xi_+^{-\lambda-1}$ $(\lambda\notin\mathbf{Z}_+)$
$(x-i0)^\lambda$	$[\sqrt{2\pi}\,e^{-i\pi\lambda/2}/\Gamma(-\lambda)]\xi_-^{-\lambda-1}$ $(\lambda\notin\mathbf{Z}_+)$
$(x\pm i0)^m=x^m$	$\sqrt{2\pi}\,i^m\delta^{(m)}$ $(m\in\mathbf{Z}_+)$
$x^{-1}\log\lvert x\rvert$	$\sqrt{\pi/2}\,i\operatorname{sgn}\xi\cdot(C+\log\lvert\xi\rvert)$
e^{-x^2}/α	$\sqrt{a/2}\,e^{-a\xi^2/4}$ $(a>0)$
$\begin{cases}e^{-ax} & (x>0)\\ 0 & (x\leqslant 0)\end{cases}$	$\dfrac{1}{\sqrt{2\pi}}\dfrac{1}{a+i\xi}$ $(a>0)$

$T\,(\in\mathcal{S}')$	$\mathcal{F}[T]\,(\in\mathcal{S}')$
$\dfrac{e^{-a\lvert x\rvert}}{\sqrt{\lvert x\rvert}}$	$\dfrac{\sqrt{a+\sqrt{a^2+\xi^2}}}{\sqrt{a^2+\xi^2}}\quad (a>0)$
$e^{-b\sqrt{x^2+a^2}}$	$\sqrt{\dfrac{2}{\pi}}\dfrac{ab}{\sqrt{\xi^2+b^2}}K_1(a\sqrt{\xi^2+b^2})\quad (a>0, b>0)$
$\log\dfrac{x^2+a^2}{x^2+b^2}$	$-\dfrac{\sqrt{2\pi}}{\lvert\xi\rvert}(e^{-a\lvert\xi\rvert}-e^{-b\lvert\xi\rvert})\quad (a\geqslant 0, b\geqslant 0)$
$\arctan(x/a)$	$-\sqrt{\pi/2}\, i\operatorname{sgn} a\cdot(e^{-\lvert a\rvert\lvert\xi\rvert}/\xi)\quad (a\in\mathbf{R}, a\neq 0)$
$\dfrac{\sin ax}{\lvert x\rvert^{1-\nu}}$	$-\dfrac{i}{\sqrt{2\pi}}\Gamma(\nu)\cos\dfrac{\nu\pi}{2}\left(\dfrac{1}{\lvert\xi-a\rvert^{\nu}}-\dfrac{1}{\lvert\xi+a\rvert^{\nu}}\right)\quad (\nu\notin\mathbf{Z})$
$\dfrac{\cos ax}{\lvert x\rvert^{1-\nu}}$	$\dfrac{1}{\sqrt{2\pi}}\Gamma(\nu)\cos\dfrac{\nu\pi}{2}\left(\dfrac{1}{\lvert\xi-a\rvert^{\nu}}-\dfrac{1}{\lvert\xi+a\rvert^{\nu}}\right)\quad (\nu\notin\mathbf{Z})$
$\dfrac{\sin ax}{x}$	$\begin{cases}\sqrt{\pi/2} & (\lvert\xi\rvert<\lvert a\rvert)\\ 0 & (\lvert\xi\rvert>\lvert a\rvert)\end{cases}$
$\begin{cases}1/(a^2-x^2)^{\nu+(1/2)} & (\lvert x\rvert<a)\\ 0 & (\lvert x\rvert>a)\end{cases}$	$\dfrac{1}{\sqrt{2}}\Gamma((1/2)-\nu)\left(\dfrac{\lvert\xi\rvert}{2a}\right)^{\nu}J_{-\nu}(a\lvert\xi\rvert)\quad (\operatorname{Re}\nu<1/2,\ a>0)$
$\begin{cases}0 & (\lvert x\rvert<a)\\ 1/(x^2-a^2)^{\nu+(1/2)} & (\lvert x\rvert>a)\end{cases}$	$-\dfrac{1}{\sqrt{2}}\Gamma((1/2)-\nu)\left(\dfrac{\lvert\xi\rvert}{2a}\right)^{\nu}N_{\nu}(a\lvert\xi\rvert)\quad (-1/2<\operatorname{Re}\nu<1/2,\ a>0)$
$\dfrac{1}{(x^2+a^2)^{\nu+(1/2)}}$	$\dfrac{\sqrt{2}}{\Gamma(\nu+(1/2))}\left(\dfrac{\lvert\xi\rvert}{2a}\right)^{\nu}K_{\nu}(a\lvert\xi\rvert)\quad (\operatorname{Re}\nu>-1/2,\ a>0)$

Case 2. $n>1$. Let $r=\sqrt{x_1^2+x_2^2+\ldots+x_n^2}$ and $\rho=\sqrt{\xi_1^2+\xi_2^2+\ldots+\xi_n^2}$, where $x=(x_i)\in\mathbf{R}^n$ and $\xi=(\xi_i)\in\mathbf{R}^n$. If $f\in L_1(\mathbf{R}^n)$ depends only on r, then $\mathcal{F}[f]$ depends only on ρ and is expressed as

$$\mathcal{F}[f](\rho)=\rho^{-(n-2)/2}\int_0^{\infty} f(r)r^{n/2}J_{(n-2)/2}(r\rho)\,dr.$$

The constant C in the table stands for Euler's number.

$T\,(\in\mathcal{S}')$	$\mathcal{F}[T]\,(\in\mathcal{S}')$
$\delta(x)$	$(2\pi)^{n/2}$
$P(x)$ (polynomial)	$(2\pi)^{n/2}P(i\partial/\partial\xi)\delta(\xi)$
$\operatorname{Pf} r^{\lambda}$	$\dfrac{2^{(n/2)+\lambda}\Gamma((n+\lambda)/2)}{\Gamma(-\lambda/2)}\operatorname{Pf}\rho^{-n-\lambda}\quad (\lambda\notin 2\mathbf{Z}_+,\ \ \lambda\notin -n-2\mathbf{Z}_+)$
r^{2m}	$(2\pi)^{n/2}(-\Delta)^m\delta(\xi)\quad (m\in\mathbf{Z}_+)$
$\operatorname{Pf} r^{-n-2m}$	$\dfrac{(-1)^m\rho^{2m}}{2^{(n/2)+2m}\Gamma((n/2)+m)m!}\left[2\log\dfrac{2}{\rho}-C+\sum_{j=1}^{m}\dfrac{1}{j}+\dfrac{\Gamma'((n/2)+m)}{\Gamma((n/2)+m)}\right]\quad (m\in\mathbf{Z}_+)$
$(1+r^2)^{\lambda}$	$\dfrac{\rho^{-[(n/2)+\lambda]}K_{(n/2)+\lambda}(\rho)}{2^{-\lambda-1}\Gamma(-\lambda)}\quad (\lambda\notin\mathbf{Z}_+)$
$(1+r^2)^m$	$(2\pi)^{n/2}(1-\Delta)^m\delta(\xi)\quad (m\in\mathbf{Z}_+)$
$\operatorname{Pf} r^{\lambda}\log\lambda$	$\dfrac{2^{(n/2)+\lambda}\Gamma((n+\lambda)/2)}{\Gamma(-\lambda/2)}\operatorname{Pf}\rho^{-n-\lambda}\left[\log\dfrac{2}{\rho}+\dfrac{1}{2}\dfrac{\Gamma'((n+\lambda)/2)}{\Gamma((n+\lambda)/2)}+\dfrac{1}{2}\dfrac{\Gamma'(-\lambda/2)}{\Gamma(-\lambda/2)}\right]$ $(\lambda\notin 2\mathbf{Z}_+,\ \ \lambda\notin -n-2\mathbf{Z}_+)$
$r^{2m}\log r$	$(-1)^{m-1}2^{(n/2)+2m-1}m!\,\Gamma((n/2)+m)\operatorname{Pf}\rho^{-n-2m}$ $+(2\pi)^{n/2}\left[\log 2-\dfrac{1}{2}C+\dfrac{1}{2}\sum_{j=1}^{m}\dfrac{1}{j}+\dfrac{\Gamma'((n/2)+m)}{\Gamma((n/2)+m)}\right](-\Delta)^m\delta(\xi)\quad (m\in\mathbf{Z}_+)$
$\operatorname{Pf} r^{-n-2m}\log r$	$\dfrac{(-1)^m}{2^{(n/2)+2m}\Gamma((n/2)+m)m!}\rho^{2m}\left[\left\{\log\dfrac{2}{\rho}-\dfrac{1}{2}C+\dfrac{1}{2}\sum_{j=1}^{m}\dfrac{1}{j}+\dfrac{1}{2}\dfrac{\Gamma'((n/2)+m)}{\Gamma((n/2)+m)}\right\}^2\right.$ $\left.+\dfrac{\pi^2}{24}+\dfrac{1}{4}\sum_{j=1}^{m}\dfrac{1}{j^2}-\dfrac{1}{4}\dfrac{\Gamma''((n/2)+m)}{\Gamma((n/2)+m)}+\dfrac{\Gamma'((n/2)+m)^2}{\Gamma((n/2)+m)^2}\right]\quad (m\in\mathbf{Z}_+)$
e^{-ar}	$\dfrac{\sqrt{2^n}}{\sqrt{\pi}}\Gamma\left(\dfrac{n+1}{2}\right)\dfrac{a}{(a^2+\rho^2)^{(n+1)/2}}\quad (a>0)$

The Fourier transform mentioned above is a transformation in the family of complex-valued functions or distributions. Similar transformations in the family of real-valued functions are frequently used in applications:

Fourier cosine transform $f_c(u)=\int_0^\infty F(t)\cos ut\,dt.$

Inverse transform $\frac{2}{\pi}\int_0^\infty f_c(u)\cos ut\,du=\begin{cases} F(t) & (t>0),\\ F(-t) & (t<0).\end{cases}$

Fourier sine transform $f_s(u)=\int_0^\infty F(t)\sin ut\,dt.$

Inverse transform $\frac{2}{\pi}\int_0^\infty f_s(u)\sin ut\,du=\begin{cases} F(t) & (t>0),\\ -F(-t) & (t<0).\end{cases}$

The Fourier transform can be expressed in terms of these transforms. For example (in $\mathbf{R}^1$),

$$\mathcal{F}[f](u)=\frac{1}{\sqrt{2\pi}}\int_0^\infty [f(t)+f(-t)]\cos ut\,dt-\frac{i}{\sqrt{2\pi}}\int_0^\infty [f(t)-f(-t)]\sin ut\,dt.$$

$F(t)$	$f_c(u)$	$f_s(u)$
$\begin{cases} 1 & (0<t<a)\\ 0 & (a<t)\end{cases}$	$\dfrac{\sin au}{u}$	$\dfrac{1-\cos au}{u}$
t^{-1}	(diverges)	$(\pi/2)\operatorname{sgn} u$
$t^{\alpha-1}\quad(0<\alpha<1)$	$\Gamma(\alpha)\cos(\pi\alpha/2)u^{-\alpha}$	$\Gamma(\alpha)\sin(\pi\alpha/2)u^{-\alpha}$
$1/(a^2+t^2)$	$\pi e^{-a\|u\|}/2a$	$[e^{-au}\mathrm{Ei}(au)-e^{au}\mathrm{Ei}(-au)]/a$ (2)
e^{-at}	$a/(a^2+u^2)$	$u/(a^2+u^2)$
$e^{-\lambda t^2}\quad(\mathrm{Re}\,\lambda>0)$	$\sqrt{\pi/4\lambda}\,e^{-u^2/4\lambda}$	$e^{-u^2/4\lambda}\varphi(u/2\sqrt{\lambda})/\sqrt{\lambda}$ (3)
$e^{-\lambda t^2}t$		$\sqrt{\pi/4\lambda}\,(u/4\lambda)e^{-u^2/4\lambda}$
$\dfrac{\sin at}{t}\quad(a>0)$	$\begin{cases} \pi/2 & (0<u<a)\\ 0 & (a<u)\end{cases}$	$\dfrac{1}{2}\log\left\|\dfrac{a+u}{a-u}\right\|$
$\tanh(\pi t/2)$		$\operatorname{cosech} u$
$\operatorname{sech}(\pi t/2)$	$\operatorname{sech} u$	
$J_\nu(t)\quad(\mathrm{Re}\,\nu>-1)$	$\begin{cases} \dfrac{\cos(\nu\arcsin u)}{\sqrt{1-u^2}} \\ -\dfrac{(u-\sqrt{u^2-1})^\nu}{\sqrt{u^2-1}}\sin\dfrac{\nu\pi}{2}\end{cases}$	$\dfrac{\sin(\nu\arcsin u)}{\sqrt{1-u^2}}\quad(0<u<1)$ $\dfrac{(u-\sqrt{u^2-1})^\nu}{\sqrt{u^2-1}}\cos\dfrac{\nu\pi}{2}\quad(1<u)$
$J_0(at)$	$\begin{cases} 1/\sqrt{a^2-u^2}\\ 0\end{cases}$	$0\quad(0\leqslant u<a)$ $1/\sqrt{u^2-a^2}\quad(a<u)$
$N_0(t)$	$\begin{cases} 0\\ -\dfrac{1}{\sqrt{u^2-1}}\end{cases}$	$\dfrac{2}{\pi}\dfrac{\arcsin u}{\sqrt{1-u^2}}\quad(0<u<1)$ $\dfrac{2}{\pi}\dfrac{\log(u-\sqrt{u^2-1})}{\sqrt{u^2-1}}\quad(1<u)$
$K_0(t)$	$\pi/2\sqrt{1+u^2}$	$(\operatorname{arcsinh} u)/\sqrt{1+u^2}$

Notes

(2) Ei is the exponential integral function (→ Table 19.II.3, this Appendix).

(3) We put $\varphi(x)=\int_0^x e^{t^2}dt.$

12. Laplace Transforms and Operational Calculus

(I) Laplace Transforms (→ 240 Laplace Transform)

Laplace transform $V(p)=\int_0^\infty e^{-pt}F(t)\,dt\quad(\mathrm{Re}\,p>0).$

Inverse transform (Bromwich integral) $\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{pt}V(p)\,dp=\begin{cases} F(t) & (t>0),\\ 0 & (t<0).\end{cases}$

$F(t)$	$V(p)$	$F(t)$	$V(p)$
1	$1/p$	$J_\nu(t)$ $(\mathrm{Re}\,\nu>-1)$	$\dfrac{(\sqrt{1+p^2}-p)^\nu}{\sqrt{1+p^2}}$
$1(t-a)=\begin{cases}0 & (0\leqslant t<a)\\ 1 & (a\leqslant t)\end{cases}$	e^{-ap}/p	$\frac{1}{t}J_\nu(at)$ $(\mathrm{Re}\,\nu>0)$	$\dfrac{(\sqrt{a^2+p^2}-p)^\nu}{\nu a^\nu}$
$[x/a]$ (integral part)	$1/p(e^{ap}-1)$		
$t^{\alpha-1}$ $(\mathrm{Re}\,\alpha>0)$	$\Gamma(\alpha)/p^\alpha$	$t^\nu J_\nu(at)\left(\mathrm{Re}\,\nu>-\frac{1}{2}\right)$	$\dfrac{(2a)^\nu\Gamma[\nu+(1/2)]}{\sqrt{\pi}\,(p^2+a^2)^{\nu+(1/2)}}$
e^{-at}	$1/(p+a)$		
$e^{-at}t^{\alpha-1}$ $(\mathrm{Re}\,\alpha>0,\ a>0)$	$(p+a)^{-\alpha}\Gamma(\alpha)$	$t^{\nu/2}J_\nu(x\sqrt{t})$ $(\mathrm{Re}\,\nu>-1)$	$\dfrac{x^\nu}{2^\nu p^{\nu+1}}e^{-x^2/4p}$
$e^{-at}F(t)\ (a>0)$	$V(p+a)$		
$(1-e^{-t})/t$	$\log(1+p^{-1})$	$J_0(t)$	$(1+p^2)^{-1/2}$
$(\pi t)^{-1/2}e^{-x^2/4t}$	$p^{-1/2}e^{-x\sqrt{p}}\ (x>0)$	$J_0(x\sqrt{t})$	$e^{-x^2/4p}/p$
$\log t$	$-(\log p+C)/p$ (1)	$N_0(t)$	$\dfrac{2}{\pi}\dfrac{\log(\sqrt{1+p^2}-p)}{\sqrt{1+p^2}}$
$\sin at$	$a/(p^2+a^2)$	$L_n(t)$ (2)	$\dfrac{1}{p}\left(\dfrac{p-1}{p}\right)^n$
$\cos at$	$p/(p^2+a^2)$		
$\sin(x\sqrt{t})$	$\dfrac{\sqrt{\pi}}{2}\dfrac{x}{p^{3/2}}e^{-x^2/4p}$	$t^\alpha L_n^{(\alpha)}(t)$	$\dfrac{\Gamma(\alpha+n+1)}{n!}\dfrac{1}{p^{\alpha+1}}\left(\dfrac{p-1}{p}\right)^n$
$t^{-1/2}\cos(x\sqrt{t})$	$\sqrt{\pi/p}\,e^{-x^2/4p}$	$H_{2n+1}(\sqrt{t})$ (3)	$\sqrt{\dfrac{\pi}{2}}\,(2n+1)!!\dfrac{(1-p)^n}{p^{n+(3/2)}}$ (4)
$t^{-1}\sin xt$	$\arctan(x/p)$		
$x^{-1}(1-\cos ax)$	$\frac{1}{2}\log[1+(a^2/p^2)]$		
$\sinh at$	$a/(p^2-a^2)$	$\dfrac{H_{2n}(\sqrt{t})}{\sqrt{t}}$	$\sqrt{\pi}\,(2n-1)!!\dfrac{(1-p)^n}{p^{n+(1/2)}}$
$\cosh at$	$p/(p^2-a^2)$		

Notes

(1) C is Euler's constant.
(2) $L_n(t)$ is a Laguerre polynomial.
(3) $H_n(t)$ is a Hermite polynomial.
(4) $(2n+1)!!=(2n+1)(2n-1)(2n-3)\ldots5\cdot3\cdot1$.

(II) Operational Calculus (→ 306 Operational Calculus)

Heaviside function (unit function) $\mathbf{1}(t)=\begin{cases}0 & (t<0)\\ 1 & (t\geqslant 0).\end{cases}$

Dirac delta function (unit impulse function) $\delta(t)=\lim\limits_{\varepsilon\to0}\dfrac{1}{2\varepsilon}[\mathbf{1}(t+\varepsilon)-\mathbf{1}(t-\varepsilon)]$.

When an operator $\Omega(p)$ operates on $\mathbf{1}(t)$ and the result is $A(t)$ we write $\Omega(p)\mathbf{1}(t)=A(t)$.
In the following table (i) of general formulas, we assume the relations $\Omega_i(p)\mathbf{1}(t)=A_i(t)$ $(i=1,2)$.

Carson's integral $\Omega(p)=p\displaystyle\int_0^\infty e^{-pt}A(t)\,dt$ $(\mathrm{Re}\,p>0)$.

Laplace transform $V(p)=\dfrac{\Omega(p)}{p}=\displaystyle\int_0^\infty e^{-pt}A(t)\,dt$.

(i) General Formulas	
$\Omega(p)$	$A(t)$
$\Omega_1(p)+\Omega_2(p)$	$A_1(t)+A_2(t)$
$a\Omega_1(p)$	$aA_1(t)$
$p\Omega_1(p)$	$A_1(0)\delta(t)+A_1'(t)$
$\frac{1}{p}\Omega_1(p)$	$\int_0^t A_1(\tau)d\tau$
$\Omega_1(ap)$	$A_1(t/a)$
$[p/(p+a)]\Omega_1(p+a)$	$e^{-at}A_1(t)$ $(\mathrm{Re}\,a \geqslant 0)$
$\frac{1}{p}\Omega_1(p)\Omega_2(p)$	$\int_0^t A_1(\tau)A_2(t-\tau)d\tau$ $=\int_0^t A_1(t-\tau)A_2(\tau)d\tau$

(ii) Examples	
$\Omega(p)=pV(p)$	$A(t)$
p	$\delta(t)$
$1/p^n$ $(n=0,1,2,\ldots)$	$(t^n/n!)\mathbf{1}(t)$
$p/(p+a)$	$(e^{-at})\mathbf{1}(t)$
$p^2/(p^2+a^2)$	$(\cos at)\mathbf{1}(t)$
$ap/(p^2+a^2)$	$(\sin at)\mathbf{1}(t)$
$a_0+\frac{a_1}{p}+\frac{a_2}{p^2}+\ldots$	$\left(a_0+a_1\frac{t}{1!}+a_2\frac{t^2}{2!}+\ldots\right)\mathbf{1}(t)$
$\sum_{k=1}^{n}\frac{B_k}{p-p_k}$	$\sum_{k=1}^{n}\frac{B_k}{p_k}(e^{p_k t}-1)\mathbf{1}(t)$ $=\Omega(0)\mathbf{1}(t)+\sum_{k=1}^{n}\frac{B_k}{p_k}e^{p_k t}$

13. Conformal Mappings

(→ 77 Conformal Mappings)

Original Domain	Image Domain	Mapping Function
$\|z\|<1$ (unit disk)	$\|w\|<1$	$w=\varepsilon\frac{z-z_0}{1-\bar{z}_0 z}$, $\|z_0\|<1$, $\|\varepsilon\|=1$ (general form)
$\mathrm{Im}\,z>0$ (upper half-plane)	$\|w\|<1$	$w=\varepsilon\frac{z-z_0}{z-\bar{z}_0}$, $\mathrm{Im}\,z_0>0$, $\|\varepsilon\|=1$ (general form)
$\mathrm{Im}\,z>0$ (upper half-plane)	$\mathrm{Im}\,w>0$	$w=\frac{az+b}{cz+d}$, a,b,c,d are real; $ad-bc>0$ (general form)
$0<\arg z<\alpha$ (angular domain)	$\mathrm{Im}\,w>0$	$w=z^{\pi/\alpha}$
$\|z\|<1, \mathrm{Im}\,z>0$ (upper semidisk)	$\mathrm{Im}\,w>0$	$w=\left(\frac{1+z}{1-z}\right)^2$
$0<\arg z<\alpha$, $\|z\|<1$ (fan shape)	$\mathrm{Im}\,w>0$	$w=\left(\frac{1+z^{\pi/\alpha}}{1-z^{\pi/\alpha}}\right)^2$
$\alpha<\arg\frac{z-p}{z-q}<\beta$ (circular triangle)	$0<\arg w<\gamma$	$w=\left(e^{-i\alpha}\frac{z-p}{z-q}\right)^{\frac{\gamma}{\beta-\alpha}}$
$0<\mathrm{Im}\,z<\eta$ (parallel strip)	$\mathrm{Im}\,w>0$	$w=e^{\pi z/\eta}$
$\mathrm{Re}\,z<0$, $0<\mathrm{Im}\,z<\eta$ (semiparallel strip)	$\mathrm{Im}\,w>0$, $\|w\|<1$	$w=e^{\pi z/\eta}$
$y^2>4c^2(x+c^2)$, $z=x+iy$, $c>0$ (exterior of a parabola)	$\mathrm{Im}\,w>0$	$w=\sqrt{z}-ic$
$y^2<4c^2(x+c^2)$, $z=x+iy$, $c>0$ (interior of a parabola)	$\mathrm{Im}\,w>0$	$w=i\sec\frac{\pi\sqrt{z}}{2ic}$

Original Domain	Image Domain	Mapping Function
$\frac{x^2}{[c+(1/c)]^2}+\frac{y^2}{[c-(1/c)]^2}>1$, $z=x+iy$, $c>1$ (exterior of an ellipse)	$\lvert w\rvert>c$	$w=\frac{z+\sqrt{z^2-4}}{2}$, $z=w+\frac{1}{w}$
$\frac{x^2}{\cos^2\alpha}-\frac{y^2}{\sin^2\alpha}<4$, $z=x+iy$, $0<\alpha<\pi/2$ (exterior of a hyperbola)	$\operatorname{Im} w>0$	$w=\left(e^{-i\alpha}\frac{z+\sqrt{z^2-4}}{2}\right)^{\pi(\pi-2\alpha)}$
$\frac{x^2}{\cos^2\alpha}-\frac{y^2}{\sin^2\alpha}>4$, $x>0$, $z=x+iy$, $0<\alpha<\pi/2$ (right-hand side interior of a hyperbola)	$\operatorname{Im} w>0$	$w=\frac{1}{2i}\left[\left(\frac{z+\sqrt{z^2-4}}{2}\right)^{\frac{\pi}{2\alpha}}+\left(\frac{z-\sqrt{z^2-4}}{2}\right)^{\frac{\pi}{2\alpha}}\right]$
$\lvert z\rvert<1$	Slit domain with boundary $\lvert\operatorname{Re} w\rvert\leqslant 2$, $\operatorname{Im} w=0$	$w=z+\frac{1}{z}$
$\lvert z\rvert<1$	Slit domain with boundary $\lvert w\rvert\geqslant 1/4$, $\arg w=\lambda$	$w=\frac{z}{(1+e^{-i\lambda}z)^2}$
$\lvert z\rvert<1$	Slit domain with boundary $\lvert w\rvert\geqslant 1/4^{1/p}$ $\arg w=\lambda+(2j\pi/p)$, $j=0,\ldots,p-1$	$w=\frac{z}{(1+e^{-ip\lambda}z^p)^{2/p}}$
$-\pi/2<\operatorname{Re} z<\pi/2$ (parallel strip)	Slit domain with boundary $\lvert\operatorname{Re} w\rvert\geqslant 1$, $\operatorname{Im} w=0$	$w=\sin z$
$-\pi<\operatorname{Im} z<\pi$ (parallel strip)	Slit domain with boundary $\operatorname{Re} w\leqslant -1$, $\operatorname{Im} w=\pm\pi$	$w=z+e^z$
Arbitrary circle or half plane	Interior of an n-gon	$w=c\int^z\prod_{j=1}^{n}(t-z_j)^{\alpha_j-1}dt+c'$ ($c\neq 0$, c' are constants), where the inverse image of the vertex with the inner angle $\alpha_j\pi$ ($j=1,\ldots,n$) is $z=z_j$. When $z_n=\infty$, we omit the factor $(t-z_n)^{\alpha_n-1}$ (Schwarz-Christoffel transformation)
Arbitrary circle or half-plane	Exterior of an n-gon	$w=c\int^z(t-p)^{-2}\prod_{j=1}^{n}(t-z_j)^{1-\alpha_j}dt+c'$ ($c\neq 0$, c' are constants), where the inverse image of the vertex with the inner angle $\alpha_j\pi$ ($j=1,\ldots,n$) is $z=z_j$, and the inverse image of ∞ is $z=p$
$\operatorname{Im} z>0$	Interior of an equilateral triangle	$w=\int_0^z\frac{1}{\sqrt[3]{t^2(1-t)^2}}dt$
$\operatorname{Im} z>0$	Interior of an isosceles right triangle	$w=\int_0^z\frac{1}{\sqrt[5]{t^2(1-t)^3}}dt$

Original Domain	Image Domain	Mapping Function
$\operatorname{Im} z>0$	Interior of a right triangle with one angle $\pi/6$	$w=\int_0^z \frac{1}{\sqrt[6]{t^3(1-t)^4}}\,dt$
$\|z\|<1$	Interior of a regular n-gon	$w=\int_0^z (1-t^n)^{-2/n}\,dt$
$0<\operatorname{Re} z<\omega_1$, $0<\operatorname{Im} z<\omega_3/i$ (rectangle)	$\operatorname{Im} w>0$	$w=\wp(z\|2\omega_1,2\omega_3)$ ($\wp$ is the Weierstrass $\wp$-function)
$-K<\operatorname{Re} z<K$, $0<\operatorname{Im} z<K'$ (rectangle)(1)	$\operatorname{Im} w>0$	$w=\operatorname{sn}(z,k)$, $z=\int_0^w \frac{1}{\sqrt{(1-t^2)(1-k^2t^2)}}\,dt$ (sn is Jacobi's sn function)
$v<\|z\|<1$ $\operatorname{Im} z<0$ (upper half-ring domain)	$\log q<\operatorname{Re} w<0$, $0<\operatorname{Im} w<\pi$ (rectangle)	$w=\log z$
$\|z\|<1$	$\frac{u^2}{A^2}+\frac{v^2}{B^2}<1$, $w=u+v$, $A>B>0$; (interior of an ellipse)	$w=\sqrt{A^2-B^2}\,\sin\left(\frac{\pi}{2K}\int_0^{2z/k(1+z^2)}\frac{dt}{\sqrt{(1-t^2)(1-k^2t^2)}}\right)$, $\frac{1}{\pi}\log\frac{A+B}{A-B}=\frac{K'}{K}$ (1)
$\operatorname{Im} z>0$	interior of a circular polygon	$\{w;z\}\equiv\frac{w'''}{w'}-\frac{3}{2}\left(\frac{w''}{w'}\right)^2=R(z)$ ($R(z)$ is a rational function)
$\|z\|<1$	Interior of an equilateral circular triangle with inner angle π/k, $1<k\leqslant\infty$	$w=\dfrac{z\int_0^1 t^{-\frac{1}{2}-\frac{1}{2k}}(1-t)^{-\frac{1}{6}+\frac{1}{2k}}(1-z^3t)^{-\frac{5}{6}+\frac{1}{2k}}\,dt}{\int_0^1 t^{-\frac{1}{2}-\frac{1}{2k}}(1-t)^{-\frac{5}{6}+\frac{1}{2k}}(1-z^3t)^{-\frac{1}{6}+\frac{1}{2k}}\,dt}$ The vertices are the images of $z=1$, $e^{2\pi i/3}$, and $e^{4\pi i/3}$; and $\left[\frac{dw}{dz}\right]_{z=0}=\frac{\Gamma[(5/6)+(1/2k)]\Gamma(2/3)}{\Gamma[(1/6)+(1/2k)]\Gamma(4/3)}$
$\operatorname{Im} z>0$	Interior of a circular triangle with inner angles $\pi\alpha$, $\pi\beta$, $\pi\gamma$, $\alpha+\beta+\gamma<1$(2)	$w=\dfrac{\int_0^1 t^{-\frac{1+\alpha+\beta+\gamma}{2}}(1-t)^{-\frac{1+\alpha-\beta-\gamma}{2}}(1-zt)^{-\frac{1-\alpha+\beta-\gamma}{2}}\,dt}{\int_0^1 t^{-\frac{1+\alpha+\beta+\gamma}{2}}(1-t)^{-\frac{1-\alpha-\beta+\gamma}{2}}(1-t+zt)^{-\frac{1-\alpha+\beta-\gamma}{2}}\,dt}$
$\|\tau\|>1$, $-1/2<\operatorname{Re}\tau<0$	$\operatorname{Im} J>0$	$J=J(\tau)$, $\tau=\omega_3/\omega_1$, $J=g_2^3/(g_2^3-27g_3^2)$ (the absolute invariant of the elliptic modular function); $J(e^{2\pi i/3})=0$, $J(i)=1$, $J(\infty)=\infty$
$\|\tau+1/2\|<1/2$, $-1<\operatorname{Re}\tau<0$	$\operatorname{Im}\lambda<0$	$\lambda=\lambda(\tau)$, $\tau=\omega_3/\omega_1$, $\lambda=(e_2-e_3)/(e_1-e_3)$; $J(\tau)\equiv\frac{4}{27}\frac{[\lambda(\tau)^2-\lambda(\tau)+1]^2}{\lambda(\tau)^2[\lambda(\tau)-1]^2}$, $\lambda(-1)=\infty$, $\lambda(0)=1$, $\lambda(\infty)=0$

Notes

(1) K, K', k' are the usual notations in the in the theory of elliptic integrals:

$$K\equiv\int_0^1\frac{1}{\sqrt{(1-t^2)(1-k^2t^2)}}\,dt,\qquad K'=\int_0^1\frac{1}{\sqrt{(1-t^2)(1-k'^2t^2)}}\,dt,\qquad k^2+k'^2=1.$$

(2) When $\alpha+\beta+\gamma=1$, the circular triangle is mapped into the ordinary linear triangle by a suitable linear transformation, and we can apply the Schwarz-Christoffel transformation. When $\alpha+\beta+\gamma>1$, we have a similar mapping function replacing the integral representations of hypergeometric functions in the formula by the corresponding integral representations of the hypergeometric functions converging at α, β, and γ.

14. Ordinary Differential Equations

(I) Solution by Quadrature

$a,b,c,\ldots$ are integral constants.

(1) Solution of the First-Order Differential Equations (→ 313 Ordinary Differential Equations).
(i) Separated type $dy/dx=X(x)Y(y)$. The general solution is

$$\int^{y}\frac{dy}{Y(y)}=\int^{x}X(x)\,dx+c.$$

(ii) Homogeneous ordinary differential equation $dy/dx=f(y/x)$. Putting $y=ux$, we have $du/dx=[f(u)-u]x$, and the equation reduces to type (i). The general solution is

$$x=c\exp\left[\int^{u}\frac{du}{f(u)-u}\right]\quad\left(u=\frac{y}{x}\right).$$

(iii) Linear ordinary differential equation of the first order. $dy/dx+p(x)y+q(x)=0$. The general solution is

$$y=\left[c-\int q(x)P(x)\,dx\right]\Big/P(x),$$

where

$$P(x)\equiv\exp\left[\int p(x)\,dx\right].$$

(iv) Bernoulli's differential equation $dy/dx+p(x)y+q(x)y^{\alpha}=0$ $(\alpha\neq 0,1)$. Putting $z=y^{1-\alpha}$, the equation is transformed into

$$dz/dx+(1-\alpha)p(x)z+(1-\alpha)q(x)=0,$$

which reduces to (iii).
(v) Riccati's differential equation $dy/dx+ay^2=bx^m$. If $m=-2,\ 4k/(1-2k)$ (k an integer), this is solved by quadrature. In general, it is reduced to Bessel's differential equation by $ay=u'/u$.
(vi) Generalized Riccati differential equation $dy/dx+p(x)y^2+q(x)y+r(x)=0$. If we know one, two, or three special solutions $y=y_i(x)$, the general solution is represented as follows. When $y_1(x)$ is one known special solution,

$$y=y_1(x)+P(x)\Big/\left[\int p(x)P(x)\,dx+c\right],$$

where

$$P(x)\equiv\exp\left[-\int\{q(x)+2p(x)y_1(x)\}\,dx\right].$$

When $y_1(x)$, $y_2(x)$ are the known solutions,

$$\frac{y-y_1(x)}{y-y_2(x)}=c\exp\left[\int p(x)\{y_2(x)-y_1(x)\}\,dx\right].$$

When $y_1(x)$, $y_2(x)$, $y_3(x)$ are known solutions,

$$\frac{y-y_1(x)}{y-y_2(x)}=c\,\frac{y_3(x)-y_1(x)}{y_3(x)-y_2(x)}.$$

(vii) Exact differential equation $P(x,y)dx+Q(x,y)dy=0$. If the left-hand side is an exact differential form, the condition is $\partial P/\partial y=\partial Q/\partial x$. The general solution is

$$\int P\,dx+\int\left(Q-\frac{\partial}{\partial y}\int P\,dx\right)dy=c.$$

(viii) Integrating factors. A function $M(x,y)$ is called an integrating factor of a differential equation $P(x,y)dx+Q(x,y)dy=0$, if $M(x,y)[P(x,y)dx+Q(x,y)dy]$ is an exact differential form $d\varphi(x,y)$. If we know an integrating factor, the general solution is given by $\varphi(x,y)=c$. If we know two independent integrating factors M and N, the general solution is given by $M/N=c$.
(ix) Clairaut's differential equation $y=xp+f(p)$ $(p=dy/dx)$. The general solution is the family of straight lines $y=cx+f(c)$, and the singular solution is the envelope of this family, which is given by eliminating p from the original equation and $x+f'(p)=0$.
(x) Lagrange's differential equation $y=x\varphi(p)+\psi(p)$ $(p\equiv dy/dx)$. Differentiation with respect to x reduces the equation to a linear differential equation $[\varphi(p)-p](dx/dp)+\varphi'(p)x+\psi'(p)=0$

with respect to x, p (see (iii)). The general solution of the original equation is given by eliminating p from the original equation and the solution of the latter linear equation. The parameter p may represent the solution. If the equation $p=\varphi(p)$ has a solution $p=p_0$, we have a solution $y=p_0x+\psi(p_0)$ (straight line). This solution is sometimes the singular solution.

(xi) Singular solutions. The singular solution of $f(x,y,p)=0$ is included in the equation resulting from eliminating p from $f=0$ and $\partial f/\partial p=0$, though the eliminant may contain various curves that are not the singular solutions.

(xii) System of differential equations.

eq. (1) $\quad dx:dy:dz=P:Q:R.$

A function $M(x,y,z)$ is called a Jacobi's last multiplier for eq. (1) if M is a solution of a partial differential equation $(\partial MP/\partial x)+(\partial MQ/\partial y)+(\partial MR/\partial z)=0$. If we know two independent last multipliers M and N, then $M/N=c$ is a solution of eq. (1). If we know a last multiplier M and a solution $f=a$ of eq. (1), we may find another solution of (1) as follows: solving $f=a$ with respect to z and inserting the solution into eq. (1), we see that $M(Qdx-Pdy)/f_z$ is an exact differential form $dG(x,y,a)$ in three variables x, y, and a. Then $G(x,y,f(x,y,z))=b$ is another solution of eq. (1).

(2) Solutions of Higher-Order Ordinary Differential Equations. The following (i)–(iv) are several examples of depression.

(i) $f(x,y^{(k)},y^{(k+1)},\ldots,y^{(n)})=0$ $(0<k\leqslant n)$. Set $y^{(k)}=z$; the equation reduces to one of the $(n-k)$th order in z.

(ii) $f(y,y',y'',\ldots,y^{(n)})=0$. This is reduced to $(n-1)$st order if we consider $y'=p$ as a variable dependent on y.

(iii) $y''=f(y)$. The general solution is given by

$$x=a\pm\int\left[2\int f(y)\,dy+b\right]^{-1/2}dy.$$

We have a similar formula for $y^{(n)}=f(y^{(n-2)})$.

(iv) Homogeneous ordinary differential equation of higher order. If the left-hand side of $F(x,y,y',\ldots,y^{(n)})=0$ satisfies the homogeneity relation $F(x,\rho y,\rho y',\ldots,\rho y^{(n)})=\rho^\alpha F(x,y,y',\ldots,y^{(n)})$, the equation is reduced to one of the $(n-1)$st order in u by $u=y'/y$.

If F satisfies $F(\rho x,\rho^l y,\rho^{l-1}y',\ldots,\rho^{l-n}y^{(n)})=\rho^\alpha F(x,y,y',\ldots,y^{(n)})$, then $u=y/x^l$, $t=\log x$ reduces the equation to one of type (ii) not containing t.

(v) Euler's linear ordinary differential equation.

$$p_n(x)x^n y^{(n)}+p_{n-1}(x)x^{n-1}y^{(n-1)}+\ldots+p_1(x)xy'+p_0(x)y=q(x)$$

is reduced to a linear equation by $t=\log x$.

(vi) Linear ordinary differential equations of higher order (exact equations). A necessary and sufficient condition that $L[y]\equiv\sum_{j=0}^n p_j(x)y^{(j)}=X(x)$ is an exact differential form is $\sum_{j=0}^n(-1)^j p_j^{(j)}=0$, and then the first integral of the equation is given by

$$\sum_{j=0}^{n-1}\sum_{k=0}^{n-j-1}(-1)^k p_{k+j+1}^{(k)}y^{(j)}=\int X(x)\,dx+c.$$

(vii) Linear ordinary differential equation of higher order (depression).

$$L[y]\equiv\sum_{j=0}^{n}p_j(x)y^{(j)}=X(x).$$

If we know mutually independent special solutions $y_1(x),\ldots,y_m(x)$ for the homogeneous linear ordinary differential equation $L[y]=0$, the equation is reduced to the $(n-m)$th linear ordinary differential equation with respect to z by a transformation $z=A(y)$, where $A(y)=0$ is the mth linear ordinary differential equation with solutions $y_1(x),\ldots,y_m(x)$. For example, if $m=1$, the equation is reduced to the $(n-1)$st linear ordinary differential equation with respect to z by the transformation

$$y(x)=y_1(x)\int z(x)\,dx.$$

Also, if $n=m=2$, the general solution is

$$y=c_1y_1+c_2y_2-y_1\int Ty_2\,dx+y_2\int Ty_1\,dx,$$

where $T(x)\equiv X(x)/[y_1(x)y_2'(x)-y_2(x)y_1'(x)]$. The denominator of the last expression is the Wronskian of y_1 and y_2.

(viii) Regular singularity. For a linear ordinary differential equation of higher order,

eq. (1) $\quad x^n y^{(n)} + x^{n-1}p_1(x)y^{(n-1)} + \ldots + p_n(x)y = 0,$

the point $x=0$ is its regular singularity if $p_1(x),\ldots,p_n(x)$ are analytic at $x=0$.

We put $p_0=1$ and

$$\sum_{\nu=0}^{\infty} f_\nu(\rho)x^\nu \equiv \sum_{j=0}^{n} p_{n-j}(x)\rho(\rho-1)\ldots(\rho-j+1).$$

If ρ is a root of the characteristic equation $f_0(\rho)=0$ and $\rho+1,\rho+2,\ldots$ are not roots, we can determine the coefficients c_ν uniquely from

eq. (2) $\quad \sum_{\nu=0}^{m} c_\nu f_{m-\nu}(\rho+\nu)=0 \qquad (m=1,2,\ldots),$

starting from a fixed value $c_0\ (\neq 0)$, and the series $y=x^\rho\sum_{\nu=0}^{\infty}c_\nu x^\nu$ converges and represents a solution of eq. (1). If the differences of all pairs of roots of the determining equation are not integers, we have n linearly independent solutions of eq. (1) applying the process for each characteristic root.

If there are roots whose differences are integers (including multiple roots), we denote such a system of roots by $\rho_1,\ldots,\rho_l$. We arrange them in increasing order, and denote the multiplicities of the roots by $e_1,\ldots,e_l$, respectively. Put $q_k=\rho_k-\rho_1$ $(k=1,2,\ldots,l; 0=q_1<q_2<\ldots<q_l)$. Take $N\geqslant q_l$ and a constant $c\ (\neq 0)$. Let λ be a parameter, and starting from $c_0=c_0(\lambda)\equiv c\prod_{k=1}^{N}f_0(\lambda+k)$, we determine $c_\nu=c_\nu(\lambda)$ uniquely by the relation (2). Putting

$m_k\equiv e_k+e_{k+1}+\ldots+e_l \quad (k=1,\ldots,l) \qquad$ for h in $\quad m_{k+1}\leqslant h\leqslant m_k-1,$

the series

eq. (3) $$y=\left[\frac{\partial^h}{\partial\lambda^h}x^\lambda\sum_{\nu=0}^{\infty}c_\nu(\lambda)x^\nu\right]_{\lambda=\rho_k}=x^{\rho_k}\sum_{\nu=0}^{\infty}x^\nu\left[\sum_{j=0}^{h}\binom{h}{j}c_\nu^{(j)}(\rho_k)(\log x)^{h-j}\right]$$

converges and gives e_k independent solutions of eq. (1). Hence for $k=1,\ldots,l$, we may have $\sum_{k=1}^{l}e_k=m_1$ mutually independent solutions of (1). Applying this process to every characteristic root, we have finally n independent solutions of (1) (Frobenius method).

In the practical computation of the solution, since it is known to have the expression (3), we often determine its coefficients successively by the method of undetermined coefficients.

(3) Solution of Linear Ordinary Differential Equations with Constant Coefficient (→ 252 Linear Ordinary Differential Equations). Let α_i, α_{jk} be constants. We consider the following linear ordinary differential equation of higher order (eq. (1)) and system of linear ordinary differential equations (eq. (2)).

eq. (1) $\quad \sum_{i=0}^{n}\alpha_i y^{(i)}=X(x).$

eq. (2) $\quad y_j'=\sum_{k=1}^{n}\alpha_{jk}y_k+X_j(x) \quad (j=1,\ldots,n).$

(i) The general solution of the homogeneous equation (cofactor) is given by the following formulas:

for eq. (1) $\quad y=x^j\exp\lambda_k x \quad (j=0,1,,\ldots,e_k-1; k=1,\ldots,m),$

for eq. (2) $\quad y_j(x)=\sum_{k=1}^{m}p_{jk}(x)\exp\lambda_k x \quad (j=1,\ldots,n),$

where $\lambda_1,\ldots,\lambda_m$ are the roots of the characteristic equation of eq. (1) or eq. (2) given by

eq. (1′) $\quad \sum_{i=0}^{n}\alpha_i\lambda^i=0,$

eq. (2′) $\quad \det(\alpha_{jk}-\lambda\delta_{jk})=0,$

respectively. We denote the multiplicities of the roots by $e_1,\ldots,e_m$ $(e_1+\ldots+e_m=n)$; $p_{jk}(x)$ is a polynomial of degree at most e_k-1 containing e_k arbitrary constants.

If all the coefficients in the original equation are real, and the root $\lambda_k=\mu_k+i\nu_k$ is imaginary, then $\bar{\lambda}_k=\mu_k-i\nu_k$ is also a root with the same multiplicity. Then we may replace $\exp\lambda_k x$ and $\exp\bar{\lambda}_k x$ by $\exp\mu_k x\cos\nu_k x$ and $\exp\mu_k x\sin\nu_k x$, and in this way we can represent the solution using real functions.

(ii) Inhomogenous equation. The solution of an inhomogeneous linear ordinary differential equation is given by the method of variation of parameters or by the method described in Section (2)(vii).

We explain the method of variation of parameters for eq. (2). First we use (i) to find a fundamental system of n independent solutions $y_j = \varphi_{jk}(x)$ $(k=1,\ldots,n)$ by (i)). Inserting $y_j = \sum_{k=1}^{n} c_k(x)\varphi_{jk}(x)$ into eq. (2), we have a system of linear equations in the $c_k'(x)$. Solving for the $c_k'(x)$ and integrating, we have $c_k(x)$.

Special forms of $X(x)$ or $X_j(x)$ determine the form of the solutions, and the parameters may be found by the method of undetermined coefficients. The following table shows some examples of special solutions for eq. (1). In the table, α, k, a, b, c, are constants, p_r, q_r are polynomials of degree r, and I_a is the operator defined by

$$I_a \cdot F = \frac{1}{a}\left[\sin ax \int \cos ax \cdot F(x)\,dx - \cos ax \int \sin ax \cdot F(x)\,dx\right] \quad (a \neq 0).$$

$X(x)$	Condition	Special Solution
$p_r(x)$	$\lambda=0$ is an m-tuple root of(1′)	$x^m q_r(x)$
$ke^{\alpha x}$	$\lambda=\alpha$ is an m-tuple root of (1′)	$cx^m e^{\alpha x}$
$e^{\alpha x}p_r(x)$	$\lambda=\alpha$ is an m-tuple root of (1′)	$x^m q_r(x)e^{\alpha x}$
$\left.\begin{matrix}\cos(ax+b)\\ \sin(ax+b)\end{matrix}\right\}$	$\left\{\begin{matrix}(1')\equiv\varphi(\lambda^2)+\lambda\psi(\lambda^2), \text{ and}\\ \varphi(-a^2)+a^2\psi(-a^2)\neq 0\end{matrix}\right\}$	$c_1\cos(ax+b)+c_2\sin(ax+b)$
$\left.\begin{matrix}\cos(ax+b)\\ \sin(ax+b)\end{matrix}\right\}$	$\left\{\begin{matrix}(1')=g(\lambda)/f(\lambda^2) \text{ and } f(\lambda^2)\\ \text{is divisible by } (\lambda^2+a^2)^m\\ (\text{but not by } (\lambda^2+a^2)^{m+1})\end{matrix}\right\}$	$c(I_a)^m\left\{\begin{matrix}\cos(ax+b)\\ \sin(ax+b)\end{matrix}\right.$

(II) Riemann's P-Function and Special Functions (→ 253 Linear Ordinary Differential Equations (Global Theory))

(1) Some Examples Expressed by Elementary Functions. A, B are integral constants.

$$P\left\{\begin{matrix} a & b & c & \\ 0 & \mu & -\mu & x \\ 1 & \mu' & -\mu' & \end{matrix}\right\} = \begin{cases} A\left(\dfrac{x-b}{x-c}\right)^{\mu} + B\left(\dfrac{x-b}{x-c}\right)^{\mu'} & (\mu\neq\mu'), \\ \left(\dfrac{x-b}{x-c}\right)^{\mu}\left[A + B\log\left(\dfrac{x-b}{x-c}\right)\right] & (\mu=\mu'). \end{cases}$$

$$P\left\{\begin{matrix} a & b & c & \\ \lambda & \mu & \nu & x \\ 0 & 0 & 0 & \end{matrix}\right\} = A + B\int (x-a)^{\lambda-1}(x-b)^{\mu-1}(x-c)^{\nu-1}dx \quad (\lambda+\mu+\nu=1).$$

These are for finite a, b, c. If $c=\infty$, $x-c$ should be replaced by 1.

$$P\left\{\begin{matrix} \overbrace{\quad\infty\quad}^{} & & 0 & \\ \alpha & 0 & 0 & x \\ \alpha' & 0 & 1 & \end{matrix}\right\} = \begin{cases} Ae^{\alpha x} + Be^{\alpha' x} & (\alpha\neq\alpha'), \\ e^{\alpha x}(Ax+B) & (\alpha=\alpha'). \end{cases}$$

$$P\left\{\begin{matrix} \overbrace{\quad\infty\quad}^{} & & 0 & \\ \alpha & 1-\sigma & \sigma & x \\ 0 & 0 & 0 & \end{matrix}\right\} = A + B\int e^{\alpha x}x^{\sigma-1}dx \quad (\alpha\neq 0).$$

$$P\left\{\begin{matrix} \overbrace{\quad\infty\quad}^{} & & 0 & \\ \alpha & -\sigma & \sigma & x \\ \alpha' & 1-\sigma' & \sigma' & \end{matrix}\right\} = x^{\sigma}e^{\alpha x}\left[A + B\int e^{(\alpha'-\alpha)x}x^{\sigma'-\sigma-1}dx\right] \quad (\alpha\neq\alpha').$$

Riemann's P-function is reduced to Gauss's hypergeometric function with parameters $\alpha=\lambda+\mu+\nu$, $\beta=\lambda+\mu+\nu'$, $\gamma=1+\lambda+\lambda'$ by transforming a, b, c to 0, 1, ∞ by a suitable linear transformation and by putting $z=x^{-\lambda}(x-1)^{-\mu}y$.

(2) Representation of Special Functions by Riemann's P-function.
(i) Gauss's hypergeometric differential equation $x(1-x)y''+[\gamma-(\alpha+\beta+1)x]y'-\alpha\beta y=0$.

$$y = P\left\{\begin{matrix} 0 & 1 & \infty & \\ 0 & 0 & \alpha & x \\ 1-\gamma & \gamma-\alpha-\beta & \beta & \end{matrix}\right\}.$$

A special solution is $F(\alpha,\beta,\gamma;x)$ (→ 206 Hypergeometric Functions).

(ii) Confluent hypergeometric differential equation $xy''-(x-\mu)y'-\lambda y=0$.

$$y=P\left\{\begin{matrix} \overbrace{\quad\infty\quad} & 0 & \\ 0 \quad \lambda & 0 & x \\ 1 \quad \mu-\lambda & 1-\mu & \end{matrix}\right\}.$$

A special solution is

$${}_1F_1(\lambda,\mu;x)\equiv\sum_{k=0}^{\infty}\frac{\Gamma(\lambda+k)}{\Gamma(\lambda)}\frac{\Gamma(\mu)}{\Gamma(\mu+k)}\frac{x^k}{k!}.$$

(iii) Whittaker's differential equation $y''+\left[-\frac{1}{4}+\frac{k}{x}+\frac{(1/4)-n^2}{x^2}\right]y=0$.

$$y=P\left\{\begin{matrix} \overbrace{\quad\infty\quad} & 0 & \\ 1/2 \quad k & 1/2+n & x \\ -1/2 \quad -k & 1/2-n & \end{matrix}\right\}.$$

Special solutions are $M_{k,n}(x), W_{k,n}(x)$.

(iv) Bessel's differential equation $x^2y''+xy'+(x^2-\nu^2)y=0$.

$$y=P\left\{\begin{matrix} \overbrace{\quad\infty\quad} & 0 & \\ i \quad 1/2 & \nu & x \\ -i \quad 1/2 & -\nu & \end{matrix}\right\}.$$

Special solutions are $J_\nu(x)$, $N_\nu(x)$ (→ 39 Bessel Functions). When $m=0,1,2,\ldots,J_{m-1/2}(x)=(-1)^m2^{m+1/2}\pi^{-1/2}x^{m-1/2}d^m(\cos x)/d(x^2)^m$.

(v) Hermite's different equation (parabolic cylindrical equation) $y''-2xy'+2ny=0$.

$$y=P\left\{\begin{matrix} \overbrace{\quad\infty\quad} & 0 & \\ 0 \quad -n/2 & 0 & x^2 \\ 1 \quad (n+1)/2 & 1/2 & \end{matrix}\right\}.$$

When $n=0,1,2,\ldots$, the Hermite polynomial $H_n(x)=(-1)^n2^{-n/2}e^{x^2}d^n(e^{-x^2})/dx^n$ is the solution.

(vi) Laguerre's differential equation $xy''+(l-x+1)y'+ny=0$.

$$y=P\left\{\begin{matrix} \overbrace{\quad\infty\quad} & 0 & \\ 0 \quad -n & 0 & x \\ 1 \quad l+n+1 & -l & \end{matrix}\right\}.$$

When $n=0,1,2,\ldots$, the Laguerre polynomial $L_n^l(x)=(1/n!)x^{-l}e^x d^n(x^{n+l}e^{-x})/dx^n$ is the solution.

(vii) Jacobi's differential equation $x(1-x)y''+[q-(p+1)x]y'+n(n+p)y=0$.

$$y=P\left\{\begin{matrix} 0 & 1 & \infty & \\ 1-q & q-p & p+n & x \\ 0 & 0 & -n & \end{matrix}\right\}.$$

When $n=0,1,2,\ldots$, the Jacobi polynomial

$$G_n(p,q;x)=\frac{\Gamma(q)x^{1-q}(1-x)^{q-p}}{\Gamma(n+q)}\frac{d^n\left[x^{q+n-1}(1-x)^{p+n-q}\right]}{dx^n}$$

is the solution.

(viii) Legendre's differential equation $(1-x^2)y''-2xy'+n(n+1)y=0$.

$$y=P\left\{\begin{matrix} 1 & -1 & \infty & \\ 0 & 0 & n+1 & x \\ 0 & 0 & -n & \end{matrix}\right\}.$$

When $n=0,1,2,\ldots$, the general solution is

$$\frac{d^n}{dx^n}\left[A(x^2-1)^n+B(x^2-1)^n\int\frac{dx}{(x^2-1)^{n+1}}\right].$$

The Legendre polynomial $P_n(x)=[d^n\{(x^2-1)^n\}/dx^n]/2^nn!$ is a special solution.

(3) Solution by Cylindrical Functions of Ordinary Linear Differential Equations of the Second Order. We denote cylindrical functions by $C_\nu(x)$ (→ 39 Bessel Functions).

Equation	Solution
$y''+\dfrac{1-2\alpha}{x}y'+\left[(\beta\gamma x^{\gamma-1})^2+\dfrac{\alpha^2-\nu^2\gamma^2}{x^2}\right]y=0$	$y=x^\alpha C_\nu(\beta x^\gamma)$
$y''+\left[\dfrac{1-2\alpha}{x}-2\beta\gamma i x^{\gamma-1}\right]y'+\left[\dfrac{\alpha^2-\nu^2\gamma^2}{x^2}-\beta\gamma(\gamma-2\alpha)ix^{\gamma-2}\right]y=0$	$y=x^\alpha\exp(i\beta x^\gamma)C_\nu(\beta x^\gamma)$
$y''+\left[\dfrac{1}{x}-2u(x)\right]y'+\left[1-\dfrac{\nu^2}{x^2}+u(x)^2-u'(x)-\dfrac{u(x)}{x}\right]y=0$	$y=\exp\left[\int u(x)\,dx\right]C_\nu(x)$
$y''+\alpha^2\nu^2x^{2\nu-2}y=0$	$y=\sqrt{x}\,C_{1/2\nu}(\alpha x^\nu)$
$y''+(e^{2x}-\nu^2)y=0$	$y=C_\nu(e^x)$
$x^2y''+xy'+(\beta^2x^2-\nu^2)y=0$	$y=C_\nu(\beta x)$
$x^2y''+xy'-(x^2+\nu^2)y=0$	$y=C_\nu(ix)$ (modified Bessel function)

(III) Transformation Groups and Invariants

Let $U\equiv\xi\dfrac{\partial}{\partial x}+\eta\dfrac{\partial}{\partial y}$ be the infinitesimal transformation of a given continuous transformation group of two variables, and $U'\equiv\xi\dfrac{\partial}{\partial x}+\eta\dfrac{\partial}{\partial y}+\zeta\dfrac{\partial}{\partial p}$ be that of its extended group.

We have

$$\zeta=\frac{\partial\eta}{\partial x}+p\left(\frac{\partial\eta}{\partial y}-\frac{\partial\xi}{\partial x}\right)-p^2\frac{d\xi}{\partial x}.$$

We put

$$p\equiv\frac{dy}{dx},\qquad r\equiv\frac{d^2y}{dx^2}.$$

Let α, β and γ be invariants of the 0th, first, and second order, respectively. The general form of the diffferential equation of the first or of the second order invariant under U is given by $\Phi(\alpha,\beta)=0$ (or $\beta=F(\alpha)$), and $\Psi(\alpha,\beta,\gamma)=0$ (or $\gamma=G(\alpha,\beta)$), respectively, where F, Φ, Ψ, G denote arbitrary functions of the corresponding variables.

Group With Infinitesimal Transformation U			Invariants			Note
ξ	η	ζ	0th	1st	2nd	
0	1	0	x	p	r	(1)
1	0	0	y	p	r	(1)
$-y$	x	$1+p^2$	x^2+y^2	$(y-xp)/(x+yp)$	$r/(1+p^2)^{3/2}$	(2)
0	y	p	x	p/y	r/y	(3)
x	0	$-p$	y	xp	x^2r	(3)
x	y	0	y/x	p	xr	
x	$-y$	$-2p$	xy	x^2p	x^3r	(4)
μx	νy	$(\nu-\mu)p$	y^μ/x^ν	$x^{1-\nu/\mu}p$ or px/y	$r\mu/x^{\nu-\mu-1}$	
μ	ν	0	$\nu x-\mu y$	p	r	
0	$h(x)$	$h'(x)$	x	$h(x)p-h'(x)y$	$h(x)r-h''(x)y$	
$k(y)$	0	$-k'(y)p^2$	y	$\dfrac{1}{p}-\dfrac{k'(y)}{k(y)}x$	$\dfrac{r}{p^3}+\dfrac{k''(y)}{k'(y)p}$	(5)
0	$k(y)$	$k'(y)p$	x	$\dfrac{p}{k(y)}$	$\dfrac{r}{k(y)}-\dfrac{k'(y)p^2}{[k(y)]^2}$	(6)
$h(x)$	0	$-h'(x)p$	y	$h(x)p$	$(h(x))^2r+h(x)h'(x)p$	

Group With Infinitesimal Transformation U			Invariants			Note
ξ	η	ζ	0th	1st	2nd	
0	$h(x)k(y)$	$h'(x)k(y)$ $+h(x)k'(y)p$	x	$\dfrac{p}{k(y)}-\dfrac{h'(x)}{h(x)}\int^y\dfrac{dy}{k(y)}$	—	(7)
$xh(x)$	$yh(x)$	$h'(x)(y-xp)$	$\dfrac{y}{x}$	$\left(p-\dfrac{y}{x}\right)h(x)$	$\left(\dfrac{x^2r}{xp-y}-1\right)h(x)$ $+h'(x)$	
y	x	$1-p^2$	x^2-y^2	$\dfrac{1-p}{1+p}\dfrac{x+y}{x-y}$ or $(x-yp)/(1+p)(x-y)$	$\dfrac{r}{(1-p^2)^{3/2}}$	

Notes
(1) Parallel translation.
(2) Rotation.
(3) Affine transformation.
(4) Similar transformation; the equation is a homogeneous differential equation.
(5) Linear differential equation.
(6) Separated variable type.
(7) When $k(y)=y^n$, the equation is Bernoulli's differential equation.

Reference

[1] A. R. Forsyth, A treatise on differential equations, Macmillan, fourth edition, 1914.

15. Total and Partial Differential Equations

(I) Total Differential Equations (→ 428 Total Differential Equations)

Suppose we are given a system of total differential equations

$$dz_j=\sum_{k=1}^{n}P_{jk}(x;z)\,dx_k\quad (j=1,2,\ldots,m).$$

A condition for complete integrability is given by

$$\frac{\partial P_{jk}(x;z)}{\partial x_l}+\sum_i\frac{\partial P_{jk}(x;z)}{\partial z_i}P_{il}(x;z)=\frac{\partial P_{jl}(x;z)}{\partial x_k}+\sum_i\frac{\partial P_{jl}(x;z)}{\partial z_i}P_{ik}(x;z).$$

Under this condition, the solution with the initial condition $(x_1^0,\ldots,x_n^0;z_1^0,\ldots,z_m^0)$ is obtained as follows: First, solve the system of differential equations $dz_j/dx_1=P_{j1}(x_1,x_2^0,\ldots,x_n^0;z)$ in x_1 with the initial condition $z_j(x_1^0)=z_j^0$, and denote the solution by $z_j=\varphi_j(x_1)$. Next, considering x_1 as a parameter, solve the system of differential equations $dz_j/dx_2=P_{j2}(x_1,x_2,x_3^0,\ldots,x_n^0;z)$ in x_2 with the initial condition $z_j(x_2^0)=\varphi_j(x_1)$, and denote the solution by $z_j=\varphi_j(x_1,x_2)$. Repeat the process, until we finally have $z_j=\varphi_j(x_1,\ldots,x_n)$, which is the solution of the original equation. Or, if we have m independent first integrals $f_j(x;z)=c_j$ of the equation $dz_j/dx_1=P_{j1}(x;z)$, we may transform the equation into $du_j=\sum_{k=1}^n Q_{jk}(x;u)dx_k$ by the transformation $u_j=f_j(x;z)$. Since the $Q_{jk}(x;u)$ do not involve x_1 and the equation is a completely integrable total differential equation, we have reduced the number of variables. We obtain the general solution by repeating this process n times.

For

$$P(x,y,z)\,dx+Q(x,y,z)\,dy+R(x,y,z)\,dz=0$$

$(n=3,m=1)$, the complete integrability condition is

$$P\left(\frac{\partial Q}{\partial z}-\frac{\partial R}{\partial y}\right)+Q\left(\frac{\partial R}{\partial x}-\frac{\partial P}{\partial z}\right)+R\left(\frac{\partial P}{\partial y}-\frac{\partial Q}{\partial x}\right)=0.$$

(II) Solution of Partial Differential Equations of First Order (→ 322 Partial Differential Equations (Methods of Integration), 324 Partial Differential Equations of First Order)

Let z be a function of x and y, and

$$p \equiv \partial z/\partial x, \quad q \equiv \partial z/\partial y, \quad r \equiv \partial^2 z/\partial x^2, \quad s \equiv \partial^2 z/\partial x \partial y, \quad t \equiv \partial^2 z/\partial y^2.$$

We consider a partial differential equation of the first order $F(x,y,z,p,q)=0$.

(1) The Lagrange-Charpit Method. We consider the auxiliary equation

$$\frac{dx}{F_p}=\frac{dy}{F_q}=\frac{dz}{pF_p+qF_q}=\frac{-dp}{F_x+pF_z}=\frac{-dq}{F_y+qF_z},$$

which is a system of ordinary differential equations. Let $G(x,y,z,p,q)=a$ be the solution of the auxiliary equation. Using this together with the original equation $F=0$, we obtain $p=P(x,y,z,a)$, $q=Q(x,y,z,a)$, and the complete solution by integrating $dz=P\,dx+Q\,dy$. If we know another solution of the auxiliary equation $H(x,y,z,p,q)=b$ independent of $G=a$, we have the complete solution $z=\Phi(x,y,a,b)$ by eliminating p and q from $F=0$, $G=a$, and $H=b$.

(2) Solution of Various Standard Forms of Partial Differential Equations of the First Order. The integration constants are a, b.
(i) $f(p,q)=0$. The complete solution is $z=ax+\varphi(a)y+b$, where the function $t=\varphi(a)$ is defined by $f(t,a)=0$.
(ii) $f(px,q)=0$, $f(x,qy)=0$, $f(p/z,q/z)=0$. These equations reduce to (i) if $x=e^X$, $y=e^Y$, $z=e^Z$, respectively.
(iii) $f(x,p,q)=0$. If we can solve for $p=F(x,q)$, the complete solution is $z=\int F(x,a)dx+ay+b$. A similar procedure applies to $f(y,p,q)=0$.
(iv) $f(z,p,q)=0$. Solve $f(z,t,at)=0$ for $t=F(z,a)$. The complete solution is then given by $x+ay+b=\int dz/F(z,a)$. If we eliminate a and b from the complete solution $\Phi(x,y,z,a,b)=0$ and $\partial\Phi/\partial a=\partial\Phi/\partial b=0$, we have the singular solution of the original equation.
(v) Separated variable type $f(x,p)=g(y,q)$. Solve the two ordinary differential equations $f(x,p)=a$ and $g(y,q)=a$ for the solutions $p=P(x,a)$ and $q=Q(y,a)$, respectively. Then the complete solution is $z=\int P(x,a)dx+\int Q(y,a)dy+b$.
(vi) Lagrange's partial differential equation $Pp+Qq=R$. Here P, Q, R are functions of x, y, and z. Denote the solutions of the system of differential equations $dx:dy:dz=P:Q:R$ by $u(x,y,z)=a$, $v(x,y,z)=b$. Then the general solution is $\Phi(u,v)=0$, where Φ is an arbitrary function. A similar method is applicable to

$$\sum_{j=1}^{n} P_j(x_1,\ldots,x_n)\frac{\partial z}{\partial x_j}=R(x_1,\ldots,x_n).$$

If we have n independent solutions $u_j(x)=a_j$ of a system of n differential equations $dx_j/P_j=dz/R$ $(j=1,\ldots,n)$, the general solution is given by $\Phi(u_1,\ldots,u_n)=0$.
(vii) Clairaut's partial differential equation $z=px+qy+f(p,q)$. The complete solution is given by the family of planes $z=ax+by+f(a,b)$. The singular solution as the envelope of the family of planes is given by eliminating p and q from the original equation and $x=-\partial f/\partial p$ and $y=-\partial f/\partial q$.

(III) Solutions of Partial Differential Equations of Second Order (→ 322 Partial Differential Equations (Methods of Integration))

(1) Quadrature. Here φ and ψ are arbitrary functions.
(i) $r=f(x)$. The general solution is $z=\int\int f(x)dx\,dx+\varphi(y)x+\psi(y)$. A similar rule applies to $t=f(y)$.
(ii) $s=f(x,y)$. The general solution is $z=\int\int f(x,y)dx\,dy+\varphi(x)+\psi(y)$.
(iii) Wave equation. $r-t=0$. The general solution is $z=\varphi(x+y)+\psi(x-y)$.
(iv) Laplace's differential equation. $r+t=0$. Let $x+iy=\zeta$ and φ, ψ be complex analytic functions of ζ. The general solution is $z=\varphi(\zeta)+\psi(\bar\zeta)$, and a real solution is $z=\varphi(\zeta)+\bar\varphi(\bar\zeta)$.
(v) $r+Mp=N$, where M and N are functions of x and y. The general solution is given by $z=\int[\int L(x,y)N(x,y)dx+\varphi(y)]/L(x,y)dx+\psi(y)$, $L(x,y)=\exp\int[M(x,y)dx]$. In the integration, y is considered a constant.
A similar method is applicable to $s+Mp=N$, $s+Mq=N$, and $t+Mq=N$.
(vi) Monge-Ampère partial differential equation. $Rr+Ss+Tt+U(rt-s^2)=V$, where R, S, T, U, V are functions of x, y, z, p, q.

First, in the case $U=0$, we take auxiliary equations

eq. (1) $\quad R\,dy^2+T\,dx^2-S\,dx\,dy=0,$

eq. (2) $\quad R\,dp\,dy+T\,dq\,dx=V\,dx\,dy.$

Equation (1) is decomposed into two linear differential forms $X_i\,dx+Y_i\,dy=0$ $(i=1,2)$. The combination with (2) gives a solution $u_i(x,y,z,p,q)=a_i$, $v_i(x,y,z,p,q)=b_i$ $(i=1,2)$, and we have intermediate integrals $F_i(u_i,v_i)=0$ $(i=1,2)$ for an arbitrary function F_i. We have the solution of the original equation by solving the intermediate integrals. If $S^2\neq 4RT$, two intermediate integrals are distinct, and hence we can solve them in the form $p=P(x,y,z)$, $q=Q(x,y,z)$, and then we may integrate $dz=P\,dx+Q\,dy$.

Next, in the case $U\neq 0$, let λ_1 and λ_2 be the solutions of $U^2\lambda^2+US\lambda+TR+UV=0$. We have two auxiliary equations

$$\begin{cases}\lambda_1 U\,dy+T\,dx+U\,dp=0,\\ \lambda_2 U\,dx+R\,dy+U\,dq=0,\end{cases}\quad\text{or}\quad\begin{cases}\lambda_2 U\,dy+T\,dx+U\,dp=0,\\ \lambda_1 U\,dx+R\,dy+U\,dq=0,\end{cases}$$

and from the solutions $u_i=a_i$, $v_i=b_i$ $(i=1,2)$, we have intermediate integrals $F_i(u_i,v_i)=0$ $(i=1,2)$. If $4(TR+UV)\neq S^2$, $\lambda_1\neq\lambda_2$, we have two different intermediate integrals $F_i=0$. Solving the simultaneous equations $F_i=0$ in $p=P(x,y,z)$, $q=Q(x,y,z)$, we may also find the solution by integrating $dz=P\,dx+Q\,dy$.

(vii) Poisson's differential equation. $P=(rt-s^2)^nQ$, where $P=P(p,q,r,s,t)$ is homogeneous with respect to r, s, t and we assume that $Q=Q(x,y,z)$ satisfies $\partial Q/\partial z\neq\infty$ for x, y, z when $rt=s^2$. The equation $P(p,\varphi(p),r,r\varphi'(p),r\{\varphi'(p)\}^2)=0$ is then an ordinary differential equation in φ as a function of p. We first solve this for φ, and then solve a partial differential equation of the first order $q=\varphi(p)$ by the method (II)(2)(i).

(2) Intermediate Integrals. Let $f(x,y,z,p,q,r,s,t)$ be polynomials with respect to r, s, t. Suppose that $f(x,y,z,p,q,r,s,t)=0$ has the first integral $u(x,y,z,p,q)=0$. We insert

$$r=-\left(\frac{\partial u}{\partial x}+p\frac{\partial u}{\partial z}+s\frac{\partial u}{\partial q}\right)\Big/\frac{\partial u}{\partial p},\qquad t=-\left(\frac{\partial u}{\partial y}+q\frac{\partial u}{\partial z}+s\frac{\partial u}{\partial p}\right)\Big/\frac{\partial u}{\partial q}$$

into the original equation, and replace all the coefficients that are polynomials of s by 0. We thus obtain a system of differential equations in u. If u and v are two independent solutions of this system, an intermediate integral of the original equation is given in the form $\Phi(u,v)=0$.

(3) Initial Value Problem for a Hyperbolic Partial Differential Equation $L[u]\equiv u_{xy}+au_x+bu_y+cu=h$.

$$u(\xi,\eta)=[(uR)_A+(uR)_B]/2+\iint_\Delta R(x,y;\xi,\eta)h(x,y)\,dx\,dy$$
$$+\int_A^B\left[\frac{1}{2}\left(u\frac{\partial R}{\partial n'}-R\frac{\partial u}{\partial n'}\right)-\{a\cos(n,x)+b\cos(n,y)\}uR\right]ds,$$

where Δ is the hatched region in Fig. 19, and the conormal n' is the mirror image of the normal n with respect to $x=y$.

$$u(\xi,\eta)=(uR)_C+\int_C^A R(u_y+au)\,dy+\int_C^B R(u_x+bu)\,dx+\iint_\square R(x,y;\xi,\eta)h(x,y)\,dx\,dy$$

(characteristic initial value problem).

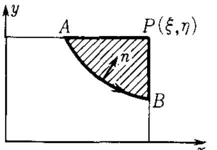


Fig. 19

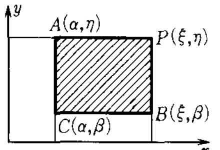


Fig. 20

Here $\square$ is the hatched rectangular region in Fig. 20. $R(x,y;\xi,\eta)$ is the Riemann function; it satisfies

$M[R(x,y;\xi,\eta)]=0,$

$R_x-bR=0$ (on $x=\xi$),

$R_y-aR=0$ (on $y=\eta$),

$R(\xi,\eta;\xi,\eta)=1.$

Example (i). $u_{xy}=h(x,y)$. $R(x,y;\xi,\eta)=1$.

$$u(\xi,\eta)=\frac{1}{2}[u_A+u_B]+\frac{1}{2}\int_A^B\left[u_y\cos(n,x)+u_x\cos(n,y)\right]ds+\iint_\Delta h(x,y)\,dx\,dy.$$

Example (ii). Telegraph equation $u_{xy}+cu=0$ $(c>0)$. $R(x,y;\xi,\eta)=J_0\left(2\sqrt{c(x-\xi)(y-\eta)}\right)$.

Example (iii). $u_{xy}+\dfrac{n}{x+y}(u_x+u_y)=0$ ($n=$ a constant >0).

$$R(x,y;\xi,\eta)=\left(\frac{x+y}{\xi+\eta}\right)^n F\left(1-n,n;1;-\frac{(x-\xi)(y-\eta)}{(x+y)(\xi+\eta)}\right).$$

(IV) Contact Transformations (→ 82 Contact Transformations)

We consider a transformation $(x_1,\ldots,x_n;z)\to(X_1,\ldots,X_n;Z)$. We put $p_j\equiv\partial z/\partial x_j$, $P_j\equiv\partial Z/\partial X_j$ $(j=1,\ldots,n)$. The transformation is called a contact transformation if there exists a function $\rho(x,z,p)\neq 0$ satisfying $dZ-\Sigma P_j dX_j=\rho(x,z,p)(dz-\Sigma p_j dx_j)$.

A transformation given by $(2n+1)$ equations $\Omega=0$, $\partial\Omega/\partial X_j+P_j\partial\Omega/\partial Z=0$, $\partial\Omega/\partial x_j+p_j\partial\Omega/\partial z=0$ generated by a generating function $\Omega(x,z,X,Z)$ is a contact transformation.

Generating Function	ρ	Transformation	Name
Σx_jX_j+z+Z	-1	$X_j=-p_j,\quad P_j=-x_j,$ $Z=\Sigma p_jx_j-z$	Legendre's transformation
$\Sigma X_j^2+Z^2-\Sigma x_jX_j-zZ$	$Z/(2Z-z)$	$X_j=-p_jZ,$ $p_j=-(2X_j-x_j)/(2Z-z)$	Pedal transformation
$\Sigma(X_j-x_j)^2+(Z-z)^2-a^2$	1	$X_j=x_j-ap_j(1+\Sigma p_j^2)^{-1/2},$ $P_j=p_j,$ $Z=z_j+a(1+\Sigma p_j^2)^{-1/2}$	Similarity
$\Sigma(X_j-x_j)^2-Z^2-z^2$	$-\dfrac{1}{\sqrt{\Sigma p_j^2-1}}$	$X_j=x_j-p_jz,$ $P_j=-p_j(\Sigma p_j^2-1)^{-1/2},$ $Z=z(\Sigma p_j^2-1)^{1/2}$	

(V) Fundamental Solutions (→ 320 Partial Differential Equations H)

A function (or a generalized function such as a distribution) T satisfying $LT=\delta$ (δ is the Dirac delta function) for a linear differential operator L is called the fundamental (or elementary) solution of L. In the following table, we put

$$\Delta\equiv\sum_{i=1}^n\frac{\partial^2}{\partial x_i^2},\quad \square\equiv\frac{\partial^2}{\partial x_n^2}-\sum_{i=1}^{n-1}\frac{\partial^2}{\partial x_i^2},\quad r^2\equiv\sum_{i=1}^n x_i^2,\quad \mathbf{1}(x)=\begin{cases}1 & (x>0)\\ 0 & (x\leqslant 0)\end{cases}\quad \text{(Heaviside function)}.$$

J_ν is the Bessel function of the first kind; K_ν and I_ν are the modified Bessel functions. (→Table 19.IV, this Appendix.)

$$s=\begin{cases}\sqrt{x_n^2-x_1^2-\ldots-x_{n-1}^2} & \text{(if } x_n>0 \text{ and the quantity under the radical sign is positive),}\\ 0 & \text{(otherwise).}\end{cases}$$

(For Pf (finite part) → 125 Distributions and Hyperfunctions.)

Operator	Fundamental Solution
d/dx	$\mathbf{1}(x)$
$\dfrac{d^m}{dx^m}$	$\begin{cases}x^{m-1}/(m-1)! & (x>0)\\ 0 & (x\leqslant 0)\end{cases}$
$\partial^n/\partial x_1\partial x_2\ldots\partial x_n$	$\mathbf{1}(x_1)\mathbf{1}(x_2)\ldots\mathbf{1}(x_n)$
$\dfrac{\partial}{\partial x}+i\dfrac{\partial}{\partial y}$	$\dfrac{1}{2\pi}\dfrac{1}{x+iy}\quad(i\equiv\sqrt{-1})$

Operator	Fundamental Solution
Δ	$\begin{cases} -\left[\Gamma\left(\frac{n}{2}\right)/2(n-2)\pi^{n/2}\right]\frac{1}{r^{n-2}} & (n\geq 3) \\ (1/2\pi)\log r & (n=2) \end{cases}$
Δ^m	$\begin{cases} \left[\Gamma\left(\frac{n}{2}\right)/2^m(m-1)!\pi^{n/2}\prod_{k=1}^{m}(2k-n)\right]\frac{1}{r^{n-2m}} & \begin{pmatrix} n-2m \text{ is a positive integer} \\ \text{or a negative odd integer} \end{pmatrix} \\ \left[\Gamma\left(\frac{n}{2}\right)/2^m(m-1)!\pi^{n/2}\prod_{\substack{k=1 \\ k\neq m-h}}^{m}(2k-n)\right]\frac{\log r}{r^{n-2m}} & \begin{pmatrix} n-2m=-2h, \\ h=0,1,2,\ldots \end{pmatrix} \end{cases}$
$\left(1-\frac{\Delta}{4\pi^2}\right)^m$	$\frac{2\pi^m}{(m-1)!}r^{m-(n/2)}K_{(n/2)-m}(2\pi r)$
$\square^m$	$(\mathrm{Pf}\, s^{2m-n})/\pi^{(n/2)-1}2^{2m-1}(m-1)!\Gamma[m+1-(n/2)]$
$(\square-\lambda)^m$ (λ is real and $\neq 0$)	$\frac{\lvert\lambda\rvert^{(n/4)-(m/2)}}{\pi^{(n/2)-1}(m-1)!2^{m-1+(n/2)}}\mathrm{Pf}\, s^{m-(n/2)}\begin{cases} I_{m-(n/2)}(\sqrt{\lambda s}) & (\lambda>0) \\ J_{m-(n/2)}(\sqrt{\lvert\lambda\rvert s}) & (\lambda<0) \end{cases}$
$\left(\frac{\partial}{\partial x_n}-\sum_{i=1}^{n-1}\frac{\partial^2}{\partial x_i^2}\right)^m$	$\begin{cases} \frac{x_n^{m-1}}{(m-1)!}\left(\frac{1}{2\sqrt{\pi x_n}}\right)^{n-1}\exp\left(-\sum_{i=1}^{n-1}x_i^2/4x_n\right) & (x_n>0) \\ 0 & (x_n\leq 0) \end{cases}$

(VI) Solution of Boundary Value Problems (→ 188 Green's Functions, 323 Partial Differential Equations of Elliptic Type, 327 Partial Differential Equations of Parabolic Type)

$$L[u]=Au_{xx}+2Bu_{xy}+Cu_{yy}+Du_x+Eu_y+Fu,$$

$$M[v]=(Av)_{xx}+2(Bv)_{xy}+(Cv)_{yy}-(Dv)_x-(Ev)_y+Fv.$$

Green's formula $$\int\int_D\{vL[u]-uM[v]\}\,dx\,dy=\int_C\left\{P\left(u\frac{\partial v}{\partial n'}-v\frac{\partial u}{\partial n'}\right)+Quv\right\}ds.$$

eq. (1) $$\begin{cases} A\cos(n,x)+B\cos(n,y)=P\cos(n',x), \\ B\cos(n,x)+C\cos(n,y)=P\cos(n',y). \end{cases}$$

$$Q=(A_x+B_y-D)\cos(n,x)+(B_x+C_y-E)\cos(n,y).$$

The integration contour C is the boundary of the domain D (Fig. 21), n is the inner normal of C, and n', called the conormal, is given by (1).

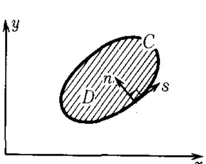


Fig. 21

(1) Elliptic Partial Differential Equation $L[u]\equiv u_{xx}+u_{yy}+au_x+bu_y+cu=h$.

$$u(\xi,\eta)=-\int_C u(s)\frac{\partial G}{\partial n}ds+\int\int_D G(x,y;\xi,\eta)h(x,y)\,dx\,dy.$$

Here $G(x,y;\xi,\eta)$ is Green's function, which satisfies $M(G(x,y;\xi,\eta))=0$ in the interior of D except at $(x,y)\neq(\xi,\eta)$, and

$$G(x,y;\xi,\eta)=-(1/2\pi)\log\sqrt{(x-\xi)^2+(y-\eta)^2}\ +\text{ a regular function},$$

$$G(x,y;\xi,\eta)=0\quad((x,y)\in C).$$

(2) Laplace's Differential Equation in the 2-Dimensional Case $\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=0.$

$u(x,y)\equiv\tilde{u}(r,\varphi)=\operatorname{Re}f(z) \quad (z\equiv x+iy=re^{i\theta}).$

(i) Interior of a disk ($r\leqslant 1$).

$$f(z)=\frac{1}{2\pi}\int_0^{2\pi}\tilde{u}(1,\varphi)\frac{e^{i\varphi}+z}{e^{i\varphi}-z}d\varphi, \qquad \tilde{u}(r,\theta)=\frac{1}{2\pi}\int_0^{2x}\tilde{u}(1,\varphi)\frac{1-r^2}{1-2r\cos(\theta-\varphi)+r^2}d\varphi$$

(Poisson's integration formula).

(ii) Annulus ($0<q\leqslant r\leqslant 1$).

$$f(z)=\frac{\omega_1}{\pi^2 i}\left\{\int_0^{2\pi}\tilde{u}(1,\varphi)\zeta_1(w)\,d\varphi-\int_0^{2\pi}\tilde{u}(q,\varphi)\zeta_3(w)\,d\varphi-a\log z\right\}$$

(Villat's integration formula).

$$w=\frac{\omega_1}{\pi}(i\log z+\varphi), \quad a=\left(\frac{1}{2\omega_3}-\frac{\eta_1}{\pi i}\right)\int_0^{2\pi}\{\tilde{u}(1,\varphi)-\tilde{u}(q,\varphi)\}\,d\varphi, \quad \frac{\omega_3}{\omega_1}=-\frac{i}{\pi}\log q.$$

Here ζ_1 and ζ_3 are the Weierstrass ζ-functions (→ 134 Elliptic Functions) with the fundamental periods $2\omega_1$ and $2\omega_3$.

(iii) Half-plane ($y\geqslant 0$). $f(z)=\frac{i}{\pi}\int_{-\infty}^{\infty}\frac{u(t,0)}{z-t}dt, \quad u(x,y)=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{u(t,0)y}{(x-t)^2+y^2}dt.$

(3) Laplace's Differential Equation in the 3-Dimensional Case.

(i) Interior of a sphere ($r\leqslant 1$).

$$\tilde{u}(r,\varphi,\theta)=\frac{1}{4\pi}\int_0^{\pi}\int_0^{2\pi}\tilde{u}(1,\Phi,\Theta)\frac{1-r^2}{(1-2r\cos\gamma+r^2)^{3/2}}\sin\Theta\,d\Theta\,d\Phi,$$

where

$$\cos\gamma=\cos\Theta\cos\theta+\sin\Theta\sin\theta\cos(\Phi-\varphi).$$

(ii) Half-space ($z\geqslant 0$).

$$u(x,y,z)=\frac{z}{2\pi}\int\int_{-\infty}^{\infty}\frac{u(\xi,\eta,0)}{\{(x-\xi)^2+(y-\eta)^2+z^2\}^{3/2}}d\xi\,d\eta.$$

(4) Equation of Oscillation (Helmholtz Differential Equation) $\Delta u+k^2u=0$. Let u_n be the normalized eigenfunction with the same boundary condition for the eigenvalue k_n. Green's function is

$$G(P,Q)=\sum\frac{u_n(P)u_n^*(Q)}{k^2-k_n^2}$$

Domain	Boundary Condition	Eigenvalue	Eigenfunction
rectangle $0\leqslant x\leqslant a,\ 0\leqslant y\leqslant b$	$u=0$	$k_{nm}=\pi\sqrt{\frac{n^2}{a^2}+\frac{m^2}{b^2}}$ $(n,\ m=1,2,\ldots)$	$\sin n\pi\frac{x}{a}\sin m\pi\frac{y}{b}$
circle $0\leqslant r\leqslant a$	$u=0$	k_{nm} is the root of $J_m(kx)=0$	$J_m(k_{nm}r)e^{\pm im\varphi}$
annulus $b\leqslant r\leqslant a$	$u=0$	k_{nm} is the root of $J_m(ka)N_m(kb)-J_m(kb)N_m(ka)=0$	$\left\{\frac{J_m(k_{nm}r)}{J_m(k_{nm}a)}-\frac{N_m(k_{nm}r)}{N_m(k_{nm}a)}\right\}e^{\pm im\varphi}$
fan shape $0\leqslant r\leqslant a,\ 0\leqslant\varphi\leqslant\alpha$	$u=0$	k_{nm} is the root of $J_\mu(ka)=0$ $(\mu=m\pi/\alpha)$	$J_\mu(k_{nm}r)\sin\mu\varphi$
rectangular parallelepiped $0\leqslant x\leqslant a,\ 0\leqslant y\leqslant b,\ 0\leqslant z\leqslant c$	$\frac{\partial u}{\partial n}=0$	$k_{nml}=\pi\sqrt{\frac{n^2}{a^2}+\frac{m^2}{b^2}+\frac{l^2}{c^2}}$	$\cos n\pi\frac{x}{a}\cos m\pi\frac{y}{b}\cos l\pi\frac{z}{c}$
sphere $0\leqslant r\leqslant a$	$\frac{\partial u}{\partial n}=0$	k_{nl} is the root of $\psi_n'(ka)=0$, where $\psi_n(\rho)\equiv\sqrt{\pi/2}\,J_{n+(1/2)}(\rho)$	$\psi_n(k_{nl}r)P_n^m(\cos\theta)^{\pm im\varphi}$

(5) Heat Equation. $\frac{\partial u}{\partial t}=\kappa\Delta u\left(\Delta=\frac{\partial^2}{\partial x_1^2}+\ldots+\frac{\partial^2}{\partial x_m^2};\ \kappa\text{ is a positive constant}\right)$. Boundary condi-

tion: $hu-k\,\partial u/\partial n=\varphi$, where h and k are nonnegative constants with $h+k=1$, and φ is a given function.

$$u(P,t)=\int_V G(P,Q,t)u(Q,0)\,dV_Q+\kappa\int_0^t d\tau\int_S\left\{\frac{\partial G(P,Q,t-\tau)}{\partial n_Q}+G(P,Q,t-\tau)\right\}\varphi(Q,\tau)\,dS_Q.$$

Here V is the domain, and S is its boundary. $G(P,Q,t)$ is the elementary solution that satisfies $\partial G/\partial t=\kappa\Delta G$ in V and $k\partial G/\partial n=hG$ on S, and further in the neighborhood of $P=Q$, $t=0$, it has the form $G(P,Q,t)=(4\pi\kappa t)^{-m/2}e^{-R^2/4\kappa t}$ + terms of lower degree ($R=\overline{PQ}$).

(i) $-\infty<x<\infty$, $G=U(x-\xi,t)$, where $U(x,t)=e^{-x^2/4\kappa t}/\sqrt{4\pi\kappa t}$ (similar in the following case (ii)).

(ii) $0\leqslant x<\infty$. $u(0,t)=0$: $G=U(x-\xi,t)-U(x+\xi,t)$.

$\dfrac{\partial u}{\partial x}=hu$: $G=U(x-\xi,t)+U(x+\xi,t)-2he^{h\xi}\displaystyle\int_{-\infty}^{-\xi}e^{h\eta}U(x-\eta,t)\,d\eta.$

(iii) $0\leqslant x\leqslant l$. $u(0,t)=u(l,t)=0$: $G=\vartheta\left(\dfrac{x-\xi}{2l}\Big|\tau\right)-\vartheta\left(\dfrac{x+\xi}{2l}\Big|\tau\right).$

$\dfrac{\partial u}{\partial x}(0,t)=\dfrac{\partial u}{\partial x}(l,t)=0$: $G=\vartheta\left(\dfrac{x-\xi}{2l}\Big|\tau\right)+\vartheta\left(\dfrac{x+\xi}{2l}\Big|\tau\right).$

$u(0,t)=\dfrac{\partial u}{\partial x}(l,t)=0$:

$$G=\vartheta\left(\frac{x-\xi}{4l}\Big|\tau\right)-\vartheta\left(\frac{x+\xi}{4l}\Big|\tau\right)+\vartheta\left(\frac{x+\xi-2l}{4l}\Big|\tau\right)-\vartheta\left(\frac{x-\xi-2l}{4l}\Big|\tau\right).$$

Here ϑ is the elliptic theta function: $\vartheta(x|\tau)=\vartheta_3(x,\tau)\equiv1+2\sum e^{i\pi\tau n^2}\cos 2n\pi x$.

(iv) $0\leqslant x<\infty$, $0\leqslant y<\infty$. $u(x,0,t)=u(0,y,t)=0$:

$$G=\left(e^{-(x-\xi)^2/4\kappa t}-e^{-(x+\xi)^2/4\kappa t}\right)\left(e^{-(y-\eta)^2/4\kappa t}-e^{-(y+\eta)^2/4\kappa t}\right)/4\pi\kappa t.$$

(v) $0\leqslant x\leqslant a$, $0\leqslant y\leqslant b$. $u=0$ on the boundary:

$$G=\frac{4}{ab}\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\exp\left\{-\kappa\pi^2t\left(\frac{m^2}{a^2}+\frac{n^2}{b^2}\right)\right\}\sin\frac{m\pi x}{a}\sin\frac{m\pi\xi}{a}\sin\frac{n\pi y}{b}\sin\frac{n\pi\eta}{b}.$$

(vi) $0\leqslant x\leqslant a$, $0\leqslant y\leqslant b$, $0\leqslant z\leqslant c$. $u=0$ on the boundary:

$$G=\frac{8}{abc}\sum_{l=1}^{\infty}\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\exp\left\{-\kappa\pi^2t\left(\frac{l^2}{a^2}+\frac{m^2}{b^2}+\frac{n^2}{c^2}\right)\right\}\times\sin\frac{l\pi x}{a}\sin\frac{l\pi\xi}{a}\sin\frac{m\pi y}{b}\sin\frac{m\pi\eta}{b}\sin\frac{n\pi z}{c}\sin\frac{n\pi\zeta}{c}.$$

(vii) $0\leqslant r<\infty$. Spherically symmetric. $|x|=r$, $|\xi|=r'$:

$$G=\left(e^{-(r-r')^2/4\kappa t}-e^{-(r+r')^2/4\kappa t}\right)/8\pi rr'(\pi\kappa t)^{1/2}.$$

(viii) $0\leqslant r\leqslant a$. Spherically symmetric. $u=0$ on the boundary:

$$G=\frac{1}{2\pi arr'}\sum_{n=1}^{\infty}e^{-\kappa n^2\pi^2t/a^2}\sin\frac{n\pi r}{a}\sin\frac{n\pi r'}{a}.$$

(ix) $a\leqslant r<\infty$. Spherically symmetric. $k\partial u/\partial r-hu=0$ on the boundary:

$$G=\frac{1}{8\pi rr'(\pi\kappa t)^{1/2}}\left[e^{-(r-r')^2/4\kappa t}+e^{-(r+r'-2a)^2/4\kappa t}-\frac{ah+k}{ak}(4\pi\kappa t)^{1/2}\times\exp\left\{\kappa t\left(\frac{ah+k}{ak}\right)^2+(r+r'-2a)\frac{ah+k}{ak}\right\}\times\operatorname{erfc}\left\{\frac{r+r'-2a}{2\sqrt{\kappa t}}+\frac{ah+k}{ak}\sqrt{\kappa t}\right\}\right]\quad\left(\operatorname{erfc}x\equiv\int_x^{\infty}e^{-t^2}dt\right).$$

(x) $0\leqslant r<\infty$. Axially symmetric: $G=e^{-(r^2+r'^2)/4\kappa t}I_0(rr'/2\kappa t)/4\pi\kappa t$.

(xi) $0\leqslant r\leqslant a$. Axially symmetric. $k\dfrac{\partial u}{\partial r}-hu=0$ on the boundary:

$$G=\frac{1}{\pi a^2}\sum_{n=1}^{\infty}\frac{J_0(r\alpha_n)J_0(r'\alpha_n)}{\{J_0(a\alpha_n)\}^2+\{J_1(a\alpha_n)\}^2}e^{-\kappa\alpha_n t/a},$$

where α_n is given by $k\alpha_nJ_1(a\alpha_n)-hJ_0(a\alpha_n)=0$.

16. Elliptic Integrals and Elliptic Functions

(I) Elliptic Integrals (→ 134 Elliptic Functions)

(1) Legendre-Jacobi Standard Form.

Elliptic integral of the first kind

$$F(k,\varphi)\equiv\int_0^{\varphi}\frac{d\psi}{\sqrt{1-k^2\sin^2\psi}}=\int_0^{\sin\varphi}\frac{dt}{\sqrt{(1-t^2)(1-k^2t^2)}}\qquad (k \text{ is the modulus}).$$

Elliptic integral of the second kind

$$E(k,\varphi)\equiv\int_0^{\varphi}\sqrt{1-k^2\sin^2\psi}\,d\psi=\int_0^{\sin\varphi}\sqrt{\frac{1-k^2t^2}{1-t^2}}\,dt.$$

Elliptic integral of the third kind

$$\Pi(\varphi,n,k)=\int_0^{\varphi}\frac{d\psi}{(1+n\sin^2\psi)\sqrt{1-k^2\sin^2\psi}}=\int_0^{\sin\varphi}\frac{dt}{(1+nt^2)\sqrt{(1-t^2)(1-k^2t^2)}}.$$

When $\varphi=\pi/2$, elliptic integrals of the first and the second kinds are called complete elliptic integrals:

$$K(k)\equiv F\left(k,\frac{\pi}{2}\right)=\int_0^{\pi/2}\frac{d\psi}{\sqrt{1-k^2\sin^2\psi}}=\int_0^{1}\frac{dt}{\sqrt{(1-t^2)(1-k^2t^2)}}=\frac{\pi}{2}F\left(\frac{1}{2},\frac{1}{2};1;k^2\right),$$

$$E(k)\equiv E\left(k,\frac{\pi}{2}\right)=\int_0^{\pi/2}\sqrt{1-k^2\sin^2\psi}\,d\psi=\int_0^{1}\sqrt{\frac{1-k^2t^2}{1-t^2}}\,dt=\frac{\pi}{2}F\left(-\frac{1}{2},\frac{1}{2};1;k^2\right),$$

where F is the hypergeometric function.

$$K(k')=K(\sqrt{1-k^2})\equiv K'(k),\qquad E(k')=E(\sqrt{1-k^2})\equiv E'(k)\qquad (k'^2=1-k^2;\ k' \text{ is the complementary modulus}).$$

$$EK'+E'K-KK'=\frac{\pi}{2}\ \text{(Legendre's relation)}.\qquad K\left(\frac{1}{\sqrt{2}}\right)=\frac{\Gamma(1/4)^2}{4\sqrt{\pi}}.$$

$$\frac{\partial F}{\partial k}=\frac{1}{k'^2}\left(\frac{E-k'^2F}{k}-\frac{\sin\varphi\cos\varphi}{\sqrt{1-k^2\sin^2\varphi}}\right),\qquad \frac{\partial E}{\partial k}=\frac{E-F}{k}.$$

(2) Change of Variables.

$$\tan(\psi-\varphi)=k'\tan\varphi:\quad F\left(\frac{1-k'}{1+k'},\psi\right)=(1+k')F(k,\varphi),$$

$$E\left(\frac{1-k'}{1+k'},\psi\right)=\frac{2}{1+k'}[E(k,\varphi)+k'F(k,\varphi)]-\frac{1-k'}{1+k'}\sin\psi.$$

$$\sin\chi=\frac{(1+k)\sin\varphi}{1+k\sin^2\varphi}:\quad F\left(\frac{2\sqrt{k}}{1-k},\chi\right)=(1+k)F(k,\varphi),$$

$$E\left(\frac{2\sqrt{k}}{1+k},\chi\right)=\frac{1}{1+k}\left[2E(k,\varphi)-k'^2F(k,\varphi)+2k\frac{\sin\varphi\cos\varphi}{1+k\sin^2\varphi}\sqrt{1-k^2\sin^2\varphi}\right].$$

k_1	$\sin\varphi_1$	$\cos\varphi_1$	$F(k_1,\varphi_1)$	$E(k_1,\varphi_1)$
$i\frac{k}{k'}$	$k'\frac{\sin\varphi}{\sqrt{1-k^2\sin^2\varphi}}$	$\frac{\cos\varphi}{\sqrt{1-k^2\sin^2\varphi}}$	$k'F(k,\varphi)$	$\frac{1}{k'}\left[E(k,\varphi)-k^2\frac{\sin\varphi\cos\varphi}{\sqrt{1-k^2\sin^2\varphi}}\right]$
k'	$-i\tan\varphi$	$\frac{1}{\cos\varphi}$	$-iF(k,\varphi)$	$i[E(k,\varphi)-F(k,\varphi)-\sqrt{1-k^2\sin^2\varphi}\ \tan\varphi]$
$\frac{1}{k}$	$k\sin\varphi$	$\sqrt{1-k^2\sin^2\varphi}$	$kF(k,\varphi)$	$\frac{1}{k}[E(k,\varphi)-k'^2F(k,\varphi)]$

(3) Transformation into Standard Form.

(i) The following are reducible to elliptic integrals of the first kind (we assume $a>b>0$ for parameters).

$AF(k,\varphi)$	A	k	φ
$\int_1^x \frac{dt}{\sqrt{t^3-1}}$	$\frac{1}{\sqrt[4]{3}}$	$\frac{\sqrt{3}-1}{2\sqrt{2}}$	$\arccos\frac{\sqrt{3}+1-x}{\sqrt{3}-1+x}$
$\int_1^x \frac{dt}{\sqrt{1-t^3}}$	$\frac{1}{\sqrt[4]{3}}$	$\frac{\sqrt{3}+1}{2\sqrt{2}}$	$\arccos\frac{\sqrt{3}-1+x}{\sqrt{3}+1-x}$
$\int_0^x \frac{dt}{\sqrt{1+t^4}}$	$\frac{1}{2}$	$\frac{1}{\sqrt{2}}$	$\arccos\frac{1-x^2}{1+x^2}$
$\int_0^x \frac{dt}{\sqrt{(a^2-t^2)(b^2-t^2)}}$	$\frac{1}{a}$	$\frac{b}{a}$	$\arcsin\frac{x}{b}$
$\int_b^x \frac{dt}{\sqrt{(a^2-t^2)(t^2-b^2)}}$	$\frac{1}{a}$	$\sqrt{1-(b/a)^2}$	$\arcsin\sqrt{\frac{1-(b/x)^2}{1-(b/a)^2}}$
$\int_a^x \frac{dt}{\sqrt{(t^2-a^2)(t^2-b^2)}}$	$\frac{1}{a}$	$\frac{b}{a}$	$\arcsin\sqrt{\frac{1-(a/x)^2}{1-(b/x)^2}}$
$\int_0^x \frac{dt}{\sqrt{(a^2+t^2)(b^2+t^2)}}$	$\frac{1}{a}$	$\sqrt{1-(b/a)^2}$	$\arctan\frac{x}{b}$
$\int_0^x \frac{dt}{\sqrt{(a^2-t^2)(b^2+t^2)}}$	$\frac{1}{\sqrt{a^2+b^2}}$	$\frac{a}{\sqrt{a^2+b^2}}$	$\arcsin\sqrt{\frac{1+(b/a)^2}{1+(b/x)^2}}$
$\int_b^x \frac{dt}{\sqrt{(a^2+t^2)(t^2-b^2)}}$	$\frac{1}{\sqrt{a^2+b^2}}$	$\frac{a}{\sqrt{a^2+b^2}}$	$\arccos\frac{b}{x}$

(ii) The following are reducible to elliptic integrals of the second kind (we assume $a>b>0$ for parameters).

$AE(k,\varphi)$	A	k	φ
$\int_0^x \sqrt{\frac{a^2-t^2}{b^2-t^2}}\,dt$	a	$\frac{b}{a}$	$\arcsin\frac{x}{b}$
$\int_x^a \sqrt{\frac{b^2+t^2}{a^2-t^2}}\,dt$	$\sqrt{a^2+b^2}$	$\frac{a}{\sqrt{a^2+b^2}}$	$\arccos\frac{x}{a}$
$\int_b^x \frac{1}{t^2}\sqrt{\frac{t^2+a^2}{t^2-b^2}}\,dt$	$\frac{\sqrt{a^2+b^2}}{b^2}$	$\frac{a}{\sqrt{a^2+b^2}}$	$\arccos\frac{b}{x}$
$\int_\infty^x t^2\sqrt{\frac{t^2+a^2}{t^2-b^2}}\,dt$	$\sqrt{a^2+b^2}$	$\frac{a}{\sqrt{a^2+b^2}}$	$\arcsin\sqrt{\frac{1+(b/a)^2}{1+(x/a)^2}}$
$\int_0^x \sqrt{\frac{a^2+t^2}{(b^2+t^2)^3}}\,dt$	$\frac{a}{b^2}$	$\frac{\sqrt{a^2-b^2}}{a}$	$\arctan\frac{x}{b}$
$\int_b^x \frac{dt}{t^2\sqrt{(t^2-b^2)(a^2-t^2)}}$	$\frac{1}{ab^2}$	$\frac{\sqrt{a^2-b^2}}{a}$	$\arcsin\sqrt{\frac{1-(b/x)^2}{1-(b/a)^2}}$

(II) Elliptic Theta Functions

(1) For $\operatorname{Im}\tau>0$, we put $q\equiv e^{i\pi\tau}$ and define

$$\vartheta_0(u,\tau)\equiv\vartheta_4(u,\tau)\equiv 1+2\sum_{n=1}^{\infty}(-1)^n q^{n^2}\cos 2n\pi u,$$

$$\vartheta_1(u,\tau)\equiv 2\sum_{n=0}^{\infty}(-1)^n q^{[n+(1/2)]^2}\sin(2n+1)\pi u,$$

$$\vartheta_2(u,\tau)\equiv 2\sum_{n=0}^{\infty} q^{[n+(1/2)]^2}\cos(2n+1)\pi u,$$

$$\vartheta_3(u,\tau)\equiv 1+2\sum_{n=1}^{\infty} q^{n^2}\cos 2n\pi u.$$

Each of the four functions ϑ_j $(j=0,1,2,3)$ as a function of two variables u and τ satisfies the following partial differential equation

$$\frac{\partial^2\vartheta(u,\tau)}{\partial u^2}=4\pi i\frac{\partial\vartheta(u,\tau)}{\partial\tau}.$$

(2) Mutual Relations.

$$\vartheta_0^4(u)+\vartheta_2^4(u)=\vartheta_1^4(u)+\vartheta_3^4(u),\qquad \vartheta_0^2(u)=k\vartheta_1^2(u)+k'\vartheta_3^2(u),$$

$$\vartheta_2^2(u)=-k'\vartheta_1^2(u)+k\vartheta_3^2(u),\qquad \vartheta_1^2(u)=k\vartheta_0^2(u)-k'\vartheta_2^2(u),$$

where k is the modulus such that $iK'(k)/K(k)=\tau$, and k' is the corresponding complementary modulus.

$$k=\vartheta_2^2(0)/\vartheta_3^2(0),\qquad k'=\vartheta_0^2(0)/\vartheta_3^2(0).$$

$$\vartheta_1'(0)=\pi\vartheta_2(0)\vartheta_3(0)\vartheta_0(0),\qquad \frac{\vartheta_1'''(0)}{\vartheta_1'(0)}=\frac{\vartheta_2''(0)}{\vartheta_0(0)}+\frac{\vartheta_3''(0)}{\vartheta_3(0)}+\frac{\vartheta_0''(0)}{\vartheta_0(0)}.$$

(3) Pseudoperiodicity. In the following table, the only variables in ϑ are u and τ. m and n are integers.

Increment of u	ϑ_0	ϑ_1	ϑ_2	ϑ_3	Exponential Factor
$m+n\tau$	$(-1)^n\vartheta_0$	$(-1)^{m+n}\vartheta_1$	$(-1)^m\vartheta_2$	ϑ_3	$\exp[-n\pi i \times(2u+n\tau)]$
$m-\frac{1}{2}+n\tau$	ϑ_3	$(-1)^{m+1}\vartheta_2$	$(-1)^{m+n}\vartheta_1$	$(-1)^n\vartheta_0$	
$m+\left(n+\frac{1}{2}\right)\tau$	$(-1)^n i\vartheta_1$	$(-1)^{m+n}i\vartheta_0$	$(-1)^m\vartheta_3$	ϑ_2	$\exp\left[-\left(n+\frac{1}{2}\right)\pi i \times\left\{2u+\left(n+\frac{1}{2}\right)\tau\right\}\right]$
$m-\frac{1}{2}+\left(n+\frac{1}{2}\right)\tau$	ϑ_2	ϑ_3	$(-1)^{m+n}i\vartheta_0$	$(-1)^n i\vartheta_1$	
Zeros $u=$	$m+\left(n+\frac{1}{2}\right)\tau$	$m+n\tau$	$m+\frac{1}{2}+n\tau$	$m+\frac{1}{2}+\left(n+\frac{1}{2}\right)\tau$	

(4) Expansion into Infinite Products. We put $Q_0\equiv\prod_{n=1}^{\infty}(1-q^{2n})$. Then we have

$$\vartheta_0(u)=Q_0\prod_{n=1}^{\infty}(1-2q^{2n-1}\cos 2\pi u+q^{4n-2}),$$

$$\vartheta_1(u)=2Q_0q^{1/4}\sin\pi u\prod_{n=1}^{\infty}(1-2q^{2n}\cos 2\pi u+q^{4n}),$$

$$\vartheta_2(u)=2Q_0q^{1/4}\cos\pi u\prod_{n=1}^{\infty}(1+2q^{2n}\cos 2\pi u+q^{4n}),$$

$$\vartheta_3(u)=Q_0\prod_{n=1}^{\infty}(1+2q^{2n-1}\cos 2\pi u+q^{4n-2}).$$

(III) Jacobi's Elliptic Functions

(1) We express the modulus k and the complementary modulus as follows.

$$k=\frac{\vartheta_2^2(0)}{\vartheta_3^2(0)},\qquad k'=\frac{\vartheta_0^2(0)}{\vartheta_3^2(0)},\qquad k^2+k'^2=1.$$

Then we have

$$K(k)=K=\frac{\pi}{2}\vartheta_3^2(0),\qquad K'(k)=K'=-i\tau K.$$

The relation between q and k is

$$q=e^{i\pi\tau}=e^{-\pi(K'/K)},$$

$$q^{1/4}=\left(\frac{k}{4}\right)^{1/2}\left[1+2\left(\frac{k}{4}\right)^2+15\left(\frac{k}{4}\right)^4+150\left(\frac{k}{4}\right)^6+1707\left(\frac{k}{4}\right)^8+\dots\right],$$

$$q=\frac{1}{2}L+\frac{2}{2^5}L^5+\frac{15}{2^9}L^9+\frac{150}{2^{13}}L^{13}+\frac{1707}{2^{17}}L^{17}+\dots,\quad\text{where}\quad L=\frac{1-\sqrt[4]{1-k^2}}{1+\sqrt[4]{1-k^2}}.$$

(2) Functions sn, cn, dn; Addition Theorem.

$$\mathrm{sn}(u,k)\equiv\frac{1}{\sqrt{k}}\frac{\vartheta_1(u/2K)}{\vartheta_0(u/2K)},\qquad \mathrm{cn}(u,k)\equiv\sqrt{\frac{k'}{k}}\frac{\vartheta_2(u/2K)}{\vartheta_0(u/2K)},\qquad \mathrm{dn}(u,k)\equiv\sqrt{k'}\frac{\vartheta_3(u/2K)}{\vartheta_0(u/2K)}.$$

$$\mathrm{sn}^2u+\mathrm{cn}^2u=1,\qquad \mathrm{dn}^2u+k^2\mathrm{sn}^2u=1.$$

$$\mathrm{sn}(u+v)=\frac{\mathrm{sn}\,u\,\mathrm{cn}\,v\,\mathrm{dn}\,v+\mathrm{sn}\,v\,\mathrm{cn}\,u\,\mathrm{dn}\,u}{1-k^2\mathrm{sn}^2u\,\mathrm{sn}^2v},\qquad \mathrm{cn}(u+v)=\frac{\mathrm{cn}\,u\,\mathrm{cn}\,v-\mathrm{sn}\,u\,\mathrm{dn}\,u\,\mathrm{sn}\,v\,\mathrm{dn}\,v}{1-k^2\mathrm{sn}^2u\,\mathrm{sn}^2v},$$

$$\mathrm{dn}(u+v)=\frac{\mathrm{dn}\,u\,\mathrm{dn}\,v-k^2\mathrm{sn}\,u\,\mathrm{cn}\,u\,\mathrm{sn}\,v\,\mathrm{cn}\,v}{1-k^2\mathrm{sn}^2u\,\mathrm{sn}^2v}.$$

$$\frac{d\,\mathrm{sn}\,u}{du}=\mathrm{cn}\,u\,\mathrm{dn}\,u,\qquad \frac{d\,\mathrm{cn}\,u}{du}=-\mathrm{sn}\,u\,\mathrm{dn}\,u,\qquad \frac{d\,\mathrm{dn}\,u}{du}=-k^2\mathrm{sn}\,u\,\mathrm{cn}\,u.$$

(3) Periodicity. In the next table, m and n are integers.

Increment of u	$\operatorname{sn} u$	$\operatorname{cn} u$	$\operatorname{dn} u$
$2mK+2niK'$	$(-1)^m \operatorname{sn} u$	$(-1)^{m+n} \operatorname{cn} u$	$(-1)^n \operatorname{dn} u$
$(2m-1)K+2niK'$	$(-1)^{m+1} \frac{\operatorname{cn} u}{\operatorname{dn} u}$	$(-1)^{m+n} k' \frac{\operatorname{sn} u}{\operatorname{dn} u}$	$(-1)^n k' \frac{1}{\operatorname{dn} u}$
$2mK+(2n+1)iK'$	$(-1)^m k^{-1} \frac{1}{\operatorname{sn} u}$	$(-1)^{m+n+1} i k^{-1} \frac{\operatorname{dn} u}{\operatorname{sn} u}$	$i(-1)^{n+1} \frac{\operatorname{cn} u}{\operatorname{sn} u}$
$(2m-1)K+(2n+1)iK'$	$(-1)^{m+1} k^{-1} \frac{\operatorname{dn} u}{\operatorname{cn} u}$	$(-1)^{m+n} i k' k^{-1} \frac{1}{\operatorname{cn} u}$	$(-1)^n i k' \frac{\operatorname{sn} u}{\operatorname{cn} u}$
Zeros $u=$	$2nK+2miK'$	$(2n+1)K+2miK'$	$(2n+1)K+(2m+1)iK'$
Poles $u=$	$2nK+(2m+1)iK'$	$2nK+(2m+1)iK'$	$2nK+(2m+1)iK'$
Fundamental periods	$4K,\quad 2iK'$	$4K,\quad 2K+2iK'$	$2K,\quad 4iK'$

(4) Change of Variables. In the next table, the second column, for example, means the relation $\operatorname{sn}(ku, 1/k) = k \operatorname{sn}(u,k)$.

u	k	sn	cn	dn
ku	$1/k$	$k \operatorname{sn}$	dn	cn
iu	k'	$i \frac{\operatorname{sn}}{\operatorname{cn}}$	$\frac{1}{\operatorname{cn}}$	$\frac{\operatorname{dn}}{\operatorname{cn}}$
$k'u$	$i\frac{k}{k'}$	$k' \frac{\operatorname{sn}}{\operatorname{dn}}$	$\frac{\operatorname{cn}}{\operatorname{dn}}$	$\frac{1}{\operatorname{dn}}$
iku	$i\frac{k'}{k}$	$ik \frac{\operatorname{sn}}{\operatorname{dn}}$	$\frac{1}{\operatorname{dn}}$	$\frac{\operatorname{cn}}{\operatorname{dn}}$
$ik'u$	$\frac{1}{k'}$	$ik' \frac{\operatorname{sn}}{\operatorname{cn}}$	$\frac{\operatorname{dn}}{\operatorname{cn}}$	$\frac{1}{\operatorname{cn}}$
$(1+k)u$	$\frac{2\sqrt{k}}{1+k}$	$\frac{(1+k)\operatorname{sn}}{1+k\operatorname{sn}^2}$	$\frac{\operatorname{cn}\operatorname{dn}}{1+k\operatorname{sn}^2}$	$\frac{1-k\operatorname{sn}^2}{1+k\operatorname{sn}^2}$ (Gauss's transformation)
$(1+k')u$	$\frac{1-k'}{1+k'}$	$(1+k')\frac{\operatorname{sn}\operatorname{cn}}{\operatorname{dn}}$	$\frac{1-(1+k')\operatorname{sn}^2}{\operatorname{dn}}$	$1-(1-k')\operatorname{sn}^2$ (Landen's transformation)
$\frac{(1+k')^2 u}{2}$	$\left(\frac{1-\sqrt{k'}}{1+\sqrt{k'}}\right)^2$	$\frac{1+\sqrt{k'}}{1-\sqrt{k'}} \frac{k^2 \operatorname{sn}\operatorname{cn}}{(1+\operatorname{dn})(k'+\operatorname{dn})}$	$\frac{\operatorname{dn}-\sqrt{k'}}{1-\sqrt{k'}} \times \sqrt{\frac{2(1+k')}{(1+\operatorname{dn})(k'+\operatorname{dn})}}$	$\frac{\sqrt{2(1+k')}}{1+\sqrt{k'}} \frac{\operatorname{dn}+\sqrt{k'}}{\sqrt{(1+\operatorname{dn})(k'+\operatorname{dn})}}$

Jacobi's transformation. $\operatorname{sn}(iu,k) = i\frac{\operatorname{sn}(u,k')}{\operatorname{cn}(u,k')}$, $\quad \operatorname{cn}(iu,k) = \frac{1}{\operatorname{cn}(u,k')}$,

$\operatorname{dn}(iu,k) = \frac{\operatorname{dn}(u,k')}{\operatorname{cn}(u,k')}$.

(5) Amplitude.
The inverse function $\varphi = \operatorname{am}(u,k)$ of $u(k,\varphi) \equiv F(k,\varphi) = \int_0^{\varphi} \frac{d\psi}{\sqrt{1-k^2\sin^2\psi}}$

is called the amplitude.

$\mathrm{sn}(u,k)=\sin\varphi=\sin\mathrm{am}(u,k),\qquad \mathrm{cn}(u,k)=\cos\varphi=\cos\mathrm{am}(u,k),$

$\mathrm{dn}(u,k)=\sqrt{1-k^2\sin^2\varphi}=\sqrt{1-k^2\mathrm{sn}^2(u,k)}\ .$

$$u(k,x)=\int_0^x\frac{dt}{\sqrt{(1-t^2)(1-k^2t^2)}},\qquad x=\mathrm{sn}(u,k).$$

$$\mathrm{am}(u,k)=\frac{\pi u}{2K}+\sum_{n=1}^{\infty}\frac{2q^n}{n(1+q^{2n})}\sin\left(n\pi\frac{u}{2K}\right)\qquad (q=e^{i\pi\tau}=e^{-\pi(K'/K)}).$$

$\mathrm{am}(\theta,1)=\mathrm{gd}\theta$ (Gudermann function).

(IV) Weierstrass's Elliptic Functions

(1) Weierstrass's $\wp$-function. For the fundamental periods $2\omega_1$, $2\omega_3$, we have

$$\wp(u)\equiv\frac{1}{u^2}+\sum_{n,m}{}'\left[\frac{1}{(u-2n\omega_1-2m\omega_3)^2}-\frac{1}{(2n\omega_1+2m\omega_3)^2}\right]$$

$$=\frac{1}{u^2}+\frac{g_2}{20}u^2+\frac{g_3}{28}u^4+\frac{g_2^2}{1200}u^6+\frac{3g_2g_3}{6160}u^8+\dots.$$

$$g_2\equiv 60\sum_{n,m}{}'\frac{1}{(2n\omega_1+2m\omega_3)^4},\qquad g_3\equiv 140\sum_{n,m}{}'\frac{1}{(2n\omega_1+12m\omega_3)^6},$$

where Σ' means the sum over all integers except $m=n=0$.

$\wp(-u)=\wp(u)$. Putting $\omega_2\equiv-(\omega_1+\omega_3)$, $e_j\equiv\wp(\omega_j)$ $(j=1,2,3)$ we have

$$e_1+e_2+e_3=0,\qquad e_1e_2+e_2e_3+e_3e_1=-g_2/4,\qquad e_1e_2e_3=g_3/4.$$

$$\wp'(u)\equiv d\wp/du=-2\sum_{m,n}\frac{1}{(u-2n\omega_1-2m\omega_3)^3}.$$

$$\wp'^2(u)=4[\wp(u)-e_1][\wp(u)-e_2][\wp(u)-e_3]=4\wp^3(u)-g_2\wp(u)-g_3.$$

Addition theorem

$$\wp(u+v)=-\wp(u)-\wp(v)+\frac{1}{4}\left[\frac{\wp'(u)-\wp'(v)}{\wp(u)-\wp(v)}\right]^2,$$

$$\wp(u+\omega_j)=e_j+\frac{(e_j-e_k)(e_j-e_l)}{\wp(u)-e_j}\qquad (j,k,l)=(1,2,3).$$

Using theta functions corresponding to $\tau=\omega_3/\omega_1$,

$$\wp(u)=-\frac{\eta_1}{\omega_1}-\frac{d^2\log\vartheta_1(u/2\omega_1)}{du^2}\qquad\left(\eta_1=\zeta(\omega_1)=-\frac{1}{12\omega_1}\frac{\vartheta_1'''(0)}{\vartheta_1'(0)}\right),$$

$$\wp'(u)=-\frac{1}{4\omega_1^3}\frac{\vartheta_1'^3(0)\vartheta_2(u/2\omega_1)\vartheta_3(u/2\omega_1)\vartheta_0(u/2\omega_1)}{\vartheta_2(0)\vartheta_0(0)\vartheta_1^3(u/2\omega_1)}.$$

The relations to Jacobi's elliptic functions are

$q\equiv\exp(i\pi\omega_3/\omega_1)$.

$$\wp\left(\frac{u}{\sqrt{e_1-e_3}}\right)=e_1+(e_1-e_3)\frac{\mathrm{cn}^2u}{\mathrm{sn}^2u}=e_2+(e_1-e_3)\frac{\mathrm{dn}^2u}{\mathrm{sn}^2u}=e_3+(e_1-e_3)\frac{1}{\mathrm{sn}^2u},$$

where the modulus is $k=\sqrt{\dfrac{e_2-e_3}{e_1-e_3}}$, $\qquad K(k)=\omega_1\sqrt{e_1-e_3}$.

(2) ζ-function.

$$\zeta(u)\equiv\frac{1}{u}+\sum_{n,m}{}'\left[\frac{1}{u-2n\omega_1-2m\omega_3}+\frac{u}{(2n\omega_1+2m\omega_3)^2}+\frac{1}{2n\omega_1+2m\omega_3}\right]$$

$$=\frac{1}{u}-\frac{g_2}{60}u^3-\frac{g_3}{140}u^5-\frac{g_2^2}{8400}u^7-\frac{g_2g_3}{18480}u^9-\dots$$

$$=(\zeta_1/\omega_1)u+d\log\vartheta_1(u/2\omega_1)/du.$$

$$\zeta'(u)=-\wp(u).$$

Pseudoperiodicity. Putting $\eta_j \equiv \zeta(\omega_j)$ $(j=1,2,3)$ we have

$$\zeta(u+2n\omega_1+2m\omega_3)=\zeta(u)+2n\eta_1+2m\eta_3 \qquad (n,m=0,\pm 1,\pm 2,\dots),$$

$$\eta_1=-\frac{1}{12\omega_1}\frac{\vartheta_1'''(0)}{\vartheta_1'(0)}, \qquad \eta_1+\eta_2+\eta_3=0,$$

$$\eta_1\omega_3-\eta_3\omega_1=\eta_2\omega_1-\eta_1\omega_2=\eta_3\omega_2-\eta_2\omega_3=\pi i/2 \text{ (Legendre's relation)}.$$

Addition theorem $\zeta(u+v)=\zeta(u)+\zeta(v)+\dfrac{1}{2}\dfrac{\zeta''(u)-\zeta''(v)}{\zeta'(u)-\zeta'(v)}$.

(3) σ-function.

$$\sigma(u)\equiv u\prod_{n,m}{}'\left(1-\frac{u}{2n\omega_1+2m\omega_3}\right)\exp\left[\frac{u}{2n\omega_1+2m\omega_3}+\frac{1}{2}\left(\frac{u}{2n\omega_1+2m\omega_3}\right)^2\right] \qquad \begin{bmatrix} n,m=0,\pm 1, \\ \pm 2,\dots, \\ (n,m)\neq(0,0)\end{bmatrix}$$

$$=u-\frac{g_2}{2^4\cdot 3\cdot 5}u^5-\frac{g_3}{2^3\cdot 3\cdot 5\cdot 7}u^7-\frac{g_2^2}{2^9\cdot 3^2\cdot 5\cdot 7}u^9-\dots$$

$$=2\omega_1\left(\exp\frac{\eta_1 u^2}{2\omega_1}\right)\frac{\vartheta_1(u/2\omega_1)}{\vartheta_1'(0)}.$$

$$\zeta(u)=\sigma'(u)/\sigma(u). \qquad \sigma(-u)=-\sigma(u).$$

Pseudoperiodicity. $\sigma(u+2n\omega_1+2m\omega_3)=(-1)^{n+m+mn}[\exp(2n\eta_1+2m\eta_3)(u+n\omega_1+m\omega_3)]\sigma(u)$.

(4) Cosigma functions σ_1, σ_2, σ_3.

$$\sigma_j(u)\equiv -e^{\eta_j u}\frac{\sigma(u+\omega_j)}{\sigma(\omega_j)}=\left(\exp\frac{\eta_1 u^2}{2\omega_1}\right)\frac{\vartheta_{j+1}(u/2\omega_1)}{\vartheta_{j+1}(0)} \qquad (j=1,2,3;\ \vartheta_4\equiv\vartheta_0).$$

$$\wp(u)-e_j=\left[\frac{\sigma_j(u)}{\sigma(u)}\right]^2, \qquad \wp'(2u)=-\frac{2\sigma_1(u)\sigma_2(u)\sigma_3(u)}{\sigma^3(u)}=-\frac{\sigma(2u)}{\sigma^4(u)}.$$

$$\operatorname{sn}u=\alpha\frac{\sigma(u/\alpha)}{\sigma_3(u/\sigma)}, \qquad \operatorname{cn}u=\frac{\sigma_1(u/\alpha)}{\sigma_3(u/\alpha)}, \qquad \operatorname{dn}u=\frac{\sigma_2(u/\alpha)}{\sigma_3(u/\alpha)}, \qquad \text{where } \alpha=\sqrt{e_1-e_3}=\frac{K}{\omega_1}.$$

References

[1] W. F. Magnus, F. Oberhettinger, and R. P. Soni, Formulas and theorems for the special functions of mathematical physics, Springer, third enlarged edition, 1966.
[2] Y. L. Luke, The special functions and their approximations I, II, Academic Press, 1969.
[3] A. Erdelyi, Higher transcendental functions I, II, III (Bateman manuscript project) McGraw-Hill, 1953, 1955.
In particular, for hypergeometric functions of two variables see
[4] P. Appell, Sur les fonctions hypergéométriques de plusieurs variables, Mémor. Sci. Math., Gauthier-Villars, 1925.
Also → references to 39 Bessel Functions, 134 Elliptic Functions, 174 Gamma Function, 389 Special Functions.

17. Gamma Functions and Related Functions

(I) Gamma Functions and Beta Functions (→ 174 Gamma Function)

In this Section (I), C means Euler's constant, B_n means a Bernoulli number, ζ means the Riemann zeta function.

(1) Gamma function. $\Gamma(z)\equiv\displaystyle\int_0^\infty e^{-t}t^{z-1}\,dt \qquad (\operatorname{Re}z>0)$

$$=\frac{1}{e^{2\pi iz}-1}\int_\infty^{(0+)}e^{-t}t^{z-1}\,dt.$$

In the last integral, the integration contour goes once around the positive real axis in the positive direction.

$$\Gamma(n+1)=n! \qquad (n=0,1,2,\ldots), \qquad \Gamma(1/2)=\sqrt{\pi}.$$

$$\Gamma(z+1)=z\Gamma(z), \qquad \Gamma(z)\Gamma(1-z)=\frac{\pi}{\sin\pi z}, \qquad \prod_{j=0}^{n-1}\Gamma\left(z+\frac{j}{n}\right)=(2\pi)^{(n-1)/2}n^{(1/2)-nz}\Gamma(nz),$$

$$\frac{1}{\Gamma(z)}=ze^{Cz}\prod_{n=1}^{\infty}\left(1+\frac{z}{n}\right)e^{-z/n}, \qquad \log\Gamma(1+z)=-\frac{1}{2}\log\frac{\sin\pi z}{\pi z}-Cz-\sum_{n=1}^{\infty}\frac{\zeta(2n+1)}{2n+1}z^{2n+1}$$

$$(|z|<1).$$

$$\left|\frac{\Gamma(x+iy)}{\Gamma(x)}\right|^2=\prod_{n=0}^{\infty}1\Big/\left(1+\frac{y^2}{(x+n)^2}\right) \qquad (x,y \text{ are real and } x>0).$$

Asymptotic expansion (Stirling formula).

$$\Gamma(z)\approx e^{-z}z^{z-1}\sqrt{2\pi z}\,\exp\left[\sum_{n=1}^{\infty}\frac{(-1)^{n-1}B_{2n}z^{1-2n}}{2n(2n-1)}\right] \qquad (|\arg z|<\pi)$$

$$=e^{-z}z^{z-(1/2)}\sqrt{2\pi}\left[1+\frac{1}{12z}+\frac{1}{288z^2}-\frac{139}{51840z^3}-\frac{571}{2488320z^4}+O(z^{-5})\right].$$

(2) Beta Function. $B(x,y)\equiv\int_0^1 t^{x-1}(1-t)^{y-1}dt \qquad (\mathrm{Re}\,x,\quad \mathrm{Re}\,y>0)$

$$=\Gamma(x)\Gamma(y)/\Gamma(x+y).$$

(3) Incomplete Gamma Function.

$$\gamma(\nu,x)\equiv\int_0^x t^{\nu-1}e^{-t}dt=\Gamma(\nu)-x^{(\nu-1)/2}e^{-x/2}W_{(\nu-1)/2,\nu/2}(x) \qquad (\mathrm{Re}\,\nu>0).$$

(4) Incomplete Beta Function. $B_\alpha(x,y)\equiv\int_0^\alpha t^{x-1}(1-t)^{y-1}dt \qquad (0<\alpha\leqslant 1).$

(5) Polygamma Functions. $\psi(z)\equiv\frac{d}{dz}\log\Gamma(z)$

$$=\frac{\Gamma'(z)}{\Gamma(z)}=\int_0^\infty\left[\frac{e^{-t}}{t}-\frac{e^{-zt}}{1-e^{-t}}\right]dt=-C+\sum_{n=0}^{\infty}\left(\frac{1}{n+1}-\frac{1}{z+n}\right).$$

$$\psi'(z)=\sum_{n=0}^{\infty}\frac{1}{(z+n)^2}, \qquad \psi^{(k)}(z)=\sum_{n=0}^{\infty}\frac{(-1)^{k+1}k!}{(z+n)^{k+1}} \qquad (k=1,2,\ldots).$$

(II) Combinatorial Problems (→ 330 Permutations and Combinations)

Factorial $n!=n(n-1)(n-2)\ldots3\cdot2\cdot1. \quad 0!=1.$

Binomial coefficient $\binom{\alpha}{r}=\frac{\alpha(\alpha-1)\ldots(\alpha-r+1)}{r!}.$

(1) Number of Permutations of n Elements Taken r at a Time.

$${}_nP_r=n(n-1)\ldots(n-r+1)=n!/(n-r)!.$$

Number of combinations of n elements taken r at a time

$${}_nC_r=\frac{{}_nP_r}{r!}=\frac{n!}{r!(n-r)!}=\binom{n}{r}.$$

$${}_nC_r={}_nC_{n-r}, \qquad {}_nC_r={}_{n-1}C_r+{}_{n-1}C_{r-1}.$$

Number of multiple permutations ${}_n\Pi_r=n^r.$

Number of multiple combinations ${}_nH_r={}_{n+r-1}C_r=\frac{(n+r-1)!}{r!(n-1)!}.$

Number of circular permutations ${}_nP_r/r.$

(2) Binomial Theorem. $(a+b)^n=\sum_{r=0}^{n}\binom{n}{r}a^{n-r}b^r.$

Multinomial theorem $(a_1+\ldots+a_m)^n=\sum\frac{n!}{p_1!\ldots p_m!}a_1^{p_1}\ldots a_m^{p_m}.$

The latter summation runs over all nonnegative integers satisfying $p_1+\ldots+p_m=n$.

References

See references to Table 16, this Appendix.

18. Hypergeometric Functions and Spherical Functions

(I) Hypergeometric Function (→ 206 Hypergeometric Functions)

(1) Hypergeometric Function. $F(a,b;c;z)\equiv\sum_{n=0}^{\infty}\frac{\Gamma(a+n)}{\Gamma(a)}\frac{\Gamma(b+n)}{\Gamma(b)}\frac{\Gamma(c)}{\Gamma(c+n)}\frac{z^n}{n!}.$

The fundamental system of solutions of the hypergeometric differential equation $z(1-z)\frac{d^2u}{dz^2}+[c-(a+b+1)z]\frac{du}{dz}-abu=0$ at $z=0$ is given by

$$u_1=F(a,b;c;z),\qquad u_2=z^{1-c}F(a-c+1,\,b-c+1;\,2-c;\,z)\qquad (c\neq 0,-1,-2,\ldots).$$

$$F(a,b;c;z)=F(b,a;c;z).\qquad dF/dz=(ab/c)F(a+1,\,b+1;\,c+1;\,z).$$

$$F(a,b;c;1)=\frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}\qquad (\mathrm{Re}(a+b-c)<0).$$

$$F(a,b;c;z)=\frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)}\int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a}dt\qquad (\mathrm{Re}\,c>\mathrm{Re}\,b>0,\quad |z|<1),$$

$$F(a,b;c;z)=\frac{1}{2\pi i}\frac{\Gamma(c)}{\Gamma(a)\Gamma(b)}\int_{-i\infty}^{i\infty}\frac{\Gamma(a+s)\Gamma(b+s)\Gamma(-s)}{\Gamma(c+s)}(-z)^s ds.$$

(2) Transformations of the Hypergeometric Function.

$$\begin{aligned}
F(a,b;\,c;\,z)&=(1-z)^{-a}F\left(a,c-b;\,c;\,\frac{z}{z-1}\right)\\
&=(1-z)^{c-a-b}F(c-a,c-b;\,c;\,z)\\
&=(1-z)^{-a}\frac{\Gamma(c)\Gamma(b-a)}{\Gamma(b)\Gamma(c-a)}F\left(a,c-b;\,a-b+1;\,\frac{1}{1-z}\right)\\
&\quad+(1-z)^{-b}\frac{\Gamma(c)\Gamma(a-b)}{\Gamma(a)\Gamma(c-b)}F\left(b,c-a;\,b-a+1;\,\frac{1}{1-z}\right)\\
&=\frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}F(a,b;\,a+b-c+1;\,1-z)\\
&\quad+(1-z)^{c-a-b}\frac{\Gamma(c)\Gamma(a+b-c)}{\Gamma(a)\Gamma(b)}F(c-a,c-b;\,c-a-b+1;\,1-z)\\
&=\frac{\Gamma(c)\Gamma(b-a)}{\Gamma(b)\Gamma(c-a)}(-z)^{-a}F\left(a,1-c+a;\,1-b+a;\,\frac{1}{z}\right)\\
&\quad+\frac{\Gamma(c)\Gamma(a-b)}{\Gamma(a)\Gamma(c-b)}(-z)^{-b}F\left(b,1-c+b;\,1-a+b;\,\frac{1}{z}\right).
\end{aligned}$$

(3) Riemann's Differential Equation (→ Table 14.II, this Appendix).

$$\frac{d^2u}{dz^2}+\left[\frac{1-\alpha-\alpha'}{z-a}+\frac{1-\beta-\beta'}{z-b}+\frac{1-\gamma-\gamma'}{z-c}\right]\frac{du}{dz}$$

$$+\left[\frac{\alpha\alpha'(a-b)(a-c)}{z-a}+\frac{\beta\beta'(b-c)(b-a)}{z-b}+\frac{\gamma\gamma'(c-a)(c-b)}{z-c}\right]\frac{u}{(z-a)(z-b)(z-c)}=0.$$

Here we have $\alpha+\alpha'+\beta+\beta'+\gamma+\gamma'=1$ (Fuchsian relation). The solution of this equation is given by Riemann's P-function

$$u=P\begin{Bmatrix} a & b & c & \\ \alpha & \beta & \gamma & z \\ \alpha' & \beta' & \gamma' & \end{Bmatrix}$$

$$=\left(\frac{z-a}{z-b}\right)^{\alpha}\left(\frac{z-c}{z-b}\right)^{\gamma}F\left(\alpha+\beta+\gamma,\alpha+\beta'+\gamma;\ 1+\alpha-\alpha';\ \frac{(c-b)(z-a)}{(c-a)(z-b)}\right)$$

$$(\alpha-\alpha',\beta-\beta',\gamma-\gamma'\neq\text{integer}).$$

We have 24 representations of the above function by interchanging the parameters a, b, c; α, α'; β, β'; γ, γ' in the right-hand side.

(4) Barnes's Extended Hypergeometric Function.

$${}_pF_q(\alpha_1,\ldots,\alpha_p;\beta_1,\ldots,\beta_q;z)\equiv\sum_{n=0}^{\infty}\frac{(\alpha_1)_n\ldots(\alpha_p)_n}{(\beta_1)_n\ldots(\beta_q)_n}\frac{z^n}{n!},\quad\text{where}\quad(\alpha)_n=\alpha(\alpha+1)\ldots(\alpha+n-1)$$

$=\Gamma(\alpha+n)/\Gamma(\alpha)$. $\quad F(a,b;c;z)={}_2F_1(a,b;c;z)$. $\quad {}_0F_0(x)=e^x$, $\quad {}_1F_0(\alpha;x)=(1-x)^{-\alpha}$.

(5) Appell's Hypergeometric Functions of Two Variables.

$$F_1(\alpha;\beta,\beta';\gamma;x,y)=\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{(\alpha)_{m+n}(\beta)_m(\beta')_n}{m!n!(\gamma)_{m+n}}x^my^n,$$

$$F_2(\alpha;\beta,\beta';\gamma,\gamma';x,y)=\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{(\alpha)_{m+n}(\beta)_m(\beta')_n}{m!n!(\gamma)_m(\gamma')_n}x^my^n,$$

$$F_3(\alpha,\alpha';\beta,\beta';\gamma;x,y)=\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{(\alpha)_m(\alpha')_n(\beta)_m(\beta')_n}{m!n!(\gamma)_{m+n}}x^my^n,$$

$$F_4(\alpha;\beta;\gamma,\gamma';x,y)=\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{(\alpha)_{m+n}(\beta)_{m+n}}{m!n!(\gamma)_m(\gamma')_n}x^my^n.$$

(6) Representation of Various Special Functions by Hypergeometric Functions.

$$(1-x)^{\nu}=F(-\nu,b;b;x),\qquad e^{-nx}=\left(\frac{\operatorname{sech}x}{2}\right)^n(\tanh x)F\left(1+\frac{n}{2},\frac{1+n}{2};\ 1+n;\ \operatorname{sech}^2x\right).$$

$$\log(1+x)=xF(1,1;\ 2;\ -x),\qquad \frac{1}{2}\log\frac{1+x}{1-x}=xF\left(\frac{1}{2},1;\ \frac{3}{2};\ x^2\right),$$

$$\sin nx=n(\sin x)F\left(\frac{1+n}{2},\frac{1-n}{2};\ \frac{3}{2};\sin^2x\right),$$

$$\cos nx=F\left(\frac{n}{2},-\frac{n}{2};\ \frac{1}{2};\ \sin^2x\right)=(\cos x)F\left(\frac{1+n}{2},\frac{1-n}{2};\ \frac{1}{2};\ \sin^2x\right),$$

$$\arcsin x=xF\left(\frac{1}{2},\frac{1}{2};\ \frac{3}{2};\ x^2\right),\qquad \arctan x=xF\left(\frac{1}{2},1;\ \frac{3}{2};\ -x^2\right).$$

$$P_{2n}(x)=(-1)^n\frac{(2n-1)!!}{(2n)!!}F\left(-n,n+\frac{1}{2};\ \frac{1}{2};\ x^2\right),$$

$$P_{2n+1}(x)=(-1)^n\frac{(2n+1)!!}{(2n)!!}xF\left(-n,n+\frac{3}{2};\ \frac{3}{2};\ x^2\right)\qquad\text{(spherical function)},$$

where

$$n=0,1,2,\ldots; \qquad m!!=\begin{cases} m(m-2)\ldots 4\cdot 2 & (m \text{ even}), \\ m(m-2)\ldots 3\cdot 1 & (m \text{ odd}), \end{cases} \qquad 0!!=(-1)!!=1.$$

$$K(x)=\frac{\pi}{2}F\left(\frac{1}{2},\frac{1}{2};1;x^2\right), \qquad E(x)=\frac{\pi}{2}F\left(-\frac{1}{2},\frac{1}{2};1;x^2\right) \quad \text{(complete elliptic integral)}.$$

$$J_\nu(x)=\frac{x}{2\Gamma(\nu+1)}{}_0F_1\left(\nu+1;-\frac{x^2}{4}\right)=\frac{x^\nu e^{-ix}}{2^\nu\Gamma(\nu+1)}{}_1F_1\left(\nu+\frac{1}{2};2\nu+1;2ix\right).$$

$$e^x=\lim_{b\to\infty}F(a,b;a;x/b)={}_1F_1(a;a;x)={}_0F_0(x).$$

(II) Legendre Function (→ 393 Spherical Functions)

(1) Legendre Functions. The generalized spherical function corresponding to the rotation group of 3-dimensional space is the solution of the following differential equation.

$$(1-z^2)\frac{d^2u}{dz^2}-2z\frac{du}{dz}+\left[\nu(\nu+1)-\frac{\mu^2}{1-z^2}\right]u=0.$$

When $\mu=0$, the equation is Legendre's differential equation, and the fundamental system of solutions is given by the following two kind of functions.

Legendre function of the first kind $\quad \mathfrak{P}_\nu(z)\equiv P_\nu(z)\equiv {}_2F_1\left(-\nu,\ \nu+1;\ 1;\ \frac{1-z}{2}\right).$

Legendre function of the second kind

$$\mathfrak{Q}_\nu(z)\equiv\frac{\Gamma(\nu+1)\sqrt{\pi}}{2^{\nu+1}\Gamma[\nu+(3/2)]}z^{-\nu-1}{}_2F_1\left(\frac{\nu+2}{2},\ \frac{\nu+1}{2};\ \nu+\frac{3}{2};\ \frac{1}{z^2}\right).$$

$$Q_\nu(x)\equiv\frac{1}{2}[\mathfrak{Q}_\nu(x+i0)+\mathfrak{Q}_\nu(x-i0)]$$

$$=\pi\frac{(\cos\nu\pi)P_\nu(x)-P_\nu(-x)}{2\sin\nu\pi} \qquad (\nu\neq\text{integer};\ -1<x<1).$$

Recurrence formulas:

$$\mathfrak{P}_\nu(z)=\mathfrak{P}_{-\nu-1}(z). \qquad \mathfrak{Q}_\nu(z)-\mathfrak{Q}_{-\nu-1}(z)=\pi(\cot\nu\pi)\mathfrak{P}_\nu(z) \qquad (\nu\neq\text{integer}).$$

$$\mathfrak{P}_\nu(-z)=e^{\pm\nu\pi i}\mathfrak{P}_\nu(z)-(2/\pi)(\sin\nu\pi)\mathfrak{Q}_\nu(z), \qquad \mathfrak{Q}_\nu(-z)=-e^{\pm\nu\pi i}\mathfrak{Q}_\nu(z) \qquad (\pm=\operatorname{sgn}(\operatorname{Im}z)).$$

$$(z^2-1)d\mathfrak{P}_\nu(z)/dz=(\nu+1)[\mathfrak{P}_{\nu+1}(z)-z\mathfrak{P}_\nu(z)],$$

$$(2\nu+1)z\mathfrak{P}_\nu(z)=(\nu+1)\mathfrak{P}_{\nu+1}(z)+\nu\mathfrak{P}_{\nu-1}(z),$$

$$(z^2-1)d\mathfrak{Q}_\nu(z)/dz=(\nu+1)[\mathfrak{Q}_{\nu+1}(z)-z\mathfrak{Q}_\nu(z)],$$

$$(2\nu+1)z\mathfrak{Q}_\nu(z)=(\nu+1)\mathfrak{Q}_{\nu+1}(z)+\nu\mathfrak{Q}_{\nu-1}(z).$$

$$\mathfrak{P}_\nu(z)=\pi^{-1/2}2^{-\nu-1}\tan\nu\pi\frac{\Gamma(\nu+1)}{\Gamma[\nu+(3/2)]}z^{-\nu-1}{}_2F_1\left(\frac{\nu}{2}+1,\ \frac{\nu+1}{2};\ \nu+\frac{3}{2};\ \frac{1}{z^2}\right)$$

$$+\pi^{-1/2}2^\nu\frac{\Gamma[\nu+(1/2)]}{\Gamma(\nu+1)}z^\nu{}_2F_1\left(\frac{1-\nu}{2},\ \frac{-\nu}{2};\ \frac{1}{2}-\nu;\ \frac{1}{z^2}\right).$$

$$P_\nu(\cos\theta)=\frac{\sin\nu\pi}{\pi}\sum_{n=0}^{\infty}(-1)^n\left(\frac{1}{\nu-n}-\frac{1}{\nu+n+1}\right)P_n(\cos\theta) \qquad (\nu\neq\text{integer};\ 0\leqslant\theta<\pi).$$

Estimation: $\quad |P_\nu(\cos\theta)|\leqslant\dfrac{2}{\sqrt{\nu\pi\sin\theta}}, \qquad |Q_\nu(\cos\theta)|\leqslant\dfrac{\sqrt{\pi}}{\sqrt{\nu\sin\theta}} \qquad (0<\theta<\pi;\ \nu>1).$

$$P_\nu(1)=1, \qquad P_\nu(0)=-\frac{\sin\nu\pi}{2\pi^{3/2}}\Gamma\left(\frac{\nu+1}{2}\right)\Gamma\left(\frac{-\nu}{2}\right),$$

$$Q_\nu(0)=\frac{1}{4\sqrt{\pi}}(1-\cos\nu\pi)\Gamma\left(\frac{\nu+1}{2}\right)\Gamma\left(\frac{-\nu}{2}\right).$$

(2) The Case $\nu=n$ $(=0,1,2,\ldots)$. In the following, the symbol !! means

$$m!!\equiv\begin{cases} m(m-2)\ldots 4\cdot 2 & (m \text{ even}), \\ m(m-2)\ldots 5\cdot 3\cdot 1 & (m \text{ odd}). \end{cases}$$

The function P_n is a polynomial of degree n (Legendre polynomial) and is represented as follows:

$$P_n(z) \equiv \frac{1}{2^n n!}\frac{d^n}{dz^n}(z^2-1)^n$$

$$= \frac{(2n)!}{2^n (n!)^2} z^n {}_2F_1\left(-\frac{n}{2},\ \frac{1-n}{2};\ \frac{1}{2}-n;\ \frac{1}{z^2}\right)$$

$$= \frac{(2n-1)!!}{n!}\left[z^n - \frac{n(n-1)}{(2n-1)} z^{n-2} + \frac{n(n-1)(n-2)(n-3)}{2\cdot 4\cdot(2n-1)(2n-3)} z^{n-4} \mp \dots\right].$$

$$P_{2m}(z) = \sum_{j=0}^{m} (-1)^{m-j}\frac{(2m+2j-1)!!}{(2j)!(2m-2j)!!} z^{2j},$$

$$P_{2m+1}(z) = \sum_{j=0}^{m} (-1)^{m-j}\frac{(2m+2j+1)!!}{(2j+1)!(2m-2j)!!} z^{2j+1}.$$

$$P_n(\cos\theta) = \frac{(2n)!}{2^{2n}(n!)^2} e^{\mp in\theta} {}_2F_1\left(\frac{1}{2},\ -n;\ \frac{1}{2}-n;\ e^{\pm 2i\theta}\right)$$

$$= \frac{2(2n-1)!!}{(2n)!!}\left[\cos n\theta + \frac{1}{1}\frac{n}{(2n-1)}\cos(n-2)\theta + \frac{1\cdot 3}{1\cdot 2}\frac{n(n-1)}{(2n-1)(2n-3)}\cos(n-4)\theta + \frac{1\cdot 3\cdot 5}{1\cdot 2\cdot 3}\frac{n(n-1)(n-2)}{(2n-1)(2n-3)(2n-5)}\cos(n-6)\theta + \dots\right]$$

$$+ \begin{cases} \left[\dfrac{(n-1)!!}{n!!}\right]^2 & (n \text{ even}), \\ 0 & (n \text{ odd}). \end{cases}$$

$$= \frac{4}{\pi}\frac{(2n)!!}{(2n+1)!!}\left[\sin(n+1)\theta + \frac{1\cdot(n+1)}{1\cdot(2n+3)}\sin(n+3)\theta + \frac{1\cdot 3\cdot(n+1)(n+2)}{1\cdot 2\cdot(2n+3)(2n+5)}\sin(n+5)\theta + \dots \text{(ad infinitum)}\right] \qquad (0<\theta<\pi).$$

Laplace-Mehler integral representation

$$P_n(\cos\theta) = \frac{1}{\pi}\int_0^\pi (\cos\theta + i\sin\theta\cos\varphi)^n\, d\varphi$$

$$= \frac{\sqrt{2}}{\pi}\int_0^\theta \frac{\cos[n+(1/2)]\varphi}{\sqrt{\cos\varphi - \cos\theta}}\, d\varphi = \frac{\sqrt{2}}{\pi}\int_\theta^\pi \frac{\sin[n+(1/2)]\varphi}{\sqrt{\cos\theta - \cos\varphi}}\, d\varphi.$$

$$P_n(x) = \frac{(-1)^n}{n!} r^{n+1}\frac{\partial^n}{\partial z^n}\left(\frac{1}{r}\right) \qquad \left(x = \frac{z}{r},\ r = \sqrt{z^2+\rho^2}\right).$$

$$P_n(1) = 1, \qquad P_n(-1) = (-1)^n, \qquad P_{2n+1}(0) = 0,$$

$$P_{2n}(0) = (-1)^n\frac{(2n)!}{2^{2n}(n!)^2} = \frac{(-1)^n(2n-1)!!}{(2n)!!}.$$

Recurrence formulas: $nP_n(z) - (2n-1)zP_{n-1}(z) + (n-1)P_{n-2}(z) = 0,$

$$(z^2-1)\frac{dP_n}{dz} = n(zP_n - P_{n-1}) = \frac{n(n+1)}{2n+1}(P_{n+1} - P_{n-1}) = (n+1)(P_{n+1} - zP_n).$$

$$\mathfrak{Q}_n(z) = \frac{1}{2^n n!}\frac{d^n}{dz^n}\left[(z^2-1)^n \log\frac{z+1}{z-1}\right] - \frac{1}{2}P_n(z)\log\frac{z+1}{z-1}$$

$$= 2^n n! \int_z^\infty \dots \int_z^\infty \frac{(dz)^{n+1}}{(z^2-1)^{n+1}}$$

$$= 2^n \int_z^\infty \frac{(t-z)^n}{(t^2-1)^{n+1}}\, dt$$

$$= (-1)^n \frac{1}{(2n-1)!!}\frac{d^n}{dz^n}\left[(z^2-1)^n \int_z^\infty \frac{dt}{(t^2-1)^{n+1}}\right] \qquad (\mathrm{Re}\, z > 1).$$

$$Q_n(\cos\theta)=\frac{2\cdot(2n)!!}{(2n+1)!!}\left[\cos(n+1)\theta+\frac{1\cdot(n+1)}{1\cdot(2n+3)}\cos(n+3)\theta\right.$$
$$\left.+\frac{1\cdot 3\cdot(n+1)(n+2)}{1\cdot 2\cdot(2n+3)(2n+5)}\cos(n+5)\theta+\ldots\right]\qquad(0<\theta<\pi).$$
$$Q_n(x)=\frac{1}{2^n n!}\frac{d^n}{dx^n}\left[(x^2-1)^n\log\frac{1+x}{1-x}\right]-\frac{1}{2}P_n(x)\log\frac{1+x}{1-x}$$
$$=\frac{1}{2}P_n(x)\log\frac{1+x}{1-x}-\sum_{j=1}^{n}\frac{1}{j}P_{j-1}(x)P_{n-j}(x).$$
$$Q_0(x)=\frac{1}{2}\log\frac{1+x}{1-x},\qquad Q_1(x)=\frac{x}{2}\log\frac{1+x}{1-x}-1,\qquad Q_2(x)=\frac{1}{4}(3x^2-1)\log\frac{1+x}{1-x}-\frac{3}{2}x.$$

(3) Generating Functions.

$$\frac{1}{\sqrt{1-2hz+h^2}}=\begin{cases}\sum_{n=0}^{\infty}h^nP_n(z) & (|h|<\min|z\pm\sqrt{z^2-1}\,|),\\ \sum_{n=0}^{\infty}\frac{1}{h^{n+1}}P_n(z) & (|h|>\max|z\pm\sqrt{z^2-1}\,|).\end{cases}$$

(If $-1\leqslant z\leqslant 1$, the right-hand side is equal to 1.)

$$\frac{1}{z-t}=\sum_{n=0}^{\infty}(2n+1)\mathfrak{P}_n(t)\mathfrak{Q}_n(z)\qquad\left(|t+\sqrt{t^2-1}\,|<|z+\sqrt{z^2-1}\,|\right).$$
$$\frac{1}{\sqrt{1-2tz+z^2}}\log\frac{z-t+\sqrt{1-2tz+z^2}}{\sqrt{z^2-1}}=\sum_{n=0}^{\infty}t^n\mathfrak{Q}_n(z)\qquad(\mathrm{Re}\,z>1,\quad |t|<1).$$
$$r=\sqrt{x^2+y^2+z^2}\,,\qquad \cos\theta=z/r,\qquad x,y\text{ real},$$
$$\frac{1}{r}+\frac{1}{2}+\sum_{m=1}^{\infty}\left[\frac{1}{\sqrt{(2mi\pi+z)^2+x^2+y^2}}+\frac{1}{\sqrt{(2mi\pi-z)^2+x^2+y^2}}\right]$$

(Here the square root of a complex number is taken so that its real part is positive.)

$$=\begin{cases}1+\sum_{n=1}^{\infty}e^{-nz}J_0\left(n\sqrt{x^2+y^2}\,\right) & (\mathrm{Re}\,z>0),\\ \frac{1}{r}+\frac{1}{2}+\sum_{n=1}^{\infty}\frac{(-1)^nB_{2n}}{(2n)!}r^{2n-1}P_{2n-1}(\cos\theta) & (0<\theta<2\pi;\,z\text{ real}).\end{cases}$$

(4) Integrals of Legendre Polynomials.

Orthogonal relations: $\int_{-1}^{+1}P_n(z)P_m(z)\,dz=\delta_{nm}\frac{2}{2n+1}$.

$$\int_{-1}^{+1}z^kP_n(z)\,dz=0\qquad(k=0,1,\ldots,n-1).$$
$$\int_0^1 z^\lambda P_n(z)\,dz=\begin{cases}\frac{\lambda(\lambda-2)\ldots(\lambda-n+2)}{(\lambda+n+1)(\lambda+n-1)\ldots(\lambda+1)} & (n\text{ even}),\\ \frac{(\lambda-1)(\lambda-3)\ldots(\lambda-n+2)}{(\lambda+n+1)(\lambda+n-1)\ldots(\lambda+2)} & (n\text{ odd})\qquad(\mathrm{Re}\lambda>-1).\end{cases}$$
$$\int_0^\pi P_n(\cos\theta)\sin m\theta\,d\theta=\begin{cases}\frac{2(m-n+1)(m-n+3)\ldots(m+n-1)}{(m-n)(m-n+2)\ldots(m+n)} & (m>n;\ m+n\text{ is odd}),\\ 0 & (\text{otherwise}).\end{cases}$$

(5) Conical Function (Kegelfunktion). This is the Legendre function corresponding to the case $\nu=-(1/2)+i\lambda$ (λ is a real parameter),

$$P_{-(1/2)+i\lambda}(\cos\theta)=1+\frac{4\lambda^2+1^2}{2^2}\sin^2\frac{\theta}{2}+\frac{(4\lambda^2+1^2)(4\lambda^2+3^2)}{2^2\cdot 4^2}\sin^4\frac{\theta}{2}+\ldots.$$
$$P_{-(1/2)+i\lambda}(x)\equiv P_{-(1/2)-i\lambda}(x).$$

(III) Associated Legendre Functions (→ 393 Spherical Functions)

(1) Associated Legendre Functions. The fundamental system of solutions of the differential equation in (II) (1) is given by the following two kind of functions when $\mu\neq 0$.

Associated Legendre function of the first kind:

$$\mathfrak{P}_\nu^\mu(z)\equiv\frac{1}{\Gamma(1-\mu)}\left(\frac{z+1}{z-1}\right)^{\mu/2}{}_2F_1\left(-\nu,\nu+1;1-\mu;\frac{1-z}{2}\right),$$

where we take the branch satisfying $\arg[(z+1)/(z-1)]^{\mu/2}=0$ for $z>1$ in the expression raised to the $(\mu/2)$th power.

Associated Legendre function of the second kind:

$$\mathfrak{Q}_\nu^\mu(z)\equiv\frac{e^{i\mu\pi}}{2^{\nu+1}}\frac{\Gamma(\nu+\mu+1)\sqrt{\pi}}{\Gamma[\nu+(3/2)]}(z^2-1)^{\mu/2}z^{-\nu-\mu-1}{}_2F_1\left(\frac{\nu+\mu+2}{2},\frac{\nu+\mu+1}{2};\nu+\frac{3}{2};\frac{1}{z^2}\right),$$

where we take the branch satisfying $\arg(z^2-1)^{\mu/2}=0$ for $z>1$ in $(z^2-1)^{\mu/2}$, and $\arg z^{-\nu-\mu-1}=0$ for $z>0$ in $z^{-\nu-\mu-1}$, respectively.

$$P_\nu^\mu(x)\equiv e^{i\mu\pi/2}\mathfrak{P}_\nu^\mu(x+i0)=e^{i\mu\pi/2}\mathfrak{P}_\nu^\mu(x-i0)$$
$$=\frac{1}{\Gamma(1-\mu)}\left(\frac{1+x}{1-x}\right)^{\mu/2}{}_2F_1\left(-\nu,\nu+1;1-\mu;\frac{1-x}{2}\right)\qquad(-1\leqslant x\leqslant 1).$$

$$Q_\nu^\mu(x)\equiv e^{-i\mu\pi}\left[e^{-i\mu\pi/2}\mathfrak{Q}_\nu^\mu(x+i0)+e^{i\mu\pi/2}\mathfrak{Q}_\nu^\mu(x-i0)\right]/2$$
$$=\frac{\pi}{2\sin\mu\pi}\left[P_\nu^\mu(x)\cos\mu\pi-\frac{\Gamma(\nu+\mu+1)}{\Gamma(\nu-\mu+1)}P_\nu^{-\mu}(x)\right]\qquad(-1\leqslant x\leqslant 1).$$

Integral representations:

$$\mathfrak{P}_\nu^{-\mu}(z)=\frac{(z^2-1)^{\mu/2}}{2^\mu\sqrt{\pi}\,\Gamma[\mu+(1/2)]}\int_{-1}^{+1}\frac{(1-t^2)^{\mu-1/2}}{\left(z+t\sqrt{z^2-1}\,\right)^{\mu-\nu}}dt$$
$$\left(\operatorname{Re}\mu>-\frac{1}{2},\ |\arg(z\pm 1)|<\pi\right).$$

$$\mathfrak{P}_\nu^{-\mu}(z)=\frac{(z^2-1)^{\mu-2}}{2^\nu\Gamma(\mu-\nu)\Gamma(\nu+1)}\int_0^\infty\frac{(\sinh t)^{2\nu+1}}{(z+\cosh t)^{\mu+\nu+1}}dt$$
$$(\operatorname{Re}z>-1,\ |\arg(z\pm 1)|<\pi,\ \operatorname{Re}\nu>-1,\ \operatorname{Re}(\mu-\nu)>0).$$

$$\mathfrak{P}_\nu^{-\mu}(z)=\sqrt{\frac{2}{\pi}}\frac{\Gamma[\mu+(1/2)](z^2-1)^{\mu-2}}{\Gamma(\mu+\nu+1)\Gamma(\mu-\nu)}\int_0^\infty\frac{\cosh[\nu+(1/2)]t}{(z+\cosh t)^{\mu+(1/2)}}dt$$
$$(\operatorname{Re}z>-1,\ |\arg(z\pm 1)|<\pi,\ \operatorname{Re}(\mu+\nu)>-1,\ \operatorname{Re}(\mu-\nu)>0).$$

$$\mathfrak{P}_\nu^\mu(\cosh\alpha)=\sqrt{\frac{2}{\pi}}\frac{(\sinh\alpha)^\mu}{\Gamma[-\mu+(1/2)]}\int_0^\alpha\frac{\cosh[\{\nu+(1/2)\}t]\,dt}{(\cosh\alpha-\cosh t)^{\mu+(1/2)}}\qquad\left(\alpha>0,\ \operatorname{Re}\mu<\frac{1}{2}\right).$$

$$P_\nu^\mu(\cos\theta)=\sqrt{\frac{2}{\pi}}\frac{(\sin\theta)^\mu}{\Gamma[-\mu+(1/2)]}\int_0^\theta\frac{\cos[\{\nu+(1/2)\}\varphi]\,d\varphi}{(\cos\varphi-\cos\theta)^{\mu+(1/2)}}\qquad\left(0<\theta<\pi,\ \operatorname{Re}\mu<\frac{1}{2}\right).$$

$$P_\nu^{-\mu}(\cos\theta)=\frac{\Gamma(2\mu+1)2^{-\mu}(\sin\theta)^\mu}{\Gamma(\mu+1)\Gamma(\mu+\nu+1)\Gamma(\mu-\nu)}\int_0^\infty\frac{t^{\mu+\nu}dt}{(1+2t\cos\theta+t^2)^{\mu+(1/2)}}$$
$$(\operatorname{Re}(\mu+\nu)>-1,\ \operatorname{Re}(\mu-\nu)>0).$$

$$P_\nu^{-\mu}(\cos\theta)=\frac{1}{\Gamma(\nu+\mu+1)}\int_0^\infty e^{-t\cos\theta}J_\mu(t\sin\theta)t^\nu dt\qquad\left(0<\theta<\frac{\pi}{2},\ \operatorname{Re}(\mu+\nu)>-1\right).$$

$$\mathfrak{Q}_\nu^\mu(z)=\frac{e^{i\mu\pi}}{2^{\nu+1}}\frac{\Gamma(\nu+\mu+1)}{\Gamma(\nu+1)}(z^2-1)^{\mu/2}\int_{-1}^{+1}(1-t)^\nu(z-1)^{-\nu-\mu-1}dt$$
$$(\operatorname{Re}(\nu+\mu)>-1,\ \operatorname{Re}\nu>-1,\ |\arg(z\pm 1)|<\pi).$$

$$\mathfrak{Q}_\nu^\mu(\cosh\alpha)=\sqrt{\frac{\pi}{2}}\frac{e^{i\mu\pi}(\sinh\alpha)^\mu}{\Gamma[-\mu+(1/2)]}\int_\alpha^\infty\frac{e^{-[\nu+(1/2)]t}dt}{(\cosh t-\cosh\alpha)^{\mu+(1/2)}}$$
$$(\alpha>0,\ \operatorname{Re}\mu<1/2,\ \operatorname{Re}(\nu+\mu)>-1).$$

Recurrence formulas:

$$(z^2-1)d\mathfrak{P}_\nu^\mu(z)/dz=(\nu-\mu+1)\mathfrak{P}_{\nu+1}^\mu(z)-(\nu+1)z\mathfrak{P}_\nu^\mu(z),$$

$$(2\nu+1)z\mathfrak{P}_\nu^\mu(z)=(\nu-\mu+1)\mathfrak{P}_{\nu+1}^\mu(z)+(\nu+\mu)\mathfrak{P}_{\nu-1}^\mu(z),$$

$$\mathfrak{P}_{-\nu-1}^\mu(z)=\mathfrak{P}_\nu^\mu(z),$$

$$\mathfrak{Q}_\nu^{-\mu}(z)=e^{-2i\mu\pi}\frac{\Gamma(\nu-\mu+1)}{\Gamma(\nu+\mu+1)}\mathfrak{Q}_\nu^\mu(z),$$

$$(1-x^2)dP_\nu^\mu(x)/dx=(\nu+1)xP_\nu^\mu(x)-(\nu-\mu+1)P_{\nu+1}^\mu(x).$$

The case when μ is an integer $m(m=0,1,2,\ldots)$ and ν is also an integer n:

$$P_n^{m+2}(x)+2(m+1)x(1-x^2)^{-1/2}P_n^{m+1}(x)+(n-m)(n+m+1)P_n^m(x)=0.$$

$$(2n+1)xP_n^m(x)-(n-m+1)P_{n+1}^m(x)-(n+m)P_{n-1}^m(x)=0 \qquad (0\leqslant m\leqslant n-2),$$

$$(x^2-1)dP_n^m(x)/dx-(n-m+1)P_{n+1}^m(x)+(n+1)xP_n^m(x)=0,$$

$$P_{n-1}^m(x)-P_{n+1}^m(x)=(2n+1)\sqrt{1-x^2}\;P_n^{m-1}(x).$$

$$\mathfrak{P}_\nu^{-\mu}(z)=\frac{\Gamma(\nu-\mu+1)}{\Gamma(\nu+\mu+1)}\left[\mathfrak{P}_\nu^\mu(z)-\frac{2}{\pi}e^{-i\mu\pi}(\sin\mu\pi)\mathfrak{Q}_\nu^\mu(z)\right],$$

$$\mathfrak{Q}_\nu^\mu(z)\sin[(\nu+\mu)\pi]-\mathfrak{Q}_{-\nu-1}^\mu(z)\sin[(\nu-\mu)\pi]=\pi e^{i\mu\pi}(\cos\nu\pi)\mathfrak{P}_\nu^\mu(z),$$

$$\mathfrak{P}_\nu^\mu(-z)=e^{\mp i\nu\pi}\mathfrak{P}_\nu^\mu(z)-(2/\pi)[\sin(\nu+\mu)\pi]e^{-i\mu\pi}\mathfrak{Q}_\nu(z) \qquad (\mp=-\operatorname{sgn}(\operatorname{Im}z)),$$

$$\mathfrak{Q}_\nu^\mu(-z)=-e^{\pm i\nu\pi}\mathfrak{Q}_\nu^\mu(z) \qquad (\pm=\operatorname{sgn}(\operatorname{Im}z)).$$

$$e^{-i\mu\pi}\mathfrak{Q}_\nu^\mu(\cosh\alpha)=\frac{\pi\Gamma(1+\mu+\nu)}{\sqrt{2\pi\sinh\alpha}}\mathfrak{P}_{-\mu-1/2}^{-\nu-1/2}(\coth\alpha) \qquad (\operatorname{Re}\cosh\alpha>0).$$

$$e^{-i\mu\pi}\mathfrak{Q}_\nu^\mu(x\pm i0)=e^{\pm i\mu\pi/2}[Q_\nu^\mu(x)\mp(i\pi/2)P_\nu^\mu(x)].$$

$$Q_{-\nu-1}^\mu(x)=\frac{\sin(\nu+\mu)\pi}{\sin(\nu-\mu)\pi}Q_\nu^\mu(x)-\frac{\pi\cos\nu\pi\cos\mu\pi}{\sin(\nu-\mu)\pi}P_\nu^\mu(x),$$

$$P_\nu^{-\mu}(x)=\frac{\Gamma(\nu-\mu+1)}{\Gamma(\nu+\mu+1)}\left[\cos\mu\pi P_\nu^\mu(x)-\frac{2}{\pi}\sin\mu\pi Q_\nu^\mu(x)\right],$$

$$P_\nu^\mu(-x)=[\cos(\nu+\mu)\pi]P_\nu^\mu(x)-(2/\pi)[\sin(\nu+\mu)\pi]Q_\nu^\mu(x),$$

$$Q_\nu^\mu(-x)=-[\cos(\nu+\mu)\pi]Q_\nu^\mu(x)+(\pi/2)[\sin(\nu+\mu)\pi]P_\nu^\mu(x).$$

$$\mathfrak{P}_\nu^m(z)=\frac{\Gamma(1+\nu+m)(z^2-1)^{m/2}}{\Gamma(1+\nu-m)m!2^m}{}_2F_1\left(m-\nu,m+\nu+1;m+1;\frac{1-z}{2}\right)=(z^2-1)^{m/2}\frac{d^m\mathfrak{P}_\nu(z)}{dz^m},$$

$$\mathfrak{P}_\nu^{-m}(z)=(z^2-1)^{-m/2}\int_1^z\cdots\int_1^z P_\nu(z)(dz)^m.$$

$$\mathfrak{Q}_\nu^m(z)=(z^2-1)^{m/2}\frac{d^m\mathfrak{Q}_\nu(z)}{dx^m}, \qquad \mathfrak{Q}_\nu^{-m}(z)=(-1)^m(z^2-1)^{-m/2}\int_z^\infty\cdots\int_z^\infty\mathfrak{Q}_\nu(z)(dz)^m.$$

$$P_\nu^m(x)=(-1)^m\frac{\Gamma(1+\nu+m)(1-x^2)^{m/2}}{\Gamma(1+\nu-m)m!2^m}{}_2F_1\left(m-\nu,m+\nu+1;m+1;\frac{1-x}{2}\right)$$
$$=(-1)^m(1-x^2)^{m/2}\frac{d^mP_\nu(x)}{dx^m},$$

$$P_\nu^{-m}(x)=(1-x^2)^{-m/2}\int_x^1\cdots\int_x^1P_\nu(x)(dx)^m=(-1)^m\frac{\Gamma(\nu-m+1)}{\Gamma(\nu+m+1)}P_\nu^m(x).$$

$$Q_\nu^m(x)=(-1)^m(1-x^2)^{m/2}\frac{d^mQ_\nu(x)}{dx^m}, \qquad Q_\nu^{-m}(x)=(-1)^m\frac{\Gamma(\nu-m+1)}{\Gamma(\nu+m+1)}Q_\nu^m(x).$$

The values at the origin are

$$P_\nu^\mu(0)=\frac{\sqrt{\pi}\,2^\mu}{\Gamma[(\nu-\mu)/2+1]\Gamma[(-\nu-\mu+1)/2]},$$

$$\frac{dP_\nu^\mu(0)}{dx} = \frac{2^{\mu+1}\sin[\pi(\nu+\mu)/2]\Gamma[(\nu+\mu+2)/2]}{\Gamma[(\nu-\mu+1)/2]\sqrt{\pi}} = \frac{\sqrt{\pi}\,2^{\mu+1}}{\Gamma[(\nu-\mu+1)/2]\Gamma[(-\nu-\mu)/2]},$$

$$Q_\nu^\mu(0) = -2^{\mu-1}\sqrt{\pi}\,\sin\left(\frac{\nu+\mu}{2}\pi\right)\frac{\Gamma[(\nu+\mu+1)/2]}{\Gamma[(\nu-\mu+2)/2]},$$

$$\frac{dQ_\nu^\mu(0)}{dx} = 2^\mu\sqrt{\pi}\,\cos\left(\frac{\nu+\mu}{2}\pi\right)\frac{\Gamma[(\nu+\mu+2)/2]}{\Gamma[(\nu-\mu+1)/2]}.$$

(2) Generating Functions.

$$(\cos\theta + i\sin\theta\sin\varphi)^n = P_n(\cos\theta) + 2\sum_{m=1}^{n}(-i)^m\frac{n!}{(n+m)!}(\cos m\varphi)P_n^m(\cos\theta).$$

$$P_\nu^{-\mu}(\cos\theta) = \frac{\sin\nu\pi}{\pi}\sum_{n=0}^{\infty}(-1)^n\left(\frac{1}{\nu-n} - \frac{1}{\nu+n+1}\right)P_n^{-\mu}(\cos\theta) \qquad (0<\theta<\pi,\ \mu\geqslant 0).$$

(3) Orthogonal Relations.

$$\int_{-1}^{+1}P_n^m(x)P_{n'}^m(x)\,dx = \frac{2}{2n+1}\frac{(n+m)!}{(n-m)!}\delta_{nn'}.$$

$$\int_0^{2\pi}d\varphi\int_0^\pi \sin\theta e^{\pm i(m-m')\varphi}P_n^m(\cos\theta)P_{n'}^{m'}(\cos\theta)\,d\theta = \frac{4\pi}{2n+1}\frac{(n+m)!}{(n-m)!}\delta_{nn'}\delta_{mm'}.$$

(4) Addition Theorems.

$$\mathfrak{P}_\nu\left(z\zeta - \sqrt{z^2-1}\,\sqrt{\zeta^2-1}\,\cos\varphi\right) = \mathfrak{P}_\nu(z)\mathfrak{P}_\nu(\zeta) + 2\sum_{m=1}^{\infty}(-1)^m\mathfrak{P}_\nu^m(z)\mathfrak{P}_\nu^{-m}(\zeta)\cos m\varphi$$

$$(\operatorname{Re} z>0,\ \operatorname{Re}\zeta>0,\ |\arg(z-1)|<\pi,\ |\arg(\zeta-1)|<\pi).$$

$$\mathfrak{Q}_\nu\left(tt' - \sqrt{t^2-1}\,\sqrt{t'^2-1}\,\cos\varphi\right) = \mathfrak{Q}_\nu(t)\mathfrak{P}_\nu(t') + 2\sum_{m=1}^{\infty}(-1)^m\mathfrak{Q}_\nu^m(t)\mathfrak{P}_\nu^{-m}(t')\cos m\varphi$$

$$(t,t' \text{ real},\ 1<t'<t,\ \nu\neq\text{negative integer},\ \varphi \text{ real}).$$

$$P_\nu(\cos\theta\cos\theta' + \sin\theta\sin\theta'\cos\varphi) = P_\nu(\cos\theta)P_\nu(\cos\theta') + 2\sum_{m=1}^{\infty}(-1)^m P_\nu^{-m}(\cos\theta')P_\nu^m(\cos\theta')\cos m\varphi$$

$$= P_\nu(\cos\theta)P_\nu(\cos\theta') + 2\sum_{m=1}^{\infty}\frac{\Gamma(\nu-m+1)}{\Gamma(\nu+m+1)}P_\nu^m(\cos\theta)P_\nu^m(\cos\theta')\cos m\varphi$$

$$(0\leqslant\theta<\pi,\ 0\leqslant\theta'<\pi,\ \theta+\theta'<\pi,\ \varphi \text{ real}).$$

$$Q_\nu(\cos\theta\cos\theta' + \sin\theta\sin\theta'\cos\varphi) = P_\nu(\cos\theta')Q_\nu(\cos\theta) + 2\sum_{m=1}^{\infty}(-1)^m P_\nu^{-m}(\cos\theta')Q_\nu^m(\cos\theta)\cos m\varphi$$

$$(0<\theta'<\pi/2,\ 0<\theta<\pi,\ \theta+\theta'<\pi,\ \varphi \text{ real}).$$

$$\mathfrak{Q}_n\left(\tau\tau' + \sqrt{\tau^2+1}\,\sqrt{\tau'^2+1}\,\cosh\alpha\right) = \sum_{m=n+1}^{\infty}\frac{1}{(m-n-1)!(m+n)!}\mathfrak{Q}_n^m(i\tau)\mathfrak{Q}_n^m(i\tau')e^{-m\alpha}$$

$$(\tau,\ \tau',\ \alpha>0).$$

(5) Asymptotic Expansions.

$$\mathfrak{P}_\nu^\mu(z) = \left[\frac{2^\nu\Gamma[\nu+(1/2)]}{\sqrt{\pi}\,\Gamma(\nu-\mu+1)}z^\nu + \frac{2^{-\nu-1}\Gamma[-\nu-(1/2)]}{\sqrt{\pi}\,\Gamma(-\mu-\nu)}z^{-\nu-1}\right]\left[1+O(z^{-2})\right]$$

$$(\nu+(1/2)\neq\text{integer},\ |\arg z|<\pi,\ |z|\gg 1).$$

$$\mathfrak{Q}_\nu^\mu(z) = \frac{\sqrt{\pi}\,e^{i\mu\pi}}{2^{\nu+1}}\frac{\Gamma(\nu+\mu+1)}{\Gamma[\nu+(3/2)]}z^{-\nu-1}\left[1+O(z^{-2})\right]$$

$$(\nu+(1/2)\neq\text{negative integer},\ |\arg z|<\pi,\ |z|\gg 1).$$

$$P_\nu^\mu(\cos\theta) = \frac{2}{\sqrt{\pi}}\frac{\Gamma(\nu+\mu+1)}{\Gamma[\nu+(3/2)]}\frac{\cos[\{\nu+(1/2)\}\theta-(\pi/4)+(\mu\pi/2)]}{\sqrt{2\sin\theta}}\left[1+O(\nu^{-1})\right]$$

$$(\varepsilon\leqslant\theta\leqslant\pi-\varepsilon,\ \varepsilon>0,\ |\nu|\gg 1/\varepsilon).$$

$$P_\nu^\mu(\cos\theta)=\frac{2\Gamma(\nu+\mu+1)}{\sqrt{\pi}\,\Gamma[\nu+(3/2)]}$$
$$\times\sum_{l=0}^{\infty}\frac{\Gamma\left(\frac{1}{2}+\mu+l\right)\Gamma\left(\frac{1}{2}-\mu+l\right)\Gamma\left(\nu+\frac{3}{2}\right)}{\Gamma\left(\frac{1}{2}+\mu\right)\Gamma\left(\frac{1}{2}-\mu\right)\Gamma\left(\nu+l+\frac{3}{2}\right)l!}\,\frac{\cos\left[\left(\nu+\frac{2l+1}{2}\right)\theta-\frac{(2l+1)\pi}{4}+\frac{\mu\pi}{2}\right]}{(2\sin\theta)^{l+(1/2)}},$$

$$Q_\nu^\mu(\cos\theta)=\sqrt{\pi}\,\frac{\Gamma(\nu+\mu+1)}{\Gamma[\nu+(3/2)]}$$
$$\times\sum_{l=0}^{\infty}(-1)^l\frac{\Gamma\left(\frac{1}{2}+\mu+l\right)\Gamma\left(\frac{1}{2}-\mu+l\right)\Gamma\left(\nu+\frac{3}{2}\right)}{\Gamma\left(\frac{1}{2}+\mu\right)\Gamma\left(\frac{1}{2}-\mu\right)\Gamma\left(\nu+l+\frac{3}{2}\right)l!}\,\frac{\cos\left[\left(\nu+\frac{2l+1}{2}\right)\theta+\frac{(2l+1)\pi}{4}+\frac{\mu\pi}{2}\right]}{(2\sin\theta)^{l+(1/2)}}.$$

(In the final two formulas the series converges when $\nu+\mu\neq$negative integer, $\nu+(1/2)\neq$negative integer, $\pi/6<\theta<5\pi/6$.)

$$\left[\left(\nu+\frac{1}{2}\right)\cos\frac{\theta}{2}\right]^\mu P_\nu^{-\mu}(\cos\theta)=J_\mu(\eta)+\sin^2\frac{\theta}{2}\left[\frac{J_{\mu+1}(\eta)}{2\eta}-J_{\mu+2}(\eta)+\frac{\eta}{6}J_{\mu+3}(\eta)\right]+O\left(\sin^4\frac{\theta}{2}\right)$$
$$(\eta=(2\nu+1)\sin(\theta/2)).$$

(6) Estimation. When $\nu\geqslant 1,\quad \nu-\mu+1>0,\quad \mu\geqslant 0,$

$$|P_\nu^{\pm\mu}(\cos\theta)|<\frac{\Gamma(\nu\pm\mu+1)}{\Gamma(\nu+1)}\left(\frac{8}{\nu\pi\sin\theta}\right)^{1/2}\frac{1}{(\sin\theta)^\mu},$$

$$|Q_\nu^{\pm\mu}(\cos\theta)|<\frac{\Gamma(\nu\pm\mu+1)}{\Gamma(\nu+1)}\left(\frac{2\pi}{\nu\sin\theta}\right)^{1/2}\frac{1}{(\sin\theta)^\mu},$$

$$|P_\nu^{\pm m}(\cos\theta)|<\frac{\Gamma(\nu\pm m+1)}{\Gamma(\nu+1)}\left(\frac{4}{\nu\pi\sin\theta}\right)^{1/2}\frac{1}{(\sin\theta)^m},$$

$$|Q_\nu^{\pm m}(\cos\theta)|\leqslant\frac{\Gamma(\nu\pm m+1)}{\Gamma(\nu+1)}\left(\frac{4}{\nu\sin\theta}\right)^{1/2}\frac{1}{(\sin\theta)^m}.$$

(7) Torus Functions. These are solutions of the differential equation

$$\frac{d^2u}{d\eta^2}+\coth\eta\frac{du}{d\eta}-\left(n^2-\frac{1}{4}+\frac{m^2}{\sinh^2\eta}\right)u=0.$$

The fundamental system of solutions is given by

$\mathfrak{P}_{n-(1/2)}^m(\cosh\eta)$, $\mathfrak{Q}_{n-(1/2)}^m(\cosh\eta)$.

The asymptotic expansion when $m=0$ is

$$\mathfrak{P}_{n-(1/2)}(\cosh\eta)$$
$$=\frac{(n-1)!\,e^{n-(1/2)\eta}}{\Gamma[n+(1/2)]\sqrt{\pi}}\left[\frac{2\Gamma^2[n+(1/2)]}{\pi n!(n-1)!}(\log 4+\eta)e^{-2n\eta}{}_2F_1\left(\frac{1}{2},n+\frac{1}{2};n+1;e^{-2\eta}\right)+A+B\right].$$

Here

$$A=1+\frac{(1/2)[n-(1/2)]}{1\cdot(n-1)}e^{-2\eta}+\frac{(1/2)(3/2)[n-(1/2)][n-(3/2)]}{1\cdot2\cdot(n-1)(n-2)}e^{-4\eta}+\dots$$
$$+\frac{(2n-3)!!(2n-1)!!}{[(2n-2)!!]^2}e^{-2(n-1)\eta},$$

$$B=\frac{\Gamma[n+(1/2)]}{\pi^{3/2}(n-1)!}\sum_{l=1}^{\infty}\frac{\Gamma[l+(1/2)]\Gamma[n+l+(1/2)]}{(n+l)!\,l!}(u_{n+l}+u_l-v_{l-(1/2)}-v_{n+l-(1/2)})e^{-2(l+n)\eta},$$

where

$$u_r\equiv1+\frac{1}{2}+\dots+\frac{1}{r},\qquad v_{r-(1/2)}\equiv\frac{2}{1}+\frac{2}{3}+\frac{2}{5}+\dots+\frac{2}{2r-1}=2u_{2r}-u_r.$$

References

See references to Table 16, this Appendix.

19. Functions of Confluent Type and Bessel Functions

(I) Hypergeometric Function of Confluent Type (→ 167 Functions of Confluent Type)

(1) Kummer Functions.

$$v(z)={}_1F_1(a;\ c;\ z)\equiv\sum_{n=0}^{\infty}\frac{\Gamma(a+n)}{\Gamma(a)}\frac{\Gamma(c)}{\Gamma(c+n)}\frac{z^n}{n!}$$

$$=\frac{\Gamma(c)}{\Gamma(a)\Gamma(c-a)}z^{1-c}\int_0^z e^t t^{a-1}(z-t)^{c-a-1}dt \qquad (0<\mathrm{Re}\,a<\mathrm{Re}\,c)$$

$$=\frac{\Gamma(c)2^{1-c}}{\Gamma(a)\Gamma(c-a)}e^{z/2}\int_{-1}^{+1}e^{zt/2}(1-t)^{c-a-1}(1+t)^{a-1}dt \qquad (0<\mathrm{Re}\,a<\mathrm{Re}\,c).$$

The fundamental system of solutions of the confluent hypergeometric differential equation (Kummer's differential equation)

$$z\frac{d^2v}{dz^2}+(c-z)\frac{dv}{dz}-av=0,$$

when $c\neq 0,-1,-2,\ldots$, is given by

$$\mathring{v}_1(z)\equiv{}_1F_1(a;c;z),\qquad \mathring{v}_2(z)\equiv z^{1-c}{}_1F_1(a-c+1;\ 2-c;\ z).$$

$$d\,{}_1F_1(a;\ c;\ z)/dz=(a/c)\,{}_1F_1(a+1;c+1;z),$$

$${}_1F_1(a;c;z)=e^z{}_1F_1(c-a;c;-z),$$

$$a\,{}_1F_1(a+1;c+1;z)=(a-c)\,{}_1F_1(a;c+1;z)+c\,{}_1F_1(a;c;z),$$

$$a\,{}_1F_1(a+1;c;z)=(z+2a-c)\,{}_1F_1(a;c;z)+(c-a)\,{}_1F_1(a-1;c;z).$$

Putting $(a)_n=a(a+1)\ldots(a+n-1)=\Gamma(a+n)/\Gamma(a)$ we have

$$\lim_{c\to-n}\frac{1}{\Gamma(c)}{}_1F_1(a;c;z)=\frac{z^{n+1}(a)_{n+1}}{(n+1)!}{}_1F_1(a+n+1;n+2;z) \qquad (n=0,1,2,\ldots).$$

Asymptotic expansion:

$$\mathring{v}_1\approx A_1z^{-a}\sum_{n=0}^{\infty}\frac{(a)_n(a-c+1)_n}{n!}(-z)^{-n}+B_1e^zz^{a-c}\sum_{n=0}^{\infty}\frac{(c-a)_n(1-a)_n}{n!}z^{n},$$

$$\mathring{v}_2\approx A_2z^{-a}\sum_{n=0}^{\infty}\frac{(a)_n(a-c+1)_n}{n!}(-z)^{-n}+B_2e^zz^{a-c}\sum_{n=0}^{\infty}\frac{(c-a)_n(1-a)_n}{n!}z^{n}$$

$$(|z|\gg|a|,\ |z|\gg|c|,\ -3\pi/2<\arg z<\pi/2, c\neq\text{integer}),$$

where

$$A_1=e^{-i\pi a}\Gamma(c)/\Gamma(c-a),\qquad B_1=\Gamma(c)/\Gamma(a),$$

$$A_2=e^{-i\pi(a-c+1)}\Gamma(2-c)/\Gamma(1-a),\qquad B_2=\Gamma(2-c)/\Gamma(a-c+1).$$

(2) The fundamental system of solutions at $z=0$ of the hypergeometric differential equation of confluent type

$$\frac{d^2u}{dz^2}+\frac{du}{dz}+\left[\frac{\kappa}{z}+\frac{(1/4)-\mu^2}{z^2}\right]u=0$$

is given by

$$z^{(1/2)\pm\mu}e^{-z}{}_1F_1[(1/2)\pm\mu-\kappa;\ \pm2\mu+1;z].$$

(II) Whittaker Functions (→ 167 Functions of Confluent Type)

(1) A pair of linearly independent solutions of Whittaker's differential equation

$$\frac{d^2W}{dz^2}+\left[-\frac{1}{4}+\frac{\kappa}{z}+\frac{(1/4)-\mu^2}{z^2}\right]W=0$$

is given by $M_{\kappa,\pm\mu}(z)=z^{\pm\mu+(1/2)}e^{-z/2}{}_1F_1[\pm\mu-\kappa+(1/2);\pm2\mu+1;z]$.

Whittaker functions:

$$W_{\kappa,\mu}(z)\equiv\frac{\Gamma(-2\mu)}{\Gamma[(1/2)-\mu-\kappa]}M_{\kappa,\mu}(z)+\frac{\Gamma(2\mu)}{\Gamma[(1/2)+\mu-\kappa]}M_{\kappa,-\mu}(z)=W_{\kappa,-\mu}(z).$$

When 2μ is an integer, the above definition of $W_{\kappa,\mu}(z)$ loses meaning, but by taking the limit with respect to μ we can define it in terms of the following integrals.

$$W_{\kappa,\mu}(z)=\frac{z^{\mu+(1/2)}e^{-z/2}}{\Gamma[\mu+(1/2)-\kappa]}\int_0^\infty e^{-z\tau}\tau^{\mu-\kappa-(1/2)}(1+\tau)^{\mu+\kappa-(1/2)}d\tau$$

$$=\frac{z^{\kappa}e^{-z/2}}{\Gamma[\mu+(1/2)-\kappa]}\int_0^\infty t^{\mu-\kappa-(1/2)}e^{-t}\left(1+\frac{t}{z}\right)^{\mu+\kappa-(1/2)}dt$$

$(\mathrm{Re}[\mu+(1/2)-\kappa]>0,\quad |\arg z|<\pi)$.

$$W_{\kappa,\mu}(z)=\frac{e^{-z/2}}{2\pi i}\int_{-i\infty}^{+i\infty}\frac{\Gamma(s-\kappa)\Gamma[-s-\mu+(1/2)]\Gamma[-s+\mu+(1/2)]}{\Gamma[-\kappa+\mu+(1/2)]\Gamma[-\kappa-\mu+(1/2)]}z^s\,ds.$$

$$M_{l+\mu+(1/2),\mu}(z)=(-1)^l z^{\mu+(1/2)}e^{-z/2}(2\mu+1)_{l\,1}F_1(-l;2\mu+1;z)\qquad(l=0,1,2,\ldots).$$

$$M_{\kappa,\mu}(z)=e^{-i\pi[\mu+(1/2)]}M_{-\kappa,\mu}(e^{i\pi}z).$$

$$M_{\kappa,\mu}(z)=\frac{\Gamma(2\mu+1)}{\Gamma[\mu+(1/2)-\kappa]}e^{i\pi\kappa}W_{-\kappa,\mu}(e^{i\pi}z)+\frac{\Gamma(2\mu+1)}{\Gamma[\mu+(1/2)+\kappa]}e^{i\pi[\kappa-\mu-(1/2)]}W_{\kappa,\mu}(z)$$

$(-3\pi/2<\arg z<\pi/2,\quad 2\mu\neq-1,-2,\ldots)$.

$$M_{\kappa,\mu}(z)=\frac{\Gamma(2\mu+1)}{\Gamma[\mu+(1/2)-\kappa]}e^{-i\pi\kappa}W_{-\kappa,\mu}(e^{-i\pi}z)+\frac{\Gamma(2\mu+1)}{\Gamma[\mu+(1/2)+\kappa]}e^{-i\pi[\kappa-\mu-(1/2)]}W_{\kappa,\mu}(z)$$

$(-\pi/2<\arg z<3\pi/2,\quad 2\mu\neq-1,-2,\ldots)$.

$$W_{\kappa,\mu}(z)=z^{1/2}W_{\kappa-(1/2),\mu-(1/2)}(z)+[(1/2)-\kappa+\mu]W_{\kappa-1,\mu}(z)$$

$$=z^{1/2}W_{\kappa-(1/2),\mu+(1/2)}(z)+[(1/2)-\kappa-\mu]W_{\kappa-1,\mu}(z).$$

$$z\,dW_{\kappa,\mu}(z)/dz=[\kappa-(z/2)]W_{\kappa,\mu}(z)-\left[\mu^2-\{\kappa-(1/2)\}^2\right]W_{\kappa-1,\mu}(z).$$

When κ is sufficiently large we have

$$M_{\kappa,\mu}(z)\sim\pi^{-1/2}\Gamma(2\mu+1)\kappa^{-\mu-(1/4)}z^{1/4}\cos\left[2(z\kappa)^{1/2}-\mu\pi-(\pi/4)\right],$$

$$W_{\kappa,\mu}(z)\sim-(4z/\kappa)^{1/4}\exp(-\kappa+\kappa\log\kappa)\sin\left[2(z\kappa)^{1/2}-\pi\kappa-(\pi/4)\right],$$

$$W_{-\kappa,\mu}(z)\sim(z/4\kappa)^{1/4}\exp\left(\kappa-\kappa\log\kappa-2(z\kappa)^{1/2}\right).$$

Asymptotic expansion:

$$W_{\kappa,\mu}(z)\approx e^{-z/2}z^{\kappa}$$

$$\times\left(1+\sum_{n=1}^{\infty}\frac{[\mu^2-\{\kappa-(1/2)\}^2][\mu^2-\{\kappa-(3/2)\}^2]\ldots[\mu^2-\{\kappa-n+(1/2)\}^2]}{n!z^n}\right).$$

(2) Representation of Various Special Functions by Whittaker Functions.

(i) Probability integral (error function) $\quad \mathrm{erf}\,x\equiv\Phi(x)\equiv\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$

$$=1-\pi^{-1/2}x^{-1/2}e^{-x^2/2}W_{-1/4,1/4}(x^2)$$

$$=\frac{2x}{\sqrt{\pi}}{}_1F_1\left(\frac{1}{2};\frac{3}{2};-x^2\right)=\frac{2}{\sqrt{\pi}}\left(x-\frac{x^3}{1!3}+\frac{x^5}{2!5}-\frac{x^7}{3!7}\pm\ldots\right).$$

Asymptotic expansion:

$$\frac{\sqrt{\pi}}{2}[1-\Phi(x)]\approx\frac{e^{-x^2}}{2x}\left(1-\frac{1}{2x^2}+\frac{1\cdot3}{(2x^2)^2}-\frac{1\cdot3\cdot5}{(2x^2)^3}\pm\ldots\right).$$

$$\frac{1}{1+i}\Phi\left(x\frac{1+i}{2}\sqrt{\pi}\right)=C(x)-iS(x),$$

where $C(x)$, $S(x)$ are the following Fresnel integrals.

$$C(x)\equiv\int_0^x\cos\frac{\pi}{2}t^2dt=\frac{1}{2}+\frac{1}{\pi x}\sin\frac{\pi}{2}x^2+O\left(\frac{1}{x^2}\right),$$

$$S(x)\equiv\int_0^x\sin\frac{\pi}{2}t^2dt=\frac{1}{2}-\frac{1}{\pi x}\cos\frac{\pi}{2}x^2+O\left(\frac{1}{x^2}\right).$$

(ii) Logarithmic integral

$$\mathrm{Li}\,z\equiv\int_0^z\frac{dt}{\log t}\qquad\text{(When } z>1\text{, take Cauchy's principal value at } t=1.)$$

$$=-(\log 1/z)^{-1/2}z^{1/2}W_{-1/2,0}(-\log z).$$

$\mathrm{Li}\,z$ is sometimes written as $\mathrm{li}\,z$.

(3) Exponential Integral

$$\mathrm{Ei}\,x\equiv\int_{-\infty}^x\frac{e^t}{t}dt\quad\text{(When } x>0\text{, take the Cauchy's principal value at } t=0 \text{ while integrating.)}$$

$$=C+\log|x|+\sum_{n=1}^{\infty}\frac{x^n}{n\cdot n!}\qquad(x\text{ real, }\neq 0)$$

$$=e^x\sum_{n=1}^{N}\frac{(n-1)!}{t^n}+N!\sum_{\substack{n=0\\n\neq N}}^{\infty}\frac{t^{n-N}}{n!(n-N)}-\left(1+\frac{1}{2}+\cdots+\frac{1}{N}\right)+C+\log|x|.$$

Cosine integral $\quad \mathrm{Ci}\,x\equiv-\int_x^\infty\frac{\cos t}{t}dt=C+\log x-\int_0^x\frac{1-\cos t}{t}dt.$

Sine integral $\quad \mathrm{Si}\,x\equiv\int_0^x\frac{\sin t}{t}dt,$

$$\mathrm{si}\,x\equiv-\int_x^\infty\frac{\sin t}{t}dt=\mathrm{Si}\,x-\frac{\pi}{2}.$$

Asymptotic expansion $\quad \mathrm{Ei}\,ix=\mathrm{Ci}\,x+i\,\mathrm{si}\,x\approx e^{ix}\left(\frac{1}{ix}+\frac{1!}{(ix)^2}+\frac{2!}{(ix)^3}+\frac{3!}{(ix)^4}+\cdots\right).$

(III) Bessel Functions (→ 39 Bessel Functions)

(1) Cylindrical Functions. A cylindrical function Z_ν is a solution of Bessel's differential equation

$$\frac{d^2Z_\nu}{dz^2}+\frac{1}{z}\frac{dZ_\nu}{dz}+\left(1-\frac{\nu^2}{z^2}\right)Z_\nu=0.$$

Recurrence formulas:

$$Z_{\nu-1}(z)+Z_{\nu+1}(z)=(2\nu/z)Z_\nu(z),\qquad Z_{\nu-1}(z)-Z_{\nu+1}(z)=2dZ_\nu(z)/dz.$$

$$\int z^{\nu+1}Z_\nu(z)dz=z^{\nu+1}Z_{\nu+1}(z),\qquad\int z^{-\nu}Z_{\nu+1}(z)dz=-z^{-\nu}Z_\nu(z).$$

As special solutions, we have the following three kinds of functions.

(i) Bessel function (Bessel function of the first kind).

$$J_\nu(z)\equiv\left(\frac{z}{2}\right)^\nu\sum_{l=0}^{\infty}\frac{(-1)^l}{l!\Gamma(\nu+l+1)}\left(\frac{z}{2}\right)^{2l}=\frac{M_{0,\nu}(2iz)}{(2iz)^{1/2}2^{2\nu}i^\nu\Gamma(\nu+1)}\qquad(|\arg z|<\pi).$$

$$J_\nu(e^{im\pi}z)=e^{im\nu\pi}J_\nu(z).$$

$$J_{-n}(z)=(-1)^nJ_n(z).$$

$$J_{n+(1/2)}(z)=\sqrt{\frac{2}{\pi}}\,z^{n+(1/2)}\left(-\frac{1}{z}\frac{d}{dz}\right)^n\left(\frac{\sin z}{z}\right)\qquad(n=0,1,2,\ldots).$$

(ii) Neumann function (Bessel function of the second kind).

$$N_\nu(z)\equiv\frac{1}{\sin\nu\pi}\left[(\cos\nu\pi)J_\nu(z)-J_{-\nu}(z)\right]\qquad(\nu\neq\text{integer};\ |\arg z|<\pi),$$

$$N_n(z)\equiv\frac{2}{\pi}J_n(z)\left(C+\log\frac{z}{2}\right)-\frac{1}{\pi}\left(\frac{z}{2}\right)^n\sum_{l=0}^{\infty}\frac{(-1)^l}{l!(n+l)!}\left(\frac{z}{2}\right)^{2l}[\varphi(l)+\varphi(l+n)]$$

$$-\frac{1}{\pi}\left(\frac{z}{2}\right)^{-n}\sum_{l=0}^{n-1}\frac{(n-l-1)!}{l!}\left(\frac{z}{2}\right)^{2l}\qquad\left(\varphi(l)\equiv\sum_{m=1}^{l}\frac{1}{m}\right),$$

$N_{-n}(z) \equiv (-1)^n N_n(z) \qquad (n=0,1,2,\ldots; \qquad |\arg z| < \pi).$

$N_\nu(e^{im\pi}z) = e^{-im\nu\pi}N_\nu(z) + 2i(\sin m\nu\pi \cot \nu\pi)J_\nu(z).$

$N_{n+(1/2)}(z) = (-1)^{n+1}J_{-[n+(1/2)]}(z).$

(iii) Hankel function (Bessel function of the third kind).

$H_\nu^{(1)}(z) \equiv J_\nu(z) + iN_\nu(z),$

$H_\nu^{(2)}(z) \equiv J_\nu(z) - iN_\nu(z).$

$H_\nu^{(1)}(iz/2) = -2ie^{-i\nu\pi/2}(\pi z)^{-1/2}W_{0,\nu}(z).$

$H_{-\nu}^{(1)}(z) = e^{i\nu\pi}H_\nu^{(1)}(z), \quad H_{-\nu}^{(2)}(z) = e^{-i\nu\pi}H_\nu^{(2)}(z). \quad \overline{H_\nu^{(2)}(x)} = H_\nu^{(1)}(x) \quad (x, \nu \text{ real}).$

(2) Integral Representation.

Hansen-Bessel formula

$$J_n(z) = \frac{1}{2\pi}\int_{-\pi}^{+\pi} e^{iz\cos t}e^{in[t-(\pi/2)]}dt = \frac{i^{-n}}{\pi}\int_0^\pi e^{iz\cos t}\cos nt\,dt = \frac{1}{\pi}\int_0^\pi \cos(z\sin t - nt)dt \qquad (n=0,1,2,\ldots).$$

Mehler's formula

$$J_0(x) = \frac{2}{\pi}\int_0^\infty \sin(x\cosh t)dt,$$

$$N_0(x) = -\frac{2}{\pi}\int_0^\infty \cos(x\cosh t)dt \qquad (x>0).$$

Poisson's formula

$$J_\nu(z) = \frac{2(z/2)^\nu}{\sqrt{\pi}\,\Gamma[\nu+(1/2)]}\int_0^{\pi/2}\cos(z\cos t)\sin^{2\nu}t\,dt \qquad \left(\mathrm{Re}\,\nu > -\frac{1}{2}\right),$$

$$N_\nu(z) = \frac{2(z/2)^\nu}{\sqrt{\pi}\,\Gamma[\nu+(1/2)]}\left[\int_0^{\pi/2}\sin(z\sin t)\cos^{2\nu}t\,dt - \int_0^\infty e^{-z\sinh t}\cosh^{2\nu}t\,dt\right]$$

$(\mathrm{Re}\,z > 0, \quad \mathrm{Re}\,\nu > -1/2).$

Schläfli's formula

$$J_\nu(z) = \frac{1}{\pi}\int_0^\pi \cos(z\sin t - \nu t)dt - \frac{\sin\nu\pi}{\pi}\int_0^\infty e^{-z\sinh t}e^{-\nu t}dt \quad (\mathrm{Re}\,z>0),$$

$$N_\nu(z) = \frac{1}{\pi}\int_0^\pi \sin(z\sin t - \nu t)dt - \frac{1}{\pi}\int_0^\infty e^{-z\sinh t}[e^{\nu t} + (\cos\nu\pi)e^{-\nu t}]dt \quad (\mathrm{Re}\,z>0).$$

$$J_\nu(z) = \frac{z^\nu}{2\pi i}\int_{c-i\infty}^{c+i\infty}\exp\left[\frac{1}{2}\left(t - \frac{z^2}{t}\right)\right]t^{-\nu-1}dt \quad (c>0, \quad |\arg z|<\pi, \quad \mathrm{Re}\,\nu > -1).$$

$$J_\nu(x) = \frac{2(x/2)^{-\nu}}{\sqrt{\pi}\,\Gamma[(1/2)-\nu]}\int_1^\infty \frac{\sin xt}{(t^2-1)^{\nu+(1/2)}}dt,$$

$$N_\nu(x) = -\frac{2(x/2)^{-\nu}}{\sqrt{\pi}\,\Gamma[(1/2)-\nu]}\int_1^\infty \frac{\cos xt}{(t^2-1)^{\nu+(1/2)}}dt \quad \left(x>0, \quad -\frac{1}{2} < \mathrm{Re}\,\nu < \frac{1}{2}\right).$$

$$J_\nu(z) = \frac{1}{2\pi i}\int_{-\infty}^{(0+)} e^{z[t-(1/t)]/2}t^{-\nu-1}dt \quad (\mathrm{Re}\,z>0).$$

(The contour goes once around the negative real axis in the positive direction.)

Sommerfeld's formula

$$J_\nu(z) = \frac{1}{2\pi}\int_{-\eta+i\infty}^{2\pi-\eta+i\infty} e^{iz\cos t}e^{i\nu[t-(\pi/2)]}dt,$$

$$H_\nu^{(1)}(z) = \frac{1}{\pi}\int_{-\eta+i\infty}^{\eta-i\infty} e^{iz\cos t}e^{i\nu[t-(\pi/2)]}dt,$$

$$H_\nu^{(2)}(z) = \frac{1}{\pi}\int_{\eta-i\infty}^{2\pi-\eta+i\infty} e^{iz\cos t}e^{i\nu[t-(\pi/2)]}dt \qquad (-\eta < \arg z < \pi-\eta, \qquad 0<\eta<\pi).$$

$$H_\nu^{(1)}(z) = -\frac{2i}{\pi}e^{-i\nu\pi/2}\int_0^\infty e^{iz\cosh t}\cosh\nu t\,dt \quad (0<\arg z<\pi; \text{ when } \nu=0, \text{ it holds also at } z=0).$$

$$H_\nu^{(1)}(z) = -\frac{2ie^{-i\nu\pi}(z/2)^\nu}{\sqrt{\pi}\,\Gamma[\nu+(1/2)]}\int_0^\infty e^{iz\cosh t}\sin^{2\nu}t\,dt$$

$(0<\arg z<\pi, \quad \mathrm{Re}\,\nu > -1/2; \text{ when } z=0, \ -1/2<\mathrm{Re}\,\nu<1/2).$

$$H_\nu^{(1)}(z) = -i\frac{e^{-i\nu\pi/2}}{\pi}\int_0^\infty e^{iz[t-(1/t)]/2}t^{-\nu-1}\,dt \qquad (0<\arg z<\pi;\ \text{when } \arg z=0,\ -1<\operatorname{Re}\nu<1).$$

(3) Generating Function.

$$\exp\left[\frac{z(t-t^{-1})}{2}\right] = J_0(z) + \sum_{n=1}^{\infty}[t^n+(-t)^{-n}]J_n(z),$$

$$\exp(iz\cos\theta) = \sum_{n=-\infty}^{\infty} i^n J_n(z)e^{in\theta} = J_0(z) + 2\sum_{n=1}^{\infty} i^n J_n(z)\cos n\theta.$$

$$\int J_\nu(z)\,dz = 2\sum_{n=0}^{\infty} J_{\nu+2n+1}(z).$$

Kapteyn's series $\quad \dfrac{1}{1-z} = 1 + 2\sum_{n=1}^{\infty} J_n(nz),$

$$\frac{1}{2}\frac{z^2}{1-z^2} = \sum_{n=1}^{\infty} J_{2n}(2nz) \qquad \left(\left|\frac{z\exp\sqrt{1-z^2}}{1+\sqrt{1-z^2}}\right|<1\right).$$

Schlömilch's series. Supposing that $f(x)$ is twice continuously differentiable with respect to the real variable x in $0\leqslant x\leqslant\pi$, we have

$$f(x) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n J_0(nx) \quad (0<x<\pi),$$

where $\quad a_0 \equiv 2f(0) + \dfrac{2}{\pi}\displaystyle\int_0^\pi du\int_0^{\pi/2} f'(u\sin\varphi)\,d\varphi,$

$$a_n \equiv \frac{2}{\pi}\int_0^\pi du\int_0^{\pi/2} uf'(u\sin\varphi)\cos n\varphi\,d\varphi.$$

$$1 = J_0(z) + 2\sum_{n=1}^{\infty} J_{2n}(z) = [J_0(z)]^2 + 2\sum_{n=1}^{\infty}[J_n(z)]^2.$$

(4) Addition Theorem. For the cylindrical function Z_ν, we have

$$e^{i\nu\psi}Z_\nu(kR) = \sum_{n=-\infty}^{\infty} J_n(k\rho)Z_{\nu+n}(kr)e^{in\varphi}$$

$$\left(R=\sqrt{r^2+\rho^2-2r\rho\cos\varphi},\quad 0<\psi<\frac{\pi}{2},\qquad e^{2i\psi}=\frac{r-\rho e^{-i\varphi}}{r-\rho e^{i\varphi}},\quad 0<\rho<r,\right.$$

k is an arbitrary complex number),

$$\frac{Z_\nu(kR)}{R^\nu} = 2^\nu k^{-\nu}\Gamma(\nu)\sum_{m=0}^{\infty}(\nu+m)\frac{J_{\nu+m}(k\rho)}{\rho^\nu}\frac{Z_{\nu+m}(kr)}{r^\nu}C_m^{(\nu)}(\cos\varphi)$$

($\nu\neq$ negative integer).

$$\frac{\exp[(-1)^{\iota+1}ikR]}{R} = \frac{\pi}{2}\frac{(-1)^{\iota+1}i}{\sqrt{r\rho}}\sum_{m=0}^{\infty}(2m+1)^{\iota}J_{m+(1/2)}(k\rho)H_{m+(1/2)}^{(\iota)}(kr)P_m(\cos\varphi)$$

$(\iota=1,2)$.

$$e^{ik\rho\cos\varphi} = \left(\frac{\pi}{2k\rho}\right)^{1/2}\sum_{m=0}^{\infty} i^m(2m+1)J_{m+(1/2)}(k\rho)P_m(\cos\varphi)$$

$$= 2^\nu\Gamma(\nu)\sum_{m=0}^{\infty}(\nu+m)i^m J_{\nu+m}(k\rho)(k\rho)^{-\nu}C_m^{(\nu)}(\cos\varphi) \qquad (\nu\neq 0,-1,-2,\ldots),$$

where P_m is a Legendre polynomial, and $C_m^{(\nu)}$ is a Gegenbauer polynomial.

(5) Infinite Products and Partial Fractions. Let $j_{\nu,n}$ be the zeros of $z^{-\nu}J_\nu(z)$ in ascending order with respect to the real part. We have

$$J_\nu(z) = \frac{(z/2)^\nu}{\Gamma(\nu+1)}\prod_{n=1}^{\infty}\left(1-\frac{z^2}{j_{\nu,n}^2}\right) \qquad (\nu\neq -1,-2,-3,\ldots).$$

Note that if ν is real and greater than -1, all zeros are real.

Kneser-Sommerfeld formula

$$\frac{\pi J_\nu(xz)}{4J_\nu(z)}[J_\nu(z)N_\nu(Xz)-N_\nu(z)J_\nu(Xz)]=\sum_{n=1}^{\infty}\frac{J_\nu(j_{\nu,n}x)J_\nu(j_{\nu,n}X)}{(z^2-j_{\nu,n}^2)J_{\nu,n}'^2(j_{\nu,n})}$$

$$(0<x<X<1,\quad \operatorname{Re} z>0).$$

(6) Definite Integrals.

$$\int_0^{\pi/2}J_\nu(z\cos\theta)\cos\theta\,d\theta=\frac{1}{2z}\int_0^{2z}J_{2\nu}(t)\,dt,\qquad \int_0^z J_\mu(t)\,dt=\frac{1}{\pi}\int_0^\pi\frac{\sin(z\sin\theta)}{\sin\theta}\cos\mu\theta\,d\theta.$$

$$\int_0^\infty e^{-at}J_\nu(bt)t^{\mu-1}\,dt=\frac{(b/2a)^\nu\Gamma(\mu+\nu)}{a^\mu\Gamma(\nu+1)}{}_2F_1\left(\frac{\mu+\nu}{2},\ \frac{\mu+\nu+1}{2};\ \nu+1;\ \frac{-b^2}{a^2}\right)$$

$$(\operatorname{Re}(a+ib)>0,\quad \operatorname{Re}(a-ib)>0,\quad \operatorname{Re}(\mu+\nu)>0).$$

$$\int_0^\infty e^{-at}J_\nu(bt)t^\nu\,dt=\frac{(2x)^\nu\Gamma[\nu+(1/2)]}{(a^2+b^2)^{\nu+(1/2)}\sqrt{\pi}}\qquad\left(\operatorname{Re}\nu>-\frac{1}{2},\quad \operatorname{Re}a>|\operatorname{Im}b|\right).$$

$$\int_0^\infty e^{-at}J_\nu(bt)\frac{dt}{t}=\frac{(\sqrt{a^2+b^2}\ -a)^\nu}{\nu b^\nu}\qquad(\operatorname{Re}\nu>0,\quad \operatorname{Re}a>|\operatorname{Im}b|).$$

Sommerfeld's formula
$$\int_0^\infty J_0(\tau r)e^{-|x|\sqrt{\tau^2-k^2}}\frac{\tau\,d\tau}{\sqrt{\tau^2-k^2}}=\frac{e^{ik\sqrt{r^2+k^2}}}{\sqrt{r^2+x^2}}$$

$$(r,x\text{ real};\quad -\pi/2\leqslant\arg\sqrt{\tau^2-k^2}\ <\pi/2,\quad 0\leqslant\arg k<\pi).$$

Weyrich's formula
$$\frac{i}{2}\int_{-\infty}^{+\infty}e^{i\tau x}H_0^{(1)}(r\sqrt{k^2-\tau^2}\ )\,d\tau=\frac{e^{ik\sqrt{r^2+x^2}}}{\sqrt{r^2+x^2}}$$

$$(r,x\text{ real};\quad 0\leqslant\arg\sqrt{k^2-\tau^2}\ <\pi,\quad 0\leqslant\arg k<\pi).$$

Weber-Sonine formula

$$\int_0^\infty J_\nu(at)e^{-p^2t^2}t^{\mu-1}\,dt=\frac{(a/2p)^\nu\Gamma[(\nu+\mu)/2]}{2p^\mu\Gamma(\nu+1)}{}_1F_1\left(\frac{\nu+\mu}{2};\ \nu+1;\ \frac{-a^2}{4p^2}\right)$$

$$(\operatorname{Re}(\mu+\nu)>0,\quad |\arg p|<\pi/4,\quad a>0),$$

$$\int_0^\infty J_\nu(at)e^{-p^2t^2}t^{\nu+1}\,dt=\frac{a^\nu}{(2p^2)^{\nu+1}}e^{-a^2/4p^2}\qquad(\operatorname{Re}\nu>-1,\quad |\arg p|<\pi/4).$$

Sonine-Schafheitlin formula

$$\int_0^\infty J_\mu(at)J_\nu(bt)t^{-\lambda}\,dt=\frac{a^\mu\Gamma[(\mu+\nu-\lambda+1)/2]}{2^\lambda b^{\mu-\lambda+1}\Gamma[(-\mu+\nu+\lambda+1)/2]\Gamma(\mu+1)}$$

$$\times{}_2F_1\left(\frac{\mu+\nu-\lambda+1}{2},\ \frac{\mu-\nu-\lambda+1}{2};\ \mu+1;\ \frac{a^2}{b^2}\right)$$

$$(\operatorname{Re}(\mu+\nu-\lambda+1)>0,\quad \operatorname{Re}\lambda>-1,\quad 0<a<b).$$

(7) Asymptotic Expansion.

(i) Hankel's asymptotic representation. We put

$$(\nu,m)\equiv\frac{[4\nu^2-1^2][4\nu^2-3^2]\dots\left[4\nu^2-(2m-1)^2\right]}{2^{2m}m!}\qquad(m=1,2,3,\dots);\qquad(\nu,0)\equiv1.$$

For $|z|\gg|\nu|,\quad |z|\gg1,$

$$J_\nu(z)=\sqrt{\frac{2}{\pi z}}\ \cos\left(z-\frac{\nu\pi}{2}-\frac{\pi}{4}\right)\left[\sum_{m=0}^{M-1}(-1)^m\frac{(\nu,2m)}{(2z)^{2m}}+O(|z|^{-2M})\right]$$

$$-\sqrt{\frac{2}{\pi z}}\ \sin\left(z-\frac{\nu\pi}{2}-\frac{\pi}{4}\right)\left[\sum_{m=0}^{M-1}(-1)^m\frac{(\nu,2m+1)}{(2z)^{2m+1}}+O(|z|^{-2M-1})\right]$$

$$(-\pi<\arg z<\pi),$$

$$N_\nu(z)=\sqrt{\frac{2}{\pi z}}\,\sin\left(z-\frac{\nu\pi}{2}-\frac{\pi}{4}\right)\left[\sum_{m=0}^{M-1}(-1)^m\frac{(\nu,2m)}{(2z)^{2m}}+O(|z|^{-2M})\right]$$

$$+\sqrt{\frac{2}{\pi z}}\,\cos\left(z-\frac{\nu\pi}{2}-\frac{\pi}{4}\right)\left[\sum_{m=0}^{M-1}(-1)^m\frac{(\nu,2m+1)}{(2z)^{2m+1}}+O(|z|^{-2M-1})\right]$$

$$(-\pi<\arg z<\pi),$$

$$H_\nu^{(1)}(z)=\sqrt{\frac{2}{\pi z}}\,\exp\left[i\left(z-\frac{\nu\pi}{2}-\frac{\pi}{4}\right)\right]\left[\sum_{m=0}^{M-1}\frac{(\nu,m)}{(-2iz)^m}+O(|z|^{-M})\right]\qquad(-\pi<\arg z<2\pi),$$

$$H_\nu^{(2)}(z)=\sqrt{\frac{2}{\pi z}}\,\exp\left[-i\left(z-\frac{\nu\pi}{2}-\frac{\pi}{4}\right)\right]\left[\sum_{m=0}^{M-1}\frac{(\nu,m)}{(2iz)^m}+O(|z|^{-M})\right]\qquad(-2\pi<\arg z<\pi).$$

(ii) Debye's asymptotic representation.

$\nu\doteqdot x,\ 1-(\nu/x)>\varepsilon,\ \nu/x=\sin\alpha$, when $1-(\nu/x)>(3/x)\nu^{1/2}$,

$$H_\nu^{(1)}(x)\sim\frac{1}{\sqrt{\pi}}\exp\left[ix\left\{\cos\alpha+\left(\alpha-\frac{\pi}{2}\right)\sin\alpha\right\}\right]$$

$$\times\left[\frac{e^{i\pi/4}}{X}+\left(\frac{1}{8}+\frac{5}{24}\tan^2\alpha\right)\frac{3e^{3\pi i/4}}{2X^3}\right.$$

$$\left.+\left(\frac{3}{128}+\frac{77}{576}\tan^2\alpha+\frac{385}{3456}\tan^4\alpha\right)\frac{3\cdot5e^{5\pi i/4}}{2^2X^5}+\ldots\right]$$

$$(X=[-x\cos(\alpha/2)]^{1/2}).$$

$\nu\doteqdot x,\quad(\nu/x)-1>\varepsilon,\quad\nu/x=\cosh\sigma,\quad$ when $\quad|\nu^2-x^2|^{1/2}\gg1,\quad|\nu^2-x^2|^{3/2}\nu^{-2}\gg1$

$$H_\nu^{(1)}(x)\sim\frac{1}{\sqrt{\pi}}\exp[x(\sigma\cosh\sigma-\sinh\sigma)]$$

$$\times\left[\frac{1}{X}+\left(\frac{1}{8}-\frac{5}{24}\coth^2\sigma\right)\frac{3}{2X^3}+\left(\frac{3}{128}-\frac{77}{576}\coth^2\sigma+\frac{385}{3456}\coth^4\sigma\right)\frac{3\cdot5}{2^2X^5}+\ldots\right]$$

$$(X=[-x\sinh(\sigma/2)]^{1/2}).$$

When $\nu\doteqdot x,\quad|x-\nu|\ll x^{1/3},\quad x\gg1,\quad x-\nu=\delta,$

$$H_\nu^{(2)}(x)\sim\frac{6^{1/3}e^{i\pi/3}}{3^{1/2}\pi}\left[\frac{\Gamma(1/3)}{x^{1/3}}-6^{1/3}e^{i\pi/3}\delta\frac{\Gamma(2/3)}{x^{2/3}}+\left(\frac{2}{5}\delta-\delta^3\right)\frac{\Gamma(4/3)}{x^{4/3}}\right.$$

$$\left.+\left(\frac{3}{140}-\frac{\delta^2}{4}+\frac{\delta^4}{4}\right)6^{1/3}e^{i\pi/3}\frac{\Gamma(5/3)}{x^{5/3}}+\ldots\right]$$

(iii) Watson-Nicholson formula. When $x,\nu>0,\quad w=[(x/\nu)^2-1]^{1/2}$,

$$H_\nu^{(\iota)}(x)=3^{-1/2}w\exp[(-1)^{\iota+1}i\{(\pi/6)+\nu(w-(w^3/3)-\arctan w)\}]H_{1/3}^{(\iota)}(\nu w^3/3)+O|\nu^{-1}|$$

$$(\iota=1,2).$$

(IV) Functions Related to Bessel Functions

(1) Modified Bessel Functions.

$$I_\nu(z)\equiv e^{-i\nu\pi/2}J_\nu(e^{i\pi/2}z)$$

$$=\sum_{n=0}^{\infty}\frac{(z/2)^{\nu+2n}}{n!\Gamma(\nu+n+1)},$$

$$K_\nu(z)\equiv\frac{i\pi}{2}e^{i\nu\pi/2}H_\nu^{(1)}(e^{i\pi/2}z)=-\frac{i\pi}{2}e^{-i\nu\pi/2}H_\nu^{(2)}(e^{-i\pi/2}z)$$

$$=\frac{\pi}{2}\frac{I_{-\nu}(z)-I_\nu(z)}{\sin\nu\pi}=\left(\frac{\pi}{2z}\right)^{1/2}W_{0,\nu}(2z).$$

Recurrence formulas:

$$I_{\nu-1}(z)-I_{\nu+1}(z)=(2\nu/z)I_\nu(z),$$
$$I_{\nu-1}(z)+I_{\nu+1}(z)=2I_\nu'(z).$$
$$K_{\nu-1}(z)-K_{\nu+1}(z)=-(2\nu/z)K_\nu(z),$$
$$K_{\nu-1}(z)+K_{\nu+1}(z)=-2K_\nu'(z),$$
$$K_{-\nu}(z)=K_\nu(z).$$

Airy's integral: $\displaystyle\int_0^\infty \cos(t^3-tx)\,dt=\frac{\pi}{3}\sqrt{\frac{x}{3}}\left[J_{1/3}\left(\frac{2x\sqrt{x}}{3\sqrt{3}}\right)+J_{-1/3}\left(\frac{2x\sqrt{x}}{3\sqrt{3}}\right)\right],$

$$\int_0^\infty \cos(t^3+tx)\,dt=\frac{1}{3}\sqrt{x}\,K_{1/3}\left(\frac{2x\sqrt{x}}{3\sqrt{3}}\right)\qquad (x>0).$$

H. Weber's formula: $\displaystyle\frac{1}{2p^2}e^{-(a^2+b^2)/4p^2}I_\nu\left(\frac{ab}{2p^2}\right)=\int_0^\infty e^{-p^2t^2}J_\nu(at)J_\nu(bt)\,t\,dt$

$$(\operatorname{Re}\nu>-1,\quad |\arg p|<\pi/4;\quad a,b>0).$$

Watson's formula: $\displaystyle J_\mu(z)N_\nu(z)-J_\nu(z)N_\mu(z)=\frac{4\sin(\mu-\nu)\pi}{\pi^2}\int_0^\infty K_{\nu-\mu}(2z\sinh t)e^{(\mu+\nu)t}\,dt$

$$(\operatorname{Re}z>0,\quad \operatorname{Re}(\mu-\nu)<1),$$

$$J_\nu(z)\frac{\partial N_\nu(z)}{\partial\nu}-N_\nu(z)\frac{\partial J_\nu(z)}{\partial\nu}=-\frac{4}{\pi}\int_0^\infty K_0(2z\sinh t)e^{-2\nu t}\,dt\qquad(\operatorname{Re}z>0).$$

Nicholson's formula: $\displaystyle J_\nu^2(z)+N_\nu^2(z)=\frac{8}{\pi^2}\int_0^\infty K_0(2z\sinh t)\cosh 2\nu t\,dt\qquad(\operatorname{Re}z>0).$

Dixon-Ferrar formula: $\displaystyle J_\nu^2(z)+N_\nu^2(z)=\frac{8\cos\nu\pi}{\pi^2}\int_0^\infty K_{2\nu}(2z\sinh t)\,dt$

$$\left(\operatorname{Re}z>0;\qquad -\frac{1}{2}<\operatorname{Re}\nu<\frac{1}{2}\right).$$

(2) Kelvin Functions. $\quad \operatorname{ber}_\nu(z)\pm i\operatorname{bei}_\nu(z)\equiv J_\nu(e^{\pm 3\pi i/4}z),$

$$\operatorname{her}_\nu(z)\pm i\operatorname{hei}_\nu(z)\equiv H_\nu^{(1)}(e^{\pm 3\pi i/4}z),$$
$$\operatorname{ker}_\nu(z)\equiv-(\pi/2)\operatorname{hei}_\nu(z),$$
$$\operatorname{kei}_\nu(z)\equiv(\pi/2)\operatorname{her}_\nu(z).$$

When ν is an integer n, $\quad \operatorname{ber}_n(x)-i\operatorname{bei}_n(x)=(-1)^nJ_n(\sqrt{i}\,x),$

$$\operatorname{her}_n(x)-i\operatorname{hei}_n(x)=(-1)^{n+1}H_n^{(1)}(\sqrt{i}\,x)\qquad(x\text{ real}).$$

(3) Struve Function. $\quad \displaystyle \boldsymbol{H}_\nu(x)\equiv\frac{2(z/2)^\nu}{\Gamma[\nu+(1/2)]\sqrt{\pi}}\int_0^{\pi/2}\sin(z\cos\theta)\sin^{2\nu}\theta\,d\theta$

$$=\sum_{m=0}^{\infty}\frac{(-1)^m(z/2)^{\nu+2m+1}}{\Gamma[m+(3/2)]\Gamma[\nu+m+(3/2)]}.$$

Anger function: $\displaystyle \boldsymbol{J}_\nu(z)\equiv\frac{1}{\pi}\int_0^\pi\cos(\nu\theta-z\sin\theta)\,d\theta.$

H. F. Weber function: $\displaystyle \boldsymbol{E}_\nu(z)\equiv\frac{1}{\pi}\int_0^\pi\sin(\nu\theta-z\sin\theta)\,d\theta.$

Putting $\displaystyle \nabla_\nu\equiv z^2\frac{d^2}{dz^2}+z\frac{d}{dz}+z^2-\nu^2,$

$$\nabla_\nu\boldsymbol{H}_\nu(z)=\frac{4(z/2)^{\nu+1}}{\Gamma[\nu+(1/2)]\sqrt{\pi}},\quad \nabla_\nu\boldsymbol{J}_\nu(z)=\frac{(z-\nu)\sin\nu\pi}{\pi},$$

$$\nabla_\nu\boldsymbol{E}_\nu(z)=-\frac{z+\nu}{\pi}-\frac{(z-\nu)\cos\nu\pi}{\pi}.$$

When ν is an integer n, $J_n(z)=J_n(z)$.

$$\int_0^z J_0(t)\,dt = zJ_0(z)+\frac{\pi z}{2}[J_1(z)\boldsymbol{H}_0(z)-J_0(z)\boldsymbol{H}_1(z)],$$

$$\int_0^z N_0(t)\,dt = zN_0(z)+\frac{\pi z}{2}[N_1(z)\boldsymbol{H}_0(z)-N_0(z)\boldsymbol{H}_1(z)].$$

(4) Neumann Polynomials. $O_n(t)=\sum_{j=0}^{[n/2]}\frac{n(n-j-1)!}{j!(t/2)^{n-2j+1}}$ (n is a positive integer),

$$O_0(t)=1/t.$$

$$\frac{1}{t-z}\equiv 1+2\sum_{n=1}^{\infty}O_n(t)J_n(z) \qquad (|t|>|z|).$$

Schläfli polynomials: $S_n(t)\equiv\frac{2}{n}\left[tO_n(t)-\cos^2\frac{n\pi}{2}\right]$ (n is a positive integer),

$$S_0(t)\equiv 0.$$

$\nabla_n S_n(x)=2n+2(x-n)\sin^2(n\pi/2)$ (∇_n is the same operator defined in (3)).

Lommel polynomials: $$R_{m,\nu}(z)\equiv\frac{\Gamma(\nu+m)}{\Gamma(\nu)(z/2)^m}\,{}_2F_3\left(\frac{1-m}{2},\ \frac{-m}{2};\ \nu,\ -m,\ 1-\nu-m;\ -z^2\right)$$

$$=(\pi z/2\sin\nu\pi)[J_{\nu+m}(z)J_{-\nu+1}(z)+(-1)^m J_{-\nu-m}(z)J_{\nu-1}(z)]$$

(m is a nonnegative integer).

References

See references to Table 16, this Appendix.

20. Systems of Orthogonal Functions (→ 317 Orthogonal Functions)

$$\int_a^b p_n(x)p_m(x)\varphi(x)\,dx=\delta_{nm}A_n$$

Name	Notation $p_n(x)$	Interval (a,b)	Weight $\varphi(x)$	Norm A_n
Legendre	$P_n(x)$	$(-1,+1)$	1	$2/(2n+1)$
Gegenbauer	$C_n^\nu(x)$	$(-1,+1)$	$(1-x^2)^{\nu-(1/2)}$	$2\pi\Gamma(2\nu+n)/2^{2\nu}(n+\nu)n![\Gamma(\nu)]^2$
Chebyshev	$T_n(x)$	$(-1,+1)$	$(1-x^2)^{-1/2}$	$\pi\,(n=0);\ \pi/2\ (n\geqslant 1)$
Hermite	$H_n(x)$	$(-\infty,+\infty)$	e^{-x^2}	$\sqrt{\pi}\cdot n!$
Jacobi	$G_n(\alpha,\gamma;x)$	$(0,1)$	$x^{\gamma-1}(1-x)^{\alpha-\gamma}$	$\dfrac{n![\Gamma(\gamma)]^2\Gamma(\alpha+n-\gamma+1)}{(\alpha+2n)\Gamma(\alpha+n)\Gamma(\gamma+n)}$
Laguerre	$L_n^\alpha(x)$	$(0,\infty)$	$x^\alpha e^{-x}$	$\Gamma(\alpha+n+1)/n!$

For Legendre polynomials $P_n(x)$ → Table 18.II, this Appendix.

(I) Gegenbauer Polynomials (Gegenbauer Functions)

$$C_n^\nu(t)\equiv\frac{\Gamma(n+2\nu)}{n!\Gamma(2\nu)}\,{}_2F_1\left(n+2\nu,\ \ -n;\ \ \nu+\frac{1}{2};\ \ \frac{1-t}{2}\right)$$

$$=\frac{\Gamma(2\nu+n)\Gamma[\nu+(1/2)]}{\Gamma(2\nu)n!}\left[\frac{1}{4}(t^2-1)\right]^{(1/4)-(\nu/2)}\mathfrak{P}^{(1/2)-\nu}_{n+\nu-(1/2)}(t).$$

Generating function $(1-2\alpha t+\alpha^2)^{-\nu}\equiv\sum_{n=0}^{\infty}C_n^{\nu}(t)\alpha^n$. $C_{n-l}^{l+(1/2)}(x)=\frac{1}{(2l-1)!!}\frac{d^lP_n(t)}{dt^l}$.

Orthogonal relation $\int_0^{\pi}(\sin^{2\nu}\theta)C_m^{\nu}(\cos\theta)C_n^{\nu}(\cos\theta)d\theta=\frac{\pi\Gamma(2\nu+n)}{2^{2\nu-1}(\nu+n)n![\Gamma(\nu)]^2}\delta_{nm}$.

(II) Chebyshev (Tschebyscheff) Polynomials

(1) Chebyshev Polynomial (Chebyshev Function of the First Kind)

$$
\begin{aligned}
T_n(x)&\equiv\cos(n\arccos x)\\
&=(1/2)\left[\left(x+i\sqrt{1-x^2}\right)^n+\left(x-i\sqrt{1-x^2}\right)^n\right]\\
&=F(n,-n;\ 1/2;\ (1-x)/2)\\
&=\sum_{j=0}^{[n/2]}(-1)^j\binom{n}{2j}x^{n-2j}(1-x^2)^j\\
&=\frac{(-1)^n(1-x^2)^{1/2}}{(2n-1)!!}\frac{d^n(1-x^2)^{n-(1/2)}}{dx^n},
\end{aligned}
$$

Chebyshev function of the second kind

$$
\begin{aligned}
U_n(x)&\equiv\sin(n\arccos x)\\
&=(1/2i)\left[\left(x+i\sqrt{1-x^2}\right)^n-\left(x-i\sqrt{1-x^2}\right)^n\right]\\
&=\frac{(-1)^{n-1}n}{(2n-1)!!}\frac{d^{n-1}(1-x^2)^{n-(1/2)}}{dx^{n-1}}.
\end{aligned}
$$

$T_n(x)$, $U_n(x)$ are mutually linearly independent solutions of Chebyshev's differential equation $(1-x^2)y''-xy'+n^2y=0$. Recurrence relations are

$$T_{n+1}(x)-2xT_n(x)+T_{n-1}(x)=0,\quad U_{n+1}(x)-2xU_n(x)+U_{n-1}(x)=0.$$

Generating function:

$$\frac{1-t^2}{1-2tx+t^2}=T_0(x)+2\sum_{n=0}^{\infty}T_n(x)t^n,\quad \frac{1}{1-2tx+t^2}=\frac{1}{\sqrt{1-x^2}}\sum_{n=0}^{\infty}U_{n+1}(x)t^n.$$

Orthogonal relation:

$$\int_{-1}^{+1}\frac{T_m(x)T_n(x)}{\sqrt{1-x^2}}dx=\begin{cases}0 & (m\neq n),\\ \pi/2 & (m=n\neq 0),\\ \pi & (m=n=0);\end{cases}\qquad \int_{-1}^{+1}\frac{U_m(x)U_n(x)}{\sqrt{1-x^2}}dx=\begin{cases}\frac{\pi}{2} & (m=n\neq 0),\\ 0 & (\text{otherwise}).\end{cases}$$

Orthogonality in finite sums. Let $u_0, u_1, \ldots, u_k$ be the zeros of $T_{k+1}(x)$. All zeros are real and situated in the interval $(-1, 1)$. Then we have

$$\sum_{i=0}^{k}T_m(u_i)T_n(u_i)=\begin{cases}0 & (m\neq n,\text{ or } m=n=k+1),\\ (k+1)/2 & (1\leqslant m=n\leqslant k),\\ k+1 & (m=n=0).\end{cases}$$

Let $p_n(x)$ be the best approximation of x^n in $-1\leqslant x\leqslant 1$ by polynomials of degree at most $n-1$. Then we have $x^n-p_n(x)=2^{-n+1}T_n(x)$.

(2) Expansions by $T_n(x)$.

$$e^{ax}=I_0(a)+2\sum_{n=1}^{\infty}I_n(a)T_n(x),$$

$$\sin ax=2\sum_{n=0}^{\infty}(-1)^nJ_{2n+1}(a)T_{2n+1}(x),$$

$$\cos ax=J_0(a)+2\sum_{n=1}^{\infty}(-1)^nJ_{2n}(a)T_{2n}(x),$$

$$\log(1+x\sin 2\alpha)=2\log\cos\alpha-2\sum_{n=1}^{\infty}\frac{1}{n}(-\tan\alpha)^n T_n(x),$$

$$\arctan x=2\sum_{n=1}^{\infty}(-1)^n\frac{(\sqrt{2}-1)^{2n+1}}{2n+1}T_{2n+1}(x).$$

(III) Parabolic Cylinder Functions (Weber Functions) (→ 167 Functions of Confluent Type)

Parabolic cylinder functions:

$$D_\nu(z)\equiv 2^{(1/4)+(\nu/2)}z^{-1/2}W_{(1/4)+(\nu/2),\,-1/4}(z^2/2)$$

$$=\sqrt{\pi}\,2^{(1/4)+(\nu/2)}z^{-1/2}\left[\frac{M_{(1/4)+(\nu/2),\,-1/4}(z^2/2)}{\Gamma[(1-\nu)/2]}+\frac{M_{(1/4)-(\nu/2),\,-1/4}(z^2/2)}{\Gamma(-\nu/2)}\right]$$

$$=2^{\nu/2}e^{-z^2/4}\sqrt{\pi}\left[\frac{1}{\Gamma[(1-\nu)/2]}{}_1F_1\left(\frac{-\nu}{2};\frac{1}{2};\frac{z^2}{2}\right)-\frac{\sqrt{2}\,z}{\Gamma(-\nu/2)}{}_1F_1\left(\frac{1-\nu}{2};\frac{3}{2};\frac{z^2}{2}\right)\right].$$

The solutions of Weber's differential equation

$$\frac{d^2u}{dz^2}+\left(\nu+\frac{1}{2}-\frac{z^2}{4}\right)u=0$$

are given by

$$D_\nu(z),\ D_\nu(-z),\ D_{-\nu-1}(iz),\ D_{-\nu-1}(-iz),$$

and the following relations hold among them.

$$D_\nu(z)=\left[\Gamma(\nu+1)/\sqrt{2\pi}\right]\left[e^{i\nu\pi/2}D_{-\nu-1}(iz)+e^{-i\nu\pi/2}D_{-\nu-1}(-iz)\right]$$

$$=e^{-i\nu\pi}D_\nu(-z)+\left[\sqrt{2\pi}/\Gamma(-\nu)\right]e^{-i(\nu+1)\pi/2}D_{-\nu-1}(iz)$$

$$=e^{i\nu\pi}D_\nu(-z)+\left[\sqrt{2\pi}/\Gamma(-\nu)\right]e^{i(\nu+1)\pi/2}D_{-\nu-1}(-iz).$$

Integral representation:

$$D_\nu(z)=\frac{e^{-z^2/4}}{\Gamma(-\nu)}\int_0^\infty e^{-zt-(t^2/2)}t^{-\nu-1}dt\quad(\mathrm{Re}\,\nu<0).$$

$$e^{-(z^2/4)-zt-(t^2/2)}=\sum_{n=0}^{\infty}\frac{(-t)^n}{n!}D_n(z)=\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}t^\nu\Gamma(-\nu)D_\nu(z)d\nu\quad(c<0,\ |\arg t|<\pi/4).$$

Recurrence formula:

$$D_{\nu+1}(z)-zD_\nu(z)+\nu D_{\nu-1}(z)=0,\quad dD_\nu(z)/dz+(1/2)zD_\nu(z)-\nu D_{\nu-1}(z)=0.$$

$$D_\nu(0)=\frac{2^{\nu/2}\sqrt{\pi}}{\Gamma[(1-\nu)/2]},\quad D_\nu'(0)=-\frac{2^{(\nu+1)/2}\sqrt{\pi}}{\Gamma(-\nu/2)}.$$

Asymptotic expansion:

$$D_\nu(z)\approx e^{-z^2/4}z^\nu\left(1-\frac{\nu(\nu-1)}{2z^2}+\frac{\nu(\nu-1)(\nu-2)(\nu-3)}{2\cdot 4z^4}\mp\ldots\right)\quad\left(|\arg z|<\frac{3}{4}\pi\right).$$

$$D_{-1}(z)=e^{z^2/4}\sqrt{\frac{\pi}{2}}\left[1-\mathrm{erf}\left(\frac{z}{\sqrt{2}}\right)\right],\quad \mathrm{erf}(x)\equiv\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt\quad\text{(error function)}.$$

(IV) Hermite Polynomials

For the parabolic cylinder functions, when ν is an integer n, we have

$$D_n(z)=(-1)^n e^{z^2/4}d^n(e^{-z^2/2})/dz^n=e^{-z^2/4}H_n(z/\sqrt{2}),$$

where $H_n(x)$ is the Hermite polynomial

$$H_n(x) \equiv 2^{-n/2}(-1)^n e^{x^2} d^n(e^{-x^2})/dx^n = e^{x^2/2} D_n(\sqrt{2}\,x).$$

A Hermite polynomial is more often defined by the following function $\mathrm{He}_n(x)$ (e.g., in W.F. Magnus, F. Oberhettinger, and R. P. Soni [1]).

$$\mathrm{He}_n(x) \equiv (-1)^n e^{x^2/2} d^n(e^{-x^2/2})/dx^n = e^{x^2/4} D_n(x) = H_n(x/\sqrt{2}).$$

The function $y = H_n(x)$ is a solution of Hermite's differential equation

$$y'' - 2xy' + 2ny = 0.$$

$H_n(x)$ is a polynomial in x of degree n, and is an even or odd function according to whether n is even or odd.

$$H_{2n}(x) = (-1)^n (2n-1)!!\,{}_1F_1(-n; 1/2; x^2),$$

$$H_{2n+1}(x) = (-1)^n (2n+1)!!\sqrt{2}\,x\,{}_1F_1(-n; 3/2; x^2).$$

Recurrence formula:

$$H_{n+1}(x) = \sqrt{2}\,xH_n(x) - nH_{n-1}(x) = \sqrt{2}\,xH_n(x) - H_n'(x)/\sqrt{2},$$

$$H_n'(x) = \sqrt{2}\,nH_{n-1}(x).$$

$$H_{2n}(0) = \frac{(-1)^n(2n)!}{2^n n!} = (-1)^n(2n-1)!!, \quad H_{2n+1}(0) = 0.$$

Generating function:

$$e^{\sqrt{2}\,tx - (t^2/2)} = \sum_{n=0}^{\infty} H_n(x)t^n/n!.$$

Orthogonal relation:

$$\int_{-\infty}^{+\infty} H_n(x)H_m(x)e^{-x^2}dx = \delta_{nm} n!\sqrt{\pi}.$$

(V) Jacobi Polynomials

$$G_n(\alpha,\gamma;x) \equiv F(-n, \alpha+n; \gamma; x)$$

$$= x^{1-\gamma}(1-x)^{\gamma-\alpha}\frac{\Gamma(\gamma+n)}{\Gamma(\gamma)}\frac{d^n}{dx^n}\left[x^{\gamma+n-1}(1-x)^{\alpha+n-\gamma}\right].$$

These satisfy Jacobi's differential equation $x(1-x)y'' + [\gamma-(\alpha+1)x]y' + n(\alpha+n)y = 0$.

Orthogonal relation:

$$\int_0^1 x^{\gamma-1}(1-x)^{\alpha-\gamma} G_m(\alpha,\gamma;x)G_n(\alpha,\gamma;x)\,dx = \frac{n!\Gamma(\alpha+n-\gamma+1)\Gamma(\gamma)^2}{(\alpha+2n)\Gamma(\alpha+n)\Gamma(\gamma+n)}\delta_{mn}$$

$(\mathrm{Re}\gamma > 0, \quad \mathrm{Re}(\alpha-\gamma) > -1).$

Representation of other functions:

$$P_n(x) = G\left(1, 1; \frac{1-x}{2}\right), \quad T_n(x) = G\left(0, \frac{1}{2}; \frac{1-x}{2}\right),$$

$$C_n^\nu(x) = (-1)^n \frac{\Gamma(2\nu+n)}{\Gamma(2\nu)\cdot n!} G_n\left(2\nu, \nu+\frac{1}{2}; \frac{1+x}{2}\right).$$

(VI) Laguerre Functions

(1) Laguerre Functions.

$$L_\nu^{(\alpha)}(z)\equiv\frac{\Gamma(\alpha+\nu+1)}{\Gamma(\alpha+1)\Gamma(\nu+1)}z^{-(\alpha+1)/2}e^{z/2}M_{[(\alpha+1)/2]+\nu,\alpha/2}(z)$$

$$=\frac{\Gamma(\alpha+\nu+1)}{\Gamma(\alpha+1)\Gamma(\nu+1)}{}_1F_1(-\nu;\alpha+1;z),$$

These satisfy Laguerre's differential equation

$$zd^2[L_\nu^{(\alpha)}(z)]/dz^2+(\alpha+1-z)d[L_\nu^{(\alpha)}(z)]/dz+\nu L_\nu^{(\alpha)}(z)=0.$$

(2) Laguerre Polynomials. When ν is an integer n ($n=0,1,2,\ldots$), the function $L_n^{(\alpha)}(x)$ reduces to a polynomial of degree n as follows.

Laguerre polynomials:

$$L_n^{(\alpha)}(x)=\frac{e^x x^{-\alpha}}{n!}\frac{d^n}{dx^n}(e^{-x}x^{n+\alpha})=\sum_{j=0}^{n}\binom{n+\alpha}{n-j}\frac{(-x)^j}{j!}.$$

$$L_n^{(0)}(x)=1,\quad L_0^{(m)}(x)=1,\quad L_{n+m}^{(-m)}(x)=\frac{(-1)^m n!}{(n+m)!}x^m L_n^{(m)}(x)\quad(m=0,1,2,\ldots).$$

Recurrence formulas:

$$nL_n^{(\alpha)}(x)=(-x+2n+\alpha-1)L_{n-1}^{(\alpha)}(x)-(n+\alpha-1)L_{n-2}^{(\alpha)}(x),$$

$$xd[L_n^{(\alpha)}(x)]/dx=nL_n^{(\alpha)}(x)-(n+\alpha)L_{n-1}^{(\alpha)}(x)\quad(n=2,3,\ldots).$$

Generating function:

$$\frac{e^{-xt/(1-t)}}{(1-t)^{\alpha+1}}=\sum_{n=0}^{\infty}L_n^{(\alpha)}(x)t^n\quad(|t|<1).$$

Orthogonal relations:

$$\int_0^\infty e^{-x}x^\alpha L_m^{(\alpha)}(x)L_n^{(\alpha)}dx=\delta_{mn}\Gamma(\alpha+n+1)/n!=\delta_{mn}\Gamma(1+\alpha)\binom{n+\alpha}{n}.$$

$$H_{2n}(x)=(-2)^n n!L_n^{(-1/2)}(x^2),\quad H_{2n+1}(x)=(-2)^n n!\sqrt{2}\,xL_n^{(1/2)}(x^2).$$

(3) Sonine Polynomials.

$$S_n^{(\alpha)}(x)\equiv\frac{(-1)^n}{\Gamma(\alpha+n+1)}L_n^{(\alpha)}(x).$$

(VII) Orthogonal Polynomials

$$P_{n,m}(x)=\sum_{k=0}^{n}(-1)^k\binom{n}{k}\binom{n+k}{k}\frac{x(x-1)\ldots(x-k+1)}{m(m-1)\ldots(m-k+1)}$$

(where n, m are positive integers and $n\leqslant m$).

We have the same polynomials if we replace x^k in $P_n(1-2x)$ by

$$x(x-1)\ldots(x-k+1)/m(m-1)\ldots(m-k+1)\quad(k=0,1,\ldots,n).$$

Orthogonality in finite sums:

$$\sum_{k=0}^{m}P_{n,m}(k)P_{l,m}(k)=\delta_{nl}\frac{(m+n+1)!(m-n)!}{(2n+1)(m!)^2}.$$

Chebyshev's q functions:

$$q_n(m,x)=\frac{(-1)^n(m-1)!}{2^n(m-n-1)!}P_{n,m-1}(x),\quad \xi_{n,m}(x)=\left[2^n(n!)^2/(2n)!\right]q_n(m,x-1).$$

For given data y_k at m points $x_k=x_1+(k-1)h$ $(k=1,\ldots,m)$ that are equally spaced with step h, the least square approximation among the polynomials $Q(x)$ of degree $n(<m)$, i.e., the polynomial that minimizes the square sum of the residues $S=\sum_{k=1}^{m}[y_k-Q(x_k)]^2$ is given by the

following formula (→ 19 Analog Computation):

$$Q(x)=\sum_{k=0}^{n}\frac{B_k}{S_k}\xi_{k,m}\left(\frac{x-x_1}{h}+1\right),\quad S=\sum_{k=1}^{m}(y_k)^2-\sum_{k=0}^{m}\frac{B_k^2}{S_k},$$

$$B_k=\sum_{i=1}^{m}y_i\xi_{k,m}(i),\quad S_k=\sum_{i=1}^{m}[\xi_{k,m}(i)]^2.$$

References

See references to Table 16, this Appendix.

21. Interpolation (→ 223 Interpolation)

(1) Lagrange's Interpolation Polynomial.

$$f(x)=\sum_{s=0}^{n}f(x_s)\frac{(x-x_0)(x-x_1)\dots(x-x_{s-1})(x-x_{s+1})\dots(x-x_n)}{(x_s-x_0)(x_s-x_1)\dots(x_s-x_{s-1})(x_s-x_{s+1})\dots(x_s-x_n)}.$$

Aitken's interpolation scheme. The interpolation polynomial $f(x)$ corresponding to the value $y_s=f(x_s)$ $(s=0,1,\dots,n)$ is given inductively by the following procedure. The order of $x_0, x_1, \dots, x_s$ is quite arbitrary.

$$p_{s,0}(x)=y_s\quad(s=0,1,\dots,n),$$

$$p_{s,k+1}(x)=\left[(x_s-x)p_{k,k}(x)-(x_k-x)p_{s,k}(x)\right]/(x_s-x_k)\quad(s=k+1,k+2,\dots,n),$$

$$f(x)\equiv p_{n,n}(x).$$

(2) Interpolation for Equally Spaced Points. When the points x_k lie in the order of their subscripts at a uniform distance h $(x_s=x_0+sh)$, we make the following difference table $(\Delta x=h)$.
Forward difference:

$$\Delta_i\equiv\Delta_i^1\equiv f_{i+1}-f_i=f(x_{i+1})-f(x_i),\quad \Delta_i^s\equiv\Delta_{i+1}^{s-1}-\Delta_i^{s-1}.$$

Variable	Value of Function	Difference (1st)	(2nd)	(3rd)	(4th)	...
...	...					
$x_0-2\Delta x$	f_{-2}	...	...			
$x_0-\Delta x$	f_{-1}	Δ_{-2}	Δ_{-2}^2	...	...	
x_0	f_0	Δ_{-1}	Δ_{-1}^2	Δ_{-2}^3	Δ_{-2}^4	...
$x_0+\Delta x$	f_1	Δ_0	Δ_0^2	Δ_{-1}^3	Δ_{-1}^4	...
$x_0+2\Delta x$	f_2	Δ_1	Δ_1^2	Δ_0^3	...	
$x_0+3\Delta x$	f_3	Δ_2	...	...		
...	...	...				

Backward difference:

$$\bar{\Delta}_i^s\equiv\bar{\Delta}_i^{s-1}-\bar{\Delta}_{i-1}^{s-1}=\Delta_{s-i}^s.$$

Central difference:

$$\delta_i^s=\delta_{i+(1/2)}^{s-1}-\delta_{i-(1/2)}^{s-1},\quad \delta_{i+(s/2)}^s=\Delta_i^s.$$

Newton interpolation formula (forward type):

$$f(x_0+u\Delta x)=f(x_0)+\frac{u}{1!}\Delta_0+\frac{u(u-1)}{2!}\Delta_0^2+\frac{u(u-1)(u-2)}{3!}\Delta_0^3$$
$$+\frac{u(u-1)(u-2)(u-3)}{4!}\Delta_0^4+\dots.$$

Gauss's interpolation formula (forward type):

$$f(x_0+u\Delta x)=f(x_0)+\frac{u}{1!}\Delta_0+\frac{u(u-1)}{2!}\Delta_{-1}^2+\frac{u(u-1)(u+1)}{3!}\Delta_{-1}^3$$
$$+\frac{u(u-1)(u+1)(u-2)}{4!}\Delta_{-2}^4+\dots.$$

Stirling's interpolation formula:

$$f(x_0+u\Delta x)=f(x_0)+\frac{u}{1!}\frac{\Delta_{-1}+\Delta_0}{2}+\frac{u^2}{2!}\Delta^2_{-1}+\frac{u(u^2-1)}{3!}\frac{\Delta^3_{-2}+\Delta^3_{-1}}{2}+\frac{u^2(u^2-1)}{4!}\Delta^4_{-2}+\dots.$$

Bessel's interpolation formula:

$$f\left(\frac{x_0+x_1}{2}+v\Delta x\right)=\frac{f(x_0)+f(x_1)}{2}+\frac{v}{1!}\Delta_0+\frac{1}{2!}\left(v^2-\frac{1}{4}\right)\frac{\Delta^2_{-1}+\Delta^2_0}{2}+\frac{v}{3!}\left(v^2-\frac{1}{4}\right)\Delta^3_{-1}$$

$$+\frac{1}{4!}\left(v^2-\frac{1}{4}\right)\left(v^2-\frac{9}{4}\right)\frac{\Delta^4_{-2}+\Delta^4_{-1}}{2}+\dots.$$

Everett's interpolation formula:

$$f(x_0+u\Delta x)=f(x_1-\xi\Delta x)=\xi f(x_0)+\frac{\xi(\xi^2-1)}{3!}\Delta^2_{-1}+\frac{\xi(\xi^2-1)(\xi^2-4)}{5!}\Delta^4_{-2}+\dots$$

$$+uf(x_1)+\frac{u(u^2-1)}{3!}\Delta^2_0+\frac{u(u^2-1)(u^2-4)}{5!}\Delta^4_{-1}+\dots.\quad (\xi=1-u)$$

(3) Interpolation for Functions of Two Variables. Let $x_m=x_0+m\Delta x$, $y_n\equiv y_0+n\Delta y$ (m and n are integers). We define the finite differences as follows:

$$\Delta_x(x_0,y_0)\equiv f(x_1,y_0)-f(x_0,y_0),$$

$$\Delta_y(x_0,y_0)\equiv f(x_0,y_1)-f(x_0,y_0),$$

$$\Delta^2_x(x_0,y_0)\equiv\Delta_x(x_1,y_0)-\Delta_x(x_0,y_0)\equiv\delta^2_x(x_1,y_0),$$

$$\Delta_{xy}(x_0,y_0)\equiv\Delta_y(x_1,y_0)-\Delta_y(x_0,y_0)=\Delta_x(x_0,y_1)-\Delta_x(x_0,y_0),$$

$$\Delta^2_y(x_0,y_0)\equiv\Delta_y(x_0,y_1)-\Delta_y(x_0,y_0)\equiv\delta^2_y(x_0,y_1),\quad\dots.$$

Newton's formula:

$$f(x_0+u\Delta x,y_0+v\Delta y)=f(x_0,y_0)+(u\Delta_x+v\Delta_y)(x_0,y_0)$$

$$+(1/2!)\left[u(u-1)\Delta^2_x+2uv\Delta_{xy}+v(v-1)\Delta^2_y\right](x_0,y_0)+\dots.$$

Everett's formula. Putting $s\equiv1-u,t\equiv1-v$ we have

$$f(x_0+u\Delta x,y_0+v\Delta y)=stf(x_0,y_0)+svf(x_0,y_1)+utf(x_1,y_0)+uvf(x_1,y_1)$$

$$-(1/6)\left[us(1+s)\{t\delta^2_x(x_0,y_0)+v\delta^2_x(x_0,y_1)\}+us(1+u)\{t\delta^2_x(x_1,y_0)+v\delta^2_x(x_1,y_1)\}\right.$$

$$\left.+vt(1+t)\{s\delta^2_y(x_0,y_0)+u\delta^2_y(x_1,y_0)\}+vt(1+v)\{s\delta^2_y(x_0,y_1)+u\delta^2_y(x_1,y_1)\}\right]+\dots.$$

References

[1] F. J. Thompson, Table of the coefficients of Everett's central-difference interpolation formula, Tracts for computers, no. V, Cambridge Univ. Press, 1921.
[2] M. Lindow, Numerische Infinitesimalrechnung, Dummler, Berlin, 1928.
[3] H. T. Davis, Tables of the higher mathematical functions I, Principia Press, Bloomington, 1933.
[4] K. Hayashi and S. Moriguti, Table of higher transcendental functions (in Japanese), Iwanami, second revised edition, 1967.

22. Distribution of Typical Random Variables

(→ 341 Probability Measures, 374 Sampling Distributions)

In the following table, Nos. 1–13 are 1-dimensional continuous distributions, and Nos. 20–21 are k-dimensional continuous distributions, for which the distribution density is the one with respect to Lebesgue measure. Nos. 14–19 are 1-dimensional discrete distributions, and Nos. 22–24 are k-dimensional discrete distributions, where the density function $P(x)$ means the probability at the point x.

The characteristic function, average, and variance are given only for those represented in a simple form.

Distribution of Typical Random Variables

No.	Name	Symbol	Density Function	Domains
1	Normal	$N(\mu,\sigma^2)$	$\frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$	$-\infty<x<\infty$
2	Logarithmic normal		$\frac{1}{(2\pi\sigma^2)^{1/2}}\frac{1}{y}\exp\left[-\frac{(\log y-\mu)^2}{2\sigma^2}\right]$	$0<y<\infty$
3	Gamma	$\Gamma(p,\sigma)$	$[\Gamma(p)]^{-1}\sigma^{-p}x^{p-1}e^{-x/\sigma}$	$0<x<\infty$
4	Exponential	$e(\mu,\sigma)$	$(1/\sigma)\exp(-(x-\mu)/\sigma)$	$\mu<x<\infty$
5	Two-sided exponential		$(1/2\sigma)e^{-\lvert x\rvert/\sigma}$	$-\infty<x<\infty$
6	Chi square	$\chi^2(n)$	$2^{-n/2}[\Gamma(n/2)]^{-1}x^{(n/2)-1}e^{-x/2}$	$0<x<\infty$
7	Beta	$B(p,q)$	$[B(p,q)]^{-1}x^{p-1}(1-x)^{q-1}$	$0<x<1$
8	F	$F(m,n)$	$2K_F x^{(m/2)-1}[1+(mx/n)]^{-(m+n)/2}$, $K_F\equiv[B(m/2,n/2)]^{-1}(m/n)^{m/2}$	$0<x<\infty$
9	z	$z(m,n)$	$K_F e^{mz}[1+(me^{2z}/n)]^{-(m+n)/2}$, $K_F\equiv[B(m/2,n/2)]^{-1}(m/n)^{m/2}$	$-\infty<z<\infty$
10	t	$t(n)$	$[\sqrt{n}\,B(n/2,1/2)]^{-1}[1+(t^2/n)]^{-(n+1)/2}$	$-\infty<t<\infty$
11	Cauchy	$C(\mu,\sigma)$	$(\pi\sigma)^{-1}\left[1+\frac{(x-\mu)^2}{\sigma^2}\right]^{-1}$	$-\infty<x<\infty$
12	One-side stable for exponent 1/2		$c(2\pi)^{-1/2}x^{-3/2}\exp(-c^2/2x)$	$0<x<\infty$
13	Uniform rectangular	$U(\alpha,\beta)$	$1/(\beta-\alpha)$	$\alpha<x<\beta$
14	Binomial	$Bin(n,p)$	$\binom{n}{x}p^xq^{n-x}$	$x=0,1,2,\ldots,n$
15	Poisson	$P(\lambda)$	$e^{-\lambda}\lambda^x/x!$	$x=0,1,2,\ldots$
16	Hypergeometric	$H(N,n,p)$	$\binom{Np}{x}\binom{Nq}{n-x}\Big/\binom{N}{n}$	x integer $0\leq x\leq Np$, $0\leq n-x\leq Nq$
17	Negative binomial	$NB(m,p)$	$\Gamma(m+x)[\Gamma(m)x!]^{-1}p^mq^x$	$x=0,1,2,\ldots$
18	Geometric	$G(p)$	pq^x	$x=0,1,2,\ldots$
19	Logarithmic		$K_Lq^x/x,\ K_L\equiv-1/\log p$	$x=1,2,3,\ldots$
20	Multidimensional normal	$N(\mu,\Sigma)$	$(2\pi)^{-k/2}\lvert\Sigma\rvert^{-1/2}$ $\times\exp[-(x-\mu)\Sigma^{-1}(x-\mu)'/2]$, $x=(x_1,\ldots,x_k)$, $\mu=(\mu_1,\ldots,\mu_k)$, $\Sigma=(\sigma_{ij})$	$-\infty<x_1,\ldots,x_k<\infty$
21	Dirichlet		$\frac{\Gamma(\nu_1+\ldots+\nu_{k+1})}{\Gamma(\nu_1)\ldots\Gamma(\nu_{k+1})}x_1^{\nu_1-1}\ldots x_{k+1}^{\nu_{k+1}-1}$, $x_{k+1}=1-(x_1+\ldots+x_k)$	$x_1,\ldots,x_k>0$, $x_1+\ldots+x_k<1$
22	Multinomial	$M(n,(p_i))$	$n!(x_1!\ldots x_{k+1}!)^{-1}p_1^{x_1}\ldots p_{k+1}^{x_{k+1}}$, $x_{k+1}=n-(x_1+\ldots+x_k)$	$x_1,\ldots,x_k=0,1,\ldots,n$, $x_1+\ldots+x_k\leq n$
23	Multidimensional hypergeometric	$H(N,n,(p_i))$	$\binom{Np_1}{x_1}\cdots\binom{Np_{k+1}}{x_{k+1}}\Big/\binom{N}{n}$, $x_{k+1}=n-(x_1+\ldots+x_k)$	$x_1,\ldots,x_k$ integers $0\leq x_i\leq Np_i$ $(i=1,\ldots,k+1)$
24	Negative polynomial		$\frac{\Gamma(m+x_1+\ldots+x_k)}{\Gamma(m)x_1!\ldots x_k!}p_0^mp_1^{x_1}\ldots p_k^{x_k}$,	$x_1,\ldots,x_k=0,1,2,\ldots$

Conditions for Parameters	Characteristic Function	Mean	Variance	No.
$-\infty<\mu<\infty$, $\sigma>0$	$\exp\left(i\mu t-\frac{\sigma^2t^2}{2}\right)$	μ	σ^2	1
$-\infty<\mu<\infty$, $\sigma>0$		$e^{\mu+(\sigma^2/2)}$	$e^{2\mu}(e^{2\sigma^2}-e^{\sigma^2})$	2
$p,\sigma>0$	$(1-i\sigma t)^{-p}$	σp	$\sigma^2 p$	3
$-\infty<\mu<\infty$, $\sigma>0$	$e^{i\mu t}(1-i\sigma t)^{-1}$	$\mu+\sigma$	σ^2	4
$\sigma>0$	$(1+\sigma^2t^2)^{-1}$	0	$2\sigma^2$	5
n positive integer	$(1-2it)^{-n/2}$	n	$2n$	6
$p,q>0$		$\frac{p}{p+q}$	$\frac{pq}{(p+q)^2(p+q+1)}$	7
m,n positive integers		$\frac{n}{n-2}\ (n>2)$	$\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}\ (n>4)$	8
m, n positive integers				9
n positive integer		$0\ (n>1)$	$n/(n-2)\ (n>2)$	10
$-\infty<\mu<\infty$, $\sigma>0$	$\exp(i\mu t-\sigma\|t\|)$	none	none	11
$0<c<\infty$	$\exp[-c\|t\|^{1/2}(1-it/\|t\|)]$	none	none	12
$-\infty<\alpha<\beta<\infty$	$(e^{i\beta t}-e^{i\alpha t})/it(\beta-\alpha)$	$(\alpha+\beta)/2$	$(\beta-\alpha)^2/12$	13
$p+q=1, p,q>0$, n positive integer	$(pe^{it}+q)^n$	np	npq	14
$\lambda>0$	$\exp[-\lambda(1-e^{it})]$	λ	λ	15
$p+q=1, p,q>0$, N, Np, n positive integers $N\geqslant n$	$(Nq)^{[n]}(N^{[n]})^{-1} \times F(-n,-Np;\ Nq-n+1;\ e^{it})$, $m^{[n]}\equiv m!/(m-n)!$	np	$\frac{npq(N-n)}{N-1}$	16
$p+q=1, p,q>0$, $m>0$	$\frac{p^m}{(1-qe^{it})^m}$	$\frac{mq}{p}$	$\frac{mq}{p^2}$	17
$p+q=1, p,q>0$	$\frac{p}{1-qe^{it}}$	$\frac{q}{p}$	$\frac{q}{p^2}$	18
$p+q=1, p,q>0$	$-K_L\log(1-qe^{it})$	K_Lq/p	$K_Lq(1-K_Lq)/p^2$	19
$-\infty<\mu_1,\ldots,\mu_k<\infty$, Σ symmetric positive definite quadratic form	$\exp\left(i\mu t'-\frac{t\Sigma t'}{2}\right)$, $t=(t_1,\ldots,t_k)$	$E(x_i)=\mu_i$	$V(x_i)=\sigma_{ii}$, $\mathrm{Cov}(x_i,x_j)=\sigma_{ij}$	20
$\nu_1,\ldots,\nu_{k+1}>0$		$E(x_i)=\frac{\nu_i}{\nu_1+\ldots+\nu_{k+1}}$	$V(x_i)=C\nu_i(\nu_1+\ldots+\nu_{k+1}-\nu_i)$, $\mathrm{Cov}(x_i,x_j)=-C\nu_i\nu_j$, $C\equiv(\nu_1+\ldots+\nu_{k+1})^{-2}\times(\nu_1+\ldots+\nu_{k+1}+1)^{-1}$	21
$p_1+\ldots+p_{k+1}=1$, $p_1,\ldots,p_{k+1}>0$ n positive integer	$(p_1e^{it_1}+\ldots+p_ke^{it_k}+p_{k+1})^n$	$E(x_i)=np_i$	$V(x_i)=np_i(1-p_i)$, $\mathrm{Cov}(x_i,x_j)=-np_ip_j$	22
$p_1+\ldots+p_{k+1}=1$, $p_1,\ldots,p_{k+1}>0$, $N,Np_1,\ldots,Np_k,n$ positive integers		$E(x_i)=np_i$	$V(x_i)=Cnp_i(1-p_i)$, $\mathrm{Cov}(x_i,x_j)=-Cnp_ip_j$, $C\equiv\frac{N-n}{N-1}$	23
$p_0+p_1+\ldots+p_k=1, p_0,p_1,\ldots,p_k>0$, $m>0$	$p_0^m(1-p_1e^{it_1}-\ldots-p_ke^{it_k})^m$	$E(x_i)=\frac{mp_i}{p_0}$	$V(x_i)=mp_i(p_0+p_i)/p_0^2$, $\mathrm{Cov}(x_i,x_j)=mp_ip_j/p_0^2$	24

Remarks
1. Reproducing property with respect to μ, σ^2.
2. $X = \log Y : N(\mu, \sigma^2)$.
3. Reproducing property with respect to p.
4. $e(0,\sigma) = \Gamma(1,\sigma)$.
6. n is the number of degrees of freedom; reproducing property with respect to n.
8. m and n are the numbers of degrees of freedom.
9. $e^{2z} = F(m,n)$.
10. n is the number of degrees of freedom.
11. $C(0,1) = t(1)$; reproducing property with respect to μ and σ.
14. Reproducing property with respect to n.
15. Reproducing property with respect to λ.
17. Reproducing property with respect to m.
18. $G(p) = NB(1,p)$.
20. Generalization of normal distribution; reproducing property with respect to μ and Σ.
22. Generalization of binomial distribution; reproducing property with respect to n.
23. Generalization of hypergeometric distribution.
24. Generalization of negative binomial distribution; reproducing property with respect to m.

23. Statistical Estimation and Statistical Hypothesis Testing

Listed below are some frequently used and well-investigated statistical procedures. (Concerning main probability distributions → 398 Statistical Decision Functions, 399 Statistical Estimation, 400 Statistical Hypothesis Testing). The following notations and conventions are adopted, unless otherwise stated.

Immediately after the heading number, the distribution is indicated by the symbol as defined in Table 22, this Appendix. It is to be understood that a random sample $(x_1, x_2, \ldots, x_n)$ is observed from this distribution. Where two distributions are involved, samples $(x_1, \ldots, x_{n_1})$ and $(y_1, \ldots, y_{n_2})$ are understood to be observed from the respective distributions.

Next, a necessary and sufficient statistic based on the sample is marked with * when it is complete, and # otherwise. Then appears the sampling distribution of this statistic. For those statistics consisting of several independent components, the distribution of these are shown. Greek lowercase letters except α and χ denote unknown parameters. Italic lowercase letters denote constants, each taking arbitrary real values. Italic capital letters denote constants whose values are specified in each procedure; repeated occurrences of the same letter under the same heading number specify a certain common real value.

Problems of point estimation, interval estimation, and hypothesis testing are presented, with corresponding estimators, confidence intervals, and tests (critical regions) as their solutions. All the confidence intervals here are those constructed from UMP unbiased tests, having $1-\alpha$ as confidence levels. Alternative hypotheses are understood to be the negations of corresponding null hypotheses. Significance levels of all the tests are α. The following symbols are attached to each procedure to describe its properties.

For estimators:
UMV: uniformly minimum variance unbiased.
ML: maximum likelihood.
AD: admissibility with respect to quadratic loss function.
IAD: inadmissibility with respect to quadratic loss function.

For tests:
UMP: uniformly most powerful.
UMPU: uniformly most powerful unbiased.
UMPI( ): uniformly most powerful invariant with respect to the product of transformation groups shown in ( ).
LR: likelihood ratio.
O: group of orthogonal transformations.
L: group of shift transformations.
S: group of change of scales.
AD: admissibility with respect to simple loss function.

IAD: inadmissibility with respect to simple loss function. (Note that UMPU implies AD.)

The following symbols denote $100(1-\alpha)\%$ points of respective distributions, α being sufficiently small.

$u(\alpha)$: standard normal distribution.
$t_f(\alpha)$: t-distribution with f degrees of freedom.
$\chi_f^2(\alpha)$: χ^2 distribution with f degrees of freedom.
$F_{f_2}^{f_1}(\alpha)$: F-distribution with (f_1,f_2) degrees of freedom.

(1) $N(\mu,b^2)$. Σx_i*. $N(n\mu,nb^2)$.
Point estimation of μ. $\bar{x}=\frac{1}{2}\Sigma x_i$: UMV, ML, AD.
Interval estimation of μ. $\left(\bar{x}\pm u(\alpha/2)\dfrac{b}{\sqrt{n}}\right)$.
Hypothesis $[\mu\leqslant k]$. $\bar{x}>k+u(\alpha)\dfrac{b}{\sqrt{n}}$: UMP, LR.
Hypothesis $[h\leqslant\mu\leqslant l]$. $\bar{x}<h-C$ or $\bar{x}>l+C$: UMPU, LR.

(2) $N(a,\sigma^2)$. $\Sigma(x_i-a)^2$*. $\sigma^2\chi_n^2$. ($\sigma^2\chi_n^2$ is the σ^2-multiplication of a random variable obeying the $\chi^2(n)$ distribution. We use similar notations in the following.)

Point estimation of σ^2. $\dfrac{\Sigma(x_i-a)^2}{n}$: UMV, ML, IAD.
Interval estimation of σ^2. $(A\Sigma(x_i-a)^2,\ B\Sigma(x_i-a)^2)$.
Hypothesis $[\sigma^2\leqslant k]$. $\Sigma(x_i-a)^2>\chi_n^2(\alpha)k$: UMP, LR.
Hypothesis $[\sigma^2=k]$. $\Sigma(x_i-a)^2<Ak$ or $\Sigma(x_i-a)^2>Bk$: UMPU.

(3) $N(\mu,\sigma^2)$. $\begin{pmatrix}\Sigma x_i\\ \Sigma(x_i-\bar{x})^2\end{pmatrix}^*$. $\begin{pmatrix}N(n\mu,n\sigma^2)\\ \sigma^2\chi_{n-1}^2\end{pmatrix}$.
Point estimation of μ. $\bar{x}$: UMV, ML, AD.
Interval estimation of μ. $\left[\bar{x}\pm t_{n-1}(\alpha/2)\dfrac{\sqrt{\Sigma(x_i-\bar{x})^2}}{\sqrt{n(n-1)}}\right]$.
Hypothesis $[\mu\leqslant k]$. $\dfrac{\bar{x}-k}{\sqrt{\Sigma(x_i-\bar{x})^2}}>\dfrac{t_{n-1}(\alpha)}{\sqrt{n(n-1)}}$: UMPU, LR.
Hypothesis $[\mu=k]$. $\dfrac{|\bar{x}-k|}{\sqrt{\Sigma(x_i-\bar{x})^2}}>\dfrac{t_{n-1}(\alpha/2)}{\sqrt{n(n-1)}}$: UMPU, LR, UMPI(S, O) for $k=0$.
Point estimation of σ^2. $\dfrac{\Sigma(x_i-\bar{x})^2}{n-1}$: UMV, IAD. $\dfrac{\Sigma(x_i-\bar{x})^2}{n}$: ML, IAD.
Point estimation of σ. $\dfrac{\Gamma[(n-1)/2]}{\sqrt{2}\,\Gamma(n/2)}\sqrt{\dfrac{\Sigma(x_i-\bar{x})^2}{n-1}}$: UMV, IAD.
Interval estimation of σ^2. $(A\Sigma(x_i-\bar{x})^2,\ B\Sigma(x_i-\bar{x})^2)$.
Hypothesis $[\sigma^2\leqslant k]$. $\Sigma(x_i-\bar{x})^2>\chi_{n-1}^2(\alpha)k$: UMP, LR.
Hypothesis $[\sigma^2=k]$. $\Sigma(x_i-\bar{x})^2<Ak$ or $\Sigma(x_i-\bar{x})^2>Bk$: UMPU.
Hypothesis $[\sigma^2\geqslant k]$. $\Sigma(x_i-\bar{x})^2<\chi_{n-1}^2(1-\alpha)k$: UMPU, UMPI(L).
Hypothesis $\left[\dfrac{\mu}{\sigma}\leqslant k\right]$. $\dfrac{\bar{x}}{\sqrt{\Sigma(x_i-\bar{x})^2}}>E$: UMPI(S), AD.

(4) $Bin(N,\theta)$. Σx_i*. $Bin(Nn,\theta)$.
Point estimation of θ. $\dfrac{\bar{x}}{N}$: UMV, ML, AD.
Hypothesis $[\theta\leqslant k]$. $\bar{x}>A$: UMP.
Hypothesis $[h\leqslant\theta\leqslant l]$. $\bar{x}<B$ or $\bar{x}>C$: UMPU.

(5) $H(N,m,\theta)$ $(n=1)$. x*.
Point estimation of θ. $\dfrac{Nx}{m}$: UMV, AD.
Hypothesis $[\theta\leqslant k]$. $x>A$: UMP.

(6) $NB(N,\theta)$. $\Sigma x_i{}^*$. $NB(Nn,\theta)$.
Point estimation of θ. $\dfrac{Nn-1}{Nn+\Sigma x_i-1}$ (1 when the denominator is 0): UMV, AD.

$\dfrac{Nn}{Nn+\Sigma x_i}$: ML.

Hypothesis $[\theta \leqslant k]$. $\Sigma x_i < A$: UMP.
Hypothesis $[h \leqslant \theta \leqslant l]$. $\Sigma x_i < B$ or $\Sigma x_i > C$: UMPU.

(7) $P(\lambda)$. $\Sigma x_i{}^*$. $P(n\lambda)$.
Point estimation of λ. $\bar{x}$: UMV, ML, AD.
Hypothesis $[\lambda \leqslant k]$. $\bar{x} > A$: UMP.
Hypothesis $[h \leqslant \lambda \leqslant l]$. $\bar{x} < B$ or $\bar{x} > C$: UMPU.

(8) $G(\theta)$. $\Sigma x_i{}^*$. $NB(n,\theta)$.
For the point estimation of θ and hypothesis testing $\rightarrow$ (6).

(9) $U[0,\theta]$. $\max x_i{}^*$.
Point estimation of θ. $\max x_i$: ML, IAD. $\dfrac{n+1}{n}\max x_i$: UMV, IAD.
Hypothesis $[\theta \leqslant k]$. $\max x_i > (1-\alpha)^{1/n}k$: UMP.
Hypothesis $[\theta = k]$. $\max x_i < k\alpha^{1/n}$ or $\max x_i > k$: UMP.

(10) $U[\xi,\eta]$. $(\min x_i, \max x_i)^*$.
Point estimation of ξ. $\dfrac{n\min x_i - \max x_i}{n-1}$: UMV, IAD. $\min x_i$: ML, IAD.
Point estimation of $\dfrac{\xi+\eta}{2}$. $\dfrac{\min x_i + \max x_i}{2}$: UMV, AD.
Hypothesis $[\eta-\xi \leqslant k]$. $\max x_i - \min x_i > k\alpha^{1/n}$: UMP.

(11) $U\left[\theta-\frac{1}{2},\theta+\frac{1}{2}\right]$. $(\min x_i, \max x_i)^\#$.
Point estimation of θ. $\dfrac{\min x_i + \max x_i}{2}$: ML, AD.
Hypothesis $[\theta \leqslant k]$. $\min x_i > k+\frac{1}{2}-\alpha^{1/n}$ or $\max x_i > k+\frac{1}{2}$: UMP.

(12) $e(\mu,\sigma)$. $\begin{pmatrix}\Sigma x_i \\ \min x_i\end{pmatrix}^*$. $\begin{pmatrix}\Gamma(n,\sigma)+n\mu \\ e(\mu,\sigma/n)\end{pmatrix}$.

Point estimation of σ. $\dfrac{\Sigma x_i - n\min x_i}{n-1}$: UMV, IAD. $\bar{x}-\min x_i$: ML, IAD.
Point estimation of μ. $\dfrac{n}{n-1}\min x_i - \dfrac{1}{n-1}\bar{x}$: UMV, IAD. $\min x_i$: ML, IAD.
Hypothesis $[\sigma \leqslant k, \mu = h]$. $\Sigma x_i < h$ or $\Sigma x_i > k\log\alpha^{-1/n}+h$: UMP.
Hypothesis $[h \leqslant \sigma \leqslant l]$. $\Sigma x_i - n\min x_i < A$ or $\Sigma x_i - n\min x_i > B$: UMPU.
Hypothesis $[\mu = k]$. $\dfrac{n\min x_i - k}{\Sigma x_i - n\min x_i} < 0$ or $\dfrac{n\min x_i - k}{\Sigma x_i - n\min x_i} > C$: UMPU.

(13) $\Gamma(p,\sigma)$. $\Sigma x_i{}^*$. $\Gamma(np,\sigma)$.
Point estimation of σ. $\dfrac{\bar{x}}{p}$: UMV, ML, IAD.
Interval estimation of σ. $(C\Sigma x_i, D\Sigma x_i)$.
Hypothesis $[\sigma \leqslant k]$. $\Sigma x_i > A$: UMP.
Hypothesis $[\sigma = k]$. $\Sigma x_i < Ck$ or $\Sigma x_i > Dk$: UMPU.

(14) $\begin{matrix} N(\mu_1,a^2) \\ N(\mu_2,b^2). \end{matrix}$ $\begin{pmatrix}\Sigma x_i \\ \Sigma y_i\end{pmatrix}^*$. $\begin{pmatrix}N(n_1\mu_1,n_1a^2) \\ N(n_2\mu_2,n_2b^2)\end{pmatrix}$.

Point estimation of $\mu_1-\mu_2$. $\bar{x}-\bar{y}$: UMV, ML, AD.

Interval estimation of $\mu_1-\mu_2$. $\left(\bar{x}-\bar{y}\pm u(\alpha/2)\sqrt{\dfrac{a^2}{n_1}+\dfrac{b^2}{n_2}}\right)$.

Hypothesis $[\mu_1-\mu_2 \leqslant k]$. $\bar{x}-\bar{y} > k+u(\alpha)\sqrt{\dfrac{a^2}{n_1}+\dfrac{b^2}{n_2}}$: UMP, LR.

Hypothesis $[\mu_1-\mu_2=k]$. $|\bar{x}-\bar{y}-k|>u(\alpha/2)\sqrt{\frac{a^2}{n_1}+\frac{b^2}{n_2}}$: UMPU, UMPI(L), LR.

(15) $N(\mu_1,\sigma^2)$ $N(\mu_2,\sigma^2)$. $\begin{bmatrix}\Sigma x_i \\ \Sigma y_i \\ s^2=\Sigma(x_i-\bar{x})^2+\Sigma(y_i-\bar{y})^2\end{bmatrix}^* . \begin{bmatrix}N(n_1\mu_1,n_1\sigma^2) \\ N(n_2\mu_2,n_2\sigma^2) \\ \sigma^2\chi^2_{n_1+n_2-2}\end{bmatrix}$.

Point estimation of $\mu_1-\mu_2$. $\bar{x}-\bar{y}$: UMV, ML, AD.

Interval estimation of $\mu_1-\mu_2$. $\left(\bar{x}-\bar{y}\pm t_{n_1+n_2-2}(\alpha/2)\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\sqrt{\frac{s^2}{n_1+n_2-2}}\right)$.

Hypothesis $[\mu_1-\mu_2\leqslant k]$. $t=\frac{(\bar{x}-\bar{y}-k)\sqrt{n_1n_2}\sqrt{n_1+n_2-2}}{\sqrt{n_1+n_2}\sqrt{s^2}}>t_{n_1+n_2-2}(\alpha)$: UMPU, UMPI(L), LR.

Hypothesis $[\mu_1-\mu_2=k]$. $|t|>t_{n_1+n_2-2}(\alpha)$: UMPU, UMPI(L), LR.

Point estimation of σ^2. $\frac{s^2}{n_1+n_2-2}$: UMV, IAD. $\frac{s^2}{n_1+n_2}$: ML, IAD.

Interval estimation of σ^2. (As^2,Bs^2).

Hypothesis $[\sigma^2\leqslant k]$. $s^2>\chi^2_{n_1+n_2-2}(\alpha)k$: UMP, LR.

Hypothesis $[\sigma^2=k]$. $s^2<Ak$ or $s^2>Bk$: UMPU.

Hypothesis $[\sigma^2\geqslant k]$. $s^2>\chi^2_{n_1+n_2-2}(1-\alpha)k$: UMPU, UMPI(L), LR.

(16) $N(\mu_1,\sigma_1^2)$ $N(\mu_2,\sigma_2^2)$. $\begin{bmatrix}\Sigma x_i, & \Sigma(x_i-\bar{x})^2 \\ \Sigma y_i, & \Sigma(y_i-\bar{y})^2\end{bmatrix}^*$.

Interval estimation of $\frac{\sigma_1^2}{\sigma_2^2}$. $\left[A\frac{\Sigma(x_i-\bar{x})^2}{\Sigma(y_i-\bar{y})^2}, B\frac{\Sigma(x_i-\bar{x})^2}{\Sigma(y_i-\bar{y})^2}\right]$.

Hypothesis $\left[\frac{\sigma_1^2}{\sigma_2^2}\leqslant k\right]$. $\frac{(n_2-1)}{(n_1-1)}\frac{\Sigma(x_i-\bar{x})^2}{\Sigma(y_i-\bar{y})^2}>F^{n_1-1}_{n_2-1}(\alpha)k$: UMPU, UMPI(L, S), LR.

(17) $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$. $\begin{bmatrix}\Sigma x_i, & \Sigma(x_i-\bar{x})^2, & \Sigma(x_i-\bar{x})(y_i-\bar{y}) \\ \Sigma y_i, & \Sigma(y_i-\bar{y})^2, & \end{bmatrix}^*$.

Point estimation of ρ. $r=\frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\Sigma(x_i-\bar{x})^2\Sigma(y_i-\bar{y})^2}}$: ML.

Hypothesis $[\rho=0]$. $|r|>\frac{t_{n-1}(\alpha/2)}{\sqrt{t_{n-1}(\alpha/2)^2+n-2}}$: UMPU, LR.

Appendix B
Numerical Tables

1. Prime Numbers and Primitive Roots (→ 297 Number Theory, Elementary

In the following table, p is a prime number and r is a corresponding primitive root.

p	r	p	r	p	r	p	r	p	r	p	r	p	r	p	r
2		79	3	191	19	311	17	439	17	577	5	709	2	857	3
3	2	83	2	193	5	313	17	443	2	587	2	719	11	859	2
5	2	89	3	197	2	317	2	449	3	593	3	727	5	863	5
7	3	97	5	199	3	331	3	457	13	599	7	733	7	877	2
11	2	101	2	211	2	337	19	461	2	601	7	739	3	881	3
13	2	103	5	223	3	347	2	463	3	607	3	743	5	883	2
17	3	107	2	227	2	349	2	467	2	613	2	751	3	887	5
19	2	109	11	229	7	353	3	479	13	617	3	757	2	907	2
23	5	113	3	233	3	359	7	487	3	619	2	761	7	911	17
29	2	127	3	239	7	367	11	491	2	631	3	769	11	919	7
31	3	131	2	241	7	373	2	499	7	641	3	773	2	929	3
37	2	137	3	251	11	379	2	503	5	643	11	787	2	937	5
41	7	139	2	257	3	383	5	509	2	647	5	797	2	941	2
43	3	149	2	263	5	389	2	521	3	653	2	809	3	947	2
47	5	151	7	269	2	397	5	523	2	659	2	811	3	953	3
53	2	157	5	271	43	401	3	541	2	661	2	821	2	967	5
59	2	163	2	277	5	409	29	547	2	673	5	823	3	971	11
61	2	167	5	281	3	419	2	557	2	677	2	827	2	977	3
67	2	173	2	283	3	421	2	563	2	683	5	829	2	983	5
71	7	179	2	293	2	431	7	569	3	691	3	839	11	991	7
73	5	181	2	307	5	433	5	571	3	701	2	853	2	997	7

†Mersenne numbers. A prime number of the form 2^p-1 is called a Mersenne number. There exist 27 such p's less than 44500: $p=2$, 3, 5, 7, 13, 17, 19, 31, 61, 89, 107, 127, 521, 607, 1279, 2203, 2281, 3217, 4253, 4423, 9689, 9941, 11213, 19937, 21701, 23209, 44497. The even perfect numbers are the numbers of the form $2^{p-1}(2^p-1)$, where 2^p-1 is a Mersenne number.

2. Indices Modulo p (→ 297 Number Theory, Elementary)

Let r be a primitive root corresponding to a prime number p. The index $l=\mathrm{Ind}_r a$ of a with respect to the basis r is the integer l in $0 \leqslant l < p-1$ satisfying $r^l \equiv a (\mathrm{mod}\, p)$. $a \equiv b (\mathrm{mod}\, p)$ is equivalent to $\mathrm{Ind}_r a \equiv \mathrm{Ind}_r b (\mathrm{mod}(p-1))$. The index satisfies the following congruence relations with respect to $\mathrm{mod}(p-1)$: $\mathrm{Ind}_r ab \equiv \mathrm{Ind}_r a + \mathrm{Ind}_r b$, $\mathrm{Ind}_r a^n \equiv n \,\mathrm{Ind}_r a$, $\mathrm{Ind}_s a \equiv \mathrm{Ind}_s r \,\mathrm{Ind}_r a$.

We can solve congruence equations using these relations. The following is a table of indices.

p	$p-1$	r	2	3	5	7	11	13	17	19	23	29	31	37	41	43	47
2	1		—														
3	2	2	1	—													
5	4	2	1	3	—												
7	$2\cdot 3$	3	2	1	5	—											
11	$2\cdot 5$	2	1	8	4	7	—										
13	$2^2\cdot 3$	2	1	4	9	11	7	—									
17	2^4	3	14	1	5	11	7	4	—								
19	$2\cdot 3^2$	2	1	13	16	6	12	5	10	—							
23	$2\cdot 11$	5	2	16	1	19	9	14	7	15	—						
29	$2^2\cdot 7$	2	1	5	22	12	25	18	21	9	20	—					
31	$2\cdot 3\cdot 5$	3	24	1	20	28	23	11	7	4	27	9	—				
37	$2^2\cdot 3^2$	2	1	26	23	32	30	11	7	35	15	21	9	—			
41	$2^3\cdot 5$	7	14	25	18	1	37	9	7	31	4	33	12	8	—		
43	$2\cdot 3\cdot 7$	3	27	1	25	35	30	32	38	19	16	41	34	7	6	—	
47	$2\cdot 23$	5	18	20	1	32	7	11	16	45	5	35	3	42	15	13	—
53	$2^2\cdot 13$	2	1	17	47	14	6	24	10	37	39	46	33	30	45	22	44
59	$2\cdot 29$	2	1	50	6	18	25	45	40	38	15	28	49	55	14	33	23
61	$2^2\cdot 3\cdot 5$	2	1	6	22	49	15	40	47	26	57	35	59	39	54	43	20
67	$2\cdot 3\cdot 11$	2	1	39	15	23	59	19	64	10	28	44	47	22	53	9	50
71	$2\cdot 5\cdot 7$	7	6	26	28	1	31	39	49	16	15	68	11	20	25	48	9
73	$2^3\cdot 3^2$	5	8	6	1	33	55	59	21	62	46	35	11	64	4	51	31
79	$2\cdot 3\cdot 13$	3	4	1	62	53	68	34	21	32	26	11	56	19	75	49	59
83	$2\cdot 41$	2	1	72	27	8	24	77	56	47	60	12	38	20	40	71	23
89	$2^3\cdot 11$	3	16	1	70	81	84	23	6	35	57	59	31	11	21	29	54
97	$2^5\cdot 3$	5	34	70	1	31	86	25	89	81	77	13	46	91	85	4	84
101	$2^2\cdot 5^2$	2	1	69	24	9	13	66	30	96	86	91	84	56	45	42	58
103	$2\cdot 3\cdot 17$	5	44	39	1	4	61	72	70	80	24	86	57	93	50	77	85
107	$2\cdot 53$	2	1	70	47	43	22	14	29	78	62	32	27	38	40	59	66
109	$2^2\cdot 3^3$	11	15	80	92	20	1	101	87	105	3	98	34	43	63	42	103
113	$2^4\cdot 7$	3	12	1	83	8	74	22	5	99	41	89	50	67	94	47	31
127	$2\cdot 3^2\cdot 7$	3	72	1	87	115	68	94	38	84	121	113	46	98	80	71	60
131	$2\cdot 5\cdot 13$	2	1	72	46	96	56	18	43	35	23	51	29	41	126	124	105
137	$2^3\cdot 17$	3	10	1	75	42	122	25	38	46	125	91	73	102	119	97	19
139	$2\cdot 3\cdot 23$	2	1	41	86	50	76	64	107	61	27	94	56	80	32	115	98
149	$2^2\cdot 37$	2	1	87	104	142	109	53	124	84	95	120	132	72	41	93	138
151	$2\cdot 3\cdot 5^2$	7	10	93	136	1	82	23	124	120	145	42	34	148	3	74	128
157	$2^2\cdot 3\cdot 13$	5	141	82	1	147	28	26	40	124	135	129	62	116	21	113	92
163	$2\cdot 3^4$	2	1	101	15	73	47	51	57	125	9	107	69	33	160	38	28
167	$2\cdot 83$	5	40	94	1	118	28	103	53	58	99	150	90	61	97	87	132
173	$2^2\cdot 43$	2	1	27	39	95	23	130	73	33	20	144	102	162	138	84	64
179	$2\cdot 89$	2	1	108	138	171	15	114	166	54	135	118	62	149	155	80	36
181	$2^2\cdot 3^2\cdot 5$	2	1	56	156	15	62	164	175	135	53	48	99	26	83	20	13
191	$2\cdot 5\cdot 19$	19	44	116	50	171	85	112	98	1	134	33	175	15	165	8	123
193	$2^6\cdot 3$	5	34	84	1	104	183	141	31	145	162	123	82	5	151	24	29
197	$2^2\cdot 7^2$	2	1	181	89	146	29	25	159	154	120	36	141	192	110	78	66
199	$2\cdot 3^2\cdot 11$	3	106	1	138	142	189	172	123	55	118	70	164	11	167	88	76
211	$2\cdot 3\cdot 5\cdot 7$	2	1	43	132	139	162	144	199	154	21	179	115	118	17	80	124
223	$2\cdot 3\cdot 37$	3	180	1	89	210	107	147	144	172	163	128	82	152	204	118	50
227	$2\cdot 113$	2	1	46	11	154	28	61	99	178	34	8	197	77	131	150	218
229	$2^2\cdot 3\cdot 19$	7	111	68	214	1	42	195	24	52	131	191	175	164	73	12	193

(table continued on following page)

p	$p-1$	r	2	3	5	7	11	13	17	19	23	29	31	37	41	43	47
233	$2^3\cdot 29$	3	72	1	165	222	197	158	103	136	112	132	182	8	85	25	139
239	$2\cdot 7\cdot 17$	7	66	74	138	1	4	43	52	155	63	160	188	31	99	15	113
241	$2^4\cdot 3\cdot 5$	7	190	182	138	1	25	47	111	85	57	154	151	73	6	219	114
251	$2\cdot 5^3$	11	135	6	80	218	1	162	184	233	134	203	226	187	64	77	85
257	2^8	3	48	1	55	85	196	106	120	125	28	94	242	219	19	207	61
263	$2\cdot 131$	5	190	50	1	79	166	62	126	43	156	221	136	170	17	154	65
269	$2^2\cdot 67$	2	1	109	208	19	230	142	105	223	176	187	259	56	200	254	32
271	$2\cdot 3^3\cdot 5$	43	266	153	220	98	92	15	16	261	75	45	222	182	156	1	213
277	$2^2\cdot 3\cdot 23$	5	147	188	1	22	7	222	103	252	208	74	47	87	126	55	218
281	$2^3\cdot 5\cdot 7$	3	204	1	186	182	253	9	166	221	197	172	62	135	23	132	75

3. Bernoulli Numbers and Euler Numbers (→ 177 Generating Functions)

B_n are Bernoulli numbers; E_n are Euler numbers.

n	Numerator of B_n	Denominator of B_n	B_n	E_n
2	1	6	0.16667	1
4	1	30	0.03333	5
6	1	42	0.02381	61
8	1	30	0.03333	1385
10	5	66	0.07576	50521
12	691	2730	0.25311	2702765
14	7	6	1.16667	199360981
16	3617	510	7.09216	19391512145
18	43867	798	54.97118	2404879675441
20	174611	330	529.12424	370371188237525
22	854513	138	6192.12319	6.934887×10^{16}
24	236364091	2730	86580.25311	1.551453×10^{19}
26	8553103	6	1425517.16667	4.087073×10^{21}
28	23749461029	870	27298231.06782	1.252260×10^{24}
30	8615841276005	14322	601580873.90064	4.415439×10^{26}

4. Class Numbers of Algebraic Number Fields

(I) Class Numbers of Real Quadratic Field (→ 347 Quadratic Fields)

Let $k=\mathbf{Q}(\sqrt{m})$, where m is a positive integer without square factor ($1<m\leqslant 501$). h is the class number (in the wider sense) of k. The $-$sign in the row of $N(\varepsilon)$ means that the norm $N(\varepsilon)$ of the fundamental unit is -1. When $N(\varepsilon)=+1$, the class number in the narrow sense is $2h$, and when $N(\varepsilon)=-1$, the class number in the narrow sense is also h.

m	h	$N(\varepsilon)$	m	h	$N(\varepsilon)$	m	h	$N(\varepsilon)$	m	h	$N(\varepsilon)$	m	h	$N(\varepsilon)$	m	h	$N(\varepsilon)$
2	1	–	85	2	–	170	4	–	253	1		335	2		421	1	–
3	1		86	1		173	1	–	254	3		337	1	–	422	1	
5	1	–	87	2		174	2		255	4		339	2		426	2	
6	1		89	1	–	177	1		257	3	–	341	1		427	6	
7	1		91	2		178	2		258	2		345	2		429	2	
10	2	–	93	1		179	1		259	2		346	6	–	430	2	
11	1		94	1		181	1	–	262	1		347	1		431	1	
13	1	–	95	2		182	2		263	1		349	1	–	433	1	–
14	1		97	1	–	183	2		265	2	–	353	1	–	434	4	

m	h	$N(\varepsilon)$	m	h	$N(\varepsilon)$	m	h	$N(\varepsilon)$	m	h	$N(\varepsilon)$	m	h	$N(\varepsilon)$	m	h	$N(\varepsilon)$
15	2		101	1	–	185	2	–	266	2		354	2		435	4	
17	1	–	102	2		186	2		267	2		355	2		437	1	
19	1		103	1		187	2		269	1	–	357	2		438	4	
21	1		105	2		190	2		271	1		358	1		439	5	
22	1		106	2	–	191	1		273	2		359	3		442	8	–
23	1		107	1		193	1	–	274	4	–	362	2	–	443	3	
26	2	–	109	1	–	194	2		277	1	–	365	2	–	445	4	–
29	1	–	110	2		195	4		278	1		366	2		446	1	
30	2		111	2		197	1	–	281	1	–	367	1		447	2	
31	1		113	1	–	199	1		282	2		370	4	–	449	1	–
33	1		114	2		201	1		283	1		371	2		451	2	
34	2		115	2		202	2	–	285	2		373	1	–	453	1	
35	2		118	1		203	2		286	2		374	2		454	1	
37	1	–	119	2		205	2		287	2		377	2		455	4	
38	1		122	2	–	206	1		290	4	–	379	1		457	1	–
39	2		123	2		209	1		291	4		381	1		458	2	–
41	1	–	127	1		210	4		293	1	–	382	1		461	1	–
42	2		129	1		211	1		295	2		383	1		462	4	
43	1		130	4	–	213	1		298	2	–	385	2		463	1	
46	1		131	1		214	1		299	2		386	2		465	2	
47	1		133	1		215	2		301	1		389	1	–	466	2	
51	2		134	1		217	1		302	1		390	4		467	1	
53	1	–	137	1	–	218	2	–	303	2		391	2		469	3	
55	2		138	2		219	4		305	2		393	1		470	2	
57	1		139	1		221	2		307	1		394	2	–	471	2	
58	2	–	141	1		222	2		309	1		395	2		473	3	
59	1		142	3		223	3		310	2		397	1	–	474	2	
61	1	–	143	2		226	8	–	311	1		398	1		478	1	
62	1		145	4	–	227	1		313	1	–	399	8		479	1	
65	2	–	146	2		229	3	–	314	2	–	401	5	–	481	2	–
66	2		149	1	–	230	2		317	1	–	402	2		482	2	
67	1		151	1		231	4		318	2		403	2		483	4	
69	1		154	2		233	1	–	319	2		406	2		485	2	–
70	2		155	2		235	6		321	3		407	2		487	1	
71	1		157	1	–	237	1		322	4		409	1	–	489	1	
73	1	–	158	1		238	2		323	4		410	4		491	1	
74	2	–	159	2		239	1		326	3		411	2		493	2	–
77	1		161	1		241	1	–	327	2		413	1		494	2	
78	2		163	1		246	2		329	1		415	2		497	1	
79	3		165	2		247	2		330	4		417	1		498	2	
82	4	–	166	1		249	1		331	1		418	2		499	5	
83	1		167	1		251	1		334	1		419	1		501	1	

One can find a table of fundamental units and representatives of ideal classes for $0<m<2025$ in E. L. Ince, *Cycles of reduced ideals in quadratic fields*, Royal Society, London, 1968.

(II) Class Numbers of Imaginary Quadratic Fields (→ 347 Quadratic Fields)

Let $k=\mathbf{Q}(\sqrt{-m})$, where m is a positive integer without square factor ($1 \leqslant m \leqslant 509$). h is the class number of k. In the present case, there is no distinction between the class numbers in the wider and narrow senses.

m	h	m	h	m	h	m	h	m	h	m	h	m	h	m	h
1	1	65	8	129	12	193	4	255	12	319	10	389	22	447	14
2	1	66	8	130	4	194	20	257	16	321	20	390	16	449	20
3	1	67	1	131	5	195	4	258	8	322	8	391	14	451	6
5	2	69	8	133	4	197	10	259	4	323	4	393	12	453	12
6	2	70	4	134	14	199	9	262	6	326	22	394	10	454	14
7	1	71	7	137	8	201	12	263	13	327	12	395	8	455	20
10	2	73	4	138	8	202	6	265	8	329	24	397	6	457	8
11	1	74	10	139	3	203	4	266	20	330	8	398	20	458	26

m	h	m	h	m	h	m	h	m	h	m	h	m	h	m	h
13	2	77	8	141	8	205	8	267	2	331	3	399	16	461	30
14	4	78	4	142	4	206	20	269	22	334	12	401	20	462	8
15	2	79	5	143	10	209	20	271	11	335	18	402	16	463	7
17	4	82	4	145	8	210	8	273	8	337	8	403	2	465	16
19	1	83	3	146	16	211	3	274	12	339	6	406	16	466	8
21	4	85	4	149	14	213	8	277	6	341	28	407	16	467	7
22	2	86	10	151	7	214	6	278	14	345	8	409	16	469	16
23	3	87	6	154	8	215	14	281	20	346	10	410	16	470	20
26	6	89	12	155	4	217	8	282	8	347	5	411	6	471	16
29	6	91	2	157	6	218	10	283	3	349	14	413	20	473	12
30	4	93	4	158	8	219	4	285	16	353	16	415	10	474	20
31	3	94	8	159	10	221	16	286	12	354	16	417	12	478	8
33	4	95	8	161	16	222	12	287	14	355	4	418	8	479	25
34	4	97	4	163	1	223	7	290	20	357	8	419	9	481	16
35	2	101	14	165	8	226	8	291	4	358	6	421	10	482	20
37	2	102	4	166	10	227	5	293	18	359	19	422	10	483	4
38	6	103	5	167	11	229	10	295	8	362	18	426	24	485	20
39	4	105	8	170	12	230	20	298	6	365	20	427	2	487	7
41	8	106	6	173	14	231	12	299	8	366	12	429	16	489	20
42	4	107	3	174	12	233	12	301	8	367	9	430	12	491	9
43	1	109	6	177	4	235	2	302	12	370	12	431	21	493	12
46	4	110	12	178	8	237	12	303	10	371	8	433	12	494	28
47	5	111	8	179	5	238	8	305	16	373	10	434	24	497	24
51	2	113	8	181	10	239	15	307	3	374	28	435	4	498	8
53	6	114	8	182	12	241	12	309	12	377	16	437	20	499	3
55	4	115	2	183	8	246	12	310	8	379	3	438	8	501	16
57	4	118	6	185	16	247	6	311	19	381	20	439	15	502	14
58	2	119	10	186	12	249	12	313	8	382	8	442	8	503	21
59	3	122	10	187	2	251	7	314	26	383	17	443	5	505	8
61	6	123	2	190	4	253	4	317	10	385	8	445	8	506	28
62	8	127	5	191	13	254	16	318	12	386	20	446	32	509	30

There are only 9 instances of m for which $h=1$, and only 18 instances of m for which $h=2$ (Baker, Stark). All these cases are in this table.

One can find a table of structures of the ideal class groups and representatives of ideal classes for $m<24000$ in H. Wada, *A table of ideal class groups of imaginary quadratic fields*, Proc. Japan Acad., 46 (1970), 401–403.

(III) Class Numbers of Cyclotomic Fields

Cyclotomic field $k=\mathbf{Q}(e^{2\pi i/l})$ $(1<l<100;\ l$ prime). h_1 is the first factor of the class number of k (→ 14 Algebraic Number Fields).

l	h_1	l	h_1	l	h_1	l	h_1	l	h_1	l	h_1
3	1	13	1	29	2^3	43	211	61	$41\cdot1861$	79	$5\cdot53\cdot377911$
5	1	17	1	31	3^2	47	$5\cdot139$	67	$67\cdot12739$	83	$3\cdot279405653$
7	1	19	1	37	37	53	4889	71	$7^2\cdot79241$	89	$113\cdot118401449$
11	1	23	3	41	11^2	59	$3\cdot59\cdot233$	73	$89\cdot134353$	97	$577\cdot3457\cdot206209$

$h_1>1$ for $l>19$ (Uchida).

5. Characters of Finite Groups; Crystallographic Groups

(I) Symmetric Groups S_n, Alternating Groups A_n $(3\leqslant n\leqslant 7)$, and Mathieu Groups M_n $(n=11,12,22,23,24)$

(1) In each table, the first column gives the representation of the conjugate class as we represent a permutation by the product of cyclic permutations. For example, $(3)(2)^2$ means the conjugate class containing (123)(45)(67).

(2) The second column gives the order of the centralizer of the elements of the conjugate class.
(3) In the table of S_n, the first row gives the type of Young diagram corresponding to each irreducible character. For example, $[3,2^2,1]$ means $T(3,2,2,1)$.
(4) In the table of A_n, when we restrict the self-conjugate character of S_n (the character with *) to A_n, it is decomposed into two mutually algebraically conjugate irreducible characters, and therefore we show only one of them. The other irreducible character of A_n is given by the restriction to A_n of the character of S_n that is not self-conjugate.
(5) In the table of M_n, each character with a bar over the degree is one of the two mutually algebraically conjugate characters.

$\varepsilon_1^{\pm}=(-1\pm\sqrt{-3})/2,\quad \varepsilon_2^{\pm}=(1\pm\sqrt{5})/2,\quad \varepsilon_3^{\pm}=(-1\pm\sqrt{-7})/2$

$\varepsilon_4^{\pm}=(-1\pm\sqrt{-11})/2,\quad \varepsilon_5^{\pm}=(-1\pm\sqrt{-15})/2,\quad \varepsilon_6^{\pm}=(-1\pm\sqrt{-23})/2$

S_3		[3]	[2, 1]*	$[1^3]$
(1)	6	1	2	1
(2)	2	1	0	−1
(3)	3	1	−1	1

A_3		[2, 1]*
(1)	3	1
(3)	3	ε_1^+
(3)	3	ε_1^-

S_4		[4]	[3, 1]	$[2^2]$*	$[2, 1^2]$	$[1^4]$
(1)	24	1	3	2	3	1
(2)	4	1	1	0	−1	−1
(3)	3	1	0	−1	0	1
(4)	4	1	−1	0	1	−1
$(2)^2$	8	1	−1	2	−1	1

A_4		$[2^2]$*
(1)	12	1
(3)	3	ε_1^+
(3)	3	ε_1^-
$(2)^2$	4	1

S_5		[5]	[4, 1]	[3, 2]	$[3, 1^2]$*	$[2^2, 1]$	$[2, 1^3]$	$[1^5]$
(1)	120	1	4	5	6	5	4	1
(2)	12	1	2	1	0	−1	−2	−1
(3)	6	1	1	−1	0	−1	1	1
(4)	4	1	0	−1	0	1	0	−1
$(2)^2$	8	1	0	1	−2	1	0	1
(3)(2)	6	1	−1	1	0	−1	1	−1
(5)	5	1	−1	0	1	0	−1	1

A_5		$[3, 1^2]$*
(1)	60	3
(3)	3	0
$(2)^2$	4	−1
(5)	5	ε_2^+
(5)	5	ε_2^-

S_6		[6]	[5, 1]	[4, 2]	$[4, 1^2]$	$[3^2]$	[3, 2, 1]*	$[2^3]$	$[3, 1^3]$	$[2^2, 1^2]$	$[2, 1^4]$	$[1^6]$
(1)	720	1	5	9	10	5	16	5	10	9	5	1
(2)	48	1	3	3	2	1	0	−1	−2	−3	−3	−1
(3)	18	1	2	0	1	−1	−2	−1	1	0	2	1
(4)	8	1	1	−1	0	−1	0	1	0	1	−1	−1
$(2)^2$	16	1	1	1	−2	1	0	1	−2	1	1	1
(3)(2)	6	1	0	0	−1	1	0	−1	1	0	0	−1
(5)	5	1	0	−1	0	0	1	0	0	−1	0	1
(6)	6	1	−1	0	1	0	0	0	−1	0	1	−1
(4)(2)	8	1	−1	1	0	−1	0	−1	0	1	−1	1
$(2)^3$	48	1	−1	3	−2	−3	0	3	2	−3	1	−1
$(3)^2$	18	1	−1	0	1	2	−2	2	1	0	−1	1

A_6		$[3,2,1]^*$
(1)	360	8
(3)	9	−1
$(2)^2$	8	0
(5)	5	ε_2^+
(5)	5	ε_2^-
(4)(2)	4	0
$(3)^2$	9	−1

S_7		[7]	[6,1]	[5,2]	$[5,1^2]$	[4,3]	[4,2,1]	$[3^2,1]$	$[4,1^3]^*$	$[3,2^2]$	$[3,2,1^2]$	$[2^3,1]$	$[3,1^4]$	$[2^2,1^3]$	$[2,1^5]$	$[1^7]$
(1)	5040	1	6	14	15	14	35	21	20	21	35	14	15	14	6	1
(2)	240	1	4	6	5	4	5	1	0	−1	−5	−4	−5	−6	−4	−1
(3)	72	1	3	2	3	−1	−1	−3	2	−3	−1	−1	3	2	3	1
(4)	24	1	2	0	1	−2	−1	−1	0	1	1	2	−1	0	−2	−1
$(2)^2$	48	1	2	2	−1	2	−1	1	−4	1	−1	2	−1	2	2	1
(3)(2)	12	1	1	0	−1	1	−1	1	0	−1	1	−1	1	0	−1	−1
(5)	10	1	1	−1	0	−1	0	1	0	1	0	−1	0	−1	1	1
(6)	6	1	0	−1	0	0	1	0	0	0	−1	0	0	1	0	−1
(4)(2)	8	1	0	0	−1	0	1	−1	0	−1	1	0	−1	0	0	1
$(2)^3$	48	1	0	2	−3	0	1	−3	0	3	−1	0	3	−2	0	−1
$(3)^2$	18	1	0	−1	0	2	−1	0	2	0	−1	2	0	−1	0	1
(5)(2)	10	1	−1	1	0	−1	0	1	0	−1	0	1	0	−1	1	−1
$(3)(2)^2$	24	1	−1	2	−1	−1	−1	1	2	1	−1	−1	−1	2	−1	1
(4)(3)	12	1	−1	0	1	1	−1	−1	0	1	1	−1	−1	0	1	−1
(7)	7	1	−1	0	1	0	0	0	−1	0	0	0	1	0	−1	1

A_7		$[4,1^3]^*$
(1)	2520	10
(3)	36	1
$(2)^2$	24	−2
(5)	5	0
(4)(2)	4	0
$(3)^2$	9	1
$(3)(2)^2$	12	1
(7)	7	ε_3^+
(7)	7	ε_3^-

M_{11}									
(1)	g	1	10	11	55	45	44	$\overline{16}$	$\overline{10}$
$(2)^4$	48	1	2	3	−1	−3	4	0	−2
$(4)^2$	8	1	2	−1	−1	1	0	0	0
$(3)^3$	18	1	1	2	1	0	−1	−2	1
$(5)^2$	5	1	0	1	0	0	−1	1	0
(8)(2)	8	1	0	−1	1	−1	0	0	$\pm i\sqrt{2}$
(8)(2)	8	1	0	−1	1	−1	0	0	$\mp i\sqrt{2}$
(6)(3)(2)	6	1	−1	0	−1	0	1	0	1
(11)	11	1	−1	0	0	1	0	ε_4^+	−1
(11)	11	1	−1	0	0	1	0	ε_4^-	−1

$g = 11 \cdot 10 \cdot 9 \cdot 8 = 7920.$

M_{12}															
(1)	g	1	11	11	55	55	55	45	54	66	99	120	144	176	$\overline{16}$
$(2)^4$	192	1	3	3	−1	−1	7	−3	6	2	3	−8	0	0	0
$(4)^2$	32	1	3	−1	3	−1	−1	1	2	−2	−1	0	0	0	0
$(3)^3$	54	1	2	2	1	1	1	0	0	3	0	3	0	−4	−2
$(5)^2$	10	1	1	1	0	0	0	0	−1	1	−1	0	−1	1	1
(8)(2)	8	1	1	−1	−1	1	−1	−1	0	0	1	0	0	0	0
(6)(3)(2)	6	1	0	0	−1	−1	1	0	0	−1	0	1	0	0	0
(11)	11	1	0	0	0	0	0	1	−1	0	0	−1	1	0	ε_4^+
(11)	11	1	0	0	0	0	0	1	−1	0	0	−1	1	0	ε_4^-
$(2)^6$	240	1	−1	−1	−5	−5	−5	5	6	6	−1	0	4	−4	4
(10)(2)	10	1	−1	−1	0	0	0	0	1	1	−1	0	−1	1	−1
$(4)^2(2)^2$	32	1	−1	3	−1	3	−1	1	2	−2	−1	0	0	0	0
$(3)^4$	36	1	−1	−1	1	1	1	3	0	0	3	0	−3	−1	1
$(6)^2$	12	1	−1	−1	1	1	1	−1	0	0	−1	0	1	−1	1
(8)(4)	8	1	−1	1	1	−1	−1	−1	0	0	1	0	0	0	0

$g = 12 \cdot 11 \cdot 10 \cdot 9 \cdot 8 = 95040.$

M_{22}

(1)	g	1	21	55	154	210	$\overline{280}$	231	385	99	$\overline{45}$
$(2)^8$	384	1	5	7	10	2	−8	7	1	3	−3
$(3)^6$	36	1	3	1	1	3	1	−3	−2	0	0
$(5)^4$	5	1	1	0	−1	0	0	1	0	−1	0
$(4)^4(2)^2$	16	1	1	−1	2	−2	0	−1	1	−1	1
$(4)^4(2)^2$	32	1	1	3	−2	−2	0	−1	1	3	1
$(7)^3$	7	1	0	−1	0	0	0	0	0	1	ε_3^+
$(7)^3$	7	1	0	−1	0	0	0	0	0	1	ε_3^-
$(8)^2(4)(2)$	8	1	−1	1	0	0	0	−1	1	−1	−1
$(6)^2(3)^2(2)^2$	12	1	−1	1	1	−1	1	1	−2	0	0
$(11)^2$	11	1	−1	0	0	1	ε_4^+	0	0	0	1
$(11)^2$	11	1	−1	0	0	1	ε_4^-	0	0	0	1

$g = 22 \cdot 21 \cdot 20 \cdot 48 = 443520.$

M_{23}

(1)	g	1	22	230	231	$\overline{770}$	1035	2024	$\overline{45}$	$\overline{990}$	$\overline{231}$	253	$\overline{896}$
$(2)^8$	2688	1	6	22	7	−14	27	8	−3	−18	7	13	0
$(3)^6$	180	1	4	5	6	5	0	−1	0	0	−3	1	−4
$(5)^4$	15	1	2	0	1	0	0	−1	0	0	1	−2	1
$(4)^4(2)^2$	32	1	2	2	−1	−2	−1	0	1	2	−1	1	0
$(7)^3$	14	1	1	−1	0	0	−1	1	ε_3^+	ε_3^+	0	1	0
$(7)^3$	14	1	1	−1	0	0	−1	1	ε_3^-	ε_3^-	0	1	0
$(8)^2(4)(2)$	8	1	0	0	−1	0	1	0	−1	0	−1	−1	0
$(6)^2(3)^2(2)^2$	12	1	0	1	−2	1	0	−1	0	0	1	1	0
$(11)^2$	11	1	0	−1	0	0	1	0	1	0	0	0	ε_4^+
$(11)^2$	11	1	0	−1	0	0	1	0	1	0	0	0	ε_4^-
$(15)(5)(3)$	15	1	−1	0	1	0	0	−1	0	0	ε_5^+	1	1
$(15)(5)(3)$	15	1	−1	0	1	0	0	−1	0	0	ε_5^-	1	1
$(14)(7)(2)$	14	1	−1	1	0	0	−1	1	$-\varepsilon_3^+$	ε_3^+	0	−1	0
$(14)(7)(2)$	14	1	−1	1	0	0	−1	1	$-\varepsilon_3^-$	ε_3^-	0	−1	0
(23)	23	1	−1	0	1	ε_6^+	0	0	−1	1	1	0	−1
(23)	23	1	−1	0	1	ε_6^-	0	0	−1	1	1	0	−1

$g = 23 \cdot 22 \cdot 21 \cdot 20 \cdot 48 = 10200960.$

M_{24}

$(1)^{24}$	g	1	23	$7\cdot36$	$23\cdot11$	$23\cdot77$	$55\cdot64$	$\overline{45}$	$\overline{22\cdot45}$	$\overline{23\cdot45}$	$23\cdot45$	$\overline{11\cdot21}$	$\overline{770}$
$(2)^8$	$21\cdot2^{10}$	1	7	28	13	−21	64	−3	−18	−21	27	7	−14
$(3)^6$	$27\cdot40$	1	5	9	10	16	10	0	0	0	0	−3	5
$(5)^4$	60	1	3	2	3	1	0	0	0	0	0	1	0
$(4)^4(2)^2$	128	1	3	4	1	−5	0	1	2	3	−1	−1	−2
$(7)^3$	42	1	2	0	1	0	−1	ε_3^+	ε_3^+	$2\varepsilon_3^+$	−1	0	0
$(7)^3$	42	1	2	0	1	0	−1	ε_3^-	ε_3^-	$2\varepsilon_3^-$	−1	0	0
$(8)^2(4)(2)$	16	1	1	0	−1	−1	0	−1	0	−1	1	−1	0
$(6)^2(3)^2(2)^2$	24	1	1	1	−2	0	−2	0	0	0	0	1	1
$(11)^2$	11	1	1	−1	0	0	0	1	0	1	1	0	0
$(15)(5)(3)$	15	1	0	−1	0	1	0	0	0	0	0	ε_5^+	0
$(15)(5)(3)$	15	1	0	−1	0	1	0	0	0	0	0	ε_5^-	0
$(14)(7)(2)$	14	1	0	0	−1	0	1	$-\varepsilon_3^-$	ε_3^+	0	−1	0	0
$(14)(7)(2)$	14	1	0	0	−1	0	1	$-\varepsilon_3^+$	ε_3^-	0	−1	0	0
(23)	23	1	0	−1	0	0	1	−1	1	0	0	1	ε_6^+
(23)	23	1	0	−1	0	0	1	−1	1	0	0	1	ε_6^-
$(12)^2$	12	1	−1	0	1	−1	0	1	1	−1	0	0	1
$(6)^4$	24	1	−1	0	1	−1	0	−1	−1	1	2	0	1
$(4)^6$	96	1	−1	0	1	−1	0	1	−2	−1	3	3	−2
$(3)^8$	$7\cdot72$	1	−1	0	1	7	−8	3	3	−3	6	0	−7
$(2)^{12}$	$15\cdot2^9$	1	−1	12	−11	11	0	5	−10	−5	35	−9	10
$(10)^2(2)^2$	20	1	−1	2	−1	1	0	0	0	0	0	1	0
$(21)(3)$	21	1	−1	0	1	0	−1	ε_3^-	ε_3^-	$-\varepsilon_3^+$	−1	0	0
$(21)(3)$	21	1	−1	0	1	0	−1	ε_3^+	ε_3^+	$-\varepsilon_3^-$	−1	0	0
$(4)^4(2)^4$	$3\cdot2^7$	1	−1	4	−3	3	0	−3	6	3	3	−1	2
$(12)(6)(4)(2)$	12	1	−1	1	0	0	0	0	0	0	0	−1	−1

$(1)^{24}$	g	$23\cdot 21$	$23\cdot 55$	$23\cdot 88$	$23\cdot 99$	$23\cdot 144$	$23\cdot 11\cdot 21$	$23\cdot 7\cdot 36$	$77\cdot 72$	$11\cdot 35\cdot 27$
$(2)^8$	$21\cdot 2^{10}$	35	49	8	21	48	49	−28	−56	−21
$(3)^6$	$27\cdot 40$	6	5	−1	0	0	−15	−9	9	0
$(5)^4$	60	−2	0	−1	−3	−3	3	1	−1	0
$(4)^4(2)^2$	128	3	1	0	1	0	−3	4	0	−1
$(7)^3$	42	0	−2	1	2	1	0	0	0	0
$(7)^3$	42	0	−2	1	2	1	0	0	0	0
$(8)^2(4)(2)$	16	−1	1	0	−1	0	−1	0	0	1
$(6)^2(3)^2(2)^2$	24	2	1	−1	0	0	1	−1	1	0
$(11)^2$	11	−1	0	0	0	1	0	−1	0	0
$(15)(5)(3)$	15	1	0	−1	0	0	0	1	−1	0
$(15)(5)(3)$	15	1	0	−1	0	0	0	1	−1	0
$(14)(7)(2)$	14	0	0	1	0	−1	0	0	0	0
$(14)(7)(2)$	14	0	0	1	0	−1	0	0	0	0
(23)	23	0	0	0	0	0	0	0	1	−1
(23)	23	0	0	0	0	0	0	0	1	−1
$(12)^2$	12	0	0	0	0	0	0	0	0	0
$(6)^4$	24	0	0	0	2	−2	0	0	0	0
$(4)^6$	96	3	−3	0	−3	0	−3	0	0	3
$(3)^8$	$7\cdot 72$	0	8	8	6	−6	0	0	0	0
$(2)^{12}$	$15\cdot 2^9$	3	−15	24	−19	16	9	36	24	−45
$(10)^2(2)^2$	20	−2	0	−1	1	1	−1	1	−1	0
$(21)(3)$	21	0	1	1	−1	1	0	0	0	0
$(21)(3)$	21	0	1	1	−1	1	0	0	0	0
$(4)^4(2)^4$	$3\cdot 2^7$	3	−7	8	−3	0	1	−4	−8	3
$(12)(6)(4)(2)$	12	0	−1	−1	0	0	1	−1	1	0

$g=24\cdot 23\cdot 22\cdot 21\cdot 20\cdot 48=244823040.$

(II) General Linear Groups $GL(2,q)$, Unitary Groups $U(2,q)$, and Special Linear Groups $SL(2,q)$ (q is a power of a prime) (→ 151 Finite Groups I)

(1) The notations are as follows. $\varepsilon=\exp[2\pi\sqrt{-1}\,/(q-1)]$, $\eta=\exp[2\pi\sqrt{-1}\,/(q^2-1)]$, $\sigma=\exp[2\pi\sqrt{-1}\,/(q+1)]$, ρ is the generator of the multiplicative group of $GF(q)-\{0\}$, ω is the generator of the multiplicative group of $GF(q^2)-\{0\}$, $\omega^{q-1}=\alpha$, B is an element of $GL(2,q)$ with order q^2-1, and $B_1=B^{q-1}$.
(2) The first column gives a representative of the conjugate class.

General Linear Group $GL(2,q)$.

	$X_n(1)$	$X_n(q)$	$Y_{m,n}$	Z_n
$\begin{pmatrix}\rho^a & \\ & \rho^a\end{pmatrix}$	ε^{2na}	$q\varepsilon^{2na}$	$(q+1)\varepsilon^{(m+n)a}$	$(q-1)\eta^{na(q+1)}$
$\begin{pmatrix}\rho^a & \\ 1 & \rho^a\end{pmatrix}$	ε^{2na}	0	$\varepsilon^{(m+n)a}$	$-\eta^{na(q+1)}$
$\begin{pmatrix}\rho^a & \\ & \rho^b\end{pmatrix}$	$\varepsilon^{n(a+b)}$	$\varepsilon^{n(a+b)}$	$\varepsilon^{ma+nb}+\varepsilon^{mb+na}$	0
B^c	ε^{nc}	$-\varepsilon^{nc}$	0	$-(\eta^{nc}+\eta^{ncq})$

(1) $1\leqslant a\leqslant q-1$, $1\leqslant b\leqslant q-1$, $a\not\equiv b \pmod{q-1}$, $1\leqslant c<q^2-1$, $c\not\equiv 0 \pmod{q+1}$.
(2) We assume that $1\leqslant n\leqslant q-1$, for $X_n(1), X_n(q)$, $1\leqslant m<n\leqslant q-1$, for $Y_{m,n}$, $1\leqslant n<q^2-1$ for Z_n, $n\not\equiv 0 (\mathrm{mod}\, q+1)$. Here, $Z_n=Z_{n'}$ when $n\equiv n'q (\mathrm{mod}\, q^2-1)$.

Unitary Group $U(2,q)$.

	$X'_n(1)$	$X'_n(q)$	$Y'_{m,n}$	Z'_n
$\begin{pmatrix}\alpha^s & \\ & \alpha^s\end{pmatrix}$	σ^{2ns}	$q\sigma^{2ns}$	$(q-1)\sigma^{(m+n)s}$	$(q+1)\sigma^{ns}$
$\begin{pmatrix}\alpha^s & \\ 1 & \alpha^s\end{pmatrix}$	σ^{2ns}	0	$-\sigma^{(m+n)s}$	σ^{ns}
$\begin{pmatrix}\alpha^s & \\ & \alpha^t\end{pmatrix}$	$\sigma^{n(s+t)}$	$-\sigma^{n(s+t)}$	$-(\sigma^{ms+nt}+\sigma^{mt+ns})$	0
$\begin{pmatrix}\omega^u & \\ & \omega^{-uq}\end{pmatrix}$	σ^{-nu}	σ^{-nu}	0	$\eta^{nu}+\eta^{-nuq}$

(1) $\begin{pmatrix}\alpha^s & \\ 1 & \alpha^s\end{pmatrix}$, $\begin{pmatrix}\omega^u & \\ & \omega^{-uq}\end{pmatrix}$ are the canonical forms of an element of $U(2,q)$ in $GL(2,q^2)$.
(2) $1 \leqslant s \leqslant q+1$, $1 \leqslant t \leqslant q+1$, $s \not\equiv t \pmod{q+1}$, $1 \leqslant u < q^2-1$, $u \not\equiv 0 \pmod{q-1}$. When $u \equiv -u'q \pmod{q^2-1}$ u, u' gives the same conjugate class.
(3) The ranges are $1 \leqslant n \leqslant q+1$ for $X_n'(1)$, $X_n'(q)$, $1 \leqslant m < n \leqslant q+1$ for $Y_{m,n}'$, $1 \leqslant n < q^2-1$ for Z_n', $n \not\equiv 0 \pmod{q-1}$. When $n' \equiv -nq \pmod{q^2-1}$, we have $Z_n' = Z_{n'}'$.

Special Linear Group $SL(2,2^n)$ (the case when $q=2^n$).

			Y_n	Z_m
$\begin{pmatrix}1 & \\ & 1\end{pmatrix}$	1	q	$q+1$	$q-1$
$\begin{pmatrix}1 & \\ 1 & 1\end{pmatrix}$	1	0	1	-1
$\begin{pmatrix}\rho^a & \\ & \rho^{-a}\end{pmatrix}$	1	1	$\varepsilon^{na}+\varepsilon^{-na}$	0
B_1^c	1	-1	0	$-(\sigma^{mc}+\sigma^{-mc})$

(1) $1 \leqslant a \leqslant (q-2)/2$, $1 \leqslant c \leqslant q/2$.
(2) $1 \leqslant n \leqslant (q-2)/2$, $1 \leqslant m \leqslant q/2$.

Special Linear Group $SL(2,q)$ ($q=$ power of an odd prime number, $e=(q-1)/2$, $e'=(q+1)/2$).

			Y_n	Z_m		
$\begin{pmatrix}1 & \\ & 1\end{pmatrix}$	1	q	$q+1$	$q-1$	$\dfrac{q+1}{2}$	$\dfrac{q-1}{2}$
$Z=\begin{pmatrix}-1 & \\ & -1\end{pmatrix}$	1	q	$(-1)^n(q+1)$	$(-1)^m(q-1)$	$(-1)^e\dfrac{q+1}{2}$	$(-1)^{e'}\dfrac{q-1}{2}$
$P_1=\begin{pmatrix}1 & \\ 1 & 1\end{pmatrix}$	1	0	1	-1	$\mu^\pm$	$\lambda^\pm$
$P_2=\begin{pmatrix}1 & \\ \beta & 1\end{pmatrix}$	1	0	1	-1	$\mu^\mp$	$\lambda^\mp$
P_1Z	1	0	$(-1)^n$	$-(-1)^m$	$(-1)^e\mu^\pm$	$(-1)^{e'}\lambda^\pm$
P_2Z	1	0	$(-1)^n$	$-(-1)^m$	$(-1)^e\mu^\mp$	$(-1)^{e'}\lambda^\mp$
$\begin{pmatrix}\rho^a & \\ & \rho^{-a}\end{pmatrix}$	1	1	$\varepsilon^{na}+\varepsilon^{-na}$	0	$(-1)^a$	0
B_1^c	1	-1	0	$-(\sigma^{mc}+\sigma^{-mc})$	0	$-(-1)^c$

(1) $1 \leqslant a \leqslant (q-3)/2$, $1 \leqslant c \leqslant (q-1)/2$, $1 \leqslant n \leqslant (q-3)/2$, $1 \leqslant m \leqslant (q-1)/2$,
$\lambda^\pm = \{-1 \pm [(-1)^e q]^{1/2}\}/2$, $\mu^\pm = \{1 \pm [(-1)^e q]^{1/2}\}/2$.
(2) The last two columns mean two characters (with the same signs), respectively.

(III) Ree group $Re(q)$, Suzuki Group $Sz(q)$, and Janko Group J.

Ree group $Re(q)$ ($q=3^{2n+1}=3m^2$).
The order of $Re(q)$ is $q^3(q^3+1)(q-1)$, $q_0=q^2-q+1$, $m_+=q+3m+1$, $m_-=q-3m+1$.

						A	B	C	X_μ
1	1	1	q_0	q^3	qq_0	$(q-1)mm_+/2$	$(q-1)mm_-/2$	$m(q^2-1)$	q^3+1
J	2	1	-1	q	$-q$	$-(q-1)/2$	$(q-1)/2$	0	$q+1$
X	3	1	$-(q-1)$	0	q	$-(q+m)/2$	$(q-m)/2$	$-m$	1
Y	9	1	1	0	0	m	m	$-m$	1
T	3	1	1	0	0	α	α	2α	1
T^{-1}	3	1	1	0	0	$\bar{\alpha}$	$\bar{\alpha}$	$2\bar{\alpha}$	1
YT	9	1	1	0	0	β	β	$-\beta$	1
YT^{-1}	9	1	1	0	0	$\bar{\beta}$	$\bar{\beta}$	$-\bar{\beta}$	1
JT	6	1	-1	0	0	γ	$-\gamma$	0	1
JT^{-1}	6	1	-1	0	0	$\bar{\gamma}$	$-\bar{\gamma}$	0	1
R^a		1	1	1	1	0	0	0	$\rho^{\mu a}+\rho^{-\mu a}$
S^b		1	3	-1	-3	1	-1	0	0
JR^a		1	-1	1	-1	0	0	0	$\rho^{\mu a}+\rho^{-\mu a}$
JS^b		1	-1	-1	1	1	-1	0	0
V^s		1	0	-1	0	-1	0	-1	0
W^t		1	0	-1	0	0	1	1	0

		X'_μ	Y_ν	Y'_λ	Z_κ	Z'_τ
1	1	q^3+1	$(q-1)q_0$	$(q-1)q_0$	$(q^2-1)m_+$	$(q^2-1)m_-$
J	2	$-(q+1)$	$3(q-1)$	$-(q-1)$	0	0
X	3	1	$2q-1$	$2q-1$	$-m_+$	$-m_-$
Y	9	1	-1	-1	-1	-1
T	3	1	-1	-1	$-3m-1$	$3m-1$
T^{-1}	3	1	-1	-1	$-3m-1$	$3m-1$
YT	9	1	-1	-1	-1	-1
YT^{-1}	9	1	-1	-1	-1	-1
JT	6	-1	-3	1	0	0
JT^{-1}	6	-1	-3	1	0	0
R^a		$\rho^{\mu a}+\rho^{-\mu a}$	0	0	0	0
S^b		0	$\sigma(\nu b)$	$\sigma'(\lambda b)$	0	0
JR^a		$-(\rho^{\mu a}+\rho^{-\mu a})$	0	0	0	0
JS^b		0	$\sigma(\nu b)$	$\sigma'(\lambda b)$	0	0
V^s		0	0	0	$-\sum_{i=0}^{2}(v^{\kappa s q^i}+v^{-\kappa s q^i})$	0
W^t		0	0	0	0	$-\sum_{i=0}^{2}(w^{\tau t q^i}+w^{-\tau t q^i})$

(1) The first column gives a representative of conjugate class, and the second column gives its order. The orders of R, S, V, W are $(q-1)/2$, $(q+1)/4$, m_-, m_+, respectively. R, S, T are commutative with J.
(2) $R^a \sim R^{-a}$, $V^s \sim V^{sq} \sim V^{sq^2} \sim V^{-s} \sim V^{-sq} \sim V^{-sq^2}$, $W^t \sim W^{tq} \sim W^{tq^2} \sim W^{-t} \sim W^{-tq} \sim W^{-tq^2}$. Here we fix an integer δ satisfying $\delta^3 \equiv 1$ [mod$(q+1)/4$], $(\delta-1, (q+1)/4)=1$.

$$S^b \sim S^{b\delta} \sim S^{b\delta^2} \sim S^{-b} \sim S^{-b\delta} \sim S^{-b\delta^2}, \qquad JR^a \sim JR^{-a}, \qquad JS^b \sim JS^{-b},$$

where $A \sim B$ means that A and B are mutually conjugate.
(3) $\rho=\exp[4\pi\sqrt{-1}/(q-1)]$, $v=\exp(2\pi\sqrt{-1}/m_-)$, $w=\exp(2\pi\sqrt{-1}/m_+)$,
$\sigma=\exp[8\pi\sqrt{-1}/(q+1)]$.
(4) $1 \leqslant \mu \leqslant (q-3)/4$, $1 \leqslant \lambda \leqslant (q-3)/8$.
Here ν is considered mod$(q+1)/4$ and

$$Y_\nu = Y_{\nu\delta} = Y_{\nu\delta^2} = Y_{-\nu} = Y_{-\nu\delta} = Y_{-\nu\delta^2},$$

κ is considered mod m_- and

$$Z_\kappa = Z_{\kappa q} = Z_{\kappa q^2} = Z_{-\kappa} = Z_{-\kappa q} = Z_{-\kappa q^2},$$

τ is considered mod m_+ and

$$Z'_\tau = Z'_{\tau q} = Z'_{\tau q^2} = Z'_{-\tau} = Z'_{-\tau q} = Z'_{-\tau q^2}.$$

(5) $\sigma(\nu b) = -\sum_{i=0}^{2}(\sigma^{\nu b\delta^i}+\sigma^{-\nu b\delta^i})$, $\sigma'(\lambda b)=\sum_{i=0}^{1}(\sigma^{\lambda b\delta^i}+\sigma^{-\lambda b\delta^i})-(\sigma^{\lambda b\delta^2}+\sigma^{-\lambda b\delta^2})$.
(6) $\alpha=\dfrac{-m+m\sqrt{-q}}{2}$, $\beta=\dfrac{-m-\sqrt{-q}}{2}$, $\gamma=\dfrac{1-\sqrt{-q}}{2}$. We show one of the two mutually complex conjugate characters, for the characters A, B, C.

Suzuki group $Sz(q)$. The order of $Sz(q)$ is $q^2(q^2+1)(q-1)$ $(q=2^{2n+1}, 2q=r^2)$.

			X_α	Y_β	Z_γ		
1	1	q^2	q^2+1	$(q-r+1)(q-1)$	$(q+r+1)(q-1)$	$r(q-1)/2$	$r(q-1)/2$
σ	1	0	1	$r-1$	$-r-1$	$-r/2$	$-r/2$
ρ	1	0	1	-1	-1	$r\sqrt{-1}/2$	$-r\sqrt{-1}/2$
ρ^{-1}	1	0	1	-1	-1	$-r\sqrt{-1}/2$	$r\sqrt{-1}/2$
π_0^i	1	1	$\varepsilon_0^{\alpha i}+\varepsilon_0^{-\alpha i}$	0	0	0	0
π_1^j	1	-1	0	$-(\varepsilon_1^{\beta j}+\varepsilon_1^{\beta jq}+\varepsilon_1^{-\beta j}+\varepsilon_1^{-\beta jq})$	0	1	1
π_2^k	1	-1	0	0	$-(\varepsilon_2^{\gamma k}+\varepsilon_2^{\gamma kq}+\varepsilon_2^{-\gamma k}+\varepsilon_2^{-\gamma kq})$	-1	-1

(1) The first column gives a representative of the conjugate class.
(2) π_0, π_1, π_2 are the elements of order $q-1$ $q+r+1$, $q-r+1$, respectively.
(3) ε_0, ε_1, ε_2 are the primitive $q-1$, $q+r+1$, $q-r+1$ roots of 1, respectively.
(4) π_0^i and π_0^{-i} are mutually conjugate elements, and hence X_α and $X_{-\alpha}$ give the same character. i, α run over the representatives of mod $q-1$, and $i, \alpha \not\equiv 0$ (mod $q-1$).

(5) $\pi_1^j, \pi_1^{-j}, \pi_1^{jq}, \pi_1^{-jq}$ are mutually conjugate, and hence Y_β, $Y_{-\beta}$, $Y_{\beta q}$, $Y_{-\beta q}$ give the same character. j, β run over the representatives of $\bmod q+r+1$, and $j, \beta \not\equiv 0 \ (\bmod q+r+1)$.
(6) $\pi_2^k, \pi_2^{-k}, \pi_2^{kq}, \pi_2^{-kq}$ are mutually conjugate, and hence Z_γ, $Z_{-\gamma}$, $Z_{\gamma q}$, $Z_{-\gamma q}$ give the same character. k, γ run over the representatives of mod $q-r+1$, and $k, \gamma \not\equiv 0$ (mod $q-r+1$).

Janko Group J.

1	1	77	133	209	133	77	77	133	76	76	56	56	120	120	120
2	1	5	5	1	-3	-3	-3	-3	4	-4	0	0	0	0	0
3	1	-1	1	-1	-2	2	2	-2	1	1	2	2	0	0	0
5	1	2	-2	-1	ε^+	$-\varepsilon^+$	$-\varepsilon^-$	ε^-	1	1	$2\varepsilon^-$	$2\varepsilon^+$	0	0	0
5	1	2	-2	-1	ε^-	$-\varepsilon^-$	$-\varepsilon^+$	ε^+	1	1	$2\varepsilon^+$	$2\varepsilon^-$	0	0	0
6	1	-1	-1	1	0	0	0	0	1	-1	0	0	0	0	0
7	1	0	0	-1	0	0	0	0	-1	-1	0	0	1	1	1
10	1	0	0	1	$-\varepsilon^+$	$-\varepsilon^+$	$-\varepsilon^-$	$-\varepsilon^-$	-1	1	0	0	0	0	0
10	1	0	0	1	$-\varepsilon^-$	$-\varepsilon^-$	$-\varepsilon^+$	$-\varepsilon^+$	-1	1	0	0	0	0	0
11	1	0	1	0	1	0	0	1	-1	-1	1	1	-1	-1	-1
15	1	-1	1	-1	ε^+	$-\varepsilon^+$	$-\varepsilon^-$	ε^-	1	1	$-\varepsilon^-$	$-\varepsilon^+$	0	0	0
15	1	-1	1	-1	ε^-	$-\varepsilon^-$	$-\varepsilon^+$	ε^+	1	1	$-\varepsilon^+$	$-\varepsilon^-$	0	0	0
19	1	1	0	0	0	1	1	0	0	0	-1	-1	λ_1	λ_2	λ_3
19	1	1	0	0	0	1	1	0	0	0	-1	-1	λ_2	λ_3	λ_1
19	1	1	0	0	0	1	1	0	0	0	-1	-1	λ_3	λ_1	λ_2

(1) The order of J is $8\cdot 3\cdot 5\cdot 7\cdot 11\cdot 19=175560$.
(2) The first column gives the order of the elements of each conjugate class.
(3) $\rho=\exp(2\pi\sqrt{-1}\,/19)$, $\lambda_1=\rho+\rho^7+\rho^8+\rho^{11}+\rho^{12}+\rho^{18}$, $\lambda_2=\rho^2+\rho^{14}+\rho^{16}+\rho^3+\rho^5+\rho^{17}$, $\lambda_3=\rho^4+\rho^9+\rho^{13}+\rho^6+\rho^{10}+\rho^{15}$, $\varepsilon^\pm=(1\pm\sqrt{5}\,)/2$.

References

For $S_n (2\leqslant n\leqslant 10)$, $A_n (3\leqslant n\leqslant 9)$:
[1] D. E. Littlewood, The theory of group characters, second edition, Oxford Univ. Press, 1950.
For $S_n (n=11,12,13)$:
[2] M. Zia-ud-Din, Proc. London Math. Soc., 39(1935), 200–204, 42 (1937), 340–355.
For S_{14}:
[3] K. Kondô, Table of characters of the symmetric group of degree 14, Proc. Phys. Math. Soc. Japan, 22 (1940), 585–593.
For M_n $(n=12,24)$:
[4] G. Frobenius, Über die Charaktere der mehrfach transitive Gruppen, S. B. Preuss. Akad. Wiss., 1904, 558–571.
For M_n $(n=11,22,23)$:
[5] N. Burgoyne and P. Fong, The Schur multipliers of the Mathieu groups, Nagoya Math. J., 27 (1966), 733–745.
For $GL(2,q)$, $SL(2,q)$:
[6] I. Schur, Untersuchungen über die Darstellung der endlichen Gruppen durch gebrochene lineare Substitutionen, J. Reine Angew. Math., 132 (1907), 85–137.
[7] H. E. Jordan, Group characters of various types of linear groups, Amer. J. Math., 29 (1907), 387–405.
For $GL(3,q)$, $GL(4,q)$, $GL(n,q)$:
[8] R. Steinberg, The representations of $GL(3,q)$, $GL(4,q)$, $PGL(3,q)$ and $PGL(4,q)$, Canad. J. Math. 3 (1951), 225–235.
[9] J. A. Green, The characters of the finite general linear groups, Trans. Amer. Math. Soc., 80 (1955), 402–447.
For $U(2,q)$, $U(3,q)$:
[10] V. Ennola, On the characters of the finite unitary groups, Ann. Acad. Sci. Fenn., 323 (1963), 1–34.
For $Sz(q)$, $Re(q)$, J:
[11] M. Suzuki, On a class of doubly transitive groups, Ann. of Math., (2) 75 (1962), 105–145.
[12] H. N. Ward, On Ree's series of simple groups, Trans. Amer. Math. Soc., 121 (1966), 62–89.
[13] Z. Janko, A new finite simple group with Abelian Sylow 2-subgroups and its characterization, J. Algebra, 3 (1966), 147–186.

For other Lie-type groups:
[14] B. Srinivasan, The characters of the finite symplectic group $Sp(4,q)$, Trans. Amer. Math. Soc., 131 (1968), 488–525.
[15] H. Enomoto, The characters of the finite symplectic group $Sp(4,q)$, $q=2^f$, Osaka J. Math., 9 (1972), 75–94.
[16] G. I. Lehrer, The characters of the finite special linear groups, J. Algebra, 26 (1973), 564–583.
[17] P. Deligne and G. Lusztig, Representations of reductive groups over finite fields, Ann. of Math., (2) 103 (1976), 103–161.
For sporadic groups:
[18] M. Hall, Jr. and D. Wales, The simple groups of order 604, 800, J. Algebra, 9 (1968), 417–450.
[19] J. S. Frame, Computation of characters of the Higman-Sims group and its automorphism group, J. Algebra, 20 (1972), 320–349.
[20] D. Fendel, A characterization of Conway's group 3, J. Algebra, 24 (1973), 159–196.
[21] D. Wright, The irreducible characters of the simple group of M. Suzuki of order 448, 345, 497, 600, J. Algebra, 29 (1974), 303–323.

(IV) Three-Dimensional Crystal Classes (→ 92 Crystallographic Groups)

Crystal System Bravais Types	Geometric Crystal Classes			Arithmetic Crystal Classes	Number of Space Groups[2]
	Schoenflies Notation	International Notation[1]			
		Short	Full		
Triclinic	C_1	1	1	$(P,1)$	1
P	$S_2(C_i)$	$\bar{1}$	$\bar{1}$	$(P,\bar{1})$	2
Monoclinic	C_2	2	2	$(P,2)\,(C,2)$	3–5
P, C	C_{1h}	m	m	$(P,m)\,(C,m)$	6–9
	C_{2h}	$2/m$	$\frac{2}{m}$	$(P,2/m)\,(C,2/m)$	10–15
Orthorhombic	$D_2(V)$	222	222	$(P,222)\,(C,222)\,(F,222)\,(I,222)$	16–24
P, C, F, I	C_{2v}	$mm2$	$mm2$	$(P,mm2)\,(C,mm2)\,(A,mm2)\,(F,mm2)\,(I,mm2)$	25–46
	$D_{2h}(V_h)$	mmm	$\frac{2}{m}\frac{2}{m}\frac{2}{m}$	$(P,mmm)\,(C,mmm)\,(F,mmm)\,(I,mmm)$	47–74
Tetragonal	C_4	4	4	$(P,4)$[3] $(I,4)$	75–80
P, I	S_4	$\bar{4}$	$\bar{4}$	$(P,\bar{4})\,(I,\bar{4})$	81–82
	C_{4h}	$4/m$	$\frac{4}{m}$	$(P,4/m)\,(I,4/m)$	83–88
	D_4	422	422	$(P,422)$[4] $(I,422)$	89–98
	C_{4v}	$4mm$	$4mm$	$(P,4mm)\,(I,4mm)$	99–110
	$D_{2d}(V_d)$	$\bar{4}2m$	$\bar{4}2m$	$(P,\bar{4}2m)\,(P,\bar{4}m2)\,(I,\bar{4}m2)\,(I,\bar{4}2m)$	111–122
	D_{4h}	$4/mmm$	$\frac{4}{m}\frac{2}{m}\frac{2}{m}$	$(P,4/mmm)\,(I,4/mmm)$	123–142
Trigonal	C_3	3	3	$(P,3)$[5] $(R,3)$	143–146
P, R	$S_6(C_{3i})$	$\bar{3}$	$\bar{3}$	$(P,\bar{3})\,(R,\bar{3})$	147–148
	D_3	32	32	$(P,312)$[6] $(P,321)$[7] $(R,32)$	149–155
	C_{3v}	$3m$	$3m$	$(P,3m1)\,(P,31m)\,(R,3m)$	156–161
	D_{3d}	$\bar{3}m$	$\bar{3}\frac{2}{m}$	$(P,\bar{3}1m)\,(P,\bar{3}m1)\,(R,\bar{3}m)$	162–167
Hexagonal	C_6	6	6	$(P,6)$[8]	168–173
P	C_{3h}	$\bar{6}$	$\bar{6}$	$(P,\bar{6})$	–174
	C_{6h}	$6/m$	$\frac{6}{m}$	$(P,6/m)$	175–176
	D_6	622	622	$(P,622)$[9]	177–182

Crystal System Bravais Types	Geometric Crystal Classes			Arithmetic Crystal Classes	Number of Space Groups[2]
	Schoenflies Notation	International Notation[1]			
		Short	Full		
Hexagonal	C_{6v}	$6mm$	$6mm$	$(P, 6mm)$	183–186
P	D_{3h}	$\bar{6}m2$	$\bar{6}m2$	$(P, \bar{6}m2)\ (P, \bar{6}2m)$	187–190
(cont.)	D_{6h}	$6/mmm$	$\frac{6}{m}\frac{2}{m}\frac{2}{m}$	$(P, 6/mmm)$	191–194
Cubic	T	23	23	$(P, 23)\ (F, 23)\ (I, 23)$	195–199
P, F, I	T_h	$m3$	$\frac{2}{m}\bar{3}$	$(P, m3)\ (F, m3)\ (I, m3)$	200–206
	O	432	432	$(P, 432)$[10] $(F, 432)\ (I, 432)$	207–214
	T_d	$\bar{4}3m$	$\bar{4}3m$	$(P, \bar{4}3m)\ (F, \bar{4}3m)\ (I, \bar{4}3m)$	215–220
	O_h	$m3m$	$\frac{4}{m}\bar{3}\frac{2}{m}$	$(P, m3m)\ (F, m3m)\ (I, m3m)$	221–230

Notes
(1) The notation is based upon *International tables for X-ray crystallography* I, Kynoch, 1969. In each crystal system, the lowest class is a holohedry.
(2) These correspond to the consecutive numbers of space groups in the book cited in (1).
(3)–(10) Enantiomorphic pairs arise from these classes: two pairs for (4), (8), (9), and one pair for the others.
For the shapes of Bravais lattices → 92 Crystallographic Groups E, Fig. 3.

6. Miscellaneous Constants

$\sqrt{2}$ = 1.41421 35623 73095, $\sqrt{10}$ = 3.16227 76601 68379.
$\sqrt[3]{2}$ = 1.25992 10498 94873, $\sqrt[3]{100}$ = 4.64158 88336 12779.
$\log_{10}2$ = 0.30102 99956 63981 = 1/3.32192 80948 87364.

(I) Base of Natural Logarithm e (1000 decimals)

e = 2.71828 18284 59045 23536 02874 71352 66249 77572 47093 69995 95749 66967 62772 40766 30353 54759
45713 82178 52516 64274 27466 39193 20030 59921 81741 35966 29043 57290 03342 95260 59563 07381
32328 62794 34907 63233 82988 07531 95251 01901 15738 34187 93070 21540 89149 93488 41675 09244
76146 06680 82264 80016 84774 11853 74234 54424 37107 53907 77449 92069 55170 27618 38606 26133
13845 83000 75204 49338 26560 29760 67371 13200 70932 87091 27443 74704 72306 96977 20931 01416
92836 81902 55151 08657 46377 21112 52389 78442 50569 53696 77078 54499 69967 94686 44549 05987
93163 68892 30098 79312 77361 78215 42499 92295 76351 48220 82698 95193 66803 31825 28869 39849
64651 05820 93923 98294 88793 32036 25094 43117 30123 81970 68416 14039 70198 37679 32068 32823
76464 80429 53118 02328 78250 98194 55815 30175 67173 61332 06981 12509 96181 88159 30416 90351
59888 85193 45807 27386 67385 89422 87922 84998 92086 80582 57492 79610 48419 84443 63463 24496
84875 60233 62482 70419 78623 20900 21609 90235 30436 99418 49164 31409 34317 38143 64054 62531
52096 18369 08887 07016 76839 64243 78140 59271 45635 49061 30310 72085 10383 75051 01157 47704
17189 86106 87396 96552 12671 54688 95703 50354.

e (in octal) = 2.55760 52130 50535 5.
$1/e$ = 0.36787 94411 71442, e^2 = 7.38905 60989 30650 = 1/0.13533 52832 36613,
$\sqrt{e}$ = 1.64872 12707 00128 = 1/0.60653 06597 12633.
$\log_e 10$ = 2.30258 50929 94046 = 1/0.43429 44819 03252,
$\log_e 2$ = 0.69314 71805 59945 = 1/1.44269 50408 88964.

(II) The Number π (1000 decimals) (→ 328 Pi(π))

π = 3.14159 26535 89793 23846 26433 83279 50288 41971 69399 37510 58209 74944 59230 78164 06286 20899
86280 34825 34211 70679 82148 08651 32823 06647 09384 46095 50582 23172 53594 08128 48111 74502
84102 70193 85211 05559 64462 29489 54930 38196 44288 10975 66593 34461 28475 64823 37867 83165
27120 19091 45648 56692 34603 48610 45432 66482 13393 60726 02491 41273 72458 70066 06315 58817
48815 20920 96282 92540 91715 36436 78925 90360 01133 05305 48820 46652 13841 46951 94151 16094
33057 27036 57595 91953 09218 61173 81932 61179 31051 18548 07446 23799 62749 56735 18857 52724
89122 79381 83011 94912 98336 73362 44065 66430 86021 39494 63952 24737 19070 21798 60943 70277
05392 17176 29317 67523 84674 81846 76694 05132 00056 81271 45263 56082 77857 71342 75778 96091
73637 17872 14684 40901 22495 34301 46549 58537 10507 92279 68925 89235 42019 95611 21290 21960
86403 44181 59813 62977 47713 09960 51870 72113 49999 99837 29780 49951 05973 17328 16096 31859
50244 59455 34690 83026 42522 30825 33446 85035 26193 11881 71010 00313 78387 52886 58753 32083
81420 61717 76691 47303 59825 34904 28755 46873 11595 62863 88235 37875 93751 95778 18577 80532
17122 68066 13001 92787 66111 95909 21642 01989.

π (in octal) = 3.11037 55242 10264 3.
$1/\pi$ = 0.31830 98861 83791, π^2 = 9.86960 44010 89359 = 1/0.10132 11836 42338,
$\sqrt{\pi}$ = 1.77245 38509 05516 = 1/0.56418 95835 47756,
$\sqrt{2\pi}$ = 2.50662 82746 31001 = 1/0.39894 22804 01433,
$\sqrt{\pi/2}$ = 1.25331 41373 15500 = 1/0.79788 45608 02865,
$\sqrt[3]{\pi}$ = 1.46459 18875 61523 = 1/0.68278 40632 55296.
$\log_{10}\pi$ = 0.49714 98726 94134, $\log_e\pi$ = 1.14472 98858 49400.

(III) Radian rad

1 rad = 57°.29577 95130 82321 = 3437′.74677 07849 393 = 20626 4″.80624 70964.
1° = 0.01745 32925 19943 rad, 1′ = 0.00029 08882 08666 rad, 1″ = 0.00000 48481 36811 rad.

(IV) Euler's Constant C (100 decimals) (→ 174 Gamma Function)

C = 0.57721 56649 01532 86060 65120 90082 40243 10421 59335 93992
35988 05767 23488 48677 26777 66467 09369 47063 29174 67495.

e^C = 1.78107 24179 90197 98522.

$$S_n = \sum_{i=1}^{n} \frac{1}{i}.$$

n	S_n	n	S_n	n	S_n	n	S_n
3	1.83333 333	6	2.45000 000	15	3.31822 899	100	5.18737 752
4	2.08333 333	8	2.71785 714	20	3.59773 966	500	6.79282 343
5	2.28333 333	10	2.92896 825	50	4.79920 534	1000	7.48547 086

7. Coefficients of Polynomial Approximations

In this table, we give some typical examples of approximation formulas for computation of functions on a digital computer (→ 19 Analog Computation, 336 Polynomial Approximation).

(I) Exponential Function

(1) Putting $\frac{x}{\log 2} + 1 = q + y + \frac{1}{2}\left(q \text{ is an integer}, -\frac{1}{2} \leqslant y < \frac{1}{2}\right)$, we have
$e^x = 2^q v(y), v(y) \fallingdotseq \Sigma a_i y^i$, which gives an approximation by a polynomial of the 7th degree, where the maximal error is 3×10^{-11}.
a_0 = 0.70710 67811 6, a_1 = 0.49012 90717 2, a_2 = 0.16986 57957 2, a_3 = 0.03924 73321 5,
a_4 = 0.00680 09712, a_5 = 0.00094 28173, a_6 = 0.00010 93869, a_7 = 0.00001 0826.

(2) An approximation by a polynomial of the 11th degree: $e^x \fallingdotseq \Sigma a_i x^i$ $(-1 \leqslant x \leqslant 0)$.
Maximal error 1×10^{-12}.
a_0 = 0.99999 99999 990, a_1 = 0.99999 99999 995, a_2 = 0.50000 00000 747,
a_3 = 0.16666 66666 812, a_4 = 0.04166 66657 960, a_5 = 0.00833 33332 174,
a_6 = 0.00138 88925 998, a_7 = 0.00019 84130 955, a_8 = 0.00002 47944 428,
a_9 = 0.00000 27550 711, a_{10} = 0.00000 02819 019, a_{11} = 0.00000 00255 791.

(3) $e^x \fallingdotseq 1 + \cfrac{x}{-\cfrac{x}{2} + \cfrac{k_0 + k_1x^2 + k_2x^4}{1 + k_3x^2}}$ $(-\log\sqrt{2} \leqslant x \leqslant \log\sqrt{2})$.

Maximal error 1.4×10^{-14}.
$k_0 = 1.00000\,00000\,00327\,1$, $k_1 = 0.10713\,50664\,56464\,2$,
$k_2 = 0.00059\,45898\,69018\,8$, $k_3 = 0.02380\,17331\,57418\,6$.

(II) Logarithmic Function

(1) An approximation by a polynomial of the 11th degree: $\log(1+x) \fallingdotseq \Sigma a_i x^i$ $(0 \leqslant x \leqslant 1)$.
Maximal error 1.1×10^{-10}.
$a_0 = 0.00000\,00001\,10$, $a_1 = 0.99999\,99654\,98$, $a_2 = -0.49999\,82537\,98$,
$a_3 = 0.33329\,85059\,64$, $a_4 = -0.24963\,72428\,65$, $a_5 = 0.19773\,31015\,60$,
$a_6 = -0.15744\,88954\,13$, $a_7 = 0.11712\,91156\,18$, $a_8 = -0.07364\,03719\,14$,
$a_9 = 0.03469\,74937\,56$, $a_{10} = -0.01046\,82295\,69$, $a_{11} = 0.00148\,19917\,22$.

(2) For $1 \leqslant x \leqslant 2$, and putting $y = \frac{x - \sqrt{2}}{x + \sqrt{2}}(3 + 2\sqrt{2})$ $(-1 \leqslant y \leqslant 1)$, then $\log x \fallingdotseq \log\sqrt{2} + \sum a_i y^{2i+1}$ gives an approximation by a polynomial of the 11th degree $(0 \leqslant i \leqslant 5)$, where the maximal error is 9.2×10^{-15}.
$a_0 = 0.34314\,57505\,07610\,6$, $a_1 = 0.00336\,70892\,56222\,5$, $a_2 = 0.00005\,94707\,04347\,4$,
$a_3 = 0.00000\,12504\,99776\,2$, $a_4 = 0.00000\,00285\,68292\,8$, $a_5 = 0.00000\,00007\,43713\,9$.

(III) Trigonometric Functions

(1) We put $\frac{x}{2\pi} = p + \frac{q}{2} + \frac{r}{4} + \frac{z}{8}$ (p is an integer; $q = 0, 1$; $r = 0, 1$; $-1 \leqslant z < 1$), and $s = \sin\frac{\pi z}{4}$, $c = \cos\frac{\pi z}{4}$.
If $r = 0$, $\sin x = (-1)^q s$, $\cos x = (-1)^q c$,
If $r = 1$, $\sin x = (-1)^q c$, $\cos x = -(-1)^q s$.
Here s and c are computed by the following approximation formulas. Putting $-z^2/2 = y$, $s(y) = \sin(\pi z/4) \fallingdotseq z\Sigma a_i y^i$, $c(y) = \cos(\pi z/4) \fallingdotseq \Sigma b_i y^i$ gives an approximation by a polynomial of the 5th degree, where the maximal errors are s: 2×10^{-15}, c: 2×10^{-13}.
$a_0 = 0.78539\,81633\,97426$, $a_1 = 0.16149\,10243\,75338$, $a_2 = 0.00996\,15782\,61200$,
$a_3 = 0.00029\,26094\,99152$, $a_4 = 0.00000\,50133\,389$, $a_5 = 0.00000\,00555\,1357$,
$b_0 = 0.99999\,99999\,999$, $b_1 = 0.61685\,02750\,601$, $b_2 = 0.06341\,73767\,885$,
$b_3 = 0.00260\,79335\,007$, $b_4 = 0.00005\,74476\,09$, $b_5 = 0.00000\,07765\,93$.

(2) $\frac{\sin(\pi x/2)}{x} \fallingdotseq \Sigma(-1)^i a_i x^{2i}$ $(-1 \leqslant x \leqslant 1)$. This gives an approximation by a polynomial of 10th degree $(0 \leqslant i \leqslant 5)$, where the maximal error is 2.67×10^{-11}.
$a_0 = 1.57079\,63267\,682$, $a_1 = 0.64596\,40955\,820$, $a_2 = 0.07969\,26037\,435$,
$a_3 = 0.00468\,16578\,837$, $a_4 = 0.00016\,02547\,767$, $a_5 = 0.00000\,34318\,696$.

(3) $\tan\frac{\pi x}{4} \fallingdotseq x\left(k_0 + \frac{x^2|}{|k_1} + \ldots + \frac{x^2|}{|k_4}\right)$ (continued fraction) $(-1 \leqslant x \leqslant 1)$.
Maximal error 9.8×10^{-12}.
$k_0 = 0.78539\,81634\,9907$, $k_1 = 6.19229\,46807\,1350$, $k_2 = -0.65449\,83095\,2316$,
$k_3 = 520.24599\,06398\,9939$, $k_4 = -0.07797\,95098\,7751$.

(IV) Inverse Trigonometric Functions

(1) An approximation by a polynomial of the 21st degree $(0 \leqslant i \leqslant 10)$:
$\arcsin x \fallingdotseq \Sigma a_i x^{2i+1}$ $(|x| \leqslant 1/\sqrt{2})$.
Maximal error 10^{-10}.
$a_0 = 1.00000\,00005\,3$, $a_1 = 0.16666\,65754\,5$, $a_2 = 0.07500\,46066\,5$, $a_3 = 0.04453\,58425\,7$,
$a_4 = 0.03175\,26509\,6$, $a_5 = 0.01176\,58281\,9$, $a_6 = 0.06921\,26185\,7$, $a_7 = -0.14821\,09628\,8$,
$a_8 = 0.32889\,76635\,2$, $a_9 = -0.35020\,41201\,5$, $a_{10} = 0.19740\,50325\,0$.

(2) Putting $x=w+u$ $(w=\frac{1}{8},\frac{3}{8},\frac{5}{8},\frac{7}{8};\ -\frac{1}{8}\leqslant u\leqslant\frac{1}{8})$, $v=\dfrac{x-w}{1+xw}$ $(|v|\leqslant\frac{1}{2})$

$\arctan x=\arctan w+t(v)$, $t(v)=\arctan v$.

The values of $\arctan w$:

$\arctan(1/8)=0.12435\,49945\,46711$, $\arctan(3/8)=0.35877\,06702\,70611$,
$\arctan(5/8)=0.55859\,93153\,43560$, $\arctan(7/8)=0.71882\,99996\,21623$.

$t(v)$ is computed by an approximation by a polynomial of the 9th degree $(0\leqslant i\leqslant 4)$, where $t(v)=\arctan v\fallingdotseq\sum(-1)^i a_i v^{2i+1}$.

Maximal error 1.6×10^{-13}.

$a_0=0.99999\,99999\,9992$, $a_1=0.33333\,33328\,220$, $a_2=0.19999\,97377\,6$,
$a_3=0.14280\,9976$, $a_4=0.10763\,60$.

(3) $\arctan x\fallingdotseq x\left(k_0+\dfrac{x^2|}{|k_1}+\ldots+\dfrac{x^2|}{|k_6}\right)$ (continued fraction) $(-1\leqslant x\leqslant 1)$.

Maximal error 3.6×10^{-10}.

$k_0=0.99999\,99936\,2$, $k_1=-3.00000\,30869\,4$, $k_2=-0.55556\,97728\,4$,
$k_3=-15.77401\,81127\,3$,
$k_4=-0.16190\,80978\,0$, $k_5=-44.57191\,79508\,8$, $k_6=-0.10810\,67493\,1$.

(V) Gamma Function

An approximation by a polynomial of the 8th degree:

$\Gamma(2+x)\fallingdotseq\sum a_i x^i$ $(-1/2\leqslant x\leqslant 1/2)$.

Maximal error 7.6×10^{-8}.

$a_0=0.99999\,9926$, $a_1=0.42278\,4604$, $a_2=0.41184\,9671$, $a_3=0.08156\,52323$,
$a_4=0.07406\,48982$. $a_5=-0.00012\,51376\,7$, $a_6=0.01229\,95771$, $a_7=-0.00349\,61289$,
$a_8=0.00213\,85778$.

(VI) Normal Distribution

(1) $\dfrac{2}{\sqrt{\pi}}\displaystyle\int_x^\infty e^{-t^2}dt\fallingdotseq\frac{1}{(1+\sum a_i x^{i+1})^{16}}$ $(0\leqslant x<\infty)$. This gives an approximation by a polynomial of the 6th degree.

Maximal error 2.8×10^{-7}.

$a_0=0.07052\,30784$, $a_1=0.04228\,20123$, $a_2=0.00927\,05272$,
$a_3=0.00015\,20143$, $a_4=0.00027\,65672$, $a_5=0.00004\,30638$.

(2) $P(x)\equiv\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_x^\infty e^{-t^2/2}dt$,

$$4P(x)(1-P(x))\fallingdotseq\left[\exp\left(-\frac{2x^2}{\pi}\right)\right]\left[1+x^4\left(a_0+\frac{a_1}{x^2+a_2}\right)\right]\qquad(0\leqslant x<\infty).$$

Maximal error 2×10^{-5}.

$a_0=0.0055$, $a_1=0.0551$, $a_2=14.4$.

(3) The inverse function of (2)

$$x\fallingdotseq\left[y\left(a_0+\frac{a_1}{y+a_2}\right)\right]^{1/2},\quad y=-\log[4P(x)(1-P(x))]\qquad(0\leqslant y<\infty).$$

Maximal error 4.9×10^{-4}.

$a_0=2.06117\,86$, $a_1=-5.72622\,04$, $a_2=11.64059\,5$.

Statistical Tables for Reference

Statistical Tables

[1] J. A. Greenwood and H. O. Hartley, Guide to tables in mathematical statistics, Princeton Univ. Press, 1962.
[2] Research Group for Statistical Sciences (T. Kitagawa and M. Masuyama, eds.) New statistical tables (Japanese), explanation p. 264, table p. 214, Kawade, 1952.
[3] R. A. Fisher and F. Yates, Statistical tables for biological, agricultural and medical research, explanation p. 30, table p. 137, Oliver & Boyd, third edition, 1948.
[4] E. S. Pearson and H. O. Hartley, Biometrika tables for statisticians, explanation p. 104, table p. 154, Cambridge Univ. Press, third edition, 1970.
[5] K. Pearson, Tables for statisticians and biometricians I, 1930, explanation p. 83, table p. 143; II, 1931, explanation p. 250, table p. 262, Cambridge Univ. Press.
[6] Statistical tables JSA-1972 (Japanese), table p. 454, explanation p. 260, Japanese Standards Association, 1972.

Tables of Special Statistical Values

[8] Harvard Univ., Tables of the cumulative binomial probability distribution, Harvard, 1955,

$\sum_{i=r}^{n} \binom{n}{i} p^i q^{n-i}$: 5 dec.,

$p=0.01(0.01)0.50$,
$n=1(1)50(2)100(10)200(20)500(50)1000$.
[9] National Bureau of Standards, NBS applied mathematical series, no. 6, Tables of the binomial probability distribution, 1950, $\binom{n}{i} p^i q^{n-i}$ and the partial sum: 7 dec., $p=0.01(0.01)0.50$, $n=2(1)49$.
[10] T. Kitagawa, Table of Poisson distribution (Japanese), Baihûkan, 1951, $e^{-m}m^i/i!$: 7-8 dec., $m=0.001(0.001)1.000(0.01)10.00$.
[11] G. J. Lieberman and D. B. Owen, Tables of the hypergeometric probability distribution, Stanford, 1961, $\binom{k}{x}\binom{N-k}{n-x} \Big/ \binom{N}{n}$: 6 dec., $N=2(1)50(10)100(100)2000$, $n=1(1)(N/2)$, $k=1(1)n$.
[12] National Bureau of Standards, NBS no. 23, Tables of normal probability functions, 1942,
$\varphi(x)=(1/\sqrt{2\pi})\exp(-\frac{1}{2}x^2)$,
$\Phi(x)=\int_{-x}^{x}\varphi(x)dx$: 15 dec.,
$x=0(0.00001)1.0000(0.001)8.285$.
[13] K. Pearson, Tables of the incomplete beta-functions, Cambridge, second edition, 1968, $I_x(p,q)$: 8 dec., $p, q=0.5(0.5)\ 11(1)50$.
[14] K. Pearson, Tables of the incomplete gamma-function, Cambridge, 1922, revised edition, 1951,
$I(u,p)=\int_0^{u\sqrt{p+1}}(1/e^v)(v^p/\Gamma(p+1))dv$:
7 dec., $p=0.0(0.1)5.0(0.2)50.0$, $u=0.1(0.1)20.0$; $p=-1.0(0.05)0.0$, $u=0.1(0.1)51.3$.
[15] N. V. Smirnov, Tables for the distribution and density function of the t-distribution, Pergamon, 1961, 6 dec., $f=1(1)35$, $t=0(0.01)3.00(0.02)4.50(0.05)6.50$.
[16] G. J. Resnikoff and G. J. Lieberman, Tables of the non-central t-distribution, Stanford, 1957.
[17] F. N. David, Tables of the ordinates and probability integral of the distribution of the correlation coefficient in small samples, Cambridge, 1938.
[18] D. B. Owen, The bivariate normal probability distribution, Sandia Corp., 1957, $T(h,a)$: 6 dec., $a=0.000(0.025)1.000, \infty$, $h=0.00(0.01)3.50(0.05)4.75$, $T(h,a)$
$=\int_0^h\int_0^{ax}\frac{1}{2\pi}\exp\left(-\frac{x^2+y^2}{2}\right)dy\,dx$.
[19] National Bureau of Standards, NBS no. 50, Tables of the bivariate normal distribution function and related functions, 1959, $L(h,k,r)$
$=\int_h^\infty\int_k^\infty\frac{1}{2\pi\sqrt{1-r^2}}$
$\times\exp\left[-\frac{x^2+y^2-2rxy}{2(1-r^2)}\right]dy\,dx$: 6 dec.,
$r=\pm 0.00(0.05)0.95(0.01)0.99$,
$h, k=0.0(0.1)4.0$.

Tables of Allocation

[20] T. Kitagawa and M. Midome, Table of allocation of elements for experimental design (Japanese), Baihûkan, 1953.
[21] R. C. Bose, W. H. Clatworthy, and S. S. Shrikhande, Tables of partially balanced designs with two associate classes, North Carolina Agric. Expt. Station Tech. Bull., 1954 (table of PBIBD).

Numerical Tables for Reference

General Tables

[1] M. Boll, Tables numériques universelles, Dunod, 1947.
[2] P. Barlow, Barlow's tables, Robinson, 1814, third edition, 1930.
[3] W. Shibagaki, 0.01% table of elementary functions (Japanese), Kyôritu, 1952.
[4] K. Hayashi, Table of higher functions (Japanese), Iwanami, second edition, 1967.
[5] E. Jahnke and F. Emde, Funktionentafeln mit Formelin und Kurven, Teubner, second edition 1933 (English translation: Tables of functions with formulae and curves, Dover, fourth edition, 1945).
[6] Y. Yoshida and M. Yoshida, Mathematical tables (Japanese), Baihûkan, 1958.
[7] M. Abramowitz and I. A. Stegun (eds.), Handbook of mathematical functions with formulas, graphs and mathematical tables, National Bureau of Standards, 1964 (Dover, 1965).
[8] A. Fletcher et al. (eds.), Index of mathematical tables I, II, Scientific Computing Service, Addison-Wesley, second edition, 1962.

Multiplication Table

[9] A. L. Crelle, Rechentafeln welche alles Mutiplicieren und Dividiren mit Zahlen unter 1000 ersparen, bei grösseren Zahlen aber die Rechung erleichtern und sicherer machen, W. de Gruyter, new edition 1944.

Table of Prime Numbers

[10] D. N. Lehmer, List of prime numbers from 1 to 10,006,721, Carnegie Institution of Washington, 1914.

Series of Tables of Functions

[11] British Association for the Advancement of Science, Mathematical tables, vol. 2, Emden functions, 1932; vol. 6, Bessel functions, pt. 1, 1937; vol. 8, Number-divisor tables, 1940; vol. 9, Tables of powers giving integral powers of integrals, 1940; vol. 10, Bessel functions, pt. 2, 1952.
[12] Harvard University, Computation Laboratory, Annals, vol. 2, Tables of the modified Hankel functions of order one-third and their derivatives, 1945; vol. 3, Tables of the Bessel functions of the first kind of orders zero and one, 1947; vol. 14, Orders seventy-nine through one hundred thirty-five, 1951; vol. 18, Tables of generalized sine- and cosine-integral functions, pt. 1, 1949; vol. 19, pt. 2, 1949; vol. 20, Tables of inverse hyperbolic functions, 1949; vol. 21, Tables of the generalized exponential-integral functions, 1949.
[13] National Bureau of Standards, Applied Mathematics Series (AMS), AMS 1, Tables of the Bessel functions $Y_0(x)$, $Y_1(x)$, $K_0(x)$, $K_1(x)$, $0 \leqslant x \leqslant 1$, 1948; AMS 5, Tables of sines and cosines to fifteen decimal places at hundredths of a degree, 1949; AMS 11, Table of arctangents of rational numbers, 1951; AMS 14, Tables of the exponential function e^x (including e^{-x}), 1951; AMS 16, Tables of $n!$ and $\Gamma(n+\frac{1}{2})$ for the first thousand values of n, 1951; AMS 23, Tables of normal probability functions, 1953; AMS 25, Tables of the Bessel functions $Y_0(x)$, $Y_1(x)$, $K_0(x)$, $K_1(x)$, $0 \leqslant x \leqslant 1$, 1952; AMS 26, Tables of Arctan x, 1953; Tables of 10^x, 1953; AMS 32, Table of sine and cosine integrals for arguments from 10 to 100, 1954; AMS 34, Table of the gamma function for complex arguments, 1954; AMS 36, Tables of circular and hyperbolic sines and cosines for radian arguments, 1953; AMS 40, Table of secants and cosecants to nine significant figures at hundredths of a degree, 1954; AMS 41, Tables of the error function and its derivative, 1954; AMS 43, Tables of sines and cosines for radian arguments, 1955; AMS 45, Table of hyperbolic sines and cosines, 1955; AMS 46, Table of the descending exponential, 1955.

Tables of Special Functions

[14] Akademiya Nauk SSSR, Tables of the exponential integral functions, 1954.
[15] J. Brownlee, Table of $\log \Gamma(x)$, Tracts for computers, no. 9, Cambridge Univ. Press, 1923.
[16] L. Dolansky and M. P. Dolansky, Table of $\log_2(1/p)$, $p \cdot \log_2(1/p)$ and $p \cdot \log_2(1/p) + (1-p)\log_2(1/(1-p))$, M.I.T. Research Lab. of Electronics tech. report 227, 1952.
[17] L. M. Milne-Thomson, Die elliptischen Funktionen von Jacobi, Springer, 1931; English translation: Jacobian elliptic function tables, Dover, 1950.
[18] H. T. Davis, Tables of the higher mathematical functions, Principia Press, vol. 1, Gamma function, 1933; vol. 2, Polygamma functions, 1935; vol. 3, Arithmetical tables, 1962.
[19] R. G. Selfridge and J. E. Maxfield, A table of the incomplete elliptic integral of the third kind, Dover, 1958.

[20] W. Shibagaki, 0.01% table of modified Bessel functions and the method of numerical computation for them (Japanese), Baihûkan, 1955.
[21] W. Shibagaki, Theory and application of gamma function, Appendix. Table of gamma function of complex variable effective up to 6 decimals (Japanese), Iwanami, 1952.
[22] L. J. Slater, A short table of the Laguerre polynomials, Proc. IEE, monograph no. 103c, 1955, 46–50.
[23] G. N. Watson, A treatise on the theory of Bessel functions, Appendix, Cambridge Univ. Press, 1922, second edition, 1958.
See also Statistical Tables for Reference.

Tables of Approximation Formulas of Functions

[24] S. Hitotumatu, Approximation formula (Japanese) Takeuti, 1963.
[25] T. Uno (ed.), Approximation formulas of functions for computers 1–3, Joint research work for mathematical sciences, pt. IV, section 5, 1961–1963 (Japanese, mimeographed).
[26] Z. Yamauti, S. Moriguti, and S. Hitotumatu (eds.), Numerical methods for electronic computers (Japanese), Baihûkan, I, 1965; II, 1967.
[27] C. Hastings (ed.), Approximations for digital computers, Princeton, 1955.
[28] A. J. W. Duijvestijn and A. J. Dekkers, Chebyshev approximations of some transcendental functions for use in digital computing, Philips Research Reports, 16 (1961), 145–174.
[29] L. A. Lyusternik, O. A. Chervonenkis, and A. P. Yanpol'skiĭ, Mathematical analysis: functions, limits, series, continued fractions, Pergamon, 1965. (Original in Russian, 1963.)
[30] J. F. Hart et al. (eds.), Computer approximations, Wiley, 1968.

Journals

Abh. Akad. Wiss. Göttingen
Abhandlungen der Akademie der Wissenschaften in Göttingen (Göttingen)

Abh. Bayer. Akad. Wiss. Math.-Nat. Kl.
Abhandlungen der Bayerschen Akademie der Wissenschaften. Mathematisch-Naturwissenschaftliche Klasse (Munich)

Abh. Math. Sem. Univ. Hamburg
Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg (Hamburg)

Acta Arith.
Acta Arithmetica, Polska Akademia Nauk, Instytut Matematyczny (Warsaw)

Acta Informat.
Acta Informatica (Berlin)

Acta Math.
Acta Mathematica (Uppsala)

Acta Math. Acad. Sci. Hungar.
Acta Mathematica Academiae Scientiarum Hungaricae (Budapest)

Acta Math. Sinica
Acta Mathematica Sinica (Peking). Chinese Math. Acta is its English translation

Acta Sci. Math. Szeged.
Acta Universitatis Szegediensis, Acta Scientiarum Mathematicarum (Szeged)

Actualités Sci. Ind.
Actualités Scientifiques et Industrielles (Paris)

Advances in Math.
Advances in Mathematics (New York)

Aequationes Math.
Aequationes Mathematicae (Basel-Waterloo)

Algebra and Logic
Algebra and Logic (New York). Translation of Algebra i Logika

Algebra i Logika
Akademiya Nauk SSSR Sibirskoe Otdelenie. Institut Matematiki. Algebra i Logika (Novosibirsk). Translated as Algebra and Logic

Algebra Universalis
Algebra Universalis (Basel)

Amer. J. Math.
American Journal of Mathematics (Baltimore)

Amer. Math. Monthly
The American Mathematical Monthly (Menasha)

Amer. Math. Soc. Colloq. Publ.
American Mathematical Society Colloquium Publications (Providence)

Amer. Math. Soc. Math. Surveys
American Mathematical Society Mathematical Surveys (Providence)

Amer. Math. Soc. Proc. Symp. Pure Math.
American Mathematical Society Proceedings of Symposia in Pure Mathematics (Providence)

Amer. Math. Soc. Transl.
American Mathematical Society Translations (Providence)

Amer. Math. Soc. Transl. Math. Monographs
American Mathematical Society Translations of Mathematical Monographs (Providence)

Ann. Acad. Sci. Fenn.
Suomalaisen Tiedeakatemian Toimituksia, Annales Academiae Scientiarum Fennicae. Series A. I. Mathematica (Helsinki)

Ann. der Phys.
Annalen der Physik (Leipzig)

Ann. Fac. Sci. Univ. Toulouse
Annales de la Faculté des Sciences de l'Université de Toulouse pour les Sciences Mathématiques et les Sciences Physiques (Toulouse)

Ann. Inst. Fourier
Annales de l'Institut Fourier, Université de Grenoble (Grenoble)

Ann. Inst. H. Poincaré
Annales de l'Institut Henri Poincaré (Paris)

Ann. Inst. Statist. Math.
Annals of the Institute of Statistical Mathematics (Tokyo)

Ann. Mat. Pura Appl.
Annali di Matematica Pura ed Applicata (Bologna)

Ann. Math. Statist.
The Annals of Mathematical Statistics (Baltimore)

Ann. Math.
Annals of Mathematics (Princeton)

Ann. Physique
Annales de Physique (Paris)

Ann. Probability
The Annals of Probability (San Francisco)

Ann. Polon. Math.
Annales Polonici Mathematici. Polska Akademia Nauk (Warsaw)

Ann. Roumaines Math.
Annales Roumaines de Mathématiques. Journal de l'Institut Mathématique Roumain (Bucharest)

Ann. Sci. Ecole Norm. Sup.
Annales Scientifiques de l'Ecole Normale Supérieure (Paris)

Ann. Scuola Norm. Sup. Pisa
Annali della Scuola Normale Superiore de Pisa, Scienze Fisiche e Matematiche (Pisa)

Ann. Statist.
The Annals of Statistics (San Francisco)

Arch. History Exact Sci.
Archive for History of Exact Sciences (Berlin-New York)

Arch. Math.
Archiv der Mathematik (Basel-Stuttgart)

Arch. Rational Mech. Anal.
Archive for Rational Mechanics and Analysis (Berlin)

Ark. Mat.
Arkiv för Matematik (Stockholm)

Ark. Mat. Astr. Fys.
Arkiv för Matematik, Astronomi och Fysik (Uppsala)

Astérique
Astérique. La Société Mathématique de France (Paris)

Atti Accad. Naz. Lincei, Mem.
Atti della Accademia Nazionale dei Lincei, Memorie (Rome)

Atti Accad. Naz. Lincei, Rend.
Atti della Accademia Nazionale dei Lincei, Rendiconti (Rome)

Automat. Control Comput. Sci.
Automatic Control and Computer Sciences (New York). Translation of Avtomatika i Vychislitel'naya Tekhnik. Akademiya Nauk Latviĭskoĭ SSR (Riga)

Bell System Tech. J.
The Bell System Technical Journal (New York)

Biometrika
Biometrika, A Journal for the Statistical Study of Biological Problems (London)

Bol. Soc. Mat. São Paulo
Boletim da Sociedade de Matemática de São Paulo (São Paulo)

Bull. Acad. Pol. Sci.
Bulletin de l'Academie Polonaise des Sciences (Warsaw)

Bull. Amer. Math. Soc.
Bulletin of the American Mathematical Society (Providence)

Bull. Calcutta Math. Soc.
Bulletin of the Calcutta Mathematical Society (Calcutta)

Bull. Math. Statist.
Bulletin of Mathematical Statistics (Fukuoka, Japan)

Bull. Nat. Res. Council
Bulletin of the National Research Council (Washington)

Bull. Sci. Math.
Bulletin des Sciences Mathématiques (Paris)

Bull. Soc. Math. Belg.
Bulletin de la Société Mathématique de Belgique (Brussels)

Bull. Soc. Math. France
Bulletin de la Société Mathématique de France (Paris)

Bull. Soc. Roy. Sci. Liège
Bulletin de la Société Royale des Sciences de Liège (Liège)

C. R. Acad. Sci. Paris
Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences (Paris)

Canad. J. Math.
Canadian Journal of Mathematics (Toronto)

Časopis Pěst. Mat.
Časopis pro Pěstováni Matematiky, Československá Akademie Věd (Prague)

Colloq. Math.
Colloquium Mathematicum (Warsaw)

Comm. ACM
Communications of the Association for Computing Machinery (New York)

Comm. Math. Phys.
Communications in Mathematical Physics (Berlin)

Comm. Pure Appl. Math.
Communications on Pure and Applied Mathematics (New York)

Comment. Math. Helv.
Commentarii Mathematici Helvetici (Zurich)

Compositio Math.
Compositio Mathematica (Groningen)

Comput. J.
The Computer Journal (London)

Crelles J.
= J. Reine Angew. Math.

CWI Newslett.
Centrum voor Wiskunde en Informatica. Newsletter (Amsterdam)

Cybernetics
Cybernetics (New York). Translation of Kibernetika (Kiev)

Czech. Math. J.
Czechoslovak Mathematical Journal (Prague)

Deutsche Math.
Deutsche Mathematik (Berlin)

Differentsial'nye Uravneniya
Differentsial'nye Uravneniya (Minsk). Translated as Differential Equations

Differential Equations
Differential Equations (New York). Translation of Differentsial'nye Uravneniya

Dokl. Akad. Nauk SSSR
Doklady Akademii Nauk SSSR (Moscow). Soviet Math. Dokl. is the English translation of its mathematics section

Duke Math. J.
Duke Mathematical Journal (Durham)

Econometrica
Econometrica, Journal of the Econometric Society (Chicago)

Edinburgh Math. Notes
The Edinburgh Mathematical Notes (Edinburgh)

Enseignement Math.
L'Enseignement Mathématique (Geneva)

Enzykl. Math.
Enzyklopädie der Mathematischen Wissenschaften mit Einschluss ihrer Anwendungen (Berlin)

Erg. Angew. Math.
Ergebnisse der Angewandte Mathematik (Berlin-New York)

Erg. Math.
Ergebnisse der Mathematik und ihrer Grenzgebiete (Berlin-New York)

Fund. Math.
Fundamenta Mathematicae (Warsaw)

Funkcial. Ekvac.
Fako de l'Funkcialaj Ekvacioj Japana Matematika Societo. Funkcialaj Ekvacioj (Serio Internacia) (Kobe, Japan)

Functional Anal. Appl.
Functional Analysis and its Applications (New York). Translation of Funktsional. Anal. Prilozhen.

Funktsional. Anal. Prilozhen.
Funktsional'nyi Analiz i ego Prilozheniya. Akademiya Nauk SSSR (Moscow). Translated as Functional Anal. Appl.

General Topology and Appl.
General Topology and its Applications (Amsterdam)

Hiroshima Math. J.
Hiroshima Mathematical Journal. Hiroshima Univ. (Hiroshima, Japan)

Hokkaido Math. J.
Hokkaido Mathematical Journal. Hokkaido Univ. (Sapporo, Japan)

IBM J. Res. Develop.
IBM Journal of Research and Development (Armonk, N.Y.)

Illinois J. Math.
Illinois Journal of Mathematics (Urbana)

Indag. Math.
Indagationes Mathematicae = Nederl. Akad. Wetensch. Proc.

Indian J. Math.
Indian Journal of Mathematics (Allahabad)

Indiana Univ. Math. J.
Indiana University Mathematics Journal (Bloomington)

Information and Control
Information and Control (New York)

Inventiones Math.
Inventiones Mathematicae (Berlin)

Izv. Akad. Nauk SSSR
Izvestiya Akademii Nauk SSSR (Moscow). Math. USSR-Izv. is the English translation of its mathematics section

J. Algebra
Journal of Algebra (New York)

J. Analyse Math.
Journal d'Analyse Mathématiques (Jerusalem)

J. Appl. Math. Mech.
Journal of Applied Mathematics and Mechanics (New York). Translation of Prikl. Mat. Mekh.

J. Approximation Theory
Journal of Approximation Theory (New York)

J. Assoc. Comput. Mach. (J. ACM)
Journal of the Association for Computing Machinery (New York)

J. Austral. Math. Soc.
The Journal of the Australian Mathematical Society (Sydney)

J. Combinatorial Theory
Journal of Combinatorial Theory. Series A and Series B (New York)

J. Comput. System Sci.
Journal of Computer and System Sciences (New York)

J. Differential Equations
Journal of Differential Equations (New York)

J. Differential Geometry
Journal of Differential Geometry (Bethlehem, Pa.)

J. Ecole Polytech.
Journal de l'Ecole Polytechnique (Paris)

J. Fac. Sci. Hokkaido Univ.
Journal of the Faculty of Science, Hokkaido University. Series I. Mathematics (Sapporo, Japan)

J. Fac. Sci. Univ. Tokyo
Journal of the Faculty of Science, University of Tokyo. Section I. (Tokyo)

J. Franklin Inst.
Journal of the Franklin Institute (Philadelphia)

J. Functional Anal.
Journal of Functional Analysis (New York)

J. Indian Math. Soc.
The Journal of the Indian Mathematical Society (Madras)

J. Inst. Elec. Engrs.
Journal of the Institution of Electrical Engineers (London)

J. Inst. Polytech. Osaka City Univ.
Journal of the Institute of Polytechnics, Osaka City University. Series A. Mathematics (Osaka)

J. London Math. Soc.
The Journal of the London Mathematical Society (London)

J. Math. Anal. Appl.
Journal of Mathematical Analysis and Applications (New York)

J. Math. and Phys.
Journal of Mathematics and Physics (Cambridge, Massachusetts, for issues prior to 1975; for 1975 and later, New York)

J. Math. Econom.
Journal of Mathematical Economics (Amsterdam)

J. Math. Kyoto Univ.
Journal of Mathematics of Kyoto University (Kyoto)

J. Math. Mech.
Journal of Mathematics and Mechanics (Bloomington)

J. Math. Pures Appl.
Journal de Mathématiques Pures et Appliquées (Paris)

J. Math. Soc. Japan
Journal of the Mathematical Society of Japan (Tokyo)

J. Mathematical Phys.
Journal of Mathematical Physics (New York)

J. Multivariate Anal.
Journal of Multivariate Analysis (New York)

J. Number Theory
Journal of Number Theory (New York)

J. Operations Res. Soc. Japan
Journal of the Operations Research Society of Japan (Tokyo)

J. Optimization Theory Appl.
Journal of Optimization Theory and Applications (New York)

J. Phys. Soc. Japan
Journal of the Physical Society of Japan (Tokyo)

J. Pure Appl. Algebra
Journal of Pure and Applied Algebra (Amsterdam)

J. Rational Mech. Anal.
Journal of Rational Mechanics and Analysis (Bloomington)

J. Reine Angew. Math.
Journal für die Reine und Angewandte Mathematik (Berlin). = Crelles J.

J. Res. Nat. Bur. Standards
Journal of Research of the National Bureau of Standards. Section B. Mathematics and Mathematical Physics (Washington)

J. Sci. Hiroshima Univ.
Journal of Science of Hiroshima University. Series A (Mathematics, Physics, Chemistry); Series A-I. (Mathematics) (Hiroshima)

J. Soviet Math.
Journal of Soviet Mathematics (New York). Translation of (1) Itogi Nauki—Seriya Matematika (Progress in Science—Mathematical Series); (2) Problemy Matematicheskogo Analiza (Problems in Mathematical Analysis); (3) Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov.

J. Symbolic Logic
The Journal of Symbolic Logic (New Brunswick)

Japan. J. Math.
Japanese Journal of Mathematics (Tokyo)

Jber. Deutsch. Math. Verein. (Jber. D.M.V.)
Jahresbericht der Deutschen Mathematiker Vereinigung (Stuttgart)

Kibernetika (Kiev)
Otdelenie Matematiki, Mekhaniki i Kibernetiki Akademii Nauk Ukrainskoĭ SSR. Kibernetika (Kiev). Translated as Cybernetics

Kôdai Math. Sem. Rep.
Kôdai Mathematical Seminar Reports (Tokyo)

Linear Algebra and Appl.
Linear Algebra and Its Applications (New York)

Linear and Multilinear Algebra.
Linear and Multilinear Algebra (New York)

Mat. Sb.
Matematicheskiĭ Sbornik (Recueil Mathématique). Akademiya Nauk SSSR (Moscow). Translated as Math. USSR-Sb.

Mat. Tidsskr. A
Matematisk Tidsskrift. A (Copenhagen)

Mat. Zametki
Matematicheskiĭ Zametki.
Akademiya Nauk SSSR (Moscow). Translated as Math. Notes

Math. Ann.
Mathematische Annalen (Berlin-Göttingen-Heidelberg)

Math. Comp.
Mathematics of Computation (Providence). Formerly Math. Tables Aids Comput.

Math. J. Okayama Univ.
Mathematical Journal of Okayama University (Okayama, Japan)

Math. Japonicae
Mathematica Japonicae (Osaka)

Math. Nachr.
Mathematische Nachrichten (Berlin)

Math. Notes
Mathematical Notes of the Academy of Sciences of the USSR (New York). Translation of Mat. Zametki

Math. Rev.
Mathematical Reviews (Ann Arbor)

Math. Scand.
Mathematica Scandinavica (Copenhagen)

Math. Student
The Mathematical Student (Madras)

Math. Tables Aids Comput. (MTAC)
Mathematical Tables and Other Aids to Computation (Washington). Name changed to Mathematics of Computation in 1960 (vol. 14ff.)

Math. USSR-Izv.
Mathematics of the USSR-Izvestiya (Providence). Translation of Izv. Akad. Nauk SSSR

Math. USSR-Sb.
Mathematics of the USSR-Sbornik (Providence). Translation of Mat. Sb.

Math. Z.
Mathematische Zeitschrift (Berlin-Göttingen-Heidelberg)

Mathematika
Mathematika, A Journal of Pure and Applied Mathematics (London)

Meed. Lunds Univ. Mat. Sem.
Meddelanden från Lunds Universitets Matematiska Seminarium = Communications du Séminaire Mathématique de l'Université de Lund (Lund)

Mem. Amer. Math. Soc.
Memoirs of the American Mathematical Society (Providence)

Mem. Coll. Sci. Univ. Kyôto
Memoirs of the College of Science, University of Kyôto. Series A (Kyoto)

Mem. Fac. Sci. Kyushu Univ.
Memoirs of the Faculty of Science, Kyushu University. Series A. Mathematics (Fukuoka, Japan)

Mémor. Sci. Math.
Mémorial des Sciences Mathématiques (Paris)

Michigan Math. J.
The Michigan Mathematical Journal (Ann Arbor)

Mitt. Math. Ges. Hamburg
Mitteilungen der Mathematischen Gesellschaft in Hamburg (Hamburg)

Monatsh. Math. Phys.
Monatschefte für Mathematik und Physik (Vienna)

Monatsh. Math.
Monatshefte für Mathematik (Vienna)

Monograf. Mat.
Monografje Matematyczne (Warsaw)

Moscow Univ. Math. Bull.
Moscow University Mathematics Bulletin (New York). Translation of the mathematics section of Vestnik Moskov. Univ., Ser. I, Mat. Mekh.

Nachr. Akad. Wiss. Göttingen
Nachrichten der Akademie der Wissenschaften in Göttingen. Math.-Phys. Kl. (Göttingen)

Nagoya Math. J.
Nagoya Mathematical Journal (Nagoya)

Nederl. Akad. Wetensch. Proc.
Koninklijke Nederlandse Akademie van Wetenschappen, Proceedings. Series A. Mathematical Sciences (Amsterdam) = Indag. Math., Proc. Acad. Amsterdam

Nieuw Arch. Wisk.
Nieuw Archief voor Wiskunde (Groningen)

Numerische Math.
Numerische Mathematik (Berlin-Göttingen-Heidelberg)

Nuovo Cimento
Il Nuovo Cimento (Bologna)

Osaka J. Math.
Osaka Journal of Mathematics (Osaka)

Osaka Math. J.
Osaka Mathematical Journal (Osaka)

Pacific J. Math.
Pacific Journal of Mathematics (Berkeley)

Philos. Trans. Roy. Soc. London
Philosophical Transactions of the Royal Society of London. Series A (London)

Phys. Rev.
The Physical Review (New York)

Portugal. Math.
Portugaliae Mathematica (Lisbon)

Prikl. Mat. Mekh.
Adademiya Nauk SSSR. Otdelenie Tekhnicheskikh Nauk. Institut Mekhaniki Prikladnaya Matematika i Mekhanika (Moscow). Translated as J. Appl. Mat. Mech.

Proc. Acad. Amsterdam
= Nederl. Akad. Wetensch. Proc.

Proc. Amer. Math. Soc.
Proceedings of the American Mathematical Society (Providence)

Proc. Cambridge Philos. Soc.
Proceedings of the Cambridge Philosophical Society (Cambridge)

Proc. Imp. Acad. Tokyo
Proceedings of the Imperial Academy (Tokyo)

Proc. Japan Acad.
Proceedings of the Japan Academy (Tokyo)

Proc. London Math. Soc.
Proceedings of the London Mathematical Society (London)

Proc. Nat. Acad. Sci. US
Proceedings of the National Academy of Sciences of the United States of America (Washington)

Proc. Phys.-Math. Soc. Japan
Proceedings of the Physico-Mathematical Society of Japan (Tokyo)

Proc. Roy. Soc. London
Proceedings of the Royal Society of London. Series A (London)

Proc. Steklov Inst. Math.
Proceedings of the Steklov Institute of Mathematics (Providence). Translation of Trudy Mat. Inst. Steklov.

Prog. Theoret. Phys.
Progress of Theoretical Physics (Kyoto)

Publ. Inst. Math.
Publications de l'Institut Mathématique (Belgrade)

Publ. Inst. Math. Univ. Strasbourg
Publications de l'Institut de Mathématiques de l'Université de Strasbourg (Strasbourg)

Publ. Math. Inst. HES
Publications Mathématiques de l'Institut des Hautes Etudes Scientifiques (Paris)

Publ. Res. Inst. Math. Sci.
Publications of the Research Institute for Mathematical Sciences (Kyoto)

Quart. Appl. Math.
Quarterly of Applied Mathematics (Providence)

Quart. J. Math.
The Quarterly Journal of Mathematics, Oxford. Second Series (Oxford)

Quart. J. Mech. Appl. Math.
The Quarterly Journal of Mechanics and Applied Mathematics (Oxford)

Rend. Circ. Mat. Palermo
Rendiconti del Circolo Matematico de Palermo (Palermo)

Rend. Sem. Mat. Univ. Padova
Rendiconti del Seminario Matematico dell'Universitá di Padova (Padua)

Rev. Mat. Hisp. Amer.
Revista Matemática Hispaño-Americana (Madrid)

Rev. Mod. Phys.
Reviews of Modern Physics (New York)

Rev. Un. Mat. Argentina
Revista de la Unión Matemática Argentina (Buenos Aires)

Rev. Univ. Tucumán
Revista Universidad Nacional de Tucumán, Facultad de Ciencias Exactas y Tecnologia. Serie A. Matemáticas y Fisica Teorica (Tucumán)

Roczniki Polsk. Towar. Mat.
Roczniki Polskiego Towarzystwa Matematycznego. Serja I. Prace Matematyczne (Krakow)

Rozprawy Mat.
Rozprawy Matematyczne, Polska Akademia Nauk, Instytut Matematyczny (Warsaw)

Russian Math. Surveys.
Russian Mathematical Surveys (London). Translation of Uspekhi Mat. Nauk

Sammul. Göschen
Sammulung Göschen (Leipzig)

Sankhyā
Sankhyā, The Indian Journal of Statistics. Series A and Series B (Calcutta)

S.-B. Berlin. Math. Ges.
Sitzungsberichte der Berliner Mathematischen Gesellschaft (Berlin)

S.-B. Deutsch. Akad. Wiss. Berlin
Sitzungsberichte der Deutschen Akademie der Wissenschaften zu Berlin, Mathematisch-Naturwissenschaftliche Klasse (Berlin)

S.-B. Heidelberger Akad. Wiss.
Sitzungsberichte der Heidelberger Akademie der Wissenschaften, Mathematisch-Naturwissenschaftliche Klasse (Heidelberg)

S.-B. Math.-Nat. Kl. Bayer. Akad. Wiss.
Sitzungsberichte der Mathematisch-Naturwissenschaflichen Klasse der Bayerischen Akademie der Wissenschaften zu München (Munich)

S.-B. Öster. Akad. Wiss.
Sitzungsberichte der Österreichische Akademie der Wissenschaften (Vienna)

S.-B. Phys.-Med. Soz. Erlangen
Sitzungsberichte der Physikalisch-Medizinischen Sozietät zu Erlangen (Erlangen)

S.-B. Preuss. Akad. Wiss.
Sitzungsberichte der Preussischen Akademie der Wissenschaften. Physikalisch-Mathematische Klasse (Berlin)

Schr. Math. Inst. u. Inst. Angew. Math. Univ. Berlin
Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin (Berlin)

Schr. Math. Inst. Univ. Münster
Schriftenreihe des Mathematischen Instituts der Universität Münster (Münster)

Sci. Papers Coll. Gen. Ed. Univ. Tokyo
Scientific Papers of the College of General Education, University of Tokyo (Tokyo)

Sci. Rep. Tokyo Kyoiku Daigaku
Science Reports of the Tokyo Kyoiku Daigaku. Section A (Tokyo)

Scripta Math.
Scripta Mathematica. A Quarterly Journal devoted to the Philosophy, History, and Expository Treatment of Mathematics (New York)

Sém. Bourbaki
Séminaire Bourbaki (Paris)

SIAM J. Appl. Math.
SIAM Journal of Applied Mathematics. A Publication of the Society for Industrial and Applied Mathematics (Philadelphia)

SIAM J. Comput.
SIAM Journal on Computing (Philadelphia)

SIAM J. Control
SIAM Journal on Control (Philadelphia)

SIAM J. Math. Anal.
SIAM Journal on Mathematical Analysis (Philadelphia)

SIAM J. Numer. Anal.
SIAM Journal on Numerical Analysis (Philadelphia)

SIAM Rev.
SIAM Review (Philadelphia)

Siberian Math. J.
Siberian Mathematical Journal (New York). Translation of Sibirsk. Mat. Zh.

Sibirsk. Mat. Zh.
Akademiya Nauk SSSR. Sibirskoe Otdelenie. Sibirskiĭ Matematicheskiĭ Zhurnal (Moscow). Translated as Siberian Math. J.

Skr. Norske Vid. Akad. Oslo
Skrifter Utgitt av det Norske Videnskaps-Akademii Oslo. Matematisk-Naturvidenskapelig Klasse (Oslo)

Soviet Math. Dokl.
Soviet Mathematics, Doklady (Providence). Translation of mathematical section of Dokl. Akad. Nauk SSSR

SRI J.
Stanford Research Institute Journal (Menlo Park)

Studia Math.
Studia Mathematica. (Wrocław)

Sûbutu-kaisi
Nihon Sûgaku-buturi-gakkai Kaisi (Tokyo)

Sûgaku
Sûgaku, Mathematical Society of Japan (Tokyo)

Summa Brasil. Math.
Summa Brasiliensis Mathematicae (Rio de Janeiro)

Tensor
Tensor (Chigasaki, Japan)

Teor. Veroyatnost. i Primenen.
Teoriya Veroyatnosteĭ i ee Primenenie. Akademiya Nauk SSSR (Moscow). Translated as Theor. Prob. Appl.

Theor. Prob. Appl.
Theory of Probability and Its Applications. Society for Industrial and Applied Mathematics. English translation of Teor. Veroyatnost. i Primenen. (Philadelphia)

Tôhoku Math. J.
The Tôhoku Mathematical Journal (Sendai, Japan)

Tôhoku-rihô
Tôhoku Teikokudaigaku Rikahôkoku (Sendai, Japan)

Topology
Topology. An International Journal of Mathematics (Oxford)

Trans. Amer. Math. Soc.
Transactions of the American Mathematical Society (Providence)

Trans. Moscow Math. Soc.
Transactions of the Moscow Mathematical Society (Providence). Translation of Trudy Moskov. Mat. Obshch.

Trudy Mat. Inst. Steklov.
Trudy Matematicheskogo Instituta im. V. A.

Steklova. Akademiya Nauk SSSR (Moscow-Leningrad). Translated as Proc. Steklov Inst. Math.

Trudy Moskov. Obshch.
Trudy Moskovskogo Matematicheskogo Obshchestva (Moscow). Translated as Trans. Moscow Math. Soc.

Tsukuba J. Math.
Tsukuba Journal of Mathematics. Univ. Tsukuba (Ibaraki, Japan)

Ukrain. Mat. Zh.
Akademiya Nauk Ukrainskoĭ SSR. Institut Matematiki. Ukrainskiĭ Matematicheskiĭ Zhurnal (Kiev). Translated as Ukrainian Math. J.

Ukrainian Math. J.
Ukrainian Mathematical Journal (New York). Translation of Ukrain. Mat. Zh.

Uspekhi Mat. Nauk
Uspekhi Matematicheskikh Nauk (Moscow-Leningrad). Translated as Russian Math. Surveys

Vestnik Moskov. Univ.
Vestnik Moskovskogo Universiteta. I, Matematika i Mekhanika (Moscow). Mathematical section translated as Moscow Univ. Math. Bull.

Vierteljschr. Naturf. Ges. Zürich
Vierteljahrsschrifte der Naturforschenden Gesellschaft in Zürich (Zurich)

Z. Angew. Math. Mech. (Z.A.M.M.)
Zeitschrift für Angewandte Mathematik und Mechanik, Ingenieurwissenschaftliche Forschungsarbeiten (Berlin)

Z. Angew. Math. Phys. (Z.A.M.P.)
Zeitschrift für Angewandte Mathematik und Physik (Basel)

Z. Wahrscheinlichkeitstheorie
Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete (Berlin)

Zbl. Angew. Math.
Zentralblatt für Angewandte Mathematik (Berlin)

Zbl. Math.
Zentralblatt für Mathematik und ihre Grenzgebiete (Berlin-Göttingen-Heidelberg)

Zh. Èksper. Teoret. Fiz.
Zhurnal Èksperimental'noĭ i Teoreticheskoĭ Fiziki (Moscow)

Publishers

Academic Press
Academic Press Inc., New York-London

Addison-Wesley
Addison-Wesley Publishing Company, Inc., Reading (Massachusetts)-Menlo Park (California)-London-Don Mills (Ontario)

Akadémiai Kiadó
A kiadásért felös: az Adadémiai Kiadó igazatója (Publishing House of the Hungarian Academy of Sciences), Budapest

Akademie-Verlag
Berlin

Akademische Verlag.
Akademische Verlagsgesellschaft, Leipzig

Allen
W. H. Allen & Co. Ltd., London

Allen & Unwin
Allen & Unwin, Inc., Winchester (Massachusetts)

Allyn & Bacon
Allyn & Bacon, Inc., Newton (Massachusetts)

Almqvist and Wiksell
Almqvist och Wiksell Förlag, Stockholm

Asakura
Asakura-syoten, Tokyo

Aschelhoug
H. Aschelhoug and Company, Oslo

Baihûkan
Tokyo

Benjamin
W. A. Benjamin, Inc., New York-London

Birkhäuser
Birkhäuser Verlag, Basel-Stuttgart

Blackie
Blackie & Son Ltd., London-Glasgow

Cambridge Univ. Press
Cambridge University Press, London-New York

Chapman & Hall
Chapman & Hall Ltd., London

Chelsea
Chelsea Publishing Company, New York

Clarendon Press
Oxford University Press, Oxford

Cremona
Edizioni Cremonese, Rome

de Gruyter
Walter de Gruyter and Company, Berlin

Dekker
Marcel Dekker, Inc., New York

Deutscher Verlag der Wiss.
Deutscher Verlag der Wissenschaften, Berlin

Dover
Dover Publications, Inc., New York

Dunod
Dunod, Editeur, Paris

Elsevier
Elsevier Publishing Company, Amsterdam-London-New York

Fizmatgiz
Gosudarstvennoe Izdatel'stvo Fiziko-Matematicheskoĭ Literatury, Moscow

Gauthier-Villars
Gauthier-Villars & C^{ie}, Editeur, Paris

Ginn
Ginn and Company, Waltham (Massachusetts)-Toronto-London

Gordon & Breach
Gordon & Breach, Science Publishers Ltd., London

Goztekhizdat
Gosudarstvennoe Izdatel'stvo Tekhniko-Teoreticheskoĭ Literatury, Moscow

Griffin
Charles Griffin and Company Ltd., London

Hafner
Hafner Publishing Company, New York

Harper & Row
Harper & Row Publishers, New York-Evanston-London

Hermann
Hermann & C^{ie}, Paris

Hirokawa
Hirokawa-syoten, Tokyo

Hirzel
Verlag von S. Hirzel, Leipzig

Holden-Day
Holden-Day, Inc., San Francisco-London-Amsterdam

Holt, Rinehart and Winston
Holt, Rinehart and Winston, Inc., New York-Chicago-San Francisco-Toronto-London

Interscience
Interscience Publishers, Inc., New York-London

Iwanami
Iwanami Shoten, Tokyo

Kawade
Kawade-syobô, Tokyo

Kinokoniya
Kinokoniya Company, Tokyo

Kyôritu
Kyôritu-syuppan, Tokyo

Lippincott
J. B. Lippincott Company, Philadelphia

Longman
Longman Group, Ltd., Harlow (Essex)

Longmans, Green
Longmans, Green and Company, Ltd., London-New York-Toronto-Bombay-Calcutta-Madras

Macmillan
The Macmillan Company, New York-London

Maki
Maki-syoten, Tokyo

Maruzen
Maruzen Company Ltd., Tokyo

Masson
Masson et C^{ie}, Paris

Math-Sci Press
Math-Sci Press, Brookline (Massachusetts)

McGraw-Hill
McGraw-Hill Book Company, Inc., New York-London-Toronto

Methuen
Methuen and Company Ltd., London

MIT Press
The MIT Press, Cambridge (Massachusetts)-London

Nauka
Izdatel'stvo Nauka, Moscow

Noordhoff
P. Noordhoff Ltd., Groningen

North-Holland
North-Holland Publishing Company, Amsterdam

Oldenbourg
Verlag von R. Oldenbourg, Munich-Vienna

Oliver & Boyd
Oliver & Boyd Ltd., Edinburgh-London

Oxford Univ. Press
Oxford University Press, London-New York

Pergamon
Pergamon Press, Oxford-London-Edinburgh-New York-Paris-Frankfurt

Polish Scientific Publishers
Państwowe Wydawnictwo Naukowe, Warsaw

Prentice-Hall
Prentice-Hall, Inc., Englewood Cliffs (New Jersey)

Princeton Univ. Press
Princeton University Press, Princeton

Random House
Random House, Inc., New York

Sibundô
Tokyo

Springer
Springer-Verlag, Berlin (-Göttingen)-Heidelberg-New York

Teubner
B. G. Teubner Verlagsgesellschaft, Leipzig-Stuttgart

Tôkai
Tôkai-syobô, Tokyo

Tokyo-tosyo
Tokyo

Tokyo Univ. Press
Tokyo University Press, Tokyo

Ungar
Frederick Ungar Publishing Company, New York

Univ. of Tokyo Press
University of Tokyo Press, Tokyo

Utida-rôkakuho
Tokyo

Van Nostrand
D. Van Nostrand Company, Inc., Toronto-New York-London

Vandenhoeck & Ruprecht
Göttingen

Veit
Verlag von Veit & Company, Leipzig

Vieweg
Friedr Vieweg und Sohn Verlagsgesellschaft mbH, Wiesbaden

Wiley
Wiley & Sons, Inc., New York-London

Wiley-Interscience
Wiley & Sons, Inc., New York-London

Zanichelli
Nicola Zanichelli Editore, Bologna

Special Notation

This list contains the notation commonly and frequently used throughout this work. The symbol * means that the same notation is used with more than one meaning. For more detailed definitions or further properties of the notation, see the articles cited.

Notation	Example	Definition	Article and Section
I. Logic			
$\forall$	$\forall x F(x)$	Universal quantifier (for all x, $F(x)$ holds)	411B, C
$\exists$	$\exists x F(x)$	Existential quantifier (there exists an x such that $F(x)$ holds)	411B, C
$\wedge$, &	$A \wedge B$, $A \& B$	Conjunction, logical product (A and B)	411B*
$\vee$	$A \vee B$	Disjunction, logical sum (A or B)	411B*
$\neg$	$\neg A$	Negation (not A)	411B
$\rightarrow$, $\supset$, $\Rightarrow$	$A \rightarrow B$, $A \Rightarrow B$	Implication (A implies B)	411B*
$\leftrightarrow$, $\Leftrightarrow$, $\rightleftarrows$	$A \Leftrightarrow B$	Equivalence (A and B are logically equivalent)	411B
II. Sets			
$\in$	$x \in X$	Membership (element x is a member of the set X)	381A
$\notin$	$x \notin X$	Nonmembership (element x is not a member of the set X)	381A
$\subset$	$A \subset B$	Inclusion (A is a subset of B)	381A
$\not\subset$	$A \not\subset B$	Noninclusion (A is not a subset of B)	381A
$\subsetneqq$	$A \subsetneqq B$	Proper inclusion (A is a proper subset of B)	381A
$\varnothing$		Empty set	381A
$\cup$, $\bigcup$	$A \cup B$, $\bigcup A_\lambda$	Union, join	381B, D*
$\cap$, $\bigcap$	$A \cap B$, $\bigcap A_\lambda$	Intersection, meet	381B, D*
^{c}, C	A^c, $C(A)$	Complement (of a set A)	381B
$-$, $\setminus$	$A - B$, $A \setminus B$	Difference ($A - B = A \cap B^c$)	381B
$\times$	$A \times B$	Cartesian product (of A and B)	381B*
R, $\sim$	xRy, $x \sim y$	Equivalence relation (for two elements x, y)	135A*
$/$	A/R	Quotient set (set of equivalence classes of A with respect to an equivalence relation R)	135B*
Π	$\Pi_\lambda A_\lambda$	Cartesian product (of the A_λ)	381E
Σ, $\coprod$	ΣA_λ, $\coprod A_\lambda$	Direct sum (of the A_λ)	381E
$\mathfrak{B}$	$\mathfrak{B}(A)$	Power set (set of all subsets of A)	381E
	B^A	Set of all mappings from A to B	381C
$\{\mid\}$	$\{x \mid P(x)\}$	Set of all elements x with the property $P(x)$	381A

Notation	Example	Definition	Article and Section
$\{\ \}$	$\{a_\lambda\}_{\lambda\in\Lambda}$	Family with index set Λ	165D
	$\{a_n\}$	Sequence (of numbers, points, functions, or sets)	165D
$\bar{\bar{\ }}$, $\lvert\ \rvert$, #	$\bar{\bar{X}}$, $\lvert X\rvert$, $\# X$	Cardinal number (of the set X)	49A*
$\aleph$	$\aleph_\beta$	Aleph (transfinite cardinal)	49E
$\rightarrow$	$f:X\rightarrow Y$	Mapping (f from X to Y)	381C*
$\mapsto$	$f:X\mapsto Y$	Mapping (where $f(X)=Y$, but not in the present volumes)	381C
1, id	1_A, id_A	Identity mapping (identity function)	381C
c, χ	$c_X(x)$, $\chi_X(x)$	Characteristic function (representing function)	381C
$\vert$	$f\vert A$	Restriction (of a mapping f to A)	381C*
$\circ$	$g\circ f$	Composite (of mappings f and g)	381C
lim sup, $\overline{\lim}$	$\limsup A_n$	Superior limit (of the sequence of sets A_n)	270C*
lim inf, $\underline{\lim}$	$\liminf A_n$	Inferior limit (of the sequence of sets A_n)	270C*
lim	$\lim A_n$	Limit (of the sequence of sets A_n)	270C*
$\varinjlim$	$\varinjlim A_\lambda$	Inductive limit (of A_λ)	210B
$\varprojlim$	$\varprojlim A_\lambda$	Projective limit (of A_λ)	210B
III. Order			
(,)	(a,b)	Open interval $\{x\vert a<x<b\}$	355C*
[,]	$[a,b]$	Closed interval $\{x\vert a\leqslant x\leqslant b\}$	355C*
(,]	$(a,b]$	Half-open-interval $\{x\vert a<x\leqslant b\}$	355C
[,)	$[a,b)$	Half-open interval $\{x\vert a\leqslant x<b\}$	355C
max	$\max A$	Maximum (of A)	311B
min	$\min A$	Minimum (of A)	311B
sup	$\sup A$	Supremum, least upper bound (of A)	311B
inf	$\inf A$	Infimum, greatest lower bound (of A)	311B
$\ll$	$a\ll b$	Very large (b is very large compared to a)	
$\cup$, $\vee$	$a\cup b$, $a\vee b$	Join of a, b in an ordered set	243A*
$\cap$, $\wedge$	$a\cap b$, $a\wedge b$	Meet of a, b in an ordered set	243A*
IV. Algebra			
mod	$a\equiv b\ (\mathrm{mod}\ n)$	Modulo (a and b are congruent modulo n)	297G
$\vert$	$a\vert b$	Divisibility (a divides b)	297A*
$\nmid$	$a\nmid b$	Nondivisibility (a does not divide b)	297A
det, $\lvert\ \rvert$	$\det A$, $\lvert A\rvert$	Determinant (of a square matrix A)	103A*

Notation	Example	Definition	Article and Section
tr, Sp	tr A, Sp A	Trace (of a square matrix A)	269F
^{t}, ^{T}, $'$	^tA; A^t, A^T, A'	Transpose (of a matrix A)	269B
I	I_n	Unit matrix (of degree n)	269A
E_{ij}		Matrix unit (matrix whose (i,j)-component is 1 and all others are 0)	269B
$\otimes$	$A \otimes B$	Kronecker product (of two matrices A and B)	269C*
$\cong$	$M \cong N$	Isomorphism (of two algebraic systems M and N)	256B
/	M/N	Quotient space (of an algebraic system M by N)	256F*
dim	dim M	Dimension (of a linear space, etc.)	256C
Im	Im f	Image (of a mapping f)	277E*
Ker	Ker f	Kernel (of a mapping f)	277E
Coim	Coim f	Coimage (of a mapping f)	277E
Coker	Coker f	Cokernel (of a mapping f)	277E
δ_{ij}, δ_i^j		Kronecker delta ($\delta_{ii}=1$ and $\delta_{ij}=0$ for $i \neq j$)	269A
(,), $\cdot$	$(\mathbf{a}, \mathbf{b})$, $\mathbf{a} \cdot \mathbf{b}$	Inner product (of two vectors **a** and **b**)	442B*
[,], $\times$	$[\mathbf{a}, \mathbf{b}]$, $\mathbf{a} \times \mathbf{b}$	Vector product (of two 3-dimensional vectors **a** and **b**)	442C*
$\otimes$	$M \otimes N$	Tensor product (of two modules M and N)	277J, 256I*
Hom	Hom(M, N)	Set of all homomorphisms (from M to N)	277B
Hom_A	$\mathrm{Hom}_A(M, N)$	Set of all A-homomorphisms (of an A-module M to an A-module N)	277E
Tor	$\mathrm{Tor}_n(M, N)$	Torsion product (of M, N)	200D
Ext	$\mathrm{Ext}^n(M, N)$	Extension (of M, N)	200G
$\wedge$, $\wedge^p$	$\wedge M$, $\wedge^p M$	Exterior algebra (of a linear space M), pth exterior product (of M)	256O
V. Algebraic Systems			
$\mathbf{N}$		Set of all natural numbers	294A
$\mathbf{Z}$		Set of all rational integers	294A
$\mathbf{Z}_m$		$\mathbf{Z}/m\mathbf{Z}$ (set of all residue classes modulo m)	297G*
$\mathbf{Q}$		Set of all rational numbers	294A
$\mathbf{R}$		Set of all real numbers	294A
$\mathbf{C}$		Set of all complex numbers	294A
$\mathbf{H}$		Set of all quaternions	29B
$GF(q)$, $\mathbf{F}_q$		Finite field (with q elements)	149M

Notation	Example	Definition	Article and Section
$\mathbf{Q}_p$		p-adic number field (p is a prime)	439F
$\mathbf{Z}_p$		Ring of p-adic integers	439F
[ ]	$k[x_1,\ldots,x_n]$	Polynomial ring (of variables $x_1, \ldots, x_n$ with coefficients in k)	369A
( )	$k(x_1,\ldots,x_n)$	Field extension (of k by $x_1, \ldots, x_n$)	149D
[[ ]], { }	$k[[x_1,\ldots,x_n]]$	Formal power series ring (with coefficients in k).	370A
		Note: The symbols **N**, **Z**, **Q**, **R**, **C**, and **H** stand for sets, each with its own natural mathematical structure	
VI. Groups			
GL	$GL(V)$, $GL(n, K)$	General linear group (over V, or over K of degree n)	60B
SL	$SL(n, K)$	Special linear group (over K of degree n)	60B
PSL	$PSL(n, K)$	Projective special linear group (over K of degree n)	60B
U	$U(n)$	Unitary group (of degree n)	60F
SU	$SU(n)$	Special unitary group (of degree n)	60F
O	$O(n)$	Orthogonal group (of degree n)	60I
SO	$SO(n)$	Special orthogonal group, rotation group (of degree n)	60I
$Spin$	$Spin(n)$	Spinor group (of degree n)	61D
Sp	$Sp(n)$	Symplectic group (of degree n)	60L
[For $PGL(n, K)$, $LF(n, K)$, $PU(n)$, $Sp(n)$, $PSp(n, K)$ → 60 Classical Groups]			
VII. Topology (Convergence)			
$\rightarrow$	$a_n \rightarrow a$	Convergence (sequence a_n converges to a)	87B, E*
$\downarrow$, $\searrow$	$a_n \downarrow a$, $a_n \searrow a$	Convergence monotonically decreasing	87B
$\uparrow$, $\nearrow$	$a_n \uparrow a$, $a_n \nearrow a$	Convergence monotonically increasing	87B
lim	$\lim a_n$	Limit (of a sequence a_n)	87B, E*
lim sup, $\overline{\lim}$	$\limsup a_n$, $\overline{\lim}\, a_n$	Superior limit (of a sequence a_n)	87C*
lim inf, $\underline{\lim}$	$\liminf a_n$, $\underline{\lim}\, a_n$	Inferior limit (of a sequence a_n)	87C*
^{a}, $^-$, Cl	E^a, $\bar{E}$, Cl E	Closure (of a set E)	425B
^{i}, $^\circ$, Int	E^i, E°, Int E	Interior (of a set E)	425B
ρ, d	$\rho(x, y)$, $d(x, y)$	Distance (between two points x and y)	273B*
$\Vert\ \Vert$	$\Vert x \Vert$	Norm (of x)	37B
l.i.m.	l.i.m. f_n	Limit in the mean (of a sequence f_n)	168B

Notation	Example	Definition	Article and Section
s-lim	s-lim x_n	Strong limit (of a sequence x_n)	37B
w-lim	w-lim x_n	Weak limit (of a sequence x_n)	37E
$\simeq$	$f \simeq g$	Homotopy (of two mappings f and g)	202B
$\approx$	$X \approx Y$	Homeomorphism (of two topological spaces X and Y)	425G

VIII. Geometry and Algebraic Topology

Notation	Example	Definition	Article and Section
E^n		Euclidean space (of dimension n)	140
P^n		Projective space (of dimension n)	343B
S^n		Spherical surface (of dimension n)	140
$\mathbf{T}^n$		Torus (of dimension n)	422E
H^n	$H^n(X, A)$	n-dimensional cohomology group (of X with coefficients in A)	201H
H_n	$H_n(X, A)$	n-dimensional homology group (of X with coefficients in A)	201G
	$H_n(C)$	(of chain complex C)	201B
π_n	$\pi_n(X)$	n-dimensional homotopy group (of X)	202J, 170
∂	∂C	Boundary (of C)	201B
δ	δf	Coboundary (of f)	201H*
Sq	$Sq^i x$	Streenrod square (of x)	64B
$\mathscr{P}$	$\mathscr{P}_p^r(x)$	Steenrod pth power (of x)	64B
$\smile$	$z_1 \smile z_2$	Cup product (of z_1 and z_2)	201I
$\frown$	$z_1 \frown z_2$	Cap product (of z_1 and z_2)	201K
$\wedge$	$\omega \wedge \eta$	Exterior product (of two differential forms ω and η)	105Q*
d	$d\omega$	Exterior derivative (of a differential form ω)	105Q
grad	grad φ	Gradient (of a function φ)	442D
rot	rot $\mathbf{u}$	Rotation (of a vector $\mathbf{u}$)	442D
div	div $\mathbf{u}$	Divergence (of a vector $\mathbf{u}$)	442D
Δ	$\Delta\varphi$	Laplacian (of a function φ)	323A
$\square$	$\square\varphi$	d'Alembertian (of a function φ)	130A
D	$D\varphi$	Differential operator	112A*
$\dfrac{D(u_1, \dots, u_n)}{D(x_1, \dots, x_n)}, \left\lvert\dfrac{\partial u_i}{\partial x_j}\right\rvert, \det\left(\dfrac{\partial u_i}{\partial x_j}\right)$		Jacobian determinant (of $(u_1, \dots, u_n)$ with respect to $(x_1, \dots, x_n)$)	208B
$\dfrac{\partial(u_1, \dots, u_n)}{\partial(x_1, \dots, x_n)}, \left(\dfrac{\partial u_i}{\partial x_j}\right)$		Jacobian matrix (of $(u_1, \dots, u_n)$ with respect to $(x_1, \dots, x_n)$)	208B

IX. Function Spaces

Notation	Example	Definition	Article and Section
C	$C(\Omega)$	Space of continuous functions (on Ω)	168B(1)

Notation	Example	Definition	Article and Section
L_p	$L_p(\Omega)$, $L_p(a,b)$	Space of functions such that $\|f\|^p$ is integrable on Ω	168B(2)
C^l	$C^l(L)(1 \leqslant l \leqslant \infty)$	Space of functions of class C^l	168B(9)
$\mathscr{D}$	$\mathscr{D}(\Omega)$	Space of C^∞ functions with compact support	168B(13)
$\mathscr{E}$	$\mathscr{E}(\Omega)$	Space of C^∞ functions	168B(13)

[For $\mathscr{A}(\Omega)$, $A(\Omega)$, $A_p(\Omega)$, $\mathscr{B}(\Omega)(=\mathrm{D}_{L^\infty}(\Omega))$, $BMO(\mathbf{R}^n)$, $BV(\Omega)$, c, C, $C_0(\Omega)$, $C_\infty(\Omega)$, $C_0^l(\Omega)$, $\mathscr{D}_{L^p}(\Omega)$, $\mathscr{D}_{\{M^p\}}(\Omega)$, $\mathscr{D}_{(M^p)}(\Omega)$, $\mathscr{E}_{\{M^p\}}(\Omega)$, $\mathscr{E}_{(M^p)}(\Omega)$, $H_p(\mathbf{R}^n)$, $H^l(\Omega)$, $H_0^l(\Omega)$, $\Lambda^S(\mathbf{R}^n)$, $\bigcap \lambda(\alpha^{(k)})$, $\sum \lambda^\times(\alpha^{(k)})$, l_p, $L_{(p,q)}(\Omega)$, m, $M(\Omega)$, $\mathscr{O}(\Omega)$, $\mathscr{O}_p(\Omega)$, $\mathscr{S}$, s, $S(\Omega)$, $W_p^l(\Omega)$ → 168 Function Spaces. For $\mathscr{B}(\Omega)$ (Space of Sato hyperfunctions), $\mathscr{D}'(\Omega)$, $\mathscr{E}'(\Omega)$, $\mathscr{O}_c$, $\mathscr{O}_M$, $\mathscr{S}'(\mathbf{R}^n)$ → 125 Distributions and Hyperfunctions]

X. Functions

$\| \ \|$	$\|z\|$	Absolute value (of a complex number z)	74B*
Re	$\mathrm{Re}\, z$	Real part (of a complex number z)	74A
Im	$\mathrm{Im}\, z$	Imaginary part (of a complex number z)	74A*
$\bar{\ }$	$\bar{z}$	Complex conjugate (of a complex number z)	74A
arg	$\arg z$	Argument (of a complex number z)	74C
$[\ ]$	$[\alpha]$	Gauss symbol (greatest integer not exceeding a real number α)	83A
O	$f(x)=O(g(x))$	Landau's notation ($f(x)/g(x)$ is bounded for $x \to \alpha$)	87G
o	$f(x)=o(g(x))$	Landau's notation ($f(x)/g(x)$ tends to 0 for $x \to \alpha$)	87G
$\sim$	$f(x) \sim g(x)$	Infinite or infinitesimal of the same order (for $x \to \alpha$)	87G*
D	$D(T)$	Domain (of an operator T)	251A
R	$R(T)$	Range (of an operator T)	251A
supp	$\mathrm{supp}\, f$	Support (of a function f)	168B(1)
p.v.	$\mathrm{p.v.}\int_a^b f(x)dx$	Cauchy's principal value (of an integral)	216D
Pf	$\mathrm{Pf}\int f(x)dx$	Finite part (of an integral)	125C
δ	$\delta(x)$, δ_x	Dirac's delta function (measure or distribution)	125C*
exp	$\exp x$	Exponential function ($\exp x = e^x$)	113D, 269H
log, Log	$\log x$, $\mathrm{Log}\, x$	Natural logarithmic function and its principal value, respectively	131D, G
$\sin x$, $\cos x$, $\tan x$, $\sec x$, $\mathrm{cosec}\, x$, $\mathrm{cotan}\, x$		Trigonometric functions	131E
$\arcsin x$, $\arccos x$, $\arctan x$		Inverse trigonometric functions	131E
$\mathrm{Arc}\sin x$, $\mathrm{Arc}\cos x$, $\mathrm{Arc}\tan x$		Principal value of inverse trigonometric functions	131E

Notation	Example	Definition	Article and Section
$\sinh x$, $\cosh x$, $\tanh x$		Hyperbolic functions	131F
$\binom{\ }{\ }$, C	$\binom{n}{r}$, ${}_nC_r$	Binomial coefficient, combination	330
P	${}_nP_r$	Permutation	330
!	$n!$	Factorial (of n)	330
φ	$\varphi(n)$	Euler function	295C*
μ	$\mu(n)$	Möbius function	295C
ζ	$\zeta(z)$	Riemann zeta function	450B*
J_ν	$J_\nu(z)$	Bessel function of the first kind	39B
Γ	$\Gamma(x)$	Gamma function	174A
B	$B(x,y)$	Beta function	174C
F	$F(\alpha,\beta,\gamma;z)$	Gauss's hypergeometric function	206A
P	$P\begin{Bmatrix} a & b & c \\ \lambda & \mu & \nu \\ \lambda' & \mu' & \nu' \end{Bmatrix} x$	Riemann's P function	253B
Li	Li(x)	Logarithmic integral	167D
XI. Probability			
P, Pr	$P(E)$, Pr(ε)	Probability (of an event)	342B*
E	$E(X)$	Mean or expectation (of a random variable X)	342C
V, σ^2	$V(X)$, $\sigma^2(X)$	Variance (of a random variable X)	342C
ρ	$\rho(X,Y)$	Correlation coefficient (of two random variables X and Y)	342C*
$P(\mid)$	$P(E\mid F)$	Conditional probability (of an event E under the condition F)	342E
$E(\mid)$	$E(X\mid Y)$	Conditional mean (of a random variable X under the condition Y)	342E
N	$N(m,\sigma^2)$	One-dimensional normal distribution (with mean m and variance σ^2)	Appendix A, Table 22
	$N(\boldsymbol{\mu},\Sigma)$	Multidimensional normal distribution (with mean vector $\boldsymbol{\mu}$ and variance matrix Σ)	Appendix A, Table 22
P	$P(\lambda)$	Poisson distribution (with parameter λ)	Appendix A, Table 22*

Systematic List of Articles

VI
Euclidean and Projective Geometry

VII
Differential Geometry

VIII
Algebraic Geometry

IX
Topology

X
Analysis

XI
Complex Analysis

XII
Functional Analysis

XIII
Differential, Integral, and Functional Equations

XIV
Special Functions

XV
Numerical Analysis

XVI
Computer Science and Combinatorics

XVII
Probability Theory

XVIII
Statistics

XIX
Mathematical Programming and Operations Research

Alphabetical List of Articles

Contributors to the Second Edition

Abe Eiichi
Akahira Masafumi
Akaike Hirotugu
Akaza Touru
Amari Shun-ichi
Ando Tsuyoshi
Araki Huzihiro
Araki Shôrô
Arimoto Suguru
Asai Akira
Fuji'i'e Tatuo
Fujiki Akira
Fujisaki Genjiro
Fujita Hiroshi
Fujiwara Daisuke
Fujiwara Masahiko
Fukushima Masatoshi
Furuya Shigeru
Hamada Yûsaku
Hasumi Morisuke
Hattori Akio
Hayashi Kiyoshi
Hida Takeyuki
Hijikata Hiroaki
Hirose Ken
Hitotumatu Sin
Huzii Mituaki
Ibaraki Toshihide
Ihara Shin-ichiro
Iitaka Shigeru
Ikawa Mitsuru
Ikebe Teruo
Ikebe Yasuhiko
Ikeda Nobuyuki
Inagaki Nobuo
Inoue Atsushi
Iri Masao
Itaya Nobutoshi
Itô Kiyosi
Itô Seizô
Ito Yuji
Itoh Mitsuhiro
Iwahori Nobuyoshi
Iyanaga Shôkichi
Kadoya Norihiko
Kageyama Sampei
Kamae Teturo
Kaneko Akira
Kaneyuki Soji
Kasahara Kōji
Kataoka Shinji
Kato Junji
Kato Mitsuyoshi
Kawai Takahiro
Kawasaki Tetsuro
Kimura Tosihusa
Kishi Masanori
Kobayashi Shoshichi
Kodama Yukihiro
Komatsu Hikosaburo
Kômura Yukio
Kondo Takeshi
Konishi Yoshio
Konno Hiroshi
Kotani Shinichi
Kubota Yoshihisa
Kumano-go Hajime
Kunita Hiroshi
Kuramoto Yuki
Kuroda Shige-Toshi
Kuroda Tadashi
Kusama Tokitake
Kusano Takaŝi
Kusunoki Yukio
Mabuchi Toshiki
Maeda Shūichiro
Maeda Yoshiaki
Maehara Shôji
Makabe Hajime
Maruyama Masaki
Matsumoto Kikuji
Matsumoto Yukio
Matsumura Mutsuhide
Mitsui Takayoshi
Mizohata Sigeru
Mizutani Tadayoshi
Mori Masatake
Morimoto Akihiko
Morimura Hidenori
Morita Shigeyuki
Morita Yasuo
Motohashi Nobuyoshi
Murasugi Kunio
Nagami Keio
Nagata Masayoshi
Nakagawa Hisao
Nakamura Tokushi
Nakaoka Minoru
Namba Kanji
Namba Makoto
Namikawa Yukihiko
Naruki Isao
Niiro Fumio
Nishikawa Seiki
Nishimura Junichi
Nishitani Kensaku
Nisio Makiko
Noguchi Hiroshi
Nozaki Akihiro
Ôaku Toshinori
Obata Morio
Ochiai Takushiro
Ogiue Koichi
Ohta Masami
Ohya Yujiro
Oikawa Kôtaro
Oka Mutsuo
Okabe Yasunori

Okuno Tadakazu
Omori Hideki
Oshima Toshio
Ozawa Mitsuru
Saito Kyoji
Saito Masahiko
Saito Tosiya
Sakai Fumio
Sakawa Yoshiyuki
Sato Fumitaka
Sato Ken-iti
Sawashima Ikuko
Shibata Katsuyuki
Shibata Keiichi
Shiga Koji
Shiga Tokuzo
Shikata Yoshihiro
Shimada Nobuo
Shimakura Norio
Shimizu Ryoichi
Shioda Tetsuji
Shiohama Katsuhiro
Shiraiwa Kenichi
Shiratani Katsumi
Sibuya Masaaki
Sugie Toru
Sugiura Mitsuo
Sugiura Shigeaki
Suita Nobuyuki
Sumihiro Hideyasu
Suzuki Masuo
Suzuki Mitsuo
Takahashi Tsuneo
Takakuwa Shoichiro
Takami Hideo
Takasu Toru
Takeuchi Kei
Takeuchi Masaru
Takeuti Gaishi
Tamura Itiro
Tanabe Kunio
Tanaka Hiroshi
Tanaka Masatsugu
Tanaka Shunichi
Tanno Shukichi
Tatsumi Tomomasa
Terada Toshiaki
Toda Hideo
Toda Hirosi
Toda Morikazu
Toda Nobushige
Tomiyama Jun
Totoki Haruo
Tugé Tosiyuki
Uchida Fuichi
Uchiyama Saburô
Ueno Kenji
Ukai Seiji
Ura Shoji
Ushijima Teruo
Wakabayashi Isao
Washizu Kyuichiro
Watanabe Kimio
Watanabe Shinzo
Watanabe Takesi
Yamaguchi Masaya
Yamamoto Koichi
Yamamoto Sumiyasu
Yamamoto Tetsuro
Yamazaki Masao
Yamazaki Tadashi
Yanagawa Takashi
Yanagiwara Hiroshi
Yanai Haruo
Yano Tamaki
Yokonuma Takeo
Yoshikawa Atsushi
Yoshida Junzo
Yosida Kôsaku

Contributors to the First Edition

Aizawa Sadakazu
Akao Kazuo
Akizuki Yasuo
Amemiya Ayao
Anzai Hirotada
Aoki Kiyoshi
Araki Huzihiro
Araki Shôrô
Arima Akito
Asaka Tetsuichi
Asano Keizo
Asatani Teruo
Azumaya Gorô
Fujinaka Megumu
Fujita Hiroshi
Fujiwara Daisuke
Fukuda Takeo
Fukutomi Setsuo
Furuya Shigeru
Goto Morikuni
Gotô Motinori
Gotō Yūzo
Hagihara Yusuke
Harada Manabu
Hasimoto Hidenori
Hattori Akio
Hattori Akira
Hayashi Chikio
Hayashi Keiichi
Hayashi Tsuyoshi
Hida Takeyuki
Hidaka Koji
Hirakawa Junkô
Hirano Tomoharu
Hirasawa Yoshikazu
Hirayama Akira
Hirose Ken
Hisatake Masao
Hitotumatu Sin
Hokari Shisanji
Homma Kiyomi
Homma Tatsuo
Honbu Hitoshi
Hong Imsik
Hosokawa Fujitsugu
Hukuhara Masuo
Husimi Kozi
Huzii Mituaki
Ichida Asajiro
Ihara Shin-ichiro
Ihara Yasutaka
Iitaka Shigeru
Ikawa Mitsuru
Ikeda Mineo
Ikeda Nobuyuki
Ikehara Shikao
Imai Isao
Inaba Eiji
Inagaki Takeshi
Inoue Masao
Inui Teturo
Iri Masao
Irie Shoji
Ise Mikio
Iseki Kanesiroo
Ishida Yasushi
Ishii Goro
Isizu Takehiko
Itô Juniti
Itô Kiyosi
Itô Noboru
Itô Seizô
Itô Teiiti
Ito Tsutomu
Itoh Makoto
Iwahashi Ryōsuke
Iwahori Nagayoshi
Iwamura Tsurane
Iwano Masahiro
Iwasawa Kenkichi
Iwata Giiti
Iyanaga Shôkichi
Izumi Shin-ichi
Kaburaki Masaki
Kamae Keiko
Kametani Shunji
Kanbe Tsunekazu
Kanitani Jōyō
Kasahara Koji
Katase Kiyoshi
Kato Tosio
Katuura Sutezo
Kawada Yukiyosi
Kawaguchi Akitsugu
Kawai Saburo
Kawakami Riiti
Kawata Tatuo
Kihara Taro
Kimura Motoo
Kimura Tosihusa
Kinoshita Shin'ichi
Kitagawa Toshio
Kiyasu Zen'ichi
Koba Ziro
Kobayashi Mikio
Kobori Akira
Koda Akira
Kodaira Kunihiko
Koga Yukiyoshi
Komatsu Hikosaburo
Komatu Atuo
Komatu Yûsaku
Kondo Jiro
Kondo Kazuo
Kondô Motokiti
Kondō Ryōji
Kondo Takeshi
Kondo Yōitsu

Kôta Osamu
Kotake Takesi
Kotani Masao
Kozai Yoshihide
Kubo Izumi
Kubo Ryogo
Kubota Tadahiko
Kudo Akio
Kudō Hirokichi
Kudo Tatsuji
Kuga Michio
Kunisawa Kiyonori
Kunita Hiroshi
Kunugi Kinjiro
Kuranishi Masatake
Kuroda Shige-Toshi
Kuroda Sigekatu
Kuroda Sige-Nobu
Kusama Tokitake
Kusano Takasi
Kusunoki Yukio
Maehara Shôji
Maruyama Gisiro
Mashio Shinji
Masuda Kyūya
Masuyama Motosaburo
Matsukuma Ryozai
Matsumoto Kikuji
Matsumoto Makoto
Matsumoto Toshizô
Matsumura Hideyuki
Matsumura Mutsuhide
Matsushima Yozo
Matsuyama Noboru
Matuda Tizuko
Matusita Kameo
Matuzaka Kazuo
Mikami Masao
Mimura Yukio
Minagawa Takizo
Mita Hiroo
Mitome Michio
Mitsui Takayoshi
Miyake Kazuo
Miyasawa Koichi
Miyashita Totaro
Miyazawa Hironari
Miyazima Tatuoki
Mizohata Sigeru
Mizuno Katsuhiko
Mogi Isamu
Mori Shigeo
Moriguti Sigeiti
Morimoto Haruki
Morimoto Seigo
Morimura Hidenori
Morishima Taro
Morita Kiiti
Morita Yuzo
Moriya Mikao
Moriya Tomijiro
Motohashi Nobuyoshi
Motoo Minoru
Murakami Shingo
Muramatu Tosinobu
Murata Tamotsu
Nabeya Seiji
Nagano Tadashi
Nagao Hirosi
Nagashima Takashi
Nagata Jun-iti
Nagata Masayoshi
Nagumo Mitio
Nakai Mitsuru
Nakai Yoshikazu
Nakamori Kanji
Nakamura Koshiro
Nakamura Masahiro
Nakamura Tokushi
Nakanishi Shizu
Nakano Shigeo
Nakaoka Minoru
Nakayama Tadasi
Nakayama Takashi
Namikawa Yukihiko
Nanba Kanji
Narita Masao
Narumi Seimatu
Niiro Fumio
Nikaido Hukukane
Nishi Mieo
Nishimura Toshio
Noguchi Hiroshi
Nomizu Katsumi
Nomoto Hisao
Nomura Yukichi
Nonaka Toshio
Noshiro Kiyoshi
Nozaki Akihiro
Nozaki Yasuo
Ochiai Kiichiro
Ogasawara Tôjirō
Ogawa Junjiro
Ogawara Masami
Ogino Shusaku
Ōhasi Saburo
Ohtsuka Makoto
Oikawa Kotaro
Ôishi Kiyoshi
Oka Syoten
Okamoto Kiyosato
Okamoto Masashi
Okubo Kenjiro
Okubo Tanjiro
Okugawa Kōtaro
Ōnari Setsuo
Ono Akimasa
Ono Katuzi
Ono Shigeru
Ono Takashi
Onoyama Takuji
Oshida Isao

Oshima Nobunori
Osima Masaru
Otsuki Tominosuke
Ozaki Shigeo
Ozawa Mitsuru
Ozeki Hideki
Saito Kinichiro
Saito Tosiya
Saito Yoshihiro
Sakai Eiichi
Sakai Takuzo
Sakamoto Heihachi
Sakata Shôichi
Sasaki Shigeo
Satake Ichiro
Sato Ken-iti
Sato Ryoichiro
Sato Tokui
Sato Yumiko
Sawashima Ikuko
Seimiya Toshio
Seki Setsuya
Shibagaki Wasao
Shiga Kôji
Shimada Nobuo
Shimizu Hideo
Shimizu Tatsujiro
Shimura Goro
Shintani Takuro
Shizuma Ryoji
Shoda Kenjiro
Shono Shigekata
Sibuya Masaaki
Simauti Takakazu
Sirao Tunekiti
Sueoka Seiichi
Suetuna Zyoiti
Sugawara Masahiro
Sugawara Masao
Sugiura Mitsuo
Sugiura Nariaki
Sugiyama Shohei
Suita Nobuyuki
Sumitomo Takeshi
Sunouchi Genichiro
Suzuki Michio
Suzuki Yoshindo
Takada Masaru
Takagi Teiji
Takagi Yutaka
Takahashi Hidetoshi
Takahashi Tsunero
Takenouchi Osamu
Takenouchi Tadao
Takeuchi Kei
Takeuti Gaisi
Takizawa Seizi
Tamagawa Tsuneo
Tamura Itiro
Tamura Jirô
Tanaka Chuji
Tanaka Hiroshi
Tanaka Hisao
Tanaka Minoru
Tanaka Sen-ichiro
Tannaka Tadao
Tatsumi Tomomasa
Tatuzawa Tikao
Terasaka Hidetaka
Toda Hideo
Toda Hirosi
Toda Morikazu
Tôki Yukinari
Tomotika Susumu
Totoki Haruo
Toyama Hiraku
Tsuboi Chuji
Tsuchikura Tamotsu
Tsuji Masatsugu
Tsujioka Kunio
Tsukamoto Yôtarô
Tsuzuku Toshiro
Tugué Tosiyuki
Tumura Yosiro
Uchiyama Saburo
Uchiyama Tatsuo
Udagawa Masatomo
Ueno Kenji
Ueno Tadashi
Ugaeri Tadashi
Umegaki Hisaharu
Umezawa Hiroomi
Umezawa Toshio
Uno Toshio
Ura Shoji
Ura Taro
Urabe Minoru
Wada Yasaku
Washimi Shinichi
Washio Yasutoshi
Washizu Kyuichiro
Watanabe Hiroshi
Watanabe Hisao
Watanabe Masaru
Watanabe Shinzo
Yamada Isamu
Yamaguti Masaya
Yamamoto Koichi
Yamanaka Takesi
Yamanoshita Tsuneyo
Yamanouti Takahiko
Yamashita Hideo
Yamauti Ziro
Yamazaki Keijiro
Yanagihara Kichiji
Yano Kentaro
Yano Shigeki
Yoneda Nobuo
Yoshizawa Hisaaki
Yosida Kôsaku
Yosida Setuzô
Yosida Yôiti

Translators (for the First Edition)

Adachi Masahisa
Aizawa Sadakazu
Akizuki Yasuo
Akô Kiyoshi
Araki Shôrô
Arima Akito
Asano Keizo
Azumaya Goro
Chiba Katsuhiro
Fujisawa Takehisa
Fujita Hiroshi
Fukada Ichiro
Fukuyama Masaru
Furuya Shigeru
Hashimoto Yasuko
Hasimoto Hidenori
Hasumi Morisuke
Hattori Akio
Hattori Akira
Hayashi Kazumichi
Hida Takeyuki
Hinata Shigeru
Hirasawa Yoshikazu
Hirose Ken
Hitotumatu Sin
Honda Taira
Hong Imsik
Hosoi Tsutomu
Hukuhara Masuo
Ihara Shin-ichiro
Ihara Yasutaka
Ikeda Mineo
Ikeda Nobuyuki
Imai Isao
Inoue Masao
Iri Masao
Ishii Goro
Itô Seizô
Ito Yuji
Iwahashi Ryōsuke
Iwahori Nagayoshi
Iwarmura Tsurane
Iwano Masahiro
Iwata Giichi
Iyanaga Kenichi
Iyanaga Shôkichi
Kametani Shunji
Kasahara Koji
Kato Junji
Kawada Yukiyosi
Kawano Sanehiko
Kimura Tosihusa
Kitagawa Toshio
Kobori Akira
Kodaira Kunihiko
Komatsu Hikosaburo
Komatu Atuo
Komatu Yûsaku
Kondo Ryōji
Kondo Takeshi
Kozai Yoshihide
Kubo Ryogo
Kubota Tomio
Kudo Akio
Kudo Hirokichi
Kumano-go Hitoshi
Kunisawa Kiyonori
Kunita Hiroshi
Kurita Minoru
Kuroda Shige-Toshi
Kusama Tokitake
Kusano Takaŝi
Maruyama Gisiro
Matsuda Michihiko
Matsumoto Kikuji
Matsumoto Makoto
Matsumura Hideyuki
Matsushima Yozo
Mimura Yukio
Miyazawa Hironari
Mizohata Sigeru
Mori Toshio
Morimoto Haruki
Morimura Hidenori
Morita Kiiti
Morita Yasuo
Motoo Minoru
Murakami Shingo
Muramatu Tosinobu
Murase Ichiro
Nagao Hirosi
Nagasawa Masao
Nagata Masayoshi
Nakai Mitsuru
Nakai Yoshikazu
Nakamura Tokushi
Nakanishi Shizu
Nakano Shigeo
Nakaoka Minoru
Namikawa Yukihiko
Namioka Isaac
Nanba Kanji
Niiro Fumio
Nikaido Hukukane
Nishi Mieo
Nishijima Kazuhiko
Nishimiya Han
Noguchi Hiroshi
Nonaka Toshio
Noshiro Kiyoshi
Nozaki Akihiro
Ôbayashi Tadao
Ogino Shusaku
Ohtsuka Makoto
Oikawa Kotaro
Okamoto Masashi
Okubo Kenjiro
Okugawa Kōtaro

Ono Katuzi
Ono Shigeru
Osima Masaru
Ozawa Mitsuru
Ozeki Hideki
Saito Tosiya
Sakai Shoichiro
Sakamoto Minoru
Sasaki Shigeo
Satake Ichiro
Sato Kenkichi
Sawashima Ikuko
Shibuya Masaaki
Shiga Kôji
Shimada Nobuo
Shintani Takuro
Shioda Tetsuji
Shiraiwa Kenichi
Shiratani Katsumi
Shizuma Ryoji
Simauti Takakazu
Sugawara Masahiro
Sugiura Mitsuo
Sugiyama Shohei
Suita Nobuyuki
Sumitomo Takeshi
Sunouchi Genichiro
Suzuki Michio
Takahashi Moto-o
Takahashi Reiji
Takahashi Tsunero
Takami Hideo
Takaoka Seiki
Takenouchi Osamu
Takeuchi Kei
Takeuti Gaisi
Tamura Itiro
Tanaka Hiroshi
Tanaka Hisao
Tanaka Minoru
Tatsumi Tomomasa
Tatuzawa Tikao
Terasaka Hidetaka
Toda Hideo
Toda Hirosi
Tsuka-da Nobutaka
Tsukamoto Yôtarô
Tugué Tosiyuki
Uchiyama Saburo
Uesu Tadahiro
Umegaki Hisaharu
Umezawa Toshio
Uno Toshio
Ura Shoji
Ura Taro
Urabe Minoru
Uzawa Hirobumi
Wada Junzo
Watanabe Hiroshi
Watanabe Hisao
Watanabe Shinzo
Watanabe Takesi
Yamaguti Masaya
Yamamoto Koichi
Yamanaka Takesi
Yamazaki Keijiro
Yano Kentaro
Yano Shigeki
Yoshizawa Taro
Yosida Kôsaku
Yosida Setuzô

Name Index

Note. Citation is to article and section, not to page. r = references; STR = Statistical Tables for Reference (to be found after App. B); NTR = Numerical Tables for Reference. * means that the article is not divided into sections.

A

B

C

E

H

I

J

K

M

Q

R

S

T

V

W

Encyclopedic
Dictionary
of
Mathematics

Second Edition

Z

Encyclopedic Dictionary of Mathematics

Second Edition

by the
Mathematical Society
of Japan

edited by
Kiyosi Itô

Volume II
O–Z
Appendices and Indexes

The MIT Press
Cambridge, Massachusetts,
and London, England

Subject Index

Note: Citation is to article and section or to a table in an appendix, not to page.

First MIT Press paperback edition, 1993

Originally published in Japanese in 1954 by Iwanami Shoten, Publishers, Tokyo, under the title *Iwanami Sūgaku Ziten*.

This book was set in Monolaser Times Roman by Asco Trade Typesetting Ltd., Hong Kong, and printed and bound by Arcata Graphics, Kingsport in the United States of America.

Library of Congress Cataloging-in-Publication Data

Iwanami sūgaku jiten (ziten). English.
Encyclopedic dictionary of mathematics.
Translation of: Iwanami sūgaku jiten.
Includes bibliographies and indexes.
1. Mathematics—Dictionaries. I. Itô, Kiyosi, 1915– . II. Nihon Sūgakkai. III. Title.
QA5.I8313 1986 510'.3'21 86-21092
ISBN 0-262-09026-0 (HB), 0-262-59020-4 (PB)

C

D

E

G

H

I

K

L

N

O

P

Q

R

S

U

W